

A escola na era pós-ChatGPT: revisão sistemática sobre as interações com IA, aprendizagem e agência no Ensino de Ciências

Schools in the post-ChatGPT era: A systematic review of human-AI interaction, learning, and agency in science education

Rodrigo Travitzki ^a

^a Faculdade de Educação, Universidade Estadual de Campinas (Unicamp), Campinas, Brasil.

Resumo. O lançamento do ChatGPT no final de 2022 inaugura uma nova era de incertezas e possibilidades na educação. Este artigo apresenta uma revisão sistemática sobre as interações de alunos e professores com ferramentas de Inteligência Artificial Generativa (IAG) no Ensino de Ciências na Educação Básica, seguindo as diretrizes do PRISMA. Analisamos 17 estudos publicados desde 2023. Os resultados mostram que a qualidade das interações depende de capacidades prévias do usuário (sobretudo conhecimento disciplinar), que são condições para que ele mantenha a agência. Quando o usuário preserva sua agência, os resultados tendem a ser melhores do que quando a agência é delegada à ferramenta. A mesma redução de esforço que pode liberar o aluno para pensar com maior profundidade pode, em outras condições, simplesmente dispensá-lo de pensar. Destacamos alguns riscos: pensamento superficial; dependência acrítica; aprofundamento de desigualdades educacionais. Os achados têm implicações para o ensino, como a importância do uso pedagógico e intencional da IAG para a aprendizagem, buscando manter a agência de alunos e professores. Sobre a pesquisa na área, identificamos algumas lacunas no recorte analisado: a ausência de estudos sobre efeitos de médio e longo prazo sobre aprendizagem; a escassez de investigações sobre efeitos potencialmente prejudiciais.

Palavras-chave:

inteligência artificial generativa, ensino de ciências, aprendizagem, agência humana, revisão sistemática.

Submetido em

21/05/2026

Aceito em

24/09/2026

Publicado em

19/10/2026

Abstract. The launch of ChatGPT in late 2022 has ushered in a new era of uncertainties and possibilities in education. This article presents a systematic review of how students and teachers interact with Generative Artificial Intelligence (GAI) tools in K-12 science education, following the PRISMA guidelines. We analyzed 17 studies published since 2023. The results show that the quality of these interactions depends on the user's prior capabilities (especially disciplinary knowledge), which are prerequisites for maintaining agency. When users preserve their agency, outcomes tend to be better than when agency is delegated to the tool. The same reduction in effort that can free students to think more deeply can, under other conditions, simply eliminate the need for them to think at all. We highlight some risks: superficial thinking, uncritical dependence, and the exacerbation of educational inequalities. The findings have implications for educational practice, emphasizing the importance of the pedagogical and intentional use of GAI for learning, aimed at preserving student and teacher agency. Regarding research in this area, we identified some gaps within the scope of this review: the absence of studies on medium- and long-term effects on learning, and the scarcity of investigations into potentially harmful effects.

Keywords: generative artificial intelligence, science education, learning, human agency, systematic review.

Introdução

O lançamento do ChatGPT no final de 2022 marca o início de uma nova era na inteligência artificial, com imprevisíveis consequências para a sociedade. Esse salto tecnológico só foi possível com a massiva disponibilidade de textos na internet e também devido a uma inovação estatística: o modelo *transformer*, baseado em mecanismos de atenção que, grosso

modo, estimam o significado de cada palavra, uma de cada vez, considerando todo seu contexto (Vaswani et al., 2017) – a palavra “laranja”, por exemplo, pode significar uma cor ou uma fruta, dependendo do contexto. Desde então, os modelos e plataformas baseadas em Inteligência Artificial Generativa (IAG) vêm sendo aprimorados, ampliados e diversificados, gerando novas questões em diversos campos da atividade humana. Surgem novas possibilidades, mas também incertezas. Sob a ótica da sociedade de risco, de Ulrich Beck (2010), a IAG ilustra como o avanço tecnológico contemporâneo é o próprio produtor de perigos sistêmicos e invisíveis. Passamos assim a conviver com ameaças fabricadas pela própria modernidade, cujas consequências em escala global desafiam a capacidade humana de prevê-las ou contê-las.

No campo da educação, os resultados são heterogêneos. Revisões sistemáticas sobre o uso de chatbots e sistemas de IA mostram que boa parte dos estudos aponta impacto positivo na aprendizagem, frequentemente superior aos chamados métodos tradicionais (Deng et al., 2025; Dizon & Prudente, 2025; Zhang et al., 2025). Também há evidências de impacto positivo na motivação e engajamento dos alunos (Sasikala & Ravichandran, 2024). Nos países de renda média-baixa, o efeito positivo tende a ser maior, sugerindo potencial da IAG na promoção de equidade (Dizon & Prudente, 2025). Os efeitos do uso de IAG dependem do desenho pedagógico, tendendo a ser maiores com a aprendizagem personalizada e aprendizagem por andaimes (*scaffolded learning*) (Dizon & Prudente, 2025). Contudo, há também diversos estudos experimentais com resultados nulos ou negativos. Em um deles, com cerca de mil estudantes de ensino médio, o acesso irrestrito ao GPT-4 melhorou o desempenho durante as tarefas, mas produziu desempenho inferior ao do grupo de controle quando a ferramenta foi retirada (Bastani et al., 2025). Outro experimento, com 405 estudantes de 14 e 15 anos, concluiu que as anotações manuais superaram o uso de LLM na compreensão e retenção de conhecimento três dias depois (Kreijkes et al., 2026), e resultados similares foram encontrados sobre a profundidade do raciocínio em investigação científica (Stadler et al., 2024). Uma revisão sistemática sobre dependência excessiva de IAG (Zhai et al., 2024) chegou a conclusões inversas às das meta-análises supracitadas. Em outro experimento, com estudantes de ensino superior em física, o grupo que contava apenas com um buscador superou o que usou ChatGPT na resolução de problemas, e 42% dos *prompts* deste último grupo consistiram simplesmente em colar o enunciado do problema (Krupp et al., 2024). Notam-se, com efeito, importantes diferenças entre o desempenho auxiliado pela ferramenta e a aprendizagem que de fato permanece quando ela é retirada. Fesler e colegas (2026) mobilizam o conceito de Zona de Desenvolvimento Proximal (ZDP) (Vygotsky, 1991), sugerindo que IAGs genéricas muitas vezes operam de forma pouco pedagógica, fazendo tudo pelo aluno ou fornecendo explicações muito complexas. Já as IAGs tutoras, que vão atuando gradualmente e sem dar respostas prontas, poderiam servir como andaimes, promovendo aprendizagem especialmente para os alunos mais iniciantes, embora possam limitar os alunos mais proficientes.

Por outro lado, há risco de descarregamento cognitivo (*cognitive offloading*) e pensamento “preguiçoso”: a dependência excessiva de ferramentas como o ChatGPT pode levar ao declínio do esforço mental e da profundidade do pensamento (Georgiou, 2025; Narayan, 2025). Os efeitos da IAG sobre os processos mentais são diversos: a IAG pode aumentar a

motivação e a satisfação, mas também pode gerar ansiedade, isolamento social e ilusão de maestria (Kundu & Bej, 2025; Rivera-Novoa & Duarte Arias, 2025). Há ainda a questão da importância da materialidade na aprendizagem, especialmente em algumas fases da vida e áreas do conhecimento. No Ensino de Ciências, por exemplo, a materialidade e a experiência direta com os fenômenos naturais são fundamentais, não basta interagir com a IA ou outras ferramentas digitais (Tang & Cooper, 2025).

Os efeitos da IAG na aprendizagem são ainda um campo recente de estudo, mas algo que já parece claro é que eles dependem do tipo de interação Humano-IAG (HIAG). Um estudo com 2.307 estudantes do 7º ano de escolas públicas tchecas não encontrou relação significativa entre desempenho acadêmico e frequência de uso de IAG (Lintner et al., 2026). Mas os pesquisadores incluíram um item para verificar a postura epistêmica no uso da IAG: os estudantes tinham que dizer o quanto concordavam com a frase "É mais importante saber fazer as perguntas certas à IA do que entender as coisas por si mesmo". Esse item apresentou relação significativa e consistente com o desempenho acadêmico, mostrando que a aprendizagem mediada pela IAG depende do tipo de postura frente à ferramenta, do tipo de uso que se faz dela. Alunos que confiam na IAG tendem a usá-la como substituto e apresentar piores resultados de aprendizagem. Por outro lado, alunos que usam a IAG como um "andaime pedagógico" (*scaffolding*) para compreensão tendem a apresentar resultados melhores. Outro estudo empírico, com desenho experimental, chegou a conclusões semelhantes (Shen & Tamkin, 2026). Delegar completamente a tarefa para a IA, por exemplo, se mostrou prejudicial para a aprendizagem, o que é confirmado por outro estudo (Georgiou, 2025). Já o uso para investigação conceitual se mostrou associado à maior aprendizagem (Shen & Tamkin, 2026).

Na mesma linha, destaca-se um estudo de revisão de escopo sobre o desenvolvimento do pensamento crítico nas interações com o ChatGPT (Raitskaya & Tikhonova, 2025). O estudo conclui que a IAG pode contribuir para o desenvolvimento do pensamento crítico quando usada como "parceiro de raciocínio", não apenas um fornecedor de respostas. Todos esses estudos se alinham ao quadro teórico descrito por Rivera-Novoa e Duarte Arias (2025). Os autores se fundamentam na tipologia de artefatos de Fasoli, que diferencia três tipos de artefatos: substitutivos (que assumem a tarefa pelo sujeito), complementares (que auxiliam sem substituir o controle do indivíduo) e constitutivos (necessários para a tarefa). Nesse sentido, no contexto pedagógico, o uso substitutivo da IAG tenderia a ser prejudicial, enquanto o uso complementar seria potencialmente benéfico. A Tabela 1 apresenta um resumo desses quatro estudos que relacionam aprendizagem com o tipo de interação HIAG.

Em síntese, a Tabela 1 reflete a importância do esforço e da intencionalidade para a aprendizagem (algo já bastante conhecido entre educadores) no contexto das interações HIAG. Cabe ressaltar que nem todo esforço leva à aprendizagem. De acordo com a Teoria da Carga Cognitiva (Sweller, 2011), o esforço relativo à carga intrínseca ao conteúdo é desejável na educação, diferente daquele decorrente da carga extrínseca. Nota-se também na Tabela 1 que, nas interações potencialmente prejudiciais para a aprendizagem, o aluno usa a IAG como substituto para ele mesmo, delegando totalmente a tarefa para a ferramenta.

Tabela 1. Interações potencialmente prejudiciais ou benéficas para a aprendizagem.

Fonte	Interações HIAG potencialmente prejudiciais para a aprendizagem	Interações HIAG potencialmente benéficas para a aprendizagem
(Rivera-Novoa & Duarte Arias, 2025)	Interação substitutiva: quando o estudante delega a tarefa inteira à IAG. Alunos que confiam demasiadamente na IAG correm o risco de aceitar alucinações e herdar vieses contidos nos dados de treinamento. Isso pode levar a uma ilusão de maestria ou ao "efeito ghostwriter", em que o aluno acredita ter competências que só a máquina executou.	Interação complementar: quando a IAG é usada para processos intermediários, como brainstorming, esclarecimento de dúvidas ou geração de pistas, mas o controle da tarefa permanece com o aluno. É importante que os alunos justifiquem ou critiquem o conteúdo gerado pela IAG. Sistemas como o de "interação socrática" (em que a IAG é configurada para não dar a resposta, mas sim fazer perguntas) promovem o pensamento crítico e o refinamento da argumentação.
(Shen & Tamkin, 2026)	- Delegação total da tarefa à IAG (copiar e colar a tarefa pronta, contendo erros ou não) - Dependência progressiva (começar independente, depois delegar tudo) - Correção de erros também delegada à IAG	- Investigação conceitual (apenas perguntas conceituais) - Delegação parcial informada (pedir para a IAG fazer a tarefa mas também pedir explicações) - Correção de erros de forma autônoma
(Lintner et al., 2026)	Focar mais em fazer as perguntas certas à IAG do que entender as coisas por si mesmo	Focar mais em entender as coisas por si mesmo do que fazer as perguntas certas à IAG.
(Raitskaya & Tikhonova, 2025)	Uso de IAG como fornecedor de respostas. O uso não mediado e o hábito de aceitar informações geradas pela IAG sem escrutínio reduzem o esforço cognitivo do aluno, levando a notas menores a longo prazo, aprendizagem passiva e questões éticas.	Uso de IAG como parceiro de raciocínio. Alunos são orientados a questionar e avaliar as respostas da IAG. Práticas como refinar prompts, identificar vieses da IAG, checar fatos e triangulação de fontes ativam camadas mais profundas de cognição.

Agência é um conceito que pode partir de diferentes perspectivas. Na Teoria Ator-Rede, por exemplo, se considera que objetos, como lombadas de rua, participam de uma agência distribuída (Latour, 2005). De modo geral, a agência está relacionada à intencionalidade, capacidade e responsabilidade, assim como autonomia e a real possibilidade de produzir efeitos. Nesse início da era pós-ChatGPT, é preciso reavaliar o papel social da IA, e a agência humana é fundamental para que a IA possa gerar – e distribuir – mais benefícios do que riscos (Krakowski, 2025).

Para Bandura (2018), ser um agente significa produzir intencionalmente certos efeitos por meio das próprias ações. O autor considera que a agência se manifesta por meio da antecipação, autorreatividade e autorreflexividade. Além disso, ele diferencia três modos de agência: individual, coletiva e “por procuração”. Neste último caso, a pessoa busca influenciar outras pessoas para agir em seu nome. Em Bandura, a agência por procuração refere-se à mobilização de outras pessoas; estamos aqui adaptando o conceito para nosso objeto de estudo. Ao tratar da delegação de agência à IAG, não atribuímos agência à ferramenta; o foco é a redução do exercício da agência humana ao delegar a tarefa. Boa parte da primeira coluna da Tabela 1 se refere à agência por procuração, enquanto a segunda coluna descreve interações com maior agência individual. Essa “perda de agência” no uso de IAG é discutida por Wenker (2023) no artigo “Quem escreveu isso?”, relativo ao contexto empresarial.

No Ensino de Ciências, as pesquisas apontam diferentes usos da IAG, como por exemplo: apoio no desenvolvimento de planos de aula na perspectiva CTS (Ciência, Tecnologia e

Sociedade) e formulação de questões sociocientíficas relevantes (Gunbatar & Sirin, 2025); a avaliação e pontuação automática de respostas dissertativas em ciências (Latif & Zhai, 2024) e o fornecimento de feedback instantâneo e personalizado em tarefas de física (Meyer et al., 2025). Além disso, a IAG permite gerar questões de ciência que incorporam línguas indígenas, conhecimentos locais e contextos culturais específicos de diferentes países, promovendo uma aprendizagem mais efetiva e equitativa (Nyaaba et al., 2024). As possibilidades abertas pela IAG são diversas, assim como os riscos e incertezas. No Brasil, um levantamento com mais de 400 professores de ciências da Educação Básica encontrou certa preocupação e ceticismo quanto à contribuição do ChatGPT para a aprendizagem, receios quanto à avaliação, além de divergências sobre ser ou não considerado infração ou plágio (Monteiro et al., 2024).

Foram encontradas três revisões sistemáticas recentes tratando do uso de IAG no Ensino de Ciências na Educação Básica: uma delas examina as percepções epistêmicas da IAG (Cheung et al., 2025); outra analisa a IAG no ensino STEM (*Science, Technology, Engineering and Mathematics*) na perspectiva do "Florescimento Humano", relacionado ao conceito de agência (Fock & Siller, 2025); a terceira revisão foca na integração da IAG em abordagens de aprendizagem ativa (como Aprendizagem Baseada em Problemas e Ensino por Investigação) (Otto et al., 2025). As revisões confirmam diversos resultados já mencionados, como o potencial da IAG para melhorar o desempenho acadêmico, aprendizagem, motivação e personalização do ensino.

Além disso, Otto e colegas destacam que o efeito benéfico da IA está associado à redução do esforço mental necessário para certas tarefas, permitindo maior foco em competências cognitivas superiores, como pensamento crítico (Otto et al., 2025). Otto e colegas também observaram uma tensão crescente entre o que os alunos conseguem produzir com o auxílio da IAG e o que eles realmente compreendem. Isso pode levar à perda da vontade e da confiança para realizar tarefas de forma autônoma, resultando em dependência excessiva e erosão da agência (Fock & Siller, 2025). Nenhuma dessas revisões, porém, relaciona os tipos de interação relatados em estudos empíricos na Educação Básica à aprendizagem e à agência; a mais próxima, de Otto et al. (2025), classifica as interações pelo papel atribuído à IAG e abrange do ensino médio à pós-graduação. Diante desse quadro, é importante compreender com maior profundidade os tipos de interação HIAG no Ensino de Ciências, seus efeitos sobre a aprendizagem e suas implicações para a agência dos estudantes e professores.

Neste artigo, buscamos contribuir para essa temática apresentando uma revisão sistemática de pesquisas recentes, publicadas desde 2023. Reunimos informações potencialmente úteis não apenas a pesquisadores, mas também a professores, especialmente de ciências. As perguntas de pesquisa são:

- 1) Como alunos e professores interagem com a IAG?
- 2) Quais fatores modulam a qualidade das interações HIAG?
- 3) Como a delegação de agência à IAG se relaciona com os resultados das interações HIAG?
- 4) Como o uso de IAG pode interferir na aprendizagem?
- 5) Quais os riscos e incertezas associados ao uso de IAG no contexto educacional?

Apresentamos inicialmente resultados relativos ao processo de triagem dos estudos, fornecendo uma visão geral da produção recente sobre IAG no Ensino de Ciências. A partir disso, focamos a análise no corpus de estudos selecionados, começando com um mapeamento mais geral dos estudos e das interações HIAG, e terminando no recorte mais específico, relativo à aprendizagem e agência.

Métodos

Esta revisão sistemática foi conduzida de acordo com as diretrizes do PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), referencial amplamente adotado para garantir transparência, reprodutibilidade e rigor metodológico (Page et al., 2021). Por se tratar de uma revisão de natureza qualitativa, adotou-se o PICO (*population, phenomena of interest, context*) em substituição ao PICO convencional (Hosseini et al., 2024). Os elementos foram definidos da seguinte forma:

- a) **P** (população): estudantes e professores da Educação Básica;
- b) **I** (fenômeno de interesse): uso educacional da inteligência artificial generativa;
- c) **Co** (contexto): Ensino de Ciências.

Estratégias de busca

As buscas foram realizadas no dia 18 de fevereiro de 2026, em três fontes: a Web of Science (em inglês), o Portal de Periódicos da Capes e o SciELO (em português). Foram aplicados filtros de período (2023 a janeiro de 2026), tipo de documento (artigos) e idioma (inglês na Web of Science e português nas demais). As strings de busca foram elaboradas a partir dos elementos do PICO. Para a Web of Science, utilizou-se:

("generative AI" OR "generative artificial intelligence" OR "ChatGPT" OR "Large Language Models" OR "LLM" OR "Claude AI" OR "Anthropic Claude" OR "Google Gemini" OR "Gemini AI" OR "Google Bard" OR "Microsoft Copilot" OR "Bing Chat") AND ("science education" OR "science teaching" OR "science learning" OR "STEM education" OR "physics education" OR "chemistry education" OR "biology education") AND ("high school" OR "secondary education" OR "secondary school" OR "middle school" OR "primary education" OR "K-12" OR "K12" OR "basic education")

Para as buscas em português, a tradução direta da string resultou em zero registros, o que motivou a adoção de uma string simplificada:

("Inteligência artificial" OR "IA generativa" OR "ChatGPT" OR "LLM" OR "Grandes Modelos de Linguagem") AND ("Ensino de Ciências" OR "Educação em Ciências" OR "Ensino de Física" OR "Ensino de Química" OR "Ensino de Biologia")

O Portal de Periódicos da Capes não é propriamente uma base de dados, mas uma interface integrada. A busca foi realizada na opção “Buscar assunto”, com filtros de tipo de documento (artigo), revisão por pares, período e idioma (português), recuperando 14 registros. O SciELO foi consultado na interface da própria base, com filtros semelhantes, recuperando 5 registros.

Na Web of Science, as strings de busca foram aplicadas no campo *Topic* (TS), que abrange título, resumo e palavras-chave, recuperando 47 registros. A opção por uma única fonte internacional decorre da cobertura e do controle editorial da Web of Science, embora restrinja o alcance da revisão.

Processo de triagem

A triagem foi conduzida em duas fases, conforme representado no diagrama PRISMA (Figura 1).

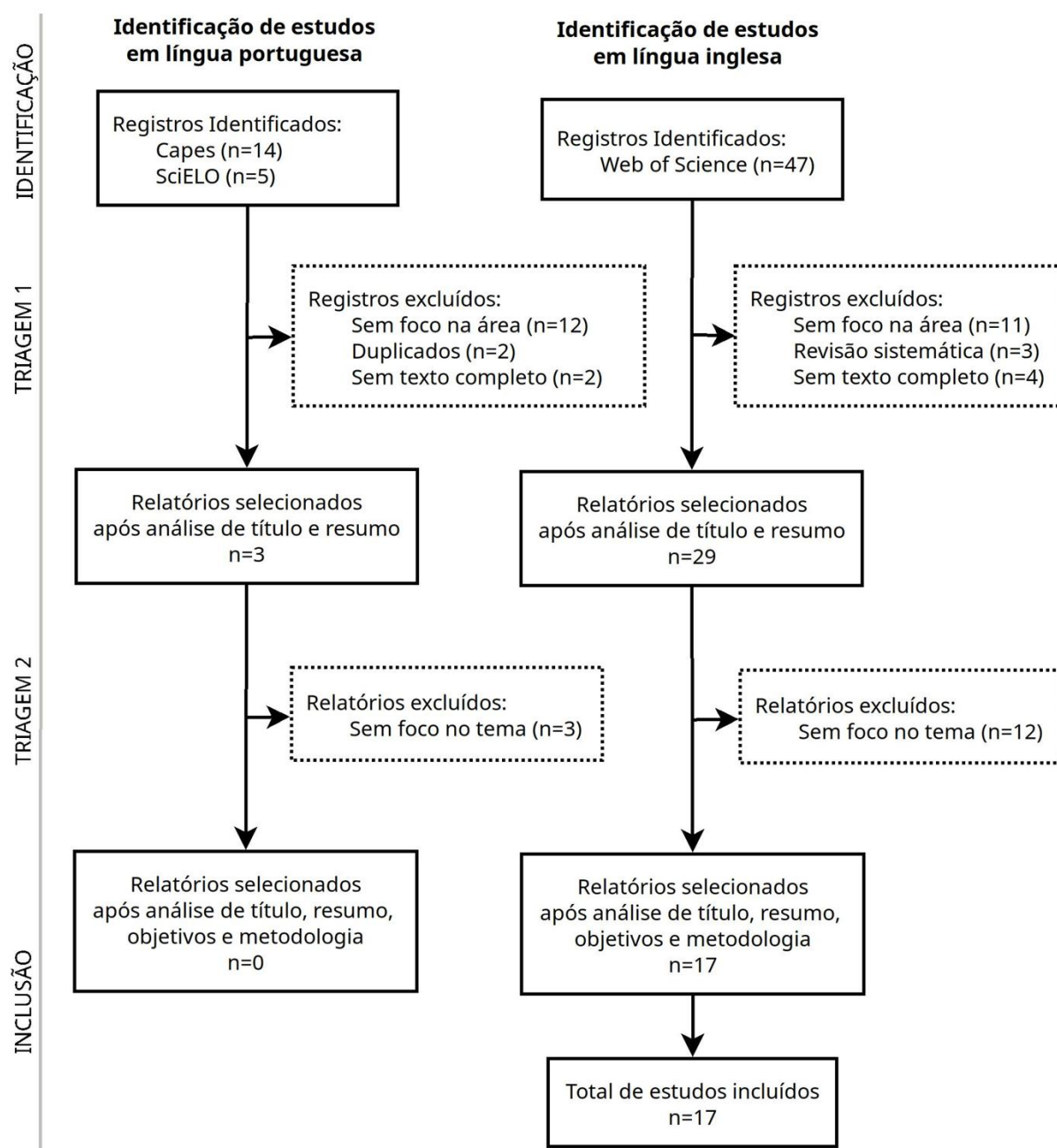


Figura 1. Diagrama PRISMA

A Triagem 1 consistiu na leitura de títulos e resumos, aplicando os seguintes critérios:

- a) C1 (Exclusão): não é estudo empírico que aborda a IAG no contexto do Ensino de Ciências na Educação Básica;
- b) C2 (Exclusão): é revisão sistemática;
- c) C3 (Exclusão): registro sem disponibilidade de acesso ao texto completo.

Cabe notar que C3 corresponde, na nomenclatura do PRISMA 2020, aos registros não recuperados: foi aplicado quando o texto completo não estava acessível para a leitura da fase seguinte.

Na Triagem 1 foram excluídos 34 registros: 2 por duplicidade, 23 por C1, 3 por C2 e 6 por C3. O C1 é um critério composto, já que delimita o objeto da revisão: um mesmo registro pode deixar de satisfazê-lo por mais de uma razão. A justificativa de cada exclusão está presente no Apêndice.

A Triagem 2 envolveu a leitura dos objetivos e metodologia dos estudos remanescentes, com o critério de selecionar apenas aqueles com dados primários sobre interações de alunos ou professores com IAG no contexto do Ensino de Ciências. Um estudo cujos pesquisadores atuavam como professores (testando estratégias de *prompt* para o planejamento de atividades) foi incluído com base nesse enquadramento.

Síntese dos dados

Ao final do processo, foram incluídos 17 estudos. A síntese foi realizada em duas etapas: (1) categorização dos tipos de interação HIAG relatados nos estudos; e (2) análise focada nos estudos com resultados empíricos relacionados à aprendizagem em ciências e à agência dos sujeitos. Os 15 estudos excluídos na segunda fase de triagem foram igualmente caracterizados, agrupados em três temáticas, com o objetivo de ampliar a transparência da revisão e identificar temas para pesquisas futuras.

A unidade de análise é o estudo, e não o episódio de interação, já que a maior parte dos artigos não disponibiliza os registros brutos. As categorias da Tabela 4 não foram definidas previamente: emergiram do corpus em leituras sucessivas, sendo que um mesmo estudo pode pertencer a mais de uma categoria.

Buscamos reduzir os vieses decorrentes de a revisão ter sido conduzida por um único pesquisador. Nenhum dos procedimentos aqui adotados substitui a triagem independente por dois revisores; buscamos apenas trabalhar da forma mais transparente e verificável possível. Os critérios de inclusão e exclusão foram escritos antes do início da triagem e não foram alterados. Os dados extraídos de cada estudo são apresentados integralmente na Tabela 3 e nas Tabelas A3 e A4 do Apêndice, permitindo confrontar cada classificação com o estudo original.

Neste trabalho, foram utilizadas ferramentas de IAG entre fevereiro e maio de 2026 (Claude, Gemini, Kimi e Notebook LM, nas versões disponíveis gratuitamente no período). Mais especificamente, as ferramentas foram utilizadas para: definição das palavras-chave de

busca, categorização temática de estudos, identificação de padrões recorrentes, tradução assistida, busca e compilação de fontes, correção gramatical e ortográfica, além de melhoria da clareza do texto. Na categorização temática e identificação de padrões, foi realizada inicialmente uma análise exploratória com o Notebook LM, tanto de forma mais “passiva” (solicitando que a ferramenta encontrasse padrões recorrentes) quanto de forma mais “ativa” (buscando padrões específicos que emergiram da leitura dos artigos). Os padrões mais frequentes foram então definidos e descritos, e a partir disso foi feita análise confirmatória, tanto na ferramenta quanto nos próprios artigos, um estudo por vez. Nenhuma classificação foi incorporada sem conferência no próprio artigo, e as divergências foram resolvidas em favor da leitura do pesquisador, de modo que a ferramenta funcionou como uma segunda opinião. A redação dos argumentos, seleção final dos estudos e interpretação dos resultados são de responsabilidade do pesquisador. Para o gerenciamento das referências e citações, foi utilizado o software livre Zotero.

Os 17 estudos apresentam desenhos heterogêneos, sendo importante ponderar a força das evidências que cada um sustenta. Para isso, construímos critérios explícitos de robustez, aplicados a todo o corpus e apresentados na Tabela A4 do Apêndice: desenho e presença de comparação, presença de pré-teste, natureza do desfecho, tamanho da amostra e duração da intervenção.

Resultados

Para esta revisão sistemática, fizemos buscas em duas fontes em português (SciELO e Portal de Periódicos da Capes) e uma fonte internacional (Web of Science). Os 66 registros resultantes dessas buscas passaram por duas fases de triagem, restando 17 estudos, todos internacionais.

Em uma perspectiva estritamente procedimental, tais etapas seriam suficientes para uma revisão sistemática. Mas buscamos caracterizar também o que foi excluído do estudo, devido ao foco nas interações HIAG. Com isso, conferimos maior transparência à revisão, proporcionamos uma visão geral dos estudos, e também apontamos lacunas para futuras revisões. Observando os 15 estudos não incluídos na análise, notamos que podem ser agrupados em três grandes temáticas: 6 deles focam na percepção docente; 5 focam no desenvolvimento de ferramentas; os 4 restantes fazem análises técnicas das capacidades da IAG¹.

A Tabela 2 descreve essas três temáticas, assim como a temática principal, proporcionando uma visão geral dos 32 estudos resultantes da primeira fase de triagem. Destes, foram selecionados os 17 estudos atribuídos à categoria 1 durante a segunda fase de triagem.

Uma observação importante diz respeito à categoria Desenvolvimento de Ferramentas: a maioria dos estudos dessa categoria utiliza dados de professores e alunos para o desenvolvimento e aprimoramento de sistemas de IAG. Isso tem uma dimensão positiva, pois

¹ Há sobreposições de categorias, nesses casos foi priorizada a categoria 1, em virtude dos objetivos deste trabalho. A classificação de todos os registros pode ser conferida no Apêndice, Tabelas A1 e A2.

indica que tais ferramentas se baseiam na realidade educacional. Contudo, há uma dimensão problemática: dados produzidos por humanos são coletados, frequentemente sem transparência nos termos de uso, e sem qualquer contrapartida para os participantes – que, em alguns casos, podem estar contribuindo involuntariamente para o desenvolvimento de tecnologias que podem vir a substituí-los, mesmo que parcialmente, na atividade profissional.

Tabela 2. Temáticas dos estudos após a primeira fase de triagem.

Temática	Descrição
1. Interações *	Estudos que investigam empiricamente interações de estudantes e/ou professores com ferramentas de IAG no contexto do Ensino de Ciências
2. Percepção docente	Estudos que capturam o que professores de ciências pensam, sentem e acreditam sobre IAG no contexto educacional, tipicamente através de questionários, entrevistas ou grupos focais
3. Desenvolvimento de ferramentas	Estudos focados na criação, aprimoramento, ajuste fino ou validação de produtos, sistemas e aplicações baseados em IAG no contexto do Ensino de Ciências
4. Avaliação técnica da IAG	Estudos que testam e avaliam as capacidades de ferramentas de IAG em tarefas de ciências, sem relação direta com contextos de ensino e aprendizagem

* Categoria dos estudos selecionados nesta revisão.

Visão geral dos estudos selecionados

A maioria dos 17 estudos trata do ensino médio, alguns se referem aos anos finais do ensino fundamental, e apenas um aos anos iniciais. Em relação às áreas científicas, predominam estudos de física e STEM, incluindo também estudos focados em química e apenas um em biologia. A ferramenta de IAG mais utilizada foi o ChatGPT, havendo também usos de Gemini, Doubao (ByteDance), GLM-4, Bing e Google Bard; dois estudos não identificam a ferramenta (Tabela A3). Treze dos 17 artigos foram publicados em 2025. A maioria foca no uso de IAG pelos alunos, enquanto alguns focam nos professores. Geograficamente, predominam os países asiáticos, havendo também contribuições da Europa, Turquia e Estados Unidos. A Tabela 3 fornece uma descrição básica de cada artigo, destacando as conclusões mais relevantes para nosso foco de estudo: as interações HIAG no Ensino de Ciências. A rastreabilidade tecnológica e os critérios de robustez de cada estudo são detalhados nas Tabelas A3 e A4 do Apêndice.

Tabela 3. Características dos estudos selecionados.

Nº	Referência	Objetivo	Método	Conclusões relevantes
01	(Chen & Chang, 2024)	Investigar a eficácia da aprendizagem baseada em jogos auxiliada por IA nos resultados de ciências, motivação intrínseca, carga cognitiva.	Estudo quase-experimental com 202 alunos do 7º ano, divididos em 3 grupos.	A integração da IAG (especialmente com exemplos) melhorou significativamente os resultados de aprendizagem e a competência percebida, reduziu a carga cognitiva e promoveu estratégias mais sistemáticas para resolver problemas científicos.

Nº	Referência	Objetivo	Método	Conclusões relevantes
02	(Huang et al., 2026)	Propor e validar o modelo de ensino IT-CSTP para o cultivo do pensamento científico no ensino de física do ensino médio.	Pesquisa quase-experimental com 80 estudantes. Coleta de dados de raciocínio científico, questionamento, criatividade e entrevistas com professores.	O uso da IAG como “parceiro de interação intelectual” foi eficaz no cultivo do pensamento científico, auxiliando na visualização de processos de investigação e reduzindo a carga cognitiva no design de projetos.
03	(Cho et al., 2025)	Explorar o processo de interação entre professores e IAG em ambiente de simulação computacional.	Estudo exploratório com 13 professores de física. Análise de <i>logs</i> de interação, questionários e grupos focais semiestruturados. Usa Taxonomia de Bloom.	Professores realizaram predominantemente interações de baixo nível cognitivo (lembrar, entender). Sequências de acompanhamento profundas foram raras, indicando necessidade de maior letramento em IA.
04	(Chen & Law, 2025)	Explorar os perfis de busca de ajuda dos alunos ao interagir com ChatGPT em ambiente de jogos digitais.	Estudo com 102 estudantes do 7º ano. Questionário de perfis de busca de ajuda, métricas de desempenho no jogo e <i>logs</i> de comportamento na IA.	Busca de ajuda instrumental (pedir dicas, não respostas) teve efeito positivo no desempenho e engajamento. Alunos sentiram-se mais confortáveis buscando ajuda da IA do que de humanos.
05	(Valeri et al., 2025)	Descrever experiências, percepções e atitudes sobre o uso do ChatGPT em disciplinas STEM.	Métodos mistos: questionário com 453 estudantes do ensino médio + 12 entrevistas semiestruturadas.	A ferramenta é amplamente usada para entender conceitos difíceis. Alunos aplicam estratégias de <i>prompting</i> para evitar alucinações e vieses; uso eficaz requer conhecimento prévio do conteúdo.
06	(Wang et al., 2025)	Otimizar o ensino de química em escolas secundárias chinesas utilizando ferramentas locais de LLM.	Amostragem aleatória de 286 alunos e 43 professores. Análise de conteúdo (codificação de questionários) após 44 horas de curso.	Ferramentas de LLM atenderam efetivamente às necessidades de aprendizagem personalizada. Professores relataram redução na carga de trabalho na preparação de aulas e na coleta de recursos.
07	(Gunbatar & Sirin, 2025)	Investigar diferentes técnicas de prompting para gerar planos de aula sobre questões sociocientíficas.	Análise documental (com 19 indicadores binários) de 14 planos gerados por GPT-3.5 e GPT-5 usando prompts variados.	Qualidade pedagógica é altamente sensível à complexidade do prompt; Cadeia de Pensamento produziu planos melhores. Dimensões éticas e conexões locais ainda são fracamente representadas pela IA.
08	(Gunbatar et al., 2025)	Examinar perfis de interação dos professores com IAG para o planejamento de aulas de química.	Estudo multicaso com 6 participantes. Dados: planos de aula, históricos de diálogo, artigos de reflexão e grupo focal.	Quatro perfis identificados: dependentes de IA, colaborativos, refinadores de planos e planejadores independentes. Experiência docente e Conhecimento Pedagógico do Conteúdo influenciaram a capacidade de avaliar criticamente as ofertas da IA.
09	(Kilde-Westberg et al., 2025)	Investigar como estudantes se engajam ativamente com ChatGPT para resolver problemas durante laboratório de física.	Estudo qualitativo com 19 alunos de ensino médio. Gravações de vídeo, transcrições, históricos de chat e entrevistas pré/pós sob a Teoria da Variação.	Alunos usaram a IA como “parceiro de laboratório” para explicações conceituais e ajuda técnica. Importância do conhecimento prévio sobre o assunto para se dispor a questionar as respostas da IAG. Importância da mediação docente e do letramento em IA para evitar “confiança cega” e antropomorfização da ferramenta.

Nº	Referência	Objetivo	Método	Conclusões relevantes
10	(Tong et al., 2025)	Explorar a colaboração Humano-IAG comparada à colaboração Humano-Humano.	Foco na resolução de problemas científicos. Amostra de 115 alunos de ensino médio. Análise de precisão em testes de conhecimento, pensamento científico e logs de diálogo.	Colaboração humana foi significativamente mais eficaz do que com IAG. ChatGPT comportou-se como aluno autoconfiante, porém inconstante, induzindo erros por falhas lógicas e de reconhecimento de imagem.
11	(Guo et al., 2025)	Investigar se a integração de LLMs em Práticas STEM promove progresso de alunos do ensino médio.	Estudo quase-experimental com 70 alunos do 8º ano. Dados: testes de conhecimento físico, pensamento computacional, resolução de problemas e entrevistas focais.	A integração da IAG teve impacto notavelmente superior no domínio de conhecimento científico, atitudes em relação à ciência, percepção de utilidade da tecnologia, habilidades de pensamento computacional e capacidades de resolução de problemas.
12	(Naumann et al., 2025)	Investigar atitudes, motivação e autoeficácia de estudantes em relação à IA no ensino de química.	Questionários pré e pós-intervenção com 46 estudantes de 9º e 10º anos (44 no pré-teste, 43 no pós-teste). Perguntas abertas de reflexão no pós-teste coletaram aspectos positivos e sugestões de melhoria.	Alunos se sentem confiantes em realizar as tarefas propostas, com apoio da IAG. Estudantes valorizam tanto o apoio da IAG quanto o do professor. Considerações éticas foram frequentes, mas a compreensão técnica da IAG se mostrou limitada.
13	(Lai, 2025)	Explorar o papel das estratégias de busca nas sequências de <i>prompt</i> dos alunos com IAG.	Amostra: 46 alunos de ensino médio. Dados: questionários de tendências, relatórios científicos e <i>prompts</i> dos alunos.	Alunos com altas estratégias de busca ("aprendizes holistas") desenvolveram modelos de conhecimento estruturados buscando inconsistências. Alunos com baixas estratégias ("atomistas") focaram em fatos isolados.
14	(Lee & So, 2025)	Investigar como os alunos usam a IAG para a Resolução Criativa de Problemas em domínios sociais e científicos.	Amostra: 38 estudantes. Análise quantitativa de conteúdo de <i>prompts</i> , análise de rede e padrões sequenciais baseada na rubrica do PISA 2022.	Grupos de alto desempenho utilizaram gama mais ampla de tipos de <i>prompt</i> e fizeram ciclos de refinamento das saídas da IAG. Grupos de baixo desempenho aceitaram passivamente as sugestões.
15	(Bitzenbauer, 2023)	Investigar atividades com IAG para promover o pensamento crítico no ensino de física quântica.	Estudo piloto com 53 alunos do 12º ano. Dados: questionário pré e pós-intervenção para medir percepções.	A intervenção teve impacto positivo nas percepções sobre os benefícios da IAG. Inesperadamente, a análise crítica das saídas da IAG levou a atitudes mais positivas em relação à IAG, não mais negativas.
16	(Potter et al., 2026)	Adaptação, com uso de IAG, de framework para apoiar a leitura e escrita científica.	Design participativo com métodos mistos. Dados: transcrições de grupos focais com 16 professores, atividades dos alunos.	Após o processo de adaptação do framework e aplicação com alunos, os professores valorizaram a IAG como apoio ao planejamento, mas se preocupam com acurácia e ética.
17	(Gunes & Yilmazlar, 2025)	Examinar o efeito de intervenção integrada à IAG no desempenho acadêmico dos alunos.	Estudo quase-experimental com 37 alunos do 5º ano. Dados: testes de desempenho acadêmico pré e pós intervenção sobre tópicos da física.	O grupo que passou pela intervenção obteve desempenho superior. No planejamento docente, a integração da IAG ajudou na criação de narrativas motivadoras. As respostas instantâneas da IAG ajudaram a guiar a investigação dos alunos.

O conjunto de 17 estudos aponta algumas tendências gerais e informações relevantes para a educação na era pós-ChatGPT. O uso da IAG como ferramenta pedagógica tem demonstrado impacto positivo no desempenho acadêmico e na compreensão conceitual dos estudantes (Chen & Chang, 2024; Gunes & Yilmazlar, 2025), auxiliando no desenvolvimento do pensamento crítico e científico (Guo et al., 2025; Huang et al., 2026). A IAG pode contribuir para maior motivação intrínseca, engajamento e percepções positivas sobre a aprendizagem (Bitzenbauer, 2023; Chen & Chang, 2024; Chen & Law, 2025) – o que pode estar relacionado ao fato de que ela atua na redução da carga cognitiva (Chen & Chang, 2024; Huang et al., 2026) e que também pode ajudar na adaptação de conteúdos (Potter et al., 2026). Contudo, a qualidade da interação com a IAG depende de capacidades prévias do usuário: conhecimento do conteúdo (Kilde-Westberg et al., 2025; Valeri et al., 2025), letramento em IA (Cho et al., 2025; Kilde-Westberg et al., 2025), sofisticação de *prompts* (Gunbatar & Sirin, 2025; Lee & So, 2025), estratégias de busca (Lai, 2025) e experiência docente (Gunbatar et al., 2025). Essas capacidades prévias acabam influenciando as formas de interação: mais superficiais e passivas, ou mais profundas e críticas.

Os estudos também reúnem evidências de que os alunos se sentem mais confortáveis buscando ajuda de IAG que de humanos (Chen & Law, 2025), mas podem antropomorfizar e depositar uma "confiança cega" na ferramenta (Kilde-Westberg et al., 2025). Professores relataram que a IAG contribuiu para a redução na carga de trabalho na preparação de aulas e na coleta de recursos (Wang et al., 2025), mas têm preocupações éticas e de acurácia sobre esse tipo de apoio (Potter et al., 2026). Além disso, permanece a importância das relações humanas dos alunos entre si (Tong et al., 2025) e com professores (Naumann et al., 2025). Em termos de abordagens pedagógicas, os estudos focam especialmente na Aprendizagem Baseada em Problemas (Kilde-Westberg et al., 2025; Lai, 2025; Lee & So, 2025) e no Ensino por Investigação (Cho et al., 2025; Huang et al., 2026), incluindo também abordagem CTS (Gunbatar & Sirin, 2025), Concepções Alternativas (Gunbatar et al., 2025) e Aprendizagem Baseada em Contexto (Gunes & Yilmazlar, 2025).

O papel do professor na era pós-ChatGPT é um tema recorrente no corpus analisado. Destaca-se que as ferramentas de IA não substituem os professores, mas servem como apoio ao ensino (Cho et al., 2025; Gunbatar et al., 2025; Naumann et al., 2025). Nesse novo contexto, o papel do professor muda de transmissor de conhecimento para mediador, facilitador, guia da aprendizagem (Huang et al., 2026; Wang et al., 2025) – uma afirmação, cabe observar, muito anterior à IAG. O conhecimento de conteúdo do professor permanece crucial para avaliar criticamente a acurácia e a qualidade das informações geradas pela IA (Valeri et al., 2025). A articulação desse conhecimento com conhecimentos pedagógicos é especialmente importante em contextos de ensino sociocientíficos e complexos (Gunbatar & Sirin, 2025). O professor “desempenha um papel crucial na identificação das necessidades e capacidades dos alunos para compreender o potencial e as limitações da IA generativa” (Kilde-Westberg et al., 2025, p. 020119-1).

Observando o conjunto de estudos, há predominância de resultados positivos da IAG na educação. Em parte, isso pode decorrer de viés de publicação. Há também um segundo mecanismo, que o próprio corpus documenta: os estudos foram em sua maioria desenhados para investigar práticas consideradas promissoras, e dificilmente captam o uso espontâneo e

desassistido predominante nas escolas – um viés de desenho. Notamos, assim, no recorte desta revisão, certa lacuna na pesquisa sobre efeitos indesejáveis da IAG no contexto do Ensino de Ciências.

Análise das interações com a IAG

A fim de iniciar com uma análise descritiva mais geral, identificamos alguns tipos de interação bastante frequentes no corpus analisado. A Tabela 4 mostra os tipos de interação e os artigos em que foram encontrados, seja como relato sobre uma pessoa, seja como conceito ou análise empírica de várias pessoas. Como em qualquer categorização qualitativa, os tipos de interação HIAG descritos na Tabela 4 não constituem uma taxonomia rígida ou definitiva, tampouco pretendem esgotar as possibilidades de interação.

Tabela 4. Tipos de interação HIAG encontrados com frequência e respectivos artigos.

Tipo de interação	Descrição	Artigos
Busca simples de informações e explicações	Usuários consultam a IAG para obter explicações de conceitos, definições e fenômenos científicos, visando facilitar a compreensão imediata. As consultas são simples, não há refinamento de prompt, nem diálogo iterativo.	1,3,5,9,13, 14,16,17
Interação dialógica e refinamento	Conversas estendidas em múltiplos turnos em que usuário e IAG refinam conclusões conjuntamente através do diálogo.	2,5,8,9,10, 13,14
Avaliação crítica da IA	Usuários analisam criticamente respostas da IAG, identificando inconsistências, imprecisões ou erros.	3,5,8,9,10, 13,15,16
Geração de ideias e soluções criativas	Uso da IAG para apoiar processos criativos, gerando alternativas, soluções inovadoras ou ideias para projetos abertos.	2,8,11,14, 16,17
Adaptação do conteúdo ao contexto local e interesses dos alunos	Professor usa IAG para adaptar conteúdo, ritmo, formato ou linguagem conforme interesses, motivações, nível de aprendizagem e características gerais dos seus alunos.	2,7,8,16, 17
Feedback instantâneo e personalizado	A IAG fornece aos alunos avaliações ou respostas personalizadas para atividades educacionais que tenham feito ou estejam fazendo.	1,2,4,10
Suporte a atividades práticas e simulações	Alunos usam a IAG durante experimentos de laboratório, atividades práticas de construção ou simulações.	2, 9,11, 17

Na Tabela 4 observa-se uma clara diferença entre os dois primeiros tipos, sendo o primeiro relativo a pedidos simples e o segundo relativo a interações dialógicas com a IAG. De modo geral, nos estudos selecionados, o primeiro tipo de interação é avaliado de forma mais negativa pelos autores dos estudos, enquanto os demais costumam ser associados a melhorias de desempenho ou aprendizagem. Essa associação, contudo, deve ser interpretada com cautela: ela pode refletir os vieses já mencionados e o recorte dos estudos, que tendem a investigar práticas consideradas promissoras e a enfatizar menos as interações problemáticas e seus riscos.

Professores e IAG

Gunbatar et al. (2025) investigaram os perfis de interação entre professores de química e IAG. O estudo identificou quatro perfis básicos de planejamento docente:

1. **Planejador dependente de IA:** prepara seus planos de aula baseando-se majoritariamente no ChatGPT. Busca ideias sobre concepções alternativas, dificuldades de aprendizagem e estratégias de ensino, atribuindo à IA o papel de "autoridade de conhecimento".
2. **Planejador colaborativo com IA:** se envolve em diálogos, focados no aprimoramento das ideias do professor, uma troca de informações com um colega para complementar pontos omissos, com o professor filtrando as sugestões da IA através do Conhecimento Pedagógico de Conteúdo que adquiriu com os anos de experiência.
3. **Refinador de plano assistido por IA:** pede para a IA revisar um plano de aula já existente e fornecer sugestões e críticas. A interação ocorre em um contexto bem definido e a IA assume o papel de "crítico" ou papel complementar.
4. **Planejador independente de IA:** usa IA apenas para pontos específicos e limitados, como verificar conhecimentos sobre concepções alternativas de um nível escolar diferente ou criar recursos visuais para economizar tempo. Costuma questionar profundamente as informações apresentadas e solicita referências acadêmicas para confirmar as respostas da IA.

Segundo o estudo, esses perfis variam conforme a experiência docente: professores menos experientes tendem a ser mais dependentes e menos críticos, enquanto os mais experientes usam a IA de forma mais específica, crítica e independente. Cabe ressaltar que a amostra é de apenas seis professores, mas o estudo utiliza métodos qualitativos robustos, como concordância de juízes e triangulação de dados.

O estudo de Gunbatar e Sirin (2025) analisa diferentes estratégias de *prompt* no apoio ao planejamento docente, mais especificamente na elaboração de planos de atividade sobre questões sociocientíficas no ensino de química. Os *prompts* foram feitos pelos pesquisadores² e inseridos no ChatGPT. Os planos resultantes foram avaliados por professores de química. O estudo analisou sete estratégias de *prompt*:

1. **Comando simples:** solicita algo pontual para a IAG sem fornecer informações adicionais ou exemplos.
2. **Atribuição de papel:** solicita que a IAG assuma um papel específico, como o de um professor de química do ensino médio.
3. **Exemplos (*few-shot*):** fornece alguns exemplos de artigos com atividades SSI (*Socio-Scientific Issues*) para que a IAG gere um novo plano similar aos exemplos.

² Nesse caso, consideramos que os pesquisadores atuaram como professores, o que justifica a inclusão desse estudo, tendo em vista o foco desta revisão sistemática.

4. **Cadeia de Pensamento:** fornece comandos passo a passo para obter uma saída mais rica, coerente e detalhada.
5. **Geração baseada em busca (RAG):** fornece fontes de dados relevantes e confiáveis para superar as limitações de conhecimento da IAG.
6. **Formatação de saída baseada em PCK (Pedagogical Content Knowledge):** solicita que o plano inclua componentes do Conhecimento Pedagógico de Conteúdo, como avaliação, estratégias instrucionais e currículo.
7. **Formatação de saída baseada no ciclo 5E:** solicita que a atividade seja organizada de acordo com as cinco fases do ciclo de aprendizagem 5E: engajamento, exploração, explicação, elaboração e avaliação.

Essa tipologia retrata, de certa forma, o primeiro estágio da IAG, que inaugura a era pós-ChatGPT em 30 de novembro de 2022: o modelo GPT-3.5. Modelos mais recentes têm internalizado algumas dessas e outras estratégias, o que vem transformando a forma de interagir com a IAG. Gunbatar e Sirin (2025) utilizaram principalmente o GPT-3.5, mas parte dos *prompts* foi repetida com o GPT-5. Comparando ambos, os autores concluíram que: comandos simples geram planos piores, e todos os planos apresentam lacunas, em ambos os modelos; o GPT-5 produz planos melhores do que o GPT-3.5 e também se beneficia de certos *prompts* estruturados, como Cadeia de Pensamento e Exemplos. O formato de saída com base no ciclo 5E apresentou resultados inconsistentes. De forma geral, o estudo reúne evidências de que professores familiarizados com as técnicas de *prompt* e com a abordagem CTS (questões sociocientíficas) produzirão resultados pedagogicamente consistentes.

Outro estudo analisou as interações de professores de física com IAG, focando nas perguntas que eles faziam durante tarefas de aprendizagem (Cho et al., 2025). As perguntas foram categorizadas em três tipos: relacionadas ao conhecimento; relacionadas à tarefa; sobre questões técnicas. Partindo da taxonomia de Bloom, o estudo encontrou predominância de questões relacionadas a “lembrar” e “entender”, que são os níveis cognitivos mais baixos da taxonomia. Além disso, a maioria das interações foi de pergunta única, sem interações posteriores.

Estudantes e IAG

Em relação às interações de alunos com IAG, um dos estudos selecionados focou nas estratégias de busca de informações online (OISS, *Online Information-Searching Strategies*), durante um curso de ciências (Lai, 2025). Foi utilizado um questionário em escala Likert para mensurar as estratégias OISS, a partir do qual foram identificados dois perfis de interação:

1. **Aprendizes holistas:** alunos com OISS maior, que “desenvolveram gradualmente seus modelos de conhecimento de conceitos científicos, organizando informações e encontrando conexões ou inconsistências entre diferentes tipos de conhecimento” (Lai, 2025, p. 241);

2. **Aprendizes atomistas:** alunos com OISS menor, que “se concentraram mais em examinar minuciosamente informações isoladas, buscando exemplos e discussões que solidificassem o discurso existente” (Lai, 2025, p. 254).

A aprendizagem foi aferida a partir dos relatórios científicos iniciais e finais, avaliados com base em rubricas. Estudantes com maiores habilidades de busca (OISS maior) superaram significativamente os de menor habilidade, em termos de acurácia do conteúdo e descrições lógicas em seus relatórios finais. Entre os estudos que relacionam tipo de interação e aprendizagem, este pode ser considerado o mais robusto no desenho metodológico: seis semanas de duração, relatório inicial como linha de base e avaliação por rubrica, com fidedignidade de 0,86 entre dois avaliadores independentes.

A diferença significativa entre os dois perfis não estava no tipo de pergunta: os 11 tipos de *prompt* foram semelhantes em ambos. O que diferiu foi a sequência de *prompts*, o encadeamento que o aluno impõe ao diálogo, segundo análise de redes (Lai, 2025).

Os tipos de *prompt* também foram investigados em outro estudo, no contexto STEM, em que alunos de 15 anos resolveram tarefas de Pensamento Criativo do PISA (*Programme for International Student Assessment*), interagindo com o ChatGPT (Lee & So, 2025). Os pesquisadores analisaram os *prompts* dos alunos, que foram categorizados em 14 tipos. Além disso, avaliaram as soluções das tarefas a partir das rubricas do PISA. O estudo conclui que os alunos com melhor desempenho na tarefa utilizaram maior diversidade de tipos de *prompts*, combinados estrategicamente com suas próprias ideias. Já aqueles com pior desempenho na tarefa “dependeram muito das sugestões do ChatGPT, com pouca ou nenhuma modificação ou originalidade”. Com efeito, a análise das interações com a IAG, comparando alunos com alto e baixo desempenho em Pensamento Criativo, levou os autores a concluir que a agência do aprendiz (*learner agency*) é um fator importante que diferencia os dois grupos, de forma que o exercício da agência está relacionado ao melhor desempenho (Lee & So, 2025, p. 254).

Um estudo com desenho quase-experimental explorou o papel da colaboração humano-IA na resolução de problemas científicos, comparando sua eficácia com a colaboração humano-humano (Tong et al., 2025). Estudantes de ensino médio fizeram tarefas respondendo no papel, com caneta preta. Depois, cerca de metade deles interagiu com outro colega, enquanto a outra metade interagiu com o ChatGPT. Por fim, cada estudante corrigiu suas respostas com caneta vermelha. Em ambos os grupos a interação melhorou o desempenho final nas tarefas, mas a interação humano-humano apresentou efeito sete vezes maior do que a humano-IA³. Os grupos não eram equivalentes no pré-teste de pensamento científico, o que recomenda cautela na comparação. Além disso, o estudo destaca que os alunos tendiam a usar o ChatGPT apenas para obter respostas, envolvendo-se pouco em explorações colaborativas mais aprofundadas.

Outro estudo analisou a interação de alunos com IAG em um jogo sobre física: um RPG de aventura focado em investigação científica sobre o movimento de objetos (Chen & Law,

³ Na colaboração humano-humano o tamanho do efeito foi $d = 0,378$, enquanto na colaboração humano-IA foi $d = 0,053$.

2025). Durante o jogo, havia sempre um botão para pedir ajuda ao ChatGPT. Depois do jogo, os alunos responderam um teste de aprendizagem sobre o tema. A pontuação no jogo foi negativamente correlacionada com o número de perguntas feitas ao ChatGPT ($r = -0,23$), sugerindo que jogadores mais proficientes dependiam menos da ferramenta externa. Os autores encontraram correlações significativas entre os perfis de interação e o desempenho no jogo, mas esses perfis não se correlacionaram com os resultados no teste de aprendizagem.

Também se destaca o estudo que analisou o papel da ferramenta como parceiro de laboratório, durante uma atividade investigativa sobre levitação acústica e velocidade do som (Kilde-Westberg et al., 2025). O roteiro foi intencionalmente desenhado para ser vago, forçando a necessidade de consultar a IAG. O estudo concluiu que a capacidade dos alunos avaliarem criticamente as respostas da IAG é altamente dependente de seu conhecimento prévio da física subjacente. Sem esse conhecimento, os alunos tendem a aceitar as "alucinações" da IAG, confiando demais nela e deixando de exercitar o pensamento crítico. O estudo também conclui que, embora a IAG possa substituir o professor em questões conceituais, ele permanece crucial para identificar as necessidades dos alunos e ajudá-los a compreender o potencial e as limitações da IA, garantindo que a atenção dos estudantes seja adequadamente direcionada.

Por fim, ressaltamos um estudo quase-experimental sobre um jogo para aprendizagem de conceitos de física (Chen & Chang, 2024). Os alunos responderam testes de aprendizagem antes e depois do jogo. Não foi encontrada diferença significativa de aprendizagem entre o grupo que jogou sem IAG e o que a usou livremente. Mas havia um terceiro grupo que usou IAG com exemplos de perguntas predefinidas, para ajudar a focar nos conceitos importantes do currículo de ciências. Nesse grupo, os resultados de aprendizagem foram significativamente maiores do que no grupo que não usou IAG.

Aprendizagem e agência

Com objetivo de realizar uma análise sistemática sobre a aprendizagem no nosso corpus, identificamos os estudos com resultados empíricos relacionados à aprendizagem em ciências – não incluindo instrumentos de autorrelato e autoavaliação, nem desempenho em jogos ou tarefas específicas. Dos 17 estudos, foram identificados seis, sendo que quatro deles utilizaram questões de múltipla escolha para avaliação da aprendizagem em ciências (Chen & Chang, 2024; Chen & Law, 2025; Gunes & Yilmazlar, 2025; Tong et al., 2025). Os dois estudos restantes utilizaram questões de preenchimento de lacunas (Guo et al., 2025) e avaliação de relatórios com base em rubricas (Lai, 2025).

Todos os seis estudos descrevem intervenções e, logo depois, a aplicação de um teste ou avaliação. Quatro deles têm desenho quase-experimental, um compara perfis de uso entre alunos que usaram IAG (Lai, 2025), e o sexto correlaciona uso e desempenho (Chen & Law, 2025), único que não aplicou medida anterior à intervenção – neste caso, além de resultados menos robustos, a inferência sobre aprendizagem é mais frágil, pois não foi avaliado o processo, apenas o resultado final. As intervenções duraram entre 2 e 6 semanas, exceto em um estudo, com duas intervenções de 20 minutos (Tong et al., 2025).

Quanto às conclusões desses seis estudos, quatro compararam grupos com e sem uso de IAG, sendo que em três deles o uso de IAG melhorou a aprendizagem (Chen & Chang, 2024; Gunes & Yilmazlar, 2025; Guo et al., 2025). Cabe notar que, nesses três estudos, a IAG foi intencionalmente integrada ao ambiente de aprendizagem, não foi apenas utilizada livremente pelos alunos.

Em Gunes e Yilmazlar (2025), porém, o desenho confunde ferramenta e estratégia: o grupo experimental recebeu a estratégia REACT integrada à IAG e o controle, o ciclo 5E, de modo que o efeito atribuível à IAG não é separável do efeito da estratégia. O quarto estudo concluiu que a interação dos alunos com IAG contribuiu para a aprendizagem, mas foi menos eficaz do que a interação dos alunos entre si (Tong et al., 2025).

Nos dois estudos restantes, todos os alunos usaram IAG, sendo assim comparados diferentes tipos de uso. Um dos estudos comparou estratégias “holistas” e “atomistas” de interação com a IAG, concluindo que as estratégias “holistas” levaram a melhores resultados em duas das três dimensões avaliadas pelas rubricas (Lai, 2025). O outro estudo, com metodologia menos robusta, não encontrou correlações significativas entre perfis de busca e resultados de aprendizagem (Chen & Law, 2025).

Em relação à agência, buscamos realizar semelhante análise sistemática, identificando os estudos com resultados empíricos que podem ser interpretados a partir da ideia de agência. Mais especificamente, a partir dos conceitos de Bandura (2018), que distingue a agência individual ou coletiva da agência “por procuração”. Nesse sentido, identificamos quatro estudos com resultados empíricos que comparam dois tipos de interação HIAG, sendo um tipo mais ativo e outro mais passivo, em que a pessoa delega a agência à IAG. A análise desses quatro estudos é descrita na Tabela 5.

Tabela 5. Interações e resultados relacionados à agência.

Estudo	i1: interação com agência humana	i2: interação com agência delegada à IAG	Resultados empíricos
(Lee & So, 2025)	“Agência de aprendiz”: tomar iniciativa pela própria aprendizagem; ação intencional de avaliar, escolher e ajustar os recursos da IAG com ideias próprias.	Comportamento de “pegar carona”: aceitação passiva das ideias da IAG e uso de comandos simples e repetitivos.	Estudantes com alto desempenho (em originalidade e diversidade de ideias, no contexto do Pensamento Criativo) interagiram com alta agência (i1), enquanto os estudantes com baixo desempenho interagiram com baixa agência (i2).
(Chen & Law, 2025)	“Busca instrumental”: Pedir dicas e pistas à IAG, buscando resolver o problema por si mesmo.	Ficar perguntando tudo para a IAG, com pouco raciocínio independente na resolução de problemas.	A busca instrumental (i1) correlacionou-se positivamente com a pontuação no jogo; já o número de perguntas à IAG (i2) correlacionou-se negativamente com a pontuação ⁴ .

⁴ Como se trata de correlação, não se pode estabelecer relações causais. Os autores interpretam essa correlação considerando que os alunos mais capacitados para o jogo não precisaram perguntar tanto à ferramenta. Nesse sentido, não seria a falta de agência que resultou em pior desempenho, mas sim que essas duas variáveis são explicadas por uma terceira: a capacidade prévia do aluno.

Estudo	i1: interação com agência humana	i2: interação com agência delegada à IAG	Resultados empíricos
(Gunbatar & Sirin, 2025)	Uso de estratégias mais complexas de <i>prompt</i> , como os Exemplos e a Cadeia de Pensamento.	Uso de <i>prompts</i> simples, delegando à IAG decisões sobre estrutura e critérios pedagógicos da atividade.	Planos mais guiados pelo usuário (i1) apresentaram maior qualidade; o plano que delegou tudo à IAG (i2) foi o pior, segundo avaliação de professores.
(Lai, 2025)	Aprendizes “holistas”: usam a IAG para estruturar o conhecimento científico, organizando informações e buscando ativamente conexões ou inconsistências entre diferentes tipos de conhecimento.	Aprendizes “atomistas”: focam as perguntas em elementos isolados, buscando a consolidação de discursos já existentes.	Nos relatórios, alunos com estratégia holista (i1) superaram os atomistas (i2) em acurácia de conteúdo e em descrições lógicas.

A Tabela 5 sugere uma tendência consistente no corpus, relacionando a agência humana e a qualidade dos resultados gerados por IAG: quando a pessoa não delega a agência à ferramenta, os resultados tendem a ser melhores. Trata-se de uma tendência interpretativa do corpus, e não de uma relação causal estabelecida. Os resultados, nesse caso, se referem a desempenho em tarefas e jogos que, isoladamente, podem ser considerados indícios, porém não evidências de aprendizagem. Apenas um desses quatro estudos (Lai, 2025) comparou resultados antes e depois da intervenção, o que permite uma inferência mais robusta a respeito da aprendizagem. Cabe ressaltar, ainda, a ausência de estudos investigando a aprendizagem em médio e longo prazo, e não apenas logo após a intervenção.

Discussão

Nossos resultados, específicos para o Ensino de Ciências, convergem com o estado da arte sobre IAG na educação em geral. A associação entre delegação de agência à IAG e piores resultados (Tabela 1) é compatível com a tendência observada no corpus (Tabela 5). A importância do desenho pedagógico intencional, apontada por Dizon e Prudente (2025) e Fesler et al. (2026), também se confirma: a IAG só melhorou a aprendizagem quando integrada de forma pedagogicamente estruturada (Chen & Chang, 2024). O risco de dependência e pensamento superficial também se manifestou nos estudos analisados: tanto alunos (Kilde-Westberg et al., 2025; Tong et al., 2025) quanto professores (Cho et al., 2025) tendem a usar a IAG para obter respostas prontas, de forma acrítica, corroborando os alertas de Otto et al. (2025) sobre a tensão entre o que o aluno produz com auxílio da IAG e o que efetivamente compreende. Além disso, a predominância de resultados positivos no corpus analisado replica, no Ensino de Ciências, um padrão já identificado na literatura mais ampla sobre IAG na educação, e que decorre sobretudo dos vieses já apontados.

Por outro lado, um dos estudos do nosso corpus encontrou evidências de que, embora a interação com IAG tenha contribuído para a aprendizagem, essa contribuição foi menor do que a proporcionada pela interação dos alunos entre si (Tong et al., 2025). Outro estudo concluiu que o uso da IAG de forma livre não apresentou resultados significativamente melhores do que não usar IAG (Chen & Chang, 2024). Na mesma linha, o estudo de Chen e

Law (2025) não encontrou associação entre uso de IAG e aprendizagem; o uso efetivo da ferramenta chegou a se correlacionar negativamente com o desempenho no jogo, enquanto o perfil autorrelatado de busca instrumental (pedir dicas, não respostas) se correlacionou positivamente com esse desempenho.

O contínuo entre o exercício e a delegação da agência (Bandura, 2018) é o eixo que organiza a leitura desses achados: a Zona de Desenvolvimento Proximal e a carga cognitiva, discutidas a seguir, ajudam a entender as relações entre agência e aprendizagem. Esse contínuo pode ser observado no estudo de Chen e Chang (2024) que, embora não cite o termo “agência”, reforça a importância do planejamento intencional do professor para guiar os alunos em seu processo de aprendizagem, fornecendo apoio quando necessário. Com a IAG não é diferente: a liberdade irrestrita de uso não garantiu, por si só, melhores resultados. A ausência de direcionamento reduziu a agência efetiva do aluno que, sem orientação, não sabia o que perguntar à ferramenta. A solução veio pela agência docente: ao reestruturar o jogo com exemplos de perguntas predefinidas, o professor criou as condições para que a IAG fosse utilizada de forma mais pedagógica.

O mesmo contínuo aparece nos perfis descritos por Lai (2025): os aprendizes holista e atomista representam polos distintos entre o exercício e a delegação da agência na aprendizagem⁵. O holista tende a exercer agência sobre a própria aprendizagem, usando a IAG para construir seu próprio conhecimento, buscando conexões e inconsistências; já o atomista, que aprendeu menos, tende a ser menos ativo, usando a ferramenta para buscas pontuais e para confirmar o que já sabe.

Além de Bandura, dois referenciais teóricos se mostraram especialmente férteis para a análise do uso educacional da IAG. Um deles é Vygotsky (1991) e seu conceito de ZDP⁶, que se refere à distância entre o que o aluno realiza sozinho e o que realiza em colaboração, e este segundo caso poderia carregar consigo um desenvolvimento individual em potencial, iminente. A aprendizagem colaborativa potencializa a aprendizagem individual. A criança aprende imitando, mas a imitação, em Vygotsky, não é cópia – ela supõe alguma compreensão das relações estruturais do problema, de modo que só se imita aquilo que já está ao alcance do próprio desenvolvimento (Chaiklin, 2011). Não se trata de reproduzir um resultado, e sim de executar a operação com um apoio que se tornará desnecessário no futuro. Daí a metáfora do andaime, proposta por Wood et al. (1976), em que o apoio pedagógico tem natureza de guia temporário para o desenvolvimento. O termo (*scaffold**) é encontrado em 12 dos 17 artigos do corpus.

No contexto da IAG, tais conceitos podem ajudar a compreender tanto as limitações da ferramenta, que é treinada antes de tudo para dar uma resposta certa, quanto possíveis formas educacionalmente interessantes de utilizar essas ferramentas. Grosso modo, quando

⁵ A agência não aparece explícita no estudo, mas além da descrição dos perfis, a própria descrição do questionário OISS (com estratégias metacognitivas, comportamentais e procedimentais) já pode ser lida a partir do conceito de agência.

⁶ A expressão russa *zona blizaihego razvitia* recebeu traduções divergentes no Brasil – desenvolvimento proximal, imediato ou iminente. Zoia Prestes (2010) defende “iminente” como a mais fiel, por preservar o caráter de possibilidade: trata-se do que pode vir a ocorrer, não do que necessariamente ocorrerá. Mantemos aqui o termo corrente na área, sem desconhecer o debate.

o professor apoia de forma pedagógica, quem executa a operação é o aluno, e o desempenho assistido indica desenvolvimento em curso. Por outro lado, quando a tarefa é delegada à ferramenta, quem executa é a ferramenta, e o aluno recebe o produto sem ter percorrido o caminho, sem esforço nem intencionalidade. Nesse caso, o desempenho assistido (pela IAG e não pelo professor) deixa de indicar o que permanecerá como aprendizagem duradoura: o ganho durante a tarefa desaparece quando a ferramenta é retirada (Bastani et al., 2025; Kreijkes et al., 2026; Shen & Tamkin, 2026). Mais do que o tipo de ferramenta, o tipo de interação se mostrou fundamental para que a IAG contribua para a aprendizagem significativa – como mostram os perfis de interação descritos por Lai (2025).

Outro referencial teórico relevante é a Teoria da Carga Cognitiva (Sweller, 2011), que distingue a carga intrínseca (associada à complexidade inerente do conteúdo) da carga extrínseca (gerada pelo formato e pelas condições da tarefa)⁷. A literatura aponta que a IAG pode liberar esforço mental para tarefas cognitivas mais elevadas, como o pensamento crítico, mas registra também o efeito inverso, com a dependência da ferramenta associada a menor desenvolvimento dessas mesmas habilidades (Otto et al., 2025). Nosso corpus aponta um aspecto ambivalente semelhante: a mesma redução de esforço que pode liberar o aluno para pensar mais profundamente pode também simplesmente dispensá-lo de pensar. O que os estudos permitem observar, contudo, não é a carga cognitiva em si, mas o comportamento: quando a IAG faz parte de um desenho baseado em andaime e o aluno segue executando a tarefa, os resultados tendem a ser melhores (Guo et al., 2025). E tendem a ser piores quando o aluno delega a tarefa inteira (Kilde-Westberg et al., 2025; Lee & So, 2025).

Cabe registrar os dois estudos do corpus que abordam a carga cognitiva. Um deles mede a carga cognitiva e encontrou o padrão oposto ao esperado: no grupo com maior aprendizagem (em medidas aplicadas logo antes e depois do jogo), houve redução da carga intrínseca, sem diferença na extrínseca (Chen & Chang, 2024). No segundo, o tema aparece em entrevistas semiestruturadas com cinco professores de física, que relatam os dois sentidos: diminuição da carga cognitiva dos alunos em conceitos abstratos, quando a IAG fornece exemplos do dia a dia relacionados a um conceito, e aumento da carga na operação de um novo ambiente digital, quando os estudantes têm dificuldades com a plataforma de laboratório, especialmente os menos proficientes no mundo digital (Huang et al., 2026).

Nossos resultados também convergem com a literatura ao enfatizar que uma interação de qualidade com a IAG exige letramento em IA e metacognição (Otto et al., 2025). Esses construtos, contudo, não têm definição convergente nos estudos selecionados, de modo que funcionam melhor como hipóteses explicativas do que como categorias de síntese. O corpus enfatiza, além disso, uma condição mais diretamente observável nos dados de interação: o conhecimento disciplinar prévio é determinante para que o usuário avalie criticamente as respostas da IAG e mantenha sua agência no processo (Gunbatar et al., 2025; Kilde-Westberg et al., 2025; Lee & So, 2025; Valeri et al., 2025). Essa condição tem uma implicação importante: a IAG tende a ser mais útil para quem já possui maior domínio do conteúdo,

⁷ Sweller (2011) trabalha com duas fontes de carga cognitiva. A distinção de uma terceira, a carga relevante (*germane*), é objeto de disputa na própria teoria, tanto em termos teóricos quanto empíricos.

sendo potencialmente prejudicial para quem mais precisaria de apoio – o oposto do que se esperaria de uma ferramenta que contribui para a democratização da educação. Esse é um risco bastante relevante para a escola na era pós-ChatGPT. Por outro lado, isso pode não ser inevitável: Fesler et al. (2026) sugerem que IAGs tutoras (desenhadas para atuar de forma gradual e sem fornecer respostas prontas) podem inverter essa lógica, beneficiando especialmente os alunos mais iniciantes. Com efeito, a distinção entre IAGs genéricas e tutoras emerge como uma questão bastante relevante para o Ensino de Ciências, não apenas em termos de eficácia, mas também de equidade.

Por fim, o próprio corpus impõe limites às conclusões desta revisão: muitos artigos não informam versão do modelo ou período de coleta, e a faixa entre 2023 e 2026 abrange modelos com capacidades bastante distintas, de modo que estudos publicados sobre “ChatGPT” podem ter avaliado intervenções tecnologicamente diferentes. Essa indeterminação, porém, não atinge todas as conclusões da mesma forma. Os resultados sobre eficácia comparada entre sistemas dependem do modelo e da versão, e vão perdendo validade com o tempo; já as conclusões mais gerais sobre interação, agência e aprendizagem parecem depender menos da ferramenta, sendo potencialmente mais úteis para a educação.

Conclusão

Nas fontes em português foram recuperados 19 registros, nenhum deles com estudo empírico sobre as interações que constituem nosso objeto (Tabela A2) – uma exclusão temática, não geográfica. Na fonte internacional (*Web of Science*) foram encontrados 17 estudos relevantes. A maioria aponta efeitos positivos do uso de IAG no Ensino de Ciências, o que parece refletir sobretudo o recorte dos estudos disponíveis, com possível viés de desenho e de publicação.

Dois referenciais teóricos se mostraram especialmente férteis para interpretar os achados. A Zona de Desenvolvimento Proximal (Vygotsky, 1991) ajuda a entender certas condições que podem promover ou prejudicar a aprendizagem mediada por IAG. Além disso, com o advento da IAG, torna-se menos claro o que o aluno consegue fazer de forma autônoma, sem auxílio do professor. Observa-se, com efeito, uma tensão entre o que ele pode fazer com ajuda de IAG e o que realmente sabe. Outro referencial importante é a Teoria da Carga Cognitiva (Sweller, 2011). Esse referencial ajuda a entender por que o efeito da IAG depende do desenho pedagógico: aliviar a tarefa favorece a aprendizagem quando o esforço poupado é o acessório, mas a prejudica quando é o esforço necessário para compreensão.

Como alunos e professores interagem com a IAG? Identificamos no corpus sete tipos recorrentes de interação no Ensino de Ciências (Tabela 4). Alguns estudos apontam que tanto alunos quanto professores tendem a interagir de forma mais pontual e acrítica (consultas únicas, *prompts* simples, aceitação passiva de respostas).

Quais fatores modulam a qualidade das interações HIAG? Os estudos apontam de forma consistente que a qualidade das interações depende fundamentalmente das capacidades prévias do usuário. O conhecimento disciplinar é necessário para avaliar criticamente as respostas da IAG e evitar que se aceitem informações imprecisas e as chamadas “alucinações”. O letramento em IA e a metacognição parecem influenciar a

sofisticação dos *prompts* e a capacidade de refinar dialogicamente as respostas da ferramenta. A experiência docente também se mostra importante para a autonomia crítica do professor. Por fim, o desenho pedagógico intencional também se mostrou fundamental, pois, sem estruturação prévia pelo professor, a liberdade irrestrita de uso da IAG não garante – e pode até comprometer – a aprendizagem.

Como a delegação de agência à IAG se relaciona com os resultados das interações HIAG? Nossa análise mostra um padrão consistente no corpus, que deve ser lido como tendência e não como relação causal estabelecida: interações em que o usuário preserva sua agência (tomando iniciativa, avaliando criticamente, ajustando e refinando as respostas da IAG com base em ideias e objetivos próprios) produzem resultados superiores aos obtidos quando a agência é delegada à ferramenta. O conceito de agência por procuração (Bandura, 2018) pode servir para enquadrar esse fenômeno em que, ao delegar à IAG decisões que deveriam ser suas, o estudante abdica do esforço cognitivo necessário para a aprendizagem. O mesmo vale para o professor, em relação à qualidade dos planejamentos produzidos com auxílio de IAG.

Como o uso de IAG pode interferir na aprendizagem? De modo geral, os estudos defendem que a IAG pode contribuir como apoio, mas não em substituição ao professor. Os seis estudos com dados empíricos sobre aprendizagem em ciências apresentam resultados diversos. Em três deles, a integração intencional da IAG ao ambiente de aprendizagem levou a ganhos significativos em relação ao grupo de controle. Em um estudo, a IAG contribuiu para a aprendizagem, mas com efeito sete vezes menor do que a interação entre pares. Nos dois restantes, compararam-se tipos de uso: em um, as estratégias mais ativas produziram melhores resultados; no outro, de desenho menos robusto, não houve associação. Cabe ressaltar, no entanto, a ausência de estudos avaliando a aprendizagem em médio e longo prazo: todos os instrumentos foram aplicados logo após a intervenção, o que limita as inferências a respeito de uma aprendizagem significativa.

Quais os riscos e incertezas associados ao uso de IAG no contexto educacional? O risco mais documentado no corpus é o da dependência e do pensamento superficial: alunos e professores tendem, espontaneamente, a usar a IAG para obter respostas prontas de forma acrítica, o que pode inibir o desenvolvimento de habilidades cognitivas de ordem superior. Quanto mais as ferramentas buscam diminuir a “fricção” para o usuário, maior é a chance de redução do esforço necessário para a aprendizagem, aquele relacionado à carga cognitiva intrínseca. O risco de confiança cega e antropomorfização da ferramenta é especialmente preocupante em contextos de baixo letramento científico e em IA. Há ainda o risco de aprofundamento de desigualdades educacionais: como a qualidade da interação depende de capacidades prévias, a IAG tende a beneficiar desproporcionalmente quem já possui maior domínio de conteúdo, letramento em IA e sofisticação metacognitiva. Quanto às incertezas, destacam-se: a ausência de estudos sobre efeitos de médio e longo prazo; a predominância de contextos asiáticos, limitando a generalização dos achados; os vieses de desenho e de publicação já apontados, que podem subestimar os efeitos negativos; e o ainda insuficiente mapeamento das condições em que IAGs tutoras podem reverter essas tendências. Por fim, há uma dimensão ética pouco explorada nos estudos analisados: a coleta de dados de alunos e professores para o desenvolvimento de sistemas de IAG, frequentemente sem transparência

nos termos de uso e sem contrapartida aos participantes, representa um risco sobre o qual as pesquisas ainda precisam se debruçar com mais atenção.

Por fim, destacamos três limitações desta revisão. A primeira é a autoria única, cujos procedimentos de mitigação estão descritos nos Métodos. A segunda é a cobertura: além de se limitar a uma fonte internacional, a estratégia de busca produziu falsos negativos, como um estudo sobre ensino de evolução, com desenho experimental robusto (Aptyka et al., 2025).

A terceira limitação é uma condição estrutural do campo: as ferramentas de IAG mudam em ciclos de meses, enquanto uma revisão sistemática leva mais de um ano entre a busca e a publicação. Qualquer síntese sobre IAG na educação descreve um estado de coisas parcialmente indeterminado e temporário. Além disso, a literatura primária raramente informa qual sistema foi usado (Tabela A3). Na área da saúde, as diretrizes para ensaios clínicos com intervenções de IA exigem que se declare a versão do algoritmo e se ela mudou durante o estudo (SPIRIT-AI e CONSORT-AI; Cruz Rivera et al., 2020; Liu et al., 2020). Propomos um registro análogo para IA na educação: ferramenta, versão, período de coleta e configuração de uso (Tabela A3). Isso permite, na pesquisa do campo, que se veja os resultados por geração tecnológica, em vez de tratá-los como evidência sobre “a IAG” em geral.

Levando em conta tais indeterminações, vale indicar alguns achados e hipóteses potencialmente mais duradouros, e que poderiam ser foco de estudos futuros: a associação entre “delegação de agência” e piores resultados (desempenho, aprendizagem); a primazia da interação humana sobre a interação com IAG; a dependência do efeito em relação ao desenho pedagógico, mais do que à ferramenta; a distância entre o desempenho assistido com IAG (e a autopercepção de aprendizagem) e a aprendizagem que permanece sem a ferramenta.

Referências

- Aptyka, H., Großschedl, J., & Hartelt, T. (2025). Bugbear or surefire success? Secondary school students' conceptual learning about evolution with ChatGPT. *International Journal of Science Education*. <https://dx.doi.org/10.1080/09500693.2025.2524083>
- Bandura, A. (2018). Toward a psychology of human agency: Pathways and reflections. *Perspectives on Psychological Science*, 13(2), 130–136. <https://dx.doi.org/10.1177/1745691617699280>
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. <https://dx.doi.org/10.1073/pnas.2422633122>
- Beck, U. (2010). *Sociedade de risco: Rumo a uma outra modernidade*. Editora 34.
- Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *Contemporary Educational Technology*, 15(3), ep430. <https://dx.doi.org/10.30935/cedtech/13176>
- Chaiklin, S. (2011). A zona de desenvolvimento próximo na análise de Vigotski sobre aprendizagem e ensino (J. C. Pasqualini, Trad.). *Psicologia em Estudo*, 16(4), 659–675. <https://dx.doi.org/10.1590/S1413-73722011000400016>
- Chen, C.-H., & Chang, C.-L. (2024). Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior. *Education and Information Technologies*, 29(14), 18621–18642. <https://dx.doi.org/10.1007/s10639-024-12553-x>

- Chen, C.-H., & Law, V. (2025). The role of help-seeking from ChatGPT in digital game-based learning. *Educational Technology Research and Development*, 73(3), 1703–1721. <https://dx.doi.org/10.1007/s11423-025-10464-3>
- Cheung, K. K. C., Long, Y., Liu, Q., & Chan, H.-Y. (2025). Unpacking epistemic insights of artificial intelligence (AI) in science education: A systematic review. *Science & Education*, 34(2), 747–777. <https://dx.doi.org/10.1007/s11191-024-00511-5>
- Cho, N., Li, H., Zhang, X., & Li, Y. (2025). Physics teachers' insights into the usability and challenges of a platform with integrated simulations and generative AI. *Journal of Science Education and Technology*. <https://dx.doi.org/10.1007/s10956-025-10280-7>
- Cruz Rivera, S., Liu, X., Chan, A.-W., Denniston, A. K., & Calvert, M. J. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. *Nature Medicine*, 26(9), 1351–1363. <https://dx.doi.org/10.1038/s41591-020-1037-7>
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://dx.doi.org/10.1016/j.compedu.2024.105224>
- Dizon, J., & Prudente, M. S. (2025). AI-powered chatbots in science education: Meta-analysis of pedagogical integration and student achievement. *International Journal of Chemistry Education Research*, 9(2), 134–150. <https://dx.doi.org/10.20885/ijcer.vol9.iss2.art4>
- Fesler, L., Claeys, J. M., Agnew, C., & Loeb, S. (2026). *The evidence base on AI in K-12: A 2026 review*. Stanford University. <https://scale.stanford.edu/sites/default/files/The%20Evidence%20Base%20on%20AI%20in%20K-12%20Report.pdf>
- Fock, A., & Siller, H.-S. (2025). Generative artificial intelligence in secondary STEM education in the light of Human Flourishing: A scoping literature review. *International Journal of STEM Education*, 12(1), 67. <https://dx.doi.org/10.1186/s40594-025-00589-5>
- Georgiou, G. P. (2025). *ChatGPT produces more “lazy” thinkers: Evidence of cognitive engagement decline* (arXiv:2507.00181). arXiv. <https://dx.doi.org/10.48550/arXiv.2507.00181>
- Gunbatar, S. A., & Sirin, G. T. (2025). Does using prompting strategies affect the production of higher-quality socio-scientific issue activity? Analysis of nuclear energy and pesticide use cases for high school chemistry teaching. *Journal of Chemical Education*, 102(11), 4676–4688. <https://dx.doi.org/10.1021/acs.jchemed.5c00684>
- Gunbatar, S. A., Sirin, G. T., Ilkyaz, O. C., & Mutlu, Y. (2025). Exploring the artificial intelligence interaction profiles of participants with different levels of teaching experience for lesson planning in the context of acids and bases. *Chemistry Education Research and Practice*, 26(4), 977–995. <https://dx.doi.org/10.1039/d5rp00064e>
- Gunes, L., & Yilmazlar, M. (2025). The effect of REACT strategy integrated with artificial intelligence on academic achievement. *International Journal of Education in Mathematics, Science and Technology*, 13(5), 1083–1096. <https://dx.doi.org/10.46328/ijemst.5698>
- Guo, Q., Zhen, J., Wu, F., He, Y., & Qiao, C. (2025). Can students make STEM progress with the large language models (LLMs)? An empirical study of LLMs integration within middle school science and engineering practice. *Journal of Educational Computing Research*, 63(2), 372–405. <https://dx.doi.org/10.1177/07356331241312365>
- Hosseini, M.-S., Jahanshahloo, F., Akbarzadeh, M. A., Zarei, M., & Vaez-Gharamaleki, Y. (2024). Formulating research questions for evidence-based studies. *Journal of Medicine, Surgery, and Public Health*, 2, 100046. <https://dx.doi.org/10.1016/j.glmedi.2023.100046>
- Huang, L., Zhou, X., & Zhu, W. (2026). The teaching model (IT-CSTP) for integrating intelligent technology into scientific thinking cultivation in physics: A practical study for high school students. *Journal of Science Education and Technology*, 35(1), 27–44. <https://dx.doi.org/10.1007/s10956-025-10225-0>

- Kilde-Westberg, S., Johansson, A., & Enger, J. (2025). Generative AI as a lab partner: A case study. *Physical Review Physics Education Research*, 21(2), 020119. <https://dx.doi.org/10.1103/ggy1-3kjk>
- Krakowski, S. (2025). Human-AI agency in the age of generative AI. *Information and Organization*, 35(1), 100560. <https://dx.doi.org/10.1016/j.infoandorg.2025.100560>
- Kreijkes, P., Kewenig, V., Kuvalja, M., Lee, M., Hofman, J. M., Vitello, S., Sellen, A., Rintel, S., Goldstein, D. G., Rothschild, D., Tankelevitch, L., & Oates, T. (2026). Effects of LLM use and note-taking on reading comprehension and memory: A randomised experiment in secondary schools. *Computers & Education*, 243, 105514. <https://dx.doi.org/10.1016/j.compedu.2025.105514>
- Krupp, L., Steinert, S., Kiefer-Emmanouilidis, M., Avila, K. E., Lukowicz, P., Kuhn, J., Küchemann, S., & Karolus, J. (2024). Unreflected acceptance – Investigating the negative consequences of ChatGPT-assisted problem solving in physics education. Em *HHAI 2024: Hybrid human AI systems for the social good* (Frontiers in Artificial Intelligence and Applications, Vol. 386, pp. 199–212). IOS Press. <https://dx.doi.org/10.3233/FAIA240195>
- Kundu, A., & Bej, T. (2025). Psychological impacts of AI use on school students: A systematic scoping review of the empirical literature. *Research and Practice in Technology Enhanced Learning*, 20, 30. <https://dx.doi.org/10.58459/rptel.2025.20030>
- Lai, C. L. (2025). Consolidate knowledge or build scientific models? The role of online information-searching strategies in students' prompt sequences with generative artificial intelligence. *Educational Technology & Society*, 28(3), 241–262. [https://dx.doi.org/10.30191/ETS.202507_28\(3\).TP03](https://dx.doi.org/10.30191/ETS.202507_28(3).TP03)
- Latif, E., & Zhai, X. (2024). Fine-tuning ChatGPT for automatic scoring. *Computers and Education: Artificial Intelligence*, 6, 100210. <https://dx.doi.org/10.1016/j.caeai.2024.100210>
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.
- Lee, C., & So, H.-J. (2025). Learners leveraging generative AI for creative problem solving: Focusing on the PISA 2022 creative thinking problems. *Educational Technology & Society*, 28(4), 241–258. [https://dx.doi.org/10.30191/ETS.202510_28\(4\).SP04](https://dx.doi.org/10.30191/ETS.202510_28(4).SP04)
- Lintner, T., Juhaňák, L., Žáková, M., Plch, L., & Ropovik, I. (2026). Generative AI use and academic achievement: A cross-sectional study of Czech seventh-grade students. *Education and Information Technologies*. <https://dx.doi.org/10.1007/s10639-026-14002-3>
- Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., & Denniston, A. K. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Nature Medicine*, 26(9), 1364–1374. <https://dx.doi.org/10.1038/s41591-020-1034-x>
- Meyer, A., Bleckmann, T., & Friege, G. (2025). Automatic feedback on physics tasks using open-source generative artificial intelligence. *International Journal of Science Education*. <https://dx.doi.org/10.1080/09500693.2025.2499220>
- Monteiro, F. F., Souza, P. V. S., Silva, M. C. da, Maia, J. R., Silva, W. F. da, & Girardi, D. (2024). ChatGPT in Brazilian K-12 science education. *Frontiers in Education*, 9, 1321547. <https://dx.doi.org/10.3389/feduc.2024.1321547>
- Narayan, N. (2025). AI, cognition, and the cost of convenience. *Journal of Technology and Systems*, 7(6), 18–34. <https://dx.doi.org/10.47941/jts.3185>
- Naumann, L., Mayr, J. L., & Markic, S. (2025). Between acceptance and scepticism: An investigation into secondary school students' attitudes toward artificial intelligence in chemistry education. *CHIMIA*, 79(10), 693–697. <https://dx.doi.org/10.2533/chimia.2025.693>
- Nyaaba, M., Zhai, X., & Faison, M. Z. (2024). Generative AI for culturally responsive science assessment: A conceptual framework. *Education Sciences*, 14(12), 1325. <https://dx.doi.org/10.3390/educsci14121325>
- Otto, S., Lavi, R., & Bertel, L. B. (2025). Human-GenAI interaction for active learning in STEM education: State-of-the-art and future directions. *Computers & Education*, 239, 105444. <https://dx.doi.org/10.1016/j.compedu.2025.105444>

- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, n160. <https://dx.doi.org/10.1136/bmj.n160>
- Potter, A. H., Arner, T., McCarthy, K. S., & McNamara, D. S. (2026). Integrating reading, writing, and digital tools in science: A participatory-design study of the InSPECT framework. *Education Sciences*, 16(1), 6. <https://dx.doi.org/10.3390/educsci16010006>
- Prestes, Z. R. (2010). *Quando não é quase a mesma coisa: Análise de traduções de Lev Semionovitch Vigotski no Brasil – repercussões no campo educacional* [Tese (Doutorado), Universidade de Brasília]. <https://repositorio.unb.br/handle/10482/9123>
- Raitskaya, L., & Tikhonova, E. (2025). Enhancing critical thinking skills in ChatGPT-human interaction: A scoping review. *Journal of Language and Education*, 11(2), 5–19. <https://dx.doi.org/10.17323/jle.2025.27387>
- Rivera-Novoa, A., & Duarte Arias, D. A. (2025). Generative artificial intelligence and extended cognition in science learning contexts. *Science & Education*. <https://dx.doi.org/10.1007/s11191-025-00660-1>
- Sasikala, P., & Ravichandran, R. (2024). Study on the impact of artificial intelligence on student learning outcomes. *Journal of Digital Learning and Education*, 4(2), 145–155. <https://dx.doi.org/10.52562/jdle.v4i2.1234>
- Shen, J. H., & Tamkin, A. (2026). *How AI impacts skill formation* (arXiv:2601.20245). arXiv. <https://dx.doi.org/10.48550/arXiv.2601.20245>
- Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, 160, 108386. <https://dx.doi.org/10.1016/j.chb.2024.108386>
- Sweller, J. (2011). Cognitive load theory. Em J. P. Mestre & B. H. Ross (Orgs.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Academic Press. <https://dx.doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Tang, K.-S., & Cooper, G. (2025). The role of materiality in an era of generative artificial intelligence. *Science & Education*, 34(2), 731–746. <https://dx.doi.org/10.1007/s11191-024-00508-0>
- Tong, D., Jin, B., Tao, Y., Ren, H., Islam, A. Y. M. A., & Bao, L. (2025). Exploring the role of human-AI collaboration in solving scientific problems. *Physical Review Physics Education Research*, 21(1), 010149. <https://dx.doi.org/10.1103/PhysRevPhysEducRes.21.010149>
- Valeri, F., Nilsson, P., & Cederqvist, A.-M. (2025). Exploring students' experience of ChatGPT in STEM education. *Computers and Education: Artificial Intelligence*, 8, 100360. <https://dx.doi.org/10.1016/j.caeai.2024.100360>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
- Vygotsky, L. S. (1991). *A formação social da mente* (4ª ed.). Martins Fontes.
- Wang, S., Chen, F., Ge, C., Shi, H., You, Y., Guo, J., Zheng, S., & Zhu, X. (2025). Using local LLM tools to optimize Chinese high school chemistry education: Practice, challenges, and future directions. *Journal of Chemical Education*, 102(10), 4368–4375. <https://dx.doi.org/10.1021/acs.jchemed.5c00220>
- Wenker, K. (2023). Who wrote this? How smart replies impact language and agency in the workplace. *Telematics and Informatics Reports*, 10, 100062. <https://dx.doi.org/10.1016/j.teler.2023.100062>
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://dx.doi.org/10.1111/j.1469-7610.1976.tb00381.x>

- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11(1), 28. <https://dx.doi.org/10.1186/s40561-024-00316-7>
- Zhang, J., Jantakoon, T., & Laoha, R. (2025). Meta-analysis of artificial intelligence in education. *Higher Education Studies*, 15(2), 189. <https://dx.doi.org/10.5539/hes.v15n2p189>

Apêndice

Tabela A1. Documentação do processo de triagem dos estudos em inglês.

1º Autor	Ano	Título	Triagem 1	Triagem 2
Bitzenbauer	2023	ChatGPT in physics education: A pilot study on easy-to-implement activities	Incluído	Incluído (Cat. 1)
Dengel	2023	Qualitative Research Methods for Large Language Models: Conducting Semi-Structured Interviews with ChatGPT and BARD on Computer Science Education	Excluído (C1)	
Adeyele	2024	Exploring the Integration of ChatGPT in Inquiry-based Learning: Teacher Perspectives	Incluído	Excluído (Cat. 2)
Beege	2024	AI in STEM education: The relationship between teacher perceptions and ChatGPT use	Incluído	Excluído (Cat. 2)
Chen	2024	Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior	Incluído	Incluído (Cat. 1)
Latif	2024	Fine-tuning ChatGPT for automatic scoring	Incluído	Excluído (Cat. 3)
Monteiro	2024	ChatGPT in Brazilian K-12 science education	Incluído	Excluído (Cat. 2)
Nyaaba	2024	Generative AI for Culturally Responsive Science Assessment: A Conceptual Framework	Incluído	Excluído (Cat. 3)
Polverini	2024	Performance of ChatGPT on the test of understanding graphs in kinematics	Incluído	Excluído (Cat. 4)
Ariza	2025	Generative AI in Engineering and Computing Education: A Scoping Review of Empirical Studies and Educational Practices	Excluído (C1)	
Chen	2025	The role of help-seeking from ChatGPT in digital game-based learning	Incluído	Incluído (Cat. 1)
Chen	2025	The Current Status and Trends of STEM Talent Cultivation in Chinese Higher Education Institutions: Based on a Holistic Analysis Model of STEM Programs	Excluído (C1)	
Cheng	2025	Leveraging generative AI in science lesson study: transforming density concept instruction through ChatGPT integration	Excluído (C3)	
Cheung	2025	Unpacking Epistemic Insights of Artificial Intelligence (AI) in Science Education: A Systematic Review	Excluído (C2)	
Cho	2025	Physics Teachers' Insights into the Usability and Challenges of a Platform with Integrated Simulations and Generative AI	Incluído	Incluído (Cat. 1)
Coskun	2025	Comparative analysis of AI and expert evaluations in engineering design pedagogy	Excluído (C1)	

1º Autor	Ano	Título	Triagem 1	Triagem 2
Fock	2025	Generative artificial intelligence in secondary STEM education in the light of Human Flourishing: a scoping literature review	Excluído (C2)	
Gunbatar	2025	Does Using Prompting Strategies Affect the Production of Higher-Quality Socio-Scientific Issue Activity? Analysis of Nuclear Energy and Pesticide Use Cases for High School Chemistry Teaching	Incluído	Incluído (Cat. 1)
Gunbatar	2025	Exploring the artificial intelligence interaction profiles of participants with different levels of teaching experience for lesson planning in the context of acids and bases	Incluído	Incluído (Cat. 1)
Guo	2025	Can Students Make STEM Progress With the Large Language Models (LLMs)? An Empirical Study of LLMs Integration Within Middle School Science and Engineering Practice	Incluído	Incluído (Cat. 1)
Gunes	2025	The Effect of REACT Strategy Integrated with Artificial Intelligence on Academic Achievement	Incluído	Incluído (Cat. 1)
Kilde-Westberg	2025	Generative AI as a lab partner: A case study	Incluído	Incluído (Cat. 1)
Kim	2025	Science Teachers' Approaches to Artificial Intelligence Integrated Science Teaching	Excluído (C1)	
Kim	2025	From co-design to co-teaching: a comprehensive approach to an integrated AI curriculum in middle school STEM education	Excluído (C1)	
Lai	2025	Consolidate knowledge or build scientific models? The role of online information-searching strategies in students' prompt sequences with Generative Artificial Intelligence	Incluído	Incluído (Cat. 1)
Lee	2025	Learners leveraging generative AI for creative problem solving: Focusing on the PISA 2022 creative thinking problems	Incluído	Incluído (Cat. 1)
Liu	2025	What are the differences between student and ChatGPT-generated pseudocode? Detecting AI-generated pseudocode in high school programming using explainable machine learning	Excluído (C1)	
Meyer	2025	Automatic feedback on physics tasks using open-source generative artificial intelligence	Incluído	Excluído (Cat. 3)
Naumann	2025	Between Acceptance and Scepticism: An Investigation into Secondary School Students' Attitudes toward Artificial Intelligence in Chemistry Education	Incluído	Incluído (Cat. 1)
Nguyen	2025	Applying Generative Artificial Intelligence to Critiquing Science Assessments	Incluído	Excluído (Cat. 3)
Otto	2025	Human-GenAI interaction for active learning in STEM education: State-of-the-art and future directions	Excluído (C2)	
Shankar	2025	Bringing Teachers in the Loop: Exploring Perspectives on Integrating Generative AI in Technology-Enhanced Learning	Incluído	Excluído (Cat. 2)
Sirisathitkul	2025	Converging objects-to-think-with: exploring microwave ovens through inquiry-based learning with ChatGPT	Excluído (C1)	

1º Autor	Ano	Título	Triagem 1	Triagem 2
Tong	2025	Exploring the role of human-AI collaboration in solving scientific problems	Incluído	Incluído (Cat. 1)
Trcková	2025	Lost in Humus: An Educational Escape Game Focused on Soil Analysis	Incluído	Excluído (Cat. 3)
Valeri	2025	Exploring students' experience of ChatGPT in STEM education	Incluído	Incluído (Cat. 1)
Wang	2025	Using Local LLM Tools to Optimize Chinese High School Chemistry Education: Practice, Challenges, and Future Directions	Incluído	Incluído (Cat. 1)
Wattanakasiwich	2025	Physics instructors' acceptance and implementation of generative AI	Excluído (C1)	
Yu	2025	Examining the use of ecology-related socio-scientific issue debate class using generative AI to improve middle school students' science and AI literacy	Excluído (C3)	
Zhao	2025	Exploring the Effects of the CER Model-Based GenAI Learning System to Cultivate Elementary School Students' Computational Thinking Core Skills in Science Courses	Excluído (C3)	
Ajose	2026	Exploring Teachers' Awareness and Perceptions of ChatGPT in K-12 STEM Education	Incluído	Excluído (Cat. 2)
García	2026	Teacher Professional Development: A Workshop Proposal for High School-University Collaboration Using Technology and AI	Excluído (C1)	
Huang	2026	The Teaching Model (IT-CSTP) for Integrating Intelligent Technology into Scientific Thinking Cultivation in Physics: A Practical Study for High School Students	Incluído	Incluído (Cat. 1)
Morbidoni	2026	Exploring the Potential of Using ChatGPT in Chemistry Education	Excluído (C1)	
Potter	2026	Integrating Reading, Writing, and Digital Tools in Science: A Participatory-Design Study of the InSPECT Framework	Incluído	Incluído (Cat. 1)
Tipple	2026	Teachers' perceptions of generative AI in gender-inclusive STEM education: a grounded theory study	Incluído	Excluído (Cat. 2)
Wang	2026	A Virtual Instructor-Led System for Assessing and Guiding Middle School Physics Experiments	Excluído (C3)	

Nota: os critérios da Triagem 1 estão descritos na seção de Métodos, e as categorias da Triagem 2 constam na Tabela 2.

Tabela A2. Documentação do processo de triagem dos estudos em português.

1º Autor	Ano	Título	Base	Triagem 1	Triagem 2
Albuquerque	2023	Uso da inteligência artificial no ensino de física: potencialidades e desafios	Capes	Excluído (C1)	
Candido	2023	Educação 4.0: a estética pós-humana na	Capes	Excluído (C1)	

1º Autor	Ano	Título	Base	Triagem 1	Triagem 2
		educação em ciências			
Leite	2023	Inteligência artificial e ensino de química: uma análise propedêutica do ChatGPT na definição de conceitos químicos	SciELO	Incluído	Excluído (Cat. 4)
Oliveira	2023	Elaboração de projetos de pesquisa com auxílio do ChatGPT: um estudo com licenciandos de matemática	Capes	Excluído (C1)	
Pereira	2023	Visualização molecular interativa para situações de estudo	SciELO	Excluído (C1)	
Prates	2023	Ensino de física mediado pelo Machine Learning: o caso do pêndulo simples	SciELO	Excluído (C1)	
Albertoni	2024	Inteligência artificial na educação inclusiva	Capes	Excluído (C3)	
Claudino	2024	Inteligência artificial em contextos restritos de recursos	Capes	Excluído (C1)	
Funaki	2024	Chatbots no ensino de física: possibilidades e desafios	Capes	Excluído (C1)	
Gonçalves Teles	2024	Percepções de professores de química da Amazônia paraense sobre a incorporação da inteligência artificial no ensino	Capes e SciELO	Excluído (C3)	
Neves	2024	Impacto do uso de tecnologias digitais na zona de desenvolvimento proximal de alunos da educação básica	Capes	Excluído (C1)	
Nunes	2024	O ChatGPT e suas influências voltadas à atual escrita científica na área de ensino	Capes	Excluído (C1)	
Padilha	2024	Comunidades de prática de professores de ciências na perspectiva da teoria do agir comunicativo e o ChatGPT	Capes	Excluído (C1)	
Silva	2024	O ChatGPT como recurso auxiliar na elaboração de aulas de ciências e matemática	Capes	Excluído (C1)	
Costa	2025	Análise do ChatGPT como ferramenta no ensino de química: explorando o tema matéria e energia	Capes	Incluído	Excluído (Cat. 4)
Olczewski	2025	ChatGPT no ensino de química: análise das explicações sobre ácidos e bases para o novo ensino médio e ensino superior	Capes e SciELO	Incluído	Excluído (Cat. 4)
Santos	2025	Tecnologia assistiva e inteligência artificial incorporadas ao desenho universal para aprendizagem	Capes	Excluído (C1)	

Nota: os critérios da Triagem 1 estão descritos na seção de Métodos, e as categorias da Triagem 2 constam na Tabela 2. Os dois registros recuperados nas duas bases (Capes e SciELO) aparecem em linha única; por isso, as 17 linhas correspondem a 19 registros.

Tabela A3. Rastreabilidade tecnológica dos estudos selecionados.

Nº	Referência	Modelo e versão	Período de coleta	Configuração
01	(Chen & Chang, 2024)	GPT-3.5-turbo	não reportado	customizada (API integrada ao jogo)
02	(Huang et al., 2026)	ChatGPT, versão não reportada	outubro a dezembro, ano não informado	genérica
03	(Cho et al., 2025)	Doubao, versão não reportada	não reportado	customizada (plataforma própria)
04	(Chen & Law, 2025)	GPT-3.5-turbo	não reportado	customizada (integrada ao jogo)
05	(Valeri et al., 2025)	ChatGPT, versão não reportada	não reportado	genérica (uso espontâneo)
06	(Wang et al., 2025)	não reportado	não reportado	não reportada (implantação local)
07	(Gunbatar & Sirin, 2025)	GPT-3.5 e GPT-5	março de 2025	genérica
08	(Gunbatar et al., 2025)	ChatGPT, versão não reportada	outono de 2024	genérica
09	(Kilde-Westberg et al., 2025)	GPT-3.5	março de 2024	genérica
10	(Tong et al., 2025)	GPT-4o	não reportado	genérica
11	(Guo et al., 2025)	GLM-4	não reportado	genérica
12	(Naumann et al., 2025)	não reportado	não reportado	genérica (plataforma Fobizz)
13	(Lai, 2025)	ChatGPT, Bing ou Google Bard, versões não reportadas	não reportado	genérica
14	(Lee & So, 2025)	GPT-3.5	não reportado	genérica
15	(Bitzenbauer, 2023)	ChatGPT, versão não reportada	não reportado	genérica
16	(Potter et al., 2026)	ChatGPT 3.5 ou Gemini (versão não reportada)	verão de 2023	genérica
17	(Gunes & Yilmazlar, 2025)	ChatGPT, versão não reportada	não reportado	genérica

Fonte: elaboração própria. “Não reportado” indica que o estudo não traz a informação; na coluna de modelo, quando a ferramenta é identificada e a versão não, isso é registrado na própria célula.

Tabela A4. Robustez dos estudos selecionados.

Nº	Desenho e comparação	Pré-teste	Natureza do desfecho	N	Duração da intervenção
01	Quase-experimental, três grupos com controle	Sim	Teste conceitual de 16 itens	202	2 semanas
02	Quase-experimental, turmas intactas	Sim	Autorrelato em escalas Likert	80	4 atividades
03	Exploratório, sem comparação	Não	Padrões de interação	13 (12 nos questionári	Uma sessão (120 minutos)

Nº	Desenho e comparação	Pré-teste	Natureza do desfecho	N	Duração da intervenção
				os)	
04	Correlacional, sem comparação	Não	Pós-teste de 16 itens, sem pré-teste	102	2 semanas
05	Métodos mistos, sem intervenção	Não	Autorrelato	453	Não se aplica
06	Intervenção sem grupo de controle	Questionário pré-aula, sem medida inicial do desfecho	Autorrelato	286	44 horas
07	Análise documental com avaliação qualitativa de 5 professores	Não se aplica	Qualidade das atividades geradas	14 atividades	Não se aplica
08	Estudo de casos múltiplos	Não	Perfis de interação	6	Não reportada
09	Estudo de caso qualitativo	Não	Análise qualitativa das interações	19	Uma sessão (70 minutos)
10	Quase-experimental, comparação com pares humanos	Sim, sem equivalência inicial	Testes de conhecimento e de raciocínio	115	2 a 3 dias
11	Quase-experimental com controle	Sim, exceto no teste de física	Escala de autorrelato, teste de conhecimento em física	70	4 semanas
12	Pré-pós de grupo único	Sim	Autorrelato	46 (44 no pré, 43 no pós)	Não reportada
13	Comparação entre perfis, ambos com IAG	Sim, relatório inicial	Relatórios científicos, avaliados com rubricas	46	6 semanas
14	Medidas repetidas, todos com IAG	Não	Desempenho na tarefa	38	Uma sessão (95 minutos)
15	Piloto de grupo único	Sim	Percepções, sem medida de aprendizagem	53	2 aulas
16	Design participativo	Não	Sem medida direta de aprendizagem	16 professores ; 8 alunos no piloto	5 sessões
17	Quase-experimental com controle	Sim	Teste de desempenho de 26 itens	37	4 semanas