Artigo de revisão

Convergências em Ciência da Informação,
Universidade Federal de Sergipe, São Cristóvão.
v. 9, 2026: e24689. ISSN 2595-4768.
DOI: 10.33467/conci.v9i.24689

Large Language Models: contribuições para a recuperação da informação em documentos

Large Language Models: contributions to information retrieval in documents

Modelos de Lenguaje de Gran Escala: contribuciones a la recuperación de información en documentos

Modèles linguistiques à grande échelle : contributions à la recherche d'informations dans les documents

Daiane Campos Procópio
Mestra em Gestão e organização do Conhecimento
Universidade Federal de Minas Gerais (UFMG)
E-mail: campos-daiane@ufmg.br. ORCID: 0000-0002-9006-191X

Patrícia Nascimento Silva
Doutora em Gestão e organização do Conhecimento
Universidade Federal de Minas Gerais (UFMG)
E-mail: patricians.prof@gmail.com. ORCID: 0000-0002-2405-8536

Renato Rocha Souza
Doutor em Gestão e organização do Conhecimento
Universidade Federal de Minas Gerais (UFMG).
E-mail: renato.rocha.souza@univie.ac.at. ORCID: 0000-0002-1895-3905

Submetido em: 13/03/2026
Aprovado em: 18/08/2026
Publicado em: 18/09/2026

Artigo submetido ao sistema de similaridade turnitin

RESUMO

A produção e o compartilhamento de informações em múltiplos formatos têm crescido exponencialmente nas últimas décadas, intensificando os desafios relacionados à organização, representação e recuperação da informação. Nesse contexto, os Large Language Models (LLMs) emergem como tecnologias capazes de aprimorar a busca semântica, interpretar consultas em linguagem natural e gerar respostas contextualizadas, ampliando as possibilidades de acesso à informação. Assim, esta pesquisa teve como objetivo analisar, na literatura científica, o uso de LLMs na recuperação da informação em documentos textuais, identificando abordagens metodológicas, contextos de aplicação, contribuições e limitações apontadas pelos estudos. Metodologicamente, realizou-se uma revisão narrativa da literatura, seguindo um protocolo, com buscas nas bases ACM Digital Library, ScienceDirect, Scopus e Web of Science. Foram selecionados artigos publicados entre 2020 e 2024, em português e inglês, com acesso aberto. Após a aplicação dos critérios de inclusão e exclusão e a leitura integral dos textos, 12 publicações compuseram o *corpus* de análise. Os estudos foram organizados em quatro eixos temáticos: automação de análises documentais em contextos regulatórios e institucionais; desenvolvimento de métodos e modelos de recuperação da informação; construção e enriquecimento de estruturas de conhecimento; e apoio à pesquisa científica. Conclui-se que os LLMs ampliam a capacidade interpretativa dos sistemas de recuperação da informação, embora apresentem limitações relacionadas a vieses, alucinações, custos computacionais e dependência de supervisão humana, demandando investigações futuras sobre transparência, avaliação de desempenho e impactos éticos e sociais.

Palavras-chave: documentos textuais; grandes modelos de linguagem; recuperação da informação.

ABSTRACT

The production and sharing of information in multiple formats have grown exponentially in recent decades, intensifying challenges related to information organization, representation and retrieval. In this context, Large Language Models (LLMs)

emerge as technologies capable of enhancing semantic search, interpreting natural language queries and generating contextualized responses, thereby expanding possibilities for information access. Thus, this study aimed to analyze, within the scientific literature, the use of LLMs in information retrieval from textual documents, identifying methodological approaches, application contexts, contributions, and limitations reported in the studies. Methodologically, a narrative literature review was conducted following a rigorous protocol, with searches carried out in the ACM Digital Library, ScienceDirect, Scopus and Web of Science databases. Articles published between 2020 and 2024, in Portuguese and English, available through open access were selected. After applying the inclusion and exclusion criteria and performing full-text reading, 12 publications composed the *corpus* of analysis. The studies were organized into four thematic axes: automation of document analysis in regulatory and institutional contexts; development of information retrieval methods and models; construction and enrichment of knowledge structures; and support for scientific research. It is concluded that LLMs enhance the interpretative capacity of information retrieval systems, although they still present limitations related to biases, hallucinations, computational costs and dependence on human supervision, requiring further research on transparency, performance evaluation and ethical and social impacts.

Keywords: information retrieval; large language models; textual documents.

RESUMEN

La producción y el intercambio de información en múltiples formatos han crecido exponencialmente en las últimas décadas, intensificando los desafíos relacionados con la organización, representación y recuperación de la información. Los Large Language Models (LLMs) emergen como tecnologías capaces de mejorar la búsqueda semántica, interpretar consultas en lenguaje natural y generar respuestas contextualizadas, ampliando las posibilidades de acceso a la información. Así, esta investigación tuvo como objetivo analizar, en la literatura científica, el uso de los LLMs en la recuperación de información

en documentos textuales, identificando enfoques metodológicos, contextos de aplicación, contribuciones y limitaciones señaladas por los estudios. Metodológicamente, se realizó una revisión narrativa de la literatura, siguiendo un protocolo riguroso, con búsquedas en las bases de datos ACM Digital Library, ScienceDirect, Scopus y Web of Science. Se seleccionaron artículos publicados entre 2020 y 2024, en portugués e inglés, de acceso abierto. Tras la aplicación de los criterios de inclusión y exclusión y la lectura completa de los textos, 12 publicaciones conformaron el corpus de análisis. Los estudios se organizaron en cuatro ejes temáticos: automatización del análisis documental en contextos regulatorios e institucionales; desarrollo de métodos y modelos de recuperación de la información; construcción y enriquecimiento de estructuras de conocimiento; y apoyo a la investigación científica. Se concluye que los LLMs amplían la capacidad interpretativa de los sistemas de recuperación de la información, aunque presentan limitaciones relacionadas con sesgos, alucinaciones, costos computacionales y dependencia de supervisión humana, lo que demanda futuras investigaciones sobre transparencia, evaluación del desempeño e impactos éticos y sociales.

Palabras clave: documentos textuales; grandes modelos de lenguaje; recuperación de la información.

RESUMÉ

La production et le partage d'informations sous de multiples formats ont connu une croissance exponentielle au cours des dernières décennies, ce qui a accentué les défis liés à l'organisation, à la représentation et à la recherche d'informations. Dans ce contexte, les grands modèles linguistiques (LLM) apparaissent comme des technologies capables d'améliorer la recherche sémantique, d'interpréter des requêtes en langage naturel et de générer des réponses contextualisées, élargissant ainsi les possibilités d'accès à l'information. Ainsi, cette étude visait à analyser, au sein de la littérature scientifique, l'utilisation des LLM dans la recherche d'informations à partir de documents textuels, en identifiant les approches méthodologiques, les contextes d'application, les

contributions et les limites rapportées dans les études. Sur le plan méthodologique, une revue narrative de la littérature a été menée selon un protocole rigoureux, avec des recherches effectuées dans les bases de données ACM Digital Library, ScienceDirect, Scopus et Web of Science. Des articles publiés entre 2020 et 2024, en portugais et en anglais, disponibles en libre accès, ont été sélectionnés. Après application des critères d'inclusion et d'exclusion et lecture intégrale des textes, 12 publications ont constitué le corpus d'analyse. Les études ont été organisées en quatre axes thématiques: l'automatisation de l'analyse documentaire dans des contextes réglementaires et institutionnels; le développement de méthodes et de modèles de recherche d'informations; la construction et l'enrichissement de structures de connaissances; et le soutien à la recherche scientifique. Il en ressort que les grands modèles linguistiques (LLM) améliorent la capacité d'interprétation des systèmes de recherche d'informations, bien qu'ils présentent encore des limites liées aux biais, aux hallucinations, aux coûts de calcul et à la dépendance vis-à-vis de la supervision humaine, ce qui nécessite des recherches supplémentaires sur la transparence, l'évaluation des performances et les impacts éthiques et sociaux.

Mots-clés: recherche d'informations; grands modèles linguistiques; documents textuels.

1 INTRODUÇÃO

A produção e o compartilhamento de informações em múltiplos formatos têm crescido de forma exponencial nas últimas décadas, configurando a era do *Big Data*. Conforme destacam Isson e Harriott (2016), esse fenômeno caracteriza-se pelo grande volume, variedade e velocidade com que os dados são gerados, o que dificulta sua análise e compreensão de maneira eficiente. Nesse contexto de sobrecarga

informacional, termos como "infoxicação" e "infodemia" passaram a ser utilizados para descrever o cansaço cognitivo e a dificuldade de localizar informações precisas. Tal cenário impõe desafios significativos à seleção, à organização e ao acesso à informação, reforçando a importância de mecanismos eficazes de gestão e recuperação da informação (Bawden; Robinson, 2008; Coneglian; Gonçalves; Santarém Segundo, 2017; Oliveira *et al.*, 2021).

No campo da Ciência da Informação (CI), essas transformações impactam diretamente os processos de organização, representação e recuperação da informação. Grande parte dos sistemas de recuperação da informação, por exemplo, é fundamentada em modelos baseados na correspondência de palavras-chave (Baeza-Yates; Ribeiro-Neto, 2013), enfrentando limitações diante de consultas formuladas em linguagem natural, ambíguas ou dependentes de contexto.

Nesse cenário, os *Large Language Models* (LLMs) surgem como uma solução tecnológica de grande potencial. Ao permitirem buscas semânticas e interpretações contextuais, os LLMs representam uma inovação relevante para o campo da Recuperação da Informação. Sua aplicação no contexto de documentos pode abrir novas possibilidades de acesso à informação, tornando-o mais eficiente (Procópio, 2025).

Assim, esta pesquisa se propõe a responder à seguinte questão: como os *Large Language Models* têm sido utilizados na recuperação da informação em documentos textuais?

O objetivo geral desta pesquisa consiste em analisar, na literatura científica, o uso de LLMs na recuperação da informação em documentos textuais. Como objetivos específicos, buscou-se: a) mapear as principais abordagens metodológicas empregadas e(b) identificar os contextos de aplicação dos LLMs na recuperação de informações em documentos.

Esta pesquisa justifica-se para aprofundar o diálogo entre a CI e a Inteligência Artificial (IA). Ao investigar o papel das tecnologias emergentes nos processos de recuperação da informação, o estudo amplia a compreensão sobre as transformações que tais tecnologias promovem no campo, fortalecendo a articulação interdisciplinar e a inovação nas práticas informacionais.

Como contribuições, o estudo propõe oferecer uma visão sistematizada da produção científica recente sobre o tema, destacando tendências, desafios e lacunas ainda pouco exploradas. Ao discutir criticamente as potencialidades e limitações dos LLMs na recuperação da informação em documentos, a pesquisa busca fornecer subsídios para futuras investigações e para o desenvolvimento de aplicações alinhadas às demandas da CI.

Por fim, destaca-se que este estudo deriva de uma pesquisa de mestrado realizada no Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais, a qual amplia e aprofunda as

discussões acerca da aplicação de LLMs na recuperação da informação em documentos textuais digitalizados.

2 *LARGE LANGUAGE MODELS*

Os *chatbots* que ganharam popularidade nos últimos anos, como o ChatGPT, *Gemin*i e *Microsoft Copilot,* são aplicações derivadas de um subcampo da IA conhecido como Inteligência Artificial Generativa (IAG). De acordo com Ozdemir (2025), LLMs são:

> Modelos de IA que geralmente (mas não necessariamente) são derivados da arquitetura Transformer e são projetados para analisar e gerar linguagem, código e muito mais. Esses modelos são treinados em grandes quantidades de dados de texto, permitindo-lhes capturar as complexidades e nuances da linguagem humana. Os LLMs podem realizar uma ampla gama de tarefas relacionadas ao idioma, desde a simples classificação de texto até a geração de texto, com alta precisão, fluência e estilo (Ozdemir, 2025, p. 11, tradução dos autores).

O funcionamento dos LLMs envolve o treinamento em grandes volumes de dados textuais de diversas fontes, como *sites*, livros, redes sociais e vídeos transcritos, com o objetivo de prever a sequência de palavras em diferentes contextos. A capacidade desses modelos está relacionada ao número de parâmetros, que pode alcançar trilhões. Esses parâmetros representam o que o modelo aprende a partir dos dados durante o treinamento, e quanto mais parâmetros ele possui, maior é a sua capacidade de capturar padrões e gerar

respostas mais precisas. Após o pré-treinamento, esses modelos passam por uma etapa chamada *fine-tuning* (ajuste fino), na qual eles são refinados com dados mais específicos e com métodos como Aprendizagem por Reforço com Feedback Humano (*Reinforcement Learning With Human Feedback*), buscando alinhar as respostas do modelo às expectativas de seus usuários. O objetivo dessa etapa pode variar desde o refinamento de habilidades conversacionais até o desempenho em tarefas como programação, resolução de problemas matemáticos ou aprovação em exames profissionais (Amaratunga, 2023; Stratis, 2024).

Porém, apesar de todos os benefícios, ainda há desafios importantes que precisam ser superados. Como esses modelos são treinados com grandes volumes de texto disponíveis na *internet*, eles acabam absorvendo e reproduzindo padrões problemáticos, incluindo discursos preconceituosos e informações falsas. Além disso, os modelos podem cometer erros, gerando respostas incorretas ou descontextualizadas que parecem convincentes, fenômeno conhecido como alucinação. Essas falhas ocorrem porque os LLMs não compreendem o conteúdo como humanos, mas apenas tentam prever a próxima palavra com base em padrões estatísticos (Dhamani; Engler, 2024).

Outro desafio refere-se ao crescente uso de dados sintéticos para o treinamento de LLMs. Eles ajudam a reduzir problemas como a falta de dados disponíveis, questões de privacidade e a baixa representatividade de determinados tipos

de dados nos conjuntos de treinamento. No entanto, quando são gerados por outros modelos de IA, também podem reproduzir erros, vieses e informações incorretas. Se esses dados forem utilizados para treinar novos modelos, esses problemas podem continuar sendo reproduzidos, afetando a qualidade e a confiabilidade das respostas. Desse modo, os dados sintéticos não devem substituir completamente os dados reais, mas ser utilizados de forma complementar e avaliados sempre que possível (Jesus; Algarve; Santarém Segundo, 2026).

3 METODOLOGIA

Para esta pesquisa, optou-se pela revisão narrativa da literatura, pois esse método oferece uma visão geral do tema, sem a pretensão de cobrir tudo, já que não segue uma busca rigorosamente sistemática, e seu valor está em permitir uma atualização rápida sobre o assunto (Cavalcante; Oliveira, 2020). Seu protocolo foi baseado em Nascimento Silva (2023).

As bases de dados científicas selecionadas para a realização das buscas foram a ACM Digital Library, ScienceDirect, Scopus e Web of Science. A escolha dessas fontes fundamenta-se em sua reconhecida qualidade científica e cobertura multidisciplinar. Essas bases reúnem periódicos de alto impacto, proporcionando acesso a estudos relevantes e atualizados em diversas áreas do conhecimento.

Como critérios de inclusão, foram selecionados artigos de pesquisas publicados entre 2020 e 2024 nos idiomas português e inglês e com acesso aberto ao seu conteúdo por meio do Portal de Periódicos da CAPES. A escolha por artigos com resultados de pesquisas ajuda a garantir a qualidade e a confiança nas informações analisadas. A opção por selecionar apenas trabalhos publicados entre 2020 e 2024 justifica-se pela rápida evolução das tecnologias de LLMs, impulsionadas por pesquisas recentes. Já a opção por textos em português e inglês amplia o número de fontes, sem comprometer a viabilidade da análise pela autora. Por fim, a opção por fontes de acesso aberto tem como objetivo viabilizar a leitura na íntegra das publicações.

Como critérios de exclusão, foram descartados trabalhos que tratem de LLMs sem o foco na recuperação da informação em documentos textuais e publicações repetidas.

Os procedimentos da pesquisa envolveram a elaboração da *string* de busca com o uso de operadores booleanos e termos em português e inglês relacionados aos LLMs e à recuperação da informação em documentos, incluindo variações terminológicas para abranger diferentes formas de expressão sobre o tema. O Quadro 1 apresenta uma síntese da metodologia adotada para a revisão narrativa da literatura neste estudo.

Quadro 1 – Metodologia para a revisão narrativa de literatura

Metodologia	
Publicações selecionadas	Artigos
Recorte temporal	2020-2024
Bases de dados	ACM Digital Library, ScienceDirect, Scopus e Web of Science
Idioma	Português e inglês
Tipo de acesso	Acesso aberto
***String* de busca**	(("large language models" OR LLM OR "modelos de linguagem de larga escala") AND ("information retrieval" OR "recuperação da informação") AND ("semantic search" OR "busca semântica") AND (documents OR documentos))
Data de execução	30 de julho de 2025

Fonte: elaborado pelos autores, 2025.

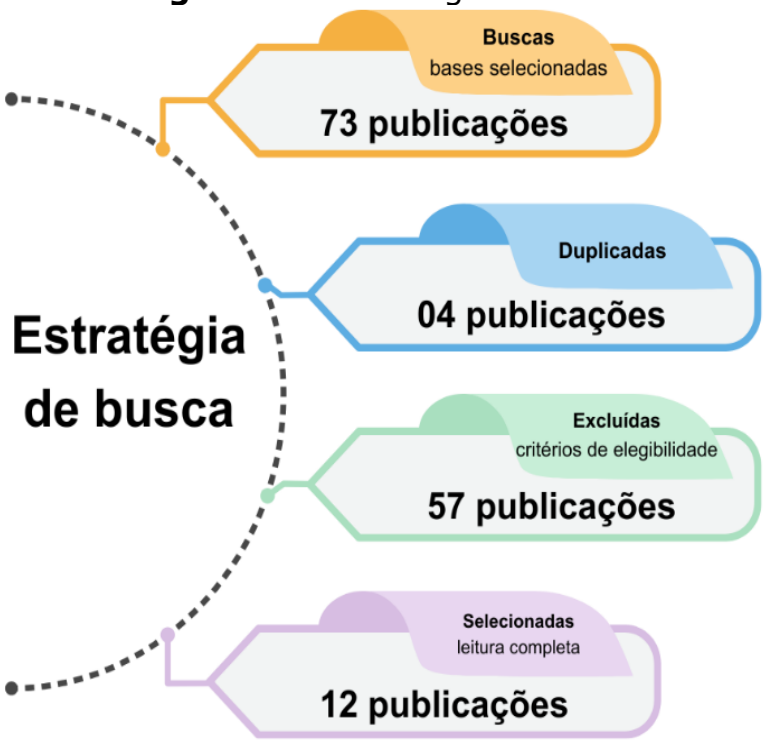
Em seguida, foi realizada a leitura dos títulos e resumos para verificar se os artigos abordavam a temática do estudo.

3.1 Resultados da busca

A estratégia de busca foi realizada nas bases ACM Digital Library, ScienceDirect, Scopus e Web of Science, utilizando a *string* de busca definida e filtros específicos, como período de publicação, acesso aberto, idioma e tipo de documento. Ao todo, foram recuperadas 73 publicações. Após a identificação de 4 registros duplicados, restaram 69 estudos para análise inicial. Em seguida, com base nos critérios de elegibilidade e na leitura de títulos e resumos, 57 publicações foram excluídas por não estarem alinhadas ao foco da pesquisa. Dessa forma, 12

estudos foram selecionados para leitura completa e análise, conforme ilustrado na Figura 1.

Figura 1 – Estratégia de busca



Fonte: elaborado pelos autores, 2025.

O Quadro 2 apresenta as 12 publicações selecionadas por terem atendido aos critérios pré-estabelecidos, com seus respectivos identificadores, títulos, autores, anos de publicação, temáticas principais e organizadas por base científica.

Quadro 2 – Publicações selecionadas

ID	Título	Autores	Ano	Tema(s)	Base
1	AgraBOT: Accelerating Third-Party Security Risk Management in Enterprise Setting through	Mert Toslali *et al.*	2024	Retrievel-Augmented-*Generation* (RAG)	ACM Digital Library

	Generative AI				
2	Scalable Range Search over Temporal and Numerical Expressions	Vebjørn Ohr e Dhruv Gupta	2024	Busca semântica, Question and Answering (Q&A)[1]	ACM Digital Library
3	A framework enabling LLMs into regulatory environment for transparency and trustworthiness and its application to drug labeling document	Leihong Wu *et al.*	2024	Busca semântica, Q&A e RAG	ScienceDirect
4	A platform-based Natural Language processing-driven strategy for digitalising regulatory compliance processes for the built environment	Ruben Kruiper *et al.*	2024	Processamento de Linguagem Natural (PLN), *web* semântica e ML	ScienceDirect
5	A Zero-Shot Strategy for Knowledge Graph Engineering Using GPT-3.5	Salvatore Carta *et al.*	2024	*Knowledge Graphs* [2]	ScienceDirect
6	AI-based decision support system for public procurement	Lucia Siciliani *et al.*	2023	Busca semântica e PLN	ScienceDirect
7	Automated Category and Trend Analysis of Scientific Articles on Ophthalmology Using Large	Hina Raja *et al.*	2024	Classificação de textos	ScienceDirect

[1] No português, Perguntas e Respostas.
[2] No português, Grafos do Conhecimento.

	Language Models: Development and Usability Study				
8	OLAF: An Ontology Learning Applied Framework	Marion Schaeffer *et al.*	2023	Ontologias	ScienceDirect
9	Recommendation system of scientific articles from discharge summaries	Adrián Alonso Barriuso *et al.*	2024	Sistema de recomendação e busca semântica	ScienceDirect
10	Blended RAG: Improving RAG (Retriever-Augmented Generation) Accuracy with Semantic Search and Hybrid Query-Based Retrievers	Kunal Sawarkar, Abhilasha Mangal e Shivam Raj Solanki	2024	RAG, busca semântica e Q&A	Scopus
11	Extractive Search for Analysis of Biomedical Texts	Daniel Clothiaux e Ravi Starzl	2022	Q&A e filtragem de documentos	Scopus
12	Seven Failure Points When Engineering a Retrieval Augmented Generation System	Scott Barnett *et al.*	2024	RAG	Scopus

Fonte: elaborado pelos autores, 2025.

Depois da triagem e da leitura integral dos documentos, procedeu-se à síntese dos textos, selecionados de forma analítica pelos autores, apresentada na próxima seção.

3.2 *Large Language Models* e recuperação da informação em documentos textuais

Nesta seção, são apresentados estudos que investigam o uso de LLMs na recuperação da informação em documentos textuais. As pesquisas destacam o potencial desses modelos para lidar com grandes volumes de dados, substituindo tarefas feitas manualmente por um processo automatizado e ágil. Para alcançar esse objetivo, foram adotadas diferentes abordagens metodológicas, incluindo a criação de *frameworks* e o uso de técnicas de IA, como Processamento de Linguagem Natural (PLN) e Machine Learning (ML). Apesar das variações, as pesquisas apontam para o mesmo propósito: transformar a difícil tarefa de buscar informações em grandes acervos em algo mais rápido, preciso, confiável e automatizado.

Toslali *et al.* (2024) [ID 1] desenvolveram o AgraBOT, uma ferramenta baseada em IAG e RAG, com o objetivo de automatizar e acelerar o processo de avaliação de riscos de segurança na contratação de serviços de empresas terceirizadas. O sistema foi capaz de reduzir o tempo médio de análise de documentos de vários dias para apenas quatro minutos, oferecendo uma interface amigável e permitindo que os avaliadores forneçam *feedback* sobre as respostas geradas pelo sistema. A solução não tem a intenção de substituir os avaliadores de riscos, mas de apoiá-los na análise dos documentos, que incluem relatórios de segurança e questionários técnicos.

Ohr e Gupta (2024) [ID 2] criaram o sistema Nash para realizar buscas por intervalos de tempo (períodos do tempo) e numéricos (faixa de valores) expressos em linguagem natural. O principal desafio abordado é a recuperação da informação a partir de expressões ambíguas, como "início dos anos 2000" ou "cerca de 10%". Para isso, o sistema utiliza curvas *z-order*, uma técnica baseada em geometria e *hashing* que permite representar e indexar essas expressões como regiões bidimensionais, facilitando consultas por operadores como "contém" e "próximo". Em testes com grandes bases de dados, incluindo Wikipédia e Common Crawl, o Nash demonstrou desempenho superior às abordagens de recuperação da informação tradicionais, além de exigir menos espaço de armazenamento.

Wu *et al.* (2024) [ID 3] propõem um *framework* chamado AskFDALabel, desenvolvido para integrar LLMs ao ambiente regulatório da Food and Drug Administration, com foco em transparência, segurança e adaptabilidade. A solução combina busca semântica com geração de texto baseada em RAG, permitindo consultas eficazes a documentos sobre medicamentos.

Kruiper *et al.* (2024) [ID 4] apresentam a plataforma i-ReC (Intelligent Regulatory Compliance) como solução para digitalizar o processo de conformidade regulatória no setor de construção civil do Reino Unido. Partindo dos problemas apontados no Relatório Hackitt (2018), o estudo destaca o uso de técnicas de PLN, ML e *web* semântica para extrair e

processar automaticamente requisitos normativos de centenas de documentos regulatórios. Entre os recursos desenvolvidos, estão o SPaR.txt, uma ferramenta para identificar expressões complexas, e um sistema de busca semântica que também apresentou bons resultados. Os autores abordam, ainda, a criação de um grafo de conhecimento com termos técnicos extraídos automaticamente, o que melhora a recuperação das informações e a interoperabilidade entre ferramentas.

Carta *et al.* (2024) [ID 5] mostram uma abordagem para a construção e o enriquecimento de grafos de conhecimento utilizando o GPT-3.5 em modo *zero-shot*, ou seja, sem a indicação de exemplos nos *prompts*. A metodologia baseia-se em uma sequência de *prompts* estruturados que permitem extrair entidades, seus tipos, suas descrições e seus relacionamentos em formato de triplas (sujeito-predicado-objeto). Os resultados mostram alta precisão na identificação de entidades e relações, além da capacidade de enriquecimento semântico dos grafos de conhecimento, especialmente em domínios com pouco suporte de bases de conhecimento.

Siciliani *et al.* (2023) [ID 6] propõem a criação de um sistema de apoio à decisão voltado para a gestão de compras públicas, com foco na análise de dados textuais estruturados e não estruturados de licitações. A solução integra IA e PLN para permitir buscas semânticas, extração de informações em triplas, resumos automáticos e visualizações em painéis interativos. A avaliação com servidores públicos mostrou bons índices de usabilidade e demonstrou o potencial da IA para

tornar os processos de compras públicas mais transparentes, eficientes e acessíveis a usuários sem conhecimento técnico na área.

Raja *et al.* (2024) [ID 7] propõem um sistema automatizado baseado em LLMs para classificar artigos científicos e identificar tendências na área de oftalmologia, com aplicação em revisões de literatura. Os testes demonstraram que o modelo é eficiente, adaptável a outras áreas e mais rápido que a classificação manual de publicações científicas.

Schaeffer *et al.* (2023) [ID 8] elaboraram um *framework* modular e automatizado para a construção de ontologias a partir de textos não estruturados. O OLAF abrange todo o processo de aprendizagem de ontologias, desde o pré-processamento até a extração de axiomas, incorporando técnicas de PLN, enriquecimento com fontes externas, como o WordNet, e métodos como a análise formal de conceitos, além de priorizar a automação, reduzindo ao máximo a intervenção humana. O *framework* foi testado na construção de uma ontologia para um mecanismo de busca semântica de produtos técnicos. Os resultados mostraram uma boa capacidade de extração de conceitos, mas dificuldades na extração de relações e axiomas, principalmente devido a ruídos gerados no pré-processamento e à limitação das técnicas de extração de relações.

Barriuso *et al.* (2024) [ID 9] apresentam o MELENDI, um sistema de recomendação de artigos científicos voltado a profissionais da saúde, que utiliza resumos de alta hospitalar

como base para identificar diagnósticos e sugerir publicações relevantes. O sistema combina modelos de similaridade semântica e estimativas automáticas de relevância, considerando tanto o conteúdo dos diagnósticos quanto fatores como reputação dos autores e número de citações. A ferramenta demonstrou viabilidade de uso em ambientes clínicos reais, com destaque para sua capacidade de auxiliar médicos a se manterem atualizados com pesquisas associadas ao seu dia a dia.

Sawarkar, Mangal e Solanki (2024) [ID 10] criaram uma abordagem para melhorar a precisão de sistemas baseados em RAG, combinando diferentes estratégias de busca: consultas híbridas com recuperação semântica e índices de vetores densos e esparsos. Os autores utilizam o termo "*blended retriever*" (recuperador híbrido), integrando as técnicas de recuperação da informação BM25, KNN e ELSER, o que resulta na melhoria do desempenho na busca e geração de respostas em conjuntos de dados como o Natural Questions e o TREC-COVID. Os resultados indicam que o uso combinado de diferentes métodos de recuperação da informação é mais eficaz do que abordagens isoladas, especialmente em contextos com grandes volumes de dados.

Clothiaux e Starzl (2022) [ID 11] desenvolveram um sistema voltado à análise de textos biomédicos com base em busca extrativa. Para isso, reuniram um conjunto de documentos relacionados aos temas de biologia, medicina e química, selecionados da base PubMed. O sistema é composto

por dois módulos principais: o primeiro realiza a busca extrativa, combinando os termos da consulta com o conteúdo dos documentos e retornando os resultados ao usuário; e o segundo é um módulo de análise que oferece diversas ferramentas para explorar os dados recuperados, como sumarização, geração de perguntas e respostas e análise de entidades. Os resultados indicam que a combinação da busca extrativa com análise semântica e modelos neurais pode ser eficaz tanto para especialistas quanto para usuários não especializados na área biomédica.

Barnett *et al.* (2024) [ID 12] apresentam uma análise prática dos principais desafios enfrentados no desenvolvimento de sistemas RAG com base em três estudos de caso aplicados nas áreas de pesquisa, educação e biomedicina. O trabalho identifica sete pontos críticos de falha que afetam a qualidade desses sistemas, incluindo erros como ausência do conteúdo necessário, documentos mal ranqueados, informações fora do contexto, falhas na extração das respostas, formatos inadequados de saída e respostas incompletas. O artigo oferece uma importante contribuição ao destacar as dificuldades práticas de aplicar RAG e propõe direções futuras de pesquisa para aprimorar seu desempenho.

4 DISCUSSÃO DOS RESULTADOS

A análise dos estudos selecionados permitiu agrupar as pesquisas em quatro eixos temáticos que refletem diferentes formas de utilização de LLMs na recuperação da informação. Essa categorização foi realizada com base no papel desempenhado pelos LLMs em cada uma das aplicações desenvolvidas nos estudos, buscando apresentar como essas tecnologias vêm sendo empregadas em diferentes contextos e finalidades. Dessa forma, os trabalhos foram classificados em: (1) automação de análises documentais em contextos regulatórios e institucionais, (2) desenvolvimento de modelos e métodos de recuperação da informação, (3) construção e enriquecimento de estruturas de conhecimento e (4) apoio à pesquisa e à prática científica. O Quadro 3 apresenta os estudos organizados de acordo com a área da pesquisa e o eixo temático.

Quadro 3 – Categorização dos estudos

Estudo	Área	Eixo
Barriuso *et al.* (2024) [ID 9]	Saúde	Apoio à pesquisa e à prática científica (4)
Clothiaux e Starzl (2022) [ID 11]		
Raja *et al.* (2024) [ID 7]		
Wu *et al*. (2024) [ID 3]		
Kruiper *et al.* (2024) [ID 4]	Construção Civil	Automação de análises documentais em contextos regulatórios e institucionais (1)
Siciliani *et al.* (2023) [ID 6]	Administração Pública	
Toslali *et al*. (2024) [ID 1]	Gestão de Riscos	
Carta *et al.* (2024) [ID 5]	Ciência da	Construção e

Schaeffer *et al.* (2023) [ID 8]	Computação	enriquecimento de estruturas de conhecimento (3)
Barnett *et al.* (2024) [ID 12]	Educação, Pesquisa e Biomedicina	Desenvolvimento de modelos e métodos de recuperação da informação (2)
Ohr e Gupta (2024) [ID 2]	Ciência da Computação	
Sawarkar, Mangal e Solanki (2024) [ID 10]		

Fonte: elaborado pelos autores, 2025.

O eixo de **automação de análises documentais em contextos regulatórios e institucionais (IDs 1, 4 e 6)** reúne os trabalhos que exploram o uso de LLMs para extrair e interpretar informações de documentos normativos, técnicos e administrativos. Esses estudos apontam para uma tendência de digitalização e automatização da recuperação da informação e da interpretação de textos regulatórios, com foco na redução do tempo de análise e no aumento da precisão em tomadas de decisão.

No eixo de **desenvolvimento de modelos e métodos de recuperação da informação (IDs 1, 10 e 12)**, estão os trabalhos que propõem inovações metodológicas para aprimorar o processamento e a recuperação de informações em documentos pertencentes a grandes bases de dados. Esses estudos contribuem para o avanço das estratégias de busca e recuperação de informações em larga escala, essenciais para o funcionamento de sistemas de IA baseados em LLMs.

Os trabalhos agrupados no eixo de **construção e enriquecimento de estruturas de conhecimento (IDs 5 e**

8) tratam do uso de LLMs para organizar e representar o conhecimento de forma estruturada, destacando o papel dos grafos e das ontologias. Ambos os estudos mostram como os LLMs podem reduzir a dependência da intervenção humana na modelagem de domínios complexos.

O eixo de **apoio à pesquisa e à prática científica (IDs 3, 7, 9 e 11)** agrupa as pesquisas voltadas à aplicação dos LLMs em contextos científicos e biomédicos, com foco na classificação, recomendação e análise automatizada da literatura. Os estudos demonstram o potencial dos LLMs para apoiar a geração e a disseminação do conhecimento científico, ampliando o acesso e a eficiência na busca e análise de informações especializadas.

De forma geral, as categorias, organizadas em eixos temáticos, revelam que as pesquisas convergem para o uso de LLMs como ferramentas de automação, apoio à decisão e aprimoramento da recuperação de informações, embora variem quanto ao domínio de aplicação. Nota-se que a maioria dos estudos é voltada à saúde e às áreas regulatórias, refletindo a necessidade de precisão e confiabilidade nas análises das informações. Em todos os eixos, há esforços para tornar os modelos mais eficientes e aplicáveis a diferentes cenários, reforçando o papel dos LLMs como tecnologias potenciais para a transformação da gestão e recuperação da informação.

Porém, apesar dos avanços e benefícios mostrados nas pesquisas analisadas, a construção e a aplicação de sistemas baseados em LLMs ainda apresentam limitações relevantes. Em

alguns estudos, mesmo com os esforços para reduzir a intervenção humana, ela continua sendo necessária. Por exemplo, Toslali *et al*. (2024) destacam que o sistema desenvolvido não é totalmente confiável para decisões autônomas, devendo ser utilizado apenas como ferramenta de apoio. De forma semelhante, Carta *et al*. (2024) trabalharam com um conjunto reduzido de documentos, já que avaliar um volume maior exigiria um esforço humano considerável. Barriuso *et al*. (2024) reforçam que LLMs devem atuar apenas como apoio, especialmente em contextos sensíveis, como o da área da saúde, devido ao risco de alucinações. Por fim, Schaeffer *et al*. (2023) alertam que a automação excessiva pode comprometer o desempenho das ontologias desenvolvidas, exigindo intervenção humana em algumas etapas do processo.

Questões relacionadas ao custo computacional e às restrições de infraestrutura também foram recorrentes. Estudos como os de Carta *et al*. (2024), Siciliani *et al*. (2023) e Raja *et al*. (2024) mostram que limitações de *hardware* e recursos financeiros influenciaram decisões técnicas e afetaram a escalabilidade e o desempenho dos sistemas propostos. Em contrapartida, o trabalho de Ohr e Gupta (2024) utiliza uma infraestrutura computacional robusta, o que pode restringir a reprodutibilidade do estudo em ambientes com menor capacidade técnica e financeira.

Outro desafio identificado em pesquisas como as de Kruiper *et al*. (2024), Raja *et al*. (2024) e Wu *et al*. (2024)

refere-se ao uso de linguagem técnica e termos específicos de domínio, que podem resultar em interpretações incorretas, vieses semânticos e erros na recuperação de informações pelos modelos. Ainda, Sawarkar, Mangal e Solanki (2024) observam que a ausência de metadados nos conjuntos de dados limitou a eficácia de consultas híbridas em comparação às básicas em seu sistema, ressaltando a importância de informações estruturadas para melhorar a precisão dos resultados.

Dessa forma, embora os estudos analisados evidenciem avanços significativos no uso de LLMs para automação e apoio à recuperação da informação, ainda persistem lacunas importantes. Observa-se que a maioria das pesquisas se concentra em contextos com infraestrutura computacional robusta, o que limita a generalização dos resultados e reduz a acessibilidade dessas soluções para instituições com recursos técnicos e financeiros mais restritos. Além disso, a dimensão social da recuperação da informação permanece pouco explorada. Embora os LLMs possam aprimorar a busca semântica e contextual, ainda há desafios no atendimento a diferentes perfis de usuários, especialmente pessoas com deficiência ou aquelas que enfrentam barreiras culturais e linguísticas.

4.1 Uso de Large Language Models no contexto da Ciência da Informação

A Ciência da Informação (CI) surgiu no contexto das transformações científicas e tecnológicas que sucederam à Segunda Guerra Mundial, impulsionadas pelo crescimento acelerado da produção de informações científicas. Um dos marcos desse processo foi o artigo *As We May Think*, publicado por Vannevar Bush, em 1945, no qual o autor chamou a atenção para a dificuldade de acesso ao crescente volume de publicações científicas e defendeu o desenvolvimento de soluções tecnológicas capazes de facilitar sua organização, seu acesso e seu uso pela comunidade científica (Bush, 1945; Saracevic, 1996).

Nas décadas seguintes, pesquisadores de diferentes áreas passaram a buscar alternativas para enfrentar esse problema, contribuindo para a consolidação da CI como um campo interdisciplinar e fortemente relacionado ao desenvolvimento das tecnologias da informação. Nesse contexto, Calvin Mooers (1951, p. 25) criou o termo "recuperação da informação", definindo-o como um processo que "abrange tanto os aspectos intelectuais da descrição da informação e de sua especificação para fins de busca quanto os sistemas, técnicas ou máquinas empregados na execução dessa operação". Desde então, a Recuperação da Informação consolidou-se como uma das principais áreas da CI, evoluindo continuamente para responder aos desafios impostos pela crescente produção e circulação de informações (Saracevic, 1996).

Segundo Borko (1968), a CI dedica-se, desde sua origem, ao estudo dos processos relacionados à informação,

abrangendo sua organização, recuperação, interpretação e seu uso. O autor destaca que o campo investiga a representação da informação tanto em sistemas naturais quanto artificiais, bem como as técnicas e os dispositivos utilizados para seu processamento. Assim, a relação entre a informação e a tecnologia acompanha a área desde sua origem. Essa característica demonstra que a incorporação de tecnologias como os LLMs representa uma continuidade da evolução da área, e não uma ruptura com seus fundamentos teóricos.

Os estudos analisados mostram que os LLMs ampliam o conceito tradicional de recuperação da informação ao incorporarem mecanismos capazes de interpretar o significado das consultas e do conteúdo recuperado. Em vez de se limitarem à correspondência entre palavras-chave, como ocorre em muitos sistemas convencionais, as soluções identificadas na literatura utilizam técnicas de busca semântica, geração de respostas contextualizadas e integração entre diferentes formas de representação do conhecimento. Essa mudança aproxima o processo de recuperação das necessidades informacionais dos usuários, especialmente em contextos nos quais consultas em linguagem natural, documentos extensos ou terminologias especializadas dificultam a localização de informações relevantes.

Sob a perspectiva da CI, os estudos mostram que a eficiência da recuperação da informação continua dependendo da qualidade da organização e da representação da informação. Como apresentado no ID 10, limitações como linguagem

excessivamente técnica, ambiguidades terminológicas, ausência de metadados e estruturas documentais inadequadas podem comprometer a interpretação realizada pelos LLMs e reduzir a qualidade das respostas recuperadas. Dessa forma, embora esses modelos ampliem significativamente a capacidade de processamento semântico, eles permanecem dependentes de informações bem estruturadas e organizadas, demonstrando que os fundamentos tradicionais da Organização e da Representação da Informação continuam desempenhando um papel essencial no desempenho desses sistemas.

Outro aspecto observado é que nenhum dos estudos analisados propõe a substituição do profissional responsável pela análise da informação. Mesmo nos sistemas mais avançados, os LLMs são empregados como ferramentas de apoio à tomada de decisão, enquanto a validação das respostas permanece sob responsabilidade humana. Esse resultado sugere que a principal contribuição dos LLMs para a CI não está na automatização integral dos processos de recuperação, mas na ampliação da capacidade de localizar, organizar, interpretar e apresentar informações de forma mais eficiente, preservando a mediação humana em atividades que exigem avaliação crítica.

5 CONSIDERAÇÕES FINAIS

Este artigo teve como objetivo analisar, na literatura científica, o uso de Large Language Models (LLMs) na recuperação da informação em documentos textuais, buscando compreender como esses modelos vêm sendo aplicados e quais contribuições apresentam para o campo. A partir da revisão da literatura e da sistematização dos estudos selecionados, foi possível responder à pergunta de pesquisa ao identificar que os LLMs têm sido utilizados principalmente para aprimorar a busca semântica, expandir consultas, gerar respostas contextualizadas e integrar estratégias como o RAG, promovendo avanços em relação aos modelos tradicionais baseados na correspondência exata de palavras.

Observou-se que as principais contribuições identificadas na literatura se concentram na ampliação da capacidade interpretativa dos sistemas de busca, na redução de ruídos informacionais e na possibilidade de interação em linguagem natural, aproximando os sistemas das necessidades reais dos usuários.

Conclui-se que os LLMs representam uma transformação significativa no campo da recuperação da informação, ampliando as possibilidades de interação, contextualização e geração de respostas em sistemas documentais. No entanto, sua adoção requer análise crítica, planejamento técnico e reflexão ética, de modo que seu potencial seja explorado de forma responsável e sustentável.

Como limitações deste estudo, destaca-se que o *corpus* analisado foi composto por 12 artigos, selecionados de acordo

com os critérios metodológicos estabelecidos para a revisão. Embora esse conjunto tenha permitido identificar tendências, contribuições e limitações relacionadas ao uso de LLMs na recuperação da informação em documentos textuais, ele restringe a abrangência dos resultados e pode não contemplar toda a diversidade de pesquisas desenvolvidas sobre o tema. Além disso, considerando o ritmo acelerado de evolução dos LLMs, os resultados aqui apresentados podem sofrer atualizações, assim como parte das discussões desenvolvidas, demandando o monitoramento constante dessas inovações tecnológicas para a atualização da literatura da área.

Recomenda-se que pesquisas futuras aprofundem as investigações sobre vieses algorítmicos, transparência e impactos éticos associados ao uso desses modelos. No contexto de países emergentes, como o Brasil, sugere-se a realização de estudos que considerem as limitações de infraestrutura, os custos de implementação e a necessidade de capacitação profissional, de modo a promover aplicações mais adequadas às realidades locais.

REFERÊNCIAS

AMARATUNGA, Thimira. **Understanding Large Language Models**: learning their underlying concepts and technologies. Nugegoda: Apress, 2023.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação de informação**: conceitos e tecnologia das máquinas de busca. 2. ed. Porto Alegre: Bookman, 2013.

BARNETT, Scott *et al*. Seven Failure Points When Engineering a Retrieval Augmented Generation System. *In*: INTERNATIONAL CONFERENCE ON AI ENGINEERING – SOFTWARE ENGINEERING FOR AI, 3., 2024, Lisbon. **Proceedings** [...] Lisbon: ACM, 2024. p. 194-199. Disponível em: https://dl.acm.org/doi/10.1145/3644815.3644945. Acesso em: 1 ago. 2026.

BAWDEN, David; ROBINSON, Lyn. The dark side of information: overload, anxiety and other paradoxes and pathologies. **Journal of Information Science**, London, v. 35, n. 2, p. 180-191, 2008. Disponível em: https://journals.sagepub.com/doi/10.1177/0165551508095781. Acesso em: 1 ago. 2026.

BARRIUSO, Adrian Alónso *et al*. Recommendation system of scientific articles from discharge sumaries. **Engineering Applications of Artificial Intelligence**, London, v. 136, Part B, 109028, Oct. 2024. Disponível em: https://www.sciencedirect.com/science/article/pii/S0952197624011862. Acesso em: 1 ago. 2026.

BORKO, Harold. Information Science: what is it? **American Documentation**, [*s. l.*], p. 3-5, Jan. 1968. Disponível em: https://cmapspublic.ihmc.us/rid%3D1KR7TR64B-NMXB04-5SVP/BORKO(1968)-533107-Borko-H-v-19-n-1-p-35-1968.pdf. Acesso em: 31 jul. 2026.

BUSH, Vannevar. As We May Think. **Atlantic Monthy**, Washington, v. 176, n. 1, p. 101–108, Jul. 1945. Disponível em: https://www.theatlantic.com/magazine/archive/1945/07/as-wemay-think/303881/. Acesso em: 31 jul. 2026.

CARTA, Salvatore *et al*. A Zero-Shot Strategy for Knowledge Graph Engineering Using GPT-3.5. **Procedia Computer Science**, Netherlands, v. 246, Issue C, p. 2235–2243, 2024. Disponível em: https://www.sciencedirect.com/science/article/pii/S1877050924026231. Acesso em: 1 ago. 2026.

CAVALCANTE, Lívia Teixeira Canuto; OLIVEIRA, Adélia Augusta Souto de. Métodos de revisão bibliográfica nos estudos científicos. **Psicologia em revista**, Belo Horizonte, v. 26, n. 1, jan./abr. 2020. Disponível em: https://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-11682020000100006. Acesso em: 1 ago. 2026.

CLOTHIAUX, Daniel; STARZL, Ravi. Extractive Search for Analysis of Biomedical Texts. *In*: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 45., 2022, Madrid. **Proceedings** [...] Madrid: SIGIR, 2022. p. 3386 – 3387. Disponível em: https://dl.acm.org/doi/10.1145/3477495.3536328. Acesso em: 1 ago. 2026.

CONEGLIAN, Caio Saraiva; GONÇALVES, Paula Regina Ventura Amorim; SANTARÉM SEGUNDO, José Eduardo. O profissional da informação na era do big data. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 22, n. 50, p. 128-143, set./dez. 2017. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n50p128. Acesso em: 1 ago. 2026.

DHAMANI, Numa; ENGLER, Maggie. **Introduction to Generative AI**. Shelter Island: Manning, 2024.

ISSON, Jean Paul; HARRIOTT, Jesse S. **People Analytics in the Era of Big Data**: Changing the Way You Attract, Acquire, Develop, and Retain Talent. Hoboken: Wiley, 2016.

JESUS, Ananda Fernanda de; ALGARVE, Wesley; SANTARÉM SEGUNDO, José Eduardo. Dados sintéticos para treinamento de Inteligência Artificial: intersecções com a Organização e Representação da Informação e do Conhecimento. **Advances on Knowledge Representation Journal**, Belo Horizonte, v. 6, n. 1, 2026. Disponível em: https://periodicos.ufmg.br/index.php/advances-kr/article/view/64683. Acesso em: 1 ago. 2026.

KRUIPER, Ruben *et al*. A platform-based Natural Language processing-driven strategy for digitalising regulatory compliance processes for the built environment. **Advanced Engineering Informatics**, Netherlands, v. 62, Part B, 102653, p. 1-14, Oct. 2024. Disponível em: https://www.sciencedirect.com/science/article/pii/S1474034624003001X. Acesso em: 1 ago. 2026.

HACKITT, Dame Judith. **Building a Safer Future**: Independent Review of Building Regulations and Fire Safety: Final Report. UK: Crown, May 2018. Disponível em: https://assets.publishing.service.gov.uk/media/5afc50c840f0b622e4844ab4/Building_a_Safer_Future_-_web.pdf. Acesso em: 16 maio 2025.

MOOERS, Calvin Northrup. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, [*s. l.*], v. 2, n. 1, p. 20-32, Jan. 1951. Disponível em: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.5090020107. Acesso em: 31 jul. 2026.

SILVA, Patrícia Nascimento. Recuperação de informação na Ciência da Informação: produção acadêmico-científica brasileira (2012-2021). **Transinformação**, Campinas, v. 35, e237336, 2023. Disponível em: https://periodicos.puc-campinas.edu.br/transinfo/article/view/7336. Acesso em: 1 ago. 2026.

OHR, Vebjørn; GUPTA, Dhruv. Scalable Range Search over Temporal and Numerical Expressions. *In*: ACM SIGIR INTERNATIONAL CONFERENCE ON THE THEORY OF INFORMATION RETRIEVAL, 2024, Washington, DC. **Proceedings** [...] Washington, DC: ACM, 2024. p. 91-100. Disponível em: https://dl.acm.org/doi/10.1145/3664190.3672509. Acesso em: 1 ago. 2026.

OLIVEIRA, Stênio Henrique *et al*. Infoxicação em tempos de infodemia: revisão integrativa. **Brazilian Journal of Health Review**, Curitiba, v. 4, n. 6, p. 25199-25211, nov./dez. 2021.

Disponível em: https://ojs.brazilianjournals.com.br/ojs/index.php/BJHR/article/view/39660. Acesso em: 1 ago. 2026.

OZDEMIR, Sinan. **Quick Start Guide to Large Language Models**: Strategies and Best Practices for ChatGPT, Embeddings, Fine-Tuning, and Multimodal AI. 2nd ed. Hoboken: Pearson, 2025.

PROCÓPIO, Daiane Campos. **Large Language Models para recuperação da informação em acervos históricos digitalizados**. 2025. 154 f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2025. Disponível em: https://hdl.handle.net/1843/3339. Acesso em: 31 jul. 2026.

RAJA, Hina *et al*. Automated Category and Trend Analysis of Scientific Articles on Ophthalmology Using Large Language Models: Development and Usability Study. **JMIR Formative Research**, [*s.l.*], v. 8, e52462, p. 1-16, 2024. Disponível em: https://pubmed.ncbi.nlm.nih.gov/38517457/. Acesso em: 1 ago. 2026.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: https://periodicos.ufmg.br/index.php/pci/article/view/22308. Acesso em: 31 jul. 2026.

SAWARKAR, Kunal; MANGAL, Abhilasha; SOLANKI, Shivam Raj. Blended RAG: Improving RAG (Retriever-Augmented Generation) Accuracy with Semantic Search and Hybrid Query-Based Retrievers. *In*: INTERNATIONAL CONFERENCE ON MULTIMEDIA INFORMATION PROCESSING AND RETRIEVAL, 7., 2024, San Jose. **Proceedings** [...] San Jose: IEEE, 2024. p. 155-161. Disponível em: https://ieeexplore.ieee.org/document/10707868. Acesso em: 1 ago. 2026.

SCHAEFFER, Marion. OLAF: An Ontology Learning Applied Framework. **Procedia Computer Science**, Netherlands, v. 225, p. 2106-2115, 2023. Disponível em: https://www.sciencedirect.com/science/article/pii/S1877050923013595?via%3Dihub. Acesso em: 1 ago. 2026.

SICILIANI, Lucia *et al*. AI-based decision support system for public procurement. **Information Systems**, Oxford, v. 119, 102284, p. 1-16, Oct. 2023. Disponível em: https://www.sciencedirect.com/science/article/pii/S0306437923001205. Acesso em: 1 ago. 2026.

STRATIS, Kyle. **Whats is Generative AI?** A Generative AI Primer for Business and Technical Leaders. Sebastopol: O'Reilly, 2024.

TOSLALI, Mert *et al*. AgraBOT: Accelerating Third-Party Security Risk Management in Enterprise Setting through Generative AI. *In*: ACM INTERNATIONAL CONFERENCE ON THE FOUNDATIONS OF SOFTWARE ENGINEERING, 32., 2024, Porto de Galinhas. **Proceedings** [...] Porto de Galinhas: ACM, 2024. p. 74-79. Disponível em: https://dl.acm.org/doi/10.1145/3663529.3663829. Acesso em: 1 ago. 2026.

WU, Leihong *et al*. A framework enabling LLMs into regulatory environment for transparency and trustworthiness and its application to drug labeling document. **Regulatory Toxicology and Pharmacology**, New York, v. 149, 105613, p. 1-9, May 2024. Disponível em: https://www.sciencedirect.com/science/article/pii/S0273230024000540. Acesso em: 1 ago. 2026.

NOTAS

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: Procópio, Daiane Campos. Silva, Patrícia Nascimento. Souza, Renato Rocha.
Coleta de dados: Procópio, Daiane Campos.
Análise de dados: Procópio, Daiane Campos. Silva, Patrícia Nascimento. Souza, Renato Rocha.

Discussão dos resultados: Procópio, Daiane Campos. Silva, Patrícia Nascimento. Souza, Renato Rocha.
Revisão e aprovação: Silva, Patrícia Nascimento. Souza, Renato Rocha.

ORIGEM DA PESQUISA
Manuscrito derivado da Dissertação de Mestrado em Gestão e Organização do Conhecimento, com autoria de Daiane Campos Procópio, sob orientação de Renato Rocha Souza da Escola de Ciência da Informação da Universidade Federal de Minas Gerais, 2025, disponível no Repositório Institucional.

AGRADECIMENTOS
À Agência Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro concedido.

FINANCIAMENTO
Cooperação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) Código de financiamento 88887.948832/2024-00.

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio à pesquisa, processo 303721/2025-1.

USO DE INTELIGÊNCIA ARTIFICIAL
Não se aplica

CONSENTIMENTO DE USO DE IMAGEM
Não se aplica

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA
Não se aplica

CONFLITO DE INTERESSES
As pessoas autoras declaram não haver conflitos de interesse.

PUBLISHER
Universidade Federal de Sergipe. Programa de Pós-graduação em Ciência da Informação. Publicação no Portal de Periódicos UFS. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

COMO CITAR
PROCÓPIO, Daiane Campos. SILVA, Patrícia Nascimento. SOUZA, Renato Rocha. Large Language Models: contribuições para a recuperação da informação em documentos. **ConCI**: Convergências em Ciência da Informação, São Cristóvão, v. 9, p. e24689, 2026. DOI: 10.33467/conci.v9i.24689