

Generative Engine Optimization en repositorios digitales: optimización de la visibilidad para la IA generativa

Generative Engine Optimization in digital repositories: optimizing visibility for generative AI

Danilo Reyes-Lillo; Alejandro Morales-Vargas; Cristòfol Rovira

Cómo citar este artículo:

Reyes-Lillo, Danilo; Morales-Vargas, Alejandro; Rovira, Cristòfol (2025). "Generative Engine Optimization en repositorios digitales: optimización de la visibilidad para la IA generativa [Generative Engine Optimization in digital repositories: optimizing visibility for generative AI]". *Infonomy*, 3(5) e25034.
<https://doi.org/10.3145/infonomy.25.034>

Artículo recibido: 15-10-2025
Artículo aprobado: 11-11-2025



Danilo Reyes-Lillo

<https://orcid.org/0000-0002-0141-8324>

<https://directorioexit.info/ficha4752>

Universitat Pompeu Fabra

Departament de Comunicació

Roc Boronat, 138

08018 Barcelona, España

daniloandres.reyes01@estudiant.upf.edu





Alejandro Morales-Vargas

<https://orcid.org/0000-0002-5681-8683>

<https://directorioexit.info/ficha5609>

Universidad de Chile

Facultad de Comunicación e Imagen, y

Dirección de Servicios de Información y Bibliotecas (SISIB)

Capitán Ignacio Carrera Pinto, 1045

7800020 Ñuñoa, Chile

amorales@uchile.cl



Cristòfol Rovira

<https://orcid.org/0000-0002-6463-3216>

<https://directorioexit.info/ficha95>

Universitat Pompeu Fabra

Departament de Comunicació

Roc Boronat, 138

08018 Barcelona, España

cristofol.rovira@upf.edu

Resumen

El auge exponencial de la Inteligencia Artificial (IA) generativa y los modelos de lenguaje a gran escala (LLMs) ha redefinido la búsqueda y recuperación de información en la web, migrando del enfoque tradicional de coincidencias de palabras clave y listas de resultados con enlaces a sitios, a la producción de respuestas sintetizadas y conversacionales. Esta transformación está representando un desafío crítico para el tráfico de los repositorios digitales académicos y la *Generative Engine Optimization* (GEO) surge como una posible respuesta. Las estrategias de optimización SEO actuales resultan insuficientes para que el contenido sea referenciado de manera efectiva por los nuevos motores generativos. El objetivo de este artículo es proponer los primeros lineamientos estratégicos de GEO aplicables a repositorios institucionales para maximizar su visibilidad. La metodología establece un marco teórico que contrasta los enfoques de búsqueda y extrapola las estrategias GEO de la industria a la gestión de repositorios. Se identifican y proponen estrategias como la optimización técnica, la estructura semántica, la calidad de los metadatos, la optimización del contenido, la interoperabilidad, el multilingüismo y la exploración de nuevas métricas. Como conclusión, se plantea que la adopción equilibrada de la GEO puede fortalecer el impacto de los repositorios sin desatender su misión de servicio público centrado en las personas usuarias y de acceso abierto a la producción científica en la era de la IA.

Palabras clave

GEO; Generative Engine Optimization; Repositorios digitales; Motores generativos; IA generativa; SEO académico; Repositorios académicos; Optimización; Visibilidad; Inteligencia artificial generativa; LLM; Reference Rate.

Abstract

The rapid growth of generative Artificial Intelligence (AI) and large-scale language models (LLMs) has transformed the way information is searched and retrieved on the web. It has shifted from traditional keyword matching and list-based results with site links to generating synthesized, conversational responses. This change presents a significant challenge to the traffic of academic digital repositories. In response, Generative Engine Optimization (GEO) emerges as a potential solution. Current SEO strategies are inadequate for ensuring that new generative engines effectively reference content. This article aims to present the first GEO strategic guidelines for institutional repositories to enhance their visibility. The methodology develops a theoretical framework that compares different search methods and adapts industry GEO strategies to repository management. Key strategies, including technical optimization, semantic structuring, metadata quality, content enhancement, interoperability, multilingual support, and new metric exploration, are identified and proposed. Ultimately, it is argued that a balanced adoption of GEO can increase the impact of repositories while maintaining their mission as user-focused public services that provide open access to scientific knowledge in the AI era.

Keywords

GEO; Generative Engine Optimization; Digital repositories; Generative engines; Generative AI; Academic SEO; Academic repositories; Optimization; Visibility; Generative artificial intelligence; LLM; Reference Rate.

1. Introducción

En los últimos tres años, la inteligencia artificial (IA) generativa ha experimentado un crecimiento exponencial, transformándose en una tecnología disruptiva, con aplicaciones que abarcan desde la creación de contenido hasta la asistencia en tareas complejas de análisis y toma de decisiones (**Bommasani et al.**, 2022; **Gozalo-Brizuela; Garrido-Merchán**, 2023). Este auge no solo redefine la manera en que producimos y consumimos información, sino que también plantea un cambio profundo en los procesos de búsqueda y recuperación de información. Tradicionalmente, los motores de búsqueda se han basado en algoritmos de indexación y relevancia; sin embargo, la irrupción de modelos generativos y sistemas conversacionales introduce un nuevo paradigma en el que la formulación de respuestas se combina con la recuperación de información, dando lugar a enfoques híbridos como la generación aumentada por recuperación o *Retrieval-Augmented Generation* (**Lewis et al.**, 2021).

Este cambio implica oportunidades y desafíos. Por un lado, la IA generativa promete ofrecer respuestas más contextuales y personalizadas, reduciendo la fricción en la interacción con grandes volúmenes de datos (**Hersh**, 2024). Por otro, surgen interrogantes sobre la precisión, la transparencia y la dependencia de modelos que, en ocasiones, pueden priorizar la fluidez sobre la veracidad (**Zhang**, 2025).

En este contexto, el tráfico y posicionamiento de los sitios web también se han visto afectados a raíz de las búsquedas impulsadas por herramientas de IA como *ChatGPT*, *Copilot*, *Gemini* o *Perplexity*, entre otras. De acuerdo con un estudio elaborado por **Handley** (2025a), se prevé que el tráfico proveniente de estas herramientas superará al tráfico orgánico tradicional en 2028, o incluso antes si *Google* adopta por defecto su “modo IA”.

Esta idea es reforzada por **McKenzie y Ofei** (2025), quienes señalan que el tráfico desde modelos de lenguaje a gran escala o *Large Language Models (LLMs)* crecerá exponencialmente y superará a *Google Search* en 2027. Sus datos muestran que las referencias desde *LLMs* han aumentado en un 800% y los clics desde *Google* disminuyen.

En este marco, surge el concepto de *Generative Engine Optimization (GEO)* u optimización para motores generativos. **Aggarwal et al.** (2024) se refieren a este término como un marco novedoso diseñado para ayudar a los creadores de contenido a optimizar la visibilidad de sus sitios web en las respuestas generadas por los *Generative Engines (GEs)* o motores generativos, los cuales combinan motores de búsqueda con modelos de lenguaje a gran escala o *Large Language Models (LLMs)* para sintetizar y resumir información. Los autores profundizan en este problema, señalando que los *GEs* proporcionan respuestas directas y sintetizadas a las consultas de los usuarios, reduciendo la necesidad de que estos visiten sitios web, generando un impacto negativo en el tráfico y la visibilidad de las páginas. Por este motivo, los métodos tradicionales para la optimización para motores de búsqueda o *Search Engine Optimization (SEO)* son ineficaces en el ámbito de los *GEs*, debido a que estos dependen de los modelos generativos en lugar de la coincidencia de palabras clave.

Debido a lo anterior, las empresas y agencias se están adaptando a la búsqueda impulsada por *LLMs* mediante la transición del SEO tradicional hacia estrategias *GEO*. **Daniels** (2025) señala que las agencias reconocen que los modelos de lenguaje a gran escala sintetizan respuestas en lugar de simplemente clasificar enlaces, por lo que están tomando las siguientes iniciativas:

La IA generativa promete ofrecer respuestas más contextuales y personalizadas, reduciendo la fricción en la interacción con grandes volúmenes de datos, pero surgen interrogantes sobre la precisión, la transparencia y la dependencia de modelos que, en ocasiones, pueden priorizar la fluidez sobre la veracidad

- Desarrollar marcos *GEO* que optimizan el contenido para mejorar la visibilidad en respuestas generativas, enfocándose en factores como la relevancia de las citas, la autoridad y la claridad, en lugar de la densidad de palabras clave.
- Experimentar estrategias de contenido, como añadir citas, estadísticas y un lenguaje autoritativo, que transmita confianza, legitimidad y pericia, para aumentar la probabilidad de ser referenciados por motores generativos.
- Adoptar enfoques integrados y basados en datos que combinan estrategia empresarial, contenido digital y experiencia de usuario para mantener la visibilidad de la marca en plataformas conversacionales.

Aunque en el ámbito empresarial y en las agencias ya se observa una respuesta activa frente a este fenómeno, en el campo de la comunicación científica, y particularmente en los repositorios digitales, apenas comienza a evaluarse el impacto que los *LLMs* tienen sobre el uso y la visibilidad de sus contenidos. Por ejemplo, el informe de la *Confederation of Open Access Repositories (COAR)* señala que la mayoría de los repositorios se están enfrentando a visitas frecuentes de rastreadores o *crawlers* agresivos, comúnmente identificados como *bots* de inteligencia artificial que recolectan grandes volúmenes de contenido para entrenar sistemas generativos (*COAR*, 2025). Esta actividad está generando interrupciones y sobrecargas en los servicios y obliga a muchos repositorios a aplicar medidas de protección que, en algunos casos, también dificultan el acceso legítimo de usuarios humanos o aplicaciones confiables. Ante este panorama, *COAR* lanzó en julio de 2025 el *Grupo de Trabajo sobre Bots de IA y Repositorios*, integrado por especialistas técnicos y representantes de repositorios, con el objetivo de documentar la magnitud del problema, recopilar estrategias de mitigación y formular recomendaciones que permitan proteger la infraestructura sin restringir el acceso abierto y legítimo a la información.

Sin embargo, como se ha señalado previamente, esto no solo se trata de un cambio tecnológico, sino también de un cambio cultural en cómo se accede a la información. Por lo tanto, este artículo tiene como objetivo principal proponer los primeros lineamientos para la optimización de repositorios ante motores de búsqueda generativos.

En primer lugar, se presenta un marco teórico que sirve como referencia conceptual para abordar el fenómeno. Posteriormente, se analizan las principales estrategias *GEO* actualmente aplicadas por expertos, para luego analizar aquellas estrategias aplicables al contexto de los repositorios digitales. Finalmente, se abordan los principales desafíos y controversias que se desprenden de este nuevo paradigma.

2. Marco teórico

La forma en que las personas buscan información ha cambiado de manera significativa en las últimas tres décadas, lo que ha exigido una constante adaptación por parte de quienes desean optimizar la visibilidad de sus contenidos. Si bien el SEO tradicional ha desempeñado un papel fundamental, los avances tecnológicos han obligado a profesionales e investigadores a replantear sus estrategias (*Oxford College of Marketing*, 2025).

La Figura 1 ilustra la evolución de los motores de búsqueda desde la década de 1990. En sus inicios, los primeros motores de búsqueda, como *AltaVista* y *Yahoo*, se basaban principalmente en la coincidencia de palabras clave. Posteriormente, en los años 2000, *Google* revolucionó el sector con algoritmos como *PageRank*, que otorgaron mayor relevancia a los enlaces entrantes o *backlinks* y a la calidad del contenido.

Durante la década siguiente, surgieron la búsqueda semántica y la búsqueda por voz, impulsadas por tecnologías como el algoritmo *Hummingbird* de *Google*. Esto trajo consigo la necesidad de implementar estrategias basadas en la web semántica, datos enlazados, optimización para dispositivos móviles y adaptación a consultas por voz.

Finalmente, en 2022, el lanzamiento de *ChatGPT*, seguido un año después por la Experiencia de Búsqueda Generativa de *Google*, marcaron un cambio radical hacia sistemas basados en inteligencia artificial, orientados a ofrecer respuestas conversacionales y síntesis de información.



Figura 1. Evolución de los motores de búsqueda

La transición tecnológica desde los motores de búsqueda tradicionales hacia la era de la inteligencia artificial generativa puede analizarse desde múltiples perspectivas. En primer lugar, mientras los primeros se basan en listas jerarquizadas de enlaces, los motores de IA sintetizan respuestas narrativas respaldadas por citas, lo que transforma la búsqueda de palabras clave en interacciones conversacionales basadas en la intención del usuario. Además, los motores de IA muestran una fuerte preferencia por fuentes de medios ganados o *earned media* –contenidos publicados orgánicamente por terceros–, excluyendo en muchos casos contenido de marcas o redes sociales, a diferencia de *Google*, que mantiene un equilibrio más amplio. Este cambio también ha modificado el comportamiento de las personas usuarias, reduciendo los

clicks externos y generando resultados sin interacción adicional ("zero-click"). Asimismo, se observan diferencias en la diversidad de dominios, la frescura del contenido y la sensibilidad al lenguaje de la consulta (Chen et al., 2025).

En este contexto, Amer y Elboghdayly (2024) identifican varios cambios clave que distinguen a ambos enfoques:

- **Recuperación de información:** los motores de búsqueda tradicionales se basan en la indexación y el posicionamiento de contenido web, ofreciendo a los usuarios una lista de enlaces que deben explorar para encontrar la información deseada. En contraste, la IA generativa proporciona respuestas directas y contextualizadas, eliminando la necesidad de revisar múltiples páginas.
- **Interacción conversacional:** mientras que los motores de búsqueda están diseñados para consultas puntuales, lo que obliga a las personas usuarias a ajustar manualmente sus términos, la IA generativa permite diálogos continuos. Esto facilita la reformulación de preguntas y la solicitud de aclaraciones en tiempo real.
- **Síntesis de información:** los motores tradicionales presentan documentos o enlaces separados, lo que requiere que el usuario integre la información por su cuenta. Por el contrario, la IA generativa combina contenido de diversas fuentes en respuestas coherentes y precisas, reduciendo la carga cognitiva.
- **Personalización:** aunque los motores de búsqueda pueden ofrecer resultados personalizados basados en factores como la geolocalización, la IA generativa va más allá, adaptando sus respuestas en función del historial de interacción con el usuario, lo que permite una experiencia más personalizada y dinámica.
- **Gestión de consultas complejas:** las búsquedas tradicionales suelen enfrentar dificultades ante preguntas matizadas o complejas, lo que obliga al usuario a formular cuidadosamente sus términos. En cambio, la IA generativa es capaz de abordar este tipo de consultas ofreciendo explicaciones detalladas y comparaciones enriquecidas.

En ese mismo marco, Kaiser et al. (2025) aplicaron un diseño experimental a gran escala para comparar el comportamiento y las preferencias de los usuarios al realizar tareas de búsqueda en línea utilizando *ChatGPT* –IA generativa– o *Google* –motor de búsqueda tradicional–. Este trabajo muestra que los usuarios de *ChatGPT* fueron más rápidos y precisos en la resolución de tareas de búsqueda, tanto tecnológicas como no tecnológicas, en comparación con los usuarios de *Google*. Además, interactuaron mucho menos con sitios web externos, lo que sugiere que perciben a *ChatGPT* como una solución integral. Sin embargo, a pesar de su mejor rendimiento, los participantes mostraron una preferencia subjetiva mayor por *Google*, cuya aceptación fue más uniforme entre distintos perfiles de personalidad. En contraste, la adopción de *ChatGPT* estuvo influenciada por rasgos individuales como la extroversión y el neuroticismo. Estas diferencias tienen implicaciones conductuales importantes, ya que el menor

uso de fuentes externas podría afectar el tráfico web y aumentar el riesgo de desinformación por errores o sesgos de la IA.

Aunque las diferencias entre los motores de búsqueda tradicionales y los motores de respuesta basados en IA generativa son evidentes, estos últimos tienden a reproducir los sesgos presentes en las consultas, lo que puede derivar en respuestas unilaterales y limitar la diversidad de perspectivas. Además, su funcionamiento poco transparente reduce la autonomía del usuario y puede afectar la confianza en la información ofrecida. Se han identificado problemas recurrentes en la IA generativa, como alucinaciones, citas incorrectas y sesgos, lo que subraya la necesidad de garantizar mayor transparencia, equilibrio informativo y control por parte del usuario en los sistemas de recuperación de información (**Venkit et al.**, 2024).

En este contexto, las estrategias tradicionales de SEO no son suficientes para estos nuevos entornos, dando lugar a técnicas de *GEO*, centradas en contenido legible por máquinas, autoridad de medios ganados y estrategias orientadas al ciclo de vida de la información (**Chen et al.**, 2025).

En el ámbito académico, el escenario no es diferente y la IA generativa está influyendo en la búsqueda de información científica. Herramientas basadas en modelos de lenguaje, como *ChatGPT* y aplicaciones especializadas como *Elicit*, han sustituido las búsquedas tradicionales por consultas en lenguaje natural, ofreciendo respuestas contextualizadas y análisis automatizados de documentos (**Arroyo-Machado**, 2024; **Font-Julián et al.**, 2024). Estas plataformas permiten realizar búsquedas semánticas, generar síntesis narrativas y extraer datos relevantes de grandes volúmenes de literatura, agilizando procesos como la revisión bibliográfica y la comparación metodológica.

Además, de acuerdo con **Buitrago-Ciro et al.** (2025), la IA generativa se ha integrado en la comunicación científica para tareas como la redacción, traducción y edición de artículos, así como para acelerar la revisión por pares. Sin embargo, este avance plantea desafíos éticos y técnicos: sesgos algorítmicos, plagio, referencias ficticias y falta de políticas globalmente aceptadas. También se observa una creciente competencia por la visibilidad en motores de búsqueda impulsados por IA, lo que podría modificar el acceso a la literatura académica. En ese sentido, autores como **Font-Julián et al.** (2024) señalan que *ChatGPT* y herramientas similares pueden ser fuertes competidores para buscadores académicos como *Google Scholar*.

Con este escenario, los repositorios digitales como fuentes de información académica también se ven afectados. La optimización de su visibilidad es clave para maximizar el uso y el impacto de la producción científica, al hacer los trabajos accesibles y detectables por una audiencia global (**Tapfuma; Hoskins**, 2022). Esto fortalece la reputación institucional, incluso llevando a la posibilidad de incrementar las citas a aquellos trabajos almacenados en repositorios (**Wheeler et al.**, 2022). Además, contribuye a la democratización del conocimiento, mejora la recuperación de información, asegura un acceso sostenible a largo plazo y facilita la difusión efectiva del contenido académico más allá de la comunidad científica, dado que su infraestructura técnica tributa a estos objetivos (**Farnel**, 2023; **Zavalina; Burke**, 2023).

Sin embargo, las estrategias de optimización de visibilidad actuales no están focalizadas en optimizar los resultados que provienen de consultas realizadas mediante herramientas de IA generativa. Dentro de las técnicas recomendadas, podemos identificar la indexación de contenido en agregadores, optimización de metadatos, uso de identificadores persistentes, activación de protocolos de interoperabilidad, integración de tecnologías de datos abiertos enlazados –*linked open data*– u optimización SEO de los repositorios (**Reyes-Lillo et al.**, 2025a). Si bien varios de estos elementos pueden tributar a una mejor recuperación por parte de los motores de IA generativa, no se identifican estrategias que apunten plenamente a abordar este nuevo desafío para los repositorios digitales.

Es en este marco, donde una aproximación hacia el *GEO* en repositorios digitales se transforma en una herramienta adicional para que los gestores de repositorios fortalezcan la visibilidad de sus contenidos en un nuevo contexto de búsquedas de información.

3. Estrategias recurrentes de GEO aplicadas por expertos

De acuerdo con lo planteado, *Generative Engine Optimization* (*GEO*) se define como el conjunto de técnicas orientadas a optimizar contenido para que sea citado, referenciado o integrado directamente en las respuestas generadas por modelos de lenguaje de gran escala (*LLMs*) (**Aggarwal et al.**, 2024). A diferencia del SEO tradicional, que busca posicionar contenido en los resultados de búsqueda, el objetivo de *GEO* es formar parte de la respuesta que ofrece una IA. Al analizar las estrategias más utilizadas actualmente por expertos en *GEO*, se pueden identificar las siguientes:

A diferencia del SEO tradicional, que busca posicionar contenido en los resultados de búsqueda, el objetivo de *GEO* es formar parte de la respuesta que ofrece una IA

3.1. Uso de estructura semántica clara

Una estructura semántica clara aporta a que los modelos de inteligencia artificial comprendan el contenido de una web y lo integren en sus respuestas generadas. El uso de Schema.org combinado con formatos como JSON-LD permite etiquetar el contenido de manera que sea legible tanto para humanos como para máquinas, facilitando su interpretación por modelos como *ChatGPT*, *Gemini* o *Perplexity*. Por ejemplo, el tipo de esquema *FAQPage* ha demostrado ser especialmente eficaz, ya que refleja el formato conversacional que los usuarios emplean al interactuar con asistentes de IA (**Sain**, 2025). Esto no solo mejora la visibilidad en motores de búsqueda tradicionales, sino que también aumenta la probabilidad de que el contenido sea citado directamente por herramientas de IA en sus respuestas.

Según expertos en SEO semántico, JSON-LD es el formato preferido por *Google* y *Bing* para representar datos estructurados y su implementación puede mejorar significativamente la comprensión del contenido por parte de los modelos de lenguaje (**Park**,

2025; **Yan**, 2025). Además, indican que los sitios que utilizan correctamente estos esquemas tienen mayor tasa de aparición en resúmenes generados por IA, lo que refuerza su autoridad y relevancia en entornos digitales.

3.2. Contenido con profundidad explicativa

En el contexto de la optimización para motores de inteligencia artificial, la profundidad explicativa del contenido es un elemento determinante para alcanzar visibilidad en las respuestas generadas por IA generativa. En esta nueva etapa de indexación semántica, no resulta suficiente ofrecer respuestas superficiales o meramente descriptivas: los modelos priorizan textos capaces de anticipar interrogantes derivadas, aportar contexto interpretativo y articular diversas perspectivas analíticas sobre un mismo tema.

El marco de trabajo **VISIBLE** (*SocialChamps*, 2025) señala que los sistemas generativos tienden a favorecer contenidos estructurados, densos en datos y caracterizados por una red coherente de vínculos semánticos, atributos que facilitan la segmentación y extracción de fragmentos relevantes para la elaboración de respuestas fiables. En la misma línea, **Shelby** (2025) advierte que los modelos de lenguaje no se basan exclusivamente en metadatos, sino que analizan la jerarquía conceptual, la coherencia narrativa y los patrones de redundancia útil como indicadores de calidad y relevancia informativa.

Esta orientación es reafirmada por *AppLabx* (2025), quienes recomiendan estructurar el contenido en tres niveles: uno superior, destinado a ofrecer respuestas directas y sintéticas; un nivel intermedio, orientado al desarrollo semántico y contextual; y un nivel inferior, dedicado a la documentación técnica y a las referencias de apoyo. De este modo, la profundidad conceptual y la arquitectura semántica del texto resultan tan decisivas como la adecuación lingüística.

3.3. Optimización para lenguaje conversacional

La optimización para lenguaje conversacional se ha consolidado como una estrategia central dentro del marco de *GEO*, en tanto que los motores de IA generativa procesan consultas que presentan una mayor longitud, naturalidad y carga contextual en comparación con las búsquedas formuladas en motores tradicionales. Según datos de **Kelly** (2025), las consultas dirigidas a herramientas de inteligencia artificial poseen una longitud media de 23 palabras, lo que evidencia una intención de búsqueda más compleja, interpretativa y orientada al diálogo.

Para que el contenido sea reconocido y priorizado por estos modelos, debe reproducir la estructura discursiva propia de una conversación. Esto implica el uso de encabezados formulados como preguntas, párrafos concisos, lenguaje claro y directo, así como transiciones semánticas fluidas que simulen la lógica de un intercambio comunicativo.

Como se indicó previamente, **Shelby** (2025) destaca que los modelos de lenguaje no se limitan al análisis de palabras clave, sino que evalúan la jerarquía semántica, la coherencia narrativa y la densidad conceptual del texto para determinar su potencial de integración en una respuesta generada. Asimismo, observa que la adopción de un

formato conversacional no solo mejora la legibilidad para los usuarios humanos, sino que también facilita la segmentación y extracción de fragmentos por parte de los sistemas generativos, aumentando así la probabilidad de que el contenido sea citado o incorporado en resúmenes producidos mediante inteligencia artificial.

3.4. Accesibilidad técnica para rastreadores de IA

La accesibilidad técnica para *crawlers* de inteligencia artificial resulta fundamental como estrategia *GEO*, dado que determina la capacidad de los modelos de lenguaje de gran escala para acceder, interpretar y utilizar el contenido de un sitio web.

Tradicionalmente, el fichero *robots.txt* ha sido empleado para regular el acceso de rastreadores de motores de búsqueda convencionales, como *GoogleBot*. Sin embargo, en la actualidad también cumple un papel crítico en la gestión del comportamiento de *bots* de IA, tales como *GPTBot*, *Claude-Web* o *CCBot*. La inclusión de directivas explícitas permite garantizar que el contenido sea indexable y utilizable por los sistemas generativos, incrementando así la probabilidad de que sea citado o referenciado en respuestas automatizadas. Un ejemplo en *robots.txt* sería:

```
User-agent: GPTBot
Allow: /
```

Adicionalmente, ha surgido un nuevo estándar denominado *llms.txt*, diseñado para orientar a los modelos de lenguaje respecto a qué secciones del sitio son relevantes para la inferencia o el entrenamiento. Este archivo funciona como un “menú curado” que guía a los *crawlers* hacia los fragmentos más valiosos del contenido. Según **Pol** (2025), la implementación de *llms.txt* puede mejorar la visibilidad en respuestas generadas por IA, al minimizar la ambigüedad y el ruido durante el proceso de rastreo. Si bien, la propia empresa reconoce que no es un estándar establecido, ya han empezado a experimentar con esta estrategia.

3.5. Velocidad y estructura técnica

La velocidad de carga y la estructura técnica de un sitio web constituyen factores determinantes para que los modelos de inteligencia artificial puedan acceder, interpretar y citar su contenido en respuestas generadas. Según *INSIDEA* (2025), si una página se ralentiza, los *crawlers* de IA, como *GPTBot* o *Claude-Web*, tienden a interrumpir el proceso de indexación, lo que reduce significativamente la visibilidad del sitio. Este comportamiento se explica por la existencia de un presupuesto de rastreo –*crawl budget*– limitado, que lleva a los modelos a priorizar páginas que ofrezcan respuestas rápidas y eficientes.

Para optimizar la accesibilidad técnica, se recomienda la implementación de *Server-Side Rendering* (SSR), que permite entregar contenido HTML completo directamente desde el servidor, facilitando su procesamiento por los *crawlers* de IA (**Reyes-Lillo et al.**, 2025b). De manera complementaria, se pueden emplear herramientas de compresión para reducir el tamaño de los archivos, lo que contribuye a acelerar la transferencia de datos y a mejorar los tiempos de carga.

Adicionalmente, la utilización de HTML semántico, mediante etiquetas como `<article>`, `<section>` o `<header>`, proporciona una jerarquía estructural clara que los modelos de lenguaje pueden interpretar para identificar y extraer de manera más eficiente la información relevante contenida en la página. En conjunto, estas prácticas técnicas no solo aumentan la probabilidad de que el contenido sea indexado, sino que también favorecen su priorización y citación en respuestas generadas por sistemas de IA.

3.6. Actualización de contenido

La actualización frecuente del contenido ha sido identificada como una práctica relevante para optimizar la visibilidad en motores de inteligencia artificial. Si bien existe cierto contenido que conserva su valor por su relevancia duradera, los modelos de lenguaje tienden a priorizar fuentes más recientes, incluso cuando estas no son las más exhaustivas.

Los modelos de lenguaje tienden a priorizar fuentes más recientes, incluso cuando estas no son las más exhaustivas

Un estudio de *Ahrefs*, que analizó más de 17 millones de citas en plataformas de IA, reveló que el contenido citado por asistentes generativos, como *ChatGPT*, es en promedio 25,7% más reciente que el contenido que aparece en los resultados orgánicos de *Google* (Law; Guan, 2025). En particular, *ChatGPT* mostró una preferencia marcada por páginas actualizadas recientemente, citando URLs con una antigüedad media de 393 días menor que las de *Google Search*. Estos hallazgos sugieren que, incluso un artículo estructurado y con profundidad conceptual, puede ser ignorado por los modelos de IA si no ha sido actualizado en los últimos años.

Frente a este escenario, se recomienda evitar actualizaciones superficiales, por ejemplo, modificar únicamente la fecha de publicación. En cambio, se recomienda priorizar revisiones sustanciales que reflejen cambios en datos, contexto o recomendaciones. En este paradigma emergente, mantener el contenido fresco y actualizado no solo optimiza el SEO tradicional, sino que también incrementa significativamente la probabilidad de ser citado en respuestas generadas por IA.

3.7. Evaluación de la tasa de referencia

En el contexto de *GEO*, no solo es importante aplicar técnicas de optimización, sino también establecer métricas para evaluar el desempeño del contenido y cómo está siendo indexado por motores de IA. En ese marco, la tasa de referencia o *Reference Rate* se ha consolidado como una métrica clave para evaluar la visibilidad en entornos de búsqueda generativa. A diferencia de las métricas SEO tradicionales, como por ejemplo el *Click-Through Rate* (CTR), que mide cuántos usuarios hacen clic en un enlace, el *Reference Rate* mide cuántas veces el contenido de un sitio es citado, mencionado o referenciado por modelos de IA en sus respuestas (SEO Juice, 2025). De acuerdo con Hall (2025), esta métrica indica el grado de autoridad que los modelos atribuyen a un dominio y se calcula de la siguiente manera:

$$\text{Reference Rate (\%)} = \frac{\text{Número de respuestas que citan tu contenido}}{\text{Total de respuestas relevantes}} \times 100$$

Por ejemplo, si un sitio web sobre migración a la nube es citado en 35 de cada 100 respuestas generadas por IA sobre ese tema, su *Reference Rate* sería del 35%, lo que indica una alta autoridad temática. Esta métrica es especialmente valiosa en el entorno actual, donde los usuarios reciben respuestas completas sin necesidad de visitar el sitio original. Un alto *Reference Rate* no solo mejora la visibilidad de marca, sino que también entrena a los modelos de lenguaje para considerar ese contenido como fuente confiable, lo que refuerza su presencia en futuras respuestas.

El *Reference Rate* mide cuántas veces el contenido de un sitio es citado, mencionado o referenciado por modelos de IA en sus respuestas

Reforzando las ideas previas, para mejorar esta métrica, se recomienda: 1) crear contenido con respuestas densas y autoexplicativas –40-80 palabras–, 2) usar datos estructurados como JSON-LD para mejorar la precisión semántica, y 3) mantener el contenido actualizado y evitar duplicados que diluyan la señal de autoridad.

4. Aplicación de GEO en repositorios digitales

En el siguiente apartado se intenta extrapolar las principales estrategias GEO aplicadas al contexto de los repositorios digitales. Si bien un repositorio digital no tiene la misma estructura ni los mismos objetivos que un sitio web tradicional, hay ciertos elementos que pueden ser adecuados para mejorar el nivel de menciones de un repositorio en aplicaciones de IA generativa. Las estrategias destacadas son las siguientes:

4.1. Optimización técnica

La optimización técnica de un repositorio digital es el primer paso para garantizar su visibilidad, accesibilidad y compatibilidad con motores de búsqueda y herramientas de inteligencia artificial. Esta optimización abarca aspectos de accesibilidad, estructura web, rendimiento y escalabilidad, todos fundamentales para que el contenido pueda ser indexado y referenciado por herramientas de IA generativa. En este caso, desglosamos la optimización técnica en tres puntos principales: 1) cumplimiento de estándares web, 2) uso de *sitemaps* y *robots.txt*, y 3) rendimiento y escalabilidad.

La optimización técnica de un repositorio digital es el primer paso para garantizar su visibilidad, accesibilidad y compatibilidad con motores de búsqueda y herramientas de inteligencia artificial

4.1.1. Cumplimiento de estándares web

El cumplimiento de los estándares técnicos, como las directrices de accesibilidad web definidas en las *Web Content Accessibility Guidelines* (WCAG) 2.1¹, especialmente en el nivel AA, contribuye a garantizar que el contenido del repositorio sea accesible para

¹ <https://www.w3.org/TR/WCAG21/>

todas las personas usuarias, incluidas aquellas con discapacidades visuales, auditivas o cognitivas. Además, estos estándares mejoran la legibilidad semántica del sitio, lo que facilita su interpretación por *crawlers* de IA

De igual modo, el uso de URLs persistentes y amigables o, aún mejor, de identificadores persistentes mejora la claridad semántica y la confiabilidad de las citas generadas por IA. Los identificadores persistentes tienen múltiples roles en el ecosistema de los repositorios digitales (**Reyes-Lillo et al.** 2025b; **Reyes-Lillo; Pastor-Ramon**, 2024), pero en este caso, ayudan a establecer confianza y perdurabilidad al acceso de cierto contenido. En contraste, las URL dinámicas o temporales pueden dificultar la verificación de fuentes por parte de los modelos de lenguaje, lo que reduce la probabilidad de ser referenciado.

En otro aspecto, no menos importante, la implementación de certificados de seguridad HTTPS válidos no solo protege la integridad de los datos, sino que también es un requisito técnico para que muchos *crawlers* de IA accedan al contenido sin restricciones.

4.1.2. Uso de *sitemaps* y *robots.txt*

Los archivos *sitemap.xml* y *robots.txt* son fundamentales para guiar a los rastreadores de IA y buscadores tradicionales. El *sitemap.xml* actúa como un índice estructurado que informa a los *bots* sobre las páginas disponibles, su jerarquía y frecuencia de actualización. Esto permite una indexación más eficiente y completa, especialmente en sitios con gran volumen de contenido. De esta manera, un sitio bien estructurado que transmita la jerarquización de su contenido a través de colecciones o comunidades, puede favorecer el acceso de herramientas de IA.

El archivo *robots.txt*, en tanto, desempeña un papel fundamental en la gestión del acceso y del rastreo automatizado de los recursos de un sitio web. A través de sus directivas, este archivo establece qué secciones pueden o no ser exploradas por los *crawlers*, configurando así un marco de control sobre la visibilidad técnica del contenido. Una configuración apropiada resulta esencial para garantizar el acceso de los agentes de indexación, incluidos los *bots* asociados a modelos de inteligencia artificial generativa, lo que incrementa la probabilidad de que el contenido sea analizado, interpretado y potencialmente incorporado en respuestas generadas por estos sistemas.

Asimismo, la inclusión de la directiva Sitemap: dentro del archivo *robots.txt*, es decir, explicitar dentro de este archivo dónde se encuentra el *sitemap*, constituye una práctica recomendada, ya que facilita el descubrimiento del mapa del sitio por parte de los *crawlers* y optimiza los procesos de indexación.

Finalmente, como se ha señalado, si bien la implementación de un archivo *llms.txt* no es un estándar establecido, algunas empresas y expertos han empezado a incorporarlo como una de las prácticas de GEO. Vale la pena revisar cómo evoluciona esta estrategia para fortalecer el acceso de motores de IA en sitios web.

4.1.3. Rendimiento y escalabilidad

El rendimiento técnico del repositorio influye directamente en su capacidad de ser indexado por IA. Páginas que tardan más de cinco segundos en cargar suelen ser ignoradas por los *crawlers*, lo que afecta negativamente la visibilidad del contenido. Para mejorar esto, se recomienda:

- Implementar *Server-Side Rendering* (SSR) para garantizar que el contenido esté disponible en el HTML inicial (**Reyes-Lillo et al.**, 2025b).
- Usar compresores, lo que reduce el tamaño de los archivos, acelerando la transferencia de datos.
- Aplicar HTML semántico (`<article>`, `<section>`, `<header>`), por ejemplo, en la personalización de las plantillas del repositorio digital, para mejorar la interpretación estructural del contenido por parte de los modelos de lenguaje.

En cuanto a escalabilidad, se recomienda el uso de almacenamiento en la nube con redundancia geográfica y sistemas de caché distribuido para consultas frecuentes, lo que mejora la disponibilidad y reduce la latencia (*Google Cloud*, 2025).

4.2. Estructura semántica del contenido

La implementación de una estructura semántica contribuye sustancialmente a que los modelos de inteligencia artificial comprendan, indexen y referencien el contenido de un repositorio digital. En el contexto de *GEO*, esto implica enriquecer las páginas de cada ítem con metadatos estructurados que describan el tipo de recurso, autores, fechas, licencias y relaciones contextuales.

La implementación de una estructura semántica contribuye sustancialmente a que los modelos de inteligencia artificial comprendan, indexen y referencien el contenido de un repositorio digital

Una de las prácticas más efectivas es el uso de Schema.org mediante el formato JSON-LD, que permite incrustar metadatos legibles por máquinas directamente en el HTML de cada página. Este enfoque no es nuevo, se ha probado que mejora la visibilidad de contenido en repositorios digitales, y es compatible con los principales motores de búsqueda y modelos de IA, y facilita la identificación de tipos de contenido como *ScholarlyArticle*, *CreativeWork*, *Dataset*, *Person* y *Organization* (**Castro et al.**, 2024; **Pekala**, 2018). Por ejemplo, al etiquetar un artículo académico con el esquema *ScholarlyArticle*, se proporciona a los modelos de lenguaje una señal clara sobre la naturaleza del contenido, lo que aumenta la probabilidad de que sea citado en respuestas generadas por IA.

Cuando JSON-LD no es viable por restricciones técnicas del sistema –como en algunos entornos XML o JSP–, se recomienda el uso de RDFa o Microdata, que permiten incrustar anotaciones semánticas directamente en las etiquetas HTML. Estas tecnologías también son compatibles con los estándares de *Linked Data* y mejoran la interoperabilidad con agregadores académicos.

La estructura semántica no solo mejora la visibilidad en motores de búsqueda tradicionales, sino que también facilita la integración del contenido en corpus de entrenamiento y sistemas de recuperación de información utilizados por modelos de IA, lo que refuerza la autoridad digital del repositorio y su impacto académico.

4.3. Calidad y estandarización de metadatos

La calidad y estandarización de los metadatos son uno de los pilares fundamentales para mejorar la visibilidad, interoperabilidad y citabilidad de los contenidos de un repositorio digital en entornos de inteligencia artificial. En el contexto de la GEO, los metadatos no solo permiten que los motores de búsqueda tradicionales indexen correctamente los recursos, sino que también facilitan que las herramientas de IA generativa comprendan, clasifiquen y referencien el contenido de manera precisa.

Para lograr esto, cada ítem del repositorio debe contar con metadatos completos y estructurados, incluyendo al menos: título, autor, fecha de publicación, resumen, palabras clave, idioma y licencia. Estos elementos son esenciales para que los sistemas de IA puedan extraer información relevante y generar respuestas confiables. Además, el uso de vocabularios estandarizados como *Dublin Core*, *DataCite*, *COAR* y las *OpenAIRE Guidelines* garantiza la interoperabilidad entre plataformas y mejora la calidad semántica de los registros (Solatorio; Dupriez, 2024).

Las *OpenAIRE Guidelines*, por ejemplo, recomiendan un perfil de aplicación que combina propiedades de *Dublin Core* y *DataCite* para lograr una mayor granularidad y precisión en la descripción bibliográfica, incluyendo información sobre proyectos financiadores, condiciones de acceso, y relaciones entre publicaciones y conjuntos de datos o *datasets* (OpenAIRE, 2018).

Las *OpenAIRE Guidelines* recomiendan un perfil de aplicación que combina propiedades de *Dublin Core* y *DataCite* para lograr una mayor granularidad y precisión en la descripción bibliográfica

Como se ha señalado previamente, la inclusión de identificadores persistentes como DOI, ORCID, RoR, Handle y ARK es crucial para asegurar la trazabilidad y permanencia de los recursos en el ecosistema digital. Estos identificadores permiten que los modelos de IA reconozcan de forma única a los autores, objetos digitales y organizaciones, lo que mejora la precisión de las citas y la atribución de contenido. Por esta razón, es fundamental incorporarlos en los metadatos.

En otro aspecto, también es recomendable incorporar licencias abiertas en los metadatos para promover la reutilización de los contenidos. Por ejemplo, integrando una licencia *Creative Commons* CC-BY, CC0 o CC-BY-SA en el campo *dc.rights* de *Dublin Core* para explicitar las condiciones de uso de este contenido.

En conjunto, la calidad de los metadatos y su estandarización no solo es fundamental para optimizar un repositorio digital, sino que también aumentan significativamente la probabilidad de que el contenido del repositorio sea citado en respuestas generadas por IA, reforzando su impacto académico y su visibilidad institucional (Reyes-Lillo et al., 2025b).

4.4. Optimización de contenido para la IA

La optimización del contenido para inteligencia artificial es una práctica esencial en el marco de la *GEO*, ya que las herramientas de IA generativa no solo indexan contenido, sino que lo sintetizan, citan y recomiendan en sus respuestas generadas. Para que el contenido de un repositorio digital sea considerado por estas herramientas, debe estar redactado y estructurado de forma que responda directamente a las necesidades de información del usuario, anticipando preguntas frecuentes y ofreciendo claridad semántica.

Una de las técnicas más efectivas en el contexto de los repositorios es redactar descripciones y resúmenes en lenguaje natural, utilizando frases completas y estructura clara. De esta manera, los autores deben estar enfocados en posicionar mejor sus trabajos a través de la redacción eficiente de sus resúmenes. Esto facilita que los modelos de IA comprendan el propósito del contenido y

lo integren en respuestas contextuales. Además, si los repositorios incluyen secciones tipo preguntas frecuentes o FAQ, HowTo y comparativas, se mejora significativamente la visibilidad en motores generativos. Según un estudio de Calla Creative (2025), los sitios que implementan correctamente secciones FAQ aparecen hasta cuatro veces más frecuentemente en resultados generados por IA como *ChatGPT* y *Google AI Overviews*.

Una de las técnicas más efectivas en el contexto de los repositorios es redactar descripciones y resúmenes en lenguaje natural, utilizando frases completas y estructura clara

Los modelos de IA priorizan contenido que responde directamente a preguntas comunes. Este tipo de formato facilita la extracción de respuestas por parte de los sistemas de *Retrieval-Augmented Generation* (RAG), que combinan recuperación de información con generación de texto.

Por último, es fundamental mantener el contenido actualizado, dada la ya expuesta priorización que tienen los motores de IA generativa hacia contenidos nuevos. Por este motivo, los repositorios deben revisar periódicamente sus registros, actualizando fechas, enlaces, licencias y resúmenes para asegurar su relevancia en el ecosistema de búsqueda generativa.

4.5. Interoperabilidad, APIs y agregadores

La interoperabilidad también es un elemento clave para optimizar la referenciación de repositorios por parte de motores de IA. En el contexto de la *GEO*, esto implica garantizar que el repositorio pueda comunicarse eficazmente con sistemas externos, tanto humanos como automatizados, mediante protocolos estandarizados y APIs abiertas.

Uno de los mecanismos más importantes es el protocolo OAI-PMH, que permite a los repositorios exponer sus metadatos de forma estructurada y accesible para servicios que busquen cosechar estos registros. Este protocolo, basado en HTTP y XML, define

seis verbos que permiten listar, identificar y recuperar registros, facilitando la integración con sistemas de descubrimiento y modelos de IA que utilizan metadatos como fuente de entrenamiento y referencia.

Además, la publicación de una API REST, un tipo de interfaz que permite la comunicación entre sistemas, permite el acceso controlado a los contenidos del repositorio, lo que es especialmente útil para desarrolladores, integradores y sistemas de IA que requieren consultas dinámicas o personalizadas. Las APIs REST son altamente escalables, independientes del cliente y utilizan métodos estándar como GET, POST, PUT y DELETE, lo que facilita su adopción en entornos académicos y tecnológicos modernos.

Por último, registrar el repositorio en agregadores académicos como BASE, CORE o Google Scholar es esencial para aumentar su visibilidad en el ecosistema de búsqueda generativa. Estos agregadores no solo amplían el alcance del contenido, sino que también alimentan los corpus utilizados por modelos de lenguaje, lo que incrementa la probabilidad de que los ítems del repositorio sean citados en respuestas generadas por IA.

4.6. Multilingüismo

En el ámbito académico y científico, donde la diversidad lingüística es un elemento estructural del ecosistema del conocimiento, garantizar la disponibilidad de metadatos y contenidos al menos en inglés y en el idioma local permite que motores de IA generativa accedan, comprendan y referencien los recursos con mayor precisión en distintos contextos de consulta.

La disponibilidad de metadatos y contenidos al menos en inglés y en el idioma local permite que motores de IA generativa accedan, comprendan y referencien los recursos con mayor precisión en distintos contextos de consulta

De acuerdo con las recomendaciones de la *COAR Task Force on Supporting Multilingualism and non-English Content in Repositories* (COAR, 2023), la declaración explícita del idioma del recurso y de sus metadatos, el uso de la etiqueta *inLanguage* en estructuras JSON-LD y la adopción de estándares para la codificación lingüística son prácticas que favorecen la indexación, la interoperabilidad semántica y la reutilización de los datos.

Asimismo, estudios recientes, como el de **Dony, Kuchma y Ševkušić** (2024), evidencian que la implementación de metadatos multilingües y la incorporación de traducciones en campos clave como título, resumen y palabras clave incrementan significativamente la capacidad de descubrimiento del contenido de un repositorio, lo que también permite que el contenido sea recuperable por motores de IA generativa.

Por lo tanto, la optimización multilingüe no solo constituye una práctica técnica orientada a la mejora de la visibilidad, sino también una estrategia de equidad epistémica que amplía el alcance global de los repositorios digitales. Al facilitar su integración en

corpus multilingües utilizados por motores de IA generativa, se promueve la generación de respuestas contextualizadas, culturalmente pertinentes y representativas de la diversidad del conocimiento global.

4.7. Nuevas métricas asociadas a IA generativa

Al cambiar las maneras de acceder a la información, también cambia la manera en que se evalúa la visibilidad y cómo los contenidos son consultados. Por esta razón, los gestores de repositorios deben explorar nuevas maneras de monitorear cómo sus contenidos son utilizados en entornos de búsqueda generativa. Si bien herramientas como *Google Search Console* siguen siendo útiles para rastrear la indexación tradicional, en el contexto de *GEO* es necesario incorporar auditorías semánticas que evalúen la calidad, estructura y actualidad de los metadatos.

Por otra parte, el éxito en *GEO* se está midiendo con métricas como el *Reference Rate*, que indica cuántas veces el contenido del repositorio es citado por motores de IA en sus respuestas. Algunas empresas como *Semrush* o *Perplexity* ofrecen herramientas para rastrear menciones, analizar la visibilidad por dominio, y comparar el rendimiento frente a servicios similares en entornos de IA. En esta línea, explorar este tipo de monitoreo permite no solo diagnosticar problemas de visibilidad, sino también ajustar estrategias editoriales, técnicas y semánticas para mejorar la presencia del repositorio en respuestas generadas por IA, lo que refuerza su impacto académico y su relevancia institucional.

5. Discusión y desafíos

Desde otra óptica, la adopción de estrategias *GEO* en repositorios no está exenta de una serie de cuestiones éticas, riesgos técnicos y desigualdades estructurales que deben ser abordadas con una perspectiva crítica.

En primer lugar, la sobreoptimización de contenido para herramientas de IA, es decir, la redacción excesivamente dirigida a satisfacer algoritmos de IA generativa, puede resultar perjudicial para la legibilidad y utilidad para usuarios humanos. Este fenómeno puede generar textos artificiales, repetitivos o carentes de profundidad, afectando la experiencia de búsqueda de un usuario. En otras palabras, la sobreoptimización podría “deshumanizar el contenido”, priorizando el alcance por los algoritmos.

La redacción excesivamente dirigida a satisfacer algoritmos de IA generativa, puede resultar perjudicial para la legibilidad y utilidad para usuarios humanos

Por otra parte, la atribución y citación de fuentes en entornos de IA generativa también supone una controversia. A diferencia de los motores de búsqueda tradicionales, los modelos de lenguaje pueden generar respuestas sin citar explícitamente las fuentes utilizadas, lo que invisibiliza el trabajo de autores e instituciones. Esta práctica plantea dilemas éticos sobre la propiedad intelectual, el reconocimiento académico y la transparencia informativa. **Orduña-Malea y Cabezas-Clavijo (2023)** señalan que también se encuentran referencias bibliográficas fantasma o falsas en las respuestas entregadas por la IA generativa, lo que supone otra perspectiva crítica.

Asimismo, la implementación efectiva de *GEO* requiere de una infraestructura técnica avanzada, personal capacitado y recursos financieros. Sin embargo, muchos repositorios, especialmente en regiones con menor financiación, enfrentan limitaciones estructurales que dificultan la adopción de estas estrategias. Esto puede ampliar la brecha entre instituciones con capacidad para optimizar su visibilidad en entornos de IA y aquellas que quedan excluidas del nuevo ecosistema de descubrimiento digital.

Finalmente, como se ha señalado previamente, el informe de *COAR* (2025) advierte sobre el impacto negativo de los *bots* de IA que acceden masivamente a los repositorios para recolectar datos, provocando caídas del sistema, ralentización del servicio y bloqueo de usuarios humanos. Más del 90% de los repositorios encuestados reportaron interrupciones frecuentes causadas por rastreadores agresivos. Esta situación plantea un dilema: ¿deberían los repositorios limitar el acceso a *bots* para proteger la estabilidad del sistema, aunque eso implique restringir el uso por parte de modelos de IA? La *Confederation of Open Access Repositories* ha lanzado un grupo de trabajo internacional para abordar esta problemática y desarrollar recomendaciones que equilibren apertura, sostenibilidad y protección técnica.

6. Conclusiones

Con la irrupción de los motores de inteligencia artificial generativa, las dinámicas de búsqueda y acceso a la información han experimentado transformaciones significativas, y el ámbito de la información científica no es ajeno a este cambio. En este contexto, los repositorios digitales también se ven impactados por estas tecnologías, que introducen nuevas formas de interacción y recuperación de contenidos.

Las estrategias de *Generative Engine Optimization* surgen como una alternativa para optimizar la visibilidad de los repositorios digitales, ofreciendo lineamientos que permiten incrementar la frecuencia con la que estos recursos aparecen en las respuestas generadas por sistemas basados en IA. Si bien muchas de estas técnicas contribuyen a mejorar la experiencia de búsqueda tanto para personas como para algoritmos, es fundamental evitar una sobrefofocalización en la optimización exclusiva hacia los motores generativos.

Numerosas prácticas *GEO* son extrapolables al contexto de los repositorios. Sin embargo, deben aplicarse de manera equilibrada, preservando la esencia del repositorio como un servicio orientado principalmente a las personas. Este trabajo constituye una primera aproximación al análisis de cómo *GEO* puede influir positivamente en la visibilidad y el impacto de los repositorios digitales. Asimismo, plantea la necesidad de continuar estudiando la evolución de estas tecnologías y la forma en que los gestores de repositorios enfrentan los desafíos derivados del cambio tecnológico.

7. Referencias

Aggarwal, Pranjal; Murahari, Vishvak; Rajpurohit, Tanmay; Kalyan, Ashwin; Narasimhan, Karthik; Deshpande, Ameet (2024). *GEO: Generative Engine Optimization*. arXiv:2311.09735v3.
<https://doi.org/10.48550/arXiv.2311.09735>

Amer, Eslam; Elboghhdadly, Tamer (2024). The end of the search engine era and the rise of Generative AI: A paradigm shift in information retrieval. In: *2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*.
<https://doi.org/10.1109/MIUCC62295.2024.10783559>

AppLabx (2025). How to optimize content for generative AI: A Step-by-step GEO checklist. AppLabx - Digital Marketing and Web Agency.
<https://blog.applabx.com/how-to-optimize-content-for-generative-ai-a-step-by-step-geo-checklist>

Bommasani, Rishi; Hudson, Drew A.; Adeli, Ehsan; Altman, Russ; Arora, Simran; Arx, Sydney von; Bernstein, Michael S.; Bohg, Jeannette; Bosselut, Antoine; Brunskill, Emma; Brynjolfsson, Erik; Buch, Shyamal; Card, Dallas; Castellon, Rodrigo; Chatterji, Niladri; Chen, Annie; Creel, Kathleen; Davis, Jared Quincy; Demszky, Dora; Donahue, Chris; Doumbouya, Moussa; Durmus, Esin; Ermon, Stefano; Etchemendy, John; Ethayarajh, Kawin; Fei-Fei, Li; Finn, Chelsea; Gale, Trevor; Gillespie, Lauren; Goel, Karan; Goodman, Noah; Grossman, Shelby; Guha, Neel; Hashimoto, Tatsunori; Henderson, Peter; Hewitt, John; Ho, Daniel E.; Hong, Jenny; Hsu, Kyle; Huang, Jing; Icard, Thomas; Jain, Saahil; Jurafsky, Dan; Kalluri, Pratyusha; Karamcheti, Siddharth; Keeling, Geoff; Khani, Fereshte; Khattab, Omar; Koh, Pang Wei; Krass, Mark; Krishna, Ranjay; Kuditipudi, Rohith; Kumar, Ananya; Ladhak, Faisal; Lee, Mina; Lee, Tony; Leskovec, Jure; Levent, Isabelle; Li, Xiang Lisa; Li, Xuechen; Ma, Tengyu; Malik, Ali; Manning, Christopher D.; Mirchandani, Suvir; Mitchell, Eric; Munyikwa, Zanele; Nair, Suraj; Narayan, Avanika; Narayanan, Deepak; Newman, Ben; Nie, Allen; Niebles, Juan Carlos; Nilforoshan, Hamed; Nyarko, Julian; Ogut, Giray; Orr, Laurel; Papadimitriou, Isabel; Park, Joon Sung; Piech, Chris; Portelance, Eva; Potts, Christopher; Raghunathan, Aditi; Reich, Rob; Ren, Hongyu; Rong, Frieda; Roohani, Yusuf; Ruiz, Camilo; Ryan, Jack; Ré, Christopher; Sadigh, Dorsa; Sagawa, Shiori; Santhanam, Keshav; Shih, Andy; Srinivasan, Krishnan; Tamkin, Alex; Taori, Rohan; Thomas, Armin W.; Tramèr, Florian; Wang, Rose E.; Wang, William; Wu, Bohan; Wu, Jiajun; Wu, Yuhuai; Xie, Sang Michael; Yasunaga, Michihiro; You, Jiaxuan; Zaharia, Matei; Zhang, Michael; Zhang, Tianyi; Zhang, Xikun; Zhang, Yuhui; Zheng, Lucia; Zhou, Kaitlyn; Liang, Percy (2022). *On the opportunities and risks of foundation models*.
<https://doi.org/10.48550/arXiv.2108.07258>

Buitrago-Ciro, Jairo; Morales-Campos, Estela; Villamizar-Romero, César-Leonardo (2025). ¿Cómo está transformando la inteligencia artificial la comunicación científica? Desafíos, oportunidades y el papel de los actores involucrados: una revisión de alcance. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, v. 39, n. 104.
<https://doi.org/10.22201/iibi.24488321xe.2025.104.59032>

Calla Creative (2025). How to write FAQ sections that actually get picked up by AI search.
<https://callacreative.com/blog/how-to-write-faq-sections-that-actually-get-picked-up-by-ai-search>

Castro, Leyla J.; Neumann, Steffen; Schneider, Gabriel (2024). *SchemaOrg and JSON-LD for rich metadata integration in the NFDI*.
<https://doi.org/10.5281/zenodo.10830175>

Chen, Mahe; Wang, Xiaoxuan; Chen, Kaiwen; Koudas, Nick (2025). *Generative Engine Optimization: How to dominate AI search*.
<https://doi.org/10.48550/arXiv.2509.08919>

COAR (2023). *Good practice advice for managing multilingual and non-English language content in repositories*.
<https://doi.org/10.5281/zenodo.10053918>

COAR (2025). *Open repositories are being profoundly impacted by AI bots and other crawlers: Report from a COAR Survey*.
<https://coar-repositories.org/news-updates/open-repositories-are-being-profoundly-impacted-by-ai-bots-and-other-crawlers-results-of-a-coar-survey>

Daniels, Chris (2025). *From SEO to GEO: How agencies are navigating LLM-driven search | Analysis*. *Campaign Asia*.
<https://www.campaignasia.com/article/from-seo-to-geo-how-agencies-are-navigating-llm-driven-search/501593>

Dony, Christophe; Kuchma, Iryna; Ševkušić, Milica (2024). *Dealing with multilingualism and non-English content in open repositories: Challenges and perspectives*. *The journal of electronic publishing*, v. 27, n. 1.
<https://doi.org/10.3998/jep.5455>

Farnel, Sharon (2023). *Understanding repository functionality and structure*. In: *Discoverability in digital repositories*, 1st edition, Routledge.
<http://doi.org/10.4324/9781003216438-4>

Font-Julián, Cristina I.; Orduña-Malea, Enrique; Codina, Lluís (2024). *ChatGPT Search as a tool for scholarly tasks: evolution or devolution?*. *Infonomy*, v. 2, n. 5.
<https://doi.org/10.3145/infonomy.24.059>

Google Cloud (2025). *Prácticas recomendadas para Cloud Storage*.
<https://cloud.google.com/storage/docs/best-practices?hl=es-419>

Gozalo-Brizuela, Roberto; Garrido-Merchán, Eduardo C. (2023). *A survey of Generative AI Applications*.
<https://doi.org/10.48550/arXiv.2306.02781>

Hall, Brody (2025). *What is Reference Rate in AI Search?* *Loganix*.
<https://loganix.com/reference-rate>

Hersh, William (2024). *Search still matters: information retrieval in the era of generative AI*. *Journal of the American Medical Informatics Association : JAMIA*, v. 31, n. 9. <https://doi.org/10.1093/jamia/ocae014>

INSIDEA (2025). How can you optimize site speed to improve AI engine indexing and UX?

<https://insidea.com/blog/seo/aieo/optimize-site-speed-to-improve-ai-engine-indexing-and-ux>

Handley, Rachel (2025). We studied the impact of AI search on SEO traffic. Here's what we learned. *Semrush Blog*.

<https://www.semrush.com/blog/ai-search-seo-traffic-study>

Kaiser, Carolin; Kaiser, Jakob; Schallner, Rene; Schneider, Sabrina (2025). A new era of online search? A large-scale study of user behavior and personal preferences during practical search tasks with generative AI versus traditional search engines. In: *Proceedings of the extended abstracts of the CHI conference on human factors in computing systems*.

<https://doi.org/10.1145/3706599.3720123>

Kelly, Brenna (2025). Investigating ChatGPT Search: Insights from 80 million clickstream records. *Semrush Blog*.

<https://www.semrush.com/blog/chatgpt-search-insights/>

Law, Ryan; Guan, Xibei (2025). New study: AI assistants prefer to cite "fresher" content (17 million citations analyzed). *SEO Blog by Ahrefs*.

<https://ahrefs.com/blog/do-ai-assistants-prefer-to-cite-fresh-content/>

Lewis, Patrick; Perez, Ethan; Piktus, Aleksandra; Petroni, Fabio; Karpukhin, Vladimir; Goyal, Naman; Küttler, Heinrich; Lewis, Mike; Yih, Wen-tau; Rocktäschel, Tim; Riedel, Sebastian; Kiela, Douwe (2021). Retrieval-augmented generation for knowledge-intensive NLP Tasks.

<https://doi.org/10.48550/arXiv.2005.11401>

Oxford College of Marketing (2025). From SEO to GEO: Why marketers must evolve with the future of search.

<https://blog.oxfordcollegeofmarketing.com/2025/06/10/from-seo-to-geo-why-marketers-must-evolve-with-the-future-of-search/>

McKenzie, Leigh; Ofei, Michael (2025). LLM Visibility: The SEO metric no one is reporting on (yet).

<https://backlinko.com/llm-visibility>

OpenAIRE (2018). OpenAIRE guidelines and application profile for repository managers and publication platforms 4.0: More detail - More connectivity.

<https://www.openaire.eu/?view=article&id=2173&catid=61>

Orduña-Malea, Enrique; Cabezas-Clavijo, Álvaro (2023). ChatGPT and the potential growing of ghost bibliographic references. *Scientometrics*, v. 128, n. 9.

<https://doi.org/10.1007/s11192-023-04804-4>

Park, Lehm (2025). *Structured data in the AI search era*.
<https://www.brightedge.com/blog/structured-data-ai-search-era>

Pekala, Shayna (2018). Microdata in the IR: A low-barrier approach to enhancing discovery of institutional repository materials in Google. *The Code4Lib Journal*, n. 39.
https://journal.code4lib.org/articles/13191?utm_source=chatgpt.com

Pol, Tushar (2025). *What Is LLMs.txt & should you use it?*
<https://www.semrush.com/blog/llms-txt/>

Reyes-Lillo, Danilo; Morales-Vargas, Alejandro; Rovira, Cristòfol (2025a). Visibility, discoverability, findability, search engine optimization (SEO) and Academic SEO in digital repositories: A scoping review. *BiD*, v. June, n. 54.
<https://doi.org/10.1344/BID2025.54.06>

Reyes-Lillo, Danilo; Pastor-Ramon, Elena (2024). Use of handle in institutional repositories and its relationship with alternative metrics: A case study in Spanish-speaking America. *Hipertext.Net*, 29, 211–222.
<https://doi.org/10.31009/hipertext.net.2024.i29.17>

Reyes-Lillo, Danilo; Rovira, Cristòfol; Morales-Vargas, Alejandro (2025b). Factores para aumentar la visibilidad en repositorios digitales: metadatos, interoperabilidad, identificadores persistentes y optimización SEO-GEO. En J. Guallar; M. Váñez; A. Ventura-Cisquella (Coords). *Comunicación digital. Tendencias y buenas prácticas* (pp. 126-141). Ediciones Profesionales de la Información.
<https://doi.org/10.3145/cuvicecom.09.esp>

Sain, Ankit (2025). *How to structure content for AI discovery: FAQs, schema & conversational format*.
<https://whitebunnie.com/blog/how-to-structure-content-for-ai-discovery-faqs-schema-conversational-format>

SEOJuice (2025). *Reference rate - Generative SEO metric explained - Generative Engine Optimization definition*. <https://seojuice.io/glossary/geo/citation/reference-rate>

Shelby, Carolyn (2025). How LLMs interpret content: How to structure information for AI search. *Search Engine Journal*.
<https://www.searchenginejournal.com/how-llms-interpret-content-structure-information-for-ai-search/544308>

Solatorio, Aivin; Dupriez, Olivier (2024). *Efficient metadata enhancement with AI for better data discoverability*.
<https://blogs.worldbank.org/en/opendata/efficient-metadata-enhancement-with-ai-for-better-data-discovera>

SocialChampsVISIBLE™ (2025). *How to build citation-worthy content that AI models prefer*. *Visible™*.
<https://govisible.ai/blog/how-to-build-citation-worthy-content-that-ai-models-prefer>

Tapfuma, Mass Masona; Hoskins, Ruth Geraldine (2022). Adoption of institutional repositories towards realization of digital libraries: The Southern African perspective. In: T. Masenya (Ed.), *Innovative technologies for enhancing knowledge access in academic libraries* (pp. 156-175). IGI Global Scientific Publishing.

<https://doi.org/10.4018/978-1-6684-3364-5.ch010>

Venkit, Pranav Narayanan; Laban, Philippe; Zhou, Yilun; Mao, Yixin; Wu, Chien-Sheng (2024). *Search engines in an AI era: The false promise of factual and verifiable source-cited responses*.

<https://doi.org/10.48550/arXiv.2410.22349>

Yan, Tony (2025). *Schema markup best practices: Structured data that boosts AI citation & SEO*.

<https://quickcreator.io/blog/schema-markup-best-practices-structured-data-ai-seo>

Zavalina, Oksana L.; Burke, Mary (2023). User searching in digital repositories. In: *Discoverability in digital repositories*, Routledge.

<http://doi.org/10.4324/9781003216438-7>

Zhang, Alex (2025). Information retrieval in the age of Generative AI: A mismatch that matters. *Legal reference services quarterly*.

<https://doi.org/10.1080/0270319X.2025.2536920>