



Information Research – Vol. 31 No. iConf (2026)

Large language models na ciência da informação: perspectivas técnicas da inteligência artificial aplicadas à organização e à arquitetura da informação (Large language models in information science: perspectives on information organisation and architecture)

Daiane Campos Procópio e Patrícia Nascimento Silva

DOI: <https://doi.org/10.47989/ir31iConf64289>

Abstract

Introduction. Despite the popularity of large language models, there are still gaps in understanding how information is organised, represented and structured within these systems. Thus, this study sought to identify and analyse aspects of information science (IS) in LLMs.

Method. This is a bibliographic study with a qualitative approach, conducted through a rigorous protocol applied to the ACM digital library, ScienceDirect, Scopus and Web of Science databases.

Analysis. A total of 53 studies published between 2022 and 2025 were examined, from which 16 categories of artificial intelligence techniques applied in LLMs were identified.

Results. These techniques were related to the four information architecture systems: organisation, labeling, navigation and search, and to classical instruments of information organisation, highlighting the convergence between principles of IS and AI innovations.

Conclusions. The articulation between information architecture and information organisation provides conceptual and practical foundations to understand how LLMs structure, represent and make data available. The study points to future research directions, reinforces the relevance of IS in the debate on AI and highlights the role of information professionals in critically mediating between technology and society.

Introdução

A ascensão dos *large language models* (LLMs) tem transformado a maneira como as pessoas acessam e interagem com a informação. No entanto, apesar de sua popularidade crescente, há uma lacuna entre o uso cotidiano dessas ferramentas e a compreensão técnica de seu funcionamento. Compreender como a informação é organizada, tratada e estruturada nesses sistemas é fundamental para promover uma cultura digital ética e transparente, implicando em pessoas mais críticas e informadas.

Essa lacuna de compreensão pode gerar riscos como a desinformação, a exposição indevida de dados pessoais, a dependência acrítica de respostas automatizadas e a redução da capacidade analítica dos usuários (Chen *et al.*, 2025; Chow, 2025; Ferreira & Oliveira, 2021; Zhou *et al.*, 2024). Além disso, o desempenho dos LLMs está diretamente relacionado à forma como os dados textuais são organizados e preparados durante seu treinamento. Neste contexto, torna-se essencial investigar os elementos que compõem a Organização e a arquitetura da informação em LLMs, sendo estes os questionamentos que norteiam esta pesquisa: Como a informação tem sido organizada, tratada e estruturada em LLMs? Quais técnicas de Inteligência Artificial (IA) são utilizadas nesse cenário?

Com isso, esta pesquisa objetiva identificar e analisar aspectos da ciência da informação (CI) em LLMs. Especificamente, buscou-se: 1) identificar estudos sobre a organização e a arquitetura da informação em LLMs e 2) discutir e relacionar aspectos da organização e da arquitetura da informação em LLMs com as técnicas de IA.

A pesquisa, de natureza bibliográfica, é uma revisão de literatura de abordagem qualitativa, que utilizou um protocolo minucioso para investigar a temática em quatro bases de dados, executado em junho de 2025.

A justificativa deste estudo reside na necessidade de ampliar o entendimento crítico sobre os LLMs e sua relação com os dados, apoiando novos projetos relacionados à governança de dados e às formas de recuperação envolvendo fontes externas. Além disso, tornar compreensível o modo como esses modelos são construídos e operam, especialmente no que se refere à estruturação e ao tratamento da informação, é essencial para fortalecer a alfabetização digital, fomentar o uso responsável da tecnologia e mitigar riscos sociais relacionados à desinformação e à privacidade.

É importante destacar que, embora a literatura ainda não estabeleça de forma explícita uma correspondência direta entre técnicas de IA e sistemas de arquitetura da informação, este trabalho propõe uma associação analítica entre esses domínios, a partir de princípios conceituais comuns relacionados à organização e ao acesso à informação.

Large language models

Segundo Kallens *et al.* (2023), LLMs são estruturas avançadas de redes neurais profundas, treinadas com grandes volumes de dados em linguagem natural, que lhes conferem a capacidade de executar múltiplas tarefas linguísticas. Os LLMs possuem ampla versatilidade, podendo gerar, resumir e traduzir textos, simular diálogos, entre outras tarefas. Além disso, eles realizam buscas semânticas, o que possibilita recuperar informações relevantes mesmo quando o usuário não utiliza as mesmas palavras na consulta (Wei *et al.*, 2025).

O desenvolvimento de um LLM ocorre em duas etapas principais: o pré-treinamento, que demanda grande volume de dados e recursos computacionais, no qual o modelo aprende padrões linguísticos gerais a partir de textos da *internet*, resultando nos chamados modelos-base; e o refinamento, fase em que esse modelo é ajustado para tarefas ou comportamentos específicos, aproveitando o conhecimento já adquirido e reduzindo custos em relação ao pré-treinamento (Alammar & Grootendorst, 2024).

Embora os LLMs tenham alcançado resultados importantes em diversas tarefas, ainda apresentam limitações relevantes, principalmente ligadas aos dados de treinamento, que podem absorver vieses sociais e estereótipos, refletindo desigualdades existentes em textos coletados da *internet*. Além disso, o volume de dados utilizados dificulta o controle de qualidade, aumentando o risco de incorporação de informações incorretas, ofensivas ou protegidas por direitos autorais. Esses problemas reforçam a necessidade de cautela no uso e na interpretação das respostas geradas pelos de LLMs, bem como de estratégias para mitigar esses problemas.

Arquitetura e organização da informação

Segundo Gabriel-Petit (2025), arquitetura da informação corresponde a uma prática de *design* voltada à definição da estrutura de ambientes digitais, com foco na organização das informações e no apoio à localização e usabilidade por meio de sistemas de rotulagem, navegação e pesquisa bem elaborados. Como ressalta Lima (2016, p. 48), sua principal função é *‘proporcionar uma estrutura lógica que possa ajudar o usuário a encontrar a informação de que necessita’*.

De acordo com Rosenfeld *et al.* (2015), a arquitetura da informação é estruturada em quatro sistemas fundamentais: organização, rotulação, navegação e busca. O sistema de organização diz respeito às formas de categorizar a informação; a rotulação trata da maneira como essa informação é representada; a navegação refere-se aos caminhos que permitem percorrê-la; e a busca está relacionada aos mecanismos utilizados para localizá-la.

A arquitetura da informação também se fundamenta nos princípios da organização da informação que, conforme Lima (2020), é a área responsável por estruturar e padronizar a representação de diferentes recursos informacionais, de modo a facilitar sua recuperação e seu uso, envolvendo tanto os processos quanto os instrumentos empregados para organizar conteúdos e garantindo que o conhecimento registrado seja acessível em sistemas de recuperação da informação e possa atender às necessidades específicas de uma comunidade de usuários, uma vez que a informação só adquire valor quando é devidamente representada e passível de ser recuperada.

Esses instrumentos, conhecidos como Sistemas de Organização do Conhecimento, são definidos por Santana *et al.* (2024, p. 3) como *‘instrumentos de representação utilizados para organizar materiais tanto objetivando a sua recuperação, como o gerenciamento da coleção’*. Entre esses sistemas, há as taxonomias, que podem ser compreendidas como uma estrutura hierárquica de termos que representa um determinado domínio do conhecimento, organizando os conceitos a partir de suas semelhanças e diferenças (Lima & Maculan, 2017); os tesouros, que são um conjunto estruturado de termos descritores que se relacionam de forma semântica e genérica, padronizados quanto à forma e ao significado e constituídos por palavras ou por grupos de palavras (Lima & Maculan, 2017); os vocabulários controlados, que são ferramenta utilizadas para padronizar e controlar os termos empregados em um sistema de informação (Cruz *et al.*, 2022); os metadados, que podem ser definidos como informações que descrevem outros dados, fornecendo elementos para sua identificação, representação, interoperabilidade e uso em sistemas de informação (Gilliland, 2016); e a indexação, que é o processo de identificar e representar o conteúdo informativo de documentos por meio da seleção de assuntos ou conceitos, organizando-os em índices com a finalidade de possibilitar sua posterior recuperação (Silva & Fujita, 2004), entre outros.

Metodologia

Neste estudo, buscou-se investigar a organização e a arquitetura da informação em LLMs a partir da análise de literatura especializada. Para isso, adotou-se uma abordagem qualitativa, de caráter exploratório e fundamentada em uma revisão de literatura (Creswell & Creswell, 2021). A opção por essa abordagem justifica-se pela natureza conceitual do objeto investigado, que envolve a análise de elementos da organização e da arquitetura da informação em LLMs, ainda em processo de consolidação teórica no campo da CI.

Ressalta-se que a relação entre as técnicas de IA e os sistemas de arquitetura da informação é aqui tratada como uma associação analítica construída pelas autoras, a partir da literatura examinada, e não como uma correspondência conceitual previamente estabelecida no campo, buscando, assim, identificar convergências conceituais.

Etapas da pesquisa

A condução metodológica da pesquisa foi estruturada em três etapas principais, descritas a seguir.

Primeira etapa: definição dos termos de busca

Na primeira etapa, procedeu-se à identificação e definição dos principais termos relacionados ao escopo da pesquisa, sendo eles *large language models*, arquitetura da informação, organização da informação e tratamento de dados. Esses termos orientaram a elaboração da *string* utilizada nas bases selecionadas, contemplando variações terminológicas em português, inglês e espanhol, de modo a ampliar a abrangência da recuperação dos estudos. A Figura 1 apresenta as *strings* utilizadas nas bases de dados para as buscas e a quantidade de publicações recuperadas. A busca foi realizada no dia 25/06/2025, utilizando o filtro de publicações de acesso aberto.

ACM Digital Library (Title)	[[Title: "modelos de linguagem de larga escala"] OR [Title: "grandes modelos de linguagem"] OR [Title: "large language models"] OR [Title: llm] OR [Title: llms] OR [Title: "modelos de lenguaje a gran escala"] OR [Title: "grandes modelos de lenguaje"] AND [[Title: "arquitetura da informação"] OR [Title: "information architecture"] OR [Title: "arquitectura de la información"] OR [Title: "organización de la información"] OR [Title: "information organization"] OR [Title: "organización de la información"] OR [Title: "tratamento de dados"] OR [Title: "data processing"] OR [Title: "tratamiento de datos"]]]	2
ACM Digital Library (Abstract)	[[Abstract: "modelos de linguagem de larga escala"] OR [Abstract: "grandes modelos de linguagem"] OR [Abstract: "large language models"] OR [Abstract: llm] OR [Abstract: llms] OR [Abstract: "modelos de lenguaje a gran escala"] OR [Abstract: "grandes modelos de lenguaje"] AND [[Abstract: "arquitetura da informação"] OR [Abstract: "information architecture"] OR [Abstract: "arquitectura de la información"] OR [Abstract: "organización de la información"] OR [Abstract: "information organization"] OR [Abstract: "organización de la información"] OR [Abstract: "tratamento de dados"] OR [Abstract: "data processing"] OR [Abstract: "tratamiento de datos"]]]	29
ACM Digital Library (Author keyword)	[[Keywords: "modelos de linguagem de larga escala"] OR [Keywords: "grandes modelos de linguagem"] OR [Keywords: "large language models"] OR [Keywords: llm] OR [Keywords: llms] OR [Keywords: "modelos de lenguaje a gran escala"] OR [Keywords: "grandes modelos de lenguaje"] AND [[Keywords: "arquitetura da informação"] OR [Keywords: "information architecture"] OR [Keywords: "arquitectura de la información"] OR [Keywords: "organización de la información"] OR [Keywords: "information organization"] OR [Keywords: "organización de la información"] OR [Keywords: "tratamento de dados"] OR [Keywords: "data processing"] OR [Keywords: "tratamiento de datos"]]]	4
ScienceDirect (Portuguese)	Title, abstract, keywords: (("Modelos de Linguagem de Larga Escala" OR "Grandes Modelos de Linguagem" OR LLM OR LLMs) AND ("Arquitetura da informação" OR "Organização da informação" OR "Tratamento de dados"))	0
ScienceDirect (English)	Title, abstract, keywords: (("Large Language Models" OR LLM OR LLMs) AND ("Information architecture" OR "Information organization" OR "Data processing"))	47
ScienceDirect (Spanish)	Title, abstract, keywords: (("Modelos de lenguaje a gran escala" OR "Grandes modelos de lenguaje" OR LLM OR LLMs) AND ("Arquitectura de la información" OR "Organización de la información" OR "Tratamiento de datos"))	0
Scopus	TITLE-ABS-KEY ((("Modelos de Linguagem de Larga Escala" OR "Grandes Modelos de Linguagem") OR ("Large Language Models" OR LLM OR LLMs) OR ("Modelos de lenguaje a gran escala" OR "Grandes modelos de lenguaje")) AND (("Arquitetura da informação" OR "Information architecture" OR "Arquitectura de la información") OR ("Organização da informação" OR "Information organization" OR "Organización de la información") OR ("Tratamento de dados" OR "Data processing" OR "Tratamiento de datos"))))	654
Web of Science	TS= (("Modelos de Linguagem de Larga Escala" OR "Grandes Modelos de Linguagem" OR "Large Language Models" OR LLM OR LLMs OR "Modelos de lenguaje a gran escala" OR "Grandes modelos de lenguaje") AND ("Arquitetura da informação" OR "Information architecture" OR "Arquitectura de la información" OR "Organização da informação" OR "Information organization" OR "Organización de la información" OR "Tratamento de dados" OR "Data processing" OR "Tratamiento de datos"))	163

Figura 1. Strings de busca.

Segunda etapa: elaboração do protocolo da revisão de literatura

Na segunda etapa, foi elaborado um protocolo para a condução da revisão de literatura, adaptado de Nascimento Silva (2023), com o objetivo de assegurar a transparência, consistência metodológica e reprodutibilidade do estudo. O protocolo definiu explicitamente: (i) o objetivo geral da revisão; (ii) as questões de pesquisa; (iii) as fontes de informação; (iv) os critérios de elegibilidade para a escolha dos documentos; (v) os critérios de inclusão e exclusão dos documentos; (vi) os campos e expressões de busca; e (vii) os procedimentos de seleção e análise dos documentos.

Foram selecionadas quatro bases de dados reconhecidas pela relevância e abrangência nas áreas de ciência da computação, tecnologia da informação e Pesquisa Multidisciplinar: ACM digital library, ScienceDirect, Scopus e Web of Science. A escolha dessas bases visou garantir a recuperação de estudos revisados por pares, com diversidade temática e rigor científico.

Não foi estabelecida delimitação temporal, considerando-se o caráter recente e dinâmico das pesquisas sobre LLMs. Foram incluídos artigos de periódicos e trabalhos completos publicados em eventos científicos, redigidos em português, inglês ou espanhol. Foram excluídas publicações duplicadas, estudos não disponíveis na íntegra via Portal de Periódicos da Capes e trabalhos cujo título e resumo não abordassem, de forma explícita, aspectos relacionados à arquitetura da informação, organização da informação ou técnicas de IA associadas aos LLMs.

O protocolo completo da revisão, com seus critérios e procedimentos, encontra-se sintetizado na Tabela 1.

Critérios	Descrição
Objetivo geral	Identificar publicações sobre os principais elementos da arquitetura da informação em LLMs.
Questões a serem resolvidas	Como a informação tem sido organizada, tratada e estruturada em LLMs?
	Quais técnicas de IA são utilizadas?
Fontes de informação	Base de dados: ACM Digital Library : por seu foco em ciência da computação e tecnologia da Informação. ScienceDirect : por reunir periódicos da Elsevier nas áreas de ciência e tecnologia. Scopus : pela cobertura multidisciplinar e ampla indexação de periódicos revisados por pares. Web of Science : por sua curadoria rigorosa e cobertura internacional de pesquisas de alto impacto.
Critérios de elegibilidade	Sem delimitação de data.
	Tipologia documental: artigos de periódicos e trabalhos de eventos, pois possuem revisão por pares.
	Idiomas: português, inglês e espanhol, a fim de aumentar a abrangência dos estudos.
Critérios de inclusão e de exclusão	Inclusão de publicações de áreas que estudam arquitetura da informação e LLMs.
	Inclusão de publicações nos idiomas inglês, português e espanhol.
	Exclusão de publicações duplicadas.
	Exclusão de publicações que não foram redigidas em português, inglês ou espanhol.
	Publicações cujo título e resumo não abordam arquitetura da informação e LLMs.

	Publicações que não estejam disponíveis na íntegra via Portal de Periódicos da Capes.
Campos de busca	Título, resumo e palavras-chave.
Expressões de busca	Large language models e suas variações E arquitetura da informação e suas variações. As expressões foram utilizadas em português, inglês e espanhol.
String	((('modelos de linguagem de larga escala' OR 'grandes modelos de linguagem') OR ('large language models' OR LLM OR LLMs) OR ('modelos de lenguaje a gran escala' OR 'grandes modelos de lenguaje')) AND (('arquitetura da informação' OR 'information architecture' OR 'arquitectura de la información' OR ('organização da informação' OR 'information organization' OR 'organización de la información') OR ('tratamento de dados' OR 'data processing' OR 'tratamiento de datos'))))
Seleção dos documentos recuperados	Leitura dos títulos e resumos das publicações recuperadas com o intuito de verificar a pertinência do conteúdo aos objetivos da pesquisa.
Procedimentos de análise	Leitura completa das publicações, registrada em planilha estruturada, a fim de identificar elementos sobre a organização e a arquitetura da informação em LLMs.
Critério de exclusão após análise dos documentos	Pesquisas que não possuem abordagem conceitual, teórica ou metodológica sobre LLMs e arquitetura da informação.
Tratamento	Software Parsifal, para a gestão do protocolo e das publicações, e Google Sheets, para o compartilhamento de dados e desenvolvimento do relatório da pesquisa entre as pesquisadoras.

Tabela 1. Protocolo da revisão de literatura.

Terceira etapa: execução das buscas, seleção e análise dos estudos

Na terceira etapa, as buscas foram executadas nas bases selecionadas, utilizando-se a *string* previamente definida. Os registros recuperados foram exportados para o *software* Parsifal (Disponível em: <https://parsif.al/>. Acesso em: 19 jun. 2025), empregado para a gestão do protocolo, organização dos resultados e aplicação dos critérios de elegibilidade, inclusão e exclusão.

Inicialmente, foram recuperados 882 registros nas bases de dados selecionadas. Após a identificação e remoção de 209 registros duplicados, procedeu-se à leitura dos títulos e resumos dos 673 documentos restantes, resultando na exclusão de 620 publicações que não atendiam aos critérios estabelecidos. As publicações consideradas pertinentes avançaram para a etapa de leitura integral. Ao final desse processo, a amostra final foi composta por 53 estudos.

A análise dos documentos selecionados foi conduzida de forma qualitativa, por meio de uma categorização facetada, a partir da leitura integral das publicações. As categorias facetadas foram definidas com base na identificação das técnicas de IA empregadas, indicadas nos documentos, bem como de aspectos relacionados à organização e arquitetura da informação associados aos LLMs. Esse procedimento permitiu classificar os estudos analisados e identificar padrões, além de mapear as principais técnicas de IA utilizadas.

A revisão de literatura foi conduzida entre os meses de junho e agosto de 2025, sendo o protocolo executado no dia 25/06/2025. As publicações selecionadas e seus respectivos identificadores, título, autores e ano de publicação estão detalhadas na Tabela 2.

Estudos (ID)	Título	Autores	Ano
1	A conceptual framework for a latest information-maintaining method using retrieval-augmented generation and a large language model in smart manufacturing: theoretical approach and performance analysis	Choi e Jeong	2025a
2	A Dutch financial large language model	Noels <i>et al.</i>	2024
3	A fine-tuning enhanced RAG system with quantized influence measure as AI judge	Rangan e Yin	2024
4	A framework enabling LLMs into regulatory environment for transparency and trustworthiness and its application to drug labeling document	Wu <i>et al.</i>	2024
5	A framework for LLM-assisted smart policing system	Sarzaeim <i>et al.</i>	2024
6	A method to promote safe cycling powered by large language models and IA agents	Costa <i>et al.</i>	2024
7	A multimodal large language model framework for intelligent perception and decision-making in smart manufacturing	Wang <i>et al.</i>	2025
8	A new pipeline for generating instruction dataset via RAG and self fine-tuning	Sung <i>et al.</i>	2024
9	A review on enhancing agricultural intelligence with large language models	Li <i>et al.</i>	2025
10	A scientific-article key-insight extraction system based on multiactor of fine-tuned open-source large language models	Song <i>et al.</i>	2025
11	A user-centered framework for data privacy protection using large language models and attention mechanisms	Zhou <i>et al.</i>	2024
12	Advancing EHR analysis predictive medication modeling using LLMs	Alghamdi e Mostafa	2025
13	Advancing Tinnitus therapeutictherapeutics: GPT-2 driven clustering analysis of cognitive behavioral therapy sessions and Google T5-based predictive modeling for THI score assessment	Jeong <i>et al.</i>	2024
14	Aggregated knowledge model enhancing domain-specific QA with fine-tuned and retrieval-augmented generation models	Liu <i>et al.</i>	2024
15	Application of retrieval-augmented generation for interactive industrial knowledge management via a large language model	Chen <i>et al.</i>	2025
16	Automating systematic literature reviews with retrieval-augmented generation: a comprehensive overview	Han <i>et al.</i>	2024
17	Batch-ICL: effective, efficient, and order-agnostic in-context learning	Zhang <i>et al.</i>	2024
18	Building lightweight domain-specific consultation systems via inter-external knowledge fusion contrastive learning	Zheng <i>et al.</i>	2024
19	Can language models automate data wrangling?	Jaimovitch-López <i>et al.</i>	2022

20	CMLLM: a novel cross-modal large language model for wind power forecasting	Zhu <i>et al.</i>	2025
21	Combining financial data and news articles for stock price movement prediction using large language models	Elahi e Taghvaei	2024
22	DataAgent: evaluating large language models' ability to answer zero-shot, natural language queries	Mishra <i>et al.</i>	2024
23	Deciphering public voices in the digital era: benchmarking ChatGPT for analyzing citizen feedback in Hamilton, New Zealand	Fu <i>et al.</i>	2024
24	Deep learning and methods based on large language models applied to Stellar light curve classification	Li <i>et al.</i>	2025
25	Development of a natural language processing (NLP) model to automatically extract clinical data from electronic health records: results from an Italian comprehensive stroke center	Badalotti <i>et al.</i>	2024
26	DocFinQA: a long-context financial reasoning dataset	Reddy <i>et al.</i>	2024
27	Domain-specific manufacturing analytics framework: an integrated architecture with retrieval-augmented generation and ollama-based models for manufacturing execution systems environments	Choi e Jeong	2025b
28	DRG-LLaMA: tuning LLaMA model to predict diagnosis-related group for hospitalized patients	Wang <i>et al.</i>	2024
29	Enhancing e-government services through state-of-the-art, modular, and reproducible architecture over large language models	Papageorgiou <i>et al.</i>	2024
30	Enhancing financial domain adaptation of language models via model augmentation	Tanabe <i>et al.</i>	2024a
31	Enriching building function classification using large language model embeddings of OpenStreetMap tags	Memduhoğlu <i>et al.</i>	2024
32	Extracting lung cancer staging descriptors from pathology reports: a generative language model approach	Cho <i>et al.</i>	2024
33	Fisher information-based efficient curriculum federated learning with large language models	Liu <i>et al.</i>	2024
34	Fishnet: financial intelligence from sub-querying, harmonizing, neural-conditioning, expert swarms, and task planning	Cho <i>et al.</i>	2024
35	JaFin: Japanese financial instruction dataset	Tanabe <i>et al.</i>	2024b
36	Large language models deconstruct the clinical intuition behind diagnosing autism	Stanley <i>et al.</i>	2025
37	Large language models for preventing medication direction errors in online pharmacies	Pais <i>et al.</i>	2024
38	LEGF-DST: LLMs-enhanced graph-fusion dual-stream transformer for fine-grained Chinese malicious SMS detection	Tong <i>et al.</i>	2025

39	LLM-assisted multi-teacher continual learning for visual question answering in robotic surgery	Du <i>et al.</i>	2024
40	LLM-assisted qualitative data analysis: security and privacy concerns in gamified workforce studies	Adeseyea <i>et al.</i>	2025
41	Military equipment entity extraction based on large language model	Liu <i>et al.</i>	2024
42	Model development for bespoke large language models for digital triage assistance in mental health care	Taylor <i>et al.</i>	2024
43	Multimodal medical image analysis: integrating LLM and RAG deep learning strategies	Yan e Shao	2025
44	Optimizing large language models for Arabic healthcare communication: a focus on patient-centered NLP applications	Mohammad <i>et al.</i>	2024
45	Privacy-preserving large language models for structured medical information retrieval	Wiest <i>et al.</i>	2024
46	RDguru: a conversational intelligent agent for rare diseases	Yang <i>et al.</i>	2024
47	Research on fine-tuning optimization strategies for large language models in tabular data processing	Zhao <i>et al.</i>	2024
48	TAT-LLM: a specialized language model for discrete reasoning over financial tabular and textual data	Zhu <i>et al.</i>	2024
49	Toward regulatory compliance: a few-shot learning approach to extract processing activities	Pragyan <i>et al.</i>	2024
50	Visual-textual integration in LLMs for medical diagnosis: A preliminary quantitative analysis	Agbareia <i>et al.</i>	2025
51	WaterGPT: training a large language model to become a hydrology expert	Ren <i>et al.</i>	2024
52	What did the occupant say? Fine-tuning and evaluating a large language model for efficient analysis of multi-domain indoor environmental	Sadick e Chinazzo	2025
53	Zero shot classification of art with large language models	Tojima e Yoshida	2025

Tabela 2. Publicações selecionadas.

Análise e discussão dos resultados

As pesquisas selecionadas, em sua maioria, apresentam estudos de caso que aplicam LLMs para solucionar problemas específicos ou promover melhorias em sistemas já existentes. Parte dos estudos precisou lidar com grandes volumes de dados nas etapas iniciais do desenvolvimento dos sistemas, como em Wu *et al.* (2024) e Elahi e Taghvaei (2024). Em alguns casos, havia dados com diferentes formatos, como em Noels *et al.* (2024) e Wang *et al.* (2025), e, em outros, coletados em tempo real, como visto em Choi e Jeong (2025a, 2025b). A utilização desses conjuntos de dados exigiu organização, rotulagem e limpeza, incluindo a remoção de dados duplicados e desnecessários para os objetivos dos sistemas propostos.

As abordagens adotadas nos estudos selecionados foram identificadas a partir da análise das técnicas de IA descritas em cada publicação e, posteriormente, classificadas em categorias

facetadas. Essas categorias foram organizadas segundo os quatro sistemas de arquitetura da informação propostos por Rosenfeld et al. (2015), sendo eles organização, rotulação, navegação e busca, conforme apresentado na Tabela 3. Já a Figura 2 ilustra como foi feito esse processo.

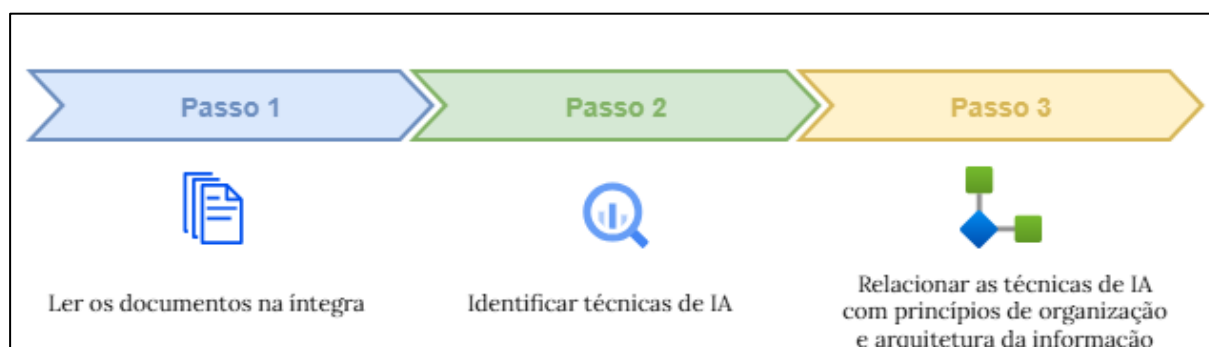


Figura 2. Processo para a categorização e associação dos estudos.

A associação entre as técnicas de IA e os sistemas de arquitetura da informação foi realizada com base em uma análise conceitual das funções desempenhadas por cada técnica no tratamento, na organização, na representação e no acesso à informação. Destaca-se que essa correspondência não é explicitamente formulada nos estudos analisados, mas constitui uma interpretação analítica das autoras, fundamentada nas características operacionais das técnicas identificadas e em sua proximidade com os instrumentos clássicos da Organização da Informação.

Sistema de Arquitetura da informação	Categorias (técnicas de IA)	Descrição	Estudos (ID)
Organização	<i>Continual learning</i> (aprendizado contínuo)	Abordagem que permite ao modelo aprender de forma incremental, incorporando novos dados e tarefas ao longo do tempo, sem esquecer conhecimentos previamente adquiridos.	39
	<i>Contrastive learning</i> (aprendizado contrastivo)	Método de treino em que o modelo aprende a aproximar exemplos que têm relação entre si e a afastar os que não têm, ajudando a melhorar buscas, classificações e o alinhamento semântico entre itens.	18
	<i>Data augmentation</i> (enriquecimento de dados)	Técnica para ampliar o conjunto de dados utilizados para o treinamento do modelo por meio de geração sintética, aumentando o conjunto sem a necessidade de coletar novos dados rotulados.	13, 38 e 41
	<i>Embeddings</i> (representações vetoriais)	Representações numéricas de textos, imagens ou outros dados em um espaço vetorial, permitindo medir a similaridade semântica entre esses elementos.	31
	<i>Fine-tuning</i> (ajuste fino ou refinamento)	Processo de ajustar a base de conhecimento de um modelo pré-treinado utilizando um conjunto de dados adicionais de um domínio, para que ele adquira maior especialização em determinado tema e/ou em uma tarefa em particular.	2, 3, 10, 12, 14, 24, 28, 32, 33, 36, 37, 41, 42, 44, 47, 48, 51 e 52

	<i>Model augmentation</i> (enriquecimento do modelo)	Uso de módulos ou ferramentas externas conectadas ao modelo para ampliar suas capacidades sem precisar treiná-lo novamente.	30
	<i>Pretraining</i> (pré-treinamento)	Etapa inicial de treinamento do modelo com grandes volumes de dados na qual o modelo aprende representações gerais antes de ser especializado.	7
Rotulação	<i>Contrastive learning</i> (aprendizado contrastivo)	Método de treino em que o modelo aprende a aproximar exemplos que têm relação entre si e a afastar os que não têm, ajudando a melhorar buscas, classificações e alinhamento semântico entre itens.	18
	<i>Prompt engineering</i> (engenharia de prompt)	Técnica voltada para a formulação de instruções que orientam o comportamento do modelo com o objetivo de gerar respostas mais alinhadas às necessidades do usuário e reduzir ocorrências de alucinação. Essa prática pode incluir, no <i>prompt</i> , o uso de exemplos e/ou cadeias de raciocínio.	6, 19, 21, 22, 23, 40, 45, 49, 50 e 53
	<i>In-context learning</i> (aprendizado em contexto)	Capacidade do modelo de executar uma tarefa a partir de instruções e poucos exemplos fornecidos no próprio <i>prompt</i> , durante a inferência, sem atualizar os pesos do modelo.	17 e 26
	<i>Instruction-tuning</i> (ajuste ou refinamento por instruções)	Processo em que o modelo é treinado ou refinado com pares de instruções e respostas para alinhar as respostas geradas a comportamentos desejados pelo usuário em tarefas de linguagem natural.	35
Navegação	<i>Active learning</i> (aprendizado ativo)	Estratégia em que o modelo escolhe os exemplos mais úteis ou incertos para serem revisados e categorizados por humanos, reduzindo a quantidade de dados necessária para treinar o modelo com qualidade.	25
	<i>Agentic architectures</i> (arquitetura agentiva)	Forma de organizar modelos para que funcionem como agentes capazes de agir sozinhos ou com pouca supervisão, interagir com ferramentas, dividir tarefas em etapas e tomar decisões com base em objetivos.	34
	<i>LLMs architecture</i> (arquitetura de LLMs)	Organização interna do modelo, normalmente em variantes <i>decoder-only</i> (apenas decodificação) ou <i>encoder-decoder</i> (codificador-decodificador), que define a sua capacidade, o seu número de camadas e a janela de contexto (espaço onde o usuário insere o <i>prompt</i>).	9
	<i>Cross-modal learning</i> (aprendizado intermodal)	Técnica que combina diferentes tipos de dados, como texto, imagem ou áudio, permitindo que o modelo realize tarefas entre variadas fontes, como buscar uma imagem a partir de uma descrição em texto.	20
	<i>Retrieval-augmented generation</i> (geração aumentada de recuperação)	Método que aprimora as respostas de um modelo ao recuperar informações de fontes externas e incorporá-las como contexto nos <i>prompts</i> , permitindo que o modelo forneça respostas específicas sem a	1, 3, 4, 5, 8, 14, 15, 16, 27, 29, 43 e 46

		necessidade de um novo treinamento.	
Busca	<i>Embeddings</i> (representações vetoriais)	Representações numéricas de textos, imagens ou outros dados em um espaço vetorial, permitindo medir a similaridade semântica entre esses elementos.	31
	<i>Attention mechanism</i> (mecanismo de atenção)	Componente que identifica a importância de cada <i>token</i> em relação aos demais, permitindo ao modelo usar o contexto e compreender dependências mesmo em trechos longos.	11
	<i>Retrieval-augmented generation</i> (geração aumentada de recuperação)	Método que aprimora as respostas de um modelo ao recuperar informações de fontes externas e incorporá-las como contexto nos <i>prompts</i> , permitindo que o modelo forneça respostas específicas sem a necessidade de um novo treinamento.	1, 3, 4, 5, 8, 14, 15, 16, 27, 29, 43 e 46

Tabela 3. Categorias e estudos organizados de acordo com os sistemas de arquitetura da informação.

Algumas técnicas de IA aparecem em mais de um sistema de arquitetura da informação porque podem atuar em diferentes etapas do tratamento da informação. Por exemplo, *embeddings* são úteis tanto na organização, ao estruturar representações vetoriais para alinhar itens semelhantes, quanto na busca, ao medir a proximidade semântica entre consultas e documentos. Da mesma forma, *contrastive learning* contribui para organizar dados aproximando exemplos relacionados, mas também auxilia na rotulação, ao tornar as categorias mais claras e distinguíveis. Já o RAG aparece na navegação e na busca, pois é capaz de integrar informações externas ao contexto do LLM para oferecer respostas mais contextualizadas e precisas.

Segundo Rosenfeld et al. (2015), os **sistemas de organização** envolvem esquemas, que determinam as características comuns entre os itens para posterior agrupamento, e estruturas, que estabelecem as relações entre esses itens e seus grupos, influenciando a forma como a informação é localizada e compreendida. Nesse âmbito, técnicas de IA como *fine-tuning*, *pretraining*, *data augmentation*, *embeddings*, *contrastive learning*, *continual learning* e *model augmentation* cumprem funções semelhantes a taxonomias e tesouros ao estruturarem dados, refinarem categorias e enriquecerem representações semânticas. Nos estudos de Rangan e Yin (2024), por exemplo, foram utilizados dados de uma única fonte, extraídos de documentos em PDF, e criados pares de perguntas e respostas para refinar um LLM para atuar como um *chatbot* que respondesse a perguntas sobre serviços sociais para pessoas em situação de vulnerabilidade.

Já os **sistemas de rotulação** dizem respeito à representação da informação e funcionam como a forma mais evidente de revelar os esquemas adotados. Para serem eficazes, devem refletir a linguagem dos usuários e representar os conteúdos com clareza. Nesse sentido, instrumentos como vocabulários controlados encontram paralelos em técnicas como *prompt engineering*, *instruction-tuning*, *contrastive learning* e *in-context learning*, pois tanto os vocabulários controlados quanto essas técnicas de IA têm como objetivo principal reduzir ruídos e tornar a representação mais precisa, aproximando a linguagem do sistema da linguagem do usuário. Em suas investigações, Elahi e Taghvaei (2024) utilizaram *prompt engineering* em dados financeiros, combinando informações rotuladas, como a descrição da empresa e seu setor, para contextualizar e refinar as respostas do modelo.

Os **sistemas de navegação** têm como objetivo orientar os usuários em ambientes informacionais, fornecendo contexto, localização e caminhos de retorno. Assim, as técnicas de RAG, *agentic architectures*, *cross-modal learning*, *active learning* e *LLM architectures* encontram correspondências com tesauros, taxonomias e metadados hierárquicos, pois ambos funcionam como mecanismos que orientam o usuário, permitindo localizar, relacionar e retornar a informações relevantes dentro de um sistema. Choi e Jeong (2025a) elaboraram um modelo conceitual baseado em RAG para uso na manufatura inteligente. A arquitetura proposta possibilitou a coleta de dados em tempo real a partir de sensores industriais, a preservação do contexto temporal e semântico e a atualização constante do conhecimento, permitindo que os usuários navegassem pelos dados de forma contextualizada e dinâmica.

Por fim, os **sistemas de busca**, essenciais para a localização de informações, tradicionalmente dependem de esquemas de indexação, vocabulários controlados e metadados. Esses instrumentos se relacionam com técnicas como *embeddings*, *attention mechanism* e RAG, que oferecem precisão e relevância nas respostas. Nos estudos de Zhou et al. (2024), o uso de *attention mechanism* demonstrou que é possível recuperar informações relevantes sem comprometer dados sensíveis, garantindo segurança na apresentação dos resultados. Memduhoğlu et al. (2024) aplicaram *embeddings* para enriquecer a classificação funcional de edificações urbanas. Com a integração de atributos físicos, espaciais e morfológicos, os autores demonstraram que os *embeddings* tornaram a busca e a recuperação mais precisas, superando técnicas tradicionais de processamento de linguagem natural.

Dessa forma, foi possível observar a conexão entre os sistemas de arquitetura da informação, os instrumentos clássicos de organização, como taxonomias, tesauros, vocabulários controlados e metadados, e as técnicas contemporâneas de IA, aplicadas ao tratamento de grandes volumes de dados. Essa relação não apenas destaca os fundamentos da CI, mas também os integra às demandas atuais da IA, oferecendo uma base teórica e prática para o desenvolvimento de soluções mais eficientes e seguras.

Considerações finais

Este estudo teve como objetivo identificar e analisar aspectos da CI relacionados à organização e à arquitetura da informação em LLMs, discutindo a relação entre os fundamentos da área e as técnicas de IA descritas na literatura. A partir dos resultados obtidos, considera-se que o objetivo proposto foi alcançado.

Por meio de uma revisão de literatura, o trabalho mapeou técnicas de IA utilizadas em LLMs e as organizou segundo os quatro sistemas da arquitetura da informação propostos por Rosenfeld et al. (2015), relacionando-as aos instrumentos da Organização da Informação. Essa análise evidenciou que os princípios da CI oferecem uma base teórica para compreender como os LLMs organizam e representam grandes volumes de dados, sendo úteis especialmente em contextos que envolvem governança e recuperação da informação.

No que se refere às contribuições para a CI, o estudo oferece subsídios teóricos ao propor uma aproximação conceitual entre técnicas de IA e os sistemas de arquitetura da informação, ampliando o escopo de aplicação desses sistemas para além dos ambientes informacionais tradicionais. Ainda, apresenta uma contribuição metodológica ao demonstrar a aplicabilidade da categorização facetada como estratégia analítica para o estudo de tecnologias emergentes, favorecendo análises sistemáticas e comparáveis. Finalmente, o trabalho também contribui para a atualização do repertório conceitual da CI ao incorporar técnicas como *embeddings*, *prompt engineering* e RAG ao debate sobre organização, representação e recuperação da informação.

Adicionalmente, o estudo reforça o papel da CI nos debates sobre IA, mostrando a relevância do profissional da informação como agente mediador entre tecnologia e sociedade. Ao favorecer a

compreensão dos mecanismos relacionados aos LLMs, o trabalho também contribui para a alfabetização informacional e digital, apoiando o uso crítico, ético e consciente de sistemas de IA generativa.

Como limitação, destaca-se que, por se tratar de uma temática recente e em rápida evolução, o estudo apresenta uma análise de caráter exploratório, indicando a necessidade de investigações futuras que aprofundem empiricamente as relações aqui discutidas e ampliem o *corpus* analisado.

Agradecimento

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), projeto APQ-01609-24.

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

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

Sobre as autoras

Daiane Campos Procópio é mestra em Gestão e Organização do Conhecimento pela Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil. Seus interesses de pesquisa incluem gestão e recuperação da informação, inteligência artificial e desenvolvimento de soluções tecnológicas aplicadas à disseminação do conhecimento. Pode ser contatada em campos-daiane@ufmg.br

Patrícia Nascimento Silva: Professora Adjunta na Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais (UFMG) - Área: tecnologia, organização e tratamento da informação. Professora e Pesquisadora no Programa de Pós-Graduação em Gestão & Organização do Conhecimento (PPGGOC) ECI/UFMG. Bolsista de Produtividade do CNPq. Fundadora e líder do grupo de pesquisa: Observatório de Dados Abertos. Representante do Comitê Gestor da Infraestrutura Nacional de Dados Abertos (CGINDA). Integra a comissão permanente de inteligência artificial da UFMG. Doutora em Gestão e Organização do Conhecimento pelo PPGOC ECI UFMG, Mestre e Bacharel em Sistemas de Informação. Atuou como Analista de Sistemas por 15 anos na área de engenharia de software. Experiência e interesse de pesquisa na área de ciência da informação e sistemas de informação, na linha de gestão e tecnologia, com destaque para: recuperação de informação, representação e organização da informação e do conhecimento, inteligência artificial, interoperabilidade, microsserviços e *application programming interface*, dados abertos e *open government data*, métodos ágeis e governança de dados. Pode ser contatada em patricians@ufmg.br

Referências

- Adeseye, A., Isoaho, J., & Mohammad, T. (2025). LLM-assisted qualitative data analysis: Security and privacy concerns in gamified workforce studies. *Procedia Computer Science*, 257, 60–67. <https://doi.org/10.1016/j.procs.2025.03.011>
- Agbareia, R., Omar, M., Soffer, S., Glicksberg, B. S., Nadkarni, G. N., & Klang, E. (2025). Visual-textual integration in LLMs for medical diagnosis: A preliminary quantitative analysis. *Computational and Structural Biotechnology Journal*, 27, 184–189. <https://doi.org/10.1016/j.csbj.2024.12.019>
- Alammar, J., & Groendorst, M. (2024). *Hands-on large language models*. O'Reilly.
- Alghamdi, H., & Mostafa, A. (2025). Advancing EHR analysis: Predictive medication modeling using LLMs. *Information Systems*, 131, 102528. <https://doi.org/10.1016/j.is.2025.102528>

- Badalotti, D., Agrawal, A., Pensato, U., Angelotti, G., & Marcheselli, S. (2024). Development of a natural language processing (NLP) model to automatically extract clinical data from electronic health records: Results from an Italian comprehensive stroke center. *International Journal of Medical Informatics*, 192, 105626. <https://doi.org/10.1016/j.ijmedinf.2024.105626>
- Chen, K., Zhou, X., Lin, Y., Feng, S., Shen, L., & Wu, P. (2025). A survey on privacy risks and protection in large language models. *arXiv*, 2505.01976v1. <https://doi.org/10.48550/arXiv.2505.01976>
- Chen, L. C., Pardeshi, M. S., Liao, Y. X., & Pai, K. C. (2025). Application of retrieval-augmented generation for interactive industrial knowledge management via a large language model. *Computer Standards & Interfaces*, 94, 103995. <https://doi.org/10.1016/j.csi.2025.103995>
- Cho, H., Yoo, S., Kim, B., Jang, S., Sunwoo, L., Kim, S., Lee, D., Kim, S., Nam, S., & Chung, J.-H. (2024). Extracting lung cancer staging descriptors from pathology reports: A generative language model approach. *Journal of Biomedical Informatics*, 157, 104720. <https://doi.org/10.1016/j.jbi.2024.104720>
- Cho, N., Srishankar, N., Cecchi, L., & Watson, W. (2024). FISHNET: Financial intelligence from sub-querying, harmonizing, neural-conditioning, expert swarms, and task planning. *Proceedings of the ACM International Conference on AI in Finance*, 5, 591-599. <https://doi.org/10.1145/3677052.3698597>
- Choi, H. & Jeong, J. (2025a). A conceptual framework for a latest information-maintaining method using retrieval-augmented generation and a large language model in smart manufacturing: Theoretical approach and performance analysis. *Machines*, 13(2), 94. <https://doi.org/10.3390/machines13020094>
- Choi, H. & Jeong, J. (2025b). Domain-specific manufacturing analytics framework: An integrated architecture with retrieval-augmented generation and Ollama-based models for manufacturing execution systems environments. *Processes*, 13(3), 670. <https://doi.org/10.3390/pr13030670>
- Chow, A. R. (2025, June 23). ChatGPT may be eroding critical thinking skills, according to a new MIT study. *Time*. <https://time.com/7295195/ai-chatgpt-google-learning-school/> (Internet Archive)
- Costa, D. G., Silva, I., Medeiros, M., Bittencourt, J. C. N., & Andrade, M. (2024). A method to promote safe cycling powered by large language models and AI agents. *MethodsX*, 13, 102880. <https://doi.org/10.1016/j.mex.2024.102880>
- Creswell, J. W. & Creswell, J. D. (2021). *Projeto de pesquisa: Métodos qualitativo, quantitativo e misto* (5ª ed.). Penso.
- Cruz, M. C. A., Ferneda, E., & Fujita, M. S. L. (2022). A disponibilização de vocabulário controlado aos usuários para a recuperação da informação. *RICI: Revista Íbero-Americana de ciência da informação*, 15(1), 266-282. <https://doi.org/10.26512/rici.v15.n1.2022.42464>
- Du, Y., Chen, K., Zhan, Y., Low, C. H., You, T., Islam, M., Guo, Z., Jin, Y., Chen, G., & Heng, P.-A. (2024). LLM-assisted multi-teacher continual learning for visual question answering in robotic surgery. *arXiv*, 2402.16664v3. <https://doi.org/10.48550/arXiv.2402.16664>
- Elahi, A. & Taghvaei, F. (2024). Combining financial data and news articles for stock price movement prediction using large language models. *arXiv*, 2411.01368v1. <https://doi.org/10.48550/arXiv.2411.01368>

- Ferreira, S. A. & Oliveira, D. A. (2021). Desinformação, crise de confiança na ciência e necessidade de políticas para divulgação científica. *Ciência da Informação Express*, 2(4), 1–6.
<https://cienciadainformacaoexpress.ufla.br/index.php/revista/article/view/28/63>
 (Internet Archive)
- Fu, X., Sanchez, T. W., Li, C., & Junqueira, J. R. (2024). Deciphering public voices in the digital era: Benchmarking ChatGPT for analyzing citizen feedback in Hamilton, New Zealand. *Journal of the American Planning Association*, 90(4), 728–741.
<https://doi.org/10.1080/01944363.2024.2309259>
- Gabriel-Petit, P. (2025). *Designing information architecture: A practical guide to structuring digital content for findability and easy navigability*. Packt Publishing.
- Gilliland, A. J. (2016). Setting the stage. In M. Baca (Ed.), *Introduction to metadata* (3rd ed.). Getty Research Institute. <https://www.getty.edu/publications/intrometadata/setting-the-stage/> (Internet Archive)
- Han, B., Susnjak, T., & Mathrani, A. (2024). Automating systematic literature reviews with retrieval-augmented generation: A comprehensive overview. *Applied Sciences*, 14(19), 9103.
<https://doi.org/10.3390/app14199103>
- Jaimovitch-López, G., Ferri, C., Hernández-Orallo, J., Martínez-Plumed, F., & Ramírez-Quintana, M. J. (2022). Can language models automate data wrangling?. *Machine Learning*, 112, 2053–2082. <https://doi.org/10.1007/s10994-022-06259-9>
- Jeong, Y., Song, J.-J., Yang, J., & Kang, S. (2024). Advancing tinnitus therapeutics: GPT-2 driven clustering analysis of cognitive behavioral therapy sessions and Google T5-based predictive modeling for THI score assessment. *IEEE Access*, 12, 52414–52429.
<https://doi.org/10.1109/ACCESS.2024.3383020>
- Kallens, P. C., Kristensen-McLachlan, R. D., & Christiansen, M. H. (2023). Large language models demonstrate the potential of statistical learning in language. *Cognitive Science*, 47(3), e13256. <https://doi.org/10.1111/cogs.13256>
- Li, H., Wu, H., Li, Q., & Zhao, C. (2025). A review on enhancing agricultural intelligence with large language models. *Artificial Intelligence in Agriculture*, 15(4), 671–685.
<https://doi.org/10.1016/j.aiaa.2025.05.006>
- Li, Y. Y., Bai, Y., Wang, C., Qu, M., Lu, Z., Soria, R., & Liu, J. (2025). Deep learning and methods based on large language models applied to stellar light curve classification. *Intelligent Computing*, 4, 0110. <https://doi.org/10.34133/icomputing.0110>
- Lima, G. A. (2016). Arquitetura da informação. In R. C. R. Miranda (Ed.), *arquitetura da informação na Câmara dos Deputados*. Câmara dos Deputados, Edições Câmara.
<http://bd.camara.leg.br/bd/handle/bdcamara/30199> (Internet Archive)
- Lima, G. A. (2020). Organização e representação do conhecimento e da informação na web: teorias e técnicas. *Perspectivas em ciência da informação*, 25, 57–97.
<https://periodicos.ufmg.br/index.php/pci/article/view/22283> (Internet Archive)
- Lima, G. A. & Maculan, B. C. M. S. (2017). Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento. *Ciência da Informação*, 46(1), 60–72.
<https://revista.ibict.br/ciinf/article/view/4014> (Internet Archive)
- Liu, F., Jung, J., Feinstein, W., D'Ambrogia, J., & Jung, G. (2024). Aggregated knowledge model: Enhancing domain-specific QA with fine-tuned and retrieval-augmented generation

- models. Proceedings of the International Conference on AI-ML Systems, 4, 1–7.
<https://doi.org/10.1145/3703412.3703434>
- Liu, J., Ren, J., Jin, R., Zhang, Z., Zhou, Y., Valduriez, P., & Dou, D. (2024). Fisher information-based efficient curriculum federated learning with large language models. arXiv, 2410.00131v2.
<https://doi.org/10.48550/arXiv.2410.00131>
- Liu, X., Yu, Z., Liu, X., Miao, L., & Yang, T. (2024). Military equipment entity extraction based on large language model. Applied Sciences, 14(19), 9063.
<https://doi.org/10.3390/app14199063>
- Memduhoğlu, A., Fulman, N., & Zipf, A. (2024). Enriching building function classification using large language model embeddings of OpenStreetMap tags. Earth Science Informatics, 17, 5403–5418. <https://doi.org/10.1007/s12145-024-01463-8>
- Mishra, M., Braham, A., Marsom, C., Chung, B., Griffin, G., Sidnerlikar, D., Sarin, C., & Rajaram, A. (2024). DataAgent: Evaluating large language models' ability to answer zero-shot, natural language queries. arXiv, 2404.00188v1. <https://doi.org/10.48550/arXiv.2404.00188>
- Mohammad, R., Alkhnbashi, O. S., & Hammoudeh, M. (2024). Optimizing large language models for Arabic healthcare communication: A focus on patient-centered NLP applications. Big Data and Cognitive Computing, 8(11), 157. <https://doi.org/10.3390/bdcc8110157>
- Nascimento Silva, P. (2023). Recuperação de informação na ciência da informação: Produção acadêmico-científica brasileira (2012–2021). Transinformação, 35, e237336.
<https://doi.org/10.1590/2318-0889202335e237336>
- Noels, S., De Blaere, J., & De Bie, T. (2024). A Dutch financial large language model. Proceedings of the ACM International Conference on AI in Finance, 5, 283–291.
<https://doi.org/10.1145/3677052.3698628>
- Papageorgiou, G., Sarlis, V., Maragoudakis, M., & Tjortjis, C. (2024). Enhancing e-government services through state-of-the-art, modular, and reproducible architecture over large language models. Applied Sciences, 14(18), 8259. <https://doi.org/10.3390/app14188259>
- Pais, C., Liu, J., Voigt, R., Gupta, V., Wade, E., & Bayati, M. (2024). Large language models for preventing medication direction errors in online pharmacies. Nature Medicine, 30(6), 1574–1582. <https://doi.org/10.1038/s41591-024-02933-8>
- Pragyan, K. C., Ghandiparsi, R., Slavin, R., Ghanavati, S., Breaux, T., & Hosseini, M. B. (2024). Toward regulatory compliance: A few-shot learning approach to extract processing activities. Proceedings of the IEEE International Requirements Engineering Conference Workshops, 32, 241–248. <https://doi.org/10.1109/REW61692.2024.00038>
- Rangan, K. & Yin, Y. (2024). A fine-tuning enhanced RAG system with quantized influence measure as AI judge. Scientific Reports, 14, 27446. <https://doi.org/10.1038/s41598-024-79110-x>
- Reddy, V., Koncel-Kedziorski, R., Lai, V. D., Krumdick, M., Lovering, C., & Tanner, C. (2024). DocFinQA: A long-context financial reasoning dataset. arXiv, 2401.06915v3.
<https://doi.org/10.48550/arXiv.2401.06915>
- Ren, Y., Zhang, T., Dong, X., Li, W., Wang, Z., He, J., Zhang, H., & Jiao, L. (2024). WaterGPT: Training a large language model to become a hydrology expert. Water, 16(21), 3075.
<https://doi.org/10.3390/w16213075>
- Rosenfeld, L., Morville, P., & Arango, J. (2015). Information architecture (4th ed.). O'Reilly.

- Sadick, A. M. & Chinazzo, G. (2025). What did the occupant say? Fine-tuning and evaluating a large language model for efficient analysis of multi-domain indoor environmental quality feedback. *Building and Environment*, 274, 112735. <https://doi.org/10.1016/j.buildenv.2025.112735>
- Santana, L. D., Martins, R. M., Chagas, L. B. R., Pereira, F. C. M., Lima, G. A., & Moura, M. A. (2024). Sistemas de Organização do Conhecimento: análise comparativa e modelagem de instrumentos de representação do conhecimento. *Encontros Bibli*, 29, e97708, 1-27. <https://doi.org/10.5007/1518-2924.2024.e97708>
- Sarzaeim, P., Mahmoud, Q. H., & Azim, A. (2024). A framework for LLM-assisted smart policing system. *IEEE Access*, 12, 74915–74929. <https://doi.org/10.1109/ACCESS.2024.3404862>
- Silva, M. R. & Fujita, M. S. L. (2004). A prática da indexação: análise da evolução de tendências teóricas e metodológicas. *Transinformação*, 16(2), 133-161. <https://periodicos.puc-campinas.edu.br/transinfo/article/view/6373> (Internet Archive)
- Song, Z., Hwang, G. Y., Zhang, X., Huang, S., & Park, B. K. (2025). A scientific-article key-insight extraction system based on multi-actor of fine-tuned open-source large language models. *Scientific Reports*, 15, 1608. <https://doi.org/10.1038/s41598-025-85715-7>
- Stanley, J., Rabot, E., Reddy, S., Belilovsky, E., Mottron, L., & Bzdo, D. (2025). Large language models deconstruct the clinical intuition behind diagnosing autism. *Cell*, 188(8), 2235–2248. <https://doi.org/10.1016/j.cell.2025.02.025>
- Sung, C., Lee, Y., & Tsai, Y. (2024). A new pipeline for generating instruction dataset via RAG and self fine-tuning. *Proceedings of the IEEE Annual Computers, Software, and Applications Conference*, 48, 2308-2312. <https://doi.org/10.1109/COMPSAC61105.2024.00371>
- Tanabe, K., Hirano, M., Matoya, K., Imajo, K., Sakaji, H., & Noda, I. (2024a). Enhancing financial domain adaptation of language models via model augmentation. *arXiv*, 2411.09249v1. <https://doi.org/10.48550/arXiv.2411.09249>
- Tanabe, K., Suzuki, M., Sakaji, H., & Noda, I. (2024b). JaFin: Japanese financial instruction dataset. *Proceedings of the IEEE Symposium on Computational Intelligence for Financial Engineering and Economics*, 1-10. <https://doi.org/10.1109/CIFEr62890.2024.10772973>
- Taylor, N., Kormilitzin, A., Lorge, I., Nevado-Holgado, A., Cipriani, A., & Joyce, D. W. (2024). Model development for bespoke large language models for digital triage assistance in mental health care. *Artificial Intelligence in Medicine*, 157, 102988. <https://doi.org/10.1016/j.artmed.2024.102988>
- Tojima, T. & Yoshida, M. (2025). Zero-shot classification of art with large language models. *IEEE Access*, 13, 17426–17435. <https://doi.org/10.1109/ACCESS.2025.3532995>
- Tong, X., Wang, J., Yang, Y., Peng, T., Zhai, H., & Ling, G. (2025). LEGF-DST: LLMs-enhanced graph-fusion dual-stream transformer for fine-grained Chinese malicious SMS detection. *Computers, Materials & Continua*, 82(2), 1902–1919. <https://doi.org/10.32604/cmc.2024.059018>
- Wang, H., Gao, C., Dantona, C., Hull, B., & Sun, J. (2024). DRG-LLaMA: Tuning LLaMA model to predict diagnosis-related group for hospitalized patients. *npj Digital Medicine*, 7, 16. <https://doi.org/10.1038/s41746-023-00989-3>

- Wang, T., Zhang, B., Jiang, D., & Li, D. (2025). A multimodal large language model framework for intelligent perception and decision-making in smart manufacturing. *Sensors*, 25(10), 3072. <https://doi.org/10.3390/s25103072>
- Wiest, I. C., Ferber, D., Zhu, J., van Treeck, M., Meyer, S. K., Juglan, R., Carrero, Z. I., Paech, D., Kleesiek, J., Ebert, M. P., Truhn, D., & Kather, J. N. (2024). Privacy-preserving large language models for structured medical information retrieval. *npj Digital Medicine*, 7, 257. <https://doi.org/10.1038/s41746-024-01233-2>
- Wu, L., Xu, J., Thakkar, S., Gray, M., Qu, Y., Li, D., & Tong, W. (2024). A framework enabling LLMs into regulatory environment for transparency and trustworthiness and its application to drug labeling document. *Regulatory Toxicology and Pharmacology*, 149, 105613. <https://doi.org/10.1016/j.yrtph.2024.105613>
- Yan, H. & Shao, D. (2025). Multimodal medical image analysis: Integrating LLM and RAG deep learning strategies. *Journal of Advances in Information Technology*, 16(4), 568–581. <https://doi.org/10.12720/jait.16.4.568-581>
- Yang, J., Shu, L., Duan, H., & Li, H. (2024). RDguru: A conversational intelligent agent for rare diseases. *IEEE Journal of Biomedical and Health Informatics*, 29(9), 6366–6378. <https://doi.org/10.1109/JBHI.2024.3464555>
- Zhao, X., Leng, X., Wang, L., & Wang, N. (2024). Research on fine-tuning optimization strategies for large language models in tabular data processing. *Biomimetics*, 9(11), 708. <https://doi.org/10.3390/biomimetics9110708>
- Zhang, K., Lv, A., Chen, Y., Ha, H., Xu, T., & Yan, R. (2024). Batch-ICL: Effective, efficient, and order-agnostic in-context learning. *arXiv*, 2401.06469v3. <https://doi.org/10.48550/arXiv.2401.06469>
- Zheng, J., Wang, H., & Yao, J. (2024). Building lightweight domain-specific consultation systems via inter-external knowledge fusion contrastive learning. *IEEE Access*, 12, 113244–113259. <https://doi.org/10.1109/ACCESS.2024.3434648>
- Zhou, B., Geißler, D., & Lukowicz, P. (2024). Misinforming LLMs: Vulnerabilities, challenges and opportunities. *arXiv*, 2408.01168v1. <https://doi.org/10.48550/arXiv.2408.01168>
- Zhou, S., Zhou, Z., Wang, C., Liang, Y., Wang, L., Zhang, J., Zhang, J., & Lv, C. (2024). A user-centered framework for data privacy protection using large language models and attention mechanisms. *Applied Sciences*, 14(15), 6824. <https://doi.org/10.3390/app14156824>
- Zhu, F., Liu, Z., Feng, F., Wang, C., Li, M., & Chua, T. S. (2024). TAT-LLM: A specialized language model for discrete reasoning over financial tabular and textual data. *Proceedings of the ACM International Conference on AI in Finance*, 5, 310–318. <https://doi.org/10.1145/3677052.3698685>
- Zhu, G., Jia, W., Xing, Z., Xiang, L., Hu, A., & Hao, R. (2025). CMLLM: A novel cross-modal large language model for wind power forecasting. *Energy Conversion and Management*, 330, 119673. <https://doi.org/10.1016/j.enconman.2025.119673>

© [CC-BY-NC 4.0](#) The Author(s). For more information, see our [Open Access Policy](#).