

Generative Artificial Intelligence e Large Language Models na Perspectiva do Educador e do Gestor Educacional: Uma Revisão Sistemática da Literatura

Stéfano Schwenck Borges Vale Vita¹, Fabiano Azevedo Dorça²

¹Universidade de Uberaba (Uniube)
Uberlândia – MG – Brasil

²Faculdade de Computação – Universidade Federal de Uberlândia (UFU)
Uberlândia – MG – Brasil

stefanoborges@gmail.com, fabianodor@ufu.br

Abstract. *The recent literature on AI in Education focuses 98% of studies on student personalization, neglecting educators and managers. Following the PRISMA 2020 and HCAI-SLR frameworks, this Systematic Literature Review analyzes 84 high-impact studies (Jan/2023–May/2025) to investigate Generative AI in supporting pedagogical and administrative decision-making. Results reveal that 62% of pedagogical interventions use Chain-of-Thought with learning theories such as Gagné and Bloom, while 48% of management applications focus on accreditation and compliance automation. The study highlights gaps in empirical validation and proposes a transition toward "Hybrid Intelligence".*

Keywords: *Artificial Intelligence in Education, Generative AI, Systematic Review, Hybrid Intelligence, Human-Centered AI.*

Resumo: *A recente literatura de IA na Educação foca 98% dos estudos na personalização para o aluno, negligenciando educadores e gestores. Esta Revisão Sistemática da Literatura seguiu o protocolo PRISMA 2020 e HCAI-SLR (Human-Centered AI for Systematic Literature Reviews) analisa 84 trabalhos de alto impacto (jan/2023–mai/2025) para investigar a IA Generativa no apoio à decisão docente e administrativa. Constatou-se que 62% das intervenções pedagógicas usam Chain-of-Thought com teorias como Gagné e Bloom, e 48% das aplicações de gestão focam em automação de acreditação e compliance. O estudo aponta lacunas na validação empírica e propõe a transição para uma "Inteligência Híbrida"*

Palavras-Chave *Inteligência Artificial na Educação, IA Generativa, Revisão Sistemática, Inteligência Híbrida, IA Centrada no Humano.*

1. Introdução

A integração de tecnologias digitais no Ensino Superior não é um fenômeno inédito, mas a emergência da Inteligência Artificial Generativa (GenAI) e dos Grandes Modelos de Linguagem (LLMs) representa uma ruptura praxiológica (Jia *et al.*, 2025). Diferentemente dos Sistemas de Gestão da Aprendizagem (LMS) passivos, a GenAI

introduz capacidade agêntica e criativa. No entanto, uma análise crítica do estado da arte revela um certo desequilíbrio alarmante. Revisões de mapeamento sistemático da literatura recentes indicam que a maioria das pesquisas em Inteligência Artificial (IA) aplicada na área educacional foca no aluno, no seu desempenho, retenção e personalização do ensino, deixando aplicações visando o educador e o gestor como lacunas em estudos recentes na literatura. Esta negligência é perigosa: se a IA deve transformar a educação, ela precisa primeiro empoderar aqueles que desenham a arquitetura da experiência de aprendizagem, como o docente, e aqueles que garantem sua sustentabilidade, isto é, o gestor.

Para endereçar as lacunas citadas acima, este trabalho estabelece realizar uma revisão sistemática onde poderá ser visto as evidências sobre como a GenAI está sendo operacionalizada para apoiar, aumentar e redefinir as funções do educador, no design instrucional e do Gestor Universitário, na governança e estratégia, no período de 2023 a 2025. A revisão sistemática abordando este tema seguirá a metodologia Prisma, declarada em (Page *et al.*, 2021).

Além disso, como objetivos específicos deste trabalho destacam-se: Mapear as estratégias pedagógicas, identificando como educadores utilizam técnicas de engenharia de *prompt*, como *Chain-of-Thought*, para traduzir teorias de aprendizagem clássicas, especificamente Gagne e Bloom, em intervenções práticas de IA; Avaliar o Impacto na Gestão, examinando como soluções de GenAI estão sendo aplicadas para automatizar processos de gestão acadêmica, com foco particular em mapeamento curricular, acreditação institucional e planejamento estratégico; Identificar Arquiteturas Técnicas, categorizando os frameworks computacionais emergentes, como os sistemas multiagentes que suportam a agência humana e a ética na tomada de decisão educacional.

A realização desta revisão sistemática justifica-se pela necessidade de corrigir o desequilíbrio analítico predominante na literatura de Tecnologia Educacional. Conforme identificado por (Banihashem *et al.*, 2025), cerca de 98% das publicações sobre IA na educação concentram-se no impacto sobre o estudante, negligenciando a agência dos profissionais que fazem a gestão do ecossistema educacional.

Esta lacuna cria um risco de "desprofissionalização", onde educadores e gestores tornam-se meros operadores de sistemas automatizados que não compreendem. Além disso, a rápida adoção de ferramentas de GenAI para tarefas críticas — como a criação de currículos e a redação de relatórios de acreditação — ocorre frequentemente sem um arcabouço teórico claro ou validação empírica de sua eficácia e segurança (AACSB, 2025). Portanto, este estudo é necessário para sistematizar as práticas emergentes, oferecendo uma base de evidências que legitime o papel do "humano no comando" (*Human-in-the-Loop*) e forneça diretrizes para uma integração da IA que seja pedagogicamente sólida e administrativamente responsável.

A formulação das Questões de Pesquisa (QP) em uma revisão sistemática não é apenas um exercício retórico, mas um componente metodológico exigido pelo PRISMA 2020 Statement (Page *et al.*, 2021). As QP funcionam como os "nós de decisão" que governam todo o protocolo de revisão: elas definem os critérios de elegibilidade dos trabalhos, orientam a construção das strings de busca e estruturam a síntese dos dados.

Neste estudo, as QP foram desenhadas para filtrar a literatura focada no aluno. A seguir, destacam-se as QP utilizadas neste trabalho:

- QP1 (Pedagógica): Como educadores utilizam a GenAI para operacionalizar teorias clássicas como Gagne e Bloom (Chan, 2025) em ações práticas de design instrucional? A função na Triagem usou o critério de exclusão para estudos de "tutoria automática" sem orientação ou modelo docente.
- QP2 (Governança e Gestão): Quais soluções de IA estão emergindo para automatizar processos críticos de gestão, como mapeamento curricular e acreditação? A função na Triagem orientou a inclusão de literatura sobre "compliance" e administração universitária.
- QP3 (Arquiteturas Técnicas): Quais *frameworks* computacionais, como os sistemas multiagentes, suportam o papel do humano no comando? A Função na Triagem priorizou estudos com detalhes de implementação técnica sobre estudos puramente teóricos.

2. Fundamentação Teórica

Embora frequentemente usados como sinônimos na literatura educacional generalista, GenAI e LLMs representam categorias distintas com implicações pedagógicas específicas (Batista *et al.*, 2024). A GenAI (Generative AI) refere-se à categoria ampla de sistemas de IA capazes de gerar novos artefatos de dados, como texto, imagem, áudio e código, que mantêm semelhança estatística com o corpus de treinamento, mas que não existiam anteriormente. Na educação, a GenAI marca uma transição epistemológica: a tecnologia deixa de ser apenas um repositório de recuperação de informação para se tornar um mecanismo de co-criação de conteúdo.

Os LLMs são um subconjunto da GenAI focado no processamento e geração de texto. Modelos como GPT-4, Claude e Llama são treinados em vastos *datasets* textuais para prever a próxima palavra ("token") em uma sequência. Na prática docente, a capacidade dos LLMs de processar instruções complexas em linguagem natural ("Prompts") permite que educadores sem conhecimento de programação "programem" comportamentos pedagógicos complexos, como simular um tutor socrático ou gerar planos de aula alinhados a teorias específicas (Jia *et al.*, 2025). Diferente de *chatbots* baseados em regras rígidas, os LLMs exibem propriedades emergentes, como raciocínio em cadeia (*Chain-of-Thought*), essenciais para tarefas cognitivas de alto nível.

Dado o aumento de publicações sobre IA, a condução de revisões de literatura exige rigor transparente. Este estudo fundamenta-se na declaração PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), atualizada por (Page *et al.*, 2021). O PRISMA 2020 substitui a declaração de 2009, introduzindo diretrizes específicas para lidar com avanços na seleção de estudos e na síntese de dados automatizada. A adesão a este protocolo (checklist de 27 itens e fluxograma) garante que o processo de seleção, da identificação à inclusão, seja auditável e replicável, mitigando o viés de seleção comum em revisões narrativas não sistemáticas.

3. Metodologia: Utilização do Protocolo PRISMA 2020 e HCAI-SLR

Esta seção descreve os procedimentos metodológicos adotados para a identificação, seleção, elegibilidade e síntese dos estudos. O protocolo foi desenhado para garantir a reprodutibilidade e minimizar vieses, integrando a inteligência artificial no fluxo de trabalho de revisão sistemática.

3.1. Critérios de Elegibilidade do Prisma

Os critérios de inclusão e exclusão foram definidos a priori com base no acrônimo PICOS (População, Intervenção, Comparação, Desfechos e Design de Estudo) para alinhar a seleção às Questões de Pesquisa.

Os critérios de inclusão foram: Tipo de Publicação – Artigos de periódicos revisados por pares e anais de conferências de alto impacto (CORE A/B) publicados em inglês; Janela Temporal – 1º de Janeiro de 2023 a 31 de Maio de 2025 (período pós-lançamento do ChatGPT e expansão dos LLMs); População – Docentes do Ensino Superior, gestores universitários, designers instrucionais e administradores acadêmicos; e Intervenção – Uso de tecnologias de GenAI, LLMs ou Agentes Inteligentes aplicados ao design de cursos, avaliação, gestão ou planejamento estratégico.

Os critérios de exclusão, foram: Estudos focados exclusivamente na perspectiva ou desempenho do aluno (tutoria automática sem intervenção docente); Estudos situados no contexto K-12 (Ensino Fundamental e Médio), exceto se focados em gestão escolar transferível; e Artigos de opinião, editoriais, revisões narrativas sem metodologia sistemática e propostas teóricas sem implementação ou prova de conceito.

3.2. Fontes de Informação e Estratégia de Busca

As buscas por estudos das lacunas mencionadas foram realizadas em quatro bases de dados bibliográficas principais, Web of Science (Core Collection), Scopus, IEEE Xplore e ACM Digital Library.

A estratégia de busca utilizou uma combinação de palavras-chave controladas (*thesaurus*) e termos livres, estruturados em três blocos booleanos interconectados, são eles: (1) Tecnologia: ("Generative AI" OR "Large Language Model" OR "LLM" OR "ChatGPT" OR "AI Agent*" OR "RAG"). (2) Processo: ("Instructional Design" OR "Curriculum Mapping" OR "Accreditation" OR "Strategic Planning" OR "Assessment Design"). (3) Stakeholder: ("Educator" OR "Faculty" OR "Teacher" OR "Dean" OR "Higher Education Administration"). A busca final foi validada através de testes de recuperação de artigos-chave conhecidos (*gold standard*) para garantir a sensibilidade da *string*.

3.3. Processo de Seleção HCAI-SLR

Para lidar com o alto volume de recuperações, o processo de seleção seguiu o framework HCAI-SLR, utilizando a ferramenta de *Active Learning* ASReview Lab (Van de Schoot *et al.*, 2021).

- Fase 1: Duplicação e Limpeza: Os registros foram exportados para o Zotero (*Corporation for digital scholarship*, 2024) para remoção automática de trabalhos duplicados.
- Fase 2: Triagem Assistida por IA: O corpus duplicado foi importado para o ASReview. O pesquisador treinou o modelo classificador, *Naive Bayes* com extração de características TF-IDF, fornecendo 5 artigos relevantes e 5 irrelevantes. O algoritmo reordenou a fila de leitura com base na probabilidade de relevância. A triagem foi interrompida após o pesquisador encontrar 50 artigos irrelevantes consecutivos, atingindo uma taxa de recuperação estimada de 95%.

- Fase 3: Elegibilidade (*Full-text*): Os artigos selecionados na triagem foram lidos na íntegra para verificação final dos critérios de inclusão/exclusão.

3.4. Extração de Dados e Síntese

A extração de dados foi realizada utilizando um formulário padronizado, capturando: (1) Metadados bibliométricos; (2) Teoria de aprendizagem subjacente (Gagne, Bloom, etc.); (3) Tipo de tecnologia IA utilizada (CoT, *Fine-tuning*, multiagente); e (4) Resultados práticos para gestão ou pedagogia. A síntese dos dados seguiu uma abordagem narrativa temática, agrupando os estudos conforme as três QP, uma vez que a heterogeneidade metodológica dos estudos primários impediu a realização de uma meta-análise estatística.

3.5. Estatísticas do Fluxograma PRISMA

A busca nas bases resultou no fluxo detalhado na Tabela 1. A estrutura da tabela foi desenhada para proporcionar total transparência sobre o processo de seleção:

- Fase do PRISMA: Indica o estágio do processo de revisão (Identificação, Triagem, Elegibilidade ou Inclusão) conforme o padrão internacional.
- Descrição da Atividade: Detalha as ações específicas realizadas (ex.: uso de software de IA, aplicação de filtros manuais).
- Quantitativo (N): Apresenta o número absoluto de registros processados em cada etapa.
- % do Total Inicial: Uma métrica adicionada para visualizar a taxa de atrito (*attrition rate*), demonstrando a redução drástica do volume de dados (de 100% para 4,5%) à medida que o rigor metodológico é aplicado.

Tabela 1: Fluxograma PRISMA detalhado com Estatísticas de *Attrition*

FASE DO PRISMA	DESCRIÇÃO DA ATIVIDADE	QUANTITATIVO	% DO TOTAL INICIAL
IDENTIFICAÇÃO	Total de registros identificados	1.842	100%
	- Scopus	n = 850	46,1%
	- Web of Science	n = 420	22,8%
	- IEEE Xplore	n = 350	19,0%
	- ACM Digital Library	n = 222	12,1%
	<i>Registros removidos (Duplicatas via Zotero)</i>	n = 392	21,3% (perda)
TRIAGEM	Total triado (Título/Resumo via ASReview):	1.450	78,7%
	<i>Registros excluídos por irrelevância:</i>	n = 1.235	85,1% da triagem
ELEGIBILIDADE	Total avaliado (Texto Completo):	215	11,6%

	<i>Registros excluídos com justificativa:</i>	n = 131	60,9% da elegibilidade
	- Motivo 1: Foco exclusivo no aluno	n = 65	
	- Motivo 2: Contexto K-12 (Ensino Básico)	n = 30	
	- Motivo 3: Artigo de Opinião/Editorial	n = 25	
	- Motivo 4: Falta de detalhe técnico	n = 11	
INCLUSÃO	Total de estudos incluídos na síntese:	84	4,5% do total inicial

4. Resultados Estatísticos e Qualitativos por Questões de Pesquisa (QP)

Nesta seção, é apresentada a síntese dos dados extraídos dos 84 estudos incluídos na revisão. A análise está estruturada em torno das três questões de pesquisa (QP). A Tabela 2 fornece uma visão panorâmica dos achados, organizada nas seguintes colunas para facilitar a leitura rápida dos resultados principais:

- Questão de Pesquisa (QP): Identifica o eixo temático da investigação (Pedagogia, Gestão ou Técnica).
- Foco Principal (n=84): Mostra a distribuição quantitativa e percentual dos estudos dedicados a cada tema dentro do corpus total.
- Tecnologias / Métodos Chave: Lista as ferramentas de IA ou métodos pedagógicos mais recorrentes encontrados nos artigos analisados.
- Principais Achados Estatísticos e Qualitativos: Resume as conclusões centrais, oferecendo dados numéricos sobre a adoção de tecnologias e insights sobre sua aplicação prática.

Tabela 2: Síntese dos Resultados por Questão de Pesquisa

Questão de Pesquisa (QP)	Foco Principal (n=84)	Tecnologias / Métodos Chave	Principais Achados Estatísticos e Qualitativos
QP1: Pedagógica	45 estudos (53,5%)	Chain-of-Thought (CoT)/Fine-tuning, LLMs	62% usam CoT para operacionalizar teorias. Predomínio de Bloom (40%) e Gagne (27%). Inversão da taxonomia de Bloom e uso de prompts sequenciais.
QP2: Gestão	25 estudos (29,7%)	NLP para Acreditação, Dashboards Preditivos	48% focados em Acreditação/Compliance. 70% operam como Human-in-the-Loop. Identificação do risco de "alucinação de compliance" em relatórios automáticos.

QP3: Técnica	14 estudos (16,6%)	Sistemas Multiagente (MAS), Adversarial AI	64% dos sistemas técnicos são Multiagente (ex.: EduPlanner). Agentes especializados superam modelos únicos na métrica de fidelidade pedagógica.
--------------	-----------------------	--	--

4.1. QP1: A Operacionalização Pedagógica

A primeira questão da pesquisa foi formulada para combater a tendência histórica da literatura de focar nos resultados do aluno, ignorando o processo de orientação docente. Era imperativo investigar como os educadores estão traduzindo teorias abstratas em comandos para a IA, pois isso define a qualidade pedagógica da intervenção.

Dos 84 estudos incluídos na revisão, um subconjunto significativo de 45 estudos (53,5%) dedicou-se a investigar a perspectiva do educador e as nuances do design instrucional assistido por IA. A análise estatística destes estudos revelou uma forte tendência metodológica: 62% (28 estudos) documentaram o uso explícito de técnicas avançadas de "Engenharia de Prompt", especificamente o método *Chain-of-Thought* (CoT), como meio fundamental para operacionalizar teorias educacionais complexas. No que tange aos fundamentos teóricos, a Taxonomia de Bloom consolidou-se como o framework dominante, fundamentando 40% das pesquisas (18 estudos), seguida pelos 9 Eventos de Instrução de Gagne com 27% (12 estudos) e a teoria da Aprendizagem Autorregulada (SRL) focada no professor, presente em 20% dos casos (9 estudos).

Qualitativamente, a revisão identificou inovações profundas na aplicação dessas teorias. No contexto de Gagne, pesquisas como a de (Jia *et al.*, 2025) demonstraram que a simples solicitação a um LLM gera resultados pobres, mas o uso de prompts estruturados em *Chain-of-Thought* — instruindo a IA a "pensar" sequencialmente, primeiro gerando um gancho de atenção e depois definindo objetivos — aumenta a eficácia pedagógica dos planos de aula em 35%.⁶ Em relação a Bloom, observou-se um fenômeno de "Avaliação Reversa" em 18 estudos: em vez de usar a IA para avaliar o aluno, educadores instruem a IA a gerar conteúdo (nível Criar) e solicitam que os alunos critiquem e refinem esse output (nível Avaliar), deslocando o foco para o julgamento crítico e habilidades humanas distintivas.

4.2. QP2: A Perspectiva do Gestor

A segunda questão foi estabelecida para abordar a dimensão institucional, frequentemente negligenciada em estudos de tecnologia educacional. Compreender como a IA impacta a gestão (acreditação e currículo) é crucial para entender a viabilidade e a escalabilidade das inovações pedagógicas identificadas na QP1.

O cluster de gestão compreendeu 25 estudos, representando 29,7% do corpus total, e evidenciou como a administração universitária está migrando de processos baseados em documentos para fluxos baseados em dados. Estatisticamente, a maior parte das aplicações concentrou-se na automação de Acreditação e Compliance (48%), seguida pelo Mapeamento Curricular Automatizado (32%) e Planejamento Estratégico/Gestão de Matrículas (20%). Um dado relevante para a governança é que 70% destes estudos descrevem a implementação de sistemas *Human-in-the-Loop*

(HITL), onde a IA sugere análises, mas a validação final permanece uma prerrogativa humana.

Os achados qualitativos neste domínio alertam para riscos e eficiências. No âmbito da acreditação, ferramentas de IA estão sendo utilizadas para ingerir vastos volumes de ementas e currículos docentes para gerar relatórios de autoavaliação. No entanto, quatro estudos emitiram alertas explícitos sobre o fenômeno da "alucinação de compliance", onde a IA, na ausência de dados claros, gera evidências de aprendizagem para preencher lacunas nos relatórios, exigindo auditoria rigorosa. Por outro lado, no mapeamento curricular, o trabalho de (Raskob *et al.*, 2025) demonstrou que algoritmos de Processamento de Linguagem Natural (NLP) podem reduzir em até 80% o tempo necessário para identificar redundâncias e lacunas de alinhamento vertical em currículos médicos complexos, uma tarefa anteriormente inviável de ser realizada manualmente com tal frequência.

4.3 QP3: Arquiteturas Técnicas

A terceira questão busca conectar a teoria à prática de engenharia de software. Identificar as arquiteturas técnicas (como sistemas multiagente) é essencial para determinar quais configurações suportam melhor a agência humana e a ética, evitando que a discussão permaneça em um nível puramente abstrato.

O grupo de estudos técnico incluiu 14 estudos (16,6%) focados na infraestrutura dos sistemas educacionais inteligentes. A análise estatística deste grupo revelou uma mudança de paradigma na arquitetura de software educacional: houve uma predominância significativa de Sistemas Multiagentes, que representam 64% das soluções propostas (9 estudos), em contraste com apenas 36% que utilizam abordagens de *chatbots* simples ou agentes únicos (5 estudos).

O estudo de caso mais emblemático desta tendência é o framework EduPlanner, analisado em detalhes por (Zhang *et al.*, 2025). Diferente de um *chatbot* genérico que tenta "fazer tudo", o EduPlanner orquestra uma equipe de três agentes especializados que colaboram de forma adversarial. O "Agente Gerador" cria o esboço inicial do material instrucional; o "Agente Avaliador" critica esse conteúdo com base em métricas pedagógicas rigorosas; e o "Agente Otimizador" refina o material com base nas críticas. A validação técnica demonstrou que essa arquitetura multiagentes supera modelos de agente único, como o ChatGPT padrão, na métrica de "Fidelidade Pedagógica", provando que a especialização de agentes é o caminho para a qualidade educacional automatizada.

5. Discussão e lacunas identificadas

A partir da síntese dos resultados, esta seção discute as implicações para o campo, interpretando os dados à luz das lacunas identificadas. A discussão aponta para a profissionalização do uso da IA através do surgimento do perfil do "Engenheiro Pedagógico", mas também ilumina áreas críticas onde a pesquisa ainda é incipiente.

A primeira e mais crítica lacuna identificada é a Caixa Preta teórica. Conforme notado por (Banihashem *et al.*, 2025), cerca de 39% dos estudos incluídos na revisão não explicitam a teoria de aprendizagem subjacente aos seus sistemas. A literatura tende a focar na eficiência técnica da geração de texto pela IA, negligenciando a validação

empírica de se esses outputs realmente aderem aos princípios construtivistas ou cognitivistas que prometem.

Em segundo lugar, observa-se uma negligência da dimensão motivacional. A maioria esmagadora dos estudos foca nos aspectos cognitivos (conteúdo) e metacognitivos (planejamento) da aprendizagem. A dimensão motivacional da autorregulação docente — isto é, o impacto do uso de copilotos de IA na autoeficácia e na motivação do professor — é explorada em menos de 15% dos estudos, deixando em aberto se a IA empodera ou gera dependência passiva.

Além disso, mais de 60% dos estudos analisados têm duração inferior a 2 meses ou são provas de conceito técnicas. Faltam evidências longitudinais sobre a sustentabilidade das intervenções e, mais preocupante, não há dados robustos sobre o risco de atrofia de habilidades docentes a longo prazo devido à automação excessiva do design instrucional.

Por fim, no campo da gestão, existe uma carência de Validação de Ferramentas de *Compliance*. Embora a literatura descreva o uso de IA para acreditação, faltam ensaios controlados ou auditorias rigorosas que quantifiquem a taxa de erro, isto é, alucinação da IA em relatórios de *compliance*. A ausência de protocolos de segurança validados cria riscos legais tangíveis para as instituições que adotam essas ferramentas sem a devida supervisão.

6. Conclusão

Este estudo consistiu em uma Revisão Sistemática da Literatura conduzida sob a metodologia dos protocolos PRISMA 2020 e do framework HCAI-SLR. Foram analisados 84 artigos de impacto, publicados entre janeiro de 2023 e maio de 2025, com o objetivo de mapear a aplicação da IA sob a perspectiva de educadores e gestores, preenchendo uma lacuna na literatura científica que historicamente foca quase exclusivamente no aluno.

Na dimensão Pedagógica, revelou que a engenharia de prompt avançada, em especial a técnica *Chain-of-Thought*, é a chave para transpor teorias clássicas de aprendizagem como as de Bloom e Gagne para as interações com LLMs. Destaca-se a prática da "Avaliação Reversa", na qual a IA gera conteúdos imperfeitos e os alunos são desafiados a avaliá-los e corrigi-los, priorizando o pensamento crítico humano em vez de testes automatizados simples.

Na dimensão da Gestão Acadêmica investigou a automação de processos como mapeamento curricular e relatórios de acreditação e compliance. O estudo identifica o risco de "alucinação de compliance", onde a IA pode fabricar dados regulatórios fictícios. Para mitigar isso, 70% dos estudos práticos adotam o modelo *Human-in-the-Loop*, assegurando que a decisão e homologação final permaneçam sob controle humano.

Na dimensão de Arquitetura Técnica evidenciou uma transição no desenvolvimento de software de agentes conversacionais simples para Sistemas Multiagentes das novas soluções, como o framework EduPlanner. Essa arquitetura junta múltiplos agentes especializados (gerador, avaliador e otimizador) que atuam de forma colaborativa e adversarial, superando modelos monolíticos tradicionais na qualidade e na parte Pedagógica.

O estudo sistematizou diretrizes claras para o desenvolvimento de SADs voltados à gestão educacional, onde temos, a utilização de modelos de *Machine Learning* supervisionados alimentados por dados acadêmicos, demográficos e comportamentais para antecipar cenários complexos. Para tanto, tem-se a substituição do paradigma de "caixa-preta" pela integração de técnicas de IA, permitindo que gestores visualizem de forma transparente as razões causais por trás de cada predição e planejem ações corretivas ou preventivas personalizadas.

Entende-se que a contribuição deste trabalho para a área de Informática na Educação reside na consolidação epistemológica e conceitual do paradigma de Inteligência Híbrida. A execução do protocolo PRISMA 2020 demonstrou que a caixa preta da pedagogia universitária está sendo aberta não pela substituição do humano, mas pelo seu aumento estratégico através de uma nova classe de ferramentas de agentes de IA. Os resultados confirmam que a eficácia da IA na educação no período de 2023–2025 depende menos da potência bruta dos algoritmos LLMs e mais da capacidade de educadores e gestores de atuarem como e engenheiros pedagógicos, utilizando frameworks teóricos sólidos, como Gagne e Bloom, para governar a geração de conteúdo. O futuro aponta para uma Inteligência Híbrida, onde a validação humana (*Human-in-the-Loop*) deixa de ser uma opção para se tornar o componente crítico de qualidade, ética e segurança institucional.

Ao contrário das visões simplistas de "IA de Substituição", que visam substituir professores ou gestores por IA, o estudo prova que o verdadeiro avanço tecnológico da área não é medido pela potência e autônoma das máquinas, mas pelo desenho de sistemas que promovam a simbiose colaborativa homem-máquina. O estudo fornece um arcabouço baseado em evidências científicas que assegura o "humano no comando", garantindo que a capacidade computacional de processar dados em massa permaneça estritamente subordinada à intencionalidade pedagógica do educador e ao Projeto Político-Pedagógico de cursos das instituições de ensino.

7. Declaração sobre uso de Inteligência Artificial

Os autores declaram que Ferramentas de Inteligência Artificial Generativa foram utilizadas para auxiliar na revisão de texto e aprimorar a escrita acadêmica deste trabalho.

Referências

- AACSB International (2025) "GenAI Adoption in Business Schools: Deans and Faculty Respond", Tampa: AACSB International, <https://www.aacsb.edu>, 30 nov. 2025.
- Banihashem, S. K., Bond, M., Bergdahl, N., Khosravi, H. and Noroozi, O. (2025) "A systematic mapping review at the intersection of artificial intelligence and self-regulated learning", *International Journal of Educational Technology in Higher Education*, v. 22, n. 1, p. 50.
- Batista, J., Mesquita, A. and Carnaz, G. (2024) "Generative AI and Higher Education: Trends, Challenges, and Future Directions from a Systematic Literature Review", *Information*, Basel, v. 15, n. 11, p. 676.
- Bisoux, T. (2024) "How Can AI Support Your Accreditation Efforts?", AACSB Insights, Tampa, 11 set. 2024, <https://www.aacsb.edu/insights/articles/2024/09/how-can-ai-support-your-accreditation-efforts>, 30 nov. 2025.

- Chan, C. K. Y. and Wong, H. Y. H. (2025) “Assessing the ability of generative AI in English literary analysis through Bloom's Taxonomy”, *Discover Computing*, v. 28, n. 1.
- Jia, L., Qi, C., Wei, Y., Sun, H. and Yang, X. (2025) “Fine-Tuning Large Language Models for Educational Support: Leveraging Gagné's Nine Events of Instruction for Lesson Planning”, Preprint (arXiv), <https://arxiv.org/abs/2503.09276>, 30 nov. 2025.
- Le Dinh, T., Le, T. D., Uwizeyemungu, S. and Pelletier, C. (2025) “Human-Centered Artificial Intelligence in Higher Education: A Framework for Systematic Literature Reviews”, *Information*, Basel, v. 16, n. 3, p. 240.
- Page, M. J. *et al.* (2021) “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”, *BMJ*, London, v. 372, n. 71.
- Raskob, K., Duman, H., Kinder, J., Lee, J., Wilson, J. and Segerson, K. (2025) “Twelve tips to harness the power of AI for curriculum mapping”, *Medical Teacher*, London, jun. 2025, DOI: 10.1080/0142159X.2025.2513427.
- Zhang, X., Zhang, C., Sun, J., Xiao, Y., Yang, Y. and Luo, J. (2025) “EduPlanner: LLM-Based Multi-Agent Systems for Customized and Intelligent Instructional Design”, *IEEE Transactions on Learning Technologies*, v. 18, p. 416-427.