

INTELIGÊNCIA ARTIFICIAL NO TREINAMENTO DE RACIOCÍNIO CLÍNICO EM ESTUDANTES DE MEDICINA: UMA REVISÃO SISTEMÁTICA

ARTIFICIAL INTELLIGENCE IN CLINICAL REASONING TRAINING FOR
MEDICAL STUDENTS: A SYSTEMATIC REVIEW

Ciências da Saúde • 27/05/2026

REGISTRO DOI: [10.70773/revistatopicos/779735289](https://doi.org/10.70773/revistatopicos/779735289)

João Pedro Paulo de Andrade¹

Guilherme de Melo Lima Medeiros²

Luciana Clédina Bezerra Lopes³

Raquel Dantas Alves Figueiredo⁴

Milena Nunes Alves de Sousa⁵

RESUMO

Introdução: A Inteligência Artificial (IA) simula o raciocínio humano na resolução de problemas, apresentando alto potencial de processamento de dados com impactos crescentes no ambiente clínico e educacional médico. **Objetivo:** Analisar a contribuição da IA no treinamento do raciocínio clínico em estudantes de medicina.

Metodologia: Revisão sistemática baseada nas diretrizes PRISMA 2020. A busca, realizada em março de 2026, incluiu ensaios clínicos e estudos observacionais (2016-2026) que avaliaram o impacto da IA no ensino e na aquisição de conhecimentos práticos por graduandos. A qualidade metodológica foi aferida por escalas específicas. **Resultados:** Foram incluídos nove estudos, totalizando 573 participantes. As evidências indicam que métodos potencializados por IA superam abordagens tradicionais no desempenho acadêmico. Plataformas personalizadas e simulações baseadas em IA favoreceram a retenção de conhecimento e o ganho de habilidades práticas através de feedbacks instantâneos. Contudo, embora modelos como o ChatGPT-4.0 demonstrem precisão em conhecimentos básicos, sua acurácia em casos complexos pode ser inferior às diretrizes clínicas consolidadas.

Conclusão: A IA possui potencial considerável para elevar a qualidade da educação médica, devendo ser integrada de forma híbrida e complementar. Ela atua como uma ferramenta que solidifica a base lógica do aluno antes da prática clínica real, sem substituir o julgamento humano e o vínculo médico-paciente. Persistem, no entanto, desafios éticos e a necessidade de supervisão crítica constante.

Palavras-chave: Aprendizado Baseado em Problemas; Competência Clínica; Educação de Graduação em Medicina; Tecnologia Educacional.

ABSTRACT

Introduction: Artificial Intelligence (AI) simulates human reasoning in problem-solving, offering high data-processing potential with growing impacts on both clinical and medical educational environments. **Objective:** To analyze the contribution of AI to clinical reasoning training among medical students. **Methods:** A systematic review based on the PRISMA 2020 guidelines. The search, conducted in March 2026, included clinical trials and observational studies (2016–2026) evaluating the impact of AI on teaching and the acquisition of practical knowledge by undergraduate students. Methodological quality was assessed using specific scales. **Results:** Nine studies were included, totaling 573 participants. Evidence indicates that AI-enhanced teaching methods outperform traditional approaches in academic performance. Personalized learning platforms and AI-based simulations fostered knowledge retention and practical skill gains through instantaneous feedback. However, while models such as ChatGPT-4.0 demonstrate precision in basic knowledge, their accuracy in complex cases may be inferior to established clinical guidelines. **Conclusion:** AI holds considerable potential to enhance the quality of medical education and should be integrated in a hybrid and complementary manner. It serves as a tool to solidify the student's logical foundation before real clinical practice, without replacing human judgment and the doctor-patient bond. Nevertheless, ethical challenges and the need for constant critical supervision persist.

Keywords: Problem-Based Learning; Clinical Competence; Undergraduate Medical Education; Educational Technology.

1. INTRODUÇÃO

A Inteligência Artificial (IA) é uma área da Ciência da Computação que busca criar dispositivos (físicos e digitais) capazes de simular o raciocínio humano aplicado à resolução de problemas ou à tomada de decisões com base em dados externos. Já existe consenso entre pesquisadores e profissionais da área médica sobre a eficácia da IA em relação à Saúde Pública em grande parte devido à sua alta capacidade de processar dados e gerar respostas cada vez mais eficazes em um ambiente clínico. Ainda existem limitações associadas à IA, principalmente no que diz respeito ao desenvolvimento e à manutenção de relações positivas entre médicos e pacientes (Ferreira *et al.*, 2026).

Não faltaram novidades envolvendo o uso de *chatbots* de IA generativa (como o *ChatGPT* da *OpenAI*) em duas áreas específicas: atendimento ao paciente e educação médica, desde o lançamento dos *chatbots* *GAIC*. De modo geral, o potencial da IA na medicina reside na sua capacidade de processar grandes volumes de dados e auxiliar na identificação de soluções para problemas relacionados à saúde (Legoff *et al.*, 2025).

O raciocínio clínico é um conjunto complexo de habilidades e processos que os médicos utilizam para observar, analisar e interpretar dados para um diagnóstico e tratamento precisos, baseando-se fortemente em um sólido conhecimento biomédico e clínico. Esse raciocínio envolve o desenvolvimento e o aprimoramento de estruturas mentais, conhecidas como "pacotes", por meio da experiência clínica, o que aumenta a eficiência diagnóstica. Nesse sentido, a incorporação da inteligência artificial surge como uma ferramenta de apoio, capaz de auxiliar na organização e análise de informações clínicas, potencializando esse processo (Ripol *et al.*, 2025).

Além disso, os algoritmos utilizados para o diálogo automatizado em sistemas de IA foram treinados em bases de dados dispersas pela internet, que frequentemente contêm dados tendenciosos e, por vezes, imprecisos. Isto pode, em última análise, comprometer a tomada de decisões médicas. Adicionalmente, a utilização eficaz da IA exige que os profissionais sejam devidamente treinados nestas ferramentas tecnológicas para manter a ética e garantir a qualidade dos cuidados prestados aos pacientes (Kuvshinchiko; Sasaki, 2024).

O avanço da IA suscita preocupações significativas quanto ao potencial de desumanização da prática médica e à erosão da autonomia profissional. A dependência de sistemas automatizados e protocolos robotizados pode resultar em cuidados menos empáticos e acolhedores. Além disso, é importante que a IA seja utilizada como complemento, e não como substituta; essa abordagem salvaguarda a autonomia do médico na tomada de decisões, permitindo-lhe determinar o curso de ação mais adequado e personalizado para seus pacientes (Legoff *et al.*, 2025).

A crescente mercantilização da saúde levou empresas de tecnologia a investir em inteligência artificial com o objetivo de auxiliar na tomada de decisões clínicas. Para aumentar a eficiência durante os plantões, muitos médicos começaram a utilizar esses recursos; embora ofereçam vantagens operacionais, também podem prejudicar a qualidade da interação humana e o atendimento individualizado. Essa situação ilustra a interseção entre a medicina mercantilizada e o modelo biomédico, onde a produtividade é cada vez mais valorizada em detrimento de abordagens centradas no paciente. Além disso, a dependência excessiva dessas ferramentas pode fomentar a dependência da IA para processos de diagnóstico e

tratamento, diminuindo, assim, as habilidades de raciocínio clínico dos profissionais de saúde (Ferreira *et al.*, 2026).

Assim sendo, este estudo apresenta a seguinte problemática: como a inteligência artificial pode contribuir no treinamento de raciocínio clínico em estudantes de medicina e quais seus principais benefícios, limitações e implicações?

Nesse sentido, é importante examinar os efeitos da inteligência artificial na prática médica, particularmente entre os médicos de família, a fim de compreender as implicações de sua aplicação na qualidade da assistência e na autonomia dos profissionais. Conseqüentemente, este estudo busca identificar os principais desafios e oportunidades associados a essa aplicação, bem como o uso pretendido no contexto profissional (Ferreira *et al.*, 2026).

A justificativa para escolha deste tema se deu pela oportunidade de pesquisar e compreender um pouco mais sobre a atuação da inteligência artificial na formação médica, sobretudo, tendo em vista o atual cenário de digitalização dos serviços de saúde, bem como ainda por meio da elevada demanda e importância de se ter profissionais mais bem capacitados tecnologicamente para atuar em determinados casos. De tal modo, pesquisar sobre esta problemática se faz necessário, visto que este estudo servirá como embasamento teórico na orientação de práticas pedagógicas mais eficazes, assegurando deste modo que o uso da IA possa advir de maneira ética e agregada junto a isso às necessidades da formação de nossos médicos.

Deste modo, este estudo tem como objetivo geral, analisar como a inteligência artificial pode contribuir no treinamento de raciocínio

clínico em estudantes de medicina.

2. METODOLOGIA

Esta pesquisa constitui-se como uma revisão sistemática (RS) de literatura, fundamentada nas diretrizes do *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) 2020. A adoção desse método viabiliza, primordialmente, a análise dos impactos de intervenções na esfera da saúde ao observar requisitos que asseguram revisões com desfechos mais detalhados e rigorosos, o que simplifica a aplicação da prática baseada em evidências (PBE) (Page *et al.*, 2022).

A RS constitui uma abordagem criteriosa que fundamenta a PBE. Seu propósito é localizar, analisar e integrar pesquisas pertinentes a uma temática delimitada, empregando procedimentos metódicos e claros que assegurem a legitimidade, a utilidade prática e a redução de erros sistemáticos. Graças ao seu plano de trabalho minucioso, o processo pode ser replicado por outros especialistas, distinguindo-se das revisões narrativas por focar em questionamentos pontuais e adotar exigências mais severas na triagem e no exame das provas científicas (De-La-Torre-Ugarte-Guanilo; Takahashi; Bertolozzi, 2011). O registro desse estudo pode ser encontrado na plataforma *Prospective Register of Systematic Reviews Online* (PROSPERO), com o seguinte ID: CRD420261382258.

Esta RS será conduzida para responder à pergunta norteadora elaborada por meio da estratégia PICO, definindo-se como população (P) estudantes da graduação de medicina. A intervenção (I) uso de ferramentas baseadas em Inteligência Artificial aplicados especificamente ao treinamento de habilidades clínicas. A

comparação (C) métodos de ensino tradicionais e desfechos (O) impactos no desenvolvimento do raciocínio clínico, na capacidade de tomada de decisão e aquisição de competências clínicas específicas.

A busca sistematizada da literatura foi conduzida em Março de 2026 nas bases de dados eletrônicas *Medical Literature Analysis and Retrieval System Online* (MEDLINE/PUBMED), *Elsevier Editora LTDA*, *Literatura Latino-Americana e do Caribe em Ciências da Saúde* (LILACS), *Cochrane Library* e *Scopus*. Para isso, foram utilizados os Descritores em Ciências da Saúde (DeCS) ("*Artificial Intelligence*" OR "*Machine Learning*") AND ("*Clinical Reasoning*" OR "*Differential*") AND ("*Education, Medical*" OR "*Students, Medical*"), com filtro de data para textos publicados entre 2016 e 2026. Inicialmente, ao todo, foram identificados 121 artigos, que foram submetidos à triagem aos pares.

Para o seguimento da seleção dos artigos, foi utilizado a plataforma Rayyan para análise final dos artigos. Primeiramente, foram removidos as duplicatas. Após isso, foram lidos os títulos e resumos dos estudos restantes e aplicados os seguintes critérios de exclusão: Não respondem à questão norteadora; população que envolve médicos especialistas ou residentes; não foram feitos em humanos; utilização de IA de forma não pedagógica, estudos que não sejam ensaios clínicos ou estudos observacionais. Com a aplicação desses critérios, restaram-se 9 artigos para a RS.

A análise do rigor metodológico dos artigos incluídos será efetuada por meio de ferramentas específicas para cada tipo de desenho de estudo. No caso dos ensaios clínicos randomizados, empregar-se-á a Escala PEDro, que mensura a qualidade metodológica a partir de 11

itens fundamentais, incluindo critérios de elegibilidade, aleatorização, sigilo de alocação, similaridade inicial entre os grupos e o emprego de cegamento. É fulcral destacar que o primeiro item, que trata da validade externa e dos critérios de elegibilidade, não será incluído no cálculo da pontuação final. Assim, o escore total de cada estudo variará de 0 a 10 pontos, refletindo exclusivamente o rigor metodológico e a qualidade da descrição estatística (PEDRO, 2022). Por outro lado, para as pesquisas observacionais, será adotado a Escala de Newcastle-Ottawa (NOS), a qual mensura a qualidade a partir de três eixos fundamentais: os critérios de seleção dos grupos, o nível de comparabilidade entre as amostras e a fidedignidade na constatação do desfecho ou da exposição analisada (Wells *et al.*, 2021). Para a avaliação do risco de viés dos estudos, será utilizada a ferramenta *Risk of Bias* (RoB) 2.

3. RESULTADOS

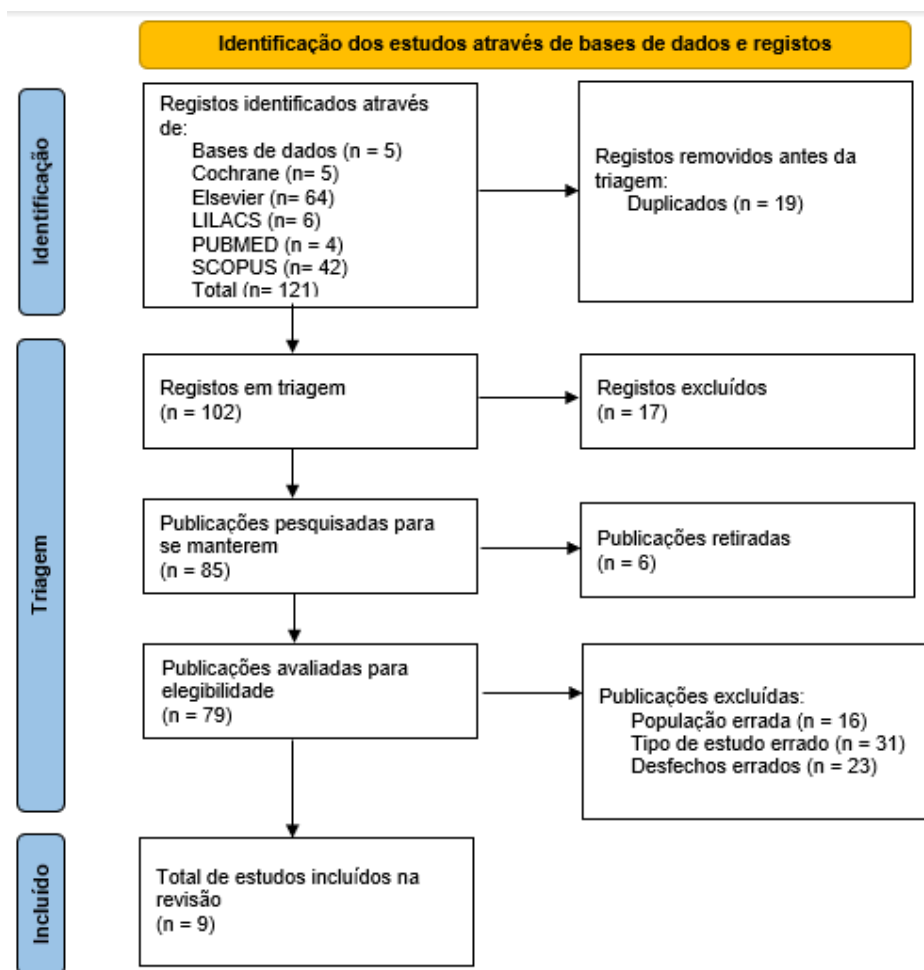
O processo de identificação e seleção dos estudos seguiu as recomendações das diretrizes PRISMA 2020, conforme detalhado na Figura 1. Inicialmente, a busca sistemática em bases de dados eletrônicas resultou em um total de 121 registros. Antes do início da triagem, foram removidos 19 registros duplicados, restando 102 artigos para a etapa de leitura de títulos e resumos.

Na fase de triagem inicial, baseada na análise de título e resumo, 17 registros foram excluídos por não atenderem aos critérios de elegibilidade ou por não estarem diretamente relacionados ao tema da Inteligência Artificial no treinamento do raciocínio clínico. Dessa forma, 17 artigos foram selecionados para a etapa de busca de texto completo. Destes, 6 artigos não foram recuperados devido à indisponibilidade de acesso ou falta de resposta dos autores,

totalizando 79 artigos avaliados para elegibilidade mediante leitura integral.

Durante a avaliação criteriosa dos textos completos, 70 estudos foram excluídos pelos seguintes motivos: população inadequada (n=16), intervenção não relacionada ao treinamento de raciocínio clínico (n=23) e desenhos de estudo não compatíveis com os critérios de inclusão (n=31). Ao final do processo, 9 estudos atenderam a todos os critérios de inclusão e foram selecionados para a síntese qualitativa e análise final desta revisão sistemática.

Figura 1: Fluxograma PRISMA de seleção dos estudos.



Fonte: Dados de pesquisa, 2026.

No Quadro 1, detalham-se os dados bibliográficos e as características estruturais dos estudos selecionados, permitindo uma visão

panorâmica da produção acadêmica sobre o tema. Observa-se que a literatura incluída é extremamente recente, com a totalidade dos artigos publicados entre os anos de 2024 e 2025, o que reflete o caráter emergente da Inteligência Artificial na educação médica. Quanto ao delineamento, predominam os ensaios clínicos randomizados (RCTs) e estudos de intervenção, indicando um esforço metodológico para estabelecer evidências de causa e efeito.

Em relação ao alcance geográfico e linguístico, os estudos abrangem diversos centros de pesquisa internacionais (como instituições na China, Coreia do Sul, Suécia, Turquia e Alemanha), evidenciando a globalização desta tecnologia, embora a linguagem de publicação predominante seja o inglês, garantindo maior circulação científica. No que diz respeito aos periódicos, as publicações concentram-se em revistas especializadas de alto impacto nas áreas de educação médica e informática em saúde, como o *BMC Medical Education* e o *JMIR Medical Education*, o que reforça a relevância e a validade acadêmica das discussões apresentadas nesta revisão.

Quadro 1: Informações gerais dos estudos incluídos.

Autores (ano)	Título	Idioma/País	Periódico	Tipo de estudo
Borg <i>et al.</i> (2025)	AI-Enhanced Social Robotic Versus Computer-Based Virtual Patients for Clinical Reasoning Training in Medical	Inglês Suécia	<i>Journal of Medical Internet Research (JMIR)</i>	Estudo de Coorte de Crossover Observacional

	Education: Observational Crossover Cohort Study			
Brügge <i>et al.</i> (2024)	Large language models improve clinical decision making of medical students through patient simulation and structured feedback: a randomized controlled trial	Inglês Alemanha	BMC Medical Education (BMC Med Educ)	Ensaio Clínico Randomiza do
Çiçek <i>et al.</i> (2025)	ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: a randomized controlled trial	Inglês Turquia	Postgraduat e Medical Journal	Ensaio clínico randomizad o
Kiyak <i>et al.</i> (2025)	AI Teaches Surgical Diagnostic Reasoning to Medical Students: Evidence from an Experiment Using a Fully Automated, Low-Cost Feedback System	Inglês Turquia	Journal of Surgical Education (J Surg Educ)	Estudo Observacio nal Prospectivo

Lee <i>et al.</i> (2025)	Comparing AI chatbot simulation and peer role-play for OSCE preparation: a pilot randomized controlled trial	Inglês Coréia do Sul	<i>BMC Medical Education (BMC Med Educ)</i>	Ensaio Clínico Randomizado
Montagna <i>et al.</i> (2025)	Impact of Clinical Decision Support Systems on Medical Students' Case-Solving Performance: Comparison Study with a Focus Group	Inglês Itália	<i>JMIR Medical Education (JMIR Med Educ)</i>	Estudo Experimental de Métodos Mistos
Wang <i>et al.</i> (2025)	Feasibility study of using GPT for history-taking training in medical education: a randomized clinical trial	Inglês China	<i>BMC Medical Education (BMC Med Educ)</i>	Ensaio Clínico Randomizado
Xu <i>et al.</i> (2025)	Virtual Standardized Patients for Improving Clinical Thinking Ability Training in Residents: Randomized Controlled Trial	Inglês China	<i>JMIR Medical Education</i>	Ensaio Clínico Randomizado

Zheng <i>et al.</i> (2024)	Application of AI-empowered scenario-based simulation teaching mode in cardiovascular disease education	Inglês China	<i>BMC Medical Education (BMC Med Educ)</i>	Ensaio Clínico Randomizado
----------------------------	---	-----------------	---	----------------------------

Fonte: Dados de pesquisa, 2026.

A análise detalhada dos estudos selecionados (Quadro 2) revela uma predominância do uso de Modelos de Linguagem de Grande Escala (LLMs), especificamente o ChatGPT (OpenAI), como a principal ferramenta de intervenção para o treinamento do raciocínio clínico. As intervenções variaram desde o fornecimento de *feedback* automatizado sobre o desempenho dos alunos até a criação de ambientes complexos de simulação com pacientes virtuais 3D.

A composição das amostras apresentou heterogeneidade significativa, variando de 16 a 178 participantes, totalizando 573 estudantes em diversos estágios da graduação. Quanto aos desfechos, os resultados apontam consistentemente para a eficácia da IA no suporte ao aprendizado, com melhorias estatisticamente significativas em exames clínicos e testes de raciocínio. Além do ganho técnico, observou-se maior entusiasmo, satisfação e redução da ansiedade entre os alunos.

Contudo, a superioridade da tecnologia não foi absoluta: enquanto a IA otimizou o tempo de resolução, o uso de diretrizes tradicionais (CPG) resultou em maior acurácia diagnóstica em cenários específicos e na comparação entre os feedbacks não foi evidenciado

diferenças significativas em relação aos de especialistas. Ademais, o aprendizado entre pares superou a IA no desenvolvimento de habilidades de anamnese, reforçando o papel da ferramenta como complementar ao julgamento humano e à interação interpessoal.

Quadro 2: Intervenções, amostras e principais desfechos dos estudos incluídos.

Autores (ano)	Intervenção	Amostra	Principais Desfechos
Borg <i>et al.</i> (2025)	Uso de SARI (interface robótica social com GPT-3.5) comparado a pacientes virtuais convencionais baseados em computador.	43 participantes	O sistema SARI superou a plataforma convencional em autenticidade, abordagem profissional, qualidade de tutoria e efeito de aprendizagem; 72% dos estudantes preferiram o robô social para o treino de raciocínio clínico.
Brügge <i>et al.</i> (2024)	Simulação de anamnese com ChatGPT-3.5, onde o grupo intervenção recebia feedback estruturado gerado pela IA.	60 participantes	O grupo que recebeu feedback superou significativamente o grupo controle em desempenho de tomada de decisão clínica após apenas quatro sessões; Houve melhoras específicas nos domínios de criação de contexto e segurança da informação.
Çiçek <i>et al.</i> (2025)	Comparação de feedback gerado pelo ChatGPT-3.5	129 participantes	Não houve diferença significativa no desempenho geral

	vs. feedback escrito por especialistas em testes formativos sobre infecções urinárias.		entre os grupos. Contudo, o feedback de especialistas foi superior em casos complexos, indicando que a IA pode carecer de nuances diagnósticas profundas.
Kiyak <i>et al.</i> (2025)	Sistema totalmente automatizado baseado em GPT-4o-mini para gerar questões de múltipla escolha e feedback personalizado sem intervenção humana.	100 participantes	Estudantes do 1º ano que usaram o sistema superaram estudantes do 2º ano (controle) no diagnóstico de condições de dor abdominal; A acurácia do conteúdo gerado pela IA foi confirmada por especialistas em 99,4% dos casos.
Lee <i>et al.</i> (2025)	Chatbot baseado em GPT-4o/Claude 3.5 para preparação de exame OSCE, fornecendo respostas simuladas e feedback automatizado.	19 participantes	Embora as pontuações no OSCE não tenham atingido diferença estatística, o grupo IA mostrou tendência de melhor desempenho em educação do paciente, enquanto o grupo de role-play humano foi melhor em anamnese.
Montagna <i>et al.</i> (2025)	Uso do ChatGPT como sistema de apoio à decisão clínica para resolver casos, comparado a diretrizes clínicas e repositórios online.	16 participantes	O grupo ChatGPT foi o mais rápido na resolução, mas o grupo que utilizou diretrizes clínicas (CPG) obteve a melhor acurácia diagnóstica em quase todos os domínios investigados.

Wang <i>et al.</i> (2025)	Simulações de pacientes baseadas em IA customizado para treinamento de habilidades de anamnese.	56 participantes	O grupo que treinou com ChatGPT obteve pontuações significativamente maiores no exame clínico estruturado pós-treino em comparação ao role-play tradicional (86,79 ± 5,46 vs 73,64 ± 4,76). Alunos relataram maior entusiasmo e menor ansiedade.
Xu <i>et al.</i> (2025)	Sistema de Pacientes Virtuais Padronizados 3D interativo para simular cenários clínicos de medicina interna.	84 participantes	O grupo experimental apresentou melhoras significativamente superiores em pontuações teóricas e testes de raciocínio clínico virtual; 90% dos estudantes acreditaram que o treino melhorou sua capacidade de pensamento clínico.
Zheng <i>et al.</i> (2024)	Ensino baseado em cenários de simulação feitos por IA (ChatGPT/Midjourney) para educação em doenças cardiovasculares.	66 participantes	Melhoria significativa no conhecimento teórico, habilidades operacionais e pensamento crítico clínico comparado ao ensino tradicional; Os estudantes relataram maior satisfação com o conteúdo e a participação individual.

Fonte: Dados de pesquisa, 2026.

No Quadro 3, apresenta-se a síntese da avaliação da qualidade metodológica dos estudos incluídos, realizada através das escalas

PEDro e NOS. De modo geral, os estudos apresentaram um rigor metodológico elevado, com as pontuações da escala PEDro variando entre 6 e 8 pontos, destacando-se os trabalhos de Xu *et al.* (2025) e Brügge *et al.* (2024) com o melhor desempenho (8/10). No que diz respeito aos estudos observacionais, a pontuação pela escala NOS situou-se entre 6 e 7 estrelas, refletindo uma boa seleção de grupos e comparabilidade, embora com limitações pontuais na aferição de desfechos a longo prazo.

Quadro 3: Escore da análise metodológica dos estudos incluídos.

Autores (ano)	Delineamento	Escala utilizada	Pontuação	Observações
Borg <i>et al.</i> (2025)	Observacional (Crossover)	NOS	7/9	Boa seleção e comparabilidade; desfecho baseado em autodeclaração/p referência.
Brügge <i>et al.</i> (2024)	Ensaio Clínico	PEDro	8/10	Estudo duplo-cego com análise de intenção de tratar e semelhança inicial entre grupos.
Çiçek <i>et al.</i> (2025)	Ensaio Clínico	PEDro	7/10	Randomização adequada; mascaramento dos avaliadores presente, mas sem mascaramento de terapeutas/alunos.

Kiyak <i>et al.</i> (2025)	Intervencio nal	PEDro	6/10	Limitações no mascaramento devido à natureza da intervenção digital.
Lee <i>et al.</i> (2025)	Ensaio Clínico	PEDro	6/10	Pontuação reduzida pelo tamanho da amostra (n=19) e falta de mascaramento total.
Montagna <i>et al.</i> (2025)	Estudo Comparativ o	NOS	(6/9)	Grupos pequenos e tempo de acompanhament o curto, impactando a representativida de.
Wang <i>et al.</i> (2025)	Ensaio Clínico	PEDro	7/10	Comparação estatística intergrupos robusta e medidas de precisão bem relatadas.
Xu <i>et al.</i> (2025)	Ensaio Clínico	PEDro	8/10	Excelente descrição de randomização e comparabilidade inicial dos grupos de residentes.
Zheng <i>et al.</i> (2024)	Experiment al	NOS	6/9	Perda de pontos na randomização (quase- experimental) e

				potencial viés de seleção.
--	--	--	--	----------------------------

Fonte: Dados de pesquisa, 2026.

Na Figura 2, observa-se a análise do risco de viés mensurado no RoB 2, em que os estudos analisados apresentaram, em sua maioria, baixo risco de viés e algumas preocupações em uma parcela. Os pontos críticos concentraram-se na randomização, pela ausência de detalhes sobre a alocação dos grupos, e no domínio de medição de desfechos, prejudicado pela inviabilidade do cegamento por parte dos participantes em contextos educacionais tecnológicos.

Figura 2: Risco de viés dos estudos incluídos.

Study ID	Experimental	Comparator	D1	D2	D3	D4	D5	Overall	
Borg et al. (2025)	Plataforma Virtual com IA	Plataforma Virtual sem IA	!	+	+	+	+	!	+ Low risk ! Some concerns - High risk
Brügge et al. (2024)	Simulação + Feedback com IA	Simulação	+	+	+	+	+	+	
Çiçek et al. (2025)	Feedback por IA	Feedback por Especialista	+	+	+	+	+	+	
Kiyak et al. (2025)	Questões + Feedback gerado por IA	Ensino tradicional	!	+	+	+	+	!	D1 Randomisation process D2 Deviations from the intended interventions D3 Missing outcome data D4 Measurement of the outcome D5 Selection of the reported result
Lee et al. (2025)	Chat Bot	Tutoria Humana	+	+	+	+	+	+	
Montagna et al. (2025)	ChatGPT	Guia de Práticas Clínicas e RO	+	+	+	+	+	+	
Wang et al. (2025)	Simulação ChatGPT	Simulação Tradicional	+	+	+	+	+	+	
Xu et al. (2025)	Treinamento Virtual	Treinamento Tradicional	+	+	+	+	+	+	
Zheng et al. (2024)	Ensino com auxílio de IA	Ensino tradicional	!	+	+	+	+	!	

Fonte: Dados de pesquisa, 2026.

4. DISCUSSÃO

A integração de ferramentas de Inteligência Artificial (IA) no processo de formação acadêmico médico promove um ganho substancial na capacidade diagnóstica e na elaboração de planos terapêuticos.

O estudo de Xu *et al.* (2025) demonstrou que a utilização de um Sistema de Pacientes Virtuais Padronizados 3D resultou em melhorias significativamente superiores nas pontuações de raciocínio clínico virtual do grupo intervenção. Segundo os resultados, a imersão em cenários clínicos simulados permitiu que

os estudantes refinassem sua lógica de investigação, refletindo-se diretamente em uma maior precisão na identificação de patologias e na proposta de tratamentos adequados.

Esses achados convergem diretamente com os resultados apresentados por Wang *et al.* (2025), cujo ensaio clínico randomizado demonstrou que o grupo submetido ao treinamento com simulações de pacientes baseadas em IA obteve pontuações significativamente maiores no exame clínico estruturado pós-treinamento em comparação ao grupo controle, que utilizou pacientes simulados de forma tradicional. O grupo que utilizou simulações baseadas utilizando o ChatGPT alcançou uma pontuação média de 86,79 (\pm 5,46), enquanto o grupo submetido à simulação com atores humanos obteve 73,64 (\pm 4,76). Esta diferença, com alto rigor estatístico ($p < 0,001$).

Essa eficácia pedagógica é corroborada historicamente pelo estudo de Botezatu *et al.* (2010), que já indicava que simuladores de pacientes virtuais são intervenções poderosas para o aprendizado efetivo e a retenção de conhecimento. O estudo de 2010 reforça que o uso de sistemas simuladores proporciona resultados de retenção superiores aos métodos tradicionais, especialmente devido ao fornecimento de feedback imediato e à prática repetitiva, elementos que consolidam as bases do raciocínio clínico e facilitam a transferência de habilidades para a prática real.

O impacto dessas ferramentas é potencializado quando a simulação é acompanhada de devolutivas analíticas, conforme demonstrado por Brügge *et al.* (2024). O estudo evidenciou que o diferencial pedagógico não reside apenas na interação com o paciente virtual, mas na oferta de feedbacks estruturados pela IA, que permitem ao

estudante identificar falhas em seu raciocínio em tempo real. Enquanto o grupo controle, que realizou apenas as conversas clínicas com a IA, manteve um desempenho estável, o grupo intervenção que recebeu feedbacks personalizados apresentou uma melhora estatisticamente significativa na tomada de decisão clínica ($p = 0,049$). Essa evidência reforça que a IA, ao atuar como uma ferramenta de correção automática, transforma o erro em uma oportunidade de aprendizado imediato, o que favorece uma curva de desenvolvimento mais sólida e consciente para os discentes.

A relevância dos feedbacks personalizados na formação médica é amplamente corroborada pela literatura, que identifica a capacidade da IA em oferecer ambientes de aprendizagem adaptativos como um fator determinante para a melhoria do desempenho acadêmico. Conforme o estudo de Uberuaga *et al.* (2024), a implementação dessas ferramentas permite a redução do hiato entre o aprendizado e a avaliação ao analisar individualmente os pontos fortes e fracos de cada aluno. Essa tecnologia destaca-se por fornecer experiências customizadas que auxiliam na identificação de lacunas de conhecimento, demonstrando um impacto positivo considerável tanto na retenção de conteúdo teórico quanto no desenvolvimento de competências clínicas.

Devido a isso, o uso de tecnologias inteligentes está se incorporando cada vez mais na educação médica e muitos estudantes veem essa situação com uma perspectiva positiva, pois afirmaram que a ferramenta auxilia na organização de estudos, na aprendizagem de conceitos médicos e no auxílio à decisão clínica diagnóstica. Segundo o estudo de Sami *et al.* (2025), cerca de 8 em cada 10 estudantes mantêm uma atitude favorável à IA, sendo que 6 em cada 10 alunos relatam que a tecnologia otimiza o tempo de estudo

e fornece respostas mais precisas para dúvidas curriculares. Nesse sentido, todos esses benefícios intensificam a imersão e o rendimento nas sessões de estudos, reduzindo o nível de estresse estudantil e impactando positivamente na saúde mental dos acadêmicos.

De acordo com Borg *et al.* (2025), o uso da IA no treinamento do raciocínio clínico traz benefícios diretos relacionados à satisfação educacional, com 72% dos estudantes preferindo simulações com robôs sociais aprimorados por IA em vez de plataformas virtuais tradicionais. Os estudantes apresentaram melhor comunicação, segurança e satisfação durante as sessões, descrevendo a IA como um "espaço seguro de experimentação" onde o erro não gera julgamento social. Dessa forma, é visível que esse tipo de intervenção pode aumentar o rendimento, reduzir os níveis de ansiedade e a insegurança, comparado ao treinamento acompanhado de professores que tendem a gerar maior pressão social sobre o aluno.

Além disso, Wang *et al.* (2025) juntamente com Zheng *et al.* (2024) mostraram que a IA proporciona um ambiente seguro e individualizado para o estudante, permitindo privacidade, maior liberdade nas escolhas e feedbacks adaptados. Os resultados práticos de Wang mostram que alunos que treinaram com IA alcançaram notas de 87 pontos em exames clínicos, enquanto o grupo tradicional ficou com 73 pontos. Graças à capacidade de comunicação da IA e ao acesso a diversos conteúdos, ela permite múltiplas sessões voltadas ao diagnóstico e à interação médico-paciente. Dessa forma, ambos os estudos deixam exposto que os usuários apresentaram altos níveis de satisfação com o ambiente de

ensino, sentindo-se imersos e automotivados, o que culmina na redução do estresse mental e na melhora do rendimento.

Todavia, mesmo com a presença cada vez mais marcante, o uso das IA's possui limitações críticas, como fornecimento de informações duvidosas. Devido ao seu acesso a conteúdos dispersos na internet, a ferramenta pode apresentar tendenciosismo nas respostas, interpretação limitada de casos complexos e a necessidade indispensável de supervisão profissional. Portanto, os estudos atuais não podem generalizar totalmente os dados obtidos, uma vez que muitas pesquisas possuem espaços amostrais reduzidos e riscos de vieses de seleção e informação (Kuvshinchikov e Sasaki, 2024).

De acordo com o estudo de Kiyak *et al.* (2025), o sistema automatizado por IA mostrou potencial ao registrar uma taxa de acerto de 87% no diagnóstico de dor abdominal no grupo que usou a ferramenta, contra apenas 15% do grupo que não usou. Entretanto, o estudo teve um número modesto de amostras, o que dificulta que esses resultados sejam aplicados de forma universal. Além disso, a falta de uma randomização completa em certos critérios leva à interpretação de que os resultados podem apresentar vieses de informação, não oferecendo ainda uma resposta científica concreta.

O estudo de Montagna *et al.* (2025) mostra que o desempenho na decisão clínica melhorou em aspectos de velocidade e construção do raciocínio, porém, a acurácia da IA (nota 3,7) foi inferior à das diretrizes médicas tradicionais (nota 4,0). Percebeu-se que o tipo de abordagem do estudante com a máquina afeta diretamente o resultado, pois comandos (prompts) imprecisos levaram a IA a cometer erros graves. Portanto, o uso da IA no treinamento ainda é impreciso para alunos iniciantes que não sabem formular perguntas

estruturadas, tornando o auxílio de um docente obrigatório para evitar erros de julgamento.

Segundo Lee *et al.* (2025), o uso da IA aprimora o raciocínio para exames práticos como o Exame Clínico Objetivo Estruturado (OSCE), mas o estudo aponta limitações severas por ser uma pesquisa piloto com apenas 19 participantes. Nesse sentido, é notório que esse tipo de intervenção ainda é limitado e seus resultados não podem ser generalizados, pois o pequeno grupo e desequilíbrios de gênero podem intervir nos dados. Nota-se ainda que a supervisão profissional é essencial, já que falhas técnicas no servidor ou imprecisões no vocabulário médico da máquina podem induzir o estudante ao erro.

Sob a ótica do estudo de Zheng *et al.* (2024), demonstra-se que a IA fornece feedbacks e estimula a procura individual do estudante para melhorar seus pontos fracos, mas a amostra de apenas 66 indivíduos limita a validade dos achados. Além do número reduzido de envolvidos, existe o limite imposto pela própria capacidade de comunicação do estudante, que muitas vezes não consegue transmitir informações à máquina de maneira adequada. Isso impede uma discussão clínica eficaz e reforça que a IA, no estágio atual da literatura, deve ser vista como um complemento e não como um substituto do ensino humano.

A partir das evidências analisadas, torna-se imperativo discutir como as instituições de ensino superior podem integrar esses dados em suas matrizes curriculares. A aplicação prática da IA nas faculdades de medicina deve ser compreendida não como um substituto para o preceptor ou para o contato humano com o paciente, mas como um modelo pedagógico complementar e robusto de preparação para a

prática clínica. Esses sistemas podem ser utilizados para preencher lacunas críticas no treinamento tradicional, como a escassez de casos clínicos variados e a dificuldade de fornecer feedback individualizado em tempo real para turmas com demanda elevada.

Ao implementar simulações baseadas em IA, as instituições oferecem um ambiente seguro para o erro, onde o estudante pode refinar seu raciocínio clínico e habilidades de anamnese antes de ser exposto a cenários reais. Essa experiência permite ao aluno chegar ao internato com uma base lógica mais sólida, permitindo que o tempo de contato humano com o paciente seja focado em aspectos que a tecnologia ainda não alcança, como o exame físico presencial, a sensibilidade clínica e a construção do vínculo médico-paciente.

A inserção da Inteligência Artificial no contexto acadêmico deve ser realizada de forma estruturada e gradual. Conforme destacam Rivera e Cedeño (2024), o maior desafio para as instituições de saúde não é apenas a tecnologia em si, mas a formação docente, que se tornou um imperativo para que os professores consigam atualizar-se e guiar os estudantes no uso ético e produtivo dessas ferramentas. Assim, a melhor forma de implementação envolve um modelo onde a IA apoie a aprendizagem adaptativa e a tutoria personalizada, garantindo que o corpo docente esteja capacitado para mediar essa inovação. Ao alinhar a atualização pedagógica com as potencialidades da IA, as instituições asseguram que a tecnologia não substitua o ensino tradicional, mas atue como um catalisador que prepara o aluno para os desafios da carreira médica.

A predominância no risco de viés dos estudos selecionados, especialmente no domínio de randomização dos participantes e medição do desfecho, reflete um desafio metodológico intrínseco às

intervenções educacionais baseadas em IA. A inviabilidade do cegamento de participantes e pesquisadores nesses cenários é um fator crítico, uma vez que a ciência do design do estudo nem sempre permite ocultar a natureza da intervenção tecnológica.

Conforme discutido por Monaghan *et al.* (2021), o cegamento é um pilar fundamental para minimizar vieses de performance e de detecção; contudo, em ensaios onde a intervenção é comportamental ou educacional, a sua implementação torna-se complexa ou impraticável. Essa transparência forçada pode introduzir um viés de engajamento, onde o entusiasmo dos alunos pela nova tecnologia influencia positivamente os resultados, ressaltando a necessidade de interpretar os ganhos de desempenho com cautela e priorizar desfechos objetivos e padronizados para mitigar tais limitações.

A produção de novos estudos é fundamental, visto que este trabalho apresenta limitações metodológicas, como risco de vieses na seleção, interpretação e registro de dados, que estão sujeitos à subjetividade dos autores. Além disso, por se tratar de um tema contemporâneo e em ascensão, a escassez de informações pode limitar a concretização dos resultados. À medida que surgirem novas descobertas nesta área ainda não totalmente explorada, a literatura científica voltada à IA e à educação médica estudantil passará por constante evolução.

5. CONCLUSÃO

A presente RS demonstra que a integração da IA no ensino médico possui um potencial transformador, sobretudo, no que tange ao fortalecimento do raciocínio clínico e à personalização da

aprendizagem. As evidências indicam que ferramentas baseadas em modelos de linguagem e simulações virtuais oferecem um ambiente seguro para o erro, permitindo que os estudantes realizem práticas deliberadas com feedbacks imediatos e adaptados às suas necessidades individuais. Essa tecnologia não apenas otimiza a retenção de conhecimento teórico, mas também prepara o discente para a prática com uma base lógica mais sólida.

Entretanto, a eficácia dessas ferramentas está intrinsecamente ligada à sua aplicação como um modelo pedagógico complementar, e não substitutivo. Embora a IA promova agilidade na resolução de casos, ela ainda pode apresentar acurácia inferior às diretrizes clínicas estabelecidas, o que impõe o risco de indução ao erro caso haja uma confiança passiva por parte do usuário. Portanto, a inserção da IA nas matrizes curriculares deve ser acompanhada por um sólido treinamento em letramento digital e pensamento crítico, capacitando o futuro médico a validar as sugestões da máquina à luz das evidências científicas e da ética profissional.

Estudos futuros são necessários para avaliar o impacto de longo prazo dessa tecnologia na performance clínica real e para estabelecer marcos regulatórios que garantam a equidade e a segurança na educação médica mediada por algoritmos.

REFERÊNCIAS BIBLIOGRÁFICAS

BORG, Alexander *et al.* AI-enhanced social robotic versus computer-based virtual patients for clinical reasoning training in medical education: observational crossover cohort study. **Journal of Medical Internet Research**, v. 27, p. e82541, 2025. DOI: 10.2196/82541.

BOTEZATU, Mihaela *et al.* Virtual patient simulation: Knowledge gain or knowledge loss? **Medical Teacher**, v. 32, n. 7, p. 562-568, 2010.

BRÜGGE, Emilia *et al.* Large language models improve clinical decision making of medical students through patient simulation and structured feedback: a randomized controlled trial. **BMC Medical Education**, v. 24, n. 1, p. 1391, 2024. DOI: 10.1186/s12909-024-06399-7.

CIÇEK, Feray Ekin *et al.* ChatGPT versus expert feedback on clinical reasoning questions and their effect on learning: a randomized controlled trial. **Postgraduate Medical Journal**, v. 101, n. 1195, p. 458-463, 2025. DOI: 10.1093/postmj/qgae170.

DE-LA-TORRE-UGARTE-GUANILO, Mônica Cecília; TAKAHASHI, Renata Ferreira; BERTOLOZZI, Maria Rita. Revisão sistemática: noções gerais. **Revista da Escola de Enfermagem da USP**, v. 45, n. 5, p. 1260–1266, out. 2011.

FERREIRA, Guilherme Freitas Bernardo *et al.* How can generative intelligence facilitate the teaching of clinical reasoning: a scoping review. **Journal of Higher Education Teaching**, v. 15, p. 1-27, 2026. DOI: [10.35699/2237-5864.2025.5833](https://doi.org/10.35699/2237-5864.2025.5833).

KIYAK, Yavuz Selim *et al.* AI teaches surgical diagnostic reasoning to medical students: evidence from an experiment using a fully automated, low-cost feedback system. **Journal of Surgical Education**, v. 82, n. 10, p. 103639, 2025. DOI: 10.1016/j.jsurg.2025.103639.

KUVSHINCHIKOV, Vladimir; DE SOUSA SASAKI, Thiago. ChatGPT na medicina do século XXI: avanços, desafios éticos e limitações

inerentes. **Health Residencies Journal**, v. 5, n. 23, 2024. DOI: 10.51723/hrj.v5i23.994.

LEE, Hye-Yoon *et al.* Comparing AI chatbot simulation and peer role-play for OSCE preparation: a pilot randomized controlled trial. **BMC Medical Education**, v. 25, n. 1, p. 1755, 2025. DOI: 10.1186/s12909-025-08308-y.

LEGOFF, Waleska Gomes da Rocha *et al.* Artificial intelligence in medicine: an experience report on the use of IA by doctors and its impacts. **Brazilian Journal of Implantology and Health Sciences**, v. 7, n. 5, p. 1257-1270, 2025. DOI: 10.36557/2674-8169.2025v7n5p1257-1270.

MONAGHAN, Thomas F *et al.* Blinding in Clinical Trials: Seeing the Big Picture. **Medicina**, v. 57, n. 7, p. 647, 24 jun. 2021. Disponível em: <https://doi.org/10.3390/medicina57070647>.

MONTAGNA, Marco *et al.* Impact of clinical decision support systems on medical students' case-solving performance: comparison study with a focus group. **JMIR Medical Education**, v. 11, p. e55709, 2025. DOI: 10.2196/55709.

RIVERA, Fernanda Lilibeth Ormeño; CEDEÑO, Michelle Antonella Sánchez. Training in the Field of Artificial Intelligence: Challenges and Opportunities in Health Science Education. **EthAlca**, [S. l.], v. 3, n. 134, 2024.

PAGE, Matthew J. *et al.* A declaração PRISMA 2020: diretriz atualizada para relatar revisões sistemáticas. **Revista Panamericana de Salud Pública**, [S. l.], v. 46, p. 1-15, 2022. Disponível em: <https://doi.org/10.26633/rpsp.2022.112>.

PEDRO. Escala PEDro. **Sydney: Physiotherapy Evidence Database**, [2022]. Disponível em: <https://pedro.org.au/portuguese/resources/pedro-scale/>. Acesso em: 17 abr. 2026.

RIPOL, Mariana de Oliveira *et al.* Artificial intelligence in medical education: a scoping review on the development of clinical reasoning. **Journal of Interprofessional Health Education**, v. 2, n. 1, 2015. DOI: [10.4025/jinterprofhealtheduc.v2i1.76833](https://doi.org/10.4025/jinterprofhealtheduc.v2i1.76833).

SAMI, Abdul *et al.* Medical students' attitudes toward AI in education: perception, effectiveness, and its credibility. **BMC Medical Education**, v. 25, n. 1, p. 82, 2025. DOI: [10.1186/s12909-025-06704-y](https://doi.org/10.1186/s12909-025-06704-y).

UBERUAGA, Vanessa *et al.* Applications and challenges of Artificial Intelligence impact on the academic performance of medical students: a Scoping Review. **JMIR Preprints**, [S. l.], 2024. Disponível em: <https://doi.org/10.2196/preprints.69724>.

WANG, Zhen *et al.* Feasibility study of using GPT for history-taking training in medical education: a randomized clinical trial. **BMC Medical Education**, v. 25, n. 1, p. 1030, 2025. DOI: [10.1186/s12909-025-07614-9](https://doi.org/10.1186/s12909-025-07614-9).

WELLS, George Anthony *et al.* The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomised studies in meta-analyses. **Ottawa: Ottawa Hospital Research Institute**, [2021]. Disponível em: <https://ohri.ca/en/who-we-are/core-facilities-and-platforms/ottawa-methods-centre/newcastle-ottawa-scale>.

XU, Liyuan *et al.* Virtual standardized patients for improving clinical thinking ability training in residents: randomized controlled trial.

ZHENG, Koulong *et al.* Application of AI-empowered scenario-based simulation teaching mode in cardiovascular disease education. **BMC Medical Education**, v. 24, p. 1003, 2024. DOI: 10.1186/s12909-024-05977-z.

¹ Discente do Curso Superior de Medicina do Instituto Centro Universitário *Campus Patos*-PB. E-mail: [acesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0009-0001-4294-0047>

² Discente do Curso Superior de Medicina do Instituto Centro Universitário *Campus Patos*-PB. E-mail: [acesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0009-0006-8873-1570>

³ Discente do Curso Superior de Medicina do Instituto Centro Universitário *Campus Patos*-PB. E-mail: [acesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0009-0002-2145-2447>

⁴ Médica de Família e Comunidade pelo Centro Universitário de Patos. Docente e Preceptora no Centro Universitário de Patos *Campus Patos*-PB. E-mail: [acesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0000-0001-6058-865X>

⁵ Doutora e Pós-Doutora em Promoção da Saúde. Pró-Reitora de Pós-graduação, Pesquisa e Extensão e Docente no Centro Universitário de Patos *Campus Patos*-PB. E-mail: [acesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0000-0001-8327-9147>

