

16º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2025

INTEGRAÇÃO DE LLMS (CHATGPT) NO ENSINO DE FÍSICA: UMA ABORDAGEM INOVADORA PARA CURSOS DE ENGENHARIA E LICENCIATURA NO IFSP

GUILHERME C. DE ALMEIDA¹, KENYA APARECIDA ALVES²

¹ Graduando em Bacharelado em Engenharia Mecânica, Bolsista PIBITI - CNPQ, IFSP, Campus São José dos Campos, guilherme.carvalho1@aluno.ifsp.edu.br

² Professora Doutora, IFSP, Campus São José dos Campos, kenya.aparecida@ifsp.edu

Área de conhecimento (Tabela CNPq): 1.03.00.00-7 Ciência da Educação

RESUMO: Este trabalho apresenta os principais resultados do projeto desenvolvido no Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação (2025), intitulado "Integração de LLMS (ChatGPT e Gemini) no Ensino de Física: Uma Abordagem Inovadora para Cursos de Engenharia e Licenciatura no IFSP". A iniciativa teve como propósito elaborar e avaliar prompts específicos para potencializar o desempenho de Modelos de Linguagem de Grande Escala (LLMs) em atividades como explicação de conceitos, resolução de problemas, proposição de experimentos e análise de dados no ensino de Física em cursos de graduação. A pesquisa explorou o uso de ChatGPT e Gemini como ferramentas de apoio pedagógico, com foco na criação de prompts otimizados baseados na técnica Chain of Thought (CoT). A metodologia foi estruturada em quatro etapas: revisão sistemática da literatura, elaboração dos prompts, testes comparativos entre os dois modelos e análise qualitativa dos resultados. Para a avaliação, foram aplicadas 115 questões de vestibulares, classificadas em três categorias: objetivas, objetivas com imagem e dissertativas. Os resultados indicaram que os prompts estruturados aumentaram significativamente a clareza das respostas e sua adequação pedagógica. Como produto final, foram definidas quatro categorias funcionais de prompts: solucionador de problemas, tutor especializado, analista de dados e direcionador de áreas de estudo, constituindo um conjunto validado de recursos para o ensino de Física mediado por Inteligência Artificial.

PALAVRAS-CHAVE: inteligência artificial educacional; modelos de linguagem; ensino de física; chain of thought; prompts pedagógicos; tecnologia educacional.

INTEGRATION OF LLMS (CHATGPT) IN PHYSICS TEACHING: AN INNOVATIVE APPROACH FOR ENGINEERING AND TEACHING COURSES AT IFSP

ABSTRACT: This work presents the main results of the project developed in the Institutional Program for Undergraduate Scholarships in Technological Development and Innovation (2025), entitled "Integration of LLMS (ChatGPT and Gemini) in Physics Teaching: An Innovative Approach for Engineering and Teaching Degree Courses at IFSP". The initiative aimed to design and evaluate specific prompts to enhance the performance of Large Language Models (LLMs) in activities such as concept explanation, problem solving, experimental design, and data analysis in undergraduate Physics teaching. The research explored the use of ChatGPT and Gemini as pedagogical support tools, focusing on the creation of optimized prompts based on the Chain of Thought (CoT) technique. The methodology was structured in four stages: systematic literature review, prompt design, comparative testing between the two models, and qualitative analysis of results. For evaluation, 115 entrance exam questions were applied, classified into three categories: multiple-choice, multiple-choice with images, and essay

questions. The results indicated that structured prompts significantly increased the clarity of responses and their pedagogical suitability. As a final product, four functional prompt categories were defined: problem solver, specialized tutor, data analyst, and study area guide, constituting a validated set of resources for AI-mediated Physics teaching..

KEYWORDS: educational artificial intelligence; language models; physics teaching; chain of thought; pedagogical prompts; educational technology.

INTRODUÇÃO

A Inteligência Artificial tem se consolidado como tecnologia transformadora no contexto educacional, especialmente com o surgimento dos Modelos de Linguagem de Grande Escala (LLMs) como ChatGPT e Gemini. No ensino de Física, disciplina que exige raciocínio lógico estruturado e análise conceitual rigorosa, esses modelos apresentam potencial significativo para apoiar a aprendizagem através de explicações personalizadas e resolução detalhada de problemas.

A técnica Chain of Thought (CoT), proposta por Wei et al. (2022), organiza prompts para estimular cadeias de raciocínios intermediários, permitindo transparência no processo de resolução. No contexto dos cursos de Engenharia e Licenciatura do IFSP, onde há ênfase na formação técnica e científica de qualidade, a investigação do potencial dos LLMs contribui para práticas educacionais inovadoras. Este trabalho teve como objetivo desenvolver e validar prompts otimizados baseados na técnica CoT para diferentes finalidades educacionais no ensino de Física, analisando comparativamente o desempenho do ChatGPT em diferentes categorias de questões acadêmicas.

MATERIAL E MÉTODOS

Foram utilizadas 115 questões de Física selecionadas de vestibulares de alta concorrência (ITA, UNESP, USP) e software de análise de dados (Excel). A metodologia foi conduzida em quatro etapas principais. Primeira etapa: revisão sistemática da literatura em bases Google Scholar e SciELO, analisando 25 artigos sobre LLMs no ensino de Física. Segunda etapa: seleção e categorização das questões em três tipos - QTO (questões objetivas, n=70), QTOI (questões objetivas com imagem, n=30) e QTDI (questões dissertativas, n=15). Terceira etapa: desenvolvimento de prompts estruturados para quatro funções pedagógicas - solucionador de problemas, tutor especializado, analista de dados e direcionador de áreas de estudo. Cada categoria seguiu estrutura genérica adaptável baseada na técnica CoT. Quarta etapa: procedimento experimental em duas fases - teste sem prompts (linha de base) e teste with prompts otimizados. A análise considerou precisão da resposta e qualidade do raciocínio, com exame qualitativo baseado na estrutura CoT para identificação de padrões de desempenho.

RESULTADOS E DISCUSSÃO

Os resultados demonstraram diferenças significativas entre as categorias de questões testadas. Nas questões objetivas (QTO), observou-se taxa de acerto de 90,2% (63 de 70 questões), com 95,7% das respostas seguindo a linha de raciocínio estruturada. A aplicação de prompts CoT melhorou a explicitação de etapas intermediárias e a adequação pedagógica das explicações. As questões objetivas com imagem (QTOI) apresentaram maior desafio, com taxa de acerto de 83,3% (25 de 30 questões), evidenciando a capacidade dos LLMs em processar elementos visuais quando adequadamente orientados pelos prompts estruturados. Verificou-se que 86,7% dessas questões tiveram o raciocínio adequadamente estruturado após a aplicação dos prompts CoT. As questões dissertativas (QTDI) apresentaram taxa de acerto de 80% (12 de 15 questões), com 93,3% das respostas demonstrando raciocínio estruturado.

Os prompts estruturados conseguiram integrar efetivamente as quatro funções pedagógicas propostas em respostas coerentes, promovendo não apenas respostas corretas, mas principalmente

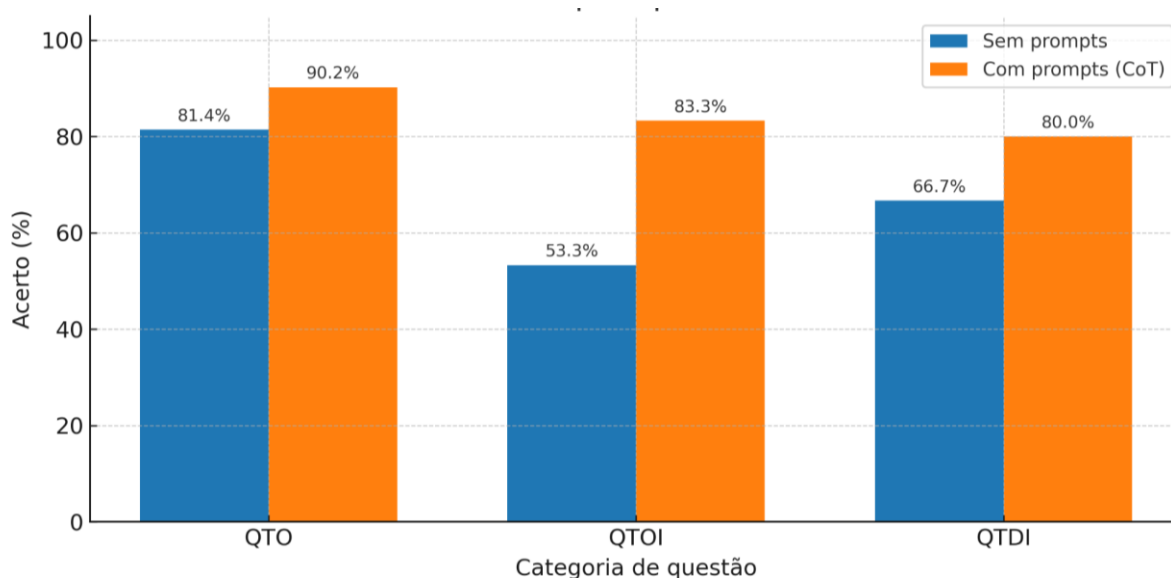
transparência no raciocínio científico, como demonstrado na Tabela 1. A comparação entre ChatGPT e Gemini revelou desempenhos similares em questões textuais, mas diferenças na capacidade de processamento multimodal, conforme ilustrado na Figura 1. Os dados indicam que a estruturação adequada de prompts pode transformar LLMs em ferramentas pedagógicas efetivas, complementando o ensino tradicional através da personalização da aprendizagem e feedback imediato.

TABELA 1. Desempenho dos LLMs por categoria de questão antes e depois da aplicação de prompts estruturados.

Categoria	Total	Sem Prompts		Com Prompts CoT	
		Acertos (%)	Raciocínio (%)	Acertos (%)	Raciocínio (%)
QTO	70	81,4	88,6	90,2	95,7
QTOI	30	53,3	76,7	83,3	86,7
QTDI	15	66,7	80	80	93,3

Fonte: Autoria Própria

FIGURA 1. Comparação de desempenho entre ChatGPT nas diferentes categorias de questões sem e com prompts otimizados.



Fonte: Autoria Própria

A Figura 1 ilustra claramente o impacto positivo dos prompts otimizados baseados na técnica Chain of Thought no desempenho do ChatGPT. Observa-se uma melhoria consistente em todas as categorias de questões, com destaque para o aumento significativo na taxa de acerto das questões objetivas com imagem (QTOI), que passou de 53,3% para 83,3%, demonstrando a eficácia dos prompts estruturados na interpretação de elementos visuais. O gráfico também evidencia que a maior melhoria ocorreu na qualidade do raciocínio estruturado, aspecto fundamental para aplicações educacionais.

CONCLUSÕES

A pesquisa demonstrou que prompts otimizados baseados na técnica Chain of Thought melhoram significativamente a qualidade das respostas dos LLMs no ensino de Física. A categorização em quatro funções pedagógicas (solucionador de problemas, tutor especializado, analista de dados e direcionador de áreas) mostrou-se efetiva para abordar diferentes aspectos do processo de ensino-aprendizagem. Os resultados evidenciaram alta eficácia em questões objetivas tradicionais e melhorias significativas no processamento de elementos visuais quando utilizados prompts adequadamente estruturados. A técnica CoT demonstrou-se fundamental para promover transparência no raciocínio, característica essencial para aplicações educacionais. Os LLMs devem ser integrados como ferramentas complementares ao ensino tradicional, potencializando a personalização da aprendizagem com supervisão pedagógica adequada. O conjunto de prompts validados constitui contribuição prática para educadores interessados em integrar IA de forma responsável no ensino de Física

AGRADECIMENTOS

Agradeço ao Conselho Nacional do Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de Iniciação em Desenvolvimento Tecnológico e Inovação (PIBITI) que viabilizou esta pesquisa e ao Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP) - Campus São José dos Campos pelo suporte institucional.

REFERÊNCIAS

- BROWN, T. **Language models are few-shot learners**. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NeurIPS), 33., 2020. Acesso em: 01 ago. 2025. Online. Disponível em: <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>.
- CRESWELL, A.; SHANAHAN, M.; HIGGINS, I. **Selection-Inference: Exploiting Large Language Models for Interpretable Logical Reasoning**. arXiv, 2022. Acesso em: 01 ago. 2025. Online. Disponível em: <https://arxiv.org/abs/2205.09712>.
- GARCÍA-MÉNDEZ, S.; DE ARRIBA-PÉREZ, F.; SOMOZA-LÓPEZ, M. C. **A Review of the Use of Large Language Models as Virtual Tutors**. Science & Education, 2024. Acesso em: 01 ago. 2025. Online. Disponível em: <https://doi.org/10.1007/s11191-024-00530-2>.
- KOJIMA, T.; et al. **Large Language Models are Zero-Shot Reasoners**. arXiv, 2022. Acesso em: 01 ago. 2025. Online. Disponível em: <https://arxiv.org/abs/2205.11916>.
- POLVERINI, G.; GREGORCIC, B. **How understanding large language models can inform the use of ChatGPT in physics education**. European Journal of Physics, 2024. Acesso em: 01 ago. 2025. Online. Disponível em: <https://doi.org/10.1088/1361-6404/ad1420>.
- WEI, J.; et al. **Chain-of-Thought Prompting Elicits Reasoning in Large Language Models**. arXiv, 2022. Acesso em: 01 ago. 2025. Online. Disponível em: <https://arxiv.org/abs/2201.11903>.
- ZHANG, Z.; ZHANG, A.; LI, M.; SMOLA, A. **Automatic Chain of Thought Prompting in Large Language Models (Auto-CoT)**. arXiv, 2022. Acesso em: 01 ago. 2025. Online. Disponível em: <https://arxiv.org/abs/2210.03493>.
- SHUM, K.; DIAO, S.; ZHANG, T. **Automatic Prompt Augmentation and Selection with Chain-of-Thought from Labeled Data (Automate-CoT)**. arXiv, 2023. Acesso em: 01 ago. 2025. Online. Disponível em: <https://arxiv.org/abs/2302.12822>.

WU, T.; TERRY, M.; CAI, C. J. **AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts**. CHI Conference on Human Factors in Computing Systems (CHI 2022), 2022. Acesso em: 01 ago. 2025. Online. Disponível em: <https://arxiv.org/abs/2110.01691>.

ZHEN, H.; SHI, Y.; HUANG, Y.; YANG, J. J.; LIU, N. **Leveraging Large Language Models with Chain-of-Thought and Prompt Engineering for Traffic Crash Severity Analysis and Inference**. Computers, 2024. Acesso em: 01 ago. 2025. Online. Disponível em: <https://doi.org/10.3390/computers13090232>.