

INTELIGÊNCIA ARTIFICIAL GENERATIVA NA EDUCAÇÃO MÉDICA: POTENCIALIDADES, LIMITAÇÕES E DESAFIOS PARA A FORMAÇÃO DE FUTUROS MÉDICOS

GENERATIVE ARTIFICIAL INTELLIGENCE IN MEDICAL EDUCATION: POTENTIALS, LIMITATIONS AND CHALLENGES FOR THE TRAINING OF FUTURE PHYSICIANS

Júlia Gabriele de Oliveira Araújo¹, Isabelle Bezerra Alves¹, Letícia Oliveira Cavalcante de Carvalho¹, Maria Clara Pereira Macário de Lima¹, Maria Isabel Jessica da Silva Dantas²,
Heloise de Matos Cavalcanti¹

¹ Acadêmicas do curso de Medicina da Faculdade de Ciências Médicas Afya Garanhuns.

² Professora do curso de Medicina da Faculdade de Ciências Médicas Afya Garanhuns.

E-mail do autor correspondente: heloise.cavalcanti@alunos.afya.com.br

Eixo Temático: Tecnologia e Inovação

RESUMO

Introdução: A inteligência artificial generativa (IAG), especialmente os modelos de linguagem de grande porte, vem sendo incorporada à educação médica para apoio ao estudo, geração de materiais, avaliação e devolutiva individualizada. **Objetivo:** Analisar as potencialidades, limitações e desafios da IAG na formação médica, considerando seus efeitos sobre aprendizagem, avaliação e desenvolvimento do pensamento crítico. **Método:** Revisão integrativa da literatura, realizada em agosto de 2026 na PubMed/MEDLINE, com descritores e termos relacionados à inteligência artificial generativa, modelos de linguagem de grande porte, educação médica, estudantes de Medicina, avaliação e aprendizagem. Foram incluídas publicações de 2023 a 2026, em inglês ou português, com foco em aplicações educacionais, resultados de aprendizagem ou riscos do uso dessas tecnologias. **Resultados:** As evidências apontam potencial para personalização do ensino, geração de questões, apoio tutorial e devolutivas rápidas. Foi identificada melhora de curto prazo em conhecimento e habilidades clínicas, porém com heterogeneidade e certeza de evidência baixa a moderada. Persistem riscos de respostas incorretas, dependência excessiva, viés, plágio, comprometimento do pensamento crítico, privacidade e desigualdade de acesso. **Conclusão:** A IAG pode complementar a educação médica, desde que utilizada com supervisão docente, verificação das informações e estratégias que preservem autonomia, integridade acadêmica e raciocínio clínico.

Palavras-chave: Inteligência Artificial Generativa; Educação Médica; Estudantes de Medicina; Aprendizagem; Avaliação Educacional.

ABSTRACT

Introduction: Generative artificial intelligence (GenAI), particularly large language models, has been incorporated into medical education to support studying, content generation, assessment, and individualized feedback. **Objective:** To analyze the potentials, limitations, and challenges of GenAI in medical training,

considering its effects on learning, assessment, and critical thinking development. **Method:** Integrative literature review conducted in August 2026 in PubMed/MEDLINE using descriptors and terms related to generative artificial intelligence, large language models, medical education, medical students, assessment, and learning. Publications from 2023 to 2026 in English or Portuguese focusing on educational applications, learning outcomes, or risks were included. **Results:** Evidence indicates potential for personalized learning, question generation, tutoring support, and rapid feedback. A recent meta-analysis found short-term improvements in knowledge and clinical skills, although heterogeneity was high and certainty of evidence ranged from low to moderate. Risks remain regarding inaccurate outputs, overreliance, bias, plagiarism, impaired critical thinking, privacy, and unequal access. **Conclusion:** GenAI may complement medical education when used with faculty supervision, information verification, and strategies that preserve autonomy, academic integrity, and clinical reasoning.

Keywords: Generative Artificial Intelligence; Education, Medical; Students, Medical; Learning; Educational Measurement.

INTRODUÇÃO

A inteligência artificial generativa (IAG) corresponde a sistemas capazes de produzir novos conteúdos a partir de padrões aprendidos em grandes conjuntos de dados. Na educação médica, sua expansão ocorreu principalmente com os modelos de linguagem de grande porte (*large language models* - LLMs), que podem responder perguntas, resumir conteúdos, simular diálogos, elaborar materiais e apoiar atividades de avaliação. Abd-Alrazaq *et al.*¹ destacam que essas ferramentas apresentam potencial para personalizar experiências de aprendizagem, apoiar estudantes e docentes e ampliar a produção de recursos educacionais.

Além da disponibilidade imediata de informações, os LLMs podem atuar como tutores virtuais, auxiliar na elaboração de planos de estudo, gerar cenários clínicos e favorecer aprendizagem interativa. Benítez *et al.*² ressaltam que essas aplicações podem contribuir para o desenvolvimento de habilidades e para o ensino de conceitos complexos, mas alertam para riscos de dependência excessiva, imprecisão, desinformação, plágio, viés algorítmico e redução do pensamento crítico.

Na avaliação educacional, Artsi *et al.*³ analisaram oito estudos sobre geração de questões de múltipla escolha por LLMs. Cinco estudos mostraram que os modelos eram capazes de produzir questões consideradas válidas para exames médicos, porém todos relataram itens inadequados ou que necessitavam de revisão. O achado reforça que a geração automática pode reduzir a carga de trabalho, mas não elimina a necessidade de supervisão especializada.

A evidência sobre impacto direto na aprendizagem ainda está em consolidação. Em ensaio cruzado randomizado, Nissen *et al.*⁴ avaliaram devolutiva individualizada produzida pelo *ChatGPT* em atividades de estudantes de Medicina. O recurso mostrou viabilidade, porém não apresentou superioridade significativa sobre o comentário de especialistas no desempenho do exame final, indicando que o benefício depende da qualidade da implementação e do desenho pedagógico.

Diante da rápida incorporação dessas ferramentas e da necessidade de formar médicos capazes de utilizá-las criticamente, torna-se relevante sintetizar as evidências disponíveis. Assim, o objetivo deste estudo foi analisar as potencialidades, limitações e desafios da inteligência artificial generativa na educação médica, com ênfase em aprendizagem, avaliação, pensamento crítico, integridade acadêmica e papel da supervisão docente.

MÉTODO

Trata-se de uma revisão integrativa da literatura, de caráter descritivo, orientada pela pergunta norteadora: “Quais são as potencialidades, limitações e desafios da inteligência artificial generativa na educação médica e quais implicações seu uso apresenta para a formação de futuros médicos?”

A busca foi realizada em agosto de 2026 na base PubMed/MEDLINE. Utilizaram-se os termos “*Generative Artificial Intelligence*”, “*Large Language Models*”, “*ChatGPT*”, “*Medical Education*”, “*Medical Students*”, “*Learning*”, “*Educational Assessment*” e “*Feedback*”, combinados pelos operadores booleanos AND e OR.

Foram incluídas publicações de 2023 a 2026, em inglês ou português, que abordassem aplicações de IAG ou LLMs na educação médica e apresentassem resultados sobre aprendizagem, avaliação, devolutiva, desempenho, pensamento crítico, integridade acadêmica ou limitações. Foram excluídos trabalhos sem relação direta com formação médica, estudos centrados exclusivamente no uso clínico de IA e publicações duplicadas. Após leitura dos estudos elegíveis, os achados foram organizados em quatro eixos: personalização e apoio ao ensino; avaliação e devolutiva; resultados de aprendizagem; e riscos éticos e pedagógicos. Por utilizar apenas literatura publicada, esta revisão não demandou apreciação por Comitê de Ética em Pesquisa.

RESULTADOS

As publicações mostram que a IAG pode atuar em diferentes etapas da formação médica. Abd-Alrazaq *et al.*¹ descrevem aplicações em desenvolvimento curricular, materiais, planos personalizados de estudo e avaliação. Vrdoljak *et al.*⁵ também apontam potencial como tutores virtuais, pacientes simulados e apoio à produção de conteúdos, desde que haja controle de qualidade e integração ao currículo.

Na elaboração de avaliações, Artsi *et al.*³ observaram que LLMs podem produzir questões de múltipla escolha potencialmente utilizáveis em exames médicos, mas todos os estudos incluídos identificaram questões defeituosas ou que exigiam modificações. Isso indica que a automação pode auxiliar docentes e discentes, porém a revisão humana continua indispensável para assegurar validade, clareza e alinhamento aos objetivos educacionais.

Quanto à devolutiva individualizada, Nissen *et al.*⁴ compararam comentários gerados pelo *ChatGPT* com comentários gerais de especialistas em tarefas escritas. O desempenho final foi semelhante entre as modalidades, sem superioridade estatisticamente significativa da devolutiva por IA. Revisão de escopo publicada por Kıyak *et al.*⁶ identificou 42 estudos sobre devolutivas geradas por LLMs na graduação médica; os resultados sugeriram viabilidade e utilidade percebida, mas a maioria dos trabalhos avaliou apenas reações ou ganhos imediatos, sem demonstrar mudanças comportamentais ou efeitos sobre pacientes.

Uma meta-análise de 20 ensaios randomizados, envolvendo 1.413 participantes, encontrou melhora de curto prazo no conhecimento e nas habilidades clínicas quando a IAG foi integrada ao ensino. Entretanto, a heterogeneidade foi elevada, a certeza da evidência variou de baixa a moderada e os autores ressaltaram que os resultados não justificam substituir métodos convencionais de ensino.⁷ Esses dados apoiam o uso da tecnologia como complemento e não como estratégia isolada.

As limitações pedagógicas e éticas aparecem de forma consistente. Benítez *et al.*² destacam dependência excessiva, comprometimento do pensamento crítico, respostas incorretas, plágio, privacidade e desigualdade; Abd-Alrazaq *et al.*¹ acrescentam viés

algorítmico e direitos autorais. Por isso, os estudantes precisam ser treinados para verificar fontes e reconhecer limites da ferramenta.

Em conjunto, as evidências indicam que a principal perspectiva é o uso supervisionado e pedagogicamente planejado da IAG. Seu valor depende de mediação docente, políticas de integridade acadêmica e manutenção de atividades que exijam raciocínio clínico independente.

Os principais achados dos estudos selecionados estão sintetizados na Tabela 1.

Tabela 1 - Síntese dos principais achados sobre inteligência artificial generativa na educação médica

Estudo	Aplicação	Principal achado
Artsi <i>et al.</i> ³	Geração de questões	Pode auxiliar na criação de avaliações, mas requer revisão por especialistas.
Nissen <i>et al.</i> ⁴	Devolutiva individualizada	Viável, porém sem superioridade significativa sobre comentário de especialistas.
Wang <i>et al.</i> ⁷	Aprendizagem	Meta-análise encontrou melhora de curto prazo em conhecimento e habilidades clínicas, com evidência ainda heterogênea.

Fonte: (Autoria própria 2026).

CONCLUSÃO

A inteligência artificial generativa apresenta potencial para complementar a educação médica por meio de personalização do estudo, geração de materiais e avaliações, simulações e devolutivas rápidas. Evidências recentes sugerem benefícios de curto prazo para conhecimento e habilidades clínicas, mas os resultados ainda são heterogêneos e não sustentam a substituição do ensino convencional ou da supervisão docente. A incorporação responsável dessas tecnologias exige letramento em IA, verificação sistemática das informações, políticas de integridade acadêmica e preservação do pensamento crítico e do raciocínio clínico. Estudos futuros devem avaliar resultados de longo prazo, impacto sobre comportamento profissional e segurança do paciente, além de comparar estratégias pedagógicas em diferentes contextos e populações.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, *et al.* Large language models in medical education: opportunities, challenges, and future directions. *JMIR Med Educ.* 2023;9:e48291. doi:10.2196/48291.
2. Benítez TM, Xu Y, Boudreau JD, Kow AWC, Bello F, Phuoc LV, *et al.* Harnessing the potential of large language models in medical education: promise and pitfalls. *J Am Med Inform Assoc.* 2024;31(3):776-783. doi:10.1093/jamia/ocad252.
3. Artsi Y, Sorin V, Konen E, Glicksberg BS, Nadkarni G, Klang E. Large language models for generating medical examinations: systematic review. *BMC Med Educ.* 2024;24:354. doi:10.1186/s12909-024-05239-y.
4. Nissen L, Rother JF, Heinemann M, Reimer LM, Jonas S, Raupach T. A randomised cross-over trial assessing the impact of AI-generated individual feedback on written online assignments for medical students. *Med Teach.* 2025;47(9):1544-1550. doi:10.1080/0142159X.2025.2451870.
5. Vrdoljak J, Boban Z, Vilović M, Kumrić M, Božić J. A review of large language models in medical education, clinical decision support, and healthcare administration. *Healthcare (Basel).* 2025;13(6):603. doi:10.3390/healthcare13060603.
6. Kıyak YS, İş-Kara T, Emekli E. Applications and outcomes of large-language-model-generated feedback in undergraduate medical education: a scoping review. *Med Sci Educ.* 2026;36(1):81-99. doi:10.1007/s40670-025-02621-3.
7. Wang C, Sun N, Pan X, Liu P, Ma Z, Wang W, *et al.* The impact of integrating generative artificial intelligence into medical education on short-term learning outcomes: a systematic review and meta-analysis of randomized controlled trials. *BMC Med Educ.* 2026;26:983. doi:10.1186/s12909-026-09320-6.