

Inteligência artificial como hipótese a ser testada: análise empírica de uma aula prática investigativa de Química no Ensino Médio

DOI: <https://doi.org/10.33871/23594381.2026.24.2.12305>

Leonardo Braga Guimarães¹, Érika Aparecida Menegardo Onhas Marques², Carlos Roberto Pires Campos³, Marize Lyra Silva Passos⁴, Deluzia Daleprane Queiroz Péres⁵

Resumo: Este artigo apresenta a análise empírica de uma aula prática investigativa de Química conduzida com 52 estudantes do 1º e 2º anos do Ensino Médio em uma escola pública de tempo integral no município de Anchieta (ES), na qual ferramentas de inteligência artificial generativa (ChatGPT e Gemini) foram integradas à determinação experimental do teor de etanol em amostras de gasolina. O diferencial metodológico consistiu em deslocar a IA da posição de fonte de respostas para a hipótese a ser testada pelo experimento. Aplicou-se questionário pós-aula com 12 itens em escala Likert e duas questões abertas, organizado em cinco constructos: aprendizagem, pensamento crítico, IA como ferramenta, uso responsável e papel do professor e do experimento. Os dados quantitativos foram tratados por estatística descritiva, análise de consistência interna (α de Cronbach) e teste de Mann-Whitney entre séries. As respostas abertas foram submetidas à análise de conteúdo temática. O instrumento apresentou confiabilidade global elevada ($\alpha = 0,815$). As médias foram positivas em todos os constructos, com destaque para "Professor e experimento" ($M = 4,59$) e "Uso responsável" ($M = 4,47$). Os estudantes do 1º ano apresentaram maior aceitação da IA como ferramenta ($p = 0,034$). A análise qualitativa identificou seis categorias, sendo as mais frequentes a omissão/incompletude das respostas (21,2%) e os erros factuais na identificação de amostras adulteradas (13,5%). Os resultados sustentam que o uso pedagógico produtivo da IA depende menos da ferramenta em si e mais do design da atividade que obriga sua verificação.

Palavras-chaves: inteligência artificial generativa; ensino de Química; pensamento crítico; aula prática investigativa.

Artificial intelligence as a hypothesis to be tested: empirical analysis of an investigative chemistry lab class in high school

Abstract: This paper presents an empirical analysis of an investigative chemistry lab class conducted with 52 first- and second-year high school students at a public full-time school in

¹ Doutorando do Programa de Pós-graduação em Educação em Ciências e Matemática do Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo, Vila Velha, ES, Brasil. bragaguimaraes.leo@gmail.com, <https://orcid.org/0000-0002-3771-3042>

² Doutoranda do Programa de Pós-graduação em Educação em Ciências e Matemática do Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo, Vila Velha, ES, Brasil. erika.onhas@educador.edu.es.gov.br, <https://orcid.org/0000-0002-7145-0047>

³ Doutor em História Social da Cultura pela Pontifícia Universidade Católica do Rio de Janeiro e Pós-Doutorado em Educação, Ciência e Tecnologia pelo CEFET-RJ. Docente do Programa de Pós-Graduação em Educação em Ciências e Matemática do Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo, Vila Velha, ES, Brasil. <https://orcid.org/0000-0001-7708-4597>

⁴ Doutora em Engenharia pela Universidade Federal do Rio Grande do Sul e Doutora em Educação pela Universidad del Norte. Docente do Programa de Pós-Graduação em Educação em Ciências e Matemática do Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo, Vila Velha, ES, Brasil. <https://orcid.org/0000-0001-7047-5018>

⁵ Doutoranda do Programa de Pós-graduação em Educação em Ciências e Matemática do Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo, Vila Velha, ES, Brasil. deluziadaleprane@gmail.com, <https://orcid.org/0000-0002-4834-1686>

Anchieta, Espírito Santo, Brazil. The activity integrated generative AI tools (ChatGPT and Gemini) into the experimental determination of ethanol content in gasoline samples. The methodological distinction lies in repositioning AI from a source of answers to a hypothesis to be tested against experimental evidence. A post-lesson questionnaire with 12 Likert-scale items and two open-ended questions was administered, organized into five constructs: learning, critical thinking, AI as a tool, responsible use, and the role of the teacher and the experiment. Quantitative data were analyzed through descriptive statistics, internal consistency (Cronbach's α), and Mann-Whitney tests between grade levels. Open responses underwent thematic content analysis. The instrument showed high overall reliability ($\alpha = 0.815$). Means were positive across all constructs, particularly for "Teacher and experiment" ($M = 4.59$) and "Responsible use" ($M = 4.47$). First-year students reported higher acceptance of AI as a tool ($p = 0.034$). Qualitative analysis identified six categories, the most frequent being omission/incompleteness of responses (21.2%) and factual errors in identifying adulterated samples (13.5%). The findings support the claim that productive pedagogical use of AI depends less on the tool itself than on activity design that demands its verification.

Keywords: generative artificial intelligence; chemistry education; critical thinking; investigative laboratory class.

Introdução

A Resolução ANP nº 807, de agosto de 2020, fixou em 27% o teor mínimo de etanol anidro para a gasolina C comercializada no Brasil. O dado afeta motoristas, postos e órgãos de fiscalização, mas também abre uma janela pedagógica pouco explorada no Ensino Médio: a partir de um princípio químico direto — diferença de polaridade entre etanol e gasolina, modulada pelo efeito *salting-out* — é possível construir uma aula investigativa, em que os estudantes verifiquem, em laboratório, a conformidade legal de combustíveis disponíveis no mercado.

A pergunta que originou esta pesquisa não adveio da literatura, mas de uma situação real: estudantes do Ensino Médio chegavam à aula de Química usando ferramentas de IA generativa para pesquisar conteúdo, sem qualquer critério para avaliar o que recebiam. O cenário tornou-se rotina depois de 2022, quando modelos de linguagem como ChatGPT e Gemini passaram a estar disponíveis gratuitamente, sem necessidade de treinamento técnico. Revisões sistemáticas recentes documentam ganhos consistentes em motivação e personalização do ensino (Garzón; Patiño; Marulanda, 2025; Lee *et al.*, 2025), mas também alertam para a aceitação acrítica de respostas incorretas, o atrofamento da autonomia intelectual e a perda da experiência sensorial não substituível pelo texto (Almasri, 2024; Leite, 2023).

Diante desse cenário, a questão pedagógica relevante não é se a IA deve entrar na sala de aula — ela já entrou —, mas como inseri-la de forma que produza aprendizagens.

A proposta aqui descrita responde por um deslocamento metodológico preciso: tratar a IA como hipótese a ser testada pelo experimento, e não como oráculo a ser consultado. Os estudantes são convidados a verificar, com vidraria e reagentes, se aquilo que a ferramenta previu corresponde ao que o experimento produziu — de fonte de respostas, a IA passa a objeto de investigação.

O objetivo deste artigo foi investigar a percepção de estudantes do Ensino Médio sobre o uso da IA generativa em uma aula prática investigativa, avaliando seus impactos sobre a aprendizagem, o pensamento crítico, a percepção de utilidade, a consciência de uso responsável e reconhecimento do papel docente.

Fundamentação teórica

IA generativa no ensino de ciências

A disponibilidade gratuita de modelos de linguagem para qualquer estudante com acesso à internet criou uma situação sem precedente nas salas de aula: a ferramenta chegou antes que os sistemas educacionais tivessem tempo de pensar em diretrizes para seu uso. Garzón, Patiño e Marulanda (2025), em revisão sistemática de 155 estudos publicados entre 2015 e 2025, documentam crescimento expressivo das pesquisas sobre IA generativa na educação a partir de 2022 e identificam as "alucinações" — respostas factualmente incorretas apresentadas com aparência de confiabilidade — como um dos principais desafios para uso pedagógico.

No ensino de Química, Leite (2023) examinou as respostas do ChatGPT a perguntas sobre conceitos fundamentais da disciplina. Na maioria dos casos, a ferramenta produziu respostas coerentes e acessíveis. Em conceitos-chave como a distinção entre as teorias ácido-base de Arrhenius e Brønsted-Lowry, porém, o modelo mesclou definições, sem sinalizar imprecisão. Um estudante sem base prévia dificilmente perceberia o erro. Almasri (2024), em revisão de estudos experimentais, observou que parte significativa dos estudantes aceita as respostas geradas sem avaliação crítica — o que impõe ao design da atividade a responsabilidade de criar situações em que a verificação seja necessária.

Aprendizagem ativa e mediação cognitiva

A integração da IA generativa ao ensino investigativo adquire sentido quando situada em relação às perspectivas clássicas da psicologia da aprendizagem que fundamentam o design de atividades ativas e mediadas, condição para avaliar o que a ferramenta pode e não pode oferecer pedagogicamente.

A estrutura da atividade apoia-se em três perspectivas teóricas clássicas. De Piaget, toma-se a ideia de aprendizagem ativa: o conhecimento se constrói pelo contato direto com o objeto, não pela transmissão de informação. A etapa experimental — em que os estudantes medem volumes, calculam percentuais e decidem qual amostra está adulterada — atende diretamente a esse princípio. De Vygotsky, o conceito de zona de desenvolvimento proximal orienta o uso da IA como mediadora inicial: a ferramenta pode dar acesso a um vocabulário e a uma explicação que ainda estão além do que o estudante produziria sozinho, sem substituir o processo de internalização, que ocorre na interação com colegas e com o professor. Almeida Neto *et al.* (2024) demonstraram esse papel em uma calculadora química baseada em ChatGPT-4, que funcionou como andaime ao propor perguntas orientadoras progressivas em lugar de fornecer a resposta pronta. De Bruner, a noção de descoberta guiada sustenta a organização das cinco etapas da atividade: o estudante não parte do zero, mas também não recebe o roteiro completo — o que mantém ativa a tensão cognitiva da qual depende a aprendizagem significativa.

Pensamento crítico e papel da mediação docente

Leite (2023) sintetiza, com precisão, o problema central: a educação do século XXI exige que o estudante pense, reflita e confronte e isso não acontece automaticamente pela introdução de ferramentas digitais — depende do design da atividade. Gomes *et al.* (2026) registram que o próprio ChatGPT reconhece não ser especialista em Química e recomenda verificação técnica, o que não é modéstia algorítmica, mas limitação real de modelos que produzem texto estatisticamente provável, não necessariamente correto. Lee e Kwon (2024), em revisão de estudos K-12, reforçam que o desenvolvimento profissional docente é condição necessária para o uso ético dessas ferramentas: em contextos nos quais os professores não receberam formação específica para mediar o uso de IA, os estudantes tenderam a empregá-la de forma passiva, como repositório de respostas prontas, em vez de instrumento de questionamento. Esse achado é diretamente relevante para o desenho da atividade aqui descrita: a professora exerceu uma função de interposição crítica indispensável — ao nomear os erros da IA, sinalizar suas lacunas e restituir o protagonismo do experimento como instância de verificação. A competência docente para explorar pedagogicamente as falhas da ferramenta é a condição que transforma o erro algorítmico em oportunidade de aprendizagem.

Metodologia

Caracterização do estudo e participantes

Trata-se de pesquisa aplicada de abordagem mista, com delineamento descritivo-exploratório e corte transversal. A coleta de dados ocorreu nos meses finais do primeiro semestre letivo de 2026 no Centro Estadual de Ensino Médio em Tempo Integral Paulo Freire, em Anchieta (ES). Participaram 52 estudantes matriculados na disciplina de Química Geral e Inorgânica, distribuídos entre 1º ano ($n = 25$; 48,1%) e 2º ano ($n = 27$; 51,9%), com idade entre 14 e 17 anos ($M = 15,75$; $DP = 0,74$). O regime de tempo integral viabilizou aulas de laboratório com duração suficiente para o ciclo investigativo proposto. A participação foi voluntária e anônima, com autorização institucional prévia e direito de desistência a qualquer momento, em observância aos preceitos éticos.

Contexto disciplinar e metodologia pedagógica

O tema — determinação do teor de etanol em gasolina — foi escolhido por sua relevância social direta: a Resolução ANP nº 807/2020 e a Portaria MAPA nº 75/2015 estabelecem limites legais (27% para gasolina C comum e 25% para Premium) que tornam a adulteração um problema concreto e verificável em laboratório. O experimento baseia-se no efeito *salting-out*: em presença de íons Na^+ e Cl^- em solução aquosa saturada, o etanol — molécula anfifílica — migra preferencialmente para a fase aquosa, separando-se da gasolina apolar. O volume final da fase aquosa após a agitação indica o teor de etanol extraído da amostra. O cálculo é direto: $\text{teor (\%)} = [(V_{\text{final}} - 50) / 50] \times 100$, em que V_{final} é o volume da fase aquosa após a mistura com 50 mL de gasolina. A atividade foi estruturada em cinco etapas sequenciais (Quadro 1), combinando o método científico investigativo com o uso intencional da IA como instância a ser testada.

Quadro 1: Etapas do ciclo investigativo com IA.

Etapas	Nome	Descrição
I	Consulta à IA	Os estudantes consultam ChatGPT ou Gemini para compreender o contexto científico do experimento e formular hipóteses iniciais sobre qual amostra de gasolina está adulterada.
II	Planejamento	Com base nas informações obtidas, os grupos organizam o experimento: identificam materiais, definem o procedimento e registram as hipóteses por escrito.
III	Experimentação	Execução do procedimento de extração líquido-líquido com solução saturada de NaCl. Medição do volume da fase aquosa e cálculo do teor de etanol em cada amostra.
IV	Confronto IA × dados	Comparação crítica entre as previsões da IA na Etapa I e os resultados experimentais. Os estudantes apresentam seus dados à IA e avaliam a qualidade das respostas recebidas.
V	Síntese e reflexão	Elaboração de conclusões, avaliação crítica da IA como ferramenta científica e redação do relatório final mediada pela professora.

Fonte: elaborado pelos autores (2026).

Instrumento de coleta

O questionário pós-aula continha três seções: (a) perfil do respondente (idade e turma); (b) doze itens em escala Likert de cinco pontos (1 = Discordo totalmente; 5 = Concordo totalmente); e (c) duas questões abertas — a primeira sobre erros ou omissões da IA durante a aula, a segunda comparando a experiência com uma aula prática sem IA. Os doze itens foram distribuídos em cinco constructos teoricamente fundamentados (Quadro 2), com referências de validação por bloco. O instrumento foi aplicado em formulário eletrônico imediatamente após a atividade, com tempo médio de resposta de seis minutos.

Quadro 2: Constructos e itens do questionário pós-aula.

Bloco	Constructo	O que mede	Itens	Referências
A	Aprendizagem	Percepção de ganho cognitivo com apoio da IA antes e durante o experimento	1, 2	Almasri (2024)
B	Pensamento crítico	Capacidade de questionar respostas da IA, comparar com evidências e perceber suas limitações	3, 4, 5	Leite (2023); Garzón <i>et al.</i> (2025)
C	IA como ferramenta	Utilidade percebida, motivação, engajamento e demanda por replicação em outras aulas	6, 7, 8	Davis (1989); Lee <i>et al.</i> (2025)
D	Uso responsável	Consciência sobre necessidade de verificação e risco de uso acrítico	9, 10	Gomes <i>et al.</i> (2026); Lee; Kwon (2024)
E	Professor e experimento	Indispensabilidade da mediação docente e da prática laboratorial concreta	11, 12	Almeida Neto <i>et al.</i> (2024); Gomes <i>et al.</i> (2026)

Fonte: elaborado pelos autores (2026).

Procedimentos de análise

Os dados Likert foram convertidos em escala numérica de 1 a 5 e submetidos à estatística descritiva (média, desvio-padrão, mediana, moda e percentual de respostas positivas — categorias "Concordo" e "Concordo totalmente"). A consistência interna foi avaliada pelo coeficiente α de Cronbach, calculado por constructo e para o instrumento global. A comparação entre os grupos de 1º e 2º anos foi conduzida via teste não paramétrico de Mann-Whitney ($\alpha = 0,05$), justificado pelo caráter ordinal da escala e pela ausência de pressuposto de normalidade. As respostas abertas foram submetidas à análise de conteúdo temática nos moldes propostos por Bardin (2011), com procedimento de unitização, categorização indutivo-dedutiva (categorias pré-definidas a partir da literatura, abertas a categorias emergentes) e quantificação das unidades de registro por categoria. A análise quantitativa foi realizada em Python 3.11 (Van Rossum; Drake, 2009) com as bibliotecas pandas (McKinney, 2010) e SciPy (Virtanen *et al.*, 2020).

Resultados e Discussão

Procedimentos de análise

Todos os 52 estudantes responderam à totalidade dos itens Likert, sem dados ausentes. O Quadro 3 apresenta as estatísticas descritivas por item e por constructo, com indicação do percentual de respostas positivas (Concordo + Concordo totalmente).

Quadro 3: Estatísticas descritivas por item e por constructo (n = 52).

Bloco	Item	Enunciado (resumido)	M	DP	Md	% CT+C
A	1	Consultar a IA antes do experimento me ajudou a entender melhor o que ia acontecer	4,17	0,58	4	90,4%
A	2	Juntar a IA com o experimento prático me fez aprender mais do que se usasse só um dos dois	4,12	0,81	4	80,8%
A		<i>Média do constructo (Aprendizagem)</i>	4,14	0,70	—	85,6%
B	3	A atividade me fez questionar se as respostas da IA estavam realmente corretas	3,83	0,96	4	65,4%
B	4	Comparar o que a IA disse com o que eu vi no experimento me fez pensar de forma mais crítica	3,90	0,89	4	67,3%
B	5	Percebi que a IA pode errar ou dar respostas incompletas sobre Química	4,12	0,81	4	76,9%
B		<i>Média do constructo (Pensamento crítico)</i>	3,95	0,89	—	69,9%
C	6	Usar a IA deixou a aula mais interessante	4,00	0,79	4	73,1%
C	7	A IA funcionou bem como ponto de partida, mas não substituiu o experimento	4,42	0,72	5	86,5%
C	8	Gostaria que outras aulas de Química também usassem IA como apoio	4,19	0,84	4	80,8%
C		<i>Média do constructo (IA como ferramenta)</i>	4,21	0,80	—	80,1%
D	9	Apreendi que é importante conferir as informações da IA com o professor ou com o livro	4,54	0,61	5	94,2%
D	10	Copiar respostas da IA sem pensar prejudica meu aprendizado	4,40	0,75	5	88,5%
D		<i>Média do constructo (Uso responsável)</i>	4,47	0,68	—	91,4%
E	11	A professora foi essencial para eu entender o conteúdo, mesmo usando a IA	4,69	0,61	5	92,3%
E	12	O experimento prático me ensinou coisas que a IA sozinha não conseguiria ensinar	4,48	0,78	5	82,7%
E		<i>Média do constructo (Professor e experimento)</i>	4,59	0,71	—	87,5%

M = média; DP = desvio-padrão; Md = mediana; % CT+C = percentual de respostas "Concordo" e "Concordo totalmente". Escala: 1 = Discordo totalmente; 5 = Concordo totalmente.

Fonte: dados da pesquisa (2026).

Todos os doze itens obtiveram médias superiores ao ponto neutro (3,0), indicando avaliação predominantemente positiva. As maiores médias concentram-se nos constructos do papel do professor e do experimento (E; M = 4,59; DP = 0,71) e do uso responsável da IA (D; M = 4,47; DP = 0,68). O item de maior média individual foi o 11 — sobre a colaboração pedagógica da professora — com M = 4,69 (76,9% de "Concordo totalmente"). A menor média coube ao pensamento crítico (B; M = 3,95), puxada pelos

itens 3 e 4, que pedem ao estudante reconhecer que questionou ativamente a IA durante a aula.

A análise de consistência interna está apresentada no Quadro 4. O instrumento como um todo apresentou $\alpha = 0,815$, valor classificado como de boa confiabilidade. Os valores por constructo, contudo, são mais heterogêneos.

Quadro 4: Consistência interna (α de Cronbach) por constructo e do instrumento global.

Constructo	Itens	α de Cronbach	Interpretação
A — Aprendizagem	1–2	0,641	Aceitável para escala curta (2 itens); confiabilidade moderada
B — Pensamento crítico	3–5	0,556	Baixa; itens medem dimensões parcialmente distintas
C — IA como ferramenta	6–8	0,684	Aceitável; consistência razoável entre itens de utilidade percebida
D — Uso responsável	9–10	0,133	Baixa; itens com pouca variância (efeito teto); ver discussão
E — Professor e experimento	11–12	0,599	Moderada; convergência satisfatória entre itens
Instrumento global	1–12	0,815	Boa; confiabilidade global elevada do instrumento

Fonte: dados da pesquisa (2026).

Os α dos constructos B (0,556) e D (0,133) ficaram abaixo do limiar usualmente aceito (0,70). No constructo D, a baixa consistência reflete efeito teto (94,2% e 88,5% de concordância nos itens 9 e 10), que reduz a variância e penaliza o coeficiente. No constructo B, sugere que os três itens medem dimensões parcialmente distintas do pensamento crítico — disposição a questionar, comparação com evidência empírica e reconhecimento de limitações —, podendo demandar refinamento em versões futuras.

A comparação entre estudantes do 1º e dos 2º anos (Quadro 5) indicou diferença estatisticamente significativa apenas no constructo C (IA como ferramenta; $U = 452$; $p = 0,034$), com mediana mais alta no 1º ano ($Mdn = 4,67$) que no 2º ano ($Mdn = 4,00$). Nos demais constructos, as diferenças entre séries não atingiram significância estatística.

Quadro 5: Comparação entre 1º e 2º anos via teste de Mann-Whitney.

Constructo	Mdn 1ª série	Mdn 2ª série	U	p	Decisão
A — Aprendizagem	4,50	4,00	408	0,183	Ns
B — Pensamento crítico	4,00	3,67	372	0,522	Ns
C — IA como ferramenta	4,67	4,00	452	0,034	p < 0,05 *
D — Uso responsável	4,50	4,50	324	0,810	Ns
E — Professor e experimento	5,00	5,00	362	0,626	Ns

ns = não significativo; * $p < 0,05$.

Fonte: dados da pesquisa (2026)

Análise qualitativa

As respostas à questão 13 — sobre erros ou omissões da IA durante a aula — foram submetidas à análise de conteúdo temática, resultando em seis categorias

mutuamente exclusivas (Quadro 6). Cada resposta foi alocada à categoria predominante; respostas em que não houve relato de problema foram categorizadas como "Sem erro/omissão detectada".

Quadro 6: Categorias temáticas da análise de conteúdo das respostas à questão 13.

Categoria temática	n	%	Exemplo representativo
Sem erro/omissão detectada	23	44,2	"Não errou e nem deixou de falar algo" [E35]; "as explicações eram completas juntamente da presença da professora" [E24]
Omissão / incompletude das respostas	11	21,2	"Ela resumiu muito, omitiu algumas informações, usou termos em inglês que eu nunca havia visto e deixou algumas respostas vagas" [E15]
Erros factuais (identificação incorreta de amostras)	7	13,5	"A IA havia dito que a mistura com mais etanol era a A, sendo que a única que continha mais quantidade de etanol era a C" [E38]
Não-contextualização ao protocolo experimental	6	11,5	"Ela enviou informações que não estavam presentes na prática. Como ex: erro de temperatura sem o uso da temperatura durante o experimento" [E10]
Generalidade / superficialidade	3	5,8	"ela disse coisas muito genéricas e rasas, não explicou muito bem" [E45]
Necessidade de correção pelo professor	2	3,8	"Sim, se não tivesse a observação da professora iríamos errar" [E47]
Total	52	100,0	

Fonte: dados da pesquisa (2026).

A categoria mais frequente foi "Sem erro/omissão detectada" (44,2%). Entre os 55,8% restantes, a categoria modal foi "Omissão/incompletude das respostas" (21,2%), seguida por "Erros factuais" (13,5%) e "Não-contextualização ao protocolo experimental" (11,5%). Os excertos a seguir ilustram cada categoria.

No grupo dos erros factuais, chamou atenção a recorrência de equívocos na identificação das amostras adulteradas — exatamente o ponto em que a verificação experimental se mostra decisiva:

"A IA havia dito que a mistura com mais etanol era a A, sendo que a única que continha mais quantidade de etanol era a C." [E38]

A categoria de omissão/incompletude reuniu relatos em que a IA forneceu respostas coerentes, porém insuficientes para o detalhe exigido pela atividade:

"Ela resumiu muito, omitiu algumas informações, usou termos em inglês que eu nunca havia visto e deixou algumas respostas vagas." [E15]

A categoria de não-contextualização ao protocolo revela um padrão expressivo: a IA mobilizou explicações genéricas, sem considerar as condições específicas do procedimento adotado:

"Ela enviou informações que não estavam presentes na prática. Como exemplo: erro de temperatura sem o uso da temperatura durante o experimento, o que acaba confundindo." [E10]

Em duas respostas, os estudantes registraram explicitamente que a mediação docente foi o que evitou que o erro da IA lhes fosse incorporado como conhecimento:

"Sim, se não tivesse a observação da professora iríamos errar." [E47]

As respostas à questão 14 — comparação com uma aula prática sem IA — revelaram quatro padrões transversais: o reconhecimento da IA como "base" de partida, suporte inicial bem avaliado para enfrentar conteúdos novos; a percepção de que o experimento e a professora cumpriram função distinta e insubstituível; a emergência espontânea de uma postura crítica; e, o mais relevante, a verbalização do risco de dependência:

"Foi melhor porque podemos observar que nem sempre a IA está certa, e sem a IA talvez não poderíamos perceber tantas diferenças entre resultados." [E05]

"A IA é como uma assistente, nunca vai substituir o professor, exemplo, no cálculo, eu nunca entenderia facilmente a explicação de uma IA igual a da professora." [E06]

"Eu acredito que o uso da IA acaba nos deixando dependente delas. Eu percebi que por mais que o uso da IA me ajudou a iniciar o trabalho, eu tive dificuldade de concluir, de pensar sozinha sem a IA." [E15]

"O melhor nessa aula foi poder ter feito a prática e ter o aprendizado que nem tudo que você vê em uma inteligência artificial está correto." [E52]

Discussão

Os resultados convergem com a literatura internacional em três pontos centrais e divergem em um aspecto. A primeira convergência está no reconhecimento dos estudantes de que a mediação docente permanece indispensável, mesmo em atividades apoiadas por IA. O constructo E — papel do professor e do experimento — obteve a maior média do instrumento ($M = 4,59$), com o item sobre a essencialidade da professora atingindo $M = 4,69$, consistente com Almeida Neto *et al.* (2024) e com Lee e Kwon (2024), que identificam o desenvolvimento profissional docente como condição para o uso ético da IA na educação básica. A análise qualitativa fortalece essa convergência: os relatos espontâneos sobre erros corrigidos pela professora (E47, E50) e a verbalização de E06 — "a IA é como uma assistente, nunca vai substituir o professor" — indicam que o reconhecimento da insubstituibilidade docente não é socialmente desejável, mas uma conclusão extraída da experiência vivida.

A segunda convergência refere-se à identificação, pelos próprios estudantes, das limitações técnicas dos modelos generativos. As categorias de omissão/incompletude (21,2%) e de não-contextualização ao protocolo (11,5%) reproduzem os padrões descritos por Leite (2023): respostas tecnicamente coerentes, porém insuficientes em profundidade ou imprecisas em conceitos específicos. O excerto de E10 — sobre a IA mencionar "erro de temperatura" em um experimento em que a temperatura não foi controlada — exemplifica o que Garzón, Patiño e Marulanda (2025) descrevem como tendência das ferramentas generativas a produzir conteúdo plausível, mas desprovido do contexto particular da aplicação.

A terceira convergência aponta para o risco de dependência digital, documentado em revisões sistemáticas (Almasri, 2024; Garzón; Patiño; Marulanda, 2025) e verbalizado por E15: "o uso da IA acaba nos deixando dependente delas". A explicitação espontânea do risco pela própria estudante sugere que o design da atividade — ao expor a IA à verificação experimental — facilitou a consciência sobre limites de uso, em vez de produzi-la apenas como discurso institucional.

A divergência relevante em relação à literatura aparece no contraste entre os constructos B (pensamento crítico) e D (uso responsável). Almasri (2024) reporta que atividades com IA tendem a produzir avaliações homogeneamente altas em medidas de pensamento crítico. No presente estudo, porém, o constructo B apresentou a menor média do instrumento ($M = 3,95$), com 65,4% e 67,3% de concordância nos itens 3 e 4, contra 94,2% e 88,5% no constructo D. A discrepância sugere que os estudantes reconhecem normativamente a necessidade de verificar a IA (constructo D), mas relatam, com menor confiança, ter exercido esse questionamento durante a aula (constructo B). Esse hiato entre norma e prática é, em si, pedagogicamente relevante: o letramento digital exige treino sistemático de questionamento, não apenas conscientização. A análise qualitativa, paradoxalmente, mostra que muitos estudantes exerceram o pensamento crítico com profundidade — identificando erros factuais na escolha da amostra adulterada, percebendo contradições e omissões —, ainda que não tenham se reconhecido plenamente nesse exercício na escala Likert.

A diferença significativa no constructo C entre as séries ($p = 0,034$) merece nota: estudantes do 1º ano avaliaram a IA mais positivamente que os do 2º ano, divergência plausivelmente atribuível ao efeito de novidade — estudantes mais jovens, com menor exposição prévia, tenderam a percebê-la como mais útil. A análise qualitativa, contudo,

não revelou diferenças temáticas entre séries, o que sugere que a percepção crítica da IA emergiu de modo análogo em ambos os grupos.

Limitações do estudo

Quatro limitações condicionam a interpretação dos achados. A primeira é a ausência de grupo controle: sem comparação direta com uma aula prática equivalente sem IA, não é possível atribuir os ganhos percebidos especificamente à ferramenta, e não ao design investigativo da atividade.

A segunda limitação é o viés de autorrelato: o instrumento mede a percepção dos estudantes sobre a experiência, não a aprendizagem efetiva mensurada por desempenho. Escalas Likert pós-aula capturam, sobretudo, satisfação imediata e tendem ao efeito teto em populações engajadas, como se observou no constructo D.

A terceira limitação é o caráter localizado do contexto: a pesquisa foi conduzida em uma única escola pública de tempo integral, com 52 estudantes, cujas condições de infraestrutura e mediação docente são particulares — a transferibilidade dos achados demanda replicação. A quarta refere-se ao instrumento: dois dos cinco constructos apresentaram α de Cronbach abaixo de 0,70, o que sugere refinamento dos itens, em particular dos blocos B e D, antes de aplicações em maior escala.

Considerações finais

Esta pesquisa partiu da hipótese de que o modo como a IA é inserida na prática importa mais do que a decisão de usá-la ou não. Os dados de 52 estudantes sustentam essa hipótese: no plano quantitativo, o instrumento mostrou-se confiável ($\alpha = 0,815$), com as maiores médias na indispensabilidade do professor e do experimento e no uso responsável; no qualitativo, 55,8% dos estudantes identificaram, espontaneamente, limitações concretas da IA, e parte significativa verbalizou consciência sobre o risco de dependência digital.

A contribuição original do estudo consiste na proposta de deslocamento epistêmico da IA: tratá-la como hipótese a ser testada experimentalmente, em lugar de fonte de respostas. Os achados sustentam a viabilidade pedagógica desse deslocamento, que promove o reconhecimento dos limites da ferramenta sem comprometer o engajamento. Para pesquisas futuras, três encaminhamentos parecem prioritários: replicar o design com grupos controle em contextos diversos; refinar os itens dos constructos B e D; e investigar, com instrumentos longitudinais, o efeito de exposições repetidas sobre o letramento digital ao longo do Ensino Médio.

Referências

ALMASRI, F. Exploring the impact of artificial intelligence in teaching and learning of science: a systematic review of empirical research. **Research in Science Education**, v. 54, n. 5, p. 977-997, 2024.

ALMEIDA NETO, R. et al. Integração da inteligência artificial na educação em química: desenvolvimento e avaliação de uma ferramenta interativa sob a perspectiva de teorias do desenvolvimento cognitivo. **Revista Caderno Pedagógico**, v. 21, n. 4, p. 1-29, 2024.

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2011.

BRASIL. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Resolução ANP nº 807, de 28 de agosto de 2020. **Diário Oficial da União**: Brasília, DF, 31 ago. 2020.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Portaria MAPA nº 75, de 5 de março de 2015. **Diário Oficial da União**: Brasília, DF, 6 mar. 2015.

DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. **MIS Quarterly**, v. 13, n. 3, p. 319-340, 1989.

GARZÓN, J.; PATIÑO, E.; MARULANDA, C. Systematic review of artificial intelligence in education: trends, benefits, and challenges. **Multimodal Technologies and Interaction**, v. 9, n. 8, art. 84, 2025.

GOMES, L. F. et al. Aplicações de inteligência artificial na educação básica: uma análise de suas contribuições e desafios. **Revista Sítio Novo**, v. 10, n. 1, p. 1-22, 2026.

LEE, H. Y. et al. Artificial intelligence (AI) learning tools in K-12 education: a scoping review. **Journal of Computers in Education**, v. 12, n. 1, p. 93-131, 2025.

LEE, S. J.; KWON, K. A systematic review of AI education in K-12 classrooms from 2018 to 2023: topics, strategies, and learning outcomes. **Computers and Education: Artificial Intelligence**, v. 6, art. 100211, 2024.

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

McKINNEY, W. Data structures for statistical computing in Python. In: PYTHON IN SCIENCE CONFERENCE, 9., 2010, Austin. **Proceedings [...]**. Austin: [s. n.], 2010. p. 56-61.

VAN ROSSUM, G.; DRAKE, F. L. **Python 3 reference manual**. Scotts Valley: CreateSpace, 2009.

VIRTANEN, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, v. 17, n. 3, p. 261-272, 2020.

Submissão: 17/06/2026. **Aprovação:** 06/07/2026. **Publicação:** 31/08/2026.