

INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO MÉDICA E NO DESENVOLVIMENTO DO RACIOCÍNIO CLÍNICO: UMA REVISÃO DE LITERATURA

Generative artificial in medical education and clinical reasoning development: a literature review

Hemilly Dantas de Sousa¹; José Adelson Alves do Nascimento Júnior²

1. Acadêmico do curso de Medicina da Faculdade de Ciências Médicas de Garanhuns; 2. Professor do curso de Medicina da Faculdade de Ciências Médicas de Garanhuns.

E-mail: hemillymed38@gmail.com

Eixo Temático: 03

RESUMO

Introdução: Desde o lançamento do ChatGPT em 2022, a IA generativa tem sido cada vez mais integrada à educação médica e à prática clínica, auxiliando no ensino e no raciocínio diagnóstico. **Objetivo:** Analisar o uso da IA generativa na educação médica e no raciocínio clínico, destacando as principais plataformas, limites e possíveis malefícios. **Método:** Revisão de literatura nas bases PubMed, Web of Science, Scopus e Google Scholar, utilizando descritores como IA generativa, ChatGPT e educação médica, priorizando publicações recentes. **Resultados:** A tecnologia mostrou-se útil como assistente de aprendizagem, simulação de pacientes e suporte diagnóstico, exemplificada pelo sistema AMIE e por redes brasileiras de eletrocardiografia na triagem da doença de Chagas. Contudo, destacam-se riscos como o viés de automação e a aceitação acrítica de erros. **Conclusão:** A IA generativa possui grande potencial como ferramenta complementar na formação e prática médica, exigindo supervisão crítica constante para evitar dependência excessiva e falhas diagnósticas.

Palavras-chave: Inteligência Artificial. Educação Médica. Raciocínio Clínico. Estudantes de Medicina. Tecnologia Educacional.

ABSTRACT

Introduction: Since the launch of ChatGPT in 2022, generative AI has been increasingly integrated into medical education and clinical practice, assisting in teaching and diagnostic reasoning. **Objective:** To analyze the use of generative AI in medical education and clinical reasoning, highlighting the main platforms, limitations, and potential harms. **Method:** Literature review in the PubMed, Web of Science, Scopus, and Google Scholar databases, using descriptors such as generative AI, ChatGPT, and medical education, prioritizing recent publications. **Results:** The technology proved useful as a learning assistant, patient simulation tool, and diagnostic support, exemplified by the AMIE system and by Brazilian electrocardiography networks in the screening of Chagas disease. However, risks such as automation bias and the uncritical acceptance of errors stand out. **Conclusion:** Generative AI has great potential as a complementary tool in medical training and practice, requiring constant critical supervision to avoid excessive dependence and diagnostic failures.

Keywords: Artificial Intelligence. Education, Medical. Clinical Reasoning. Students, Medical. Educational Technology.

INTRODUÇÃO

A inteligência artificial (IA) generativa e os modelos de linguagem de grande porte passaram a integrar o cotidiano da medicina de forma acelerada desde o lançamento público do ChatGPT, em novembro de 2022, capaz de gerar textos coerentes e de responder a perguntas complexas em linguagem natural, com potencial de aplicação na pesquisa, na prática clínica,

na interação com pacientes e, na educação médica¹. Uma nova categoria de ferramentas de IA médica surgiu nesse contexto: modelos como o GPT-4 passaram a ser discutidos quanto aos seus benefícios, limites e riscos específicos para a prática clínica, incluindo tarefas que de análise de imagens e priorização de pacientes de alto risco². Essa transformação tem sido acompanhada por revisões que sistematizam as principais aplicações desses modelos na medicina, como suporte à decisão clínica, documentação médica e educação, ao mesmo tempo em que apontam desafios relacionados a vieses de algoritmo e à necessidade de validação clínica rigorosa antes de sua adoção em larga escala³.

No campo específico da formação médica, um número crescente de aplicações vem sendo mapeado, incluindo o apoio ao ensino teórico, a geração de casos clínicos simulados, a criação de questões de avaliação e o fornecimento de feedback automatizado aos estudantes⁴, além de potencial para a atualização de conhecimento em subespecialidades médicas, como demonstrado em análise sobre o uso do ChatGPT no ensino da cardiologia⁵. Uma revisão sistemática de maior escala, que analisou 166 estudos e selecionou 40 diretamente relacionados ao tema, identificou o aprendizado personalizado e o apoio à tomada de decisão clínica como os principais benefícios da IA generativa na formação médica, ao lado de desafios relacionados à precisão do conteúdo gerado e ao risco de uso acrítico por parte dos estudantes⁶.

Especificamente em relação ao raciocínio clínico, o uso de chatbots no ensino dessa competência já foi objeto de revisão de escopo dedicada, que identificou estratégias como autoexplicação, reflexão estruturada, prática com casos clínicos e fornecimento de feedback como as aplicações educacionais mais descritas na literatura⁷, achado esse que foi reforçado por revisão sobre ensino baseado em simulação, que constatou melhora nos desfechos de raciocínio clínico em quatro de seis estudos analisados, ainda que a maioria carecesse de referenciais teóricos explícitos⁸. Apesar do crescente número de publicações, outra revisão sistemática sobre modelos conversacionais de IA na área da saúde apontou que grande parte dos estudos ainda carece de desenhos metodológicos robustos e de avaliação padronizada de desfechos, o que limita conclusões definitivas sobre sua efetividade educacional real⁹.

Diante da rápida incorporação da IA generativa ao cotidiano da formação e da prática médica, e da ainda escassa consolidação de evidências sobre seu real impacto no desenvolvimento do raciocínio clínico, seus limites de aplicação e os potenciais malefícios de seu uso inadequado no atendimento, o presente estudo teve como objetivo analisar, por meio

de uma revisão da literatura, o uso da inteligência artificial generativa como ferramenta de apoio à educação médica e ao desenvolvimento do raciocínio clínico.

MÉTODO

Trata-se de uma revisão de literatura, com buscas realizadas nas bases PubMed/MEDLINE, Web of Science, Scopus e Google Scholar, utilizando os descritores "Inteligência Artificial Generativa" (Generative Artificial Intelligence), "ChatGPT", "Educação Médica" (Medical Education) e "Raciocínio Clínico" (Clinical Reasoning). Foram incluídos artigos originais, revisões sistemáticas e de escopo publicados nos últimos cinco anos (2021 a 2026), priorizando periódicos de bom fator de impacto, com texto disponível na íntegra nos idiomas português, inglês ou espanhol. Foram excluídos estudos duplicados entre as bases consultadas, cartas ao editor, editoriais e artigos de opinião sem embasamento empírico. Os artigos selecionados foram lidos na íntegra e organizados segundo o tipo de aplicação da IA generativa e os desfechos educacionais e clínicos relatados.

Por se tratar de uma revisão de literatura fundamentada exclusivamente em dados secundários de domínio público, sem coleta direta de dados junto a seres humanos, o presente estudo dispensa apreciação por Comitê de Ética em Pesquisa, em conformidade com a Resolução nº 466/2012 do Conselho Nacional de Saúde¹⁰.

RESULTADOS

A IA generativa mostra-se aplicável como assistente de aprendizagem e criação de conteúdo educacional: a geração de casos clínicos simulados, questões de avaliação e resumos de literatura estão entre as aplicações mais recorrentes descritas na literatura, ressalvada a dependência da qualidade dos comandos (prompts) fornecidos pelo usuário para a obtenção de respostas clinicamente relevantes^{1,4}. Ainda que esses modelos apresentem desempenho comparável ao humano em tarefas de conhecimento médico estruturado, seu uso continua a exigir supervisão, dado o risco de respostas plausíveis, porém factualmente incorretas, fenômeno descrito na literatura como "alucinação"³.

Uma segunda frente de aplicação relevante é a simulação de pacientes virtuais e a tutoria para tomada de decisão clínica, por meio de estratégias como autoexplicação, reflexão estruturada e prática com casos simulados⁷, associadas a melhora nos desfechos de raciocínio clínico dos estudantes em parte dos estudos analisados⁸. Nesse mesmo sentido, a revisão

publicada em Medical Education identificou o aprendizado personalizado como um dos benefícios mais consistentemente relatados na literatura, ao lado de preocupações quanto à supervisão docente necessária para o uso seguro dessas ferramentas⁶.

Para além dos chatbots de propósito geral, observa-se hoje o desenvolvimento de plataformas de IA médica especializadas, construídas especificamente para tarefas clínicas de alta complexidade. O sistema AMIE (Articulate Medical Intelligence Explorer), desenvolvido pela Google, exemplifica essa nova geração de ferramentas, o qual foi treinado especificamente para conduzir diálogos diagnósticos de múltiplos turnos, sendo comparado a médicos de atenção primária em consultas simuladas com pacientes-atores e apresentando desempenho equivalente ou superior em precisão diagnóstica, coleta de história clínica e elaboração de plano terapêutico, ainda que sob condições controladas e não em ambiente assistencial real¹¹.

No Brasil, iniciativa semelhante antecedeu a própria popularização dos modelos de linguagem conversacionais: pesquisadores brasileiros treinaram uma rede neural profunda capaz de diagnosticar automaticamente anormalidades em eletrocardiogramas de 12 derivações a partir de mais de dois milhões de exames coletados pela Rede de Telessaúde de Minas Gerais, alcançando desempenho comparável ao de cardiologistas especialistas¹². Essa mesma base tecnológica fundamentou, mais recentemente, um desafio científico internacional voltado à detecção da doença de Chagas por meio de inteligência artificial aplicada a eletrocardiogramas, iniciativa que reuniu mais de 600 participantes de 111 equipes ao redor do mundo e que tem potencial relevante para regiões endêmicas como o Brasil, onde o teste sorológico confirmatório é frequentemente subutilizado por limitações de acesso, permitindo priorizar pacientes para investigação diagnóstica a partir de um exame simples e amplamente disponível¹³.

Essas plataformas ilustram como a IA generativa pode auxiliar o raciocínio clínico em diferentes etapas do processo diagnóstico: na coleta estruturada de história clínica por meio de diálogo natural, na geração de hipóteses diagnósticas a partir de dados clínicos complexos, na priorização de exames complementares em contextos de recursos limitados e no apoio à tomada de decisão terapêutica baseada em diretrizes atualizadas^{11,13}. No atendimento médico real, contudo, o uso dessas ferramentas ainda se concentra majoritariamente em tarefas de apoio administrativo e de documentação clínica, com adoção mais cautelosa em decisões diagnósticas e terapêuticas diretas, dado o caráter ainda experimental dos sistemas e a ausência de validação em larga escala fora de ambientes controlados⁹.

Quanto aos limites de utilização, a literatura converge para o entendimento de que a IA generativa deve atuar como ferramenta de apoio, e não como substituta do julgamento clínico. Artigo publicado no JAMA descreve o fenômeno do viés de automação, no qual profissionais de saúde tendem a aceitar sugestões geradas por sistemas de apoio à decisão mesmo quando estas se mostram incorretas, o que reduz a vigilância clínica e amplia o risco de erros diagnósticos, sobretudo entre profissionais menos experientes ou em contextos de sobrecarga assistencial¹⁴. Esse risco soma-se ao fenômeno da alucinação discutido³, reforçando que a ausência de supervisão crítica constante constitui o principal malefício incorporação da IA generativa ao atendimento médico real, com consequências que vão do erro diagnóstico isolado à erosão progressiva da própria capacidade de raciocínio clínico autônomo do profissional que depende excessivamente da ferramenta.

Quanto à qualidade metodológica dos estudos incluídos, observa-se predominância de revisões sistemáticas e de escopo publicadas em periódicos de expressivo fator de impacto^{3,6,9,14}, ao lado de estudos experimentais controlados envolvendo comparação direta entre desempenho humano e de sistemas de IA^{11,12}; ainda assim, a heterogeneidade dos desenhos, das ferramentas avaliadas e das medidas de desfecho utilizadas entre os estudos limita comparações diretas e a realização de metanálises, sendo recomendável cautela na generalização dos achados para diferentes contextos de ensino e atendimento médico, sobretudo o brasileiro.

CONCLUSÃO

A revisão mostrou que a IA generativa vem sendo cada vez mais incorporada à educação e à prática médica, com aplicações que vão da criação de conteúdo educacional e simulação de pacientes a plataformas de apoio ao raciocínio diagnóstico, como o sistema AMIE e as redes neurais brasileiras de eletrocardiografia usadas na triagem da doença de Chagas^{6,7,8,11,12,13}. No entanto, essas ferramentas apresentam limites importantes, como o risco de respostas incorretas e o viés de automação, que podem comprometer a vigilância clínica quando utilizadas sem supervisão crítica constante^{3,14}. Conclui-se que a IA generativa deve ser vista como ferramenta complementar — e não substituta — do julgamento clínico e do processo de ensino-aprendizagem, sendo necessárias pesquisas futuras com desenhos mais robustos e maior representatividade de instituições brasileiras para consolidar evidências sobre seu real impacto na formação médica.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Fatima A, Shafique MA, Alam K, Fadlalla Ahmed TK, Mustafa MS. ChatGPT in medicine: a cross-disciplinary systematic review of ChatGPT's (artificial intelligence) role in research, clinical practice, education, and patient interaction. *Medicine (Baltimore)*. 2024;103(32):e39250.
2. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med*. 2023;388(13):1233-9.
3. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930-40.
4. Tozsın A, Ucmak H, Soyutürk S, Aydın A, Gozen AS, Fahim MA, et al. The role of artificial intelligence in medical education: a systematic review. *Surg Innov*. 2024;31(4):415-23.
5. Sharma A, Medapalli T, Alexandrou M, Brilakis E, Prasad A. Exploring the role of ChatGPT in cardiology: a systematic review of the current literature. *Cureus*. 2024.
6. Lucas HC, Upperman JS, Robinson JR. A systematic review of large language models and their implications in medical education. *Med Educ*. 2024;58(11):1276-85.
7. Freitas Bernardo Ferreira G, Moura AS, Ribeiro LMC, Turci MA, Mamede S. Como a inteligência artificial generativa pode facilitar o ensino do raciocínio clínico: uma revisão de escopo. *Rev Docência Ens Super*. 2026;15:1-27.
8. Liu YM, Chou CC, Jaing TH, Okoli CTC. Generative AI for clinical reasoning: a scoping review. *Teach Learn Nurs*. 2025.
9. Wang L, Wan Z, Ni C, Song Q, Li Y, Clayton EW, et al. Applications and concerns of ChatGPT and other conversational large language models in health care: systematic review. *J Med Internet Res*. 2024;26:e22769.
10. Brasil. Ministério da Saúde. Conselho Nacional de Saúde. Resolução nº 466, de 12 de dezembro de 2012. Dispõe sobre diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. *Diário Oficial da União*. Brasília, DF; 2013 jun 13.
11. Tu T, Schaekermann M, Palepu A, Saab K, Freyberg J, Tanno R, et al. Towards conversational diagnostic artificial intelligence. *Nature*. 2025;642(8067):442-50.
12. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*. 2020;11(1):1760.
13. Reyna MA, Koscova Z, Pavlus J, Weigle J, Saghafi S, Gomes P, et al. Detection of Chagas disease from the ECG: the George B. Moody PhysioNet Challenge 2025. *Comput Cardiol*. 2025;52:1-4.
14. Khera R, Simon MA, Ross JS. Automation bias and assistive AI: risk of harm from AI-driven clinical decision support. *JAMA*. 2023;330(23):2255-7.