



## **AVALIAÇÃO DO DESEMPENHO DO CHATGPT 3.5 EM QUESTÕES DE ENGENHARIA MECÂNICA: UM ESTUDO COMPARATIVO USANDO O ENADE 2019**

DOI: 10.37702/2175-957X.COBENGE.2024.5068

**Autores:** FERNANDO CÉSAR MEIRA MENANDRO, FRANSENGIO LEITE DA CUNHA

**Resumo:** Este artigo aborda o uso da ferramenta de inteligência artificial ChatGPT e suas implicações no contexto educacional. Embora o ChatGPT seja uma ferramenta avançada de processamento de linguagem natural, sua utilização inadequada como fonte absoluta de informações pode levar a resultados equivocados. O artigo destaca as limitações do ChatGPT na sua versão 3.5 em compreender o texto, fornecer informações atualizadas e acessar dados específicos. Além disso, discute-se a importância dos usuários, especialmente os alunos, compreenderem que o ChatGPT não substitui a pesquisa, o pensamento crítico e a consulta a fontes confiáveis. O estudo propõe a aplicação de uma prova de Engenharia Mecânica, o ENADE 2019, ao ChatGPT, a fim de comparar seus resultados com os obtidos pelos estudantes reais, demonstrando que o ChatGPT, na versão estudada, ainda não é adequado para fornecer respostas precisas e aprofundadas nesse campo.

**Palavras-chave:** ChatGPT, inteligência artificial, grandes modelos de linguagem, uso adequado do ChatGPT na sala de aula, limitações do ChatGPT, ENADE.

# AVALIAÇÃO DO DESEMPENHO DO CHATGPT 3.5 EM QUESTÕES DE ENGENHARIA MECÂNICA: UM ESTUDO COMPARATIVO USANDO O ENADE 2019

## 1 INTRODUÇÃO

Recentemente, o uso de inteligência artificial tem se tornado comum em várias aplicações do cotidiano. Isso se deve à crescente modernização e automação de diversos processos, desde o controle de redes sociais até a automação de grandes empresas, como ocorre com o advento da indústria 4.0. A facilidade de lidar com grandes quantidades de dados (big data), encontrar padrões e fornecer correlações por meio de ferramentas de aprendizado de máquina (machine learning) está cada vez mais presente em diferentes níveis de tratamento de dados.

Em 30 de novembro de 2022 os desenvolvedores da empresa OpenAI apresentaram uma ferramenta de inteligência artificial capaz de interpretar e processar a linguagem textual de forma natural e amigável ao usuário. Essa ferramenta foi denominada ChatGPT (Generative Pre-trained Transformer for Chat) e rapidamente se tornou um recurso computacional amplamente utilizado em escolas e universidades, tanto por alunos quanto por professores, para auxiliar na geração de textos, artigos, publicações e até mesmo automatizar alguns processos. Porém, devido à facilidade de uso, essa ferramenta tem sido frequentemente mal interpretada, sendo utilizada como uma fonte absoluta de informações, o que não foi sua finalidade principal. Esse uso incorreto tem causado alguns problemas de ordem conceitual, especialmente no meio acadêmico, nas escolas e em outros aspectos da sociedade.

Entretanto, assim como qualquer ferramenta, o ChatGPT pode ser utilizado de maneira produtiva e ser muito útil para o processo de ensino-aprendizagem. Por outro lado, pode ser usado de forma inadequada e causar resultados indesejáveis. Cabe à sociedade, representada pelos usuários, saber utilizá-la adequadamente. Ao longo da história, existiram vários exemplos de tecnologias que, inicialmente, foram criticadas no meio acadêmico e educacional, como o uso da calculadora, do computador, do acesso à internet, de ferramentas de busca como o Google, de dispositivos móveis e redes sociais. Todas essas tecnologias são exemplos que foram e, em alguns casos, ainda são criticadas pela pedagogia. No entanto, quando utilizadas de forma adequada, são ferramentas importantes para o processo de ensino-aprendizagem. Por outro lado, se não houver um gerenciamento adequado do uso dessas ferramentas pelos professores e pela equipe pedagógica, elas podem se tornar um problema e atrapalhar mais do que ajudar no aprendizado dos alunos.

Esse conceito também se aplica ao uso do ChatGPT. Por vezes, os alunos acreditam que o ChatGPT é um oráculo moderno e, principalmente, um provedor de conhecimento. No entanto, sua real finalidade não é essa. Um uso comum do ChatGPT pelos alunos é tentar resolver questões que os professores passam em sala de aula, como exercícios de fixação ou trabalhos avaliativos, resolvidor de problemas matemáticos e físicos, e esses são exemplos de uso inadequado dessa ferramenta. O ChatGPT tem como característica desenvolver diálogos, interpretar textos e fornecer respostas em forma de texto, mas nem sempre essas respostas são fiéis à realidade ou à verdade.

O ChatGPT é um modelo de linguagem desenvolvido pela OpenAI, que significa *Generative Pre-trained Transformer*. Ele foi treinado em uma ampla variedade de dados textuais coletados da internet e de outras fontes, com o objetivo de aprender a prever a próxima palavra em uma sequência de texto, utilizando modelos de redes neurais de transformadores.

Durante o processo de treinamento, o modelo busca identificar padrões e estruturas linguísticas com base no contexto fornecido pelos dados de treinamento. Dessa forma, quando uma mensagem ou pergunta é inserida no ChatGPT, ele utiliza as informações aprendidas para gerar uma resposta adequada, tentando prever a sequência de palavras mais provável para seguir a entrada e produzir uma resposta coerente. É importante destacar, no entanto, que o ChatGPT, assim como outros modelos de linguagem, possui suas limitações. Ele não possui uma compreensão real do texto nem conhecimento factual, e suas respostas são baseadas nas correlações identificadas nos dados de treinamento. Isso significa que ocasionalmente o modelo pode gerar respostas incorretas, incoerentes ou fora de contexto.

A OpenAI desenvolveu o ChatGPT com o intuito de fornecer uma poderosa ferramenta de linguagem natural capaz de auxiliar em diversas tarefas, como geração de texto, respostas a perguntas e interações em bate-papos (chats) virtuais. No entanto, é essencial que os usuários compreendam suas limitações e utilizem o ChatGPT de maneira adequada, considerando-o como uma ferramenta auxiliar e não como uma fonte absoluta de informações. É relevante ressaltar também que o ChatGPT, na versão estudada - a 3.5, possui uma data de corte do conhecimento, sendo treinado com dados disponíveis até o ano de 2022. Portanto, ele não possui informações sobre eventos ocorridos após essa data.

Conforme mencionado anteriormente, o ChatGPT é uma tecnologia avançada que utiliza inteligência artificial para gerar respostas textuais com base no contexto fornecido. No entanto, sua utilização deve ser feita com cautela, considerando suas limitações e garantindo que seja aplicado de maneira apropriada para obter resultados satisfatórios. Além disso, é importante destacar que o ChatGPT 3.5 não possui a capacidade de acesso à internet. Ao contrário de um mecanismo de busca, ele não pode buscar informações atualizadas em tempo real.

A tentativa de utilizar o ChatGPT para gerar respostas a determinadas questões pode resultar em respostas imprecisas. Apesar dessa informação, alunos de diferentes níveis estão utilizando essa ferramenta para esse propósito. O objetivo deste estudo é demonstrar e comprovar que, com a tecnologia do ChatGPT testada (versão 3.5), os resultados das respostas às questões não são confiáveis o suficiente para que os alunos (e educadores) as utilizem indiscriminadamente na busca de respostas para as questões propostas pelos docentes em sala de aula.

Com o objetivo de validar a hipótese de que o ChatGPT 3.5 ainda não é adequado para fornecer respostas a questões que envolvem cálculos, raciocínio lógico e um conhecimento aprofundado, os pesquisadores propuseram neste estudo a realização de um teste. Esse teste consistiu na aplicação de uma prova de conhecimentos relacionados à Engenharia Mecânica, a fim de comparar os resultados obtidos pelo ChatGPT com as respostas fornecidas pelos estudantes de Engenharia Mecânica no estágio final de seu curso. Para ser mais específico, utilizou-se a prova de conhecimento proposta pelo Ministério da Educação do Brasil (MEC), o Exame Nacional de Desempenho dos Estudantes (ENADE) de 2019, aplicado aos alunos que estavam concluindo o curso de Engenharia Mecânica em todo o país na época do referido exame. Uma vez que o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP - órgão do MEC responsável pela aplicação desta avaliação), disponibiliza um relatório contendo os

resultados da prova, permitindo a comparação dos dados questão por questão, resultados estes que se apresentam como parâmetros ideais para avaliar e comparar os resultados obtidos pelo ChatGPT.

## 2 REVISÃO BIBLIOGRÁFICA

Logo após o lançamento mundial do ChatGPT, e ainda durante a estrondosa popularidade do modelo linguístico, Chung Lo (2023) produziu uma primeira revisão dos trabalhos publicados envolvendo o chatbot ChatGPT e suas implicações no campo da educação. Ainda que este artigo tenha sido produzido apenas 3 meses após o lançamento da ferramenta, o levantamento bibliográfico selecionou 50 artigos (apenas artigos acadêmicos, em língua inglesa, sobre educação e o ChatGPT, entre janeiro de 2022 e fevereiro de 2023) para análise de conteúdo. O artigo chama a atenção para o risco de plágio (ou mesmo autoplágio) na utilização indiscriminada do software, tanto por estudantes quanto por acadêmicos. Polat e outros (2024) também realizaram um levantamento bibliométrico da pesquisa sobre o ChatGPT na educação. Este trabalho levantou, após as devidas exclusões seguindo a metodologia adotada, 212 publicações. A análise bibliométrica aponta, interessantemente, para a “educação em engenharia” como um dos temas motores no mapa temático desenvolvido a partir dos dados (Polat, et al., 2024, p. 75).

Especificamente nas aplicações em engenharia mecânica, Mustapha, Yap e Abakr (2024) se dedicaram a analisar os diferentes desdobramentos no desenvolvimento dos grandes modelos de linguagem (LLM, do inglês Large Language Models) e ferramentas de Inteligência Artificial Generativa as oportunidades e limitações dessas ferramentas. Nesse trabalho foram comparadas as ferramentas ChatGPT e Bard para resolução de questões discursivas de engenharia, mostrando relativa superioridade nas respostas do ChatGPT, e que soluções inconclusivas e erradas podem ser fornecidas devido à falta de compreensão do conhecimento específico de engenharia.

Nikolic e outros oito autores de sete universidades Australianas “examinam como o ChatGPT pode afetar a avaliação na educação em engenharia analisando as respostas do ChatGPT em avaliações existentes em 10 matérias...” (Nikolic et al., 2023). O estudo mostrou fragilidade nas resoluções do modelo de linguagem tanto em assuntos variados quanto em modelos variados de avaliação. Os autores sugerem a adoção de modelos variados de avaliação de rendimento a fim de garantir a acurácia destas avaliações e também fornecem algumas sugestões de utilização da ferramenta para melhorar o processo de ensino e aprendizagem em engenharia.

Já Newton e Xiromeriti (2023) compararam a performance da ferramenta, em diferentes versões, em exames de múltipla escolha em diversas áreas do ensino superior. A comparação de performances entre a versão 3.5 e a versão 4 mostra que a nova apresenta uma performance muito superior à da versão anterior (Newton; Xiromeriti, 2023, p. 17-20). Isto seria de se esperar, uma vez que o número de parâmetros utilizados para treinamento nos grandes modelos de linguagem vem crescendo exponencialmente, passando de 20 bilhões no ChatGPT-3 para cerca de 100 trilhões no ChatGPT-4 (Mustapha; Yap; Abakr, 2024, p.3).

Pursnani, Sermet, e Demir (2023) realizaram um estudo da performance do ChatGPT no Exame Estadunidense de Fundamentos de Engenharia Ambiental, aplicado por uma organização não-governamental sem fins lucrativos (National Council of Examiners for Engineering and Surveying - NCEES) para licenciamento de engenheiros ambientais. O estudo mostra o desenvolvimento da qualidade da resolução do exame em sucessivas versões da ferramenta.

A educação em engenharia na era da Inteligência Artificial Generativa é abordada, também, no artigo de Qadir (2023). O professor Junaid Qadir dialoga com a ferramenta na proposição de temas e na discussão de problemas decorrentes da introdução desta no ambiente acadêmico.

O Exame Nacional de Desempenho (ENADE) dos alunos dos cursos de graduação no Brasil foi criado em 2003, no governo Luiz Inácio Lula da Silva (BRASIL, 2003), a partir de um exame nacional anterior, denominado Provão criado em 1996, durante o governo Fernando Henrique Cardoso (Schwartzman, 2010). Este exame busca medir o progresso dos estudantes durante os cursos de graduação nas diferentes áreas do conhecimento (Melguizo; Wainer, 2016). O ENADE aponta, através de seus resultados para o benefício aos estudantes decorrente da formação durante os cursos de graduação (Lokalka, *et al.* 2014).

### 3 METODOLOGIA

Com base nas informações e restrições mencionadas, a metodologia empregada para avaliar o desempenho do ChatGPT neste trabalho consistiu inicialmente na seleção do exame a ser utilizado. Os pesquisadores optaram por utilizar o Exame Nacional de Desempenho dos Estudantes (ENADE) de 2019, direcionado aos estudantes de Engenharia Mecânica no estágio final de sua formação.

As questões objetivas e dissertativas do exame foram aplicadas ao chat, excluindo-se, evidentemente, as questões anuladas. No caso das questões com figuras, foi preparada uma descrição detalhada da figura, de forma a possibilitar compreensão de dados gráficos pelo ChatGPT-3.5, que não possui essa habilidade, assim como descrito por Nikolic (2023).

Outra informação de relevância reside no fato de que, embora a versão utilizada do ChatGPT possua a habilidade de fornecer informações e executar tarefas fundamentadas no conhecimento adquirido durante seu treinamento anterior, é incapaz de atualizar seu conhecimento em tempo real ou acessar dados específicos de outros usuários ou websites. Como já explicado, sua funcionalidade está restrita à geração de texto, baseada em padrões previamente aprendidos. Isso implica que cada interação é tratada de forma autônoma, sem existência de uma conexão direta entre elas (conversas ou chats), inclusive no caso de um mesmo usuário, reforçando, para a versão utilizada. Cada chat é encarado como uma sessão independente, desprovida de acesso a informações ou histórico de conversas anteriores. O modelo estudado não possui conhecimento contextual ou familiaridade com questões levantadas em interações passadas, mesmo dentro da mesma sessão pertencente a um único usuário. Dessa forma, quando um usuário cria dois chats distintos (ou até mesmo chats de diferentes usuários), cada chat é considerado como uma nova sessão individual, sem qualquer prévio conhecimento das interações ocorridas em outros chats. Cada chat é iniciado a partir do zero, sem acesso a informações ou respostas provenientes de outros chats realizados pelo mesmo usuário (ou por usuários diferentes). É crucial ter em mente que o ChatGPT não possui a capacidade de "lembrar" conversas anteriores ou acompanhar o histórico do usuário, ou mesmo de outros usuários, resultando em cada chat representando uma experiência isolada e independente.

Cada aplicação do exame foi conduzida, então, em um chat separado, garantindo que o ChatGPT não tivesse acesso aos resultados obtidos em aplicações anteriores. Essa estratégia assegurou a independência e imparcialidade dos resultados de cada teste realizado.

A amostragem estatística definida foi de 10 provas. Ou seja, o exame foi aplicado 10 vezes ao ChatGPT, pelos dois autores deste artigo. Essa amostra, apesar de pequena, permitiu uma análise confiável do desempenho da ferramenta.

Após cada aplicação, os resultados obtidos para cada questão foram registrados e comparados com os resultados alcançados pelos alunos finalistas de Engenharia Mecânica que participaram do ENADE 2019 apresentados em relatório fornecido pelo INEP (Brasil, 2019). Essa comparação estabeleceu uma base de referência para avaliar o desempenho relativo do ChatGPT em relação aos estudantes reais.

Por fim, os resultados obtidos foram analisados e estão apresentados neste estudo. A comparação entre as respostas fornecidas pelo ChatGPT e as respostas dos estudantes do ENADE 2019 permitiu uma avaliação objetiva e quantitativa do desempenho do sistema em relação ao conhecimento requerido no campo da Engenharia Mecânica.

#### 4 RESULTADOS

Nas dez realizações da prova do ENADE 2019 foram obtidos os resultados apresentados na Tabela 1.

Tabela 1 – Desempenho nas diferentes realizações.

Realização	Número de acertos	Percentual de acertos
1	9	29,03%
2	7	22,58%
3	11	35,48%
4	9	29,03%
5	9	29,03%
6	12	38,71%
7	12	38,71%
8	11	35,48%
9	10	32,26%
10	8	25,81%
<b>Média</b>	<b>9,8</b>	<b>31,61%</b>
<b>Mediana</b>	<b>9,5</b>	<b>30,65%</b>
<b>Desvio Padrão</b>	<b>1,687</b>	<b>5,44%</b>

Fonte: Elaborado pelos autores.

O resultado médio obtido foi inferior à média nacional no ENADE 2019, de 36,4% (Brasil, 2019). Não foi realizada a ponderação com base no índice de facilidade das questões conforme o procedimento do ENADE nos cálculos da Tabela 1, mas considera-se que, na média, a comparação indica a fragilidade da ferramenta.

A comparação entre os resultados das provas mostrou a quantidade de acertos obtidos nas dez realizações da prova, por área de conhecimento. O Quadro 1 apresenta as áreas do conhecimento em cada questão, separadas por cor, e o número de acertos de cada uma das questões nas dez realizações da prova. As questões de número 1, 9, 21 e 28 foram anuladas, não sendo, por isso, computadas.

Analisando os resultados do Quadro 1 pode-se notar que, em 10 realizações da prova, o ChatGPT somente foi capaz de acertar todas (100%) as realizações de 3 questões: duas de formação geral e uma da área de Engenharia Térmica e de Fluidos. Em duas das principais áreas da Engenharia Mecânica, Sistemas Mecânicos e

Engenharia Térmica e de Fluidos, a ferramenta foi capaz de obter apenas cerca de 25% de acertos, sendo que em apenas duas questões (de 10) em Sistemas Mecânicos acertou mais da metade das vezes. A performance na área de Materiais e Processos de Fabricação, bem como na área de Formação Geral e Meio Ambiente foi ligeiramente melhor que nas duas outras áreas, ainda que tenham obtido resultado acima de 50% somente 3 questões de Formação Geral, uma de Materiais e Processos de Fabricação e uma de Meio Ambiente.

Quadro 1 – Número de acertos nas 10 realizações da prova.

ITEM	AREA PREDOMINANTE	ACERTOS
QUESTÃO 2	FORMAÇÃO GERAL	10
QUESTÃO 3	FORMAÇÃO GERAL	2
QUESTÃO 4	FORMAÇÃO GERAL	6
QUESTÃO 5	FORMAÇÃO GERAL	0
QUESTÃO 6	FORMAÇÃO GERAL	0
QUESTÃO 7	FORMAÇÃO GERAL	0
QUESTÃO 8	FORMAÇÃO GERAL	10
QUESTÃO 10	MATERIAIS E PROCESSOS DE FABRICAÇÃO	3
QUESTÃO 11	MATERIAIS E PROCESSOS DE FABRICAÇÃO	4
QUESTÃO 12	TERMICA E FLUIDOS	0
QUESTÃO 13	TERMICA E FLUIDOS	0
QUESTÃO 14	SISTEMAS MECÂNICOS	6
QUESTÃO 15	TERMICA E FLUIDOS	2
QUESTÃO 16	SISTEMAS MECÂNICOS	4
QUESTÃO 17	SISTEMAS MECÂNICOS	0
QUESTÃO 18	AMBIENTAL	9
QUESTÃO 19	AMBIENTAL	0
QUESTÃO 20	TERMICA E FLUIDOS	0
QUESTÃO 22	SISTEMAS MECÂNICOS	8
QUESTÃO 23	SISTEMAS MECÂNICOS	1
QUESTÃO 24	SISTEMAS MECÂNICOS	2
QUESTÃO 25	MATERIAIS E PROCESSOS DE FABRICAÇÃO	7
QUESTÃO 26	TERMICA E FLUIDOS	10
QUESTÃO 27	TERMICA E FLUIDOS	2
QUESTÃO 29	SISTEMAS MECÂNICOS	0
QUESTÃO 30	SISTEMAS MECÂNICOS	1
QUESTÃO 31	SISTEMAS MECÂNICOS	5
QUESTÃO 32	MATERIAIS E PROCESSOS DE FABRICAÇÃO	1
QUESTÃO 33	ELÉTRICA	0
QUESTÃO 34	MATERIAIS E PROCESSOS DE FABRICAÇÃO	5
QUESTÃO 35	SISTEMAS MECÂNICOS	0

Fonte: Elaborado pelos autores.

Observando, agora, os índices de facilidade das questões, constata-se que o ChatGPT não conseguiu resolver corretamente nenhuma vez a única questão considerada difícil na área de Formação Geral (Brasil, 2019, p. 151), assim como a única questão considerada muito difícil entre as áreas de Conhecimento Específico (Brasil,

2019, p. 155). A única questão fácil resolvida corretamente pela ferramenta é a de número 2. Ainda entre as áreas de Conhecimento Específico, somente uma questão com índice de dificuldade médio (questão 18) obteve mais de 5 acertos.

Como comentários adicionais, é interessante notar que houve ao menos um caso em que o ChatGPT respondeu de forma discursiva a pergunta corretamente mas indicou a alternativa errada, o que indicaria um erro no preenchimento do gabarito, se se tratasse de uma pessoa. Em outros casos a ferramenta claramente não saberia a resposta, mas indicou corretamente a opção, correspondendo a um chute. Como o modelo linguístico apenas analisa a possibilidade estatística de uma palavra se seguir a outra, esses resultados são esperados, assim como a falta de coerência nas respostas em diferentes oportunidades de resolução da prova.

## 5 CONSIDERAÇÕES FINAIS

Embora o ChatGPT seja uma ferramenta impressionante, é fundamental reconhecer que seu uso indiscriminado como uma fonte absoluta de informações pode levar a resultados equivocados. Os alunos, em particular, devem compreender que o ChatGPT não é uma fonte do conhecimento e não substitui a pesquisa, o pensamento crítico e a consulta a fontes confiáveis. Utilizá-lo para resolver questões propostas pelos professores ou como base exclusiva para trabalhos acadêmicos é considerado um mau uso dessa ferramenta.

Como ocorre com qualquer tecnologia, o ChatGPT tem seu papel no processo de ensino-aprendizagem quando utilizado de maneira apropriada. Os educadores devem orientar os alunos sobre a melhor forma de aproveitar essa ferramenta, incentivando o pensamento crítico, a investigação independente e a análise de diferentes perspectivas. Os professores desempenham um papel fundamental na orientação dos alunos sobre como utilizar o ChatGPT como uma ferramenta auxiliar, equilibrando seu uso com outras habilidades e recursos.

No entanto, é necessário um cuidadoso gerenciamento do uso do ChatGPT, ou de outros grandes modelos de linguagem, ou outros aplicativos de inteligência artificial, nas instituições de ensino. As escolas e universidades devem estabelecer diretrizes claras sobre o uso adequado dessa tecnologia, promovendo uma abordagem responsável e consciente. Isso inclui incentivar o uso crítico, fornecer orientações sobre a validação de informações e enfatizar a importância da criatividade, do pensamento original e da pesquisa acadêmica sólida.

Como futuros trabalhos, os autores pretendem realizar o mesmo estudo, comparando os resultados apresentados com as versões mais recentes do ChatGPT. Até a finalização da redação deste artigo, a versão 4.0 ainda não é gratuita, entretanto, no dia 13 de maio de 2024, a OpenAI anunciou a sua nova versão, a GPT-4o, com muitos novos recursos disponíveis, interessantes a serem utilizados no contexto deste trabalho, e também que a versão 4.0 será gratuita. Obtendo recursos, os autores pretendem comparar o resultado deste trabalho com estas duas versões, inclusive com a interpretação automática de imagens, gráficos, acesso à internet e outras facilidades, que podem modificar as conclusões apresentadas.

## REFERÊNCIAS

BRASIL, MEC. Sistema Nacional de Avaliação e Progresso da Educação Superior: uma nova sistemática de avaliação do ensino superior brasileiro. 2003.

BRASIL, MEC, INEP. **Relatório Síntese de Área: Engenharia Mecânica. 2019**

LO, Chung Kwan. What is the impact of ChatGPT on education? A rapid review of the literature. **Education Sciences**, v. 13, n. 4, p. 410, 2023.

LOYALKA, Prashant et al. Factors affecting the quality of engineering education in the four largest emerging economies. **Higher Education**, v. 68, p. 977-1004, 2014.

MELGUIZO, Tatiana; WAINER, Jacques. Toward a set of measures of student learning outcomes in higher education: evidence from Brazil. **Higher Education**, v. 72, p. 381-401, 2016.

MUSTAPHA, Khameel B.; YAP, Eng Hwa; ABAKR, Yousif Abdalla. Bard, ChatGPT and 3DGPT: a scientometric analysis of generative AI tools and assessment of implications for mechanical engineering education. **Interactive Technology and Smart Education**, 2024.

NEWTON, Philip M.; XIROMERITI, Maira. ChatGPT performance on MCQ exams in higher education. A pragmatic scoping review. **EdArXiv. February**, v. 21, 2023.

NIKOLIC, Sasha et al. ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. **European Journal of Engineering Education**, v. 48, n. 4, p. 559-614, 2023.

POLAT, Hamza et al. A bibliometric analysis of research on ChatGPT in education. **International Journal of Technology in Education (IJTE)**, v. 7, n. 1, 2024.

PURSNANI, V., SERMET, Y., DEMIR, I. Performance of ChatGPT on the US Fundamentals of Engineering Exam: Comprehensive Assessment of Proficiency and Potential Implications for Professional Environmental Engineering Practice, *arXiv preprint arXiv:2304.12198*, 2023.

QADIR, Junaid. Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In: **2023 IEEE Global Engineering Education Conference (EDUCON)**. IEEE, 2023. p. 1-9.

SCHWARTZMAN, Simon. The national assessment of courses in Brazil. **Em DILL, David D.; BEERKENS, Maarja (Ed.). Public policy for academic quality: Analyses of innovative policy instruments**. Dordrecht: Springer, p. 293-312, 2010.

## PERFORMANCE EVALUATION OF CHATGPT 3.5 IN MECHANICAL ENGINEERING QUESTIONS: A COMPARATIVE STUDY USING ENADE 2019

**Abstract:** This article discusses the use of the artificial intelligence tool ChatGPT and its implications in the educational context. Although ChatGPT is an advanced natural language processing tool, its inappropriate use as an absolute source of information can lead to erroneous results. The article highlights the limitations of ChatGPT in its 3.5 version in understanding text, providing up-to-date information and accessing specific data. It also discusses the importance of users, especially students, understanding that ChatGPT is not a substitute for research, critical thinking and consulting reliable sources. The study proposes the application of a Mechanical Engineering test, the ENADE 2019, to ChatGPT in order to compare its results with those obtained by real students, demonstrating that ChatGPT, in the version studied, is not yet suitable for providing accurate and in-depth answers in this field.

**Keywords:** *ChatGPT, artificial intelligence, large language models, proper use of ChatGPT in the classroom, limitations of ChatGPT, ENADE.*

