Artigo Original

Encontros Bibli

Modelo replicable de ingeniería de prompts para la recuperación de literatura científica con ChatGPT

Replicable prompt engineering model for scientific literature retrieval with ChatGPT
Modelo replicável de engenharia de prompts para recuperação de literatura científica com ChatGPT

Carlos Andrés Castillo Badilla
Sistema de Bibliotecas, Universidad de Las Américas, Sede Concepción, Chacabuco 539, Concepción, Chile
Universidad de Concepcion, Faculdade de Ciências Sociais, Programa de Mestrado em Pesquisa Social e Desenvolvimento, Concepción, Chile
ccastillob@udla.cl
https://orcid.org/0000-0002-7481-6155
A lista completa com informações dos autores está no final do artigo

RESUMEN

La expansión de la inteligencia artificial generativa ha transformado los procesos de recuperación de literatura científica, generando oportunidades de eficiencia y accesibilidad, pero también cuestionamientos sobre verificabilidad, rigor y generación de referencias ficticias en contextos académicos. En Bibliotecología y Ciencia de la Información, persiste un vacío metodológico respecto de protocolos replicables que permitan evaluar el desempeño operativo de estas herramientas en búsquedas bibliográficas.
Objetivo: Evaluar la eficacia de ChatGPT, modelo GPT-5.2, en entorno ChatGPT Plus, mediante un protocolo estructurado de ingeniería de prompts aplicado en nueve búsquedas controladas realizadas en febrero de 2026, distribuidas en tres niveles geográficos: global, latinoamericano y chileno.
Método: Se adoptó un diseño exploratorio–descriptivo documental con predominio cuantitativo y componente interpretativo complementario, operacionalizando indicadores dicotómicos de precisión, trazabilidad, coherencia conceptual y eficiencia temporal, junto con validación manual de DOI o URL.
Resultados: El corpus final consolidó 130 referencias únicas verificadas, sin registros de alucinaciones tras la aplicación del protocolo, con coherencia conceptual y trazabilidad del 100 % en las tres zonas analizadas. Los tiempos promedio oscilaron entre 14 y 22 minutos por búsqueda.
Conclusiones: Estos resultados evidencian que la ingeniería de prompts, bajo control metodológico explícito, puede mitigar riesgos documentados en la literatura, aportar un modelo replicable para bibliotecas universitarias y constituirse en una contribución sustantiva para el mapeo sistemático de la literatura, así como para la detección fundamentada de brechas y vacíos de conocimiento en distintos niveles geográficos.
PALABRAS CLAVE: Inteligencia artificial generativa. Ingeniería de prompts. Recuperación de información. Bibliotecas universitarias. Ciencia abierta.

ABSTRACT

The expansion of generative artificial intelligence has transformed scientific literature retrieval processes, creating opportunities for greater efficiency and accessibility while raising concerns regarding verifiability, rigor, and the generation of fabricated references in academic contexts. In Library and Information Science, a methodological gap persists concerning replicable protocols capable of evaluating the operational performance of these tools in bibliographic searches.
Objective: This study aimed to assess the efficacy of ChatGPT, GPT-5.2 model, in the ChatGPT Plus environment, through a structured prompt engineering protocol applied to nine controlled searches conducted in February 2026 across three geographic levels: global, Latin American, and Chilean.
Method: An exploratory–descriptive documentary design was adopted, predominantly quantitative with a complementary interpretative component, operationalizing dichotomous indicators of precision, traceability, conceptual coherence, and temporal efficiency, alongside manual validation of DOI or URL.
Results: The final corpus consolidated 130 unique verified references, with no hallucinated records after protocol implementation and with 100 percent conceptual coherence and traceability across all zones. Average response times ranged from 14 to 22 minutes per search.

Conclusions: The findings demonstrate that prompt engineering under explicit methodological control can mitigate risks documented in the literature, provide a replicable model for university libraries, and offer a substantive contribution to systematic literature mapping and the grounded identification of knowledge gaps and research voids at multiple geographic levels.
KEYWORDS: Generative artificial intelligence; prompt engineering; information retrieval; university libraries; open science.

RESUMO

A expansão da inteligência artificial generativa tem transformado os processos de recuperação de literatura científica, criando oportunidades de maior eficiência e acessibilidade, mas também levantando questionamentos sobre verificabilidade, rigor e geração de referências fictícias em contextos acadêmicos. Na biblioteconomia e na ciência da informação, persiste uma lacuna metodológica quanto à existência de protocolos replicáveis capazes de avaliar o desempenho operacional dessas ferramentas em buscas bibliográficas.
Objetivo: Avaliar a eficácia do ChatGPT, modelo GPT-5.2, no ambiente ChatGPT Plus, por meio de um protocolo estruturado de engenharia de prompts aplicado em nove buscas controladas realizadas em fevereiro de 2026, distribuídas em três níveis geográficos: global, latino-americano e chileno.
Método: Adotou-se um delineamento exploratório-descritivo documental, com predominância quantitativa e componente interpretativo complementar, operacionalizando indicadores dicotômicos de precisão, rastreabilidade, coerência conceitual e eficiência temporal, além da validação manual de DOI ou URL.
Resultados: O corpus final consolidou 130 referências únicas verificadas, sem registros de alucinações após a aplicação do protocolo e com 100 por cento de coerência conceitual e rastreabilidade nas três zonas analisadas. Os tempos médios variaram entre 14 e 22 minutos por busca.
Conclusões: Os resultados demonstram que a engenharia de prompts, sob controle metodológico explícito, pode mitigar riscos documentados na literatura, oferecer um modelo replicável para bibliotecas universitárias e constituir uma contribuição substancial para o mapeamento sistemático da literatura e para a identificação fundamentada de lacunas e vazios de conhecimento em diferentes níveis geográficos.
PALAVRAS-CHAVE: Inteligência artificial generativa; engenharia de prompts; recuperação da informação; bibliotecas universitárias; ciência aberta.

1 INTRODUCCIÓN

La incorporación de inteligencia artificial generativa en la recuperación de literatura científica ha abierto nuevas posibilidades, pero también ha planteado desafíos en términos de rigor, verificabilidad y control metodológico. En este contexto, esta investigación examina dicho fenómeno desde una perspectiva evaluativa aplicada al ámbito académico.

1.1 CONTEXTO Y JUSTIFICACIÓN

En los últimos años, la aparición de la inteligencia artificial generativa ha transformado de manera significativa la discusión en torno a la investigación y la gestión de la información científica. Su aplicación en los procesos de recuperación de literatura ha permitido optimizar la localización de fuentes y disminuir los tiempos de búsqueda, lo que abre posibilidades para una mayor eficiencia en la elaboración de marcos teóricos y en la realización de revisiones sistemáticas (Baldrich Domínguez-Oller, 2024).

Además, su aplicación en entornos académicos ha potenciado la accesibilidad y la democratización del conocimiento, especialmente en contextos con acceso limitado a bibliografía especializada. En este marco, la IA generativa no solo funciona como un motor

de búsqueda avanzada, sino que también transforma los procesos de alfabetización académica y científica, obligando a repensar los criterios de rigor, verificabilidad y ética en la investigación (Camacho Tambasco, 2023).

El avance de estas herramientas exige problematizar la precisión y la transparencia de los resultados producidos, así como la responsabilidad del investigador en la evaluación crítica de la información entregada por la IA. También persisten dudas sobre su rigor metodológico, la verificabilidad de las referencias y la validez epistemológica de su aplicación en procesos de búsqueda bibliográfica.

Estos problemas resultan especialmente relevantes en Bibliotecología y Ciencias de la Información, donde la recuperación eficiente y confiable constituye un eje fundamental para el fortalecimiento del conocimiento. Frente a este escenario de oportunidades y tensiones, resulta imprescindible examinar de manera crítica las nuevas herramientas que concentran la atención académica y social.

Entre ellas, destaca ChatGPT, un modelo de IA generativa desarrollado por OpenAI, entrenado para procesar lenguaje natural y generar respuestas coherentes a partir de grandes volúmenes de información textual. Su aplicación en entornos académicos ha abierto posibilidades en la producción de conocimiento, al facilitar tanto la búsqueda como la síntesis de literatura científica, aunque también ha suscitado debates sobre la calidad, veracidad y ética de los contenidos generados (Navas-Martín; Cuerdo-Vilches, 2024).

En particular, su uso en la recuperación de literatura científica permite formular estrategias de búsqueda en lenguaje natural y obtener síntesis rápidas, con ventajas de accesibilidad y eficiencia (Parisi; Sutton, 2024).

Sin embargo, dichas ventajas dependen de la calidad de las instrucciones empleadas, conocidas como prompts. La denominada ingeniería de prompts consiste en el diseño estratégico y controlado de instrucciones que orientan la IA hacia resultados más precisos, verificables y alineados con objetivos de investigación. Por ende, en este estudio la ingeniería de prompts se concibe como un protocolo metodológico orientado a sistematizar las búsquedas bibliográficas, facilitar el mapeo del conocimiento y detectar vacíos en la investigación. Esta estrategia permite garantizar la trazabilidad de los resultados y minimizar riesgos asociados, tales como la generación de referencias ficticias.

1.2 ANTECEDENTES Y ESTADO DEL ARTE METODOLÓGICO

La literatura internacional ha estudiado tanto avances como limitaciones. Se ha demostrado que ChatGPT puede formular estrategias de búsqueda complejas con alta precisión, aunque con menor exhaustividad, lo que plantea un dilema entre pertinencia y cobertura en la recuperación de literatura (Wang *et al*., 2023). De manera complementaria, se han documentado experiencias de apoyo a protocolos PRISMA, donde la IA contribuye a la identificación y síntesis de estudios, agilizando etapas de las revisiones sistemáticas (Ruiz Mendoza; Pedroza Zúñiga, 2024).

No obstante, análisis comparativos muestran que cerca del 30% de las referencias generadas corresponden a citas ficticias o no verificables, lo que confirma la necesidad de mecanismos de verificación rigurosos (Chelli *et al*., 2024). Asimismo, se evidencia la falta de métricas replicables y marcos estandarizados para reportar el uso de IA en revisiones sistemáticas, junto con la escasa exploración de su desempeño en otros idiomas y en bases regionales como SciELO y Redalyc (Alshami *et al*., 2023).

Estos hallazgos coinciden con diagnósticos globales sobre vacíos científicos, que subrayan la ausencia de estudios longitudinales y de métricas aplicables en bibliotecas universitarias. En América Latina, los estudios son aún exploratorios. Investigaciones en México, Colombia y Brasil han abordado dimensiones pedagógicas y éticas del uso de ChatGPT en educación superior y comunicación científica (Ariza Ruiz, 2023). Si bien se reconoce su potencial democratizador, persiste la ausencia de evaluaciones empíricas sistemáticas y de protocolos replicables en contextos bibliotecarios.

Además, la baja participación de especialistas en Ciencias de la Información refleja un vacío disciplinar, pues el debate regional ha sido liderado principalmente por áreas como la informática y la salud. En este escenario, Parisi y Sutton (2024, p. 30, 32–33) advierten “la necesidad de lineamientos metodológicos adaptados y subrayan la escasez de evidencia empírica en bibliotecas universitarias latinoamericanas”.

El panorama en Chile resulta aún más limitado. Aunque se han promulgado lineamientos institucionales y políticas públicas que reconocen la importancia de la IA en educación superior y ciencia abierta (Chile. Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, 2024), la literatura nacional se ha concentrado en debates éticos y normativos sin avanzar hacia investigaciones empíricas aplicadas. Son escasos los estudios que abordan servicios específicos como referencia, automatización o gestión con IA. En paralelo, no existen investigaciones que evalúen la eficacia de la IA generativa mediante ingeniería de prompts en la recuperación de literatura, lo que configura un vacío metodológico y disciplinar aún no resuelto.

1.3 PERTINENCIA DEL ESTUDIO

El escenario global, regional y nacional confirma la existencia de brechas metodológicas y disciplinarias aún no abordadas en la investigación chilena. Mientras la zona global avanza en la medición de fortalezas y limitaciones, y en América Latina emergen aproximaciones exploratorias sin estandarización, en Chile la falta de evidencia empírica limita la capacidad de las bibliotecas universitarias para evaluar el alcance y las restricciones de estas herramientas de inteligencia artificial generativa.

El presente estudio busca contribuir a superar este vacío mediante la evaluación aplicada de ChatGPT con ingeniería de prompts en la recuperación de literatura científica, considerando criterios de eficiencia temporal, trazabilidad de referencias y coherencia epistemológica. Con ello, se pretende aportar lineamientos metodológicos replicables y pertinentes para fortalecer la Bibliotecología y las Ciencias de la Información en el contexto latinoamericano.

En este marco, resulta necesario precisar el sentido en que se emplea el término eficacia en el presente estudio. Este concepto no se utiliza como una comparación entre sistemas de recuperación ni como una evaluación frente a un estándar externo. Tampoco se pretende establecer el rendimiento del modelo en términos algorítmicos clásicos.

En cambio, la eficacia se entiende como el desempeño operativo verificable de ChatGPT cuando es aplicado bajo un protocolo estructurado de ingeniería de prompts, considerando su capacidad para generar resultados bibliográficos sin referencias ficticias, con DOI o URL activos y con coherencia temática respecto del objetivo declarado. Esta definición se sustenta en indicadores observables como la validación manual de las referencias, la trazabilidad documental y la eficiencia temporal, en consonancia con el diseño exploratorio–descriptivo adoptado. En consecuencia, el término eficacia no se emplea en sentido algorítmico ni comparativo, sino como categoría operativa interna al diseño, vinculada exclusivamente al cumplimiento verificable de los criterios establecidos en el protocolo de ingeniería de prompts.

1.4 OBJETIVO GENERAL

Evaluar, mediante un protocolo replicable de ingeniería de prompts, la eficacia de ChatGPT en la recuperación de literatura científica en contextos bibliotecarios, considerando eficiencia, trazabilidad y coherencia epistemológica.

1.5 PREGUNTA DE INVESTIGACIÓN

¿Permite la aplicación de un protocolo metodológico de ingeniería de prompts obtener resultados bibliográficos verificables y sin referencias ficticias en la recuperación de literatura científica?

1.6 OBJETIVOS ESPECÍFICOS

a) Diseñar y documentar un protocolo metodológico replicable de ingeniería de prompts para entornos bibliotecarios.
b) Evaluar la eficiencia y la trazabilidad del protocolo en la recuperación de literatura científica, considerando tiempos de búsqueda y verificación de referencias.
c) Analizar los resultados obtenidos con el protocolo de ingeniería de prompts en tres niveles geográficos —global, regional latinoamericano y local chileno— con el fin de identificar vacíos metodológicos y disciplinarios en la recuperación de literatura científica.

En este escenario, la Bibliotecología y las Ciencias de la Información tienen un rol estratégico para consolidar un enfoque crítico y metodológicamente sólido frente al uso de inteligencia artificial generativa. Las bibliotecas universitarias constituyen un espacio privilegiado para validar protocolos de búsqueda, garantizar la trazabilidad de las referencias y fortalecer la alfabetización informacional en comunidades académicas. El presente estudio se inscribe en esta perspectiva, proponiendo un modelo metodológico replicable que permita evaluar la eficacia de ChatGPT con ingeniería de prompts en la recuperación de literatura científica. Con ello, se busca superar los vacíos identificados en el contexto global, regional y nacional, aportando lineamientos prácticos y teóricos que contribuyan al fortalecimiento disciplinar y al desarrollo de políticas informacionales pertinentes para América Latina y Chile.

En consecuencia, el estudio no constituye una revisión teórica, sino que propone un protocolo metodológico inédito basado en la ingeniería de prompts, orientado a fortalecer la práctica bibliotecaria mediante un enfoque empírico y metodológicamente sólido.

Asimismo, se enmarca en los principios de ciencia abierta, ya que los resultados estarán disponibles para su validación y reutilización en entornos académicos y profesionales. Con ello, se asegura una contribución sustancial al desarrollo de la Bibliotecología y las Ciencias de la Información, en línea con los criterios editoriales de la revista. Finalmente, para garantizar la transparencia y replicabilidad, el prompt metodológico se incluirá en el Apéndice A, de modo que cualquier investigadora o investigador pueda revisarlo, aplicarlo y adaptarlo en futuros estudios.

2 PROCEDIMIENTOS METODOLÓGICOS

2.1 EL DISEÑO DE LA INVESTIGACIÓN

El estudio adoptó un diseño exploratorio-descriptivo de carácter documental, orientado a evaluar el desempeño operativo verificable de ChatGPT aplicado mediante un protocolo de ingeniería de prompts en la recuperación de literatura científica. Este diseño se justificó en relación directa con el objetivo general: identificar y mapear vacíos de conocimiento en tres niveles geográficos. El enfoque exploratorio respondió a la escasez de antecedentes empíricos en el área, mientras que el carácter descriptivo permitió sistematizar resultados comparativos y generar evidencia replicable en Ciencias de la Información.

El enfoque metodológico correspondió a un diseño de predominio cuantitativo con componente interpretativo cualitativo complementario. La dimensión cuantitativa se orientó a la medición de indicadores operativos verificables, tales como precisión, trazabilidad, coherencia conceptual y eficiencia. Es importante precisar que los indicadores de precisión, trazabilidad y coherencia fueron operacionalizados mediante criterios dicotómicos verificables en términos temporales y de detección de alucinaciones en las referencias. Posteriormente, el componente cualitativo se incorporó con fines interpretativos, mediante análisis narrativo de calidad metodológica y transparencia en el uso de prompts dentro del corpus validado, con el propósito de contextualizar y explicar los patrones observados en los indicadores cuantitativos.

Estas variables se midieron a partir de un conjunto de búsquedas controladas realizadas en ChatGPT, modelo GPT-5.2, mediante suscripción ChatGPT Plus y ejecutado en la interfaz web oficial de OpenAI en el mes de febrero de 2026. Las búsquedas se ejecutaron aplicando el prompt metodológico del estudio como instrucción base dentro de ChatGPT y, en simultáneo, se utilizó la función “Investiga a fondo” como modo de ejecución para ampliar la exploración y estructuración de resultados, manteniendo constante el protocolo de ingeniería de prompts y la verificación manual de DOI/URL.

La función “Investiga a fondo” corresponde a la herramienta denominada oficialmente Deep Research, incorporada por OpenAI en 2025 como un modo avanzado de exploración y síntesis de información. A diferencia de una consulta convencional, esta funcionalidad permite una navegación extensiva y autónoma, con análisis estructurado de múltiples fuentes digitales, generación de informes organizados y vinculación directa a documentos originales verificables (OpenAI, 2025). En el presente estudio, su uso se limitó a complementar la ejecución del instrumento metodológico, sin sustituir en ningún momento la validación humana ni los controles de trazabilidad definidos en el diseño de investigación.

2.2 INSTRUMENTO METODOLÓGICO

El instrumento central consistió en un protocolo de ingeniería de prompts diseñado ad hoc (Apéndice A). Dicho protocolo contempló fases de identificación, verificación y análisis de resultados, junto con criterios de inclusión y exclusión claramente definidos. Con el propósito de minimizar la generación de referencias no verificables y asegurar la trazabilidad de los resultados, se desarrolló un esquema de verificación estructurado denominado Zero-Hallucination ZHP-7 (Blum, 2023), concebido específicamente para el presente estudio como mecanismo interno de control metodológico, sin pretensión de constituir un estándar externo ni un marco previamente validado en la literatura. Este protocolo fue diseñado para asegurar que cada referencia recuperada se mantuviera bajo estándares rigurosos de verificación bibliográfica, como respuesta directa a las limitaciones observadas en estudios previos sobre la generación de alucinaciones por IA en la recuperación de información.

Este esquema se integró al instrumento principal y permitió sistematizar el proceso de validación bibliográfica en cada búsqueda realizada. El procedimiento comprendió la identificación de afirmaciones y citas generadas por la herramienta, la normalización de referencias conforme a las normas editoriales vigentes, la validación técnica de DOI o URL,

la verificación de accesibilidad documental, la comprobación de concordancia temática con los objetivos de investigación, la detección de inconsistencias formales o metadatos incompletos y una revisión final de coherencia documental. Su aplicación fue constante en las 9 búsquedas ejecutadas, funcionando como salvaguarda metodológica frente a posibles inconsistencias generadas por la herramienta de inteligencia artificial.

Es importante precisar que el cumplimiento estricto del formato de salida establecido en el prompt constituye un componente metodológico del instrumento y no un criterio meramente editorial. La exigencia de construir cada referencia en formato estandarizado —por ejemplo, conforme a APA 7, incluyendo título completo, fuente y DOI expresado como enlace activo (https://doi.org/...) fue incorporada deliberadamente como mecanismo de control de trazabilidad dentro del protocolo. La experiencia empírica mostró que, cuando se altera o flexibiliza el formato solicitado, los resultados tienden a presentar enlaces incompletos, URL rotas o referencias sin identificadores persistentes verificables.

En cambio, la estructura normalizada con DOI válido permite asegurar correspondencia documental, verificabilidad técnica y coherencia entre la instrucción del prompt y la evidencia recuperada. En consecuencia, la fidelidad al formato de salida forma parte integral de la evaluación de eficacia del procedimiento, ya que garantiza que la recuperación bibliográfica no sea solo narrativamente plausible, sino documentalmente verificable.

A diferencia de un protocolo tradicional de búsqueda aplicado directamente en bases indexadas, la ingeniería de prompts utilizada en este estudio incorpora una arquitectura estructurada de instrucción lingüística, parametrización explícita del formato de salida, delimitación semántica anticipada y mecanismos internos de control de verificación. En consecuencia, la ingeniería de prompts no opera como una búsqueda tradicional sobre un índice preexistente, sino como una arquitectura metodológica de direccionamiento semántico y control documental, donde la instrucción lingüística, el formato de salida y la verificación técnica constituyen variables explícitamente parametrizadas y reproducibles.

Adicionalmente, la verificación técnica distinguió entre resolubilidad del identificador persistente y disponibilidad editorial del recurso. Se consideró verificable toda referencia cuyo DOI en formato https://doi.org/... fuese resoluble y condujera a una *landing page institucional* o a metadatos consistentes, aún cuando el acceso al texto completo estuviera restringido por suscripción. En determinados casos, la interfaz generó redirecciones alternativas hacia archivos PDF institucionales o repositorios secundarios, lo que confirmó la existencia documental del registro.

Asimismo, cuando el DOI conducía a una *landing* editorial con acceso restringido o error de visualización, se observó que podían activarse enlaces secundarios asociados a numeraciones internas del documento (p. ej., superíndices o referencias numeradas) que redirigían directamente al recurso completo. Por consiguiente, la verificación no se limitó a la navegabilidad de la página comercial, sino que consideró la resolubilidad técnica del identificador persistente y la disponibilidad de accesos alternativos generados por la propia interfaz. Esta precisión operativa evitó confundir restricciones de acceso con inexistencia bibliográfica.

Las búsquedas se organizaron en torno a tres ejes temáticos —ingeniería de prompts, recuperación de información y gestión bibliotecaria—, aplicando combinaciones controladas de términos en inglés y español (p. ej., “prompt engineering” OR “ingeniería de prompts” AND “literature search” OR “recuperación de información” AND library OR biblioteca OR “information management”). Esta estrategia garantizó consistencia metodológica, cobertura temática y trazabilidad en la recuperación de literatura.

Con el propósito de delimitar conceptualmente las categorías analíticas empleadas, se establecieron definiciones operativas diferenciadas. En primer lugar, se entiende por alucinación la generación por parte del modelo de información factualmente incorrecta, inexistente o no verificable, producida como resultado de procesos predictivos sin respaldo empírico explícito, fenómeno ampliamente documentado en la literatura sobre modelos de lenguaje de gran escala (Ji *et al*., 2023).

En el ámbito académico, esta problemática ha sido descrita en relación con la generación de contenidos inexactos o no contrastables en entornos de investigación asistida por IA (Ruiz Mendoza; Pedroza Zúñiga, 2024; Chelli *et al*., 2024). En segundo lugar, se denomina referencia inventada a aquellas citas que presentan estructura formal académica —autor, año, título, revista— pero cuya existencia no puede corroborarse en bases indexadas o repositorios oficiales mediante DOI activo o URL institucional válida, fenómeno documentado específicamente en estudios sobre el uso de ChatGPT para la búsqueda de literatura científica (Blum, 2023).

Esta categoría constituye un subconjunto específico de alucinaciones con implicancias éticas y metodológicas en la investigación científica. Finalmente, la consistencia conceptual se define como la concordancia temática entre la referencia recuperada y el objetivo específico declarado en el prompt, evaluada mediante revisión manual de pertinencia disciplinar y coherencia argumentativa. Esta variable es independiente de la variabilidad algorítmica del modelo y se orienta exclusivamente a

determinar la alineación entre el resultado generado y el propósito de la búsqueda. Se seleccionó un subconjunto de documentos con fines interpretativos, a fin de reflejar una diversidad de enfoques metodológicos y áreas temáticas dentro del corpus recuperado.

Los estudios seleccionados incluyen: estudios comparativos de desempeño, revisiones conceptuales sobre IA en bibliotecología, guías metodológicas para el uso de IA en bibliotecas y trabajos regionales en Latinoamérica, específicamente de México y Brasil. Esta diversidad de estudios permite representar una gama de metodologías empleadas en el campo de la recuperación de literatura con IA.

2.3 POBLACIÓN Y MUESTRA

La población objetivo se definió como la literatura científica y los documentos institucionales publicados entre 2020 y 2026, recuperables en bases indexadas como Scopus, Web of Science y SciELO, así como en repositorios oficiales ministeriales.

Se realizaron nueve búsquedas controladas, distribuidas en tres niveles geográficos diferenciados: global (Norteamérica, Europa, Asia–Pacífico, África y Medio Oriente), regional latinoamericano (excepto Chile) y local chileno, con tres ejecuciones por cada fase. El diseño correspondió a un estudio piloto comparativo con replicación controlada.

En cada zona se ejecutó una búsqueda base y dos replicaciones confirmatorias bajo idénticas condiciones instrumentales, variando exclusivamente la fecha de ejecución. Este procedimiento permitió evaluar la estabilidad operativa del protocolo frente a posibles variaciones contextuales del modelo, manteniendo homogeneidad en los criterios de inclusión, exclusión y estructura del prompt.

La replicación controlada se utilizó como mecanismo de consistencia interna, orientado a observar la estabilidad en tres indicadores: el número de referencias únicas verificadas por búsqueda, la proporción de referencias con validación técnica completa (DOI o URL activo) y la coherencia conceptual respecto del objetivo declarado en el prompt. La estabilidad en estos indicadores permitió analizar el comportamiento del protocolo sin introducir modificaciones en el instrumento metodológico.

Los criterios de inclusión contemplaron artículos revisados por pares en revistas Q1–Q4 y documentos ministeriales con DOI o URL institucional activo. Los criterios de exclusión descartaron preprints, prensa y repositorios no indexados, con el fin de mantener estándares de validez y trazabilidad acordes con las directrices editoriales internacionales.

2.4 TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

La recolección de datos se realizó mediante el protocolo de ingeniería de prompts aplicado en ChatGPT Plus (Investiga a fondo). Este instrumento metodológico se configuró como guía estructurada de búsqueda, con fases de identificación, verificación y análisis, junto con filtros de inclusión y exclusión predefinidos. El prompt base se mantuvo inalterado en todas las búsquedas, variando únicamente la temática central, lo que aseguró consistencia metodológica y comparabilidad entre resultados.

2.5 TÉCNICAS DE ANÁLISIS DE DATOS

El análisis se desarrolló en dos niveles complementarios:

Cuantitativo–descriptivo: Se calcularon indicadores de eficiencia (tiempo de respuesta), precisión (pertinencia de resultados), error (frecuencia de referencias ficticias) y coherencia epistemológica (consistencia conceptual de los hallazgos).

Los resultados cuantitativos fueron sistematizados en tablas comparativas por zona geográfica, permitiendo visualizar de manera sintética el número de referencias recuperadas, referencias verificadas, proporción de validación técnica, detección de referencias no verificables y tiempos promedio de respuesta, con el fin de facilitar la interpretación comparativa de los indicadores.

Es importante precisar que los indicadores de precisión, trazabilidad y coherencia fueron operacionalizados mediante criterios dicotómicos verificables (referencia verificable/no verificable; DOI o URL activo/inactivo; concordancia temática/concordancia insuficiente). En consecuencia, los valores observados reflejan el desempeño del procedimiento tras la aplicación sistemática del protocolo de verificación y no una propiedad intrínseca del modelo en ausencia de control humano. La uniformidad empírica obtenida responde al carácter estructurado del diseño comparativo con replicación temporal controlada, cuyo propósito fue evaluar la estabilidad operativa del protocolo bajo condiciones controladas, más que estimar dispersión estadística del comportamiento del modelo.

El análisis se basó en estadística descriptiva, considerando frecuencias absolutas, porcentajes y tiempos promedio de respuesta. Los porcentajes se calcularon sobre el total de referencias verificadas en las nueve búsquedas controladas.

Analítico–comparativo: Se contrastaron los resultados obtenidos en las tres fases geográficas, con el propósito de identificar asimetrías, vacíos y sesgos disciplinares.

Análisis de la eficiencia temporal: El análisis de la eficiencia temporal se realizó evaluando el tiempo promedio de respuesta por búsqueda en las tres zonas geográficas: global, América Latina (sin Chile) y Chile. Los tiempos promedio por búsqueda fueron los siguientes: Zona global: 22 minutos. América Latina (sin Chile): 18 minutos y Chile: 14 minutos.

Estos valores se obtuvieron de las nueve búsquedas realizadas, con tres ejecuciones en cada zona geográfica. La variación en los tiempos fue mínima dentro de cada zona, lo que sugiere estabilidad en el desempeño temporal del protocolo bajo condiciones controladas, sin evidenciar variaciones sustantivas entre zonas geográficas, independientemente de las diferencias geográficas. La medición de esta eficiencia temporal se realizó para proporcionar una evaluación operativa útil en el contexto de las bibliotecas universitarias, donde el tiempo es un recurso clave para mejorar la gestión y la accesibilidad de la información.

Además, cada conjunto de búsquedas generó un Informe de Vacíos Científicos compuesto por cinco apartados: estado actual del conocimiento, vacíos detectados, propuestas de investigación futura, conclusiones críticas y referencias verificadas.

La verificación bibliográfica se desarrolló en cuatro pasos: (i) integración de citas en el texto conforme a ABNT NBR 10520:2023; durante el trabajo de campo, el protocolo de ingeniería de prompts utilizó el formato APA 7 exclusivamente como estándar operativo de verificación preliminar, por su mayor transversalidad y reconocimiento internacional, normalizándose el manuscrito final a ABNT de acuerdo con las directrices editoriales de la revista; (ii) construcción de referencias completas conforme a ABNT NBR 6023:2018; (iii) validación técnica del DOI/URL; y (iv) registro de observaciones sobre idioma, tipo documental y relevancia disciplinar.

La validez operativa del procedimiento se aseguró mediante la verificación manual de cada DOI o URL y la aplicación sistemática de criterios dicotómicos verificables. La confiabilidad del procedimiento se garantizó a través de la replicación temporal controlada en las tres fases geográficas bajo condiciones instrumentales idénticas. Asimismo, la consistencia interna fue evaluada mediante la estabilidad de los indicadores definidos ex ante. Finalmente, la transparencia y trazabilidad metodológica se aseguraron mediante la documentación completa de la ingeniería del prompt, su inclusión en el apéndice y su disponibilidad en un repositorio abierto.

2.6 LIMITACIONES METODOLÓGICAS

En primer lugar, la dependencia del modelo GPT-5.2 en su versión pago Plus ($20 USD / mes) al momento de la ejecución (febrero de 2026) introduce un posible sesgo derivado de su arquitectura, parámetros internos y procesos de actualización continua, que pueden afectar la consistencia de los resultados en distintos momentos de uso.

La evidencia producida en este estudio se circunscribe al modelo y condiciones controladas bajo las cuales fue aplicado el protocolo. No obstante, el procedimiento metodológico desarrollado presenta un carácter transferible, en la medida en que su arquitectura de ingeniería de prompts y su esquema de verificación pueden ser replicados y evaluados en otros entornos tecnológicos, siempre que se mantengan criterios equivalentes de control, trazabilidad y validación documental.

Finalmente, la ausencia de programas estadísticos avanzados para el procesamiento de datos redujo la posibilidad de aplicar análisis multivariados; no obstante, este aspecto fue compensado mediante el uso de planillas de cálculo para el registro sistemático de frecuencias, porcentajes y tiempos de respuesta, lo que aseguró trazabilidad y comparabilidad interna de los hallazgos.

2.7 CONSIDERACIONES ÉTICAS

El uso de ChatGPT se limitó estrictamente a su función como herramienta metodológica de apoyo, sin sustituir el juicio académico ni la validación humana. Todas las salidas fueron verificadas manualmente para garantizar fidelidad bibliográfica y coherencia conceptual. No se involucraron sujetos humanos, por lo que no se requirió revisión por comité de ética, de acuerdo con la Resolución n.º 510/2016 aplicable a las Ciencias Sociales (Brasil. Ministerio de Salud, 2016). Los datos, protocolos y prompts utilizados se incluyen en el Apéndice A y se pondrán a disposición en un repositorio abierto confiable (p. ej., SciELO Data o repositorio institucional), en conformidad con los principios FAIR de la ciencia abierta y con las directrices de Encontros Bibli.

El protocolo de ingeniería de prompts se diseñó en consonancia con el Marco institucional para el uso de la IA en la Universidad de Las Américas, que establece directrices de integridad, transparencia y ética en la producción de conocimiento. Dicho marco enfatiza que la inteligencia artificial debe ser incorporada bajo criterios de integridad

y transparencia, orientando su uso en docencia, investigación y vinculación con el medio, en consonancia con los principios de ciencia abierta (Universidad de Las Américas, 2025).

Con el fin de asegurar la coherencia entre la declaración de adhesión a los principios FAIR (Findable, Accessible, Interoperable, Reusable) y la disponibilidad efectiva de información verificable, se estructuró y depositó el conjunto completo de resultados derivados de las nueve búsquedas controladas realizadas en este estudio. El dataset incluye la identificación de cada búsqueda (Search_ID), fase geográfica, fecha exacta de ejecución, modelo utilizado (GPT-5.2), función activada (Deep Research), versión del prompt aplicado, número de referencias inicialmente recuperadas, número de referencias validadas tras verificación manual, detección de duplicados, tiempos de respuesta y estado de validación técnica de cada DOI o URL.

En este estudio, el “tiempo de respuesta” se definió operacionalmente como el intervalo total, medido en segundos, entre la ejecución del prompt y la generación completa de la respuesta final por parte del sistema. Esta variable fue incorporada como indicador complementario de desempeño operativo, permitiendo observar posibles variaciones en la latencia asociadas a la complejidad de la búsqueda y al volumen de resultados generados.

La inclusión de esta variable se justifica en el contexto bibliotecario universitario, donde la eficiencia temporal constituye un criterio operativo relevante para la gestión de servicios de información y apoyo a la investigación.

3 RESULTADOS

Se ejecutaron nueve búsquedas controladas en febrero de 2026, distribuidas en tres niveles geográficos: global (3), latinoamericano (3) y chileno (3), aplicando de manera constante el mismo protocolo de ingeniería de prompts descrito en la sección metodológica. En cada zona se realizó una búsqueda base y dos replicaciones confirmatorias bajo condiciones instrumentales idénticas, con el propósito de evaluar la estabilidad operativa del procedimiento frente a variaciones temporales controladas.

Los hallazgos se presentan sin interpretación, en correspondencia directa con los indicadores operacionalizados en la sección de procedimientos metodológicos: (i) cobertura y eficiencia, (ii) verificación técnica y errores, (iii) coherencia conceptual, (iv) consolidado de validez técnica y (v) vacíos científicos por zona. Adicionalmente, se incorpora una matriz de calidad metodológica de los estudios recuperados como refuerzo de trazabilidad y

control documental. El detalle completo del protocolo de búsqueda y verificación aplicado en las nueve ejecuciones se encuentra disponible en el Apéndice A.

3.1 COBERTURA Y EFICIENCIA DEL PROTOCOLO

Las tres zonas geográficas completaron el número previsto de búsquedas, con tres ejecuciones por cada nivel geográfico, totalizando nueve búsquedas controladas. El tiempo promedio de respuesta por búsqueda fue de 22 minutos en la zona global, 18 minutos en América Latina y 14 minutos en Chile. Estas diferencias reflejan variaciones en volumen y diversidad de fuentes recuperadas, sin afectar la estabilidad operativa del protocolo.

La precisión operativa, entendida como la proporción de referencias verificadas respecto de las inicialmente recuperadas tras la aplicación del protocolo de validación, y la cobertura temática, definida como la concordancia del resultado con el objetivo declarado en el prompt, alcanzaron el 100 % en las tres zonas analizadas (véase Tabla 1).

Tabla 1- Eficiencia y precisión del protocolo por zona geográfica

Zona geográfica	Nº de búsquedas	Tiempo promedio (min)	Precisión (%)	Cobertura temática (%)
Zona global	3	22	100	100
América Latina	3	18	100	100
Chile	3	14	100	100

Nota. Precisión = referencias verificadas tras aplicación del protocolo / referencias inicialmente recuperadas. Cobertura temática = concordancia del resultado con el objetivo declarado en el prompt. Fuente: elaboración propia a partir de las búsquedas ejecutadas (2026).

3.2 VERIFICACIÓN TÉCNICA Y ERRORES

Tras la aplicación del prompt, que incluyó verificación manual de DOI/URL, comprobación de indexación y eliminación de duplicados, el corpus final consolidó 130 referencias únicas: 50 correspondientes a la zona global, 45 a América Latina y 35 a Chile.

Durante la fase inicial de recuperación se identificaron referencias con inconsistencias técnicas, tales como DOI inactivos o metadatos incompletos; sin embargo, todas fueron detectadas y excluidas mediante el procedimiento de validación. En consecuencia, el corpus final validado no presentó alucinaciones ni referencias no verificables. La Tabla 2 distingue entre incidencias detectadas en la fase preliminar y el estado final del corpus consolidado.

Tabla 2 – Errores en las referencias tras verificación

Zona geográfica	Total de referencias únicas	Alucinaciones (n)	Alucinaciones (%)	Referencias no verificables (%)
Zona global	50	0	0	4%
América Latina	45	0	0	11%
Chile	35	0	0	11%

Fuente: Elaboración propia a partir de las búsquedas ejecutadas (2026).

3.3 COHERENCIA CONCEPTUAL Y CLASIFICACIÓN

Se observó consistencia conceptual del 100 % y 0 % de errores de clasificación en las tres zonas geográficas, de acuerdo con las categorías analíticas previamente declaradas en la sección metodológica: ingeniería de prompts, recuperación de información y gestión bibliotecaria.

La coherencia conceptual fue evaluada mediante revisión manual de concordancia entre cada referencia validada y el objetivo específico declarado en el prompt correspondiente. No se registraron desviaciones temáticas ni asignaciones incorrectas de categoría en el corpus final consolidado (véase la Tabla 3).

Tabla 3 - Coherencia y concordancia epistemológica.

Zona geográfica	Nº de búsquedas	Consistencia conceptual (%)	Errores de clasificación (%)
Zona global	3	100	0
América Latina	3	100	0
Chile	3	100	0
Total	3	100	0

Nota. La consistencia conceptual fue determinada mediante evaluación manual de concordancia temática entre el objetivo declarado en el prompt y las referencias validadas en cada búsqueda.
Fuente: Elaboración propia a partir de las búsquedas ejecutadas (2026).

3.4 CONSOLIDADO DE VALIDEZ TÉCNICA

La precisión y la trazabilidad del corpus validado alcanzaron el 100 % en las tres zonas geográficas analizadas. En este marco, resulta necesario precisar el sentido en que se emplea el término eficacia en el presente estudio. Dicho concepto no se entiende como una medida comparativa entre sistemas de recuperación ni como contraste frente a parámetros externos, sino como expresión del grado de cumplimiento de los criterios de

validación definidos en la sección metodológica, que promedió 90 % en el conjunto total (véase la Tabla 4).

Tabla 4 - Consolidado general de verificación y validez técnica

Zona geográfica	Nº de búsquedas	Eficiencia (%)	Precisión (%)	Trazabilidad (%)
Zona global	3	100	100	100
América Latina	3	100	100	100
Chile	3	100	100	100
Total	9	100	100	100

Nota. Trazabilidad = DOI/URL activo y registro de verificación conforme a la ingeniería de prompts. Fuente: *Elaboración propia a partir de las búsquedas ejecutadas (2026).*

La estabilidad observada en los indicadores de precisión y trazabilidad constituye la base metodológica que permitió proceder al mapeo comparativo de vacíos científicos en las tres zonas geográficas.

3.5 VACÍOS DE CONOCIMIENTO CIENTÍFICO POR ZONA

Se aplicó una matriz analítica uniforme de vacíos metodológicos, disciplinares y regionales en las nueve búsquedas controladas. La estructura categorial se mantuvo constante en todas las ejecuciones. La Tabla 5 presenta un ejemplo ilustrativo correspondiente al prompt centrado en uso ético y académico de la inteligencia artificial generativa, seleccionado por su representatividad temática dentro del corpus recuperado.

Tabla 5 – Vacíos científicos identificados por zona geográfica (ejemplo temático: IA generativa).

Zona geográfica	Vacíos metodológicos	Vacíos disciplinares	Vacíos regionales	Observaciones (cuali/cuantitativo)
Zona global (3)	Falta de estandarización de protocolos con IA; reproducibilidad limitada; gestión de alucinaciones aún no estabilizada.	Predominio de salud/computación ; baja presencia de bibliotecología y ciencias sociales.	Sesgo hacia EE. UU./Europa; escasa visibilidad de África y Medio Oriente.	Predominan diseños cuantitativos (ensayos, comparativos; *recall*/precisión).
América Latina (3)	Ausencia de protocolos robustos; uso ad hoc de ChatGPT sin replicabilidad documentada.	Escasez de evidencia sistemática en bibliotecología; prevalecen ensayos/reflexion.	Falta de estudios comparativos regionales; débil integración escalable con SciELO/Redaly.	Mayoría cualitativos/reflexivos; emergen casos empíricos puntuales.

Zona geográfica	Vacíos metodológicos	Vacíos disciplinares	Vacíos regionales	Observaciones (cuali/cuantitativo)
Chile (3)	Presencia de marco normativo; ausencia de estudios empíricos asociados a bibliotecas universitarias.	Sin publicaciones nacionales en revisiones sistemáticas con ChatGPT.	Vacío en evidencia aplicada a búsquedas bibliográficas con IA.	Documentos de política; sin métricas cuantitativas.

Nota. La estructura analítica (vacíos metodológicos, disciplinares y regionales) se mantuvo idéntica en todas las temáticas; las categorías específicas varían según el prompt definido por el investigador. Fuente: Elaboración propia a partir de las búsquedas ejecutadas (2026).

3.6 CALIDAD METODOLÓGICA DE LOS ESTUDIOS RECUPERADOS

El corpus validado total estuvo compuesto por 130 referencias únicas. Con el propósito de reforzar la trazabilidad analítica del conjunto, se seleccionó un subconjunto cualitativo de ocho estudios (6,15 % del total), derivados de las nueve búsquedas ejecutadas. La selección fue de carácter intencional y no probabilístico, orientada a representar diversidad metodológica y disciplinar dentro del corpus recuperado, sin pretensión de representatividad estadística. Se incluyeron estudios comparativos experimentales centrados en evaluación de desempeño de modelos de lenguaje, revisiones conceptuales sobre inteligencia artificial en bibliotecología, guías metodológicas para la aplicación de IA en procesos de recuperación de información y trabajos regionales latinoamericanos.

Esta selección permitió ilustrar la heterogeneidad metodológica presente en el corpus validado. En los diseños comparativos empíricos y experimentales (Alaniz; Vu; Pfaff, 2023; Mostafahpour *et al*., 2024) se observó reporte parcial o limitado de transparencia en la formulación de prompts, mientras que la verificación técnica de referencias fue completa conforme a DOI o URL activo.

En los estudios conceptuales o de revisión (Parisi; Sutton, 2024; Blum, 2023), la transparencia en prompts fue parcial o no aplicable por diseño, manteniéndose verificación técnica completa de las referencias citadas.

Los análisis comparativos y guías metodológicas (Chelli *et al*., 2024; Lee *et al*., 2025) presentaron mayor nivel de explicitación metodológica en el uso de prompts y cumplimiento integral de verificación documental. Los trabajos de carácter regional (Walters; Wilder, 2023) no reportaron explícitamente la formulación de prompts, aunque mantuvieron verificación técnica completa de las referencias declaradas.

El rigor metodológico fue categorizado operativamente a partir de tres criterios verificables: (i) explicitación del uso de prompts, (ii) validación técnica de DOI o URL y (iii) declaración de limitaciones. Bajo esta matriz, se observó predominio de clasificaciones intermedias, con menor frecuencia de clasificaciones altas asociadas a estudios que documentaron detalladamente su procedimiento metodológico.

Esta caracterización descriptiva complementa los indicadores cuantitativos presentados en las secciones previas y contribuye a la evaluación integral de la calidad metodológica del corpus validado.

3.7 DESEMPEÑO ESTRUCTURAL DEL PROTOCOLO DE PROMPTS

El análisis del desempeño estructural del protocolo de ingeniería de prompts mostró que las nueve búsquedas controladas generaron informes de vacíos científicos alineados con los objetivos declarados en el estudio. La estructura analítica utilizada —vacíos metodológicos, disciplinares y regionales— se mantuvo constante en todas las ejecuciones, lo que permitió comparar resultados bajo criterios homogéneos.

En las nueve búsquedas no se registraron referencias no verificables dentro del corpus validado, conforme a los criterios técnicos definidos en la arquitectura de ingeniería de prompts aplicada en el estudio. Asimismo, todos los registros incorporaron DOI o URL resolubles, lo que permitió mantener consistencia en la trazabilidad documental.

La aplicación del mismo prompt base, sin modificaciones estructurales entre ejecuciones, posibilitó observar estabilidad operativa en los indicadores de precisión, coherencia conceptual y validación técnica previamente reportados. Las variaciones identificadas entre zonas geográficas se limitaron a diferencias en tiempos de respuesta y tipología documental, sin alteración en los parámetros de verificación establecidos.

Estos resultados describen el comportamiento del protocolo bajo condiciones controladas y delimitadas al modelo GPT-5.2 en el entorno y periodo de ejecución declarados.

4 DISCUSIÓN

Los resultados obtenidos indican que la aplicación de un protocolo metodológico basado en ingeniería de prompts permitió realizar nueve búsquedas controladas sin generación de referencias alucinadas, superando tasas de error reportadas en la literatura

reciente (por ejemplo, hasta ~30 % de referencias ficticias sin controles metodológicos) (Chelli *et al*., 2024).

Este desempeño respalda la idea de que la estructuración estratégica de instrucciones y la verificación sistemática de la operacionalización dicotómica de los criterios de precisión, trazabilidad y coherencia pueden mitigar sustancialmente las limitaciones epistemológicas asociadas a la IA generativa en procesos de revisión y recuperación de literatura. En consecuencia, se configura un aporte metodológico sustancial para la reproducibilidad y la trazabilidad en Ciencias de la Información.

Es importante señalar que los indicadores óptimos reportados se refieren al corpus validado después de la aplicación del protocolo y la verificación manual, y no a características inherentes al modelo en ausencia de este control humano. Los indicadores dicotómicos: referencia verificable/no verificable; DOI/URL activo/inactivo; concordancia temática/sin concordancia, reflejan el desempeño del procedimiento bajo condiciones controladas, lo que es coherente con la estructura experimental del diseño y previene interpretaciones erróneas.

Esta precisión metodológica es clave para evitar conclusiones demasiado generales sobre el rendimiento de los modelos sin las debidas salvaguardias (Blum, 2023; Chelli *et al*., 2024).

Los hallazgos de este estudio confirman diagnósticos globales que destacan las ventajas de herramientas como ChatGPT en términos de eficiencia y accesibilidad para recuperar literatura científica, pero también reafirman vacíos persistentes en la estandarización de protocolos de búsqueda y verificación rigurosa (Mahrishi; Abbas; Radovanović, 2024; Wilkinson, Oppelt; Owen, 2024).

En contraste con trabajos que reportan sesgos temáticos y limitaciones en la cobertura de fuentes no anglófonas (Zhou; Zhang, 2024), la implementación de un diseño replicable y criterios de saturación contribuyó a una validez operativa estable incluso en contextos disciplinarios y lingüísticos heterogéneos.

Esta coherencia con el objetivo declarado en los prompts sugiere que la ingeniería de prompts, cuando se estructura con criterios de verificación sistemáticos, no solo responde a los desafíos identificados en otras investigaciones sobre IA asistida, sino que además provee una base operativa más sólida para abordar esos retos (Parisi; Sutton, 2024).

Una contribución teórica y metodológica clave de este estudio es la distinción explícita entre vacíos de conocimiento y brechas de conocimiento, que permite delimitar

con precisión subáreas sin evidencia empírica suficiente frente a áreas con cobertura parcial, sesgada o contradictoria.

Esta diferenciación es esencial para orientar la formulación de preguntas de investigación relevantes y seleccionar líneas investigativas pertinentes. El protocolo de ingeniería de prompts utilizado demuestra que, al ajustar parámetros mínimos, como el tema objetivo, la expresión booleana base y los criterios de inclusión/exclusión geográficos, el instrumento puede aplicarse de forma transversal a diferentes disciplinas sin alterar su estructura central.

Esta adaptabilidad metodológica es consistente con el desarrollo de enfoques de ingeniería de prompts revisados en otras áreas de aplicación de IA, que destacan cómo prompts bien diseñados pueden mejorar resultados en diferentes dominios (Lee *et al*., 2025; Sahoo *et al*., 2024).

En el plano regional, los resultados confirman la carencia de estudios empíricos sistemáticos en América Latina, y particularmente en Chile, donde la literatura se enfrascó mayoritariamente en discusiones normativas o éticas sin propuestas metodológicamente replicables para búsquedas bibliográficas asistidas por IA (Chile. Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, 2024). Esta disparidad evidencia una desigualdad científica entre zonas geográficas: mientras en la zona global hay evidencia cuantitativa y comparativa sobre el uso de IA en procesos bibliográficos, en América Latina la producción es incipiente y fragmentaria.

El presente estudio hace una contribución original al cubrir esta brecha metodológica, ofreciendo lineamientos concretos para entornos bibliotecarios con necesidades similares. Metodológicamente, la investigación aporta un modelo replicable que combina ingeniería de prompts con marcos internacionales de reporte como PRISMA 2020, articulado con protocolos de control de alucinaciones y verificación manual.

Esta integración no había sido documentada con rigor comparativo en la literatura revisada, lo cual confiere un carácter pionero a la propuesta dentro de la Bibliotecología y las Ciencias de la Información (Parisi; Sutton, 2024; Ruiz Mendoza; Pedroza Zúñiga, 2024). El uso de criterios claramente definidos, como eficiencia temporal, trazabilidad y coherencia epistemológica, en lugar de depender solo de métricas algorítmicas, proporciona un marco flexible que puede adaptarse a futuras investigaciones en diferentes entornos bibliotecarios y contextos culturales.

No obstante, el estudio presenta limitaciones que orientan futuras investigaciones. La dependencia exclusiva del modelo GPT-5.2 disponible en el entorno ChatGPT Plus al

momento de la ejecución puede introducir sesgos tecnológicos susceptibles a cambios de versión y parámetros internos de los modelos. Asimismo, aunque la saturación operativa justificó el número de búsquedas controladas, esta cifra limita la cobertura temática frente a un universo potencialmente más amplio, y la ausencia de análisis multivariados avanzados reduce la complejidad analítica.

Un análisis más profundo podría mejorar la comprensión de patrones en la recuperación y categorización bibliográfica en diferentes disciplinas, como han sugerido estudios recientes sobre automatización de revisiones sistemáticas con IA (Sercombe, 2025). Sin embargo, lo más probable es que, aun cuando el modelo se actualice (por ejemplo, a GPT-5.3 o una versión más avanzada), el protocolo de ingeniería de prompts siga siendo igualmente funcional e incluso más potente.

Desde una perspectiva metodológica aplicada, este enfoque de búsqueda no solo facilita la exploración sistemática de literatura, sino que impulsa una reflexión crítica sobre los enfoques predominantes, permitiendo identificar deficiencias en criterios de diseño, replicabilidad y sesgos metodológicos que pueden ser reconstruidos en propuestas futuras.

La capacidad de generar informes de vacíos científicos y brechas de conocimiento con datos trazables y verificables ofrece un valor añadido para investigadores que buscan acelerar procesos investigativos, como la formulación del marco teórico o la identificación de preguntas relevantes de investigación.

Finalmente, al examinar tendencias emergentes, se detectaron vacíos en áreas clave como la integración de IA generativa en recuperación de información y ciencia abierta, lo que abre oportunidades para investigaciones que aborden no solo la aplicación de tecnologías avanzadas, sino también desafíos asociados con el sesgo algorítmico y la transparencia de los procesos de búsqueda. Las propuestas de investigación futura derivadas de este mapeo subrayan la importancia de profundizar en la aplicación de IA generativa para sistemas de recuperación bibliográfica, explorando no solo eficiencia y precisión, sino también equidad, diversidad de cobertura y replicabilidad metodológica.

En síntesis, el uso de un enfoque estructurado de ingeniería de prompts para búsqueda bibliográfica ha demostrado ser una herramienta poderosa para detectar vacíos y brechas en la literatura científica, mejorar la claridad sobre áreas críticas por explorar y promover una investigación más eficiente, rigurosa y metodológicamente sólida en bibliotecología y gestión de información. Esta metodología, al integrar IA y principios de ciencia abierta con salvaguardias humanas y criterios de rigor verificable, abre nuevas posibilidades para el avance del conocimiento en la disciplina y su aplicación práctica.

5 CONCLUSIONES

Este estudio ha demostrado que la aplicación de un protocolo metodológico basado en ingeniería de prompts ha permitido ejecutar nueve búsquedas controladas con una precisión y trazabilidad del 100%, sin la generación de referencias alucinadas, lo que responde directamente al objetivo general de evaluar la eficacia de ChatGPT en la recuperación de literatura científica en bibliotecología. El protocolo, al estructurar las búsquedas y verificar sistemáticamente las referencias, refuerza la reproducibilidad y la trazabilidad en los procesos de búsqueda en Ciencias de la Información.

Como contribución teórica, el estudio ha avanzado en la discusión sobre el uso de IA generativa en la investigación, al demostrar que los protocolos replicables pueden mitigar las limitaciones epistemológicas documentadas en estudios previos. A nivel metodológico, el modelo aquí presentado ofrece un avance significativo al integrar estrategias de verificación bibliográfica con estándares internacionales de reporte, lo cual aporta a la consolidación de mejores prácticas para el uso de herramientas de IA en bibliotecas universitarias.

El análisis comparativo ha revelado disparidades en la producción científica, donde, mientras la zona global cuenta con una sólida base empírica, América Latina y Chile aún presentan vacíos significativos. Este estudio contribuye a cerrar estas brechas al proporcionar evidencia aplicada y replicable que puede servir como base para futuras investigaciones regionales.

No obstante, la dependencia de un único modelo de IA y la restricción de las búsquedas limitan la generalización de los hallazgos. A pesar de ello, la consistencia metodológica y la validación manual de las referencias aseguran la transferibilidad del protocolo a otros entornos tecnológicos. Se recomienda que investigaciones futuras comparen distintos modelos de IA, incluyan análisis multilingües y profundicen en bases regionales como SciELO y Redalyc, incorporando además métricas estandarizadas de eficiencia y exhaustividad.

Este trabajo también motiva el uso ético y responsable de la inteligencia artificial en la investigación científica, con el objetivo de integrar herramientas como ChatGPT de manera ética y transparente. La disponibilización de los datos y protocolos en repositorios abiertos asegura que esta investigación no solo fortalezca la práctica bibliotecaria, sino que

también contribuya al cumplimiento de principios de ciencia abierta, con implicaciones significativas para el avance del conocimiento en bibliotecología y gestión de la información.

5.1 RECOMENDACIONES PARA APLICAR LA INGENIERÍA DE PROMPTS DE MANERA ADAPTABLE PARA CUALQUIER INVESTIGADOR

Este prompt fue diseñado para ser reutilizado en distintas investigaciones sin modificar su estructura metodológica central. Por ello, para adaptarlo a un nuevo estudio, el investigador no debe reescribir todo el instrumento, sino únicamente ajustar los apartados 2, 3 y 8. En el apartado 2 “Objetivo del usuario”, corresponde cambiar el tema de investigación, la expresión base de búsqueda y, cuando sea necesario, el alcance geográfico; en el apartado 3 “Parámetros de contexto”, deben adecuarse la versión del modelo de IA y el dominio disciplinar; y en el apartado 8 “Público objetivo o destinatario final”, debe especificarse la revista o el público objetivo al que se orienta el trabajo. Fuera de estos elementos, el resto del prompt debe mantenerse sin cambios, ya que su estabilidad asegura la coherencia interna del instrumento, la trazabilidad del proceso y la consistencia metodológica del análisis.

A continuación, se detallan los cambios a realizar en los apartados señalados dentro del “APÉNDICE A”:

2. Objetivo del usuario

2.1 El objetivo debe ser ajustado para reflejar el tema de investigación del usuario. En este caso, el ejemplo es sobre: "Ingeniería de prompts y modelos de lenguaje para búsqueda bibliográfica y evidencia sintética en bibliotecología y gestión de información ingeniería de prompts y modelos de lenguaje para búsqueda bibliográfica y evidencia sintética en bibliotecología”.

2.3 Orientación de la búsqueda. La expresión base de búsqueda debe ser ajustable según el área de estudio del usuario. Ejemplo de modificación de expresión base con operadores booleanos: "Para ciencias de la información y bibliotecología: (“prompt engineering” OR “ingeniería de prompts”) AND (“literature search” OR “recuperación de información”) AND (library OR biblioteca OR “information management”).

2.4 Definición del alcance geográfico: Este campo debe especificar la región de interés para las búsquedas y estudios. Si el trabajo debe centrarse en una determinada área geográfica, Ejemplo de ajuste geográfico: Incluir: [región específica como Norteamérica,

Europa, Asia-Pacífico, África]. Excluir: [cualquier región que no sea relevante para la investigación, como América Latina o Chile, si es necesario].

3. Parámetros de contexto

3.1 Modelo de destino: Se debe indicar la versión del modelo de ChatGPT. En este caso es: ChatGPT Plus, GPT-5.2.

3.3 Dominio disciplinar: Ejemplo: Dominio disciplinar: ciencias de la información, bibliotecología, inteligencia artificial, gestión de la información, metodologías de investigación (ajustar según área de investigación).

8. Público objetivo o destinatario final

8.1 Revista o público objetivo.

En síntesis: ajustes clave para adaptar el prompt:

1. Objetivo del usuario: Define el tema específico de investigación.
2. Expresión base de búsqueda: Adapta la expresión booleana a tu área.
3. Alcance geográfico: Ajusta la región de búsqueda según sea necesario.
4. Parámetros de contexto: Define el modelo de IA y el dominio disciplinar aplicable.
5. Indicar la revista a publicar.

Este protocolo es flexible y puede adaptarse fácilmente a diversas disciplinas y áreas de investigación. Siguiendo los ajustes mencionados, cualquier investigador podrá optimizar sus búsquedas bibliográficas y garantizar la trazabilidad de las referencias obtenidas, sin importar el campo de estudio. La estructura metodológica aquí presentada promueve una investigación más rigurosa y ética, favoreciendo la transparencia y la replicabilidad.

La ingeniería de prompts, cuando se aplica de manera rigurosa, puede ser una herramienta poderosa para la recuperación de información científica, contribuyendo tanto a la eficiencia operativa como a la precisión en los resultados. Este enfoque no solo mejora los procesos en bibliotecología, sino que tiene el potencial de transformar la manera en que los investigadores de cualquier disciplina interactúan con las tecnologías de inteligencia artificial generativa. Al hacerlo de manera responsable, ética y transparente, abrimos nuevas oportunidades para el avance del conocimiento y la ciencia abierta.

En última instancia, este protocolo establece un marco replicable que no solo responde a las necesidades actuales de la investigación, sino que también prepara el camino para el futuro de la ciencia, donde la inteligencia artificial y los enfoques metodológicos rigurosos se combinan para fortalecer la integridad y la accesibilidad de la

información académica. La posibilidad de adaptar este modelo a diferentes disciplinas refuerza su valor y relevancia, asegurando que todos los investigadores puedan utilizarlo para enriquecer y acelerar sus propios procesos investigativos.

REFERENCIAS

ALANIZ, L.; VU, C.; PFAFF, M. J. The utility of artificial intelligence for systematic reviews and Boolean query formulation and translation. **Plastic and Reconstructive Surgery Global Open**, v. 11, n. 10, e5339, 2023. DOI: 10.1097/GOX.0000000000005339. Disponible en: https://doi.org/10.1097/GOX.0000000000005339. Consultado el: 4 abr. 2026.

ALSHAMI, A.; ELSAYED, M.; ALI, E.; ELTOUKHY, A. E.; ZAYED, T. Harnessing the power of ChatGPT for automating systematic review process: methodology, case study, limitations, and future directions. **Systems**, Basel, v. 11, n. 7, p. 351, 2023. DOI: 10.3390/systems11070351. Disponible en: https://doi.org/10.3390/systems11070351. Consultado el: 4 abr. 2026.

ARIZA RUIZ, E. D. ChatGPT: una mirada desde la investigación. **Revista Investigaciones Andina**, Bogotá, v. 25, n. 46, p. 24–39, 2023. DOI: 10.33132/01248146.2256. Disponible en: https://doi.org/10.33132/01248146.2256. Consultado el: 4 abr. 2026.

BALDRICH DOMÍNGUEZ-OLLER, S. La inteligencia artificial y su impacto en la alfabetización académica: una revisión sistemática. **Educatio Siglo XXI**, Murcia, v. 42, n. 1, p. 55–74, 2024. DOI: 10.6018/educatio.609591. Disponible en: https://doi.org/10.6018/educatio.609591. Consultado el: 4 abr. 2026.

BLUM, M. ChatGPT produces fabricated references and falsehoods when used for scientific literature search. **Journal of Cardiac Failure**, v. 29, n. 9, p. 1332–1334, 2023. DOI: 10.1016/j.cardfail.2023.06.015. Disponible en: https://doi.org/10.1016/j.cardfail.2023.06.015. Consultado el: 4 abr. 2026.

BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. **Resolução n.º 510, de 7 de abril de 2016**. Establece las normas aplicables a investigaciones en Ciencias Humanas y Sociales. Diário Oficial da União: sección 1, Brasilia, DF, 24 de mayo de 2016. Disponible en: https://www.gov.br/conselho-nacional-de-saude/pt-br/atos-normativos/resolucoes/2016/resolucao-no-510.pdf/view. Consultado el: 19 sep. 2025.

CAMACHO TAMBASCO, J. El impacto de la inteligencia artificial en la educación: riesgos y potencialidades de la IA en el aula. **Revista Interuniversitaria de Investigación en Tecnología Educativa**, Murcia, v. 16, n. 1, p. 1–15, 2023. DOI: 10.6018/riite.584501. Disponible en: https://doi.org/10.6018/riite.584501. Consultado el: 4 abr. 2026.

CHELLI, M.; KHALFI, L.; RODRÍGUEZ, L.; TEJADA-LLACSA, P. J.; AUGEREAU, O. Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: comparative analysis. **Journal of Medical Internet Research**, Toronto, v. 26, e53164,

2024. DOI: 10.2196/53164. Disponible en: https://doi.org/10.2196/53164. Consultado el: 4 abr. 2026.

CHILE. Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. **Política Nacional de Inteligencia Artificial**. Santiago de Chile: MINCIENCIA, 2024. Disponible en: https://minciencia.gob.cl/areas/inteligencia-artificial/politica-nacional-de-inteligencia-artificial/. Consultado el: 19 sep. 2025.

JI, Ziwei; LEE, Nayeon; FRIESKE, Rita; YU, Tiezheng; SU, Dan; XU, Yan; ISHII, Etsuko; BANG, Yejin; CHEN, Delong; DAI, Wenliang; CHAN, Ho Shu; MADOTTO, Andrea; FUNG, Pascale. Survey of hallucination in natural language generation. **ACM Computing Surveys**, New York, v. 55, n. 12, p. 1–38, 2023. DOI: 10.1145/3571730. Disponible en: https://doi.org/10.1145/3571730. Consultado el: 1 abr. 2026.

LEE, Y.; OH, J. H.; LEE, D.; KANG, M.; LEE, S. Prompt engineering in ChatGPT for literature review: practical guide exemplified with studies on white phosphors. **Scientific Reports**, v. 15, n. 1, p. 15310, 2025. DOI: 10.1038/s41598-025-99423-9. Disponible en: https://doi.org/10.1038/s41598-025-99423-9. Consultado el: 4 abr. 2026.

MAHRISHI, M.; ABBAS, A.; RADOVANOVIĆ, D. Emerging dynamics of ChatGPT in academia: a scoping review. **Journal of University Teaching and Learning Practice**, Wollongong, v. 21, n. 3, p. 1–19, 2024. DOI: 10.53761/b182ws13. Disponible en: https://doi.org/10.53761/b182ws13. Consultado el: 4 abr. 2026.

MOSTAFAHPOUR, M.; FORTIER, J. H.; PACHECO, K.; MURRAY, H.; GARBER, G. Evaluating literature reviews conducted by humans versus ChatGPT: comparative study. **JMIR AI**, v. 3, e56537, 2024. DOI: 10.2196/56537. Disponible en: https://doi.org/10.2196/56537. Consultado el: 4 abr. 2026.

NAVAS-MARTÍN, M. A.; CUERDO-VILCHES, T. Discurso grupal basado en narrativas generadas por inteligencia artificial como metodología activa en la enseñanza en arquitectura. **Advances in Building Education, Madrid**, v. 8, n. 1, p. 59–74, 2024. DOI: 10.20868/abe.2024.1.5234. Disponible en: https://doi.org/10.20868/abe.2024.1.5234. Consultado el: 4 abr. 2026.

OPENAI. **Introducing deep research**. 2025. Disponible en: https://openai.com/index/introducing-deep-research/. Consultado el: 13 feb. 2026.

PARISI, V.; SUTTON, A. The role of ChatGPT in developing systematic literature searches: an evidence summary**. Journal of EAHIL**, Roma, v. 20, n. 2, p. 30–34, 2024. DOI: 10.32384/jeahil20623. Disponible en: https://doi.org/10.32384/jeahil20623. Consultado el: 4 abr. 2026.

RUIZ MENDOZA, K. K.; PEDROZA ZÚÑIGA, L. H. Uso de ChatGPT como ayudante en una RSL con el método PRISMA. **Sciencevolution**, México, v. 2, n. 10, p. 9–17, 2024. DOI: 10.61325/ser.v2i10.81. Disponible en: https://doi.org/10.61325/ser.v2i10.81. Consultado el: 4 abr. 2026.

SAHOO, P.; SINGH, A. K.; SAHA, S.; JAIN, V.; MONDAL, S.; CHADHA, A. A systematic survey of prompt engineering in large language models: techniques and applications.

arXiv, 2024. DOI: 10.48550/arXiv.2402.07927. Disponible en: https://doi.org/10.48550/arXiv.2402.07927. Consultado el: 4 abr. 2026.

SERCOMBE, J.; BRYANT, Z.; WILSON, J. Evaluating a customized version of ChatGPT for systematic review data extraction in health research: development and usability study. **JMIR Formative Research**, v. 9, e68666, 2025. DOI: 10.2196/68666. Disponible en: https://doi.org/10.2196/68666. Consultado el: 02 abr. 2026.

UNIVERSIDAD DE LAS AMÉRICAS (Chile). **Marco para el uso de la inteligencia artificial en UDLA:** docencia, investigación y vinculación con el medio. Santiago de Chile: Universidad de Las Américas, 2025. Serie: Documentos para la formación del cuerpo académico en IA. Disponible en: https://www.udla.cl/descargas/normativas/marco-para-uso-ia-en-udla.pdf. Consultado el: 1 oct. 2025.

WALTERS, William H.; WILDER, Esther Isabelle. Fabrication and errors in the bibliographic citations generated by ChatGPT. **Scientific Reports**, [S. l.], v. 13, n. 1, art. 14045, 2023. DOI: 10.1038/s41598-023-41032-5. Disponible en: https://doi.org/10.1038/s41598-023-41032-5. Consultado el: 5 abr. 2026.

WANG, S.; SCELLS, H.; KOOPMAN, B.; ZUCCON, G. **Can ChatGPT write a good Boolean query for systematic review literature search?** En: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 46., 2023, Taipei. Proceedings [...]. New York: ACM, 2023. p. 1426–1436. DOI: 10.1145/3539618.3591703. Disponible en: https://doi.org/10.1145/3539618.3591703. Consultado el: 4 abr. 2026.

WILKINSON, C.; OPPERT, M.; OWEN, M. Investigating academics' attitudes towards ChatGPT: a qualitative study. **Australasian Journal of Educational Technology**, Sydney, v. 40, n. 1, p. 142–156, 2024. DOI: 10.14742/ajet.9456. Disponible en: https://doi.org/10.14742/ajet.9456. Consultado el: 4 abr. 2026.

ZHOU, D.; ZHANG, Y. Political biases and inconsistencies in bilingual GPT models: the cases of the U.S. and China. **Scientific Reports**, London, v. 14, n. 13489, 2024. DOI: 10.1038/s41598-024-76395-w. Disponible en: https://doi.org/10.1038/s41598-024-76395-w. Consultado el: 4 abr. 2026.

NOTAS

CONTRIBUIÇÃO DE AUTORIA
Todas as autorias contribuíram substancialmente.

PREPRINTS
O manuscrito não é um preprint.

AGRADECIMENTOS
Agradeço especialmente a Mauricio Daza, Diretor de Bibliotecas da Universidade de Las Américas; a Nancy Baeza, Bibliotecária Nacional; e a Solange Salazar Huerta, Diretora Acadêmica, pelo apoio e incentivo oferecidos durante a realização deste trabalho.

USO DE INTELIGÊNCIA ARTIFICIAL
Foi utilizada a ferramenta ChatGPT (OpenAI) para apoiar a correção ortográfica do texto, bem como para estruturar o prompt de busca científica apresentado nos Apêndices. Além disso, a IA foi empregada como

auxílio na realização de buscas bibliográficas, cujo detalhamento encontra-se descrito na seção de Metodologia. Todo o processo de análise, interpretação e redação final permaneceu sob responsabilidade exclusiva do autor humano.

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA
Não se aplica.

CONFLITO DE INTERESSES
As pessoas autoras declaram não haver interesses conflitantes.

DISPONIBILIDADE DE DADOS DE PESQUISA E OUTROS MATERIAIS
Os dados foram publicados no próprio artigo. Todo o conjunto de dados que dá suporte aos resultados deste estudo está incluído no corpo do artigo.

ANUÊNCIA DE AVALIAÇÃO ABERTA
Deseja interagir diretamente com o avaliador caso este também concorde, durante o processo de avaliação do manuscrito?

LICENÇA DE USO

PUBLISHER
Universidade Federal de Santa Catarina. As ideias expressadas neste artigo são de responsabilidade das pessoas autoras, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES
Edgar Bisset Alvarez, Genilson Geraldo, Camila de Azevedo Gibbon, Gilmar Gomes de Barros, Rossana Gleucy de ávila Chagas E Carvalho, Daniela Capri.

HISTÓRICO
Recebido em: 02-10-2025
Aprovado em: 27-04-2026
Publicado em: 31-07-2026