

# ANÁLISE DE DADOS COM INTELIGÊNCIA ARTIFICIAL GENERATIVA:

Uma comparação com Deepseek e ChatGPT - dados iniciais

## Edna Karina da Silva Lira

Universidade Federal do Rio Grande (FURG); Universidade Federal de Santa Catarina (UFSC), Brasil


[liraa.karina@gmail.com](mailto:liraa.karina@gmail.com)

 <https://orcid.org/0000-0001-5543-3792>

## Jônatas Edison da Silva

Universidade Federal de Santa Catarina (UFSC), Brasil

[janesspechtm@gmail.com](mailto:janesspechtm@gmail.com)

 <https://orcid.org/0000-0001-5892-6736>

## Franciesca Goulart Santos

Universidade Federal do Rio Grande (FURG), Brasil

[franciescaggoulart@gmail.com](mailto:franciescaggoulart@gmail.com)

 <https://orcid.org/0009-0000-6568-2063>

## Angélica Conceição Dias Miranda

Universidade Federal do Rio Grande (FURG), Brasil

[angelicacdm@gmail.com](mailto:angelicacdm@gmail.com)

 <https://orcid.org/0000-0003-3624-4616>

## Thiago Magela Rodrigues Dias

Universidade Federal de Santa Catarina (UFSC);  
Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)

[thiogomagela@gmail.com](mailto:thiogomagela@gmail.com)

**DOI:** 10.22477/xiv.biredial.381

**EJE TEMÁTICO:** Infraestructura tecnológica

## RESUMEN

Pesquisadores desenvolveram estudos mencionando que Modelos que utilizam Large Language Model (LLM) como Deepseek e ChatGPT são treinados com grandes volumes de dados para gerar respostas em linguagem natural. A partir do contexto, a pergunta de pesquisa que permeia este estudo é: Qual o desempenho do ChatGpt e Deepseek podem análise de dados? Este trabalho teve como objetivo comparar o desempenho do DeepSeek e do ChatGPT na análise de dados iniciais, avaliou critérios como precisão, velocidade de processamento, capacidade de contextualização e compreensão dos sistemas em relação aos dados. Foram aplicadas duas leis da bibliometria, lei de Lotka e lei de Zipf para extrair dados de artigos publicados em um periódico científico da área da educação, que está hospedado no Portal de Periódicos Científicos da Universidade Federal do Rio Grande. Os documentos foram alocados para o Zotero e depois extraído em formato csv. Para realizar a análise, foram criados dois prompts: 1. Extração dos autores mais produtivos e das palavras-chave que mais se repetem; e 2. Faça a extração dos dez autores que mais se repetem e extração das palavras-chave. Estes foram elaborados de maneira objetiva para que o sistema o entendesse. Com isso, foi enviada a planilha com as informações e a seguir os comandos. O quadro 1 apresenta os resultados. Os resultados indicaram que ambas as inteligências artificiais apresentam limitações quanto ao tratamento e extração de informações, especialmente em contextos de dados não estruturados. No entanto, o DeepSeek demonstrou maior resposta na interpretação dos prompts, respondendo com maior precisão às solicitações propostas.

**Palabras-clave:** Publicação Científica; Inteligência Artificial; Análise de dados.

## ABSTRACT

Researchers have developed studies mentioning that Models utilizing Large Language Models (LLMs) such as Deepseek and ChatGPT are trained with large volumes of data to generate responses in natural language. Based on this context, the research question underlying this study is: What is the performance of ChatGPT and Deepseek in data analysis? This work aimed to compare the performance of DeepSeek and ChatGPT in initial data analysis, evaluating criteria such as accuracy, processing speed, contextualization capability, and the systems' comprehension in relation to the data. Two bibliometric laws, Lotka's law and Zipf's law, were applied to extract data from articles published in a scientific journal in the field of education, hosted on the Scientific Journals Portal of the Federal University of Rio Grande. The documents were allocated to Zotero and subsequently extracted in CSV format. To perform the analysis, two prompts were created: 1. Extraction of the most productive authors and the most frequently repeated keywords; and 2. Perform the extraction of the ten most frequently repeated authors and the extraction of keywords. These were elaborated in an objective manner for the system to understand them. Following this, the spreadsheet containing the information and then the commands were sent. Table 1 presents the results. The results indicated that both artificial intelligences exhibit limitations regarding the treatment and extraction of information, especially in contexts of unstructured data. However, DeepSeek demonstrated a superior response in interpreting the prompts, responding with greater accuracy to the proposed requests.

**Keywords:** .

## BIBLIOGRAFÍA

Carvalho Júnior, C. F., & Carvalho, K. R. S. A. (2018). Chatbot: uma visão geral sobre aplicações inteligentes. *Revista Sítio Novo*, 2(2), 68–84. <https://doi.org/10.47236/2594-7036.2018.v2.i2.68-84p>

DeepSeek AI. (2025). *DeepSeek AI | Leading AI language models & solutions*. <https://deepseek.ai/blog/what-is-deepseek-ai>

Haque, R. R., Brito, I. M., & Frade, S. (2024). *Copilot e ChatGPT: Como utilizar ferramentas de IA na análise de dados*. In M. Prada (Coord.), *Caderno de laboratório: Guia prático para investigadores/as* (Vol. 1, pp. 96–106). LAPSO—Laboratório de Psicologia, Iscte-Instituto Universitário de Lisboa. <https://repositorio.iscte-iul.pt/handle/10071/31862>

Ramos, A. S. M. (2023). Inteligência artificial generativa baseada em grandes modelos de linguagem: Ferramentas de uso na pesquisa acadêmica [Preprint]. *SciELO Preprints*. <https://doi.org/10.1590/SciELOPreprints.6105>

Russell, S., & Norvig, P. (2004). *Inteligência artificial* (2. ed.). Campus.

Sallam, M., Al-Mahzoum, K., Sallam, M., & Mijwil, M. M. (2025). DeepSeek: Is it the end of generative AI monopoly or the mark of the impending doomsday? *Mesopotamian Journal of Big Data*, 2025, 26–34. <https://doi.org/10.58496/MJBD/2025/002>

Sampaio, R. C., Sabbatini, Marcelo & Limongi, R. (2024). *Diretrizes para o uso ético e responsável da inteligência artificial generativa: um guia prático para pesquisadores*. São Paulo: Sociedade Brasileira de Estudos Interdisciplinares da Comunicação - Intercom.



## ANEXO 1

### RESUMEN BIOGRÁFICO DE LOS AUTORES

#### **Edna Karina da Silva Lira**

Doutoranda e Mestre em Ciência da Informação na Universidade Federal de Santa Catarina (2022-2024). Bibliotecária pela Universidade Federal do Rio Grande (FURG) com período sanduíche em Ciência da Informação na Universidade de Coimbra, em Portugal (2021-2022). Pesquisadora membro do Núcleo de Pesquisa e Estudos em Arquivos Contemporâneos - NUPEAC - Linha de pesquisa: Mercado de Trabalho em Ciência da Informação.

#### **Jônatas Edison da Silva**

Doutorando (2022 - presente) e Mestre (2020 - 2022) em Ciência Informação, pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina (PGCIN/UFSC). Bacharel em Arquivologia pela Universidade Federal de Santa Catarina (2016 - 2019). Atualmente, é Bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Com experiências em instituições públicas e privadas, na organização de documentos, processos administrativos, elaboração de instrumentos de gestão documental. Experiência na área de Ciência da Informação, atuando principalmente nos seguintes temas: Informação; Desinformação; Fake News; Pós-verdade; Redes Sociais; e Ciência da Informação. Atualmente é editor da Revista científica Encontros Bibli. Faz parte do grupo de pesquisa Humanos, Informação, Tecnologia (HIT).

#### **Franciesca Goulart Santos**

Graduanda em Biblioteconomia pela Universidade Federal do Rio Grande - FURG, atua como estagiária no Repositório Institucional da FURG.

#### **Angélica Conceição Dias Miranda**

Professora do Instituto de Ciências Humanas e da Informação (ICHI), da Universidade Federal do Rio Grande (FURG). Doutorado em Engenharia e Gestão do Conhecimento em 2010, mestrado em Engenharia de Produção em 2003, ambos pela Universidade Federal de Santa Catarina. Pós-doutorado em Educação em Ciências : química da vida e da Saúde, pela FURG. É professora associada, Dedicção Exclusiva - DE, na Universidade Federal do Rio Grande - FURG. Professora permanente no Programa de Pós-Graduação em Administração (orientação de mestrado) e no



Programa de Pós-Graduação em Educação em Ciências (orientação de mestrado e doutorado), ambos na FURG. Lider e Pesquisadora do Grupo de Pesquisa CITEG - Ciência, Informação e Tecnologia e Gestão - FURG. Coordena o Portal de Periódicos eletrônicos da FURG (<http://www.seer.furg.br>) e o Repositório Institucional ([repositorio.furg.br](http://repositorio.furg.br))

## Thiago Magela Rodrigues Dias

Doutor em Modelagem Matemática e Computacional pelo CEFET-MG (2016) tendo trabalhado com Bibliometria, Extração de Dados Científicos e Análise de Redes de Colaboração Científica. Mestre em Modelagem Matemática e Computacional pelo CEFET-MG (2008), trabalhando com Arquitetura Orientada a Serviços e Web Mining com o tema: Uma Arquitetura Orientada a Serviços para emprego em Sistemas de Mineração de Dados na Web. Possui graduação em Ciência da Computação pelo Centro Universitário de Formiga - UNIFOR (2004), além de Especialização em Produção de Software - com Ênfase em Software Livre pela UFLA (2007) e Especialização em Melhoria do Processo de Software, UFLA (2007). Atua como Professor no Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) lecionando disciplinas na Graduação e Pós-graduação da instituição. Professor do Programa de Pós-graduação em Modelagem Matemática e Computacional do CEFET-MG e Professor Permanente do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Santa Catarina (UFSC). Também tem atuação como Colaborador em Projetos no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Membro da Sociedade Brasileira de Computação (SBC) e da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB).