



() CTS () CA () EAM () ENF () EAP () EX () FP () HFS () IDD () LEQ () MD () PEQ (X) TIC

Aplicação da análise do discurso no estudo da confiabilidade da implementação de IAs no ensino de química

Gabriel Souza Setúbal *(IC)
UESC, gssetubal.lqu@uesc.br

Eric Teles Souza (IC)
UESC, etsouza.bqi@uesc.br

Milena Duarte Lima(PQ)
UESC, mdlima@uesc.br

Resumo

O avanço das tecnologias cognitivas no ambiente educacional tem impulsionado o uso de assistentes virtuais baseados em inteligência artificial (IA), provocando mudanças significativas na forma como o conhecimento é acessado, processado e validado. Este trabalho investigou a eficácia de cinco plataformas amplamente utilizadas ChatGPT, Gemini, Aria, Copilot e Deepseek na resolução de trinta questões de química do ensino médio, com o objetivo de avaliar seus limites e potencial pedagógico. As questões foram selecionadas a partir de temas sorteados da matriz curricular oficial, abrangendo conteúdos como ligações químicas, sistemas de soluções e separação de misturas. Cada IA foi submetida às mesmas questões em dois formatos: entrada por digitação direta e por arquivo PDF. Os resultados foram analisados quanto à correção conceitual, clareza das explicações e coerência lógica das respostas. As análises revelaram um alto índice de erros conceituais, omissões e incoerências, especialmente quando as IAs lidaram com cálculos ou elementos simbólicos típicos da linguagem química. Observou-se também que o desempenho das plataformas variou consideravelmente entre os formatos de entrada, sendo o PDF mais propenso a falhas interpretativas. Conclui-se que, embora apresentem potencial como ferramentas auxiliares, essas tecnologias ainda não são capazes de substituir a mediação docente, exigindo uso criterioso e supervisionado. O estudo propõe que futuras pesquisas explorem a integração crítica de IAs em outras disciplinas e contextos pedagógicos.

Palavras-chave: Inteligência artificial. Ensino de química. Confiabilidade.

Introdução

A presença de tecnologias cognitivas no ambiente educacional tem se intensificado de maneira exponencial. Elas provocam profundas transformações na forma como o conhecimento é produzido, transmitido e apropriado. Entre essas inovações, destacam-se os sistemas baseados em Inteligência Artificial (IA), que vêm ocupando um espaço crescente nas práticas pedagógicas. Mecanismos computacionais também referidos como algoritmos inteligentes, plataformas automatizadas ou soluções digitais autônomas. Eles são capazes de executar tarefas que antes exigiam raciocínio humano, como: resolver problemas, interpretar dados e gerar textos. Em particular, o uso dessas ferramentas no contexto escolar suscita debates sobre sua eficácia, limites éticos e impactos formativos. O avanço dessas tecnologias no campo educacional é analisado por diversos autores. Fernandes (2023), por exemplo, aponta a popularização de sistemas como ChatGPT, DALL-E e ChatPDF, que já são amplamente utilizados por docentes e estudantes nas atividades acadêmicas. Apesar de reconhecer seu potencial para dinamizar o ensino, o autor adverte para os riscos associados ao uso indiscriminado dessas plataformas. A reprodução de informações imprecisas, a banalização do plágio e a dependência tecnológica são alguns dos elementos citados. Tais preocupações são reiteradas por Santos e Silva (2025), em pesquisa com professores da educação básica, que destacam tanto as possibilidades quanto os entraves da utilização dessas tecnologias na sala de aula.

No mesmo sentido, a revisão sistemática conduzida por Silva et al. (2023) aponta que o uso de programas autônomos de ensino-aprendizagem pode favorecer a personalização do conteúdo e a adaptação ao ritmo de cada estudante. Contudo, o estudo também ressalta que tais benefícios só se concretizam plenamente quando há integração entre a mediação humana e o suporte técnico. A ausência de supervisão docente, por sua vez, pode comprometer a qualidade do processo formativo. Isso torna o estudante um consumidor passivo de respostas automatizadas.

Essa discussão é ampliada por Azambuja e Silva (2024), ressalta a necessidade de refletir sobre o papel da escola e da universidade em um mundo saturado de tecnologias inteligentes. Para os autores, a presença desses sistemas computacionais na educação desafia não apenas métodos pedagógicos, mas também valores formativos. Além de exigir a construção de sujeitos éticos, criativos e capazes de interpretar criticamente os recursos digitais. Ou seja, não basta incorporar softwares de processamento linguístico ou plataformas de geração de conteúdo. É preciso formar cidadãos capazes de compreender seus funcionamentos, limitações e implicações. Ao tratar especificamente do ensino de Química, autores como Leite (2023) e Rabelo e Fuini (2022) observam que as ferramentas digitais inteligentes podem contribuir significativamente para o enfrentamento de dificuldades

históricas da disciplina, como a abstração conceitual e a desmotivação dos alunos. Simuladores, assistentes virtuais e recursos de visualização tridimensional são capazes de tornar mais acessíveis conceitos como reações químicas, estequiometria e estrutura atômica. No entanto, ambos alertam que, sem um projeto pedagógico consciente, essas tecnologias podem se limitar à reprodução de conteúdos descontextualizados e tecnicistas. Com isso, afasta o estudante do pensamento científico autêntico. Ademais, há um ponto central que atravessa todos esses estudos: a necessidade de compreender os recursos inteligentes como instrumentos auxiliares e não substitutivos do processo educativo. A depender de como são utilizados, esses sistemas podem tanto estimular a autonomia discente quanto promover uma dependência nociva. Como enfatizam os estudos analisados, cabe ao professor o papel de orientar, problematizar e conduzir criticamente o uso dessas soluções computacionais no ambiente de aprendizagem.

Nesse contexto, o presente trabalho buscou ir além das discussões teóricas ao investigar, na prática, a eficácia de algumas plataformas baseadas em inteligência computacional, como ChatGPT, Copilot, Bard (descontinuado pelo google e substituído pelo Gemini) e similares, no que diz respeito à resolução de questões de Química obtidas com base em parâmetros curriculares do ensino médio. Por meio da formulação de perguntas objetivas sobre temas específicos da disciplina, foram analisados aspectos como correção conceitual, clareza das explicações, pertinência pedagógica e adequação ao nível de ensino. A proposta foi avaliar criticamente até que ponto essas soluções digitais podem, de fato, apoiar o processo educativo, e onde começam a falhar seja por limitações técnicas, imprecisões, ou pela incapacidade de considerar o contexto pedagógico em que operam.

Portanto, o foco desta investigação é verificar empiricamente os alcances e os limites de diferentes assistentes virtuais no ensino de Química, fornecendo subsídios para que educadores, gestores e pesquisadores possam tomar decisões mais fundamentadas sobre a integração de tecnologias inteligentes no cotidiano escolar.

Metodologia

Esta pesquisa foi conduzida por meio de uma abordagem quantitativa, de cunho básico, com delineamento documental e descritivo. Tendo como referência a metodologia da análise estatística descritiva. Essa metodologia se propõe a ir além da leitura superficial dos dados, permitindo compreender os significados construídos a partir da linguagem utilizada, por meio da identificação de padrões, relações e sentidos emergentes no discurso analisado. A escolha dessa abordagem fundamenta-se na necessidade de explorar, de forma aprofundada, os aspectos que podem influenciar a confiabilidade atribuída às inteligências artificiais (IAs)

como suporte na resolução de exercícios de química, compreendendo as possíveis barreiras à sua implementação no contexto do ensino escolar.

Assim, optou-se por desenvolver um percurso metodológico próprio, ancorado na análise estatística descritiva, para examinar o desempenho das IAs. Esse percurso é organizado em três momentos articulados. (I) a classificação geral das respostas como “Certa” ou “Errada”; (II) a análise detalhada das respostas divergentes do gabarito para identificar os fatores que levaram ao erro; (III) tipificação desses erros em “Erro Fundamental” e “Erro de Marcação”. Essa estrutura permitiu quantificar, comparar e interpretar os padrões de acerto e erro. Assim, transformando dados brutos em informações significativas. Com isso, a metodologia adotada não se limita a reproduzir técnicas existentes, construindo um caminho analítico próprio, adequado às especificidades do estudo das respostas das inteligências artificiais.

O estudo concentrou-se em materiais gerados por cinco ferramentas de IA: chat gpt-4o, deepseek, copilot (microsoft AI), aria (opera aAI) e Gemini (google AI, anteriormente conhecido como bard). Essas ferramentas foram selecionadas com base em dois critérios principais: a facilidade de acesso gratuito ou aberto e a ampla divulgação de suas funcionalidades no espaço digital. Foi levado em consideração sua representatividade no cotidiano dos usuários e o potencial de uso por estudantes do ensino médio. A escolha buscou garantir que o corpus da pesquisa refletisse a experiência real de estudantes que utilizam essas tecnologias em seus estudos.

Para a coleta do material de análise, o corpus do estudo, elaborou-se uma coletânea de questões contendo trinta exercícios de química de nível médio. As questões presentes nesta obra foram obtidas em provas de vestibular, que reúnem exercícios aplicados nos principais vestibulares do Brasil e no ENEM. A seleção foi feita a partir de provas aplicadas entre os anos de 2010 e 2024. O processo de construção do questionário envolveu cinco etapas: levantamento dos principais conteúdos presentes na grade curricular do ensino médio; enumeração desses conteúdos em ordem alfabética; sorteio eletrônico de três tópicos; filtragem de questões correspondentes a esses tópicos nas provas; sorteio final das trinta questões que compuseram o questionário utilizado. A seleção foi realizada por meio da plataforma digital “sorteio.com”, assegurando aleatoriedade no processo de escolha.

Cada IA foi submetida duas vezes ao mesmo conjunto de questões. Entre as duas rodadas, os históricos de uso foram apagados, com o intuito de evitar qualquer influência das respostas anteriores sobre as respostas subsequentes. Na primeira, as questões foram inseridas uma a uma por meio de digitação direta no campo de entrada da ferramenta. Para os itens que exigiam representações visuais, foram adicionadas as respectivas imagens. Já na segunda, foi

enviado um arquivo em formato PDF contendo a lista completa de questões. Com formatação semelhante à encontrada em apostilas, provas escolares ou materiais de estudo digitalizados. As mesmas foram orientadas a apresentar o raciocínio adotado em cada questão, de modo que fosse possível acompanhar o caminho lógico utilizado para justificar cada resposta. Procedimento adotado em ambas as formas de entrada, visando ampliar a compreensão da forma como as ferramentas processam e constroem suas respostas.

A análise do corpus foi organizada em três etapas:

I. Classificação geral:

O gabarito gerado pela aplicação teve suas escolhas classificadas como “Certa” ou “Errada”, com base no gabarito fornecido pelos bancos de dados (gabarito base).

II. Avaliação da natureza do grupo errada:

As questões com respostas divergentes do gabarito base foram revisadas individualmente por meio de seu desenvolvimento lógico. Isto, para a determinação dos fatores que possivelmente desviaram o raciocínio do caminho da alternativa correta. Viabilizando a identificação de outras categorias analíticas;

III. Classificação como “Erro fundamental” ou “Erro de marcação”:

Concluído o segundo passo, identificou-se duas categorias. “Erro fundamental” (EF): referente à falha lógica derivada da criação de falsas verdades a respeito da teoria por trás do tema. Além da desconsideração de uma ou mais etapas de desenvolvimento do raciocínio gerado pela IA, essencial para a conclusão lógica da resolução dos exercícios. “Erro de marcação” (EM): se trata do cenário em que a IA desenvolve a lógica e encontra a resposta corretamente, mas retorna a alternativa errada. O que, então, define a classe EM como um subgrupo da classe EF, pois, no retorno, desconsidera toda a lógica construída.

Essa metodologia foi estruturada com o objetivo de oferecer uma compreensão crítica sobre o funcionamento das IAs no campo educacional. Enfatizando as especificidades do ensino de química, que envolve raciocínios quantitativos, compreensão de linguagem simbólica e domínio de conceitos científicos. Ao focar nas formas de entrada textual e em PDF, buscou-se identificar possíveis limitações operacionais e pedagógicas, além de levantar subsídios para discussões futuras sobre o uso dessas ferramentas em ambientes escolares.

Resultados e Discussão

Seguindo a proposta metodológica, se enumeram os tópicos indicados pela grade curricular de química. Em sequência procedeu o sorteio dos conteúdos que seriam abordados

pelas inteligências artificiais (IAs) na resolução de questões. Resultando na seleção dos tópicos: ligações químicas, sistemas de solução e separação de misturas. Elementos recorrentes em avaliações externas e frequentemente abordados no ensino médio. Com base nesses eixos temáticos, trinta questões foram selecionadas de vestibulares e ENEMs entre os anos de 2010 á 2024, revisadas e organizadas em formato PDF para aplicação. Buscando, assim, simular a forma como professores e alunos normalmente lidam com questões de livros, apostilas ou provas digitalizadas.

A Tabela 01 resume o desempenho responsivo das IAs (GPT, Gemini, ARIA, Copilot e Deepseek) quando submetidas às questões via digitação direta e via leitura integral do arquivo PDF. A interpretação simbólica representativa utilizada (Tabela 02) foi fundamental para classificar os resultados. Formas como “●” representam o acerto, “▲” erro fundamental (EF), “△” erro de marcação (EM) e “-” ausência de resposta. A aplicação sistemática dessa simbologia viabilizou uma leitura comparativa, permitindo distinguir tanto o grau de eficiência quanto a natureza dos equívocos cometidos por cada modelo.

Tabela 01- *Classificação do gabarito desenvolvido por cada IA em situação de digitação, questão a questão, e de importação integral da lista em formato de PDF*

	Entrada por digitação					Entrada por PDF				
	GPT	Gemini	ARIA	Copilot	Deepseek	GPT	Gemini	ARIA	Copilot	Deepseek
1	●	●	●	●	●	●	●	●	●	●
2	●	●	▲	●	●	●	△	▲	▲	△
3	●	▲	●	●	●	●	-	●	▲	●
4	●	●	●	●	●	●	●	●	●	●
5	●	●	▲	●	△	●	△	▲	●	△
6	●	●	△	△	●	●	●	▲	●	△
7	●	●	●	●	●	●	●	●	●	●
8	●	●	●	●	●	△	△	●	▲	△
9	●	●	●	●	●	●	-	●	●	●
10	●	●	●	●	●	●	△	▲	●	●
11	●	●	●	●	●	●	●	△	●	●
12	●	●	●	●	●	●	●	△	●	●
13	●	●	●	●	●	●	△	▲	●	●
14	●	●	●	●	●	●	-	▲	▲	●
15	●	▲	▲	●	●	●	▲	▲	▲	●
16	●	●	●	●	●	●	●	△	●	●
17	●	△	▲	●	△	●	△	▲	●	△

18	●	●	●	●	●	▲	●	▲	●	●
19	●	●	▲	●	△	△	-	-	▲	△
20	●	●	●	●	●	△	-	●	●	●
21	●	△	●	●	●	●	-	△	●	●
22	●	●	●	●	●	●	-	△	●	●
23	●	●	●	●	●	●	-	△	●	●
24	●	●	▲	●	△	●	-	△	●	●
25	●	●	●	●	△	●	-	▲	●	●
26	●	▲	●	●	●	●	-	△	●	△
27	●	●	●	●	●	●	-	●	●	●
28	●	●	●	●	●	●	-	△	●	●
29	●	●	●	●	●	●	-	▲	●	●
30	●	●	▲	●	△	▲	-	△	△	●

Fonte: Autoral (2025).

Tabela 02 - *legenda para a Tabela 02*

Classe	Símbolo
Certa	●
Erro Fundamental	▲
Erro de Marcação	△
Sem Resposta	-

Fonte: Autoral (2025).

A primeira e mais evidente limitação foi a dificuldade de compreensão contextual por parte das IAs. Apesar de demonstrarem familiaridade com termos técnicos, muitos modelos apresentaram respostas fora do escopo conceitual esperado. Essa tendência revela um problema estrutural nos modelos de linguagem: eles operam com base em regularidades estatísticas da linguagem (CURY et al., 2022), e não necessariamente por meio de inferência lógica ou interpretação pedagógica. Em outras palavras, eles reconhecem palavras, mas não compreendem a intenção pedagógica por trás da pergunta nem os dados necessários para resolvê-la de forma correta.

A gravidade dessa limitação aumenta quando se trata de questões que envolvem cálculos. A química, enquanto ciência natural, requer raciocínios encadeados que transitam entre

conceitos teóricos e cálculos quantitativos. Itens que abordam pH, concentração molar, energia de ligação ou rendimento reacional exigem domínio de etapas sucessivas, que vão desde a leitura correta dos dados até a aplicação adequada de equações. Fatores que, com uma leitura equivocada, desenvolvem respostas falaciosas. Fator contribuinte das preocupações de Santos e Silva (2025). A Tabela 03 deixa claro que, salvo o GPT (com 0% de erros na digitação), os demais modelos apresentaram sérios problemas. O Gemini, por exemplo, cometeu 10% de erros fundamentais e 6,67% de marcação. ARIA, por sua vez, obteve 23,67% de erros fundamentais, revelando sua fragilidade mesmo com entrada textual direta. O Deepseek registrou 20% de erros fundamentais, enquanto o Copilot, com apenas 3% de erro de marcação. Mostrando, dessa forma, desempenho relativamente equilibrado.

Ao analisar o comportamento das IAs frente à leitura de PDFs, os resultados se tornam ainda mais preocupantes. O GPT, que antes apresentava excelência com 100% de acerto na digitação, passa a errar 16,67% das questões no PDF sendo 10% de EM e 6% de EF. Isso indica uma perda de desempenho ligada à forma de entrada. O modelo passa a apresentar respostas sem o detalhamento adequado, omitindo o processo de resolução e limitando-se a enunciar a conclusão, muitas vezes sem fundamentação. Essa mudança compromete a aprendizagem, já que o raciocínio lógico se perde.

Tabela 03- Relação em porcentagem de das respostas divergentes do esperado, em relação à lista aplicada, onde Σ é a porcentagem total de divergência em uma aplicação.

	Entrada por digitação					Entrada por PDF				
	GPT	Gemini	ARIA	Copilot	Deepseek	GPT	Gemini	ARIA	Copilot	Deepseek
▲	0	10,00	23,67	0,00	20,00	6,00	3,00	36,67	20,00	0,00
△	0	6,67	3,00	3,00	0,00	10,67	20,33	33,33	3,33	23,33
-	0	0,00	0,00	0,00	0,00	0,00	50,00	%	0,00	0,00
Σ	0	16,67	26,67	3,00	20,00	16,67	73,33	73,33	23,33	23,33

Fonte: Autoral (2025).

O desempenho do Gemini em PDF é ainda mais alarmante. O modelo deixou 50% das questões sem resposta e cometeu 20% de EM e 3% de EF. Isso resulta em 73,33% de resultados inadequados. A pior performance registrada dentre todos os modelos avaliados. A

leitura do arquivo PDF compromete significativamente a capacidade do modelo de identificar corretamente os dados e resolver as questões com coerência. A ARIA também apresentou uma forte queda. Ainda que tenha respondido a quase todas as questões (apenas 3% ficaram sem resposta), seus erros aumentaram drasticamente, alcançando 36,67% de EF e 33,33% de EM. Uma possível explicação é a necessidade de capturas de tela no lugar do PDF original, visto que a plataforma durante o decorrer do estudo não aceita importação direta de arquivos. A leitura de imagem, mesmo com softwares de reconhecimento ótico de caracteres (ORCs), adiciona uma camada de ambiguidade que compromete a precisão das respostas. O Copilot, que havia tido bom desempenho por digitação (3,33% de EM), passou a cometer 20% de EF ao ser submetido ao PDF, evidenciando um padrão de perda de acurácia similar ao de outros modelos. Já o Deepseek apresentou uma inversão nos tipos de erro: dos 20% de EF na entrada por digitação, migrou para 23,33% de EM ao ler o PDF. Esse padrão pode indicar um modelo sensível a erros de conversão de unidades ou formatação, o que compromete a interpretação numérica dos enunciados.

Esses achados confirmam o argumento de Rodrigues et al. (2021), segundo o qual os modelos de linguagem atuais não são desenhados para executar raciocínios algébricos complexos, típicos de questões químicas. Sua arquitetura, baseada na predição sequencial de palavras, falham ao lidar com cadeias de raciocínio matemático, ainda mais quando o conteúdo está inserido em estruturas não formatadas ou simbólicas como ocorre frequentemente em PDFs de conteúdo didático.

Outro aspecto relevante é a limitação desses modelos em interpretar e reproduzir linguagem simbólica e diagramática, comum na química. Representações como estruturas de Lewis, orbitais moleculares, configurações eletrônicas ou mecanismos reacionais são fundamentais para a compreensão conceitual. No entanto, por dependerem exclusivamente da linguagem verbal, as IAs muitas vezes substituem tais elementos por descrições vagas ou imprecisas. Isso se torna especialmente problemático no contexto educacional. Pois, conceitos abstratos requerem esquemas representativos visuais para sua devida internalização. A ausência desses recursos visuais compromete não apenas a clareza das respostas, mas também sua funcionalidade pedagógica. Além disso, observa-se um fenômeno preocupante: a confabulação. Como já discutido por Morin (2014), os modelos de linguagem têm a tendência de “inventar” respostas. Estas, acompanhadas de justificativas com aparência lógica, mas que são conceitualmente equivocadas ou baseadas em fontes inexistentes. A fluência textual desses modelos cria uma falsa impressão de confiabilidade. No ambiente educacional, isso é extremamente perigoso. O estudante, sem uma base crítica consolidada, tende a aceitar essas explicações erradas como corretas, internalizando erros que podem comprometer sua

formação. Apesar das limitações evidenciadas, não se deve descartar o uso das IAs na educação química. Na verdade, o potencial dessas ferramentas é inegável, desde que utilizadas com mediação adequada. Como destaca Abreu e Lima (2024), as IAs podem atuar como ferramentas de apoio à personalização do ensino, à criação automatizada de exercícios e à curadoria de conteúdo. No entanto, é imprescindível que o professor assuma o papel de filtro crítico. Promovendo a validação e contexto das respostas fornecidas por essas tecnologias.

No campo do desenvolvimento tecnológico, é possível vislumbrar modelos híbridos, que combinem a fluidez linguística dos LLMs (Large Language Models) com motores simbólicos especializados em química. Isso permitiria que a IA não apenas reconhecesse comandos verbais, mas também realizasse cálculos estequiométricos, consultas de bases de dados. Além disso, realizar análises estruturais moleculares e aplicar os princípios da termodinâmica. Esse tipo de avanço pode aproximar as IAs de um raciocínio verdadeiramente científico. Outra possibilidade de aprimoramento envolve o treinamento de modelos com incorporações específicas da área. Como a importação de livros didáticos validados, artigos revisados por pares e bancos de questões oficiais (ENEM, vestibulares, olimpíadas). Isso reduziria a margem de erro e aumentaria a precisão conceitual das respostas. Do ponto de vista pedagógico, o desenvolvimento de metadados educacionais também se faz necessário. Atualmente, uma mesma pergunta, redigida de forma diferente, pode gerar respostas contraditórias. A IA, também, não reconhece o perfil do estudante ou identifica o nível de complexidade adequado. Com ajustes nesse sentido, seria possível entregar respostas mais coerentes com os objetivos da aprendizagem em cada etapa de ensino.

Portanto, mais do que substituir o professor, a IA deve ser pensada como uma ferramenta a ser integrada de forma crítica ao processo educativo. O uso indiscriminado pode reforçar equívocos e empobrecer o raciocínio dos estudantes. Contudo, se bem orientado, seu uso pode enriquecer as práticas pedagógicas, ampliar a autonomia dos alunos e diversificar as estratégias de ensino.

Conclusão

As transformações provocadas pela presença crescente de tecnologias cognitivas no ambiente educacional, especialmente as plataformas baseadas em inteligência artificial, trazem consigo um duplo desafio. Primeiro, a ampliação das possibilidades de acesso, visualização e aprofundamento do conhecimento. Segundo, a necessidade de compreender criticamente as limitações e os riscos implicados no uso dessas ferramentas. Os dados apresentados ao longo deste trabalho demonstram que, embora os assistentes virtuais como

ChatGPT, Gemini, Aria, Copilot e Deepseek apresentem certo potencial para atuar como ferramentas de apoio no ensino de Química, seu desempenho prático ainda está aquém do necessário para uma mediação pedagógica autônoma e confiável.

A análise dos resultados revelou que essas plataformas frequentemente não conseguem lidar satisfatoriamente com questões que exigem raciocínio lógico estruturado, domínio de linguagem simbólica ou interpretação precisa de comandos pedagógicos. Muitas vezes, as respostas fornecidas pelos algoritmos apresentam erros conceituais, desenvolvimentos incoerentes ou simplesmente ignoram os dados fornecidos. O que evidencia a ausência de compreensão real dos conteúdos. Essa limitação compromete diretamente a qualidade do ensino. Isso, quando essas ferramentas são utilizadas de forma indiscriminada e sem a devida supervisão docente.

A dificuldade das IAs em lidar com conteúdos químicos, como cálculos estequiométricos, conceitos de concentração, pH ou energia de ligação, revela não apenas falhas operacionais, mas também fragilidades na relação conceitual. A Química é uma ciência marcada por sua linguagem simbólica e por sua exigência de encadeamento lógico rigoroso. Ao substituir a compreensão por previsibilidade linguística, os algoritmos analisados tendem a gerar respostas que podem parecer corretas, mas que ocultam erros graves de fundamento. Esses dados alertam para o risco da chamada ilusão de confiabilidade. Como as respostas das IAs são, em geral, escritas com fluência e segurança, podem convencer o estudante. Mesmo quando estão equivocadas. Isso reforça a necessidade de se entender essas tecnologias como instrumentos auxiliares no processo educativo, e não como substitutas do professor. A mediação pedagógica continua sendo essencial para validar o conteúdo, estimular o pensamento crítico e garantir a aprendizagem significativa. Mais do que responder corretamente às questões, é fundamental que essas ferramentas sejam capazes de fomentar o raciocínio, a argumentação lógica e a construção autônoma do conhecimento. A simples introdução de tecnologias inteligentes no cotidiano escolar, sem planejamento pedagógico, não garante avanços reais no ensino de Química ou em qualquer outra disciplina. Por fim, este estudo abre espaço para novas investigações. Recomenda-se que futuras pesquisas explorem o desempenho das inteligências artificiais em diferentes áreas do conhecimento. Além de verificar os impactos do uso dessas tecnologias em contextos reais de sala de aula. Em paralelo, analisar como a mediação do professor pode potencializar ou corrigir o uso dessas ferramentas. Além disso, seria pertinente avaliar modelos de IA voltados exclusivamente para fins educacionais, treinados com bases científicas validadas e que considerem o nível de escolaridade dos usuários. A integração crítica e responsável das IAs à prática pedagógica é um caminho possível e desejável. No entanto, essa integração só será efetiva se orientada por

princípios educacionais sólidos, comprometidos com a formação integral, ética e autônoma dos estudantes.

Referências

Azambuja, Celso Candido de; Silva, Gabriel Ferreira. Novos desafios para a educação na Era da Inteligência Artificial. *Unisinos Journal of Philosophy*, v. 25, n 1, p 1-16, 2024.

Cury, Carlos H. et al. Inteligência artificial e os desafios cognitivos no ensino de Ciências. *Revista Ciência & Ensino*, v. 4, n. 1, p. 34-52, 2022

FERNANDES, Afonso Fonseca. Inteligência artificial e educação. *BIUS-Boletim Informativo Unimotrisaúde em Sociogerontologia*, v. 39, n. 33, p. 1-3, 2023.

Morin, Edgar. *A cabeça bem-feita: repensar a reforma, reformar o pensamento*. 10. ed. Rio de Janeiro: Bertrand Brasil, 2014.

Leite, Bruno S. INTELIGÊNCIA ARTIFICIAL E ENSINO DE QUÍMICA: UMA ANÁLISE PROPEDEÚTICA DO CHATGPT NA DEFINIÇÃO DE CONCEITOS QUÍMICOS. *Quim. Nova*, V. 46, n. 9, 915-923, 2023.

Rabelo, Ana Paula Brescancini; Fuini, Lucas Lambigarlini. *Uso de inteligência artificial no ensino de química: uma pesquisa bibliográfica*. Instituto Federal de São Paulo, campus São João da Boa Vista. 2022.

Rodrigues, Juliana C. et al. Limites e possibilidades da IA na aprendizagem científica. *Educação em Foco*, v. 26, n. 3, p. 78-95, 2021.

SANTOS, Mayke Franklin da Cruz; SILVA, Cleber Cezar da. *Desafios e possibilidades da inteligência artificial generativa para a educação básica: um estudo com docentes do IFTM – Campus Paracatu*. SciELO Preprints, 2025. Disponível em: <https://doi.org/10.1590/SciELOPreprints.9628>.

DA SILVA, Luis André Ferreira; SIQUEIRA, Nadilson; RODRIGUES, Vinicius Brasil. *O uso da inteligência artificial como ferramenta para educação no Brasil*. *Revista Sociedade Científica*, v. 7, n. 1, p. 3546-3568, 2024. doi:10.61411/rsc202455317.

ALBUQUERQUE, José Gicelmo Melo; ABREU, Mirella Teresinha Corrêa de; LIMA, Ivanilton Neves de. O impacto da Inteligência Artificial na personalização do ensino. *Revista Brasileira de Ensino e Aprendizagem*, v. 9, p. 182-192, 2024. Disponível em: <https://rebena.emnuvens.com.br/revista/index>.