

LARGE LANGUAGE MODELS NA INDEXAÇÃO AUTOMÁTICA DE RECURSOS EDUCACIONAIS

Análise comparativa entre linguagem autoral e representação bibliotecária

LARGE LANGUAGE MODELS IN AUTOMATIC INDEXING OF EDUCATIONAL RESOURCES

A comparative analysis between author language and bibliographic representation

Adriléia de Moura Lima

Universidade Federal de Minas Gerais, UFMG

E-mail: adrileia.lima@gmail.com

ORCID: 0000-0002-3086-0015

Currículo Lattes: 4615913744009144

Viviane LÍlian dos Santos Barrozo

Universidade Federal de Minas Gerais, UFMG

E-mail: vivibarrozoead@gmail.com

ORCID: 0000-0002-5893-4354

Currículo Lattes: 4339720406172473

Elisângela Cristina Aganette

Universidade Federal de Minas Gerais, UFMG

E-mail: elisangelaaganette@gmail.com

ORCID: 0000-0003-4357-8016

Currículo Lattes: 6224307970906388

Veríssimo Amaral Matias

Instituto Federal de Minas Gerais, IFMG

E-mail: verissimo.amaral89@gmail.com

ORCID: 0009-0008-8611-2869

Currículo Lattes: 5986561116506986

Frederico Cesar Mafra Pereira

Universidade Federal de Minas Gerais, UFMG

E-mail: professorfredericomafra@gmail.com

ORCID: 0000-0002-1971-8069

Currículo Lattes: 4271090306897767

RESUMO

Investiga-se a capacidade dos Grandes Modelos de Linguagens (do inglês Large Language Models - LLMs) na indexação automática de recursos educacionais. Para tanto, desenvolveu-se um estudo de caso envolvendo ChatGPT Plus, DeepSeek e Gemini, comparando-os aos desempenhos dos autores científicos e dos bibliotecários. A abordagem metodológica inclui a adoção de critérios de correspondência que permitiu

XII Congresso Internacional em Tecnologia e Organização da Informação

Guimarães, 24 de junho a 10 de julho de 2026

ISSN: 2965-4130

DOI: 10.5281/zenodo.23102801

estabelecer diferentes níveis de abrangência e de progressão. Os dados foram analisados à luz das medidas de precisão, abrangência e F-measure. Os modelos apresentaram resultados semelhantes às palavras-chave atribuídas pelos autores. Por outro lado, a correspondência com os descritores incluídos pelos bibliotecários apresentou a menor correspondência. Ou seja, os LLMs se aproximaram mais da indexação derivada (extração direta do documento) do que da indexação por atribuição (linguagem artificial), empregada pelos bibliotecários. Conclui-se que, com base nos resultados, esses modelos são capazes de realizar a indexação semântica. Contudo, há dificuldade em reproduzir a lógica padronizada, institucional ou técnica da indexação bibliotecária, sendo necessário ampliar os estudos.

Palavras-chave: Grandes Modelos de Linguagens, Inteligência Artificial, Indexação Automática, Classificação Bibliográfica.

ABSTRACT

This study investigates the ability of Large Language Models (LLMs) to automatically index educational resources. To this end, a case study was conducted involving ChatGPT Plus, DeepSeek, and Gemini, comparing their performance to that of academic authors and librarians. The methodological approach includes the adoption of matching criteria that allowed for the establishment of different levels of coverage and progression. The data were analyzed using precision, coverage, and F-measure metrics. The models produced results similar to the keywords assigned by the authors. On the other hand, the correspondence with the descriptors included by librarians showed the lowest match. In other words, the LLMs were closer to derived indexing (direct extraction from the document) than to indexing by attribution (artificial language), as employed by librarians. It is concluded that, based on the results, these models are capable of performing semantic indexing. However, there is difficulty in reproducing the standardized, institutional, or technical logic of library indexing, making it necessary to expand the studies.

Keywords: Large Language Models, Artificial Intelligence, Automatic Indexing, Bibliographic Classification.

REFERÊNCIAS

ACESSO ABERTO. **Dados científicos:** como construir metadados, descrição, README, dicionário de dados e mais, 2026.

XII Congresso Internacional em Tecnologia e Organização da Informação
Guimarães, 24 de junho a 10 de julho de 2026
ISSN: 2965-4130
DOI: 10.5281/zenodo.23102801

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676**: Métodos para análise de documentos — determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992.

BANH, L.; STROBEL, G. Generative artificial intelligence. **Electronic Markets**, v. 33, n. 63, p. 1-17, 2023. DOI: <https://doi.org/10.1007/s12525-023-00680-1>. Acesso em: 22 maio 2026.

DENG, F.; BUSCHEK, D. Examining autocompletion as a basic concept for interaction with generative AI. **I-Com**, v. 19, n. 3, p. 251-264, 2020.

FERREIRA, M. H. W.; CORRÊA, R. F. Sistematização da obtenção de indicadores temáticos de informação científica. *Encontros Bibli* 28, 1-30, 2023.

FUJITA, M. S. L.; SOUSA, N. M. T. Vocabulário controlado e inteligência artificial na indexação: uma revisão bibliográfica. **Perspectivas em Ciência da Informação** 30, e56745, 2025. DOI: <https://doi.org/10.1590/1981-5344/56745>. Acesso em: 22 maio 2026.

GRECO, B. C.; EVEDOVE, P. R.; NHACUONGUE, J. A. O potencial dos modelos de linguagem de grande escala (LLMs) na indexação automática. **ISKO Brasil**, v. 14, n. 2, p. 59-81, 2026.

KE, W.; ZHENG, Y.; LI, Y.; XU, H.; NIE, D.; WANG, P.; HE, Y. Large language models in document intelligence: a comprehensive survey, recent advances, challenges, and future trends. **ACM Transactions on Information Systems**, v. 44, n. 1, p. 1-64, 2025.

MINAEE, S.; MIKOLOV, T.; NIKZAD, N.; CHENAGHLU, M.; SOCHER, R.; AMATRIAIN, X.; GAO, J. **Large language models**: a survey. arXiv preprint arXiv:2402.06196, 2024.

TOMCZAK, J. M. **Deep Generative Modeling**. Springer International Publishing, Cham.

TORRES, A. A. L.; MACULAN, B. C. M. S.; ROCHA, A. A.; ASSUNÇÃO, F. M.; MARQUES, F. B.; SILVA, G. R. Inteligência artificial e a indexação de imagens com o Método Iconográfico de Panofsky. **ISKO Brasil**, Canela, RS, 2025.