

**INTELIGÊNCIA ARTIFICIAL
NA AVALIAÇÃO DE
COMPETÊNCIAS TÉCNICAS
E RACIOCÍNIO
DIAGNÓSTICO NA
EDUCAÇÃO MÉDICA**

**ARTIFICIAL INTELLIGENCE IN THE ASSESSMENT OF TECHNICAL SKILLS
AND DIAGNOSTIC REASONING IN MEDICAL EDUCATION**

Ciências da Saúde • 17/05/2026

REGISTRO DOI: [10.70773/revistatopicos/778904160](https://doi.org/10.70773/revistatopicos/778904160)

Maria Izadora de Caldas Francelino¹

Maria Eduarda Dantas Peronico Sobral²

Lara Vitória Evangelista de Araújo³

Iago Brenner Farias Leal⁴

Fernanda Clara Marinheiro Oliveira⁵

Milena Nunes Alves de Sousa⁶

RESUMO

Introdução: A educação médica está diretamente ligada à aquisição de um conjunto de competências técnicas e habilidades clínicas. Entre as ferramentas para realização dessas habilidades existem os métodos tradicionais e, como uma alternativa, está sendo aplicado o uso da inteligência artificial (AI). **Objetivo:** Analisar as aplicações da Inteligência Artificial na verificação de competências técnicas comparado aos métodos tradicionais de avaliação, na melhora da precisão e a eficácia na aferição das competências técnicas e do raciocínio diagnóstico para estudantes da área médica. **Método:** Foi realizada uma revisão sistemática, tendo como base a pergunta PICO: “Para estudantes da área médica, o uso de Inteligência Artificial, comparado aos métodos tradicionais de avaliação, melhora a precisão e a eficácia na avaliação das competências técnicas e do raciocínio diagnóstico?”. A busca dos textos foi efetuada em seis bases de dados distintas: *U.S. National Institutes of Health's National Library of Medicine* (NIH/NLM - PubMed), *Scientific Electronic Library Online* (Scielo), Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), *Medical Literature Analysis and Retrieval System Online* (MEDLINE), Biblioteca Virtual em Saúde (BVS) e Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS), e seguiu as diretrizes da *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) para otimizar o rigor metodológico. **Resultados:** Os estudos demonstraram que a IA mostra-se como uma ferramenta complementar ao ensino tradicional, apesar de apresentar vantagens, requer supervisão humana, não substituindo plenamente o ensino tradicional. **Conclusão:** Diante do exposto, notou-se que apesar do potencial inovador do uso da IA, ainda existem limitações importantes a serem superadas. As evidências indicam que um modelo de ensino híbrido e progressivo, que

combine recursos tecnológicos e métodos tradicionais, é o mais adequado.

Palavras-chave: Inteligência Artificial; Medicina; Treinamento por Simulação.

ABSTRACT

Introduction: Medical education is directly linked to the acquisition of a set of technical skills and clinical abilities. Among the tools for developing these skills are traditional methods and, as an alternative, the use of artificial intelligence (AI) is being applied.

Objective: To analyze the applications of Artificial Intelligence in verifying technical skills compared to traditional assessment methods, in improving the accuracy and effectiveness of measuring technical skills and diagnostic reasoning for medical students.

Method: A systematic review was conducted, based on the PICO question: "For medical students, does the use of Artificial Intelligence, compared to traditional assessment methods, improve the accuracy and effectiveness in evaluating technical skills and diagnostic reasoning?". The search for texts was conducted in six distinct databases: U.S. National Institutes of Health's National Library of Medicine (NIH/NLM - PubMed), Scientific Electronic Library Online (SciELO), Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Medical Literature Analysis and Retrieval System Online (MEDLINE), Virtual Health Library (BVS), and Latin American and Caribbean Literature in Health Sciences (LILACS), and followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to optimize methodological rigor. **Results:** The studies demonstrated that AI is a complementary tool to traditional teaching, although it presents advantages, it requires human supervision and does not fully replace traditional teaching.

Conclusion: In light of the above, it was noted that despite the innovative potential of using AI, there are still important limitations to overcome. Evidence indicates that a hybrid and progressive teaching model, combining technological resources and traditional methods, is the most appropriate.

Keywords: Artificial Intelligence; Medicine; Simulation Training.

INTRODUÇÃO

O aprendizado na medicina está diretamente ligado à aquisição de um conjunto diversificado e complexo de competências técnicas, incluindo o desenvolvimento de *hard skills*, que categorizam-se em habilidades clínicas essenciais e mensuráveis, como a formação de raciocínio diagnóstico, execução de exames ou procedimentos. Para além disso, avalia-se, a abordagem verbal e atitudinal entre pacientes e colegas, que são consideradas *soft skills*. Para a verificação da obtenção dessas habilidades, exigem-se avaliações contínuas, cujos métodos devem ter constante evolução e atualização. (Levy; Taquette; Souza, 2026; Mendes *et al.*, 2023).

Portanto, visando assegurar a aquisição dessas habilidades e conseqüentemente, melhora na qualidade no cuidado, um monitoramento rigoroso é mantido durante toda a formação. Apesar das atualizações curriculares e normativas, a simulação de situações encontradas na prática médica se consolida como uma ferramenta global, aplicada a todos os envolvidos no processo de ensino na medicina (Ney; Martins; Firmida, 2026).

Entre as ferramentas para realização dessas avaliações, têm-se os métodos tradicionais, como os testes de múltipla escolha que verificam conhecimento teórico. Por outro lado, para atender às

exigências das Diretrizes Curriculares Nacionais (DCNs), aplica-se provas como o Exame Clínico Objetivo Estruturado (OSCE), que avalia o estudante de maneira multidimensional e proporcionando experiências práticas estruturadas e passíveis de repetição, garantindo uma perícia prévia que melhora a competência de quem realizará a ação posteriormente (Fraga *et al.*, 2025).

Apesar de a OSCE já possuir uma eficácia comprovada e os testes de múltipla escolha serem amplamente utilizados, ainda são ferramentas limitadas, exigindo alto investimento financeiro e tempo disponibilizado. No caso da simulação, é necessária a realização de treinamento da equipe que a aplicará, aquisição de equipamentos especializados. Além disso, outra desvantagem é a suscetibilidade a vieses humanos, como fadiga, variabilidade de critérios cognitivos entre os avaliadores e atraso na divulgação de resultados que ainda podem ser subjetivos, limitando uma correção imediata no aprendizado (Elendu, 2024).

Assim, a inteligência artificial (AI) é visualizada como uma alternativa para auxiliar na avaliação de competências técnicas na educação médica. Pois, as tecnologias disponíveis como, Processamento de Linguagem Natural (PLN), aprendizado de máquina (AM) e visão computacional, tem capacidade de aprimorar diversas lacunas dos métodos tradicionais, tornar a experiência de simulação mais acessível, além de oferecer uma análise rápida e objetiva do ponto avaliado (Elendu, 2024).

Todavia, a transição para métodos de avaliação guiados por IA não ocorre sem ajustes técnicos, pedagógicos, curriculares e éticos. Afinal, sem supervisão adequada pode induzir preconceitos que afetem pacientes ou indução a respostas incertas para o contexto

em que está inserido. Portanto, é necessária cautela na instalação de novos modelos que não prezam pelas *soft skills* (Mendes *et al.*, 2023; Weidner; Fisher, 2024).

O presente artigo tem como objetivo analisar as aplicações da Inteligência Artificial na verificação de competências técnicas comparado aos métodos tradicionais de avaliação, na melhora da precisão e a eficácia na aferição das competências técnicas e do raciocínio diagnóstico para estudantes da área médica. Falta, na literatura vigente, uma síntese rigorosa que avalie criticamente a eficácia, a validade e as limitações dessas tecnologias automáticas quando comparados ao padrão ouro. Fornecendo, assim, um panorama claro que direcione futuras pesquisas.

MÉTODO

O presente estudo consiste em uma revisão sistemática, utilizando de ensaios clínicos randomizados (ECRs) para verificar as aplicações da Inteligência Artificial na avaliação de competências técnicas e raciocínio diagnóstico na educação médica. Uma revisão sistemática é responsável por considerar as mais relevantes evidências científicas disponíveis, permitindo subsidiar pesquisas futuras (Linares-Espinós *et al.*, 2018). A filtragem dos dados foi realizada de acordo com as diretrizes da Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Galvão; Tiguman; Sarkis-Onofre, 2022) visando assegurar a confiabilidade dos resultados da pesquisa. Foi feito o registro do protocolo na plataforma PROSPERO 2026 CRD420261386895, sob o endereço <https://www.crd.york.ac.uk/PROSPERO/view/CRD420261386895>.

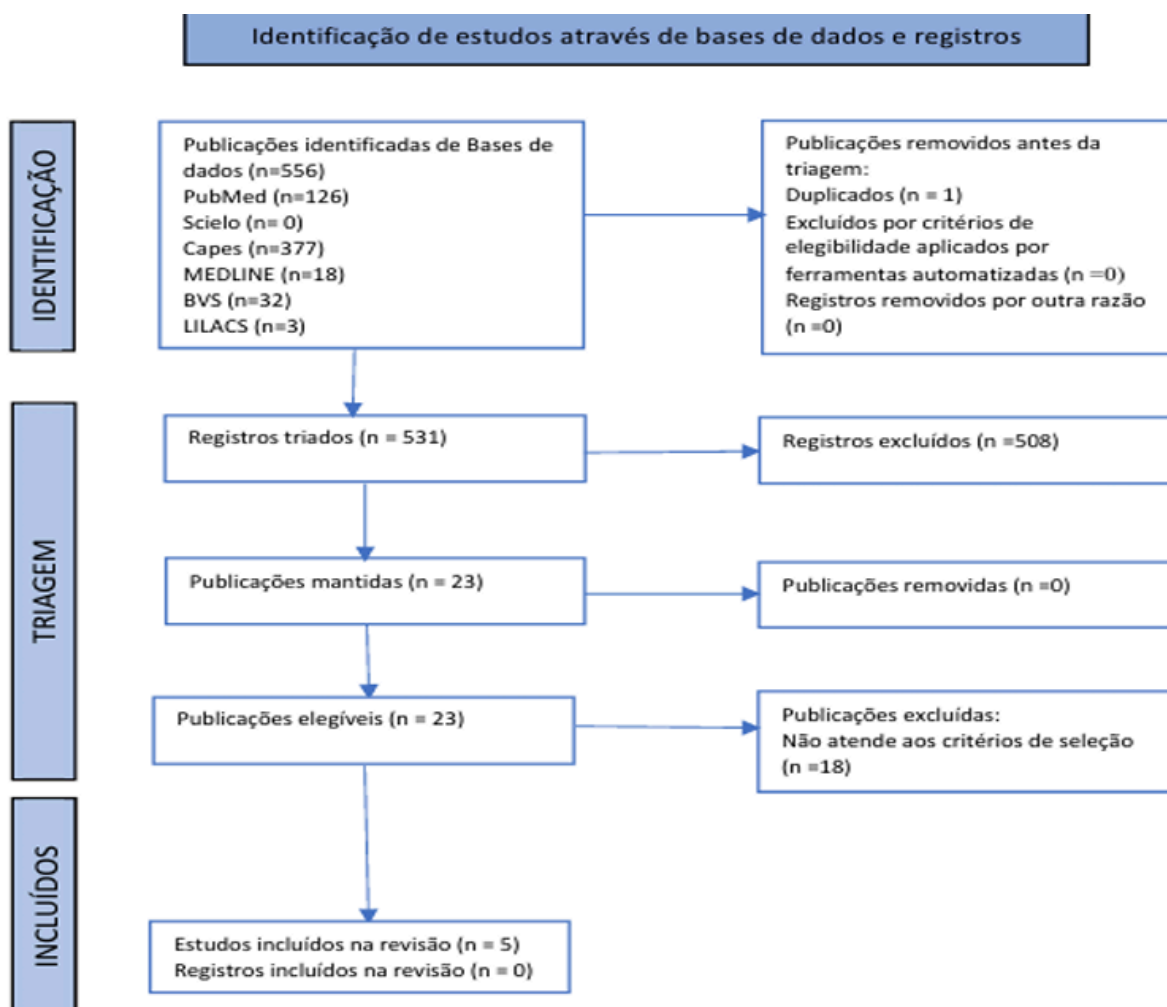
Diante disso, a questão de pesquisa foi elaborada com base na estrutura do acrônimo PICO (Paciente, Intervenção, Controle, Outcomes). Nessa perspectiva, a pergunta estruturada foi: “Para estudantes da área médica, o uso de Inteligência Artificial, comparado aos métodos tradicionais de avaliação, melhora a precisão e a eficácia na avaliação das competências técnicas e do raciocínio diagnóstico?” Nesse contexto, os elementos do PICO foram estabelecidos como: P - Estudantes da área médica; I - uso de Inteligência Artificial na avaliação; C - Métodos avaliativos tradicionais; e O - precisão, eficácia e melhoria na avaliação das competências clínicas e do raciocínio diagnóstico.

A busca bibliográfica dos artigos foi realizada em cinco bases de dados, *U.S. National Institutes of Health's National Library of Medicine* (NIH/NLM - PubMed), *Scientific Electronic Library Online* (Scielo), Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), *Medical Literature Analysis and Retrieval System Online* (MEDLINE), Biblioteca Virtual em Saúde (BVS) e Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS). Foram utilizados os Descritores em Ciências da Saúde (DeCS) em inglês e o operador booleano adequados para a pesquisa, incluindo: "*Artificial Intelligence*" AND "*Medicine*" AND "*Simulation Training*". Foi avaliado títulos e resumos por dois pesquisadores de forma independente e cega, com o objetivo de alcançar maior precisão nos estudos, como preconizam Donato e Donato (2019).

Os critérios de inclusão utilizados foram ensaios clínicos randomizados publicados nos últimos 5 anos e textos completos e gratuitos, e os de exclusão foram estudos duplicados e os que não atendem à questão da pesquisa. Após a etapa inicial de busca,

identificaram-se 532 publicações nas bases de dados, antes da aplicação dos critérios, sendo 128 artigos na Pubmed, 0 na Scielo, 377 no Capes, 23 na Medline, 32 BVS e 3 na LILACS. Após a realização da triagem dos artigos, utilizando os critérios de inclusão totalizaram 38 artigos. Foram excluídos 527 artigos por não atenderem aos critérios do estudo. Foi removido 1 estudo duplicado e 526 que não atendiam à questão da pesquisa ou aos critérios estabelecidos. Após a triagem efetuada, obteve uma amostra final de 5 estudos que serão considerados para a análise final.

Figura 1: Fluxograma de seleção dos estudos conforme Recomendação Prisma.



Fonte: Os autores, 2026.

Para examinar o rigor metodológico dos estudos selecionados, empregou-se a escala desenvolvida por meio da ferramenta RoB 2,

proposta pela Cochrane (Sterne *et al.*, 2019), a qual considera cinco domínios, como: processo de randomização, desvios da intervenção pretendida, dados ausentes, mensuração do desfecho e seleção do resultado relatado. A avaliação de cada domínio é feita através de perguntas sinalizadoras padronizadas que determinam o julgamento do risco de viés como “baixo risco”, “algumas preocupações” ou “alto risco”. A estratificação final do estudo é determinada a partir da síntese dos julgamentos em cada domínio realizada em ferramenta do Excel para implementação e desenvolvimento gráfico dos resultados.

Além disso, também foi utilizada a escala GRADE, , proposta pelo GRADE Working Group, avaliando a qualidade das evidências científicas e a força das recomendações em saúde. É composto pela classificação em quatro níveis: alta, moderada, baixa e muito baixa. Ademais, também analisa a força das recomendações, avaliadas como fortes ou fracas (Guyatt *et al.*, 2008).

RESULTADOS

No quadro 1, nota-se que 100% dos artigos estão escritos em inglês (n= 5,), com prevalência de 100% dos trabalhos publicados nos últimos 5 anos (n= 5), e 100% deles foi obtido na base de dados PubMed (n= 5), 40% dos artigos produzidos no Canadá (n=2). Além disso, a porção majoritária, 80%, dos periódicos tem foco na temática da educação médica (n= 4).

Quadro 1. Descrição dos artigos de acordo com autor/ano, periódico e título.

Autores/a no	Título	Base de dados	Periódico	Pais	Idioma
-----------------	--------	------------------	-----------	------	--------

Faziollahi et al. (2022)	Effect of Artificial Intelligence Tutoring vs Expert Instruction on Learning Simulated	PubMed	JAMA Network Open	Canadá	Inglês
--------------------------	--	--------	-------------------	--------	--------

⚠ Esta tabela possui muitas colunas e foi cortada para impressão. Para visualizá-la completa, acesse o artigo original em:

<https://revistatopicos.com.br/artigos/inteligencia-artificial-na-avaliacao-de-competencias-tecnicas-e-raciocinio-diagnostico-na-educacao-medica?noblockage>

Fonte: Dados de pesquisa, 2026.

No quadro 2 são denotadas algumas características encontradas nos estudos utilizados nesta revisão, como o estudo com maior número de participantes teve o número de amostra = 396 estudantes de medicina do 3º ano e o mais baixo com um número = 19 estudantes de medicina. Os grupos analisados são de médicos em formação ou alunos da graduação que se submeteram à realidade virtual e treinamentos convencionais, como por exemplo, *feedbacks* e simulações baseadas em IA, instruções de profissionais e OSCEs. O tempo de seguimento dos estudos variaram de 75 minutos até 3 dias. Parte equivalente a 40% dos estudos (n=2), não destacaram desfechos negativos de forma clara, já outros 40% (n=2) mostraram que o desfecho clínico do tratamento com a IA demonstrou-se com resultados mais negativos, e 20% da amostra demonstrou um pior resultado para os participantes com a intervenção humana (n=1).

Quadro 2. Características do tratamento, resultados dos ensaios clínicos encontrados.

Autores/a no	Tipo de estudo	População / grupos	Intervenç ão	Tempo	Desfec negath
Fazlollahi et al. (2022)	Ensaio clínico randomizado	70 estudantes de medicina, 23 estudantes no grupo VOA, 24	Virtual Operative Assistant (VOA): Feedback audiovisual baseado em	1 treinamento de 75 minutos, composto por 5 sessões de prática (5	Não fo encont as diferen signific as emoçõ negativ

△ Esta tabela possui muitas colunas e foi cortada para impressão. Para visualizá-la completa, acesse o artigo original em:

<https://revistatopicos.com.br/artigos/inteligencia-artificial-na-avaliacao-de-competencias-tecnicas-e-raciocinio-diagnostico-na-educacao-medica?noblackage>

Fonte: Dados de pesquisa, 2026.

Ao avaliar o quadro 3, percebe-se que em 40% dos estudos (n=2) a intervenção que envolvia IA foi igualitária ou superior à seus comparativos, sejam do grupo treinado por especialistas ou pelo grupo controle que não tinha nenhum *feedback* durante as simulações.

Além disso, fica o destaque a 40% dos textos (n=2) que a combinação do uso da IA com o método tradicional, com orientação de especialistas na área, foi o mais adequado, destacando que é possível cooperar com a tecnologia para complementar a compreensão na educação desses estudantes. Porém, apenas 20% dos trabalhos tiveram como melhor intervenção no quesito educacional o modelo de simulação tradicional com atores e avaliação de profissionais treinados, mas na perspectiva financeira o custo benefício a IA se mostrou mais vantajosa (n=1).

Quadro 3. Caracterização metodológica dos artigos selecionados.

Autores/A no	Grupo analisado	Intervenções aplicadas	Melhor Intervenção	Tempo
Fazlollahi <i>et al.</i> (2022)	70 estudantes de medicina, 23 estudantes no grupo VOA, 24 estudantes no grupo instructor e 23 estudantes no grupo controle.	As intervenções incluíram 5 sessões de feedback, com 5 minutos cada, durante um único treinamento de 75 minutos, incluindo 5 sessões práticas seguidas por 1 simulação realista de ressecção de tumor cerebral em realidade virtual. Os 3 grupos de intervenção incluíram 2 grupos de tratamento: feedback audiovisual baseado em métricas por IA (grupo VOA) e debriefing e instruções verbais	A Tutoria por IA (Grupo VOA). Ela demonstrou resultados de desempenho e transferência de habilidades superiores (avaliados pelo Expertise Score) comparada ao instrutor e ao controle, com avaliações globais equivalentes na escala OSATS e sem piora nas respostas emocionais e cognitivas.	1 treino de 75 min, composto por 5 sessões de prática (5 min de feedback em cada).

		<p>síncronas com roteiro, ministradas por um especialista remoto (grupo instrutor); e um grupo de controle que não recebeu feedback.</p>		
<p>Lee <i>et al.</i> (2025)</p>	<p>19 estudantes de medicina, um grupo de chatbot com IA (n = 9) ou um grupo de PRP (n = 10).</p>	<p>1. Grupo Intervenção: Prática com Chatbot de IA (feedback estruturado e repetição autônoma). 2. Grupo Controle: Peer Role-play (PRP - simulação entre colegas).</p>	<p>O estudo concluiu que ambas são complementares. O chatbot foi superior em autonomia e feedback estruturado, enquanto o role-play foi melhor para realismo e ambiente de prova. Não houve diferença estatisticamente significativa nos escores finais do OSCE.</p>	<p>1 hora de treino específico, precedido por 30 minutos de prática de exame físico comum a ambos os grupos</p>
<p>McCarrick <i>et al.</i> (2025)</p>	<p>90 estudantes em treinamento cirúrgico, 45 alocados</p>	<p>2 grupos paralelos: <i>Grupo de Intervenção:</i> Aprendizagem</p>	<p>O uso da IA associado à aprendizagem padrão. O grupo que interagiu</p>	<p>O grupo de intervenção o completou 3 simulações</p>

	<p>a cada grupo por meio de amostragem aleatória por conglomerados</p>	<p>experiencial padrão + 3 sessões estruturadas com um Modelo de Linguagem Profunda (ChatGPT / OpenAI) a atuar como "Paciente Simulado" (SP) para treino de história clínica. As transcrições foram depois submetidas para avaliação do tutor.</p> <p><i>Grupo de Controle:</i> Apenas aprendizagem experiencial padrão (ensino médico convencional).</p>	<p>com o ChatGPT como doente simulado melhorou muito mais as suas pontuações no Exame Clínico Objetivo Estruturado (OSCE) quando testado frente a um "paciente simulado humano", em comparação com o grupo de controle.</p>	<p>de casos clínicos baseadas em IA ao longo de 3 dias após o OSCE inicial.</p>
<p><u>Tyrrell et al.</u>(2025)</p>	<p>396 estudantes de medicina do 3º ano, randomizados em 2 grupos (A e B) usando um gerador de números</p>	<p>Comparação entre IA-CST (asíncrona, baseada em web) vs. AB-CST (presencial, com atores treinados e facilitadores).</p>	<p>Para eficácia educacional e satisfação, a simulação com atores (AB-CST) foi superior. Para custo-benefício e escalabilidade</p>	<p>Uma sessão de treinamento de 3 horas. O estudo foi um ensaio cruzado de curto prazo.</p>

	aleatórios computado rizado na proporção de 1:1.		e, a IA foi a melhor intervenção (custo de £33.48 vs £61.75 por aluno).	
Yilmaz <i>et al.</i> (2024)	99 médicos em treinamento ou formação, alocados aleatoriamente em três grupos: 1 - feedback de IA em tempo real (n= 33), 2 - instrução presencial por especialistas (n= 32) e 3 - sem feedback em tempo real (n= 32).	Foram divididos em 3 grupos: 1.Feedback de IA em Tempo Real : Sistema de tutoria inteligente (ICEMS) com métricas e orientações contínuas durante a cirurgia. 2.Instrução presencial orientada por um especialista humano. 3. Controle: Nenhum feedback fornecido ao longo da tarefa.	A Tutoria por IA em Tempo Real. O grupo da IA apresentou resultados de desempenho estatisticamente superiores tanto em comparação com o grupo de controle ($p < 0,001$) como em comparação com o ensino com instrutor humano ($p = 0,005$). A IA também ajudou a evitar o uso de força excessiva, e os resultados na avaliação qualitativa por especialistas humanos cegos (OSATS) foram tão	1 sessão de simulação de 90 min

			bons quanto a tutoria presencial (4,30 vs. 4,11).	
--	--	--	---	--

Fonte: Dados de pesquisa, 2026.

Para avaliar a qualidade dos artigos no estudo, utilizou-se a Escala do sistema GRADE (Grading of Recommendations Assessment, Development and Evaluation) para avaliar fatores relacionados à qualidade da evidência e a força das recomendações na saúde, que são inicialmente classificados como evidência alta (4 pontos) por serem ECRs, e reajusta as notas podendo adicionar pontos se presença de fatores que contribuem com o aumento da evidência o que é visto na tabela a seguir. Alternativamente, conforme os riscos de viés, imprecisão ou magnitude do efeito, pode deduzir 1 a 2 pontos. O resultado classifica a confiança no resultado em quatro níveis, de alta a muito baixa, para orientar recomendações clínicas.

O quadro 4 apresenta as pontuações dos estudos conforme critério de adesão. Observa-se que 40% obtiveram uma pontuação alta (n=2), indicando qualidade metodológica superior. Somente 20% dos trabalhos avaliados tiveram categorização de Baixo a Moderado sem pontuações adicionais (n=1).

Quadro 4. Avaliação da qualidade dos estudos selecionados nesta pesquisa conforme escala do sistema GRADE – fatores relacionados ao aumento da evidência

Autores/Autor	Grande Magnitude do Efeito	Grande dose-resposta	Confundidores ou vieses que reduziria o efeito encontrado	Nível de evidência de acordo com os critérios do sistema GRADE
Fazlollahi <i>et al.</i> (2022)	O grupo da IA (VOA) melhorou significativamente a pontuação prática em 0,66 pontos em relação ao instrutor e 0,65 em relação ao controle (em um algoritmo que vai de -1,00 a 1,00). O valor de P < ,001 comprova superioridade do modelo preditivo	Como foi uma intervenção única de 75 minutos, o desenho do estudo não permite avaliar gradientes de dose	O algoritmo de IA utilizado neste estudo não conseguiu detectar melhorias de desempenho no grupo de instrutores, de acordo com as avaliações OSATS para prática e cenários realistas	Moderado
Lee <i>et al.</i> (2025)	Não, os resultados entre os grupos foram similares ($p > 0.05$).	Não avaliado, o estudo utilizou um único treinamento .	O entusiasmo dos alunos com a tecnologia (efeito novidade) poderia inflar positivamente a satisfação com a IA, mas o design	Baixo a Moderado

			controlado tentou mitigar isso.	
McCarrick <i>et al.</i> (2025)	Sim, enquanto as avaliações básicas eram semelhantes, os alunos do grupo da IA apresentaram uma subida nas notas com grande relevância estatística ($p < 0,001$) e um tamanho de efeito consideravelmente superior ao controle (Cohen's D: 0,37 vs. 0,19).	Não é aplicável, tratou-se de um protocolo fechado de 3 sessões, sem variação de quantidade para avaliar um gradiente de resposta.	A avaliação final foi realizada por avaliadores humanos rigorosamente e cegados em relação ao grupo de origem do aluno, minimizando o favoritismo na atribuição da nota.	Alto
Tyrrell <i>et al.</i> (2025)	Não, a diferença entre os grupos foi pequena (0.36 pontos em uma escala de 10).	Não avaliado, o estudo utilizou um único treinamento	O design crossover ajudou a reduzir confundidores biossociais, mas o fato de ser um estudo "aberto" (estudantes sabiam qual era a intervenção) pode ter gerado um viés de	Moderado

			preferência pelo método tradicional (humano).	
Yilmaz <i>et al.</i> (2024)	A IA em tempo real demonstrou ser significativamente superior com um <i>P</i> forte na melhoria dos <i>scores</i> compostos de performance em relação aos outros dois braços do ensaio.	Não se aplica, pois, tratou-se de uma única intervenção padronizada de 90 minutos para todos os participantes.	O estudo reduziu fortemente os potenciais confundidores através da randomização e, sobretudo, ao ter as avaliações finais da escala clínica OSATS realizadas por peritos que desconheciam a que grupo o aluno pertencia.	Alto

Fonte: Dados de pesquisa, 2026.

No quadro 5, levou-se em consideração os fatores que reduzem o nível de evidência especialmente em ECRs, eles se dividem em 5 categorias e mantêm a classificação de muito baixo a alto. Com relação às limitações metodológicas, 80% dos artigos (n=4), informou sobre a impossibilidade de impedir que o aluno examinado soubesse que seu feedback era fornecido por IA ou por um instrutor humano.

Entre as inconsistências, um dos artigos declarou que o conhecimento prévio, ou a falta dessa informação, por parte do aluno participante poderia interferir com os resultados da pesquisa, o outro dos artigos afirmou que o sucesso do trabalho variou subjetivamente entre desempenho e satisfação do colaborador. Todavia, cerca de 60% dos trabalhos (n=3), não registaram inconsistências e justificaram-se por ter apenas um episódio de simulação. Apenas 40% dos artigos declararam presença da evidência indireta (n=2).

Com relação a imprecisão, 80% dos artigos (n=4) manifestaram-se como baixa, pois tinham tamanhos de amostra e delineamento de estudo adequados. Ao contrário de 20% (n=1) dos trabalhos que declararam uma imprecisão alta, devido ao baixo número de participantes, tem um intervalo de confiança amplo. Ademais, o grupo preponderante, 60% da amostra (n=3), tem baixo viés de publicação, mas 40% (n=2) não tiveram informações claras para fazer essa classificação.

Por fim, o padrão geral de qualidade, cerca de 80% (n=4), foi moderado com apenas 20% (n=1) categorizando como moderado a alto, com uma breve superioridade com relação aos outros trabalhos averiguados.

Quadro 5. Avaliação da qualidade dos estudos – fatores responsáveis pela diminuição do nível de evidência (GRADE)

Autores/a no	Limitaçõe s metodoló gicas	Inconsistê ncia	Evidência indireta	Imprecisã o	Viés publica o
-----------------	-------------------------------------	--------------------	-----------------------	----------------	----------------------

Faziollahi et al. (2022)	A plataforma não abrange todo o	Não avaliada isoladamente. Internamente.	Presente. Trata-se de um modelo de simulação	Baixa, os intervalos de confiança (IC 95%)	O risco baixo, o estudo foi devidamente
--------------------------	---------------------------------	--	--	--	---

⚠ Esta tabela possui muitas colunas e foi cortada para impressão. Para visualizá-la completa, acesse o artigo original em:

<https://revistatopicos.com.br/artigos/inteligencia-artificial-na-avaliacao-de-competencias-tecnicas-e-raciocinio-diagnostico-na-educacao-medica?noblockage>

Fonte: Dados de pesquisa, 2026.

O quadro 6 consiste na escala de RoB-2, que avalia o risco de viés para um desfecho específico do estudo. Nos fatores 1, 2, 3 e 5 100% dos artigos (n=5) foram considerados de baixo risco. Todavia, apenas 20% (n=1) da amostra apresentou mais de um fator com dúvidas ou com informações pouco claras, levando a uma classificação de alto risco.

Figura 2: Avaliação do risco de viés dos estudos selecionados nesta pesquisa conforme escala de RoB-2.

<u>Study ID</u>	<u>D1</u>	<u>D2</u>	<u>D3</u>	<u>D4</u>	<u>D5</u>	<u>Overall</u>	
Fazlollahi et al. (2022)	+	+	+	+	+	+	+
Lee et al. (2025)	+	+	+	+	+	+	!
McCarrick et al. (2025)	+	+	+	+	+	+	-
Tyrrell et al.(2025)	+	+	+	-	+	-	
Yilmaz et al. (2024)	+	+	+	+	+	+	
							D1 Randomisation process
							D2 Deviations from the intended interventions
							D3 Missing outcome data
							D4 Measurement of the outcome
							D5 Selection of the reported result

Fonte: Dados de pesquisa, 2026.

DISCUSSÃO

A integração da Inteligência Artificial (IA) e da Realidade Virtual (RV) na educação médica transformou a aquisição de competências técnicas, permitindo que sistemas automatizados igualem ou superem o desempenho de instrutores humanos em tarefas procedimentais de alta complexidade (Fazlollahi *et al.*, 2022). Essa superioridade manifesta-se principalmente na capacidade da tecnologia de fornecer uma avaliação quantitativa contínua e objetiva, algo que desafia os limites da percepção humana (Davidovic *et al.*, 2025; Yilmaz *et al.*, 2024).

Os estudos indicam que a integração de tecnologias de IA e RV reconfigura o ensino médico ao introduzir sistemas automatizados, como o *Virtual Operative Assistant* (VOA) e o ICEMS, cujos resultados demonstram ser iguais ou superiores aos métodos de instrução presencial por especialistas na aquisição de competências técnicas. A eficácia dessas ferramentas fundamenta-se na quantificação precisa de métricas objetivas monitoradas em tempo real, como a força aplicada aos tecidos e a economia de movimentos, o que mitiga a subjetividade humana e viabiliza a implementação do modelo de *mastery learning* (Mirchi *et al.*, 2020; Fazlollahi *et al.*, 2022).

Esse monitoramento contínuo permite que o estudante pratique de forma autônoma até atingir níveis de proficiência de especialistas, reduzindo significativamente a incidência de erros críticos. Adicionalmente, enquanto evidências de metanálises sugerem que a RV apresenta desempenho superior no desenvolvimento de habilidades práticas, a IA destaca-se no aprimoramento do conhecimento teórico e na estruturação do raciocínio clínico em

ambientes reflexivos (McCarrick *et al.*, 2025). Sob a perspectiva institucional, tais ferramentas oferecem uma vantagem econômica relevante, com custos de implementação estimados em aproximadamente metade do valor das simulações tradicionais baseadas em atores (Tyrrell *et al.*, 2025).

No que diz respeito à precisão e ao monitoramento de métricas, os métodos tradicionais mostram-se limitados, sobretudo quando confrontados com a capacidade dos sistemas modernos de promover mensuração objetiva e contínua de variáveis imperceptíveis ao olho humano, especialmente em contextos dinâmicos e de grande escala (Rodrigues, 2025). Enquanto a avaliação por instrutores dependem de observações subjetivas e qualitativas, sistemas como o ICEMS (*Intelligent Continuous Expertise Monitoring System*) avaliam o desempenho em frações de segundo, permitindo a quantificação precisa de parâmetros como distância tridimensional entre instrumentos e até mesmo estimativas de perda sanguínea, aspectos notoriamente difícil de aferir em ambientes tradicionais, mas que, em simuladores com IA, pode ser monitorado em tempo real, reduzindo a incidência de erros críticos e promovendo maior segurança no aprendizado (Fazlollahi *et al.*, 2022).

No contexto da evolução dos modelos de ensino e avaliação em ambientes técnicos complexos, destaca-se uma mudança paradigmática na forma como o desempenho é acompanhado e mensurado. Tradicionalmente, esse processo esteve atrelado à atuação direta de tutores humanos, cuja capacidade avaliativa, embora qualificada, encontra limitações inerentes à condição humana, como a fadiga, a oscilação de critérios e as restrições da atenção sustentada ao longo do tempo (Farias, 2018). É nesse

cenário que emergem os sistemas inteligentes, introduzindo um novo padrão de consistência e estabilidade na avaliação do desempenho.

Tal característica reduz de forma significativa a incidência de erros, como lesões teciduais em ambientes simulados, e favorece a implementação do modelo de *mastery learning*, no qual o estudante pode praticar de maneira autônoma, repetitiva e orientada por métricas objetivas, progredindo apenas após atingir níveis de proficiência compatíveis com o padrão de especialistas, antes da transição para a prática clínica real (Mirchi *et al.*, 2020; Yilmaz *et al.*, 2024).

Já no domínio das competências interpessoais e do raciocínio clínico, a aplicação de Modelos de Linguagem Profunda (DLMs), como o ChatGPT, tem demonstrado eficácia significativa tanto na adequação da linguagem utilizada em registros de anamnese quanto na preparação de estudantes para avaliações estruturadas, como o OSCE (Hutchison; De Oliveira, 2025). Estudos indicam que o uso dessas ferramentas como pacientes simulados, por exemplo, está associado a melhorias estatisticamente significativas no desempenho acadêmico, com elevação da média nas avaliações, resultado que se explica pela capacidade da inteligência artificial de atuar como um ambiente de experimentação cognitiva, no qual o aluno pode exercitar o raciocínio clínico de forma reflexiva e deliberada, sem a pressão do julgamento humano (Lee *et al.*, 2025; McCarrick *et al.*, 2025).

Contudo, a literatura pondera que, embora a IA seja eficiente na estruturação lógica da consulta, ela ainda enfrenta desafios severos para replicar a autenticidade da interação humana e o

desenvolvimento de habilidades humanísticas (Cavalheiro Neto; Dal Mollin, 2024). Quando comparada à dramatização entre pares (PRP) ou simulações com atores reais, esses instrumentos são vistos como limitados no ensino de "habilidades corporificadas", que envolvem a gestão de situações inesperadas, a comunicação não verbal e a resposta imediata a reações imprevisíveis de um paciente real. Ademais, simuladores virtuais baseados em voz podem falhar em captar pistas sutis de empatia, limitando-se a reconhecer gestos básicos, o que pode comprometer o desenvolvimento da sensibilidade interpessoal do futuro médico (Lee *et al.*, 2025; Tyrrell *et al.*, 2025).

Diante desse panorama, consolida-se a proposta de um modelo híbrido de instrução, no qual a atuação humana é potencializada pelo uso de sistemas inteligentes (De Jesus *et al.*, 2024). Nesse arranjo, o tutor passa a utilizar dados pormenorizados de falhas técnicas, gerados por ferramentas como o *Virtual Operative Assistant* (VOA) e o *Machine Learning to Assess Surgical Expertise* (MLASE), que empregam algoritmos para identificar erros específicos que escapam à percepção subjetiva humana, como a aplicação excessiva de força ou imprecisões na manipulação tecidual. Essa integração permite que o instrutor direcione sua atuação de forma mais estratégica, incorporando o feedback automatizado ao padrão-ouro da *Objective Structured Assessment of Technical Skills* (OSATS), o que aprimora a mentoria qualitativa e otimiza a curva de aprendizado, reduzindo a frequência de correções necessárias ao longo do treinamento (Mirchi *et al.*, 2020; Yilmaz *et al.*, 2024; Davidovic *et al.*; 2025).

Todavia, a busca por maior precisão e adaptatividade exige cautela quanto à carga cognitiva extrínseca associada ao uso de sistemas de

IA em tempo real, a qual pode superar aquela observada na instrução humana presencial isolada. Para mitigar tais riscos pedagógicos, a literatura recomenda o monitoramento do estado afetivo do aprendiz por meio da *Medical Emotions Scale* (MES), instrumento que possibilita mensurar a intensidade das emoções durante a simulação (Fazlollahi *et al.*, 2022). Essa abordagem visa assegurar que o feedback tecnológico não resulte em frustração ou sobrecarga mental, preservando, assim, a efetividade do modelo de aprendizagem por domínio (*mastery learning*).

Em síntese, a convergência entre a precisão analítica da tecnologia e o julgamento clínico do instrutor humano configura um ambiente de aprendizagem mais seguro, eficiente e orientado por evidências, em consonância com os princípios das metodologias ativas, que privilegiam a prática deliberada, o feedback contínuo e o protagonismo do estudante na construção do conhecimento (Garbin, 2025).

Apesar do potencial transformador, destacam-se limitações críticas que ainda precisam ser superadas, como as instabilidades técnicas, evidenciadas por falhas recorrentes no reconhecimento de voz, e o risco persistente associado aos modelos de linguagem profunda (DLMs), que podem gerar respostas incorretas, imprecisas ou até mesmo inteiramente fabricadas, o que torna a supervisão humana indispensável para garantir a precisão e a segurança do conteúdo clínico apresentado (McCarrick *et al.*; Tyrrell *et al.*, 2025). Para mitigar a percepção da IA como uma "caixa preta" e reduzir a incerteza ou frustração dos alunos, emerge uma demanda por Inteligência Artificial Explicável (XAI), exemplificada por sistemas como o VOA, que promovem a transparência ao permitir que tanto aprendizes

quanto educadores compreendam detalhadamente os critérios e pesos das métricas de avaliação utilizados (Mirchi *et al*, 2020).

As evidências apontam para um modelo de aprendizagem mista e gradual como o padrão pedagógico ideal, integrando estrategicamente os pontos fortes da tecnologia e do ensino tradicional (Wang *et al.*, 2026). Nesse paradigma, os estudantes utilizam sistemas de IA e RV em estágios iniciais para consolidar o raciocínio clínico e as competências técnicas básicas em um "ambiente de experimentação cognitiva" de baixo risco (Lee *et al.*, 2025). A progressão posterior para simulações e interações humanas de alta fidelidade permite que o foco educacional se desloque para o desenvolvimento de competências humanísticas complexas, como a empatia, o profissionalismo e a gestão de interações imprevisíveis, domínios onde a autenticidade humana permanece, por ora, insubstituível (Zhou *et al.*, 2013; Eklics *et al.*, 2024).

Todavia, a robustez desses resultados deve ser analisada com cautela, visto que parte significativa dos estudos apresenta limitações metodológicas importantes. Entre os desafios, destacam-se as amostras reduzidas, a curta duração do acompanhamento e a heterogeneidade entre os protocolos de intervenção e avaliação. Ademais, muitos achados derivam de ambientes simulados de realidade virtual, o que restringe a extrapolação direta para a prática clínica real com pacientes. Soma-se a isso o risco de vieses, incluindo a ausência de cegamento em algumas intervenções tecnológicas, a influência da familiaridade prévia dos participantes com ferramentas digitais e a dependência de métricas proprietárias de cada sistema.

CONCLUSÃO

Este estudo possibilitou avaliar as aplicações de inteligência artificial na avaliação de competências técnicas e raciocínio diagnóstico na educação médica. Dessa forma, notou-se que a inteligência artificial (IA) pode ser utilizada como possibilidade na contribuição para as avaliações de competências técnicas da educação médica, já que apresentam um potencial de aprimoramento nas lacunas presentes nos métodos tradicionais, como a avaliação rápida e objetiva dos parâmetros analisados. É digno de nota que todos os estudos obtiveram a classificação máxima na escala de Jadad, indicativo da excelência metodológica dos estudos selecionados.

Todavia, embora o uso da IA seja eficaz, ainda apresenta desafios para reproduzir o relacionamento humano e as habilidades humanísticas. Além disso, também possuem a necessidade de alto investimento financeiro e tempo disponibilizado para realização de treinamento da equipe, visto que, necessitam de supervisão adequada através de ajustes técnicos, pedagógicos, curriculares e éticos para seu uso.

Portanto, notou-se que apesar do seu potencial inovador, ainda existem limitações importantes a serem superadas, como falhas técnicas e riscos relacionados aos modelos de linguagem, tornando essencial a supervisão humana ou a associação de técnicas tradicionais com as técnicas fornecidas pela IA. Em síntese, as evidências indicam que um modelo de ensino híbrido e progressivo, que combine recursos tecnológicos e métodos tradicionais, é o mais adequado.

REFERÊNCIAS BIBLIOGRÁFICAS

CAVALHEIRO NETO, Afonso; MOLIN, Beatriz Helena Dal. Inteligência artificial e as novas experiências de aprendizagem. **In:** BARRETO, Alfredo Rafael Belinato (org.). **Educação e tecnologia: o novo paradigma do ensino.** Cascavel: Editora FAG, 2024. p. 189-205. Disponível em: <https://www.fag.edu.br/upload/revista/editora/681275a58a9a9.pdf#page=189>. Acesso em: 16 abr. 2026.

CRUESS, Sylvia R. Professionalism and medicine's social contract with society. **Clinical Orthopaedics and Related Research**, v. 449, p. 170-6, 2006. Disponível em: <https://doi.org/10.1097/01.blo.0000229275.66570.97>. Acesso 13 abr. 2026.

DAVIDOVIC, Vanja *et al.* Effect of Artificial Intelligence-Augmented Human Instruction on Feedback Frequency and Surgical Performance During Simulation Training. **Journal of Surgical Education**, v. 82, n. 11, p. 103743, 2025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1931720425003241>. Acesso em: 15 abr. 2026.

DE JESUS, Everaldo Antônio et al. A convergência da inteligência artificial e da educação: Perspectivas para o aprendizado do século XXI. **Revista Acadêmica Online**, v. 10, n. 51, p. 1-11, 2024. Disponível em: <https://revistaacademicaonline.com/index.php/rao/article/view/119>. Acesso em: 13 abr. 2026.

DONATO, Helena; DONATO, Mariana. Etapas na condução de uma revisão sistemática. **Acta médica portuguesa**, v. 32, n. 3, pág. 227-235, 2019. Disponível em:

<https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/11923>. Acesso 13 abr. 2026.

EKLICS, Kata *et al.* Diverse integration of simulated patients in medical education for communication, language, and clinical skills in Hungary. **Advances in Medical Education and Practice**, p. 301-312, 2024. Disponível em: <https://www.tandfonline.com/doi/full/10.2147/AMEP.S438102>. Acesso em: 15 abr. 2026.

ELENDU, Chukwuka *et al.* The impact of simulation-based training in medical education: A review. **Wolters Kluwer**, v. 103, n. 27, 2024. Disponível em: <https://www.periodicos.capes.gov.br/index.php/acervo/buscaador.html?task=detalhes&source=all&id=W4400348041>. Acesso 13 abr. 2026.

FARIAS, Mércia Inara Rodrigues de. **Métodos avaliativos no ensino de ciências:** uma análise da percepção discente sobre a avaliação. 2018. Trabalho de Conclusão de Curso (Graduação em Ciências Biológicas) – Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, João Pessoa, 2018. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/3519>. Acesso em: 13 abr. 2026.

FAZLOLLAHI, Ali M. *et al.* Effect of AI tutoring versus expert instruction on the learning of simulated surgical skills among medical students. A randomized clinical trial. **JAMA Network Open**, v. 5, n. 2, p. e2149008, 2022. Disponível em: <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2789268>. Acesso em: 10 abr. 2026.

FRAGA, Andrea de Melo Alexandre; BELLUOMINI, Fernando. Uso da Simulação no Ensino Médico: Uma Ferramenta Contemporânea para o Desenvolvimento de Habilidades Técnicas e Não Técnicas. **Unicamp**, v.13, n.4, 2025. Disponível em: <https://portal.fcm.unicamp.br/artigo/uso-da-simulacao-no-ensino-medico-uma-ferramenta-contemporanea-para-o-desenvolvimento-de-habilidades-tecnicas-e-nao-tecnicas/>. Acesso 14 abr. 2026.

GALVÃO, Taís Freire; TIGUMAN, Gustavo Magno Baldin; SARKIS-ONOFRE, Rafael. A declaração PRISMA 2020 em português: recomendações atualizadas para o relatório de revisões sistemáticas. **Epidemiologia e serviços de saúde**, v. 31, p. e2022364, 2022. Disponível em: <https://www.scielo.org/article/ress/2022.v31n2/e2022364/>. Acesso 12 abr. 2026.

GARBIN, Mônica Cristina. **Metodologias ativas de aprendizagem: fundamentos e práticas**. São Paulo: Editora Senac São Paulo, 2025.

GUYATT, Gordon H. *et al.* **GRADE**: an emerging consensus on rating quality of evidence and strength of recommendations. *Bmj*, v. 336, n. 7650, p. 924-926, 2008.

HUTCHISON, Marcos Paulo Carneiro Vieira; DE OLIVEIRA, Naila Albertina. Integração da Inteligência Artificial na educação médica: desenvolvimento de um modelo baseado em GPT para o ensino de anamnese e documentação de prontuários médicos. **Caderno Pedagógico**, v. 22, n. 9, p. e18414-e18414, 2025. Disponível em: <https://ojs.studiespublicacoes.com.br/ojs/index.php/cadped/article/view/18414>. Acesso 12 abr. 2026.

LEE, Hye-Yoon *et al.* Comparison between AI chatbot simulation and peer role-playing for OSCE preparation: a randomized controlled pilot study. **BMC Medical Education**, 2025. Disponível em: <https://link.springer.com/article/10.1186/s12909-025-08308-y>. Acesso em: 10 abr. 2026.

LEVY, Marcia de Almeida; TAQUETTE, Stella Regina; SOUZA, Luciana Maria Borges da Matta. Aprendizado do componente atitudinal na graduação em Medicina: uma revisão de literatura. **Revista brasileira de educação médica**, v. 50, p. e033, 2026. Disponível em: <https://www.scielo.br/j/rbem/a/V4LjtsWDBnMR7BfpQ8GjvBR/?format=html&lang=pt>. Acesso 13 abr. 2026.

LINARES-ESPINÓS, Estefanía *et al.* Methodology of a systematic review. **Actas Urológicas Españolas (English Edition)**, v. 42, n. 8, p. 499-506, 2018. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S2173578618301070>. Acesso 13 abr. 2026.

MCCARRICK, Cathleen A. *et al.* Um ensaio clínico randomizado controlado de uma ferramenta de simulação baseada em modelo de aprendizado profundo de linguagem para estudantes de medicina de graduação em cirurgia A Randomized Controlled Clinical Trial of a Natural Language Deep Learning Model-Based Simulation Tool for Undergraduate Medical Students in Surgery. **Journal of Surgical Education**, v. 82, n. 9, p. 103629, 2025. Disponível: <https://www.sciencedirect.com/science/article/pii/S1931720425002107>. Acesso em: 10 abr. 2026.

MENDES, Lorrane Alves da Silva *et al.* O ensino sobre dor além do conhecimento técnico: uma reflexão sobre soft skills para

profissionais de saúde. **Brazilian Journal of Pain**, v. 6, n. 2, p. 105-106, abr./jun. 2023. DOI: 10.5935/2595-0118.20230049-pt. Disponível em: <https://www.scielo.br/j/brjp/a/xYy4MHncPWgXWK8TgYsshCK/?lang=pt>. Acesso em: 16 abr. 2026.

MIRCHI, Nykan *et al.* The virtual operative assistant: an explainable artificial intelligence tool for simulation-based training in surgery and medicine. **PloS one**, v. 15, n. 2, p. e0229596, 2020. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229596>. Acesso em: 16 abr. 2026.

NEY, Márcia Silveira; MARTINS, Mirian Teresa de Sá Leitão; FIRMIDA, Mônica de Cássia. Reflexões sobre as mudanças na formação médica no Brasil. **Revista Brasileira de Educação Médica**, v. 50, n. 1, e017, 2026. DOI: 10.1590/1981-5271v50.1-2024-0229. Disponível em: <https://www.scielo.br/j/rbem/a/ShQJpR4vyc6GLDgjJYjPL9Q/>. Acesso em: 16 abr. 2026.

RODRIGUES, Paulo Ricardo Fernandes. **Mapeamento sistemático sobre técnicas de monitoramento de faltas baseadas em mineração de dados**. 2025. Monografia (Graduação em Ciência da Computação) - Centro de Ciências, Universidade Federal do Ceará, Fortaleza, 2025. Disponível em: <https://repositorio.ufc.br/handle/riufc/80322>. Acesso em: 15 abr. 2026.

STERNE, Jonathan AC *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. **bmj**, v. 366, 2019. Disponível em: <https://www.bmj.com/content/366/bmj.l4898>. Acesso em: 03 mai. 2026.

TYRRELL, Edward G. *et al.* Web-Based AI-Driven Virtual Patient Simulator Versus Actor-Based Simulation for Teaching Consultation

Skills: Multicenter Randomized Crossover Study. **JMIR Formative Research**, v. 9, p. e71667, 2025. Disponível em: <https://formative.jmir.org/2025/1/e71667>. Acesso em: 10 abr. 2026.

WEIDENER, Lukas; FISCHER, Michael. Proposing a Principle-Based Approach for Teaching AI Ethics in Medical Education. **JMIR Medical Education**, v. 10, n. 1, e55368, 9 fev. 2024. DOI: 10.2196/55368. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10891487/>. Acesso em: 16 abr. 2026.

YILMAZ, Recai *et al.* Real-Time multifaceted artificial intelligence vs In-Person instruction in teaching surgical technical skills: a randomized controlled trial. **Scientific reports**, v. 14, n. 1, p. 15130, 2024. Disponível em: <https://www.nature.com/articles/s41598-024-65716-8>. Acesso em: 10 abr. 2026.

ZHOU, Yuefang *et al.* How do medical students respond to emotional cues and concerns expressed by simulated patients during OSCE consultations?—a multilevel study. **PLoS One**, v. 8, n. 10, p. e79166, 2013. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0079166>. Acesso em: 16 abr. 2026.

¹ Centro Universitário de Patos - UNIFIP. E-mail: [acesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0009-0007-1196-201X>

² Centro Universitário de Patos - UNIFIP. E-mail: [macesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0000-0002-7003-6282>

³ Centro Universitário de Patos - UNIFIP. E-mail: [acesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0009-0007-0959-2937>

⁴ Centro Universitário de Patos - UNIFIP. E-mail: [acesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0000-0002-8279-7257>

⁵ Centro Universitário de Patos - UNIFIP. E-mail: [acesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0000-0002-9134-9771>

⁶ Administradora (2007), Enfermeira (2008), Licenciada em Educação Física (2021) Mestre em Ciências da Saúde (2009), Doutora em Promoção de Saúde (2014), Pós-doutora em Promoção de Saúde (2016) e Pós-doutora em Sistemas Agroindustriais pela Universidade Federal de Campina Grande (2018). Atualmente é professora em tempo integral no Centro Universitário de Patos (UNIFIP), Pró-Reitora de Pesquisa, Extensão e Pós-Graduação do UNIFIP, Coordenadora do Eixo Práticas Investigativas em Saúde e do Trabalho de Conclusão de Curso do Curso de Medicina do UNIFIP. E-mail: [acesse o artigo original para visualizar o e-mail](#). ORCID: <https://orcid.org/0000-0001-8327-9147>