ECRONICON OPEN ACCESS

EC NEUROLOGY
Research Article

Artificial Intelligence in Medical Education: A Narrative Review with Emphasis on Empirical Evidence and Study Quality

Daniel Finney Sankuru[1], Chinnababu Sunkavalli[2], Dommeti Sri Lakshmi[3], Guduri Ganga Bhavani[4] and Raghavendrarao MV[5]*

[1]Senior Resident, Department of Community Medicine, Government Medical College and Hospital, Eluru, Andhra Pradesh, India

[2]Senior Consultant Surgical and Robotic Oncologist, Clinical Director - Surgical Oncology, Yashoda Hospital, Hitech City, Hyderabad, Telangana, India

[3]Assistant Professor, Department of Community Medicine, Government Medical College and Hospital, Eluru, Andhra Pradesh, India

[4]Professor and HOD, Department of Community Medicine, Government Medical College and Hospital, Eluru, Andhra Pradesh, India

[5]Professor, Microbiology, World Academy of Medical Sciences, Netherlands

***Corresponding Author:** Raghavendrarao MV, Professor, Microbiology, World Academy of Medical Sciences, Netherlands.

Received: July 16, 2026; **Published:** July 31, 2026

Abstract

Introduction: Artificial intelligence (AI) is being integrated into medical education at an accelerating pace, yet the empirical evidence supporting its effectiveness remains heterogeneous and insufficiently synthesized. Much of the strongest quantitative evidence currently available concerns simulation-based education broadly rather than AI-specific applications, and this distinction is maintained throughout the review.

Objective: To critically synthesize and compare quantitative evidence on different AI applications in medical education, with emphasis on effect sizes, study quality, and the specificity of evidence to AI (as opposed to simulation technology in general).

Methods: Narrative review of peer-reviewed literature from PubMed and Scopus (2019-2026), prioritizing systematic reviews and meta-analyses that reported effect sizes or objective outcomes, supplemented by recent primary studies where AI-specific evidence was otherwise unavailable. This is a narrative, not a systematic, review: searches were not exhaustive, dual-reviewer screening was not performed, and no formal quality-appraisal instrument (e.g. MERSQI) was applied.

Results: Foundational meta-analyses of technology-enhanced simulation (not AI-specific) show large effects on knowledge and skills versus no intervention (Hedges' $g \approx 1.1$-1.2) and small-to-moderate effects versus traditional instruction ($g \approx 0.30$-0.66). Virtual patient meta-analyses show low-to-moderate positive effects on clinical reasoning and skills; a 2025 crossover study specifically comparing an AI-enhanced virtual patient platform to a conventional one found higher learner ratings across several design domains. Evidence for AI-driven adaptive learning and automated assessment tools is more limited, derived mainly from single-institution studies with non-standardized outcome measures (e.g. Cohen's $d = 0.78$ for study engagement in one adaptive-learning trial; up to 87.5% grading-time reduction in one automated-assessment study). Significant gaps remain in long-term clinical outcomes and

Introduction

Medical education continues to face persistent structural challenges, including variability in clinical exposure, limited faculty time, and difficulties in delivering truly individualized and scalable training. These issues have been intensified by growing clinical complexity and the shift toward competency-based accreditation systems, which explicitly require more granular, longitudinal assessment of trainees than earlier time-based models [1].

Artificial intelligence has emerged as a potential solution, with tools now being developed and deployed for adaptive learning, virtual simulated patients, automated assessment and feedback, and data-driven curriculum design. A 2024 scoping review mapping this literature identified a rapidly expanding but highly heterogeneous evidence base, spanning dozens of distinct AI applications with widely varying levels of methodological rigor [2]. Proponents argue that AI can overcome traditional limitations by providing personalized, scalable, and efficient educational experiences.

However, the rapid adoption of AI in medical education has outpaced rigorous, independent evaluation: the same 2024 scoping review found that the great majority of available studies were single-institution, outcomes-light feasibility or perception studies rather than comparative effectiveness research [2]. Many institutions are implementing these technologies based on small pilots or theoretical benefits, often without strong evidence of educational effectiveness or long-term impact. A further complication, addressed directly in this review, is that much of the quantitative evidence cited in favor of AI-enhanced education is in fact evidence about simulation-based education in general, generated before AI-specific tools existed; the two bodies of evidence are frequently conflated in the literature. Far fewer studies have specifically examined AI-enhanced applications or compared the strength of evidence across different AI use cases (such as virtual patients versus adaptive learning versus automated assessment).

This evidence gap is particularly important because medical education directly influences future patient care. Without a clear synthesis of where AI delivers meaningful benefits, where the supporting data remain weak, and where "AI evidence" is really repurposed simulation evidence, there is a risk of inefficient implementation or unintended consequences. This narrative review was conducted to address that need.

Methods

This is a narrative review, not a systematic review or meta-analysis. Searches were conducted in PubMed and Scopus for English-language peer-reviewed literature published between January 2019 and June 2026, supplemented by hand-searching reference lists and, where AI-specific evidence was sparse, by including recent (2024-2025) primary studies. Priority was given to systematic reviews, meta-analyses, and studies reporting effect sizes (e.g. Hedges' g, Cohen's d) or objective performance outcomes. Screening was performed by a single reviewer; no dual-reviewer verification, formal risk-of-bias tool, or PRISMA-style flow diagram was used, consistent with the narrative rather than systematic design. Thematic synthesis focused on virtual patients and simulation, adaptive learning, assessment tools, and methodological limitations, with explicit attention to whether cited evidence concerned AI specifically or simulation/education technology more broadly.

Results

Virtual patients and simulation-based training

Simulation-based education, including AI-enhanced virtual patient platforms, currently sits atop the quantitative evidence base for technology in medical education - although most of this evidence predates AI and concerns simulation technology generally rather than AI specifically. The foundational meta-analysis of technology-enhanced simulation reported a pooled Hedges' g of 1.20 for knowledge outcomes compared with no intervention (a large effect), with similarly large effects for skills and behavior [3]. A companion meta-analysis restricting comparisons to simulation versus non-simulation instruction (rather than no intervention) found smaller effects: g = 0.30 for knowledge and g = 0.30 - 0.66 across skill measures, in comparison with other instructional methods [4]. These two studies are often conflated in secondary literature; they are distinguished here because the comparator (no intervention vs. other instruction) changes the size of the reported effect substantially.

A separate meta-analysis specific to virtual patient simulations found low-to-moderate positive effects on clinical reasoning and skill outcomes compared with traditional education, with more consistent benefit for skills than for knowledge [5]. Evidence specifically incorporating AI into virtual patient design is newer and more limited: a 2025 observational crossover cohort study of 178 medical students found that an AI-enhanced, socially embodied virtual patient platform was rated more favorably than a conventional computer-based platform on authenticity, professionalism, coaching, and perceived learning effect [6]. This is a genuinely AI-specific finding, but it reflects learner-perceived design quality in a single institution rather than a standardized learning-outcome effect size, and it has not yet been replicated.

Adaptive learning and personalized pathways

Evidence for AI-driven adaptive learning platforms is still emerging and remains preliminary. A 2025 single-institution study of an AI-driven personalized learning platform reported a moderate effect on students' self-directed study time (Cohen's d = 0.78) attributable to dynamic pacing of content [7]. This study is illustrative of the current state of the field: available evidence is drawn mostly from single institutions, frequently lacks control groups, and rarely reports outcomes in a form (e.g. standardized effect sizes on validated knowledge or skills measures) that permits direct comparison with the simulation literature.

Assessment, feedback, and faculty efficiency

AI tools for automated assessment and feedback generation have shown clear efficiency benefits in institutional settings. A 2025 analysis of an AI-based Automated Systems Protocol (using GPT-4) for grading medical students' post-encounter notes across four academic years (n = 684 students) found grading-time reductions of up to 87.5% and a return on investment as high as ~800% relative to legacy human-rater models, once initial development costs were amortized [8]. However, this and comparable studies have evaluated operational efficiency rather than whether faster or cheaper grading translates into improved educational outcomes for learners, a question that remains largely unaddressed.

Summary of key quantitative findings

Application	Key Quantitative Findings	Strength of Evidence	References
Simulation-Based Training (general)	Hedges' g = 1.20 vs. no intervention (large); g = 0.30 vs. traditional instruction (small-moderate), for knowledge	Strong (multiple large meta-analyses)	[3,4]
Virtual Patients	Low-to-moderate positive effects on clinical reasoning and skills vs. traditional education	Moderate	[5]
AI-Enhanced Virtual Patients	Higher learner ratings of authenticity, coaching, and learning effect vs. conventional VP platform (single crossover study)	Emerging	[6]
AI-Driven Adaptive/Personalized Learning	Cohen's d = 0.78 for self-directed study time (single institution)	Preliminary	[7]
AI Assessment and Feedback Tools	Up to 87.5% reduction in grading time; ROI up to ~800% vs. legacy human grading	Moderate (institutional data; process outcomes only)	[8]

Table 1: *Summary of quantitative findings across AI and AI-adjacent applications in medical education. Note that the simulation-based training row reflects general simulation evidence rather than AI-specific evidence; rows 3-5 reflect AI-specific studies.*

Discussion

This review highlights important differences in the strength of evidence across AI and AI-adjacent applications in medical education, and, just as importantly, differences in how directly that evidence bears on AI itself. The most robust quantitative data - large pooled effect sizes from meta-analyses of technology-enhanced simulation - concern simulation technology broadly and were generated well before AI-specific tools were introduced [3,4]. These foundational effects have long supported the value of deliberate, structured practice in simulation environments [9], and they provide a plausible ceiling for what AI-augmented simulation might achieve, but they are not themselves evidence that AI adds value beyond conventional simulation.

Evidence that is genuinely specific to AI is more limited but growing. Virtual patient research shows a credible trajectory from general virtual patient effectiveness [5] toward AI-specific comparisons, such as the 2025 finding that an AI-enhanced social robotic virtual patient outperformed a conventional platform on several design and engagement domains [6]. This is encouraging but preliminary: a single crossover study at one institution, measuring perceived design quality rather than a standardized learning outcome, cannot yet be generalized.

AI-driven adaptive learning platforms occupy the weakest evidentiary position of the applications considered here. The single available effect-size estimate for a personalized learning platform, while moderate in magnitude ($d = 0.78$), comes from one institution, uses study-time as a proxy outcome rather than a validated knowledge or skills measure, and lacks replication [7]. This methodological gap makes it difficult to assess whether adaptive learning systems deliver benefits of similar magnitude or reliability to simulation-based approaches.

AI tools for assessment and feedback show clear advantages in operational efficiency, with documented reductions in faculty grading time and favorable return on investment [8]. However, unlike simulation research, these studies have focused on process improvement rather than direct educational effectiveness. This represents a notable and recurring difference in research focus across application areas: efficiency is measured far more often, and far more rigorously, than learning gain.

Another consistent limitation across the literature is the limited attention to equity and algorithmic bias. Professional bodies have begun to issue guidance on responsible AI use in medical education precisely because empirical study of differential effects on learner subgroups remains scarce [10], in contrast to more mature simulation research that has more thoroughly examined how instructional design features influence outcomes.

Overall, comparative analysis suggests that the strongest and most consistent evidence remains within simulation environments generally, while AI-specific evidence is real but still early-stage, most developed for virtual patients, and least developed for adaptive learning. Readers should be cautious of literature (including earlier drafts of this review) that cites general simulation effect sizes as though they were AI-specific findings.

Conclusion

AI applications in medical education show measurable promise, but the strongest quantitative support currently available is for simulation-based training in general rather than AI specifically. Evidence that directly implicates AI is strongest for virtual patient applications and weakest for adaptive learning and automated assessment tools, which remain preliminary, single-institution, and efficiency-focused. Significant gaps remain in long-term outcome data, AI-specific comparative studies, and equity research. Stronger methodological designs, standardized effect-size reporting, and greater attention to equity and clinical transfer are needed to guide responsible implementation.

Conflicts of Interest

None declared.

Funding Support

None.

Author Contributions

Conceptualization, analysis, and manuscript preparation by the author.

Bibliography

1. Nasca TJ., *et al*. "The next GME accreditation system - rationale and benefits". *New England Journal of Medicine* 366.11 (2012): 1051-1056.
2. Gordon M., *et al*. "A scoping review of artificial intelligence in medical education: BEME Guide No. 84". *Medical Teacher* 46.4 (2024): 446-470.
3. Cook DA., *et al*. "Technology-enhanced simulation for health professions education: a systematic review and meta-analysis". *Journal of the American Medical Association* 306.9 (2011): 978-988.
4. Cook DA., *et al*. "Comparative effectiveness of technology-enhanced simulation versus other instructional methods: a systematic review and meta-analysis". *Simulation in Healthcare* 7.5 (2012): 308-320.
5. Kononowicz AA., *et al*. "Virtual patient simulations in health professions education: systematic review and meta-analysis by the Digital Health Education Collaboration". *Journal of Medical Internet Research* 21.7 (2019): e14676.
6. Borg A., *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* 27 (2025): e82541.
7. Y Chen. "Evaluation of the impact of an AI-driven personalized learning platform on medical students' learning performance". *Frontiers in Medicine (Lausanne)* 12 (2025): 1610012.
8. Campbell KK., *et al*. "Applying state-of-the-art artificial intelligence to grading in simulation-based education: assessment, feedback, and ROI". *Discover Artificial Intelligence* 5 (2025): 202.
9. McGaghie WC., *et al*. "Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence". *Academic Medicine* 86.6 (2011): 706-711.
10. Association of American Medical Colleges. "Principles for the responsible use of artificial intelligence in and for medical education". Washington, DC: AAMC (2025).

Volume 18 Issue 7 July 2026