

Estratégias Pedagógicas e Impactos dos LLMs na Integração Interdisciplinar: Uma Revisão Sistemática

Martony Demes da Silva¹, André Castelo Branco Soares¹

¹Programa de Pós-Graduação Doutorado em Ciência da Computação
Associação UFMA/UFPI DCCMAPI/CCET
Universidade Federal do Piauí (UFPI)
Teresina – PI – Brasil

`martony.silva@ufpi.edu.br, andre.soares@ufpi.edu.br`

Abstract. *This Systematic Literature Review (SLR) investigates pedagogical strategies and the impacts of Large Language Models (LLMs) on interdisciplinary learning. Following the PRISMA protocol and rigorous quality auditing, the research analyzes the state of the art of AI integration in diverse educational contexts. Results indicate that the transition from vanilla models to specialized architectures, such as Retrieval-Augmented Generation (RAG) and Agentic systems, is essential to ensure academic rigor and mitigate factual hallucinations. The study highlights that AI orchestration optimizes instructional planning in PBL and STEAM models, while student agency is preserved through digital scaffolding and human-in-the-loop curation. It is concluded that the viability of AI-mediated interdisciplinarity depends on adopting high-fidelity architectures that balance technological automation with authentic cognitive engagement.*

Resumo. *Esta Revisão Sistemática da Literatura (RSL) investiga as estratégias pedagógicas e os impactos dos Modelos de Linguagem de Grande Escala (LLMs) na aprendizagem interdisciplinar. Fundamentada no protocolo PRISMA e em auditoria de qualidade rigorosa, a pesquisa analisa o estado da arte da integração da IA em contextos educacionais diversos. Os resultados indicam que a transição de modelos genéricos para arquiteturas especializadas, como a Geração Aumentada por Recuperação (RAG) e sistemas agênticos, é fundamental para assegurar o rigor acadêmico e mitigar alucinações factuais. Observou-se que a orquestração via IA otimiza o planejamento docente em modelos PBL e STEAM, enquanto a agência do estudante é sustentada por mecanismos de suporte digital gradativo e curadoria humana. Conclui-se que a viabilidade da interdisciplinaridade mediada por IA depende da adoção de arquiteturas de alta fidelidade que equilibrem a automação tecnológica com o esforço cognitivo autêntico.*

1. Introdução

O avanço da Inteligência Artificial Generativa (GenAI), impulsionado por Modelos de Linguagem de Grande Escala (LLMs), reconfigurou as práticas acadêmicas ao oferecer suporte flexível na resolução de problemas interdisciplinares [Hassan et al. 2025]. Ferramentas comerciais evoluíram para ambientes instrucionais capazes de processar volumes massivos de dados desestruturados [Bou Zakhem et al. 2025]. Sob a ótica da **Teoria da**

Carga Cognitiva, o suporte digital eficaz mitiga a “**infoxicação**” acadêmica — o colapso cognitivo pelo excesso de informações [Sweller 2011]. Esse fenômeno acentua-se frente à produção científica global de 3,3 milhões de registros anuais [Lazarev and Sedov 2025, Hassan et al. 2025], que inviabiliza a seleção manual de evidências em desenhos curriculares e satura a capacidade mental do estudante.

No entanto, a literatura aponta que o uso em estado puro (*vanilla*) de LLMs, a interação direta e padrão via interfaces de chat comerciais tradicionais, sem camadas adicionais de personalização técnica ou engenharia avançada de instruções, que é insuficiente para garantir o rigor acadêmico em domínios transversais. Evidencia-se a necessidade de transição para arquiteturas assistidas, como a **Geração Aumentada por Recuperação** (*Retrieval-Augmented Generation* – RAG, técnica que restringe as respostas da IA a uma base documental selecionada pelo docente) e **Sistemas Agênticos** (arquiteturas baseadas em componentes de software autônomos que operam coordenadamente para executar buscas e verificações).

Essas estruturas integram bases externas e interpretadores de código para assegurar a **fidelidade técnica** — alinhamento factual, rastreabilidade e eliminação de alucinações [Lu and Li 2025, Sreekanth et al. 2025]. No contexto instrucional, essa blindagem salvaguarda a segurança epistêmica do aprendiz e previne a fixação de conceitos errôneos na memória de longo prazo [Sweller 2011]. Em cenários controlados, o acoplamento dessas camadas assistidas eleva a acurácia de recuperação conceitual em até **14,8%** [Lu and Li 2025, Sreekanth et al. 2025], validando tais arquiteturas no suporte à avaliação formativa e ao autoestudo complexo.

Assim, os LLMs passam a atuar em uma **orquestração pedagógica de alta fidelidade**, que viabiliza o **scaffolding digital** (andaime), que é o fornecimento de suporte computacional adaptativo para gerenciar o esforço mental do estudante [Lazarev and Sedov 2025]. Essencial a esse fluxo é o **fading** (desvanecimento) gradativo do suporte tecnológico, requisito para prevenir o **viés de automação** (historicamente associado à inércia de pensamento), fenômeno no qual a facilidade de respostas prontas induz à passividade cognitiva e ao desengajamento intelectual do aprendiz [Yan et al. 2024, Rahimi 2025].

Contudo, há uma lacuna científica na análise de como essa mediação tecnológica reduz a sobrecarga de trabalho docente. Esse fator é crítico, dado que **70,8% os educadores** apontam a fragmentação de recursos instrucionais como o principal entrave para a viabilização prática de abordagens ativas como *STEAM* (*Science, Technology, Engineering, Arts, and Mathematics*) e *PBL* (*Problem-Based Learning / Aprendizagem Baseada em Problemas*) [Lu and Li 2025, Ravi et al. 2025].

Diante deste cenário, esta Revisão Sistemática da Literatura (RSL) analisa um *corpus* de 24 artigos primários, selecionados a partir de um levantamento rigoroso de **1.258 registros**, visando responder às seguintes Questões de Pesquisa (QPs):

- **QP1 (Fundação):** De que maneira os LLMs facilitam o diálogo entre áreas distintas e o desenvolvimento de competências transversais?
- **QP2 (Técnica):** Quais arquiteturas de IA (RAG, Sistemas Agênticos e Interpretadores de Código) apoiam a geração de conteúdo interdisciplinar com rigor acadêmico?

- **QP3 (Prática):** Como a orquestração de LLMs potencializa as diferentes práticas e abordagens pedagógicas interdisciplinares identificadas na literatura?
- **QP4 (Crítica):** Quais os principais obstáculos técnicos e pedagógicos identifi-
cados e de que maneira o risco de “inércia de pensamento” se manifesta nessas
interações?
- **QP5 (Solução):** De que forma as estratégias de *scaffolding digital* (**suporte tem-
porário**) e os mecanismos de *fading* (**retirada gradual desse suporte**) garantem
a agência do estudante e o esforço cognitivo autêntico?

O objetivo deste trabalho é consolidar um roteiro de evidências para o uso ético e integrado de LLMs na educação, superando visões instrumentalistas por meio de uma arquitetura de co-criação baseada em *scaffolding* (andaime) digital temporário. Esse arranjo estimula o **esforço cognitivo autêntico**, o engajamento mental profundo do estudante na resolução de problemas complexos, livre de tarefas operacionais redundantes [Li et al. 2025, Zlotnikova et al. 2025]. O **principal diferencial** desta RSL reside na articulação entre o desenho instrucional de modelos ativos (STEAM e PBL) e as ciências cognitivas. Ao correlacionar a mitigação da sobrecarga cognitiva dos aprendizes com soluções baseadas em arquiteturas avançadas de IA (sistemas agênticos e RAG), este mapeamento transforma a IA de ferramenta utilitária em um mecanismo estruturado de suporte ao desenho curricular interdisciplinar.

2. Metodologia

Esta pesquisa constitui uma **Revisão Sistemática da Literatura (RSL)**, método de pesquisa secundária que utiliza uma abordagem objetiva e auditável para identificar, avaliar e sintetizar as evidências de estudos primários sobre uma questão específica [Kitchenham and Charters 2007, Dermeval et al. 2020]. O delineamento seguiu as diretrizes de Kitchenham e Charters (2007), adaptadas para a área de Informática na Educação (IE) por Dermeval et al. (2020), estruturando-se em três fases: planejamento, condução e relatório [Dermeval et al. 2020, Kitchenham and Charters 2007].

2.1. Protocolo PICOC e Definição do Escopo

O planejamento desta RSL foi delimitado pelo framework **PICOC** (População, Intervenção, Comparação, Resultados e Contexto) para assegurar o rigor na formulação das Questões de Pesquisa (QPs) [Kitchenham and Charters 2007, Dermeval et al. 2020]. A população investigada compreende estudantes e docentes em contextos de Educação Básica e Ensino Superior, visando a integração pedagógica de LLMs para a superação de silos acadêmicos [Hassan et al. 2025, Li et al. 2025]. A Tabela 1 detalha os componentes do protocolo que orientaram a extração das evidências.

Tabela 1. Protocolo de Pesquisa Baseado no Framework PICOC

Elemento	Definição Metodológica
População	Contextos educacionais (Educação Básica e Ensino Superior).
Intervenção	Integração de LLMs e arquiteturas assistidas (RAG, Agentes).
Comparação	Uso simples (<i>vanilla</i>) vs. Orquestração de Sistemas Complexos.
Outcome	Interdisciplinaridade e Competências Transversais.
Contexto	Abordagens pedagógicas interdisciplinares (PBL, STEAM, IBL, ESP).

A delimitação do Contexto fundamentou-se na identificação de abordagens pedagógicas onde a interdisciplinaridade é um requisito intrínseco da prática, abrangendo desde modelos consolidados como **STEAM** e **PBL** até abordagens baseadas em investigação (*Inquiry-Based Learning* - IBL) e ensino dialógico [Hassan et al. 2025, Ravi et al. 2025]. A decisão de focar nestas orquestrações visou qualificar a síntese de evidências para cenários de alta complexidade cognitiva, onde a literatura aponta que o uso genérico (*vanilla*) de LLMs é insuficiente e pode induzir à “**inércia de pensamento**” [Yan et al. 2024, Lu and Li 2025].

Nesta perspectiva, os LLMs são analisados sob a lente do **scaffolding digital (andaime)**, suporte temporário que auxilia o aprendiz em tarefas além de sua capacidade individual, devendo ser gradualmente removido (*fading*) à medida que a competência interdisciplinar é adquirida [Lazarev and Sedov 2025, Yan et al. 2024]. Assim, o afunilamento no protocolo permitiu investigar como tecnologias de alta fidelidade, como **RAG** e **Sistemas Agênticos**, atuam como esse tecido conectivo essencial para sustentar o diálogo entre áreas díspares e garantir o rigor acadêmico exigido por esses cenários [Lu and Li 2025, Sreekanth et al. 2025, Rahimi 2025].

2.2. Estratégia de Busca e Triagem de Estudos

A RSL iniciou-se com busca automatizada nas bases **ACM Digital Library** e **Scopus**, cobrindo a janela temporal de 2022 a 2025. Esse conjunto de bases foi selecionado por sua abrangência e capacidade de indexação cruzada na Computação Aplicada à Educação: a ACM DL assegura a centralidade na Ciência da Computação, enquanto a Scopus atua como indexador unificado de periódicos, conferências internacionais e publicações-chave da SBC, além de permitir mapear estudos de *venues* importantes da IEEE e Elsevier via metadados. Essa abordagem funcionou como filtro de qualidade metodológica, minimizando redundâncias e garantindo estudos revisados por pares, cujo afunilamento delineou o escopo final focado no biênio 2024-2025.

A construção das *strings* de busca seguiu um processo iterativo de refinamento de sinônimos para maximizar a revocação das evidências [Dermeval et al. 2020, Kitchenham and Charters 2007], retornando um montante inicial de **1.258 registros** ($n_{ACM} = 613$ e $n_{Scopus} = 645$). Para a ACM DL, utilizou-se a consulta: ("Large Language Models"OR "Generative AI"OR "ChatGPT") AND ("Interdisciplinary"OR "Transversal Competencies") AND ("Pedagogical Strategies"OR "PBL"). Para a Scopus, aplicou-se: TITLE-ABS-KEY(("LLM"OR "Generative Artificial Intelligence") AND ("Interdisciplinarity"OR "Critical Thinking") AND ("Active Methodologies")). A lista completa dos registros e metadados extraídos está disponível publicamente¹.

Os dados foram importados para a ferramenta **Rayyan**, que removeu 12 duplicatas, restando 1.246 artigos. Para assegurar a integridade do processo, aplicou-se a triagem independente por dois revisores em **Modo Cego (Blind Mode)** para mitigar o viés de seleção [Dermeval et al. 2020, Kitchenham and Charters 2007]. Nessa fase, 1.219 artigos foram excluídos por não atenderem aos critérios preestabelecidos, e as divergências pon-

¹Planilha de registros da RSL disponível em: <https://docs.google.com/spreadsheets/d/1D9z2DKRr1wj3CW1N4KNk7SB2MQUAzmsQ4xRXRdFkBA/edit?usp=sharing>

tuais foram resolvidas por reuniões de consenso. Conforme ilustrado no fluxo PRISMA e de decisão da Figura 1, 27 artigos foram selecionados para leitura integral. Aplicou-se também *snowballing* retrospectivo nas referências desses 27 trabalhos, confirmando a alta cobertura da busca inicial sem a inclusão de novos artigos. Por fim, após submeter os pré-selecionados à avaliação de qualidade (Seção 2.3), com a exclusão de 3 trabalhos, o *corpus* final consolidou-se em **24 estudos primários**.

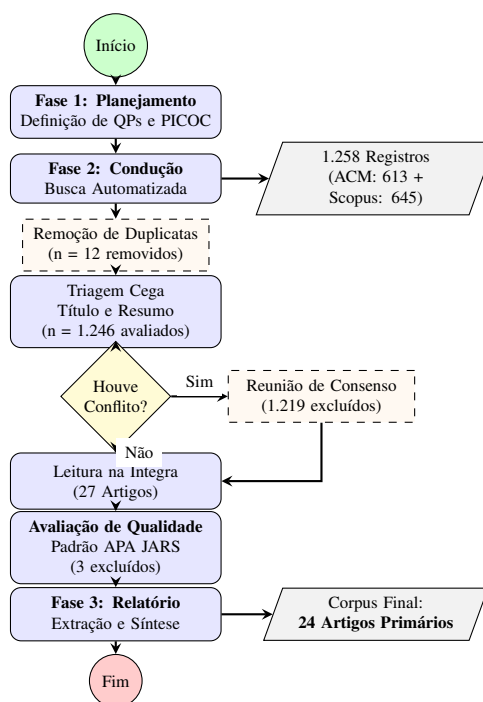


Figura 1. Metodologia detalhada da RSL integrando o fluxo PRISMA e critérios de qualidade com as respectivas contagens de artigos.

2.3. Avaliação de Qualidade (AQ)

Os 27 artigos pré-selecionados foram auditados por dois revisores independentes (com consenso para discrepâncias) sob o padrão **APA JARS** [Appelbaum et al. 2018, Dermeval et al. 2020]. Avaliaram-se cinco critérios (objetivos, método, instrumentos, resultados e discussão) com notas de 0 a 1, adotando-se o **limiar de 50% de conformidade** (mínimo de 2,5 pontos) para inclusão, conforme validado na área [Dermeval et al. 2020, Raabe and Bombasar 2020]. Três estudos foram excluídos por pontuação insuficiente em dados empíricos sobre arquiteturas complexas, consolidando o *corpus* em 24 artigos. A planilha com as notas individuais está disponível publicamente².

2.4. Extração e Síntese de Dados

A extração de dados foi operacionalizada por meio de um formulário estruturado e de um dicionário de dados específico para capturar a complexidade da orquestração pedagógica [Dermeval et al. 2020]. O formulário visou mapear seis dimensões centrais: (i)

²Auditoria de Qualidade detalhada (24 artigos): <https://docs.google.com/spreadsheets/d/1D9z2DKRr1wj3CW1N4KNk7SB2MQUAzmsQ4xRXRdFkbA/edit?gid=1347801051#gid=1347801051>

delineamento do processo (como a prática é feita); (ii) arquitetura técnica (uso de RAG, Agentes ou *vanilla*); (iii) domínios interdisciplinares; (iv) tipos de dados multimodais; (v) objetivos pedagógicos (ex: mitigação da inércia de pensamento); e (vi) artefatos entregues [Lu and Li 2025, Hassan et al. 2025].

A síntese seguiu uma abordagem de **análise temática indutiva**, categorizando as evidências para responder às cinco Questões de Pesquisa (QPs) [Hassan et al. 2025, Silva et al. 2023]. O processo envolveu a triangulação de dados técnicos e pedagógicos em três eixos analíticos:

1. **Fidelidade Técnica:** cruzamento entre arquiteturas assistidas e acurácia factual em áreas STEM [Lu and Li 2025];
2. **Agência e Mediação:** análise da relação entre estratégias de *scaffolding/fading* e o desenvolvimento do esforço cognitivo autêntico [Yan et al. 2024, Rahimi 2025];
3. **Viabilidade Pedagógica:** avaliação do impacto da GenAI na redução da sobrecarga docente e na escalabilidade de projetos interdisciplinares [Ravi et al. 2025, Sreekanth et al. 2025].

Essa metodologia permitiu identificar não apenas as tendências atuais (pontos aquecidos), mas também as lacunas na literatura sobre o suporte adaptativo em larga escala [Dermeval et al. 2020, Batista et al. 2024].

3. Resultados e Análise

A síntese das evidências extraídas dos 24 estudos primários selecionados revela que os Modelos de Linguagem de Grande Porte (LLMs) atuam como um elemento de integração interdisciplinar. A análise promovida nesta seção supera a mera descrição de funcionalidades de engenharia de software, concentrando-se em como tais arquiteturas interagem diretamente com constructos da Psicologia Cognitiva, especificamente na modulação da carga de trabalho mental durante a resolução de problemas complexos em cenários STEAM e PBL.

3.1. Visão Geral e Perfil Demográfico do Corpus

A análise cronológica identifica um pico de maturidade entre 2024 e 2025, período em que o foco das publicações migrou da exploração genérica (*vanilla*) para a validação de arquiteturas de alta fidelidade, como **RAG e Sistemas Agênticos** [Batista et al. 2024, Lu and Li 2025]. Observa-se um interesse global pelo tema, com predominância de pesquisas conduzidas na China, Estados Unidos e Brasil [Silva et al. 2023, Li et al. 2025].

Quanto ao contexto educacional, 67% das investigações concentram-se no **Ensino Superior**, ambiente que exige maior rigor lógico e onde os silos disciplinares são mais acentuados [Hassan et al. 2025]. A Tabela 2 consolida a distribuição desses estudos.

Tabela 2. Perfil Demográfico e Metodológico do Corpus Final (n = 24)

Método de Pesquisa	Qtd.	Contexto Educacional	%
Estudo de Caso	10	Ensino Superior	67%
Experimento Controlado	8	Educação Básica / Técnico	25%
Proposta de Framework	6	Educação Corporativa	8%

A predominância do **Estudo de Caso (42%)** reflete a necessidade da literatura em observar o fenômeno da co-criação humano-IA em ambientes reais, enquanto os **Experimentos Controlados (33%)** focam na validação estatística da precisão algorítmica em domínios verticais [Lu and Li 2025, Sreekanth et al. 2025]. Esse panorama demográfico estabelece a base empírica para a análise quantitativa e qualitativa das Questões de Pesquisa formuladas no protocolo.

3.2. Mediação de LLMs no Diálogo Interdisciplinar (QP1)

As evidências coletadas indicam que os LLMs atuam na conversão semântica de linguagens técnicas, simplificando a transferência de conhecimento entre campos díspares como tecnologia, economia e artes criativas [Hassan et al. 2025, Li et al. 2025]. Fundamentado na **Teoria da Carga Cognitiva**, o uso de LLMs reduz o esforço mental extrínseco ao processar e organizar volumes massivos de informação (infoxicação), permitindo que o estudante direcione seus recursos cognitivos para processos de análise e criação de alto nível [Sweller 2011, Bou Zakhem et al. 2025].

Quatro práticas principais foram mapeadas no *corpus*: (i) **Framework AI Sci-Brief** [Lazarev and Sedov 2025], atuando como estratégia de *onboarding* em investigação (*IBL*) ao gerar sínteses científicas que reduzem a "página em branco"; (ii) **CustomAI-zEd** [Pua et al. 2025], que customiza tarefas de programação com exemplos práticos de biologia e economia; (iii) **Síntese Multimodal** [Shiny Karunya et al. 2025], integrando LLMs e difusão visual para reinterpretação de obras literárias; e (iv) **Diálogo Transversal** [Hassan et al. 2025], que viabiliza a geração automatizada de estudos de caso intersetoriais. Nesse âmbito, essa capacidade de re-enquadramento de problemas (*problem reframing*) reduz o isolamento disciplinar ao recontextualizar conceitos abstratos (ex: estatística) sob o escopo de domínios aplicados como direito e medicina [Einarsson et al. 2024, Hassan et al. 2025].

3.3. Arquiteturas de IA e Rigor Acadêmico (QP2)

Um achado crítico desta revisão é a transição obrigatória de modelos *vanilla* (uso direto via chat comercial) para arquiteturas assistidas e acopladas, mitigando desvios informacionais em domínios científicos [Lu and Li 2025, Sreekanth et al. 2025]. Conforme ilustrado na Figura 2, a integração de **RAG (Geração Aumentada por Recuperação)** com **Interpretadores de Código** eleva a acurácia em até **14,8%** no dataset TQA (livros didáticos), mitigando alucinações factuais em áreas STEM [Lu and Li 2025].

Contudo, sob a ótica desta revisão, é necessário moderar as pontes inferenciais entre métricas de engenharia e desfechos educacionais: embora testes empíricos em cenários computacionais controlem alucinações semânticas, tais dados aferem o desempenho puramente técnico do sistema de Processamento de Linguagem Natural (NLP), e não necessariamente um ganho direto ou automático na eficácia pedagógica da aprendizagem interdisciplinar dos estudantes. A robustez da arquitetura funciona como uma salvaguarda epistêmica, uma pré-condição de fidelidade e segurança dos dados, mas o sucesso do processo instrucional permanece condicionado à qualidade do desenho de mediação pedagógica adotado.

A partir dessa diferenciação, a literatura mapeada consolida três pilares arquiteturais de suporte de dados para aplicação educacional:

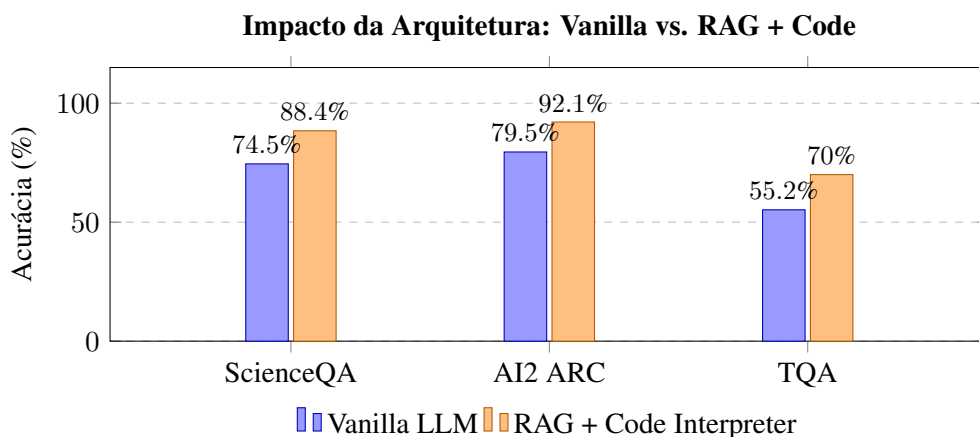


Figura 2. Ganho de precisão acadêmica por arquitetura. Fonte: Adaptado de Lu e Li (2025) [Lu and Li 2025].

1. **Sistemas Agênticos Modulares:** O framework *Agentic AI Quiz* utiliza agentes autônomos de busca, seleção e verificação para atingir **93,33% de acurácia** na geração técnica de itens de avaliação complexos [Sreekanth et al. 2025].
2. **Context Caching:** O uso de *caches* de contexto de longo prazo reduziu o tempo de computação de **85s para 21s** por item, viabilizando a escalabilidade operacional da automação para grandes turmas interdisciplinares [Sreekanth et al. 2025].
3. **KAG (Knowledge-Augmented Generation):** Evolução do RAG que integra LLMs a **Grafos de Conhecimento (KG)**, garantindo raciocínio lógico estruturado e feedback explicável em cursos complexos, como a Teoria TRIZ de inovação [Gao et al. 2025].

3.4. Orquestração e Práticas Pedagógicas Ativas (QP3)

Os dados coletados demonstram que a potencialização de práticas interdisciplinares não é um efeito colateral genérico da tecnologia, mas depende diretamente da **maturidade da orquestração técnica** aplicada ao desenho instrucional [Lu and Li 2025, Gao et al. 2025]. A IA atua como um requisito de viabilidade pedagógica ao mitigar o problema crônico de fragmentação de recursos, relatado por 70,8% dos docentes no contexto educacional [Heilala et al. 2023].

No ambiente de **Aprendizado Baseado em Projetos (PBL)** e no modelo **STEAM**, a automação de rubricas e roteiros estruturados reduziu a carga de trabalho burocrático e operacional docente em até **40 horas por ciclo** de planejamento [Ravi et al. 2025]. Isso permite que o professor transicione do papel de gestor administrativo para o de mentor ativo.

No contexto de *problem reframing*, os LLMs demonstraram viabilidade ao adaptar conceitos estatísticos abstratos para Direito e Biologia, preservando 77,1% das premissas matemáticas originais [Einarsson et al. 2024]. Contudo, esse dado advém de um piloto inicial com preferências mistas de estudantes, onde os ganhos educacionais atingiram 73,6% e métricas de satisfação não garantem proficiência conceitual estável. Em contrapartida, Lu e Li (2025) [Lu and Li 2025] apontam a síntese cross-disciplinar como um desafio em aberto de alta complexidade. Assim, o impacto prático em abordagens ativas não decorre da mera automação operacional, mas da capacidade crítica humana em

mediar as lacunas cognitivas inerentes aos modelos [Gao et al. 2025, Heilala et al. 2023].

3.5. Obstáculos Técnicos, Pedagógicos e Inércia Cognitiva (QP4)

O principal risco pedagógico identificado no *corpus* é a “**inércia de pensamento**” (*cognitive offloading*), fenômeno no qual a facilidade de obter respostas prontas induz o estudante ao desengajamento cognitivo e à aceitação passiva de soluções prontas, atrofiando o pensamento crítico e analítico [Yan et al. 2024, Lu and Li 2025]. Esse risco torna-se crítico quando a tecnologia é implantada de forma simplista como um “oráculo” factual, e não como um parceiro dialógico estruturado [Rahimi 2025].

Os dados apontam três obstáculos fundamentais que comprometem a viabilidade das propostas: (i) **Qualidade Desigual dos Artefatos (60,4%)**: a inconsistência nos materiais gerados de forma livre gera uma sobrecarga de curadoria para o professor, anulando o ganho de tempo caso não sejam adotadas arquiteturas baseadas em RAG [Heilala et al. 2023, Lu and Li 2025]; (ii) **Opacidade e Viés Algorítmico**: o viés eurocêntrico e a falta de transparência nos dados de treinamento dos modelos demandam a inclusão urgente de um **letramento crítico em IA** no currículo dos estudantes [Zlotnikova et al. 2025, Batista et al. 2024]; e (iii) **Dependência de Andaimos**: a ausência de estratégias explícitas de **fading (desvanecimento)** transforma o suporte digital em uma muleta cognitiva permanente, bloqueando a autonomia intelectual do aprendiz [Yan et al. 2024].

3.6. Scaffolding Digital e Estratégias de Fading (QP5)

As evidências indicam que a agência e o protagonismo do estudante (definidos pela autonomia e poder de escolha) são preservados quando o design técnico adota o modelo de **scaffolding digital** com mecanismos de **fading gradativo** do suporte [Lazarev and Sedov 2025, Yan et al. 2024, Rahimi 2025]. Sob a ótica da **Teoria da Carga Cognitiva**, o desvanecimento planejado garante que a IA absorva apenas o esforço mental extrínseco (como a formatação e a busca bruta de dados), preservando intacta a carga cognitiva pertinente (*germane load*) voltada à reflexão e à síntese interdisciplinar autêntica [Sweller 2011, Bou Zakhem et al. 2025].

A literatura aponta três pilares para sustentar essa agência humana na interação:

- **Questionamento Socrático Dialógico**: Sistemas configurados para atuar como um Outro Mais Experiente (*MKO*) fornecem pistas adaptativas e perguntas reflexivas, forçando o aluno a atuar como curador, em vez de receptor passivo [Rahimi 2025, Hassan et al. 2025].
- **Humano no Loop (HITL)**: A inclusão do estudante como validador obrigatório de cada etapa intermediária eleva a acurácia de sistemas de busca acadêmica para **99,09%**. Esse ganho é fruto de um processo iterativo de expansão mediado pela crítica humana, eliminando o desprendimento cognitivo [Zhong et al. 2025, Sreekanth et al. 2025].
- **Prompting Estruturado via RIE**: O uso do framework RIE (Papel, Instruções e Objetivo Final) capacita o aluno a orquestrar a IA como um parceiro de co-criação síncrono, retendo o controle metodológico da atividade [Ravi et al. 2025].

Para transparência metodológica, a Tabela 3 consolida achados e estimativas da literatura. Ressalta-se que esses dados advêm de estudos metodologicamente heterogêneos, não devendo ser interpretados como uma meta-síntese convergente. Nota-se que,

embora ganhos de acurácia técnica e redução de latência apresentem percentuais expressivos, o ganho real de aprendizagem dos estudantes resultou em um tamanho de efeito pequeno (Cohen’s $d = 0,27$) em ambientes controlados [Sreekanth et al. 2025], exigindo cautela ao generalizar os benefícios dos LLMs na educação.

Tabela 3. Consolidado de Achados Heterogêneos por Estudo na RSL

Métrica / Variável Avaliada	Valor / Impacto Constatado	Estudo de Referência
Ganho de Acurácia (RAG+Code)	+14,8% (Dataset TQA - Desempenho PLN)	Lu e Li (2025) [Lu and Li 2025]
Sucesso em <i>Problem Reframing</i>	77,1% de preservação teórica (Piloto)	Einarsson et al. (2024) [Einarsson et al. 2024]
Precisão em Sistemas Agênticos	93,33% (Geração de Itens MCQ)	Sreekanth et al. (2025) [Sreekanth et al. 2025]
Redução de Latência (<i>Cache</i>)	85s → 21s por item gerado	Sreekanth et al. (2025) [Sreekanth et al. 2025]
Economia de Tempo Docente	40 horas/ciclo de planejamento (PBL)	Ravi et al. (2025) [Ravi et al. 2025]
Tamanho de Efeito (<i>Cohen’s d</i>)	0,27 (Impacto Prático na Aprendizagem)	Sreekanth et al. (2025) [Sreekanth et al. 2025]

4. Discussão

Os resultados sintetizados nesta Revisão Sistemática da Literatura (RSL) indicam que a integração de LLMs na educação transcendeu a fase de experimentação isolada, consolidando-se como um campo de investigação sobre a **orquestração pedagógica de alta fidelidade** [Hassan et al. 2025, Batista et al. 2024]. A discussão a seguir triangula as evidências sob três eixos fundamentais, extraídos do cruzamento entre arquiteturas técnicas, processos instrucionais e produtos entregues.

4.1. Fidelidade Técnica e a Segurança da Ponte Interdisciplinar (QP1 e QP2)

O cruzamento analítico entre o **Conteúdo Abordado** e a **Arquitetura de LLM** revela que a eficácia do diálogo entre áreas é diretamente dependente do rigor factual da arquitetura escolhida [Lu and Li 2025].

- **Segurança em Exatas (STEM):** A literatura demonstra que em domínios que exigem precisão lógica (Engenharia, Física e Cálculo), o uso de modelos *vanilla* é insuficiente. A transição para arquiteturas de **RAG + Interpretador de Código** é um requisito funcional, elevando a acurácia em até **14,8%** [Lu and Li 2025].
- **Ponte com Humanidades:** Em práticas que unem literatura e tecnologia, a orquestração foca na **Síntese Multimodal**, integrando linguagens verbais e visuais para fomentar o letramento crítico e a reinterpretação criativa [Shiny Karunya et al. 2025].
- **Dedução Analítica:** A potencialização da interdisciplinaridade não é um efeito genérico da IA, mas sim da maturidade técnica. Sistemas agênticos modulares, que atingem **93,33% de precisão**, garantem que o "tecido conectivo" entre disciplinas seja baseado em informações confiáveis, conferindo validade acadêmica aos artefatos gerados [Sreekanth et al. 2025, Gao et al. 2025].

4.2. Mediação Instrucional e o Combate à Inércia Cognitiva (QP4 e QP5)

Um dos achados mais críticos do *corpus* é o risco da “**inércia de pensamento**”, onde a facilidade de obtenção de respostas prontas induz ao desprendimento intelectual do estudante [Yan et al. 2024]. A proteção da agência humana exige um design instrucional deliberado fundamentado em dois pilares:

1. **Scaffolding Digital com Fading:** A IA deve atuar como um andaime temporário que auxilia na gestão da carga cognitiva extrínseca, mas que deve sofrer um desvanecimento (*fading*) planejado para que o aluno assuma o esforço cognitivo autêntico [Lazarev and Sedov 2025, Yan et al. 2024].
2. **MKO Dialógico (Outro Mais Experiente):** A literatura recomenda que a IA assuma um papel socrático (tutor) em vez de oráculo. Ao utilizar pistas adaptativas e contra-argumentos, o sistema força o estudante a atuar como curador e validador crítico [Rahimi 2025, Hassan et al. 2025].

A integração do **Humano no Loop (HITL)** na co-criação eleva a acurácia da geração de conhecimento para **99,09%** [Zhong et al. 2025]. Apesar dessa vantagem, é preciso atenuar o viés de confirmação tecnológica: em tarefas de reformulação de problemas (*Problem Reframing*), ferramentas de IA não demonstraram utilidade incremental significativa frente ao esforço humano exclusivo, sinalizando que a orquestração complexa pode gerar sobrecarga técnica sem contrapartida pedagógica proporcional.

A partir dessa articulação teórica, consolida-se uma **Agenda de Pesquisa** emergente para guiar o desenho de futuros sistemas educacionais baseados em GenAI:

- **Modelagem do Fading Automatizado:** Detecção algorítmica do momento ideal para a retirada gradual do suporte computacional.
- **Métricas de Esforço Cognitivo:** Mensuração da carga de trabalho mental pertinente por meio da análise de logs de interação.
- **Protocolos HITL Estruturados:** Definição de *checkpoints* obrigatórios de curadoria crítica humana sobre o ecossistema.
- **Mediação Socrática Ativa:** Parametrização de agentes focados em contra-argumentação dialógica e refutação analítica.

4.3. Viabilidade Pedagógica e Escalabilidade de Metodologias Ativas (QP3)

A análise do processo instrucional em relação ao **Ganho de Produtividade** responde à viabilidade de modelos como **PBL e STEAM**. Tradicionalmente, a fragmentação de recursos é relatada por 70,8% dos docentes como o principal entrave para práticas interdisciplinares [Heilala et al. 2023].

- **Eficiência no Planejamento:** A orquestração via LLMs automatiza a geração de rubricas e roteiros complexos, economizando até **40 horas por ciclo docente** [Ravi et al. 2025].
- **Re-enquadramento de Problemas:** A capacidade de traduzir conceitos técnicos para contextos práticos (*problem reframing*) atinge **77,1% de sucesso** na preservação do rigor matemático, permitindo que o professor migre do papel de criador de conteúdo para o de mentor [Einarsson et al. 2024, Pua et al. 2025].

Conclui-se que a redução da latência cognitiva na geração de itens (de 85s para 21s) é o fator técnico que torna sustentável a implementação de currículos transversais em larga escala [Sreekanth et al. 2025, Ravi et al. 2025].

4.4. Implicações Éticas e Letramento Crítico

A integridade acadêmica depende da superação da opacidade algorítmica e dos vieses eurocêntricos [Zlotnikova et al. 2025]. Esta RSL reforça que a integração interdisciplinar requer um novo **Letramento Crítico em IA**, onde estudantes e docentes compreendam as limitações das arquiteturas agênticas para garantir que a tecnologia sustente a diversidade cultural e o rigor científico [Batista et al. 2024, Zlotnikova et al. 2025].

4.5. Ameaças à Validade e Limitações do Estudo

A principal ameaça refere-se ao viés de seleção pela restrição inicial às bases Scopus e ACM Digital Library, com potencial omissão de repositórios como o IEEE Xplore. Essa limitação foi mitigada pela capacidade de indexação cruzada da Scopus — que engloba metadados de veículos da IEEE e Elsevier — e pelo uso de uma *string* abrangente. Adicionalmente, o *snowballing* retrospectivo nos 27 artigos lidos na íntegra atuou como auditoria interna: embora não tenha incluído novos trabalhos (já capturados na busca automatizada), o procedimento confirmou a alta taxa de revocação e a robustez do levantamento.

4.6. Diretrizes Práticas por Nível de Ensino

Recomenda-se para a **Educação Básica** o uso da IA focado em *scaffolding* com *fading* rápido e blindagem via RAG institucional para evitar desvios factuais sob tutoria socrática. Já no **Ensino Superior**, indica-se a orquestração de Sistemas Agênticos Multimodais com o estudante como validador ativo no loop (*HITL*), priorizando o letramento crítico e a curadoria em projetos PBL.

5. Conclusão

Esta RSL evidenciou a transição de interações genéricas (*vanilla*) para arquiteturas de alta fidelidade como **RAG e Sistemas Agênticos** [Lu and Li 2025, Sreekanth et al. 2025], que atuam como infraestrutura computacional para reduzir barreiras da integração interdisciplinar [Li et al. 2025, Hassan et al. 2025].

A eficácia pedagógica da GenAI depende de uma **orquestração técnica e instrucional** com *scaffolding* digital multinível [Lazarev and Sedov 2025]. Camadas assistidas por interpretadores de código elevam a acurácia factual em até **14,8%** e, com validação crítica do estudante (*Human-in-the-loop*), a consistência atinge **99,09%** [Lu and Li 2025, Zhong et al. 2025]. Para mitigar a fragmentação de recursos (relatada por 70,8% dos docentes) e o **viés de automação**, a implementação exige o desvanecimento gradual (*fading*) e o letramento crítico em IA [Rahimi 2025, Yan et al. 2024, Lu and Li 2025].

O diferencial deste estudo é demonstrar que o sucesso de modelos ativos (PBL e STEAM) via LLMs requer arquiteturas que mitiguem a sobrecarga cognitiva extrínseca, economizando até **40 horas por ciclo de planejamento** docente [Ravi et al. 2025]. Como trabalhos futuros, sugere-se investigar métricas de **“esforço cognitivo autêntico”** e desenvolver tutores socráticos agênticos. Em suma, viabilizar LLMs no ensino exige migrar de interações puras para frameworks centrados na agência do estudante e no *fading* do suporte digital.

Declaração de Uso de Inteligência Artificial

Durante a preparação deste trabalho, os autores utilizaram *Google Gemini* para auxiliar na revisão gramatical e na estruturação do código TikZ da Figura 1. Toda a análise crítica dos dados, síntese teórica e conclusões metodológicas foram desenvolvidas exclusivamente por inteligência humana, assumindo os autores total responsabilidade pelo conteúdo publicado.

Referências

- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., and Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The apa publications and communications board task force report. *American Psychologist*, 73(1):3–25.
- Batista, J., Mesquita, A., and Carnaz, G. (2024). Generative ai and higher education: Trends, challenges, and future directions from a systematic literature review. *Information*, 15(11):676.
- Bou Zakhem, N., Bou Diab, M., and Tahan, S. (2025). A cross-disciplinary academic evaluation of generative ai models in hr, accounting, and economics: Chatgpt-5 vs. deepseek. *Administrative Sciences*, 15(11):412.
- Derneval, D., Coelho, J. A. P. M., and Bittencourt, I. I. (2020). Mapeamento sistemático e revisão sistemática da literatura em informática na educação. In Jaques, P. A., Siqueira, S., Bittencourt, I., and Pimentel, M., editors, *Metodologia de Pesquisa Científica em Informática na Educação: abordagem quantitativa*, chapter 3, pages 1–24. SBC, Porto Alegre.
- Einarsson, H., Lund, S. H., and Jónsdóttir, A. H. (2024). Application of chatgpt for automated problem reframing across academic domains. *Computers and Education: Artificial Intelligence*, 6:100194.
- Gao, C., Liu, Z., Li, X., and Zhang, L. (2025). Knowledge-augmented generation technology enabling boppps teaching: An example from the triz course. In *Procedia Computer Science*, volume 266, pages 963–970. Elsevier.
- Hassan, A., Tayan, O., Almiladi, A., and Sen, A. A. A. (2025). Ai-enabled assessed learning from an interdisciplinary educational perspective: An empirical case study using chatgpt llm. *International Journal of Advanced Computer Science and Applications*, 16(11):150–165.
- Heilala, J., Shibani, A., and de Freitas, A. G. (2023). The requirements for heutagogical attunement within steam education. *International Journal of Emerging Technologies in Learning*, 18(16):19–35.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University.
- Lazarev, A. and Sedov, D. (2025). Ai scibrief as a gateway to research: A framework for onboarding students into new research areas. In *Proceedings of the 5th International Conference on Technology Enhanced Learning in Higher Education (TELE)*, pages 365–369. IEEE.
- Li, Y., Geng, Y., and Yu, X. (2025). Research on a generative ai-enabled system for cultivating interdisciplinary innovative thinking and competencies in applied technology talents. In *Proceedings of the 13th International Conference on Orange Technology (ICOT)*. IEEE.
- Lu, J. and Li, J. (2025). A novel framework for educational q&a: Leveraging rag and code interpreters for knowledge retrieval and logical computation. *PLoS ONE*, 20(12):e0337361.

- Pua, L. X., Ramesh, R., Natarajan, P., and Iyer, G. N. (2025). Customaized: Bridging interdisciplinary gaps in ai education with customized content using llms. In *Proceedings of the IEEE Global Engineering Education Conference (EDUCON)*, pages 410–417. IEEE.
- Raabe, A. L. A. and Bombasar, J. R. (2020). Mensuração e testes em informática na educação. In Jaques, P. A., Siqueira, S., Bittencourt, I., and Pimentel, M., editors, *Metodologia de pesquisa científica em Informática na Educação: abordagem quantitativa*, chapter 5, pages 34–50. SBC, Porto Alegre.
- Rahimi, M. (2025). Synergising dialogic teaching with competencies-trained genai dialoguing for critical thinking and communication competencies. *Journal of University Teaching and Learning Practice*, 22(1):45–68.
- Ravi, P., Masla, J., Kakoti, G., Lin, G. C., Anderson, E., Taylor, M., Ostrowski, A. K., Breazeal, C., Klopfer, E., and Abelson, H. (2025). Co-designing large language model tools for project-based learning with k12 educators. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM.
- Shiny Karunya, T., Grace, V. B., Kharmawlong, A., Abisha, A. P., Preha, C., and Wilson, D. D. (2025). Illustrations and interdisciplinary pedagogical implications in brian jacques’ selected work using midjourney, chatgpt and dall-e. *International Research Journal of Multidisciplinary Scope*, 6(3):1067–1079.
- Silva, L. et al. (2023). Chatgpt na educação: Um mapeamento sistemático da literatura. In *Anais do XXXIV Simpósio Brasileiro de Informática na Educação (SBIE 2023)*, pages 1245–1256. SBC.
- Sreekanth, D., Gopi, S., and Dehbozorgi, N. (2025). Agentic ai quiz-based learning system: Enhancing mcq generation via long-context cached retrieval-augmented generation. In *Proceedings of the Frontiers in Education Conference (FIE)*, pages 215–222. IEEE.
- Sweller, J. (2011). *Cognitive Load Theory*. Springer.
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., and Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1):1–25.
- Zhong, C., Ye, F., Wang, Z., Jigeer, A., and Zhan, Z. (2025). Interdisciplinary-qg: An llm-based framework for generating high-quality interdisciplinary test questions with knowledge graphs and chain-of-thought reasoning. In *Proceedings of the 14th International Conference on Educational and Information Technology (ICEIT)*, pages 89–96. IEEE.
- Zlotnikova, I., Hlomani, H., Mokgetse, T., and Bagai, K. (2025). Establishing ethical standards for genai in university education: a roadmap for academic integrity and fairness. *Journal of Information, Communication and Ethics in Society*, 23(2):188–217.