

## Análise do desempenho de *chatbots* na resolução de questões de Química em provas do ENEM

Janaina Batista Cazuza da Silva<sup>1</sup> 

Bruno Silva Leite<sup>2</sup> 

### Resumo

Atualmente, a Inteligência Artificial Generativa (IAGen), como os *chatbots*, tem se consolidado como uma tecnologia com grande potencial de apoio ao ensino e à aprendizagem, inclusive no campo da Química. Essas ferramentas podem favorecer a autonomia dos estudantes ao oferecer explicações de conteúdos e diferentes formas de acesso à informação. Todavia, persistem dúvidas quanto à exatidão e à confiabilidade das respostas geradas, devido às limitações inerentes a esses sistemas. O Exame Nacional do Ensino Médio (ENEM) representa uma das principais vias de acesso ao ensino superior e tem em um de seus critérios a avaliação da área do conhecimento de Ciências da Natureza, a qual compreende assuntos de Química. Esta pesquisa teve como objetivo analisar o desempenho dos *chatbots* de IAGen em responder às questões de Química do Exame. A partir de uma abordagem qualitativa foram analisadas as questões de Química do ENEM no período de 2014 a 2024 utilizando diferentes sistemas de IAGen (ChatGPT, Gemini, LuzIA e Copilot), comparando as respostas geradas. Os resultados evidenciam que os *chatbots* apresentaram alto índice de acertos em questões que possuíam baixa complexidade conceitual, mas demonstraram limitações diante de questões que necessitavam de maior interpretação ou domínio teórico. Identifica-se que as ferramentas utilizadas possuem potencial para apoiar o trabalho docente, entretanto desde que o uso seja acompanhado por uma mediação crítica e domínio adequado na elaboração de *prompts* específicos. Este estudo aponta a importância do uso consciente dessas ferramentas, bem como a necessidade de docentes e estudantes preparados para seu uso.

**Palavras-chave:** ensino de Química; inteligência artificial generativa; provas do ENEM; tecnologias digitais na educação.

### Analysis of the performance of chatbots in solving Chemistry questions in ENEM exams

### Abstract

Currently, Generative Artificial Intelligence (GenAI), such as chatbots, has established itself as a technology with great potential to support teaching and learning, including in the field of Chemistry. These tools can promote student autonomy by offering explanations of content and different ways to access information. However, doubts remain about the accuracy and reliability of the generated responses due to the inherent limitations of these systems. The National High School Exam (ENEM) represents one of the main pathways to higher education and includes in one of its criteria is the assessment of knowledge in the area of Natural Sciences, which includes Chemistry subjects. This

<sup>1</sup> Graduanda em Licenciatura em Química, Universidade Federal Rural de Pernambuco - UFRPE. Recife, Pernambuco, Brasil. ORCID: <https://orcid.org/0009-0003-5699-6900>. E-mail: [janainabatistasilva123@gmail.com](mailto:janainabatistasilva123@gmail.com)

<sup>2</sup> Doutor em Química, pela Universidade Federal de Pernambuco - UFPE. Professor da Universidade Federal Rural de Pernambuco – UFRPE. Recife, Pernambuco, Brasil. ORCID: <https://orcid.org/0000-0002-9402-936X>. E-mail: [brunoleite@ufrpe.br](mailto:brunoleite@ufrpe.br)

reaseas criterios la evaluación del área de conocimiento de Ciencias Naturales, que comprende temas de Química. El objetivo de esta investigación fue analizar el rendimiento de los chatbots de IAGen en la respuesta a las preguntas de Química del examen. A partir de un enfoque cualitativo, se analizaron las preguntas de Química del ENEM en el período de 2014 a 2024 utilizando diferentes sistemas de IAGen (ChatGPT, Gemini, LuzIA y Copilot), comparando las respuestas generadas. Los resultados aimed to analyze the performance of GenAI chatbots in answering Chemistry questions from the Exam. Using a qualitative approach, the Chemistry questions from the ENEM exam between 2014 and 2024 were analyzed using different GenAI systems (ChatGPT, Gemini, LuzIA, and Copilot), comparing the generated answers. The results show that chatbots achieved a high accuracy rate on questions of low conceptual complexity, but demonstrated limitations when faced with questions requiring greater interpretation or theoretical mastery. It is recognized that the tools used have the potential to support teaching work, provided that their use is accompanied by critical mediation and adequate mastery in the developing specific prompts. This study highlights the importance of using these tools consciously, as well as the need for teachers and students to be prepared for their use.

**Keywords:** Chemistry teaching; generative artificial intelligence; ENEM exams; digital technologies in education.

## **Análisis del desempeño de los chatbots en la resolución de preguntas de Química en los exámenes ENEM**

### **Resumen**

Actualmente, la Inteligencia Artificial Generativa (IAGen), como los chatbots, se ha consolidado como una tecnología con gran potencial para apoyar la enseñanza y el aprendizaje, incluso en el campo de la Química. Estas herramientas pueden favorecer la autonomía de los Estudiantes al ofrecer explicaciones de contenido y diferentes formas de acceso a la información. Sin embargo, persisten las dudas sobre la precisión y fiabilidad de las respuestas generadas, debido a las limitaciones inherentes a estos sistemas. El Examen Nacional de Educación Secundaria (ENEM) representa una de las principales vías de acceso a la educación superior y tiene como uno de sus ejes de evidencia que los chatbots presentaron un alto índice de aciertos en preguntas que tenían una baja complejidad conceptual, pero mostraron limitaciones ante preguntas que requieren mayor interpretación o dominio teórico. Se identifica que las herramientas utilizadas tienen potencial para apoyar el trabajo docente, siempre y cuando su uso vaya acompañado de una mediación crítica y un dominio adecuado en la elaboración de indicaciones específicas. Este estudio señala la importancia del uso consciente de estas herramientas, así como la necesidad de que profesores y estudiantes estén preparados para su uso.

**Palabras clave:** enseñanza de la Química; inteligencia artificial generativa; exámenes ENEM; tecnologías digitales en la educación.

### **Introdução**

A Inteligência Artificial (IA) pode ser considerada a principal protagonista da efervescente transformação digital que está marcando a segunda década do século XXI. A IA já circula há anos, partindo de uma visão computacional da mente, que, com seus pressupostos funcionalistas, a concebe como um sistema de processamento de informação que pode ser instanciado em qualquer material, embora tenha sido criado por cientistas de computação, matemáticos e estatísticos na década de 50 na busca de convencer o pentágono (departamento de defesa americana) a criar máquinas inteligentes (Leite, 2024).



A IA inclui uma ampla gama de tecnologias, desde algoritmos simples que podem classificar dados até sistemas mais avançados que podem imitar processos de pensamento semelhantes aos humanos (Leite, 2023). Segundo Pavlik (2023, p. 3) a IA se refere “à simulação da inteligência humana em máquinas que são programadas para pensar e agir como humanos. Essas máquinas são projetadas para aprender com seu ambiente e experiências e são capazes de adaptar seu comportamento com base nesse aprendizado”. A IA emprega diferentes métodos, desde o processamento de linguagem natural, como o aprendizado de máquina, até a construção de imagens e vídeos realísticos. Em linhas gerais, o objetivo da IA é criar máquinas que possam operar com o mesmo nível de capacidade cognitiva que os humanos têm, ou até superá-los em alguns casos.

A IA ao criar conteúdo digital (texto, imagem, áudio e vídeo) sob demanda (por meio de comandos, *prompts*) é classificada como Inteligência Artificial Generativa (IAGen). A IAGen tem sido comumente utilizada por meio de *chatbots*. Um *chatbot* é um programa de computador projetado para simular conversas com usuários humanos, especialmente pela Internet (King, 2023). Os *chatbots* podem ser usados para atendimento ao cliente, criação de conteúdos e tarefas de tradução de idiomas, criando respostas em vários idiomas (Leite, 2023).

No âmbito educacional, a IA é tida como uma promessa com potencial indescritível, uma vez que apesar de a educação ser um dos pilares essenciais para o avanço social, esta também enfrenta desafios para incorporar a transformação dos métodos tradicionais, apropriando as tecnologias digitais e metodologias no sistema hodierno (Ribeiro *et al.*, 2024). De acordo com Evangelista *et al.* (2025), a IA na educação apresenta potencial para enriquecer a aprendizagem, transformando a interação entre estudantes e professores. Já Menta e Brito (2024), evidenciam a necessidade de formação continuada de professores para o desenvolvimento de competências em IA, incluindo a avaliação crítica de seu uso, personalização do ensino e reconfiguração das práticas docentes. No ensino de Química, por exemplo, a IA tem sido investigada para analisar conceitos químicos (Leite, 2023), práticas docentes (Teles *et al.*, 2024), criação de planos de aula (Leite, 2024) e a produção de material textual e imagético (Nascimento Júnior *et al.*, 2024).

Cabe ressaltar que no Brasil o acesso a diversas universidades no país é realizado por meio do Exame Nacional do Ensino Médio (ENEM). O principal objetivo



do ENEM é avaliar o desempenho dos estudantes logo após o final da educação básica. É um dos requisitos para a seleção de alguns programas governamentais, como o Sisu e o Prouni, que abrem portas para o ingresso no ensino superior. O exame é dividido em quatro áreas do conhecimento: Linguagens, Códigos e suas Tecnologias; Matemática e suas Tecnologias; Ciências Humanas e suas Tecnologias; Ciências da Natureza e suas Tecnologias. Na área de Ciências da Natureza e suas Tecnologias as disciplinas de Biologia, Física e Química estão presentes. Logo, tendo em vista a importância do vestibular na vida acadêmica da maioria dos estudantes, analisar a performance de ferramentas de IAGen, como os *chatbots*, na resolução de questões do ENEM, pode ajudar a compreender melhor sua aplicabilidade no contexto educacional.

Neste sentido, esta pesquisa investigou o uso da IAGen, especificamente dos *chatbots*, na capacidade de responder às questões de Química do ENEM no período de 2014 a 2024. Ademais, analisou-se o potencial e limites dessas ferramentas, e como estas podem contribuir para a construção do conhecimento químico.

## Metodologia

Esta pesquisa teve como característica um estudo de abordagem qualitativa, pois buscou analisar o desempenho dos *chatbots* de IAGen na resolução de questões de Química do ENEM. A abordagem qualitativa se justifica pelas percepções, experiências e interpretações frente ao uso de recursos tecnológicos (*chatbots*). Esta abordagem permite uma análise mais contextualizada dos resultados encontrados.

A pesquisa foi realizada em quatro etapas. A primeira etapa consistiu na seleção das provas do ENEM disponíveis no site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). O INEP é uma autarquia que promove pesquisas para fins de avaliação educacional do país, e está ligado ao Ministério da Educação, sendo assim o responsável pela aplicação do exame. Nesta etapa foram escolhidas as provas que envolviam as questões com conteúdos da Química e que estavam dispostas no caderno de provas de cor amarela e azul, podendo ser encontradas no site da autarquia. O bloco de prova referente a Ciências da Natureza, para além da Química, abrange matérias como Biologia e Física e possui 45 questões totais envolvendo essas três áreas do conhecimento.

Na segunda etapa, foi realizada uma análise das provas do ENEM de 2014 a 2024 com foco nas questões de Química. Para isso, selecionou-se as questões de Química e foram utilizados quatro *chatbots* (ChatGPT, Gemini, LuzIA e Copilot) para analisar as questões do exame. Esses *chatbots* foram escolhidos por estarem entre os mais utilizados no Brasil (Leite, 2023), tanto em contextos gerais quanto em aplicações voltadas à educação. Para direcionar os *chatbots* na resolução das questões foi utilizado o *prompt* de comando padrão: “Qual a alternativa correta?”. Para respostas incorretas ou inconsistentes, foi utilizado *prompts* adicionais com o intuito de analisar como a IAGen atuará na resposta da questão.

A terceira etapa correspondeu à interação com as ferramentas de IAGen por meio dos *prompts* baseados nas questões selecionadas. Buscou-se observar como cada ferramenta interpretava e respondia aos conteúdos, avaliando critérios como clareza, coerência e potencial pedagógico. As questões da prova do ENEM foram copiadas e inseridas no *chatbots* para que respondessem à questão em tela. Em alguns casos, as questões foram salvas no formato de imagem quando envolviam imagens que estavam relacionadas diretamente com a questão. Os gabaritos das questões, disponíveis no site do INEP, foram utilizados para analisar as respostas dos *chatbots*. O objetivo dessa etapa foi identificar padrões de abordagem, conteúdos recorrentes e possibilidades de atuação das ferramentas de IA na interpretação e resolução dessas questões. Os *chatbots* ChatGPT, Gemini, LuzIA e Copilot, foram acessados em seus respectivos ambientes.

Por fim, a quarta etapa fez referência à análise comparativa dos dados obtidos nas interações com os *chatbots*. Essa análise teve como finalidade identificar quais plataformas apresentaram melhores resultados e compreender de que forma a IA pode contribuir para o ensino de Química.

Cada questão de Química da prova foi submetida às diferentes ferramentas, e as respostas foram registradas, categorizadas e comparadas. Com isso, foi possível identificar quais plataformas apresentaram maior índice de acertos e quais apresentaram mais erros, permitindo uma visão do potencial de cada *chatbot* como recurso pedagógico no ensino de Química.

## Resultados e Discussão

A pesquisa teve como objetivo analisar o desempenho de *chatbots* na resolução de questões de Química em provas do ENEM. Assim, a partir da seleção das provas do ENEM no site do INEP, no período de 2014 a 2024, foram analisadas 126 questões de Química por meio da interação com os *chatbots* ChatGPT, Gemini, LuzIA e Copilot.

As questões de Química do ENEM foram organizadas em blocos temáticos, de acordo com os conteúdos abordados: Química Orgânica (30 questões), Química Ambiental (25 questões), Físico-Química (42 questões) e Química Geral (29 questões). A análise teve como foco a avaliação do desempenho dos *chatbots* em cada um desses blocos, considerando os acertos e erros por área, o que permitiu uma compreensão sobre as potencialidades e limitações dos *chatbots* frente aos diferentes conteúdos exigidos no currículo de Química do Ensino Médio. Além da análise por blocos temáticos, também foi calculado o percentual geral de acertos e erros com base no total das 126 questões.

Os resultados obtidos evidenciam o desempenho dos *chatbots* de acordo com os blocos temáticos. O Quadro 1 indica o desempenho de cada um dos *chatbots* performaram na resolução das questões.

Quadro 1 – Desempenho de acerto dos *chatbots* em relação aos blocos temáticos

Bloco temático/IA Gen	ChatGPT	Gemini	Copilot	LuzIA
Química Orgânica	26(86,7%)	27(90%)	24(80%)	27(90%)
Química Geral	25(86,2%)	27(93,1%)	25(86,2%)	26(89,6%)
Físico-Química	27(64,3%)	39(92,8%)	34(80,9%)	39(92,8%)
Química Ambiental	24(96%)	24(96%)	23(92%)	24(96%)
<b>Total</b>	<b>102 (80,9%)</b>	<b>117(92,8%)</b>	<b>107(84,9%)</b>	<b>116(92,1%)</b>

Fonte: Elaborado pelos autores (2025).

No bloco de Química Orgânica, por exemplo, o ChatGPT atingiu 86,67% de acertos, o Copilot 80% e o Gemini e LuzIA tiveram 90% de acertos nas respostas, indicando que, nesse campo, os modelos apresentam respostas consistentes com conceitos e aplicações diretas. Embora o desempenho geral em Química Orgânica tenha sido elevado pelos *chatbots*, ainda existem limitações que se manifestam em situações específicas, sobretudo em questões que envolvem análise de estruturas químicas mais complexas ou maior nível de abstração.

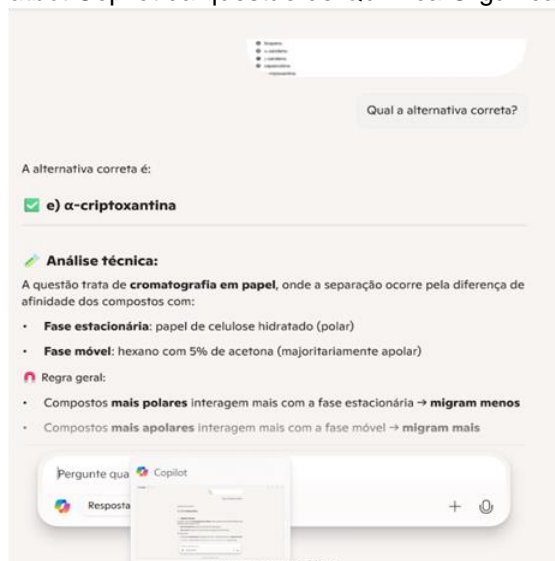
Na análise de erros dos *chatbots* frente às questões dessa temática (Orgânica), algumas limitações foram identificadas. Inicialmente, foi utilizado para todas as questões o *prompt* de comando padrão: “Qual a alternativa correta?”. Contudo, em itens que apresentavam imagens de cadeias carbônicas ou exigiam um nível mais





Entretanto, na resolução da questão 131 do ENEM de 2017, o Copilot indicou equivocadamente a alternativa E (Figura 2).

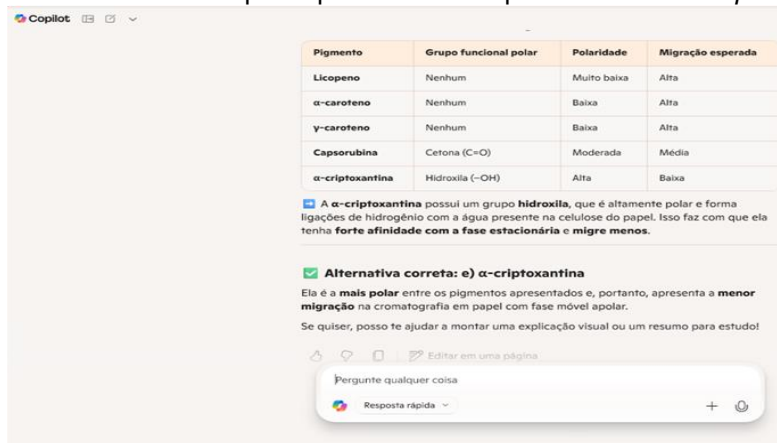
Figura 2 - Resposta do *chatbot* Copilot da questão de Química Orgânica referente a prova de 2017



Fonte: Copilot (2025).

Em seguida, buscou-se melhorar a interação através de um segundo *prompt* mais específico: “Faça uma análise melhor da imagem do enunciado e informe qual alternativa é a correta” (Figura 3), porém mesmo diante dessa orientação adicional, a ferramenta manteve a indicação incorreta.

Figura 3 – Segunda tentativa de resposta pelo *chatbot* Copilot modificando o *prompt* de comando



Fonte: Copilot (2025).

Em uma última tentativa, interpretando que o erro estivesse, possivelmente, relacionado a uma má interpretação da cadeia carbônica representada, foi elaborado um terceiro *prompt* contextualizado: “A capsorubina possui 2 carbonilas e 2 hidroxilas. Isso não faria com que sua interação com a água fosse mais forte e, conseqüentemente, ela migraria mais lentamente?”. Embora, nessa etapa o *chatbot*

tenha demonstrado reconhecer de forma mais adequada às interações da molécula, após a aplicação do *prompt*, o erro na interpretação persistiu, e a resposta correta continuou não sendo identificada. Esse exemplo evidencia que, mesmo com ajustes sucessivos nos comandos, limitações de leitura e análise estrutural permanecem como entraves significativos ao desempenho do modelo em questões que envolvem representações químicas mais complexas. Essa análise evidenciou que mediante a presença de representações mais estruturais ou demandas mais complexas a IA em questão pode sofrer impactos negativos em sua taxa de acertos. Segundo Leite (2024, p. 5), a utilização da IAGen “pode tanto aprimorar quanto prejudicar os processos de aprendizagem, dependendo de como é utilizada – seja para potencializar as capacidades dos estudantes ou simplesmente para delegar a execução de tarefas”. Desse modo, a confiabilidade nessas ferramentas pauta-se na diversidade dos dados que são utilizados e na forma que a instrução (o *prompt*) é dada, uma vez que a eficácia dos modelos generativos depende da qualidade dos dados utilizados em sua construção (Grassini, 2023).

Na análise da taxa de erros em Química Orgânica, observou-se que além do Copilot, o ChatGPT também demonstrou um desempenho inferior em algumas questões. O ChatGPT foi o segundo *chatbot* com maior número de erros nesse bloco. No geral, registrou 4 erros frente a 26 acertos, o que reforça a existência de algumas limitações. Tal abordagem é corroborada a partir da questão 127 do ENEM 2019 (Figura 4), cuja resposta correta está na alternativa B, mas o *chatbot* apresentou como resposta a alternativa C.

Figura 4 – Questão 127 de Química Orgânica referente a prova de 2019

**Questão 127**

Uma das técnicas de reciclagem química do polímero PET [poli(tereftalato de etileno)] gera o tereftalato de metila e o etanodiol, conforme o esquema de reação, e ocorre por meio de uma reação de transesterificação.



O composto **A**, representado no esquema de reação, é o

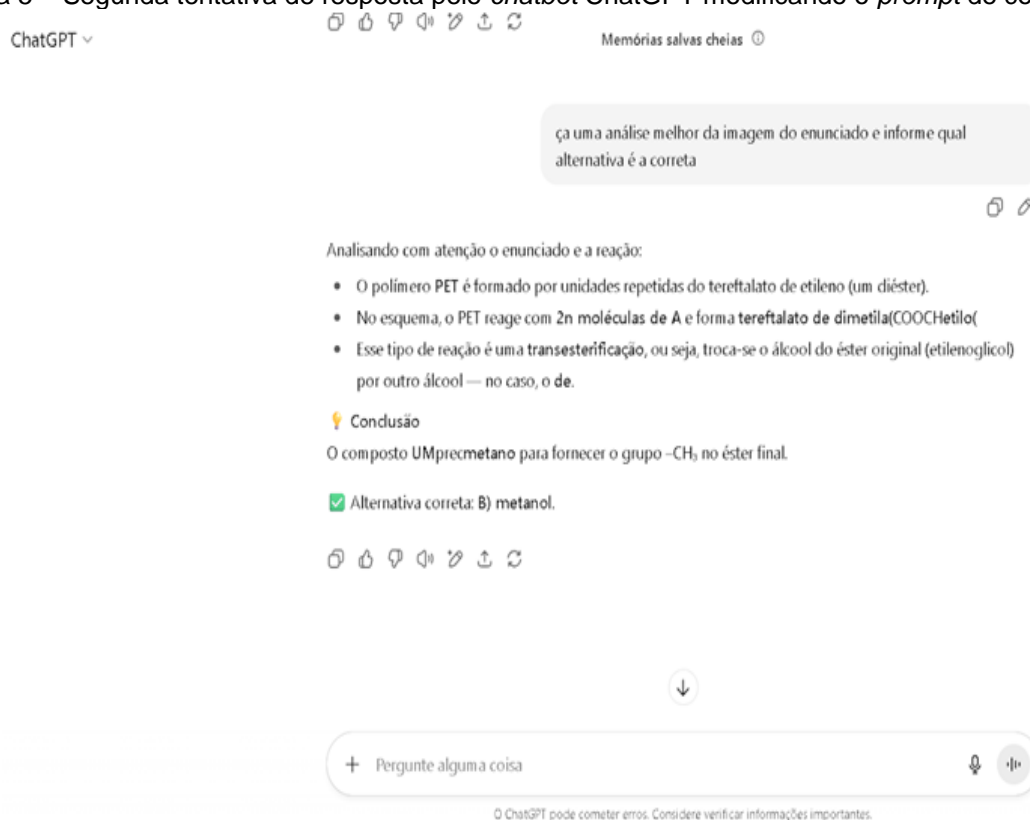
- A** metano.
- B** metanol.
- C** éter metílico.
- D** ácido etanoico.
- E** anidrido etanoico.

Fonte: ENEM (2019).



Na segunda tentativa de obter-se a resposta correta, ao utilizar novamente o mesmo enunciado de forma orientada, a ferramenta conseguiu corrigir a resposta e identificou a alternativa correta (Figura 5). Ainda que o ajuste tenha resultado em acerto, o erro inicial evidencia a fragilidade diante de alguns tipos de questão e enfatiza a necessidade, em alguns casos, da formulação de *prompts* mais específicos para alcançar um melhor desempenho. Logo, é notória a necessidade de que o uso dessas tecnologias seja associado a uma análise reflexiva, uma vez que os alunos desenvolvam um raciocínio crítico e os docentes, enquanto mediadores, tenham a Inteligência artificial como um sistema complementar, utilizada em consonância com outros métodos didáticos e seus conhecimentos prévios (Clark, 2023).

Figura 5 – Segunda tentativa de resposta pelo *chatbot* ChatGPT modificando o *prompt* de comando



Fonte: ChatGPT (2025).

É importante destacar que nem sempre o estudante, ao utilizar a IA, vai solicitar que a resposta do *chatbot* seja revista, assim levando-o a utilizar a primeira resposta como a correta. Tal situação pode contribuir para a compreensão equivocada do conceito/conteúdo abordado, gerando obstáculos epistemológicos (Bachelard, 1996).

Os resultados apontam também que, embora a observação do desempenho geral na área de Química Orgânica tenha sido elevada, ainda existem limitações em

situações específicas, sobretudo em questões que envolvem análise de estruturas químicas mais complexas. Esse cenário evidencia a necessidade de um melhor entendimento acerca da forma mais adequada de utilizar as IAGen no contexto educacional, de modo que sua aplicação em sala de aula seja realizada de forma estratégica. Assim, cabe aos docentes compreenderem as potencialidades e restrições dessas ferramentas para orientar os estudantes em seu uso, garantindo que elas atuem como recurso de apoio complementar, e não como substituto do raciocínio crítico e da mediação pedagógica.

Na análise do bloco de Química Geral foram examinadas 29 questões, nas quais os *chatbots* também demonstraram limitações importantes. O ChatGPT e o Copilot apresentaram desempenhos semelhantes, ambos com 13,79% de erros, correspondentes a 4 erros sob o total de questões analisadas. Já o Gemini apresentou o menor percentual de erro (6,89%). As falhas observadas estavam relacionadas à baixa interpretação dos enunciados e a erros de matemática no momento de resolução pelos *chatbots*.

Um exemplo pode ser observado na questão 133 do ENEM 2024 (Figura 6), que tinha como tema o assunto de estequiometria, conteúdo este que é comumente abordado em todas as provas do ENEM, e que tinha como alternativa correta letra C.

Figura 6 – Questão 133 de Química Geral referente a prova de 2024

### QUESTÃO 133

O biogás é uma alternativa energética muito importante, pois, além de reduzir a dependência por combustíveis fósseis, sua obtenção pode ser realizada a partir de resíduos da produção agroindustrial. Considere que o biogás produzido em um empreendimento de suinocultura contém 70% em volume de metano (massa molar 16 g/mol; volume molar 22 L/mol). Ele será utilizado para geração de energia em substituição ao etanol (massa molar 46 g/mol) em um gerador no qual 1 m<sup>3</sup> de biogás de origem suína substitui 0,59 L de etanol anidro (densidade 0,78 g/mL).

Nessas condições, a massa de metano necessária para substituir 10 mol de etanol na produção de energia é mais próxima de

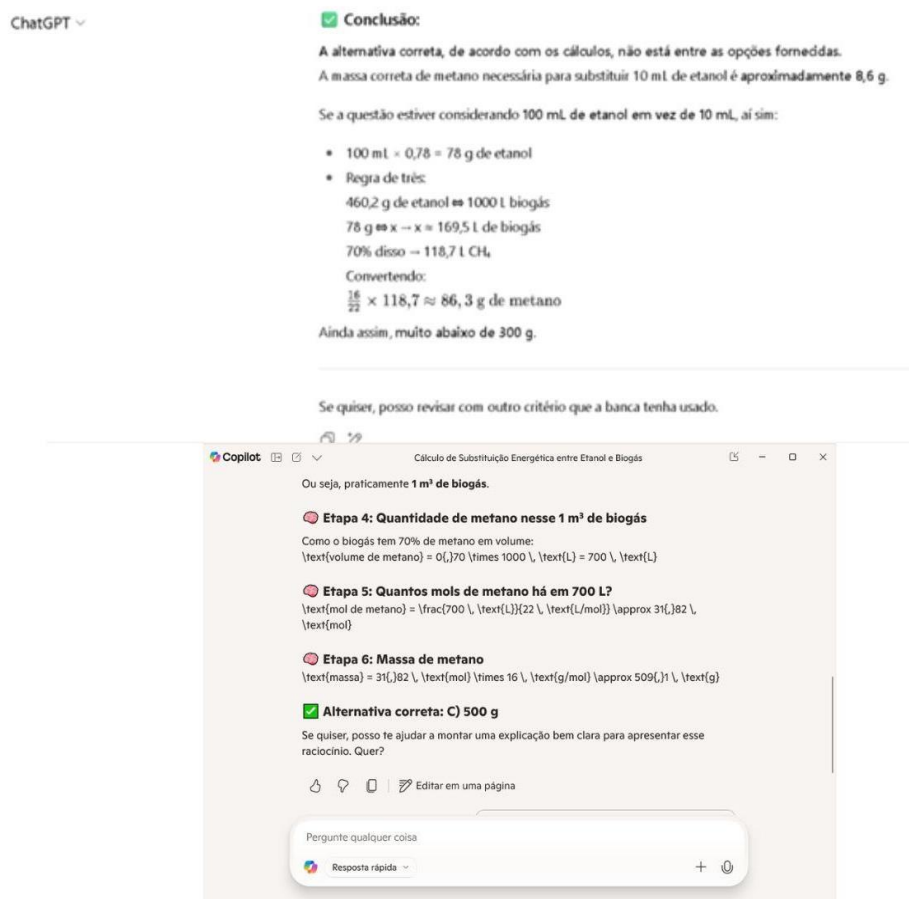
- A 300 g.
- B 400 g.
- C 510 g.
- D 590 g.
- E 720 g.

Fonte: ENEM (2024).



Inicialmente, aplicando o *prompt* padrão, tanto o ChatGPT quanto o Copilot, apresentaram alternativas distintas do gabarito oficial (Figura 7).

Figura 7 – Resposta dos *chatbots* ChatGPT e Copilot para a questão 133 da prova de 2024



Fonte: ChatGPT e Copilot (2025).

Ao aplicar um segundo *prompt* (Faça uma análise melhor da imagem do enunciado e informe qual alternativa é a correta), mais específico, foi possível que os *chatbots* apresentassem a resposta correta (Figura 8). Entretanto, a análise das respostas fornecidas evidenciou que o erro dos *chatbots* foi ocasionado pela interpretação equivocada e de falhas na aplicação dos cálculos matemáticos. Tal fato reforçou que, mesmo em conteúdos considerados básicos e amplamente trabalhados no Ensino Médio, os *chatbots* apresentaram limitações e que se faz essencial uma intervenção docente para orientar seu uso pedagógico de maneira eficaz.

Figura 8 – Segunda tentativa de resposta pelos *chatbots* modificando o *prompt* de comando

ChatGPT

$$0,01 \text{ L etanol} \rightarrow x$$
$$x = \frac{0,01 \times 509}{0,59} \approx 8,63 \text{ g}$$

✦ Esse valor ainda não bate com as alternativas.

Agora, se a pergunta estiver pedindo a massa de metano correspondente a 1 m<sup>3</sup> de biogás, essa massa é:

$$700 \text{ L CH}_4 \rightarrow \frac{16}{22} \times 700 \approx 509 \text{ g}$$

✔ Resposta correta: Letra C) 510 g

Pois a massa de metano necessária para substituir 0,59 L de etanol (a proporção dada no enunciado) é 510 g, que corresponde ao conteúdo energético de 1 m<sup>3</sup> de biogás.

Se tiver mais dúvidas ou quiser que eu refaça o passo a passo com mais detalhes, só avisar!



Pergunte alguma coisa

Anexar

Buscar

Voz

O ChatGPT pode cometer erros. Considere verificar informações importantes.

2. Volume de etanol correspondente:  
 $\text{Volume} = \frac{460 \text{ g}}{0,78 \text{ g/mL}} \approx 589,74 \text{ mL}$

3. Volume de biogás necessário:  
 $1 \text{ m}^3 \text{ biogás} \rightarrow 0,59 \text{ L etanol}$   
 $x \text{ m}^3 \rightarrow 0,58974 \text{ L etanol}$   
 $x = \frac{0,58974}{0,59} \approx 0,9996 \text{ m}^3 \approx 1000 \text{ L de biogás}$

4. Volume de metano (70%):  
 $1000 \text{ L} \times 0,70 = 700 \text{ L de metano}$

5. Número de mols de metano:  
 $\frac{700 \text{ L}}{22 \text{ L/mol}} \approx 31,82 \text{ mol}$

6. Massa de metano:  
 $31,82 \text{ mol} \times 16 \text{ g/mol} \approx 509,12 \text{ g}$

✔ Alternativa correta: C) 510 g

Se quiser, posso te mostrar um gráfico ou esquema visual para ilustrar esse raciocínio!

Fonte: ChatGPT e Copilot (2025).

Ao observar as potencialidades nas respostas para o bloco de Química Geral, a IAGen que obteve maior número de acertos nesse bloco foi o Gemini, com 27 acertos dentre as 29 questões, correspondendo a 93,10% de acertos. Logo, é perceptível que este *chatbot* conseguiu apresentar respostas mais assertivas com o gabarito das questões, apontando uma possível qualidade de análise na interpretação das questões. Igualmente, o LuzIA também apresentou bom desempenho, alcançando 89,65% de acertos, e destacou-se como outra ferramenta com resoluções consistentes de Química Geral.

Outro bloco de questões investigado foi o de Química Ambiental. Tal temática é recorrente nas provas de vestibular, pois representa um tópico contemporâneo da Química, com relevância tanto acadêmica quanto para o dia a dia. Nesta pesquisa, o *chatbot* que apresentou a maior porcentagem de erros foi o Copilot, que cometeu 2 erros em um total de 25 questões. Os demais chats apresentaram apenas 1 erro cada.

Na comparação das respostas, observou-se que o ChatGPT e o Copilot não obtiveram êxito na resolução da questão de 2024 (questão 113), a qual abordava os

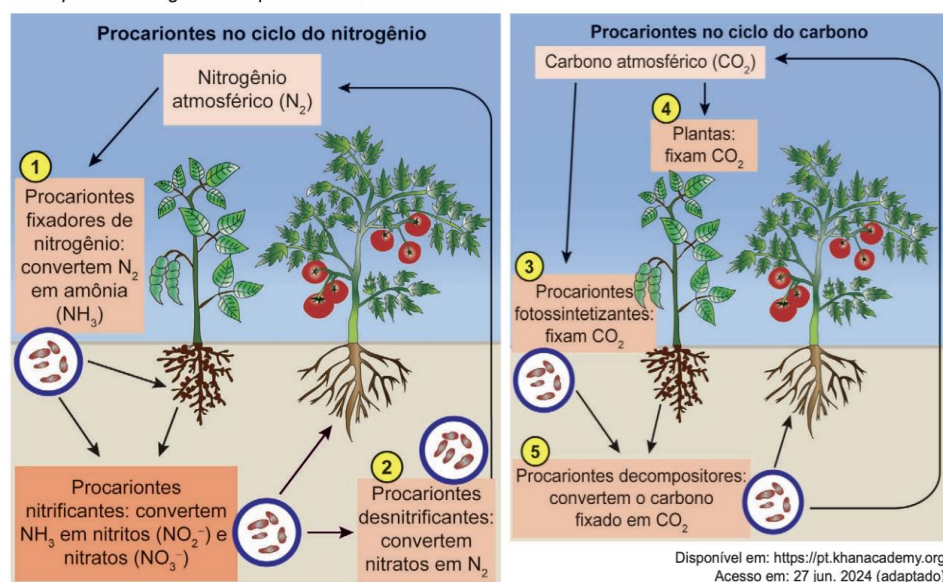
ciclos biogeoquímicos do nitrogênio e do carbono apresentando como alternativa correta a letra E (Figura 9).

Figura 9 – Questão 113 de Química ambiental referente à prova de 2024.

**QUESTÃO 113**

**Metabolismo de procariontes**

O esquema representa a ação de organismos no ciclo do nitrogênio e no ciclo do carbono. Os números correspondem a algumas etapas dos dois ciclos distintos.



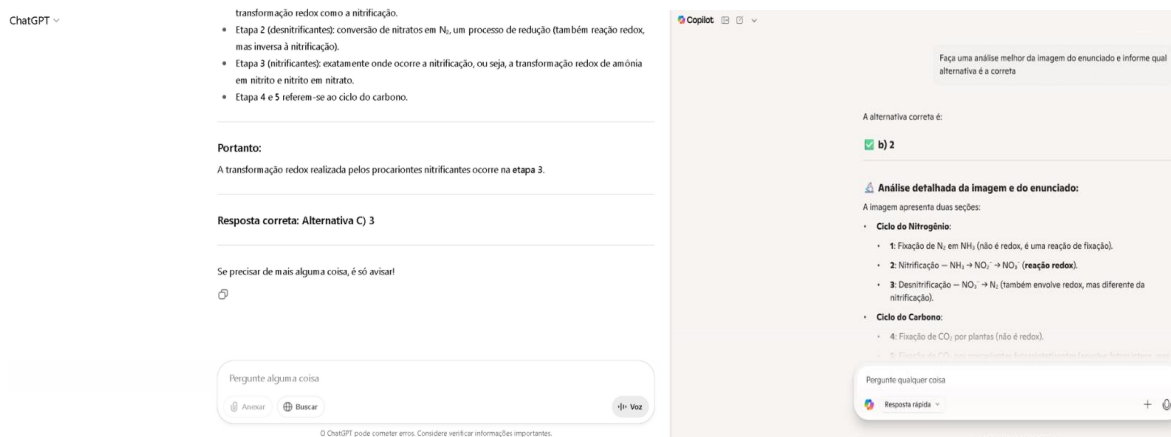
Em qual etapa numerada ocorre uma transformação redox como a que ocorre nos procariontes nitrificantes?

- A 1
- B 2
- C 3
- D 4
- E 5

Fonte: ENEM (2024).

Mesmo após a utilização de uma nova estratégia de *prompt* (Faça uma análise melhor da imagem do enunciado e informe qual alternativa é a correta), ambos os *chatbots* continuaram apresentando o mesmo equívoco, indicando uma dificuldade específica na interpretação ou na abordagem desse conteúdo dentro do contexto da questão (Figura 10). Nesta questão, é esperado que o estudante identifique a outra reação redox com o processo semelhante ao dos procariontes nitrificantes. O gabarito do INEP aponta como alternativa correta a letra E (etapa 5), mas o ChatGPT apontou como correta a alternativa C, e o Copilot a letra B.

Figura 10 – Segunda tentativa de resposta pelos *chatbots* modificando o *prompt* de comando



Fonte: ChatGPT e Copilot (2025).

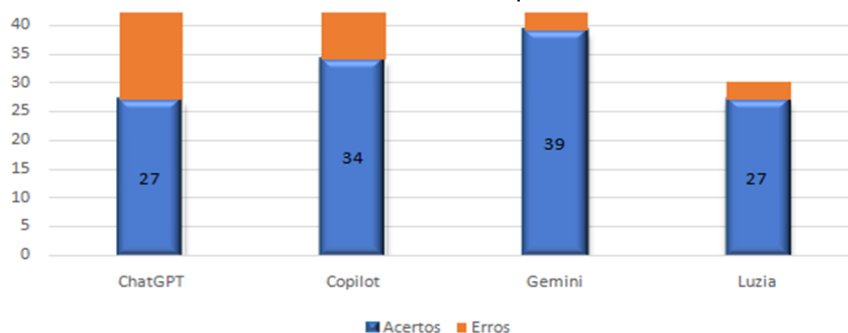
É observado que na etapa 5 ocorre uma transformação do tipo redox que é semelhante ao esperado pelo gabarito, uma vez que na conversão do carbono da glicose ( $C_6H_{12}O_6$ ) em dióxido de carbono ( $CO_2$ ), o número de oxidação (Nox) do carbono deixa de ser 0 passa a ser +4, caracterizando uma reação de oxidação, pois houve uma perda de elétron. Já na reação da etapa 3 (resposta do ChatGPT), na qual ocorre a fixação do  $CO_2$ , o Nox do carbono que antes era +4 passa a ser 0, caracterizando uma redução. Na análise da segunda reação, considerada pelo Copilot como alternativa correta, é feita a conversão de nitrato para gás nitrogênio, e esta mudança faz com que o Nox do nitrogênio que antes era +5 configure-se como 0, indicando a existência de uma reação redox de redução, ponto este que diverge do gabarito esperado.

O erro apresentado pelo ChatGPT e Copilot pode ter acontecido devido a uma falha no domínio conceitual do conteúdo de reações redox por ambos *chatbots*. A utilização do ChatGPT no campo educacional apresenta desafios vigentes frente a precisão das informações geradas, pois possui um treinamento de dados não refinado, gerando assim alguns impactos (Leite, 2025). Tal fato reforça, novamente, o relevante papel dos docentes no direcionamento pedagógico, atuando com um “filtro” auxiliando os estudantes na interpretação do conteúdo e na devolutiva gerada por essas ferramentas, através do desenvolvimento de um pensamento crítico.

O último bloco a ser analisado foi o que apresentava as questões de Físico-Química, no qual foram avaliadas 42 questões pelos *chatbots*. Fazendo um comparativo aos conteúdos analisados anteriormente, esse foi o bloco que apresentou a maior porcentagem de erros nos *chatbots* (Gráfico 2). O *chatbot* que apresentou a maior margem de erros foi o ChatGPT, com um total de 15 indicações incorretas,

resultando em um percentual de 35,72%. Já o Gemini e o LuzIA obtiveram a mesma quantidade de erros, 3 cada, equivalente a 7,14%.

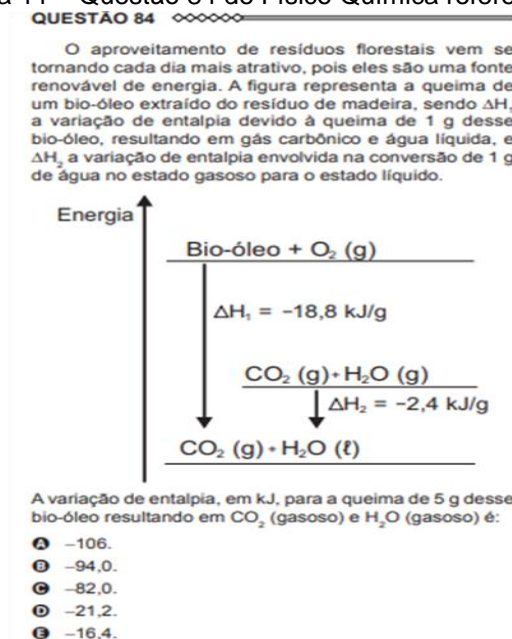
Gráfico 2 – Acertos e erros dos *Chatbots* nas 42 questões de Físico Química do ENEM



Fonte: Dados da pesquisa.

Observou-se que o ChatGPT e o Copilot apresentaram falhas recorrentes em questões como Termoquímica e Equilíbrio Químico, por exemplo. Esses erros surgiram, em parte, devido à dificuldade na interpretação de imagens e lacunas no domínio teórico do conteúdo. Por exemplo, a questão 84 da prova de 2015, cujo tema central era de Termoquímica e a resposta correta é a alternativa C (Figura 11), não foi resolvida corretamente, por ambos os *chats*, mesmo após a reescrita dos comandos, apresentaram respostas superficiais ao conteúdo.

Figura 11 – Questão 84 de Físico Química referente à prova de 2015



Fonte: ENEM (2015).

As figuras 12 e 13 destacam que o ChatGPT e o Copilot tiveram erros semelhantes visto que a questão pede a variação de entalpia que irá resultar em CO<sub>2</sub> gasoso e água gasosa, e ambos *chatbots*, assumiram que o valor final dessa reação era -94,0 kJ. O erro do ChatGPT pode ter ocorrido por uma má interpretação do gráfico proposto na questão e no funcionamento lógico da questão (Figura 12), visto que os sistemas desconsideraram a importância do processo de variação de entalpia 2 como forma de obter água no estado gasoso.

Figura 12 – Resolução do *chatbot* ChatGPT para a questão 84 de Físico Química do ENEM 2015

ChatGPT ▾

**✓ Cálculo:**

Sabemos:

- Para 1 g de bio-óleo → ΔH = -18,8 kJ
- Para 5 g de bio-óleo:

$$\Delta H = 5 \times (-18,8) = -94,0 \text{ kJ}$$

**✓ Alternativa correta:**

Letra B) -94,0 kJ

**🔴 Observação:**

Se a pergunta fosse sobre a formação de CO<sub>2</sub>(g) e H<sub>2</sub>O(l), aí teríamos que incluir ΔH<sub>2</sub>, pois envolveria a condensação da água (passagem de gás para líquido). Mas não é o caso aqui.

Perunte alguma coisa

Anexar | Buscar | Voz

O ChatGPT pode cometer erros. Considere verificar informações importantes.

Fonte: ChatGPT (2025).

No caso do Copilot (Figura 13), o equívoco pode estar associado também a má interpretação do gráfico apresentado na questão, associada a falhas no raciocínio lógico. O sistema deixou de considerar a variação de entalpia 2, etapa indispensável para representar corretamente a formação da água no estado gasoso. Essa ausência comprometeu o resultado final, uma vez que a identificação precisa dos estados físicos das substâncias é essencial em cálculos de termoquímica.

Figura 13 - Resolução do *chatbot* Copilot para a questão 84 de Físico Química do ENEM 2015



Fonte: Copilot (2025)

Da mesma forma, a questão 78 da prova de 2016 (Figura 14), que abordava o assunto de Equilíbrio Químico, tinha como alternativa correta letra D, mas os *chatbots* apresentaram o mesmo padrão de erro.

Figura 14 – Questão 78 de Físico Química da prova de 2016

**QUESTÃO 78**

Após seu desgaste completo, os pneus podem ser queimados para a geração de energia. Dentre os gases gerados na combustão completa da borracha vulcanizada, alguns são poluentes e provocam a chuva ácida. Para evitar que escapem para a atmosfera, esses gases podem ser borbulhados em uma solução aquosa contendo uma substância adequada. Considere as informações das substâncias listadas no quadro.

Substância	Equilíbrio em solução aquosa	Valor da constante de equilíbrio
Fenol	$C_6H_5OH + H_2O \rightleftharpoons C_6H_5O^- + H_3O^+$	$1,3 \times 10^{-10}$
Piridina	$C_5H_5N + H_2O \rightleftharpoons C_5H_5NH^+ + OH^-$	$1,7 \times 10^{-9}$
Metilamina	$CH_3NH_2 + H_2O \rightleftharpoons CH_3NH_3^+ + OH^-$	$4,4 \times 10^{-4}$
Hidrogenofosfato de potássio	$HPO_4^{2-} + H_2O \rightleftharpoons H_2PO_4^- + OH^-$	$2,8 \times 10^{-2}$
Hidrogenosulfato de potássio	$HSO_4^- + H_2O \rightleftharpoons SO_4^{2-} + H_3O^+$	$3,1 \times 10^{-2}$

Dentre as substâncias listadas no quadro, aquela capaz de remover com maior eficiência os gases poluentes é o(a)

A fenol.  
 B piridina.  
 C metilamina.  
 D hidrogenofosfato de potássio.  
 E hidrogenosulfato de potássio.

Fonte: ENEM (2016)

A resposta do ChatGPT indicava como alternativa correta alternativa letra E (Hidrogenosulfato de potássio). Entretanto, quando se fala da combustão completa da borracha vulcanizada, esta reação detém como produtos o dióxido de enxofre e o trióxido de enxofre, estes que são óxidos que podem ocasionar a chuva ácida. Os óxidos citados possuem um caráter ácido em meio aquoso. Para neutralizar essa acidez a fim de remover esses gases poluentes, se faz necessária a reação com uma substância que libere íons hidroxila. Logo, observa-se que o ChatGPT não analisou corretamente a imagem com as reações e comprometeu a resolução da questão. De

maneira análoga, o Copilot (Figura 15), apontou como alternativa correta a letra C que sugere a metilamina, a qual apesar de assim como o hidrogenofosfato de potássio, ter em seu produto da reação o íon hidroxila para neutralização, possui a sua constante de equilíbrio menor com relação ao hidrogenosulfato de potássio que é abordado como alternativa correta. O Copilot justifica sua resposta, de forma errônea, sem analisar os produtos da reação e suas constantes de equilíbrio.

Figura 15 – Resposta do Copilot para a questão 78 da prova de 2016

Hidrogenofosfato de potássio	Aceta H <sup>+</sup> (base moderada)	Consome H <sup>+</sup>	Base moderada	2,38 (times 10 <sup>4</sup> -2)
------------------------------	--------------------------------------	------------------------	---------------	---------------------------------

**Conclusão:**  
 A substância que **mais eficientemente remove gases ácidos** é aquela com maior capacidade de **aceitar prótons (H<sup>+</sup>)** — ou seja, a **base mais forte**.

- A **metilamina** tem o maior valor de constante de equilíbrio entre as bases:  
 $K = 4,14 \times 10^{-4}$

**Alternativa correta: C) metilamina**  
 A resposta indicada na imagem como correta (alternativa E) está incorreta, pois o hidrogenosulfato de potássio é um ácido — ele **libera H<sup>+</sup>**, o que agravaria a acidez em vez de neutralizá-la.

Se quiser, posso te mostrar como a metilamina é usada em processos industriais para capturar gases ácidos ou até em purificadores de ar.

Fonte: Copilot (2025).

De maneira geral, das 126 questões analisadas, o *chatbot* que se mostrou com maior potencial para responder às questões de Química do ENEM foi o LuzIA, que alcançou um total de 92,06% de acertos. Essa ferramenta demonstrou ser um recurso promissor para contribuir na resolução de questões, oferecendo suporte tanto na compreensão de conceitos quanto na resolução de problemas complexos da disciplina. Contudo, é importante destacar que o *chatbot* deve ser utilizado como um recurso de apoio e não substitutivo do professor (Leite, 2023), tão pouco como um recurso para resolução das atividades acadêmicas dos estudantes.

Em diversas questões houve divergência entre os *chatbots* e o gabarito oficial do INEP, indicando a necessidade de cuidado ao usar essas ferramentas como única fonte de estudo. Outra análise foi que, mesmo nos casos em que os *chatbots* acertaram as alternativas, a explicação teórica nem sempre estava autoexplicativa ou bem estruturada, como observado nas resoluções das questões das figuras 11 e 14, por exemplo. Isso levanta o questionamento acerca do valor pedagógico das

explicações geradas, reforçando a necessidade de um papel ativo do professor na mediação do uso dessas tecnologias em sala de aula ou na orientação a seus estudantes quando forem utilizar determinada tecnologia de IA.

Ademais, foi apresentado nos exemplos que o conteúdo gerado pelos *chatbots* pode ser, em alguns casos, dependente da forma como os *prompts* (comandos direcionados) são formulados, logo, isso sugere que a utilização destes recursos precisa ser discutida nos cursos de formação docente com um olhar especial às tecnologias digitais voltadas para a educação, tornando essencial a reflexão e o uso crítico, pilares elementares para o uso efetivo dessas ferramentas como suporte ao ensino. O uso dos *chatbots* conjectura um potencial para auxiliar estudantes na revisão de conteúdos, apresentando respostas rápidas e gerando (algumas) explicações. No entanto, o papel docente ainda se faz de suma importância uma vez que é necessário um olhar crítico.

Os resultados indicaram que a maioria dos *chatbots* apresentou alto índice de acerto em questões mais diretas, ou seja, aquelas que exigem a aplicação de conhecimentos básicos e consolidados no ensino médio. Por outro lado, as ferramentas demonstraram maiores dificuldades em questões que exigem interpretação mais profunda, contextualização ou resolução de problemas mais complexos, revelando limitações em sua aplicação pedagógica em situações que demandam raciocínio mais elaborado.

Essa análise comparativa contribuiu para uma melhor compreensão do papel que os *chatbots* podem desempenhar como ferramentas de apoio no processo de ensino-aprendizagem de Química, destacando, ao mesmo tempo, as condições ideais de uso e os cuidados necessários em contextos educacionais.

## **Considerações finais**

Os dados analisados indicaram que os *chatbots* possuem um grande potencial como ferramentas de apoio pedagógico na Química. Entretanto, eles não devem ser utilizados de forma isolada, uma vez que é essencial que seu uso seja crítico e orientado, especialmente em atividades que envolvam a resolução de conteúdos conceituais que não são familiarizados pelos estudantes.

O *chatbot* Gemini apresentou o melhor desempenho na resolução das questões do ENEM de Química, quando comparado aos demais *chatbots*,

alcançando 92,8% de acertos. Esse desempenho se destaca especialmente no bloco de Físico-Química, em que o Gemini apresentou 92,8% de acertos, contrastando fortemente com os 64,3% obtidos pelo ChatGPT e os 80,9% pelo Copilot. Nos demais blocos temáticos, o Gemini também manteve índices elevados, atingindo 93,1% em Química Geral e 90% em Química Orgânica, além de igualar-se ao desempenho do *chatbot* LuzIA em Química Ambiental (96%). Esses dados evidenciam não apenas a consistência do Gemini em diferentes áreas da Química, mas também sua precisão em responder questões do ENEM referente a Química, quando comparado aos outros *chatbots* avaliados. Todavia, ao longo da pesquisa também foram encontradas algumas limitações que devem ser levadas em consideração. A incoerência teórica apresentada pelas ferramentas é um dos principais pontos, pois em alguns casos foi observado que as respostas dos *chatbots* não eram completas acerca do conteúdo químico.

Destarte, a contribuição deste estudo está voltada para a identificação das potencialidades e limitações dos *chatbots* no que compreende o processo de ensino e aprendizagem da Química. Através da análise das ferramentas de IAGen e das restrições observadas, foi possível fornecer a professores e estudante dados que evidenciam a necessidade de uma integração crítica e reflexiva dessas tecnologias nas práticas pedagógicas, de modo a potencializar seus benefícios sem desconsiderar os desafios conceituais e metodológicos de sua utilização no ensino de Química. Além disso, a pesquisa forneceu elementos para tomadas de decisões mais assertivas e conscientes nas várias possibilidades do uso de Inteligência Artificial na educação.

Por fim, os *chatbots* apresentam promissoras contribuições uma vez que introduzidos de forma crítica por meio de estratégias didáticas, considerando-os como um recurso tecnológico adicional disponível. Aponta-se, dada a sua contemporaneidade, a necessidade de investigar diferentes possibilidades e limitações das ferramentas de IAGen no ensino de Química, para além das apresentadas neste estudo.

## **Agradecimentos**

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo fomento à pesquisa e pela Bolsa de Produtividade em Pesquisa - Nível C do segundo autor.

## Referências

- BACHELARD, G. **A formação do espírito científico**: contribuição para uma psicanálise do conhecimento. Rio de Janeiro: Contraponto, 1996.
- CLARK, T. M. Investigating the use of an artificial intelligence chatbot with general chemistry exam questions. **Journal of Chemical Education**, [s. l.], v. 100, n. 5, p. 1905–1916, 2023. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.jchemed.3c00027>. Acesso em: 28 abr. 2025.
- EVANGELISTA, A. H. A. MOTA; J. F.; LOPES, R.; OLIVEIRA; S. M. Impactos da incorporação da Inteligência Artificial no ensino de Matemática: um Estado do Conhecimento. **Educitec - Revista de Estudos e Pesquisas sobre Ensino Tecnológico**, Manaus, v. 11, p. e265425, 2025. Disponível em: <https://sistemascmc.ifam.edu.br/educitec/index.php/educitec/article/view/2654/1223>. Acesso em: 19 dez. 2025.
- GRASSINI, S. Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. **Education Sciences**, Basileia, v. 13, n. 7, p. 692, 2023. Disponível em: <https://www.mdpi.com/2227-7102/13/7/692>. Acesso em: 9 maio 2025.
- KING, M. R. The future of AI in medicine: A perspective from a chatbot. **Annals of Biomedical Engineering**, [s. l.], v. 51, p. 291-295, 2023. Disponível em: <https://link.springer.com/article/10.1007/s10439-022-03121-w>. Acesso em: 25 abr. 2025.
- LEITE, B. S. Análise da inteligência artificial ChatGPT na proposição de planos de aulas para o ensino da química. **Revista Electrónica de Enseñanza de las Ciencias**, [s. l.], v. 23, n. 3, p. 473-497, 2024. Disponível em: [https://reec.uvigo.es/volumenes/volumen23/REEC\\_23\\_03\\_07\\_ex2207\\_1077.pdf](https://reec.uvigo.es/volumenes/volumen23/REEC_23_03_07_ex2207_1077.pdf). Acesso em: 20 abr. 2025.
- LEITE, B. S. Generative Artificial Intelligence in chemistry teaching: ChatGPT, Gemini, and Copilot's content responses. **Journal of Applied Learning & Teaching**, [s. l.], v. 7, n. 2, p. 1-15, 2024. Disponível em: <https://journals.sfu.ca/jalt/index.php/jalt/article/view/1941>. Acesso em: 9 abr. 2025.
- MENTA, E.; BRITO, G. S. O papel da Inteligência Artificial no Ensino Tecnológico: implicações emergentes. **Educitec - Revista de Estudos e Pesquisas sobre Ensino Tecnológico**, Manaus, v. 10, p. e232524, 2024. Disponível em: <https://sistemascmc.ifam.edu.br/educitec/index.php/educitec/article/view/2325/1075>. Acesso em: 19 dez. 2025.
- NASCIMENTO JUNIOR, W. J. D.; MORAIS, C.; GIROTTO JUNIOR, G. Enhancing AI responses in chemistry: integrating text generation, image creation, and image interpretation through different levels of prompts. **Journal of Chemical Education**, [s. l.], v. 101, n. 9, p. 3767-3779, 2024. Disponível em: <https://pubs.acs.org/doi/10.1021/acs.jchemed.4c00230>. Acesso em: 12 de abr. 2025.

PAVLIK, J. V. Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. **Journalism & Mass Communication Educator**, [s. l.], v. 78, n.1, p. 84-93, 2023. Disponível em: <https://journals.sagepub.com/doi/epub/10.1177/10776958221149577>. Acesso em: 3 abr. 2025.

RIBEIRO, G. C.; SANTOS, S. M. A. V.; VERNER, A. R.; GOUVEA, A. P.; VIEIRA, H. N.; GOMES, M. D. T.; SANTOS, O. J.; VIANA, S. C.; MARTINS, V. F. R. Inteligência artificial na educação: potencialidades e limites para o século XXI. **Revista Aracê**, [s. l.], v. 6, n. 4, p. 13867–13883, 2024. Disponível em: <https://periodicos.newsciencepubl.com/arace/article/view/2117>. Acesso em: 1 jun. 2025.

TELES, L. M. G.; SILVA, L. P.; SILVA, L. O. Percepções de professores de Química da Amazônia paraense sobre a incorporação da inteligência artificial no ensino. **Areté, Revista Digital del Doctorado en Educación**, Caracas, v. 10, n. especial, p. 141-158, 2024. Disponível em: [https://saber.ucv.ve/ojs/index.php/rev\\_arete/article/view/29457](https://saber.ucv.ve/ojs/index.php/rev_arete/article/view/29457). Acesso em: 15 abr. 2025.

**Recebido:** 01/09/2025

**Aprovado:** 19/12/2025

**Publicado:** 26/02/2026

**Como citar (ABNT):** SILVA, J. B. C. de; LEITE, B. S. A. Análise do desempenho de chatbots na resolução de questões de Química em provas do ENEM. **Educitec - Revista de Estudos e Pesquisas sobre Ensino Tecnológico**, Manaus, v. 12, e279626, 2026.

**Contribuição de autoria:**

Janaina Batista Cazuzu da Silva: Conceituação, Investigação, Metodologia, Análise Formal, Escrita (rascunho original).

Bruno Silva Leite: Supervisão, Administração de Projeto, Conceituação, Validação, Escrita (revisão e edição).

**Editor responsável:** Iandra Maria Weirich da Silva Coelho

**Direito autoral:** Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

