

Análise do desempenho de *chatbots* de inteligência artificial generativa na resolução de exercícios do Enem envolvendo conceitos da Química

An Analysis of the Performance of Generative Artificial Intelligence Chatbots in Solving ENEM Questions Involving Chemistry Concepts

Richardison Rufino de **CASTILHO**

Programa de Pós-graduação em Ensino de Ciências e Matemática – PPGECCMat/UFMG

Fernando Cesar **FERREIRA**

Programa de Pós-graduação em Ensino de Ciências e Matemática – PPGECCMat/UFMG

Correspondência do autor:

ri.07.ma@gmail.com

RESUMO

O presente trabalho tem como objetivo analisar o desempenho de quatro *chatbots* de inteligência artificial generativa: *ChatGPT*, *Gemini*, *CoPilot* e *DeepSeek*; na resolução de questões envolvendo conteúdos da disciplina de química retiradas da prova do Enem. A pesquisa, de natureza qualitativa e exploratória, avaliou 89 questões do caderno amarelo de provas realizadas entre 2019 e 2024. Foram utilizadas imagens das questões enviadas como anexos aos *chatbots*, cujas respostas foram comparadas aos gabaritos oficiais. Os resultados demonstram que, embora os modelos acertassem a maioria das questões, apresentaram limitações importantes, sobretudo em temas que exigiam raciocínio quantitativo, interpretação de imagens e visualização de estruturas químicas. Destacam-se dificuldades dos modelos com questões envolvendo conteúdos de estequiometria, química orgânica e representação de fórmulas. Conclui-se que, apesar de possuírem potencial como ferramentas de apoio ao estudo, o uso dos *chatbots* deve ser mediado criticamente, enfatizando a necessidade de supervisão docente e desenvolvimento da autonomia crítica dos estudantes.

Palavras-chave: Tecnologias Digitais, Resolução de Problemas, Aprendizagem Assistida por IA.

ABSTRACT

This study aims to analyze the performance of four generative artificial intelligence chatbots: ChatGPT, Gemini, Copilot, and DeepSeek; in solving chemistry questions taken from Brazil's National High School Exam (ENEM). The research, qualitative and exploratory in nature, evaluated 89 questions from the yellow test booklet administered between 2019 and 2024. Images of the questions were submitted to the chatbots as attachments, and their responses were compared with the official answer keys. The results show that although the models correctly answered most of the questions, they exhibited significant limitations, particularly in topics requiring quantitative reasoning, image interpretation, and the visualization of chemical structures. Notable difficulties were observed in questions involving stoichiometry, organic chemistry, and formula representation. The findings suggest that, despite their potential as study support tools, the use of chatbots should be



critically mediated, highlighting the need for teacher supervision and the development of students' critical autonomy.

Keywords: Digital Technologies; Problem Solving; AI-Supported Learning.



INTRODUÇÃO

As tecnologias digitais atualmente são muitas e muitas vezes acessíveis de forma ampla, possibilitando acesso instantâneo a todo tipo de conteúdo e contribuindo para os processos de aprendizagem quando utilizadas para essa finalidade. Segundo Araya, Gibin e Souza Filho (2021, p.13), “as Tecnologias Digitais de Informação e Comunicação (TDIC) têm um grande potencial de uso no ensino de ciências, possibilitando desenvolver novas metodologias de ensino”.

Conforme Leite (2023, p. 915), “Professores e estudantes utilizam aplicativos, softwares, vídeos, jogos digitais, internet etc.” e a utilização desses recursos no processo de ensino apresenta desafios. Isso porque, conforme o autor, historicamente à medida que a tecnologia evolui, adaptações no processo educacional se fazem necessárias conforme a inserção de tais ferramentas no ato de ensinar e aprender ocorre. Nesse contexto, o uso da inteligência artificial generativa (IAGen) representa um novo desafio quando se trata da reformulação das práticas pedagógicas e na construção do conhecimento, oferecendo novas possibilidades para o ensino e a aprendizagem bem como novos desafios.

Isso porque, apesar das contribuições advindas da utilização das ferramentas de IAGen nos processos de ensino e aprendizagem, é preciso também considerar os desafios e limitações relacionados ao seu uso. Dessa forma surgem preocupações sobre os impactos que o uso de tais ferramentas possam ter sobre o processo de aprendizagem, principalmente quando utilizadas de forma a executarem tarefas pelos alunos como, por exemplo, na resolução de exercícios, testes e avaliações.

Monteiro *et al.* (2024) argumentam que a IAGen possui potencialidades que podem beneficiar tanto professores quanto alunos na execução das atividades relacionadas à rotina escolar. Contudo, apesar disso, os autores destacam pontos a serem considerados na utilização de chatbots de IAGen como

ferramenta auxiliar nos processos de ensino e aprendizagem. Isso porque as ferramentas de inteligência artificial generativa encontram-se em constante evolução e apresentam limitações (Borba; Balbino Junior, 2023), exemplo disso é o fornecimento de respostas que parecem plausíveis, mas são incorretas ou sem sentido (OpenAI, 2022).

Assim, surgem discussões acerca do conteúdo gerado pelos modelos de IAGen em tarefas escolares. Monteiro *et al.* (2024) levantam preocupações acerca do plágio e da capacidade dos alunos em identificar quando uma resposta fornecida por um chatbot está conceitualmente errada, por exemplo. Costa Júnior *et al.* (2023) apontam para preocupações sobre a confiabilidade das informações fornecidas por tais ferramentas.

No contexto da aprendizagem da disciplina de química tais preocupações acerca do conteúdo fornecido pelos modelos de IA generativa emergem uma vez que, a abordagem de conceitos químicos utiliza-se na maioria das vezes de abstrações ao se considerar a natureza microscópica dos conhecimentos em química. Assim, é necessário dos alunos uma capacidade de generalização e abstração para que ocorra a compreensão dos conceitos o que gera frequentemente dificuldades de aprendizagem na matéria (Costa, Passerino, Zaro, 2012).

Conforme Rudolph, Tan e Tan (2023) apesar de ser capaz de fornecer respostas úteis sobre os mais variados tópicos, a qualidade das respostas fornecidas pelos chatbots irá depender em grande parte do tipo e da complexidade da pergunta feita, dependendo do nível de dificuldade da sua consulta podendo até mesmo conter informações incorretas, sendo assim recomendado pelas próprias ferramentas que os usuários realizem uma verificação da resposta fornecida.

Dessa forma o trabalho objetiva analisar o desempenho de quatro chatbots de inteligência artificial generativa: *ChatGPT*, *Gemini*, *CoPilot* e *DeepSeek*, na resolução de questões envolvendo conceitos da disciplina de química retiradas do Exame Nacional do Ensino Médio



(Enem), de forma a identificar suas potencialidades e limitações enquanto ferramentas auxiliares no processo de ensino e aprendizagem.

FUNDAMENTAÇÃO TEÓRICA

A compreensão da Inteligência Artificial (IA) parte, necessariamente, da problematização do conceito de inteligência para muitos autores. Coppin (2013) e Russell e Norvig (2013; 2021) iniciam essa discussão questionando “o que é inteligência?”, uma vez que a definição de IA depende diretamente da compreensão desse conceito.

Coppin (2013, p. 4) define inteligência como a “capacidade de solucionar problemas, de responder a questões, de engendrar planos e assim por diante”, enfatizando sua dimensão funcional e pragmática. Russell e Norvig (2013; 2021) associam inteligência à capacidade de tomada de decisões racionais, entendendo que um agente inteligente é aquele que seleciona a melhor ação possível com base nas informações disponíveis tendo a racionalidade como elemento central do comportamento inteligente.

Partindo dessas concepções, a IA é compreendida como o campo científico e tecnológico voltado ao desenvolvimento de sistemas capazes de realizar tarefas que, quando desempenhadas por humanos, são consideradas inteligentes. Jackson Jr. (1985, p. 1) define IA como “a habilidade de máquinas de fazerem coisas que as pessoas dizem exigir inteligência”, caracterizando-a também como um campo de pesquisa dedicado a descrever e reproduzir aspectos da inteligência humana em sistemas computacionais. De maneira semelhante, McCarthy (1995, p. 2) afirma que a IA é “a ciência e a engenharia de criar máquinas inteligentes, especialmente programas de computador inteligentes”, ressaltando que tais sistemas não precisam se limitar a métodos biologicamente observáveis.

Coppin (2013, p. 4) também define IA como “o estudo dos sistemas que agem de um modo que a um observador qualquer pareceria ser inteligente”. O autor propõe que a IA envolve a utilização de métodos baseados no comportamento inteligente de humanos e outros animais para a solução de problemas complexos evidenciando a dimensão observacional e metodológica do campo.

Russell e Norvig (2013; 2021) sistematizam as múltiplas definições de IA em quatro abordagens: “pensar como humanos”, “pensar racionalmente”, “agir como humanos” e “agir racionalmente”. Essa sistematização em categorias busca organizar as diferentes perspectivas teóricas do campo, ora enfatizando a modelagem da cognição humana, ora privilegiando a racionalidade formal e a maximização de desempenho.

Observa-se então, a partir das diferentes definições sobre IA, que essa engloba principalmente os sistemas computacionais voltados principalmente à classificação, regressão ou tomada de decisão (Feuerriegel et al., 2023). De forma complementar, Coppin (2013) realiza uma distinção entre IA forte (que pressupõe a possibilidade de máquinas totalmente conscientes) e IA fraca, que se concentra na simulação de comportamentos inteligentes sem implicar em consciência genuína. Dessa forma, as ferramentas analisadas no presente artigo inserem-se no âmbito da chamada IA fraca.

Partindo desse campo mais amplo da ciência da computação, emerge a Inteligência Artificial Generativa (IAGen), como um subcampo. Conforme Feuerriegel et al. (2023), a IAGen se caracteriza pela capacidade de gerar conteúdos inéditos, como textos, imagens e áudios, a partir de padrões aprendidos em grandes volumes de dados.

A IAGen, utiliza-se de técnicas de aprendizado profundo (*deep learning*), utilizando redes neurais artificiais com múltiplas camadas capazes de identificar padrões complexos em grandes conjuntos de dados. Tal desenvolvimento foi impulsionado pela



arquitetura *Transformer* que possibilitou a criação de modelos de larga escala, conhecidos como *Large Language Models* (LLMs) que são treinados com grandes quantidades de materiais textuais. Ao utilizar esse método de treinamento, o modelo aprende a estimar a probabilidade condicional de uma palavra dado o contexto, possibilitando assim a geração de textos coerentes e contextualmente adequados (Vaswani et al., 2017; Feuerriegel et al., 2023).

Conforme Feuerriegel et al. (2023) é importante compreender que a IAGen não opera com compreensão semântica no sentido humano, mas com modelagem estatística de padrões. Assim, ainda que os resultados possam aparentar coerência conceitual e sofisticação argumentativa, o processo subjacente baseia-se na previsão probabilística de sequências, o que explica a ocorrência de fenômenos como inconsistências factuais ou “alucinações”, nas quais o sistema gera informações incorretas com elevado grau de fluidez linguística, uma vez que a lógica operacional dos modelos generativos, priorizam a probabilidade estatística em detrimento da verificação da veracidade do conteúdo produzido.

Por sua vez, os *chatbots* consistem em sistemas computacionais desenvolvidos para interagir com usuários por meio de linguagem natural, simulando diálogos humanos em ambientes digitais. Embora frequentemente associados às tecnologias mais recentes de IAGen, os *chatbots* operavam a partir de regras pré-programadas e correspondência de padrões. Com o avanço das técnicas de IA e, posteriormente, da IAGen, passaram a incorporar modelos de aprendizado de máquina e, mais recentemente, grandes modelos de linguagem (Adamopoulou e Moussiades, 2020).

Conforme Adamopoulou e Moussiades (2020), historicamente, os primeiros *chatbots* baseavam-se em regras predefinidas e padrões sintáticos simples, operando por meio de correspondência de palavras-chave (*pattern matching*). Esses sistemas, embora limitados, demonstraram a viabilidade da interação homem-máquina via linguagem natural.

Progressivamente, os *chatbots* evoluíram com a incorporação de técnicas de Processamento de Linguagem Natural (PLN) e, posteriormente, de métodos baseados em aprendizado de máquina.

Dessa forma, os *chatbots* atuais podem ser classificados em duas categorias: (I) *chatbots* baseados em regras e (II) *chatbots* baseados em IA. Os primeiros operam a partir de fluxos conversacionais estruturados, com respostas previamente programadas. Já os segundos utilizam técnicas de aprendizado de máquina e modelos estatísticos para interpretar intenções, extrair entidades e gerar respostas mais flexíveis e adaptativas (Adamopoulou e Moussiades, 2020).

Assim, enquanto a Inteligência Artificial constitui o campo amplo que estuda os sistemas capazes de agir ou pensar de maneira inteligente; e a Inteligência Artificial Generativa refere-se ao subcampo voltado à produção de novos conteúdos (textos, imagens, áudios, etc.); os *chatbots* configuram uma aplicação tecnológica específica, caracterizada pela interface conversacional. Em outras palavras, *chatbot* não é sinônimo de Inteligência Artificial Generativa, mas pode ser implementado a partir dela. Contudo, no presente artigo, o termo *chatbot*, será adotado como sinônimo de IAGen ao se referir as ferramentas aqui analisadas.

METODOLOGIA

A pesquisa é de natureza qualitativa do tipo exploratória (Gil, 2010), uma vez que o objetivo é o de analisar a capacidade de quatro *chatbots* de IAGen na resolução de questões do Enem que envolvem conhecimentos e conteúdos da disciplina de química de forma a considerar seus usos e aplicações no contexto educacional como ferramenta auxiliar no processo de ensino e aprendizagem a partir do entendimento de suas limitações quando utilizado como ferramenta para resolução de exercícios.

Durante o mês de março de 2025, foram analisadas 89 questões do caderno amarelo do



segundo dia de aplicação de provas do Enem das edições dos anos de 2019, 2020, 2021, 2022, 2023 e 2024 da prova de Ciências da Natureza. Foram consideradas questões que envolvessem conhecimentos e conteúdos relativos à disciplina de química para sua resolução. Isso porque, a prova do Enem costuma apresentar questões interdisciplinares que envolvem conhecimentos de diferentes disciplinas. No caso da prova de ciências da natureza, que engloba a disciplina de química, as questões podem exigir dos estudantes conhecimentos das três disciplinas da área: biologia, física e química.

O Quadro 1 apresenta a edição e a quantidade de questões analisadas de cada prova do respectivo ano/edição:

Quadro 1 - Ano/Edição do Exame Nacional do Ensino Médio – Enem e quantidade de questões analisadas.

Ano	Quantidade de questões
2019	13
2020	17
2021	16
2022	15
2023	15
2024	13

Fonte: Os autores.

Após a seleção das questões a serem submetidas a resolução por meio dos *chatbots* de IAGen foram feitas capturas de telas (*PrtsSc*) das questões e as imagens enviadas para resolução por meio de anexos. Junto ao anexo das imagens contendo as questões foi enviado o seguinte “prompt”: “A imagem contém a questão de número xxx do caderno amarelo do segundo dia de provas do Enem xxxx. Proponha uma resolução para a questão e indique a alternativa correta.”

A captura de tela de cada questão junto com o prompt foi enviado aos seguintes chatbots de IA generativa: *ChatGPT*, *Gemini*, *CoPilot* e *DeepSeek*. As versões de cada *chatbot* utilizada foram as versões gratuitas disponíveis por meio

de busca na web, não sendo necessário a criação de uma conta para sua utilização. A escolha por essas versões se deu em virtude de tentar simular uma possível utilização dessas ferramentas realizada por um aluno que utilizaria, por exemplo, a IA generativa para a resolução de uma lista de exercícios de química.

As respostas obtidas foram comparadas aos gabaritos oficiais divulgados pelo INEP. As questões cuja resposta final (alternativa indicada) divergiram do gabarito oficial foram então selecionadas para análise a fim de observar o desempenho do chatbot na resolução e identificar possíveis limitações dos mesmos.

RESULTADOS E DISCUSSÕES

Desempenho e Limitações do *ChatGPT* na Resolução de Questões do Enem

O Quadro 2 apresenta a quantidade, as questões respondidas incorretamente e os conteúdos das questões por ano/edição da prova do Enem de acordo com o desempenho do *ChatGPT*:

Quadro 2 - Quantidade de questões respondidas incorretamente pelo *ChatGPT* por ano/edição do Enem.

Ano	Nº questão	Conteúdos abordados
2019	93, 98, 123 e 133	Equilíbrios ácido-base, reações de neutralização, estequiometria, fundamentos da teoria atômica, química orgânica, reações químicas
2020	113*, 115 e 116	Cálculo de densidade, química ambiental (OD e DBO), estequiometria
2021	94 e 103	Química orgânica (polímeros), química



		orgânica (isomeria)
2022	106	Estequiometria, cinética química
2023	97 e 129	Gases ideais, estequiometria
2024	96, 122 e 123	Química orgânica (grupos funcionais), química ambiental (ciclos biogeoquímicos), estequiometria

Fonte: Os autores.

O *ChatGPT* errou 15 das 89 questões submetidas a resolução, sendo a edição de 2019 a com maior incidência de erros (4 questões) e a de 2022 a com menor (1 questão).

A análise das questões respondidas incorretamente demonstra uma concentração nas questões envolvendo conteúdos que abordam cálculos estequiométricos e conceitos de química orgânica. A dificuldade em estequiometria pode estar associada à complexidade das múltiplas etapas lógicas e à necessidade de interpretação precisa de unidades e proporções. Nesse sentido, Leite (2023), destaca limitações do *ChatGPT* em relação a capacidade de resolução de cálculos químicos e balanceamento de equações apontando erros cometidos pelo modelo na resolução desse tipo de exercício. Uçar, Lopez-Gazpio e Lopez-Gazpio (2025) também apontam limitações significativas na capacidade de raciocínio das ferramentas de IA generativa em análises envolvendo cálculos de estequiometria.

Em química orgânica (isomeria, polímeros e grupos funcionais), os erros sugerem uma limitação do modelo, baseado em linguagem, em lidar com a visualização de estruturas moleculares e o reconhecimento de padrões sutis como verificado nas questões 94/2021 e 103/2021.

Figura 1 - Questão 94 do Enem 2021.

Questão 94 - enem2021

O Prêmio Nobel de Química de 2000 deveu-se à descoberta e ao desenvolvimento de polímeros condutores. Esses materiais têm ampla aplicação em novos dispositivos eletroluminescentes (LEDs), células fotovoltaicas etc. Uma propriedade-chave de um polímero condutor é a presença de ligações duplas conjugadas ao longo da cadeia principal do polímero.

ROCHA FILHO, R. C. Polímeros condutores: descoberta e aplicações. Química Nova na Escola, n. 12, 2000 (adaptado).

Um exemplo desse polímero é representado pela estrutura

Fonte: Inep

Na questão, é necessário escolher a alternativa que representa a estrutura correta da cadeia de um polímero. Como as alternativas trazem informações visuais como resposta o *ChatGPT* indica uma alternativa incorreta (C) no lugar da alternativa correta (D) mesmo realizando uma explicação conceitual aparentemente coerente como justificativa de resolução.

Essa dificuldade se estende à interpretação de imagens, gráficos e tabelas, uma vez que questões com menor volume de texto foram frequentemente respondidas de forma incorreta: Q. 115/2020, Q. 129/2023 e Q. 123/2024. As questões 115/2020 e 123/2024, trazem gráficos com informações que compõem a análise para a resolução da questão de modo que a alternativa correta possa ser encontrada. Já a questão 129/2023, fornece informações no enunciado por meio da equação de reação do cloreto de mercúrio e da massa de cloreto de sódio em uma tabela, devendo o aluno identificar qual alternativa contém o gráfico correto de acordo com o cálculo estequiométrico realizado para chegar a resposta correta.



Figura 2 - Questão 129 do Enem 2023.

QUESTÃO 129

Um assistente de laboratório precisou descartar sete frascos contendo solução de nitrato de mercúrio(II) que não foram utilizados em uma aula prática. Cada frasco continha 5,25 g de $Hg_2(NO_3)_2$ dissolvidos em água. Tendo a toxicidade do mercúrio e sabendo que o Hg_2Cl_2 tem solubilidade muito baixa, o assistente optou por retirar o mercúrio da solução por precipitação com cloreto de sódio (NaCl), conforme a equação química:

$$Hg_2(NO_3)_2(aq) + 2 NaCl(aq) \longrightarrow Hg_2Cl_2(s) + 2 NaNO_3(aq)$$

Na dúvida sobre a massa de NaCl a ser utilizada, o assistente aumentou gradativamente a quantidade adicionada em cada frasco, como apresentado no quadro.

Frasco	I	II	III	IV	V	VI	VII
Massa de NaCl em grama (g)	0,2	0,4	0,6	0,8	1,0	1,2	1,4

O produto obtido em cada experimento foi filtrado, secado e teve sua massa aferida. O assistente organizou os resultados na forma de um gráfico que correlaciona a massa de NaCl adicionada com a massa de Hg_2Cl_2 obtida em cada frasco. A massa molar do $Hg_2(NO_3)_2$ é 525 g mol^{-1} , a do NaCl é 58 g mol^{-1} e a do Hg_2Cl_2 é 472 g mol^{-1} .

Qual foi o gráfico obtido pelo assistente de laboratório?

Fonte: Inep

Observa-se que a questão 129/2023 envolve a interpretação de informações contidas em tabelas e gráficos e também a realização de cálculos estequiométricos, habilidades já apontadas como limitantes para o *ChatGPT* anteriormente (Leite, 2023; Uçar, Lopez-Gazpio e Lopez-Gazpio, 2025).

Destaca-se ainda a questão 113/2020, onde o *ChatGPT* forneceu uma resolução conceitualmente correta, inclusive indicando a resposta correta mas indicando ao final a alternativa incorreta (E), no lugar da alternativa correta (D), demonstrando assim uma falha na finalização da resolução da questão (Figura X). Nessa perspectiva Santos *et al.* (2023) argumentam que o *ChatGPT* pode apresentar limitações conceituais o que é corroborado por Sigot e Tassotti (2025) que apontam deficiências na precisão e profundidade das explicações envolvendo conceitos da disciplina de Química por parte dos *chatbots* de IAGen.

Figura 3 - Questão 113 do Enem 2020 e sua resolução feita pelo *ChatGPT*.

Questão 113

As moedas despertam o interesse de colecionadores, numismatas e investidores há bastante tempo. Uma moeda de 100% cobre, circulante no período do Brasil Colônia, pode ser bastante valiosa. O elevado valor gera a necessidade de realização de testes que validem a procedência da moeda, bem como a veracidade de sua composição. Sabendo que a densidade do cobre metálico é próxima de 9 g cm^{-3} , um investidor negocia a aquisição de um lote de quatro moedas A, B, C e D fabricadas supostamente de 100% cobre e massas 26 g, 27 g, 10 g e 36 g, respectivamente. Com o objetivo de testar a densidade das moedas, foi realizado um procedimento em que elas foram sequencialmente inseridas em uma proveta contendo 5 mL de água, conforme esquematizado.

Com base nos dados obtidos, o investidor adquiriu as moedas

A A e B.
B A e C.
C B e C.
D B e D.
E C e D.

Para resolver essa questão, devemos verificar quais moedas possuem densidade próxima a 9 g/cm^3 , que é a densidade do cobre metálico.

Passos para a solução:

- Fórmula da densidade:

$$d = \frac{m}{V}$$
 onde:
 - d é a densidade (g/cm^3).
 - m é a massa (g).
 - V é o volume deslocado na proveta (cm^3).
- Analisando os deslocamentos de volume das moedas (dados da imagem):
 - Moeda A: Massa = 26 g, Volume deslocado = 2,8 mL

$$d_A = \frac{26}{2,8} \approx 8,93 \text{ g/cm}^3$$
 - Moeda B: Massa = 27 g, Volume deslocado = 3,0 mL

$$d_B = \frac{27}{3,0} = 9,0 \text{ g/cm}^3$$
 - Moeda C: Massa = 10 g, Volume deslocado = 1,5 mL

$$d_C = \frac{10}{1,5} \approx 6,67 \text{ g/cm}^3 \text{ (muito abaixo de } 9 \text{ g/cm}^3)$$
 - Moeda D: Massa = 36 g, Volume deslocado = 4,0 mL

$$d_D = \frac{36}{4,0} = 9,0 \text{ g/cm}^3$$
- Conclusão:

As moedas que apresentam densidade próxima a 9 g/cm^3 (densidade do cobre) são B e D.

Alternativa correta:
 B e D.
 Resposta: Letra E.

Fonte: Inep/*ChatGPT*



Desempenho e Limitações do Gemini na Resolução de Questões do Enem

O Quadro 3 apresenta a quantidade, as questões respondidas incorretamente e os conteúdos das questões por ano/edição da prova do Enem de acordo com o desempenho do Gemini:

Quadro 3 - Quantidade de questões respondidas incorretamente pelo Gemini por ano/edição do Enem.

Ano	Nº questão	Conteúdos abordados
2019	133	Equilíbrios ácido-base, reações de neutralização, estequiometria
2020	115	Química ambiental (OD e DBO)
2021	94, 103 e 108	Química orgânica (polímeros), química orgânica (isomeria), química orgânica (grupos funcionais), ácidos e bases
2022	-	-
2023	129 e 134	Estequiometria, química orgânica (isomeria)
2024	-	-

Fonte: Os autores.

O Gemini apresentou o menor número de erros (7 de 89 questões), dentre os modelos analisados, com estes distribuídos entre as edições de 2019, 2020, 2021 e 2023.

A maior incidência de erros concentrou-se nos conteúdos de química orgânica, com destaque para as questões envolvendo conceitos de isomeria nas edições de 2021 e 2023 (Figuras 4 e 5). Esse padrão pode reforçar a hipótese de que modelos baseados em linguagem possuem dificuldade com a visualização e distinção de estruturas moleculares, conforme apontado por Girotto Junior, Nascimento Junior e Morais

(2024), onde, os autores, observaram imprecisões e erros conceituais tanto na interpretação quanto na geração de imagens envolvendo conteúdos da disciplina de Química.

Figura 4 - Questão 103 do Enem 2021

QUESTÃO 103

A icterícia, popularmente conhecida por amarelão, é uma patologia frequente em recém-nascidos. Um bebê com icterícia não consegue metabolizar e excretar de forma eficiente a bilirrubina. Com isso, o acúmulo dessa substância deixa-o com a pele amarelada. A fototerapia é um tratamento da icterícia neonatal, que consiste na irradiação de luz no bebê. Na presença de luz, a bilirrubina é convertida no seu isômero lumirrubina que, por ser mais solúvel em água, é excretada pela bile ou pela urina. A imagem ilustra o que ocorre nesse tratamento.

MOREIRA, M. et al. O recém-nascido de alto risco: teoria e prática de cuidar [on-line]. Rio de Janeiro: Focuz, 2004 (adaptado).

Na fototerapia, a luz provoca a conversão da bilirrubina no seu isômero

ótico.
 funcional.
 de cadeia.
 de posição.
 geométrico.

Fonte: Inep

Figura 5 - Questão 134 do Enem 2023

QUESTÃO 134

Entre os medicamentos mais comuns consumidos para o alívio da dor está o ibuprofeno, um composto quiral com ação anti-inflamatória e efeito analgésico, que é comercializado como fármaco opticamente puro, ou seja, sem a mistura com outro isômero óptico. A fórmula estrutural plana do ibuprofeno é:

Além do ibuprofeno, destacam-se também os princípios ativos a seguir, presentes em outros medicamentos para o alívio da dor:

fenacetina.
 paracetamol.
 dipirona sódica.
 diclofenaco sódico.
 butilbrometo de escopolamina.

Fonte: Inep

O Gemini não registrou erros nas provas de 2022 e 2024. De forma geral, o modelo demonstra fragilidade em questões que exigem raciocínio abstrato, interpretação de estruturas químicas, imagens (gráficos, tabelas) e análise interdisciplinar, de forma similar aos demais modelos analisados.

Desempenho e Limitações do CoPilot na Resolução de Questões do Enem

O Quadro 4 apresenta a quantidade, as questões respondidas incorretamente e os conteúdos das questões por ano/edição da prova



do Enem de acordo com o desempenho do *CoPilot*:

Quadro 4 - Quantidade de questões respondidas incorretamente pelo *CoPilot* por ano/edição do Enem.

Ano	Nº questão	Conteúdos abordados
2019	93, 94, 98, 123 e 125,	Fundamentos da teoria atômica, reações químicas, química orgânica (grupos funcionais), estequiometria, termoquímica
2020	113, 115, 116 e 123	Cálculo de densidade, Química ambiental (OD e DBO), estequiometria, química orgânica (polímeros)
2021	103 e 108	Química orgânica (isomeria e grupos funcionais)
2022	-	-
2023	129	Estequiometria
2024	120 e 122	Eletroquímica, reações químicas, química ambiental (ciclos biogeoquímicos)

Fonte: Os autores.

O *CoPilot* apresentou desempenho semelhante ao do *ChatGPT*, com 15 respostas incorretas, com o maior volume de questões incorretas concentradas nas edições de 2019 e 2020 com 5 e 4 erros respectivamente.

No caso do *CoPilot* a maior quantidade de questões respondidas incorretamente concentrou-se nos anos de 2019 e 2020. No ano de 2019, os conteúdos abordados pelas questões respondidas incorretamente são mais diversos, abrangendo conteúdos sobre fundamentos da teoria atômica (Q. 93), reações químicas (Q. 94

e Q. 123), química orgânica (Q.98), estequiometria e termoquímica (Q. 125), o que pode indicar certa dificuldade do modelo em trabalhar com a diversidade teórica de conteúdos da disciplina de química. Essa dificuldade é evidenciada por Sigot e Tassotti (2025) que observaram que os *chatbots* podem, muitas vezes apresentar falhas na precisão e profundidade das explicações que envolvem conceitos químicos.

Assim como os outros modelos, o *CoPilot* demonstrou fragilidade em conteúdos que exigem raciocínio quantitativo (estequiometria) e visualização de estruturas moleculares (química orgânica).

Figura 6 - Questão 116 do Enem 2020

Questão 116

O crescimento da frota de veículos em circulação no mundo tem levado à busca e desenvolvimento de tecnologias que permitam minimizar emissões de poluentes atmosféricos. O uso de veículos elétricos é uma das propostas mais propagandeadas por serem de emissão zero. Podemos comparar a emissão de carbono na forma de CO_2 (massa molar igual a 44 g mol^{-1}) para os dois tipos de carros (a combustão e elétrico). Considere que os veículos tradicionais a combustão, movidos a etanol (massa molar igual a 46 g mol^{-1}), emitem uma média de $2,6 \text{ mol}$ de CO_2 por quilômetro rodado, e os elétricos emitem o equivalente a $0,45 \text{ mol}$ de CO_2 por quilômetro rodado (considerando as emissões na geração e transmissão da eletricidade). A reação de combustão do etanol pode ser representada pela equação química:

$$\text{C}_2\text{H}_5\text{OH} (\text{l}) + 3 \text{O}_2 (\text{g}) \rightarrow 2 \text{CO}_2 (\text{g}) + 3 \text{H}_2\text{O} (\text{g})$$

Foram analisadas as emissões de CO_2 envolvidas em dois veículos, um movido a etanol e outro elétrico, em um mesmo trajeto de $1\,000 \text{ km}$.

CHARADIA, C. A. Estudo da viabilidade da implantação de frotas de veículos elétricos e híbridos elétricos no atual cenário econômico, político, energético e ambiental brasileiro. Guaratinguetá: Unesp, 2015 (adaptado).

A quantidade equivalente de etanol economizada, em quilograma, com o uso do veículo elétrico nesse trajeto, é mais próxima de

A. 50.
 B. 60.
 C. 95.
 D. 99.
 E. 120.

Fonte: Inep

Os erros observados na prova de 2020 se assemelham aos conteúdos dos erros observados na prova do ano de 2019 (estequiometria e química orgânica). Destaca-se ainda, a edição de 2022 não teve questões respondidas incorretamente pelo *CoPilot*.

Uma particularidade observada no *CoPilot* foi a dificuldade na representação visual de fórmulas. Foram observados problemas na



formatação de fórmulas químicas e expressões de cálculo nas resoluções fornecidas pelo modelo (como nas Questões 104/2020, 113/2021), com símbolos mal posicionados, o que compromete a clareza e o entendimento lógico da resolução (Figuras 7 e 8). Conforme Costa Junior (2025) é possível perceber a ocorrência de informações confusas e erros na resolução de problemas realizados por *chatbots*, o que por sua vez demanda do usuário revisão e questionamento a respeito do conteúdo fornecido pela ferramenta. A má representação de fórmulas em química é especialmente problemática, pois a disposição espacial e a clareza na notação são essenciais para a compreensão conceitual.

Figura 7 – Resolução da questão 104/2020.

A questão menciona as propriedades do dióxido de carbono e seus impactos ambientais, com foco nos organismos aquáticos. O dióxido de carbono (CO_2) é solúvel em água, e seu aumento na atmosfera afeta o equilíbrio ambiental ao influenciar o pH da água.

Quando CO_2 dissolve-se em água, forma ácido carbônico (H_2CO_3), que se dissocia parcialmente, liberando íons hidrogênio (H^+). Esse processo reduz o pH da água, tornando-a mais ácida, o que é prejudicial aos organismos aquáticos, especialmente aqueles com estruturas calcárias, como corais e moluscos.

Portanto, a alternativa correta é:

A) redução do potencial hidrogeniônico da água.

Se precisar de mais explicações ou ajuda com outros tópicos, é só chamar! 😊

Fonte: *CoPilot*

Figura 8 – Resolução da questão 113/2021.

Vamos resolver a questão passo a passo:

1. Identificar a proporção de glicose para etanol na reação química: De acordo com a equação: $\text{C}_6\text{H}_{12}\text{O}_6 \rightarrow 2 \text{C}_2\text{H}_5\text{OH} + 2 \text{CO}_2$

A cada 1 mol de glicose ($\text{C}_6\text{H}_{12}\text{O}_6$), são obtidos 2 mols de etanol ($\text{C}_2\text{H}_5\text{OH}$).

2. Converter a quantidade de glicose em mols: Sabemos que a massa molar da glicose é 180 g/mol. Com 50 kg de glicose (50.000 g), temos: $\frac{\text{massa de glicose}}{\text{massa molar da glicose}} = \frac{50.000}{180} \approx 277,78$ mol.

3. Determinar a quantidade de etanol em mols: Como cada mol de glicose produz 2 mols de etanol: $2 \times 277,78 = 555,56$ mol.

4. Calcular a massa de etanol produzida: A massa molar do etanol é 46 g/mol. Logo: $555,56 \times 46 \approx 25.555,56$ g.

5. Considerar a conversão de 80%: A conversão da glicose em etanol é de 80%. Portanto: $25.555,56 \times 0,8 \approx 20.444,45$ g.

6. Determinar o volume de etanol: Sabemos que a densidade do etanol é 0,80 g/mL. Para converter massa em volume: $\frac{\text{Volume de etanol}}{\text{Massa final de etanol}} = \frac{20.444,45}{0,80} \approx 25.555,56$ mL. Convertendo para litros: $\frac{25.555,56}{1000} \approx 25,56$ L.

Resposta final: O volume de etanol obtido é mais próximo de **25 litros**. Portanto, a alternativa correta é: **C) 25**.

Caso precise de mais explicações, estou aqui! 😊

Fonte: *CoPilot*

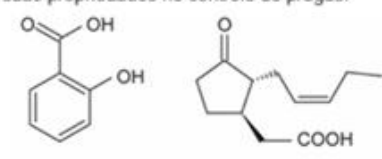
Além disso, o *CoPilot* demonstrou não ser

capaz de analisar a imagem da questão 96/2024, informando que, por ser um assistente de texto, não podia visualizar os compostos que faziam parte da resolução da questão (Figura 9). Essa é uma limitação importante do modelo, já que a prova do Enem costuma apresentar elementos visuais como imagens, gráficos e representações moleculares em suas questões para interpretação e resolução do exercício.

Figura 9 - Questão 96 do Enem 2024 e sua resolução pelo *CoPilot*.

QUESTÃO 96

Os pesticidas naturais vêm sendo utilizados no controle de pragas e doenças agrícolas como substituintes de pesticidas sintéticos tradicionais, por serem menos nocivos ao ambiente, biodegradáveis e minimizarem custos e riscos relativos à lavoura. Por exemplo, os compostos 1 e 2 estão envolvidos nas respostas de defesa das plantas. Os grupos funcionais presentes nesses compostos são importantes para suas propriedades no controle de pragas.



Composto 1 **Composto 2**

PINTO-ZEVALLOS, D. M.; ZARBIN, P. H. G. A química na agricultura: perspectivas para o desenvolvimento de tecnologias sustentáveis. *Química Nova*, n. 10, 2013 (adaptado).

Qual é a função orgânica correspondente ao grupo funcional comum presente nesses dois compostos?

A) Ácido carboxílico.
B) Cetona.
C) Alceno.
D) Álcool.
E) Fenol.

Vamos analisar a questão.

A questão 96 do Enem 2024 fala sobre pesticidas naturais e sua importância no controle de pragas agrícolas. A questão apresenta dois compostos químicos, Composto 1 e Composto 2, envolvidos nas respostas de defesa das plantas. Nossa tarefa é identificar a função orgânica correspondente ao grupo funcional comum presente nesses dois compostos.

Para isso, vamos observar as estruturas dos compostos e identificar o grupo funcional comum. Infelizmente, como sou um assistente de texto, não posso visualizar as imagens dos compostos diretamente. No entanto, se você puder descrever as estruturas dos compostos, posso ajudar na identificação do grupo funcional.

Se você puder fornecer uma descrição das estruturas dos compostos ou os grupos funcionais que aparecem neles, poderei guiá-lo na resolução da questão e indicar a alternativa correta.

Estou aqui para ajudar!

Fonte: *Inep/CoPilot*

Desempenho e Limitações do *DeepSeek* na Resolução de Questões do Enem

O Quadro 5 apresenta a quantidade, as questões respondidas incorretamente e os conteúdos das questões por ano/edição da prova do Enem de acordo com o desempenho do *DeepSeek*:



Quadro 5 - Quantidade de questões respondidas incorretamente pelo *DeepSeek* por ano/edição do Enem.

Ano	Nº questão	Conteúdos abordados
2019	93 e 119	Fundamentos da teoria atômica, ligações químicas, estrutura de Lewis, propriedades periódicas
2020	107, 113, 115* e 119	Química orgânica (grupos funcionais), cálculo de densidade, química ambiental (OD e DBO), propriedades físicas das substâncias, separação de misturas
2021	94, 103 e 108*	Química orgânica (polímeros, isomeria, grupos funcionais)
2022	-	-
2023	129	Estequiometria
2024	96, 108*, 120* e 122	Química orgânica (grupos funcionais), estequiometria, eletroquímica, reações químicas, química ambiental (ciclos biogeoquímicos)

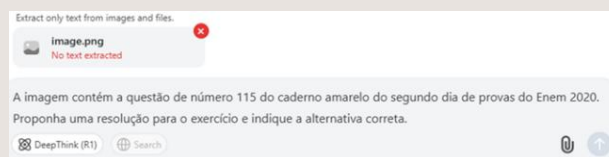
Fonte: Os autores.

O *DeepSeek* demonstrou dificuldade em conteúdos envolvendo fundamentos da teoria atômica e propriedades periódicas, na prova de 2019 e também nos conteúdos já observados nos

modelos anteriores: química orgânica (2020, 2021 e 2024) e estequiometria (2023 e 2024).

Outra limitação observada no *DeepSeek* foi a dificuldade de interpretar informações visuais ou semi-textuais pelo modelo. Em quatro ocasiões (Q. 115/2020, Q. 108/2021, Q. 108/2024 e Q. 120/2024), o modelo não conseguiu responder a questão, apresentando a mensagem de erro "*no text extracted*", conforme a figura abaixo. Essa falha em extrair informações de gráficos, imagens e tabelas pode comprometer a utilidade do modelo como ferramenta de apoio ao estudo na disciplina de Química que possui a análise de informações visuais como crucial para o entendimento de diversos conceitos e conteúdos da disciplina.

Figura 10 - Mensagem de erro do *DeepSeek* na questão 115 do Enem 2020.



Fonte: *DeepSeek*

CONSIDERAÇÕES FINAIS

A partir dos resultados observados no desempenho dos modelos de *chatbots* de IAGen na resolução de exercícios do Enem envolvendo conceitos da disciplina de química, foi possível constatar que, embora os *chatbots* tenham demonstrado capacidade de resolução na maioria das questões, todos apresentam limitações importantes a serem consideradas. As limitações observadas se relacionam principalmente aos conteúdos da disciplina de Química que exigem raciocínio quantitativo, interpretação de imagens, visualização de estruturas químicas e análise interdisciplinar.

Nesse sentido, as maiores fragilidades dos modelos analisados foram observadas nas questões dos conteúdos de estequiometria e de química orgânica, sugerindo que habilidades para abstração espacial e desenvolvimento de cálculos em etapas são particularmente



desafiadoras para os modelos considerados. Também, a dificuldade do *DeepSeek* na análise de informações contidas em gráficos e imagens, e do *CoPilot* na formatação de fórmulas e expressões indicam possíveis barreiras práticas para o uso dessas ferramentas de forma autônoma por estudantes no auxílio de sua aprendizagem.

Apesar do grande potencial das ferramentas de IAGen como ferramenta de apoio ao ensino e aprendizagem, o uso dessas ferramentas no contexto educacional impõe uma reflexão crítica acerca dos seus impactos pedagógicos. A facilidade na obtenção de respostas automatizadas pode levar a uma dependência que impede o desenvolvimento do raciocínio complexo por parte do estudante, especialmente nas áreas em que foram observadas que os *chatbots* apresentam fragilidades.

Dessa forma, o professor representa figura central na integração dessas ferramentas assumindo o papel de agente crítico, orientando os estudantes não apenas sobre as limitações técnicas dos modelos, apenas nos casos dos erros que foram observados, mas, sobretudo, fomentando o senso crítico necessário para avaliar a confiabilidade da informação apresentada pela ferramenta de IAGen.

REFERÊNCIAS

ADAMOPOULOU, E.; MOUSSIADES, L. **An Overview of Chatbot Technology**. In: IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), 16th. 2020, Neos Marmaras, Greece. p. 373-383.

ARAYA, A. M. O.; GIBIN, G. B.; SOUZA FILHO, M. P. Uma introdução às tecnologias digitais de Informação e Comunicação no ensino de Ciências. In: ARAYA, A. M. O.; GIBIN, G. B.; SOUZA FILHO, M. P. (orgs.). **O ensino de Ciências e as tecnologias digitais da Informação e Comunicação (TDIC): pesquisas desenvolvidas na educação básica**.

São Paulo: Editora Unesp, 2021. Disponível em: <https://books.scielo.org/id/cwcpz>. Acesso em: 13 jan. 2024.

BORBA, M. C.; BALBINO JUNIOR, V. R. O ChatGPT e Educação matemática. **Educ. Matem. Pesq.**, São Paulo, v.25, n.3, p.142-156, 2023.

COPPIN, B. **Inteligência Artificial**. 1st ed. Rio de Janeiro: LTC, 2013.

COSTA JUNIOR, I. L. C. Análise do ChatGPT como ferramenta no ensino de química: explorando o tema matéria e energia. **ACTIO Docência em Ciências**, v. 10, n. 2, p. 1-23, 2025.

COSTA JÚNIOR, J. F.; LIMA, U. F.; LEME, M. D.; MORAES, L. S.; COSTA, J. B.; BARROS, D. M.; SOUSA, M. A. M. A.; OLIVEIRA, L. C. F. A inteligência artificial como ferramenta de apoio no ensino superior. **REBENA Revista Brasileira de Ensino e Aprendizagem**, v. 6, p. 246-269, 2023.

COSTA, R. G. D.; PASSERINO, L. M.; ZARO, M. A. Fundamentos teóricos do processo de formação de conceitos e suas implicações para o ensino e aprendizagem de química. **Ensaio Pesquisa em Educação em Ciências**, Belo Horizonte, v. 14, n. 1, p. 271–281, 2012.

FEUERRIEGEL, S.; HARTMANN, J.; JANIESCH, C.; ZSCHECH, P. Generative AI. **Business & Information Systems Engineering**, v. 66, n. 1, p. 111–126, 2023.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas São Paulo, 2010.

GIROTTI JUNIOR, G.; NASCIMENTO JUNIOR, W. J. D.; MORAIS, C. Avaliação de tecnologias de inteligência artificial generativa



na produção de elementos textuais e imagéticos sobre conceitos de química. **CHALLENGES 2024**: Espaços e Caminhos OnLife, p. 299-311, 2024.

JACKSON, JR., P. C. **Introduction to Artificial Intelligence**. 2nd, enlarged. ed. New York: Dover Publications, Inc., 1985. ISBN 0-486-24864-X.

LEITE, B. S. Inteligência artificial e ensino de química: uma análise propedêutica do ChatGPT na definição de conceitos químicos. **Quim. Nova**, v. 46, n. 9, 915-923, 2023.

MCCARTHY, John. What is artificial intelligence? Stanford: Stanford University, 1995. Disponível em: <https://www-formal.stanford.edu/jmc/whatisai.pdf>. Acesso em: 14 jan. 2024.

MONTEIRO, F. F.; SOUZA, P. V. S.; SILVA, M. C.; MAIA, J. R.; SILVA, W. F.; GIRARDI, D.; ChatGPT in Brazilian K-12 science education. **Frontiers in Education**, v. 9, n. 1321547, p. 01-08, 2024.

OPENAI. Introducing ChatGPT. Disponível em: <https://openai.com/blog/chatgpt>. Acesso em: 14 jan 2024.

RUDOLPH, J.; TAN, S.; TAN S. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. **Journal of Applied Learning & Teaching**, v.6, n.1, 2023.

RUSSEL, S. J.; NORVIG, P. **Inteligência Artificial**. Tradução de: Artificial Intelligence, 3rd ed. Rio de Janeiro: Elsevier, 2013. 1324 p. ISBN 978-85-352-3701-6. E-book (1324 p.).

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 4th ed. New Jersey: Pearson, 2021. ISBN 978-0134610993.

SANTOS, C. M. DOS et al. O processo de

transposição didática pelo ChatGPT acerca dos modelos atômicos: inteligência artificial como professor do Ensino Médio. **Anais... XII EPPEQ - Encontro Paulista de Pesquisa em Ensino de Química**, v. 1, n. 12, 29 dez. 2023.

SIGOT, M.; TASSOTI, S. A Missed Opportunity for No-Code Chatbots? Current Challenges in Publicly Available Chemistry GPTs. **Journal of Chemical Education**, v. 102, n. 5, p. 2151–2159, 2025.

UÇAR, S-Ş.; LOPEZ-GAZPIO, I.; LOPEZ-GAZPIO, J. Evaluating and challenging the reasoning capabilities of generative artificial intelligence for technology-assisted chemistry education. **Education and Information Technologies**, v. 30, n. 8, p. 11463–11482, 2025.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. **Attention is all you need**. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NIPS), 31. 2017, Long Beach, CA. Proceedings [...]. Long Beach, CA: Curran Associates, Inc., 2017. p. 5998–6008.

