EduVision: Journal of Innovations in Pedagogy and Educational Advancements
Volume 2, Issue 6, June 2026
brightmindpublishing.com
ISSN (E): 3061-6972

ARTIFICIAL INTELLIGENCE IN MEDICAL TRAINING: A FRAMEWORK FOR MANAGING HALLUCINATIONS, BIAS, AND STUDENT OVER-RELIANCE

Nigora Kh.Yuldasheva
Tashkent State Medical University, Tashkent, Uzbekistan
nigoray@inbox.ru

Nodiraxon M.Tulyganova
Center for the Development of Professional
Qualification of Medical Workers, Tashkent, Uzbekistan

Abstract

The rapid development of large language models (LLMs) has transformed the landscape of medical education. AI-powered tutoring systems are increasingly used by medical students for concept explanation, self-assessment, examination preparation, and clinical case discussions. While these technologies offer substantial educational benefits, their integration into medical training raises concerns regarding hallucinations, fabricated references, automation bias, algorithmic bias, and excessive dependence on AI-generated recommendations. This review examines current evidence regarding the responsible implementation of LLM-based educational tools in clinical reasoning training. Literature published between 2011 and 2024 was reviewed, with emphasis on studies conducted after the emergence of advanced generative AI systems. Available evidence suggests that AI tutors can enhance learning efficiency, accessibility, and personalized education. However, documented limitations include factual inaccuracies, reference fabrication, propagation of cognitive biases, and reduction of independent analytical thinking. The review proposes a framework for responsible integration based on transparency, human oversight, evidence verification, active learning strategies, and institutional governance. The future role of AI tutors in medicine should focus on augmenting rather than replacing human teaching and clinical reasoning.

Keywords: Artificial intelligence, medical education, large language models, clinical reasoning, automation bias, hallucinations, AI governance, medical training.

EduVision: Journal of Innovations in Pedagogy and Educational Advancements
Volume 2, Issue 6, June 2026
brightmindpublishing.com
ISSN (E): 3061-6972

Introduction

Artificial intelligence has emerged as one of the most influential technological developments affecting higher education during the past decade. The release of advanced large language models (LLMs), including ChatGPT and similar systems, has accelerated the adoption of AI-assisted learning across medical schools worldwide (4,8). Unlike traditional educational software, modern LLMs can engage in conversational interactions, explain complex concepts, generate clinical scenarios, provide feedback, and adapt responses to individual learning needs (13).

Medical education appears particularly suited for AI-supported learning because it requires continuous acquisition of large volumes of knowledge, development of clinical reasoning skills, and repeated exposure to case-based learning. AI tutors can provide immediate responses, generate unlimited practice questions, and simulate patient encounters at virtually no additional cost (10).

Despite these advantages, concerns have emerged regarding the reliability and educational consequences of AI-generated content. Unlike conventional databases, LLMs generate responses probabilistically rather than retrieving verified facts. Consequently, they may produce plausible but incorrect information, a phenomenon commonly referred to as hallucination. In healthcare education, where inaccurate information may affect future patient care, such errors represent a significant challenge (11).

The purpose of this review is to examine current evidence regarding the benefits and risks of AI tutoring systems in medical education and to propose evidence-based strategies for their responsible integration into clinical reasoning training.

Evolution of AI in medical education

Digital technologies have long supported medical education through online learning platforms, simulation environments, and computerized testing systems. However, previous educational technologies were generally limited to predefined content and rule-based interactions.

The emergence of transformer-based neural networks fundamentally changed this paradigm. Modern LLMs can process large volumes of medical literature, generate coherent explanations, summarize evidence, and engage in sophisticated dialogue. Consequently, students increasingly use AI systems as personalized tutors rather than merely as information retrieval tools (6,12).

EduVision: Journal of Innovations in Pedagogy and Educational Advancements
Volume 2, Issue 6, June 2026
brightmindpublishing.com
ISSN (E): 3061-6972

Recent surveys indicate that medical students commonly employ AI applications for clarification of difficult concepts, examination preparation, generation of study materials, clinical case discussions, summarization of scientific literature, assistance with academic writing.

This shift has created opportunities for more personalized learning but has simultaneously raised questions about educational quality and academic integrity.

Educational benefits of AI tutors

Personalized learning. One of the most significant advantages of AI tutors is their ability to adapt explanations according to the learner's level of understanding. Students can request simplified or advanced explanations, repeat questions without embarrassment, and learn at their own pace. Unlike traditional classroom settings, AI systems provide immediate responses and unlimited opportunities for practice. Such flexibility may improve engagement and support self-directed learning (6).

Enhanced accessibility. AI tutoring systems are available continuously and can support learners regardless of geographical location or institutional resources. This characteristic may be particularly valuable in low-resource educational environments where access to experienced faculty members is limited. The democratization of educational support represents one of the strongest arguments for responsible AI integration in global medical education (9).

Support for clinical reasoning development. Clinical reasoning is a complex cognitive process involving data interpretation, hypothesis generation, differential diagnosis, and decision-making. Modern AI systems can simulate patient encounters, generate evolving clinical scenarios, and encourage diagnostic discussions. Preliminary studies suggest that AI-assisted case-based learning may improve diagnostic thinking and facilitate exposure to diverse clinical situations (7).

Furthermore, LLMs have demonstrated notable performance on medical licensing examinations, including United States Medical Licensing Examination (USMLE)-style questions. These findings suggest potential educational value when AI systems are used as supplementary learning tools (6).

EduVision: Journal of Innovations in Pedagogy and Educational Advancements
Volume 2, Issue 6, June 2026
brightmindpublishing.com
ISSN (E): 3061-6972

Risks associated with AI tutors

Hallucinations and fabricated references. Perhaps the most widely discussed limitation of LLMs is their tendency to generate false information presented with high confidence.

In academic settings, hallucinations may manifest as: fabricated scientific references; incorrect citations; nonexistent clinical guidelines; inaccurate statistical data; misleading medical explanations. Students lacking sufficient expertise may fail to recognize these inaccuracies and incorporate erroneous information into their learning process. The problem is particularly concerning in medicine because educational errors may ultimately affect patient care (2).

Automation bias. Automation bias refers to excessive reliance on automated recommendations, resulting in reduced critical evaluation. Research demonstrates two major forms: errors of commission (users follow incorrect AI recommendations despite contradictory evidence) and errors of omission (users fail to consider alternative actions because the AI system did not suggest them).

Medical students remain especially vulnerable because they often possess insufficient experience to challenge apparently authoritative responses. Over time, repeated acceptance of AI-generated conclusions may weaken independent analytical skills and diminish diagnostic vigilance.

Bias transfer and cognitive distortion. AI systems inherit biases present within training datasets and may reproduce systematic distortions. Recent experimental studies demonstrate that users frequently adopt biased AI recommendations even after AI assistance is removed. This phenomenon, known as bias transfer, suggests that AI systems may influence future human decision-making patterns.

Over-reliance and dependency. A growing concern is the possibility that students may increasingly substitute independent learning with AI-generated answers. Clinical reasoning develops through active engagement with uncertainty, evidence evaluation, and reflective practice. Excessive dependence on AI-generated responses may bypass these cognitive processes and reduce opportunities for deeper learning (1,3).

EduVision: Journal of Innovations in Pedagogy and Educational Advancements
Volume 2, Issue 6, June 2026
brightmindpublishing.com
ISSN (E): 3061-6972

Strategies for risk mitigation

Retrieval-augmented generation (RAG) systems retrieve information from verified external sources before generating responses. This approach improves factual grounding and provides traceable evidence. Reducing hallucination rates, improving transparency, enhancing source verification, greater educational trustworthiness is potential benefits of systems.

Human-in-the-loop supervision. AI tutors should function under faculty oversight rather than independently. Educators remain responsible for verifying educational accuracy, monitoring learning outcomes, identifying misconceptions, providing contextual interpretation.

Effective AI tutors should encourage hypothesis generation, request justification of reasoning, challenge assumptions, promote reflective thinking.

Students should understand how AI systems generate responses, limitations of training data, risks of hallucinations, uncertainty in generated outputs.

The integration of AI into medical education raises important **ethical issues.** Students must never upload identifiable patient information into public AI systems.

Educational institutions should establish clear policies consistent with data-protection regulations and related international standards (14, 15).

Successful AI integration requires governance frameworks addressing quality assurance, faculty training, ethical oversight, risk monitoring, periodic evaluation. Institutions should treat AI implementation as a strategic educational initiative rather than a purely technological intervention.

Future AI tutors will likely become increasingly multimodal, capable of integrating text, images, laboratory data, and clinical simulations into unified educational environments. Advances in retrieval systems, explainable AI, and domain-specific medical models may significantly reduce hallucination rates and improve reliability (5).

However, the ultimate goal should not be replacement of educators. Instead, AI should augment human teaching by expanding access, increasing efficiency, and supporting personalized learning.

EduVision: Journal of Innovations in Pedagogy and Educational Advancements
Volume 2, Issue 6, June 2026
brightmindpublishing.com
ISSN (E): 3061-6972

Conclusion

Large language models have significant potential to enhance medical education through personalized learning, support for clinical reasoning, and improved accessibility. However, important challenges remain, including hallucinations, fabricated references, automation bias, bias transfer, and over-reliance on AI-generated content. AI tutors responsible implementation requires source verification, active learning approaches, human oversight, transparency, privacy protection, and institutional governance.

AI should augment, not replace, medical education. When used responsibly, AI tutors can support learning while preserving the critical thinking, professional judgment, and ethical standards essential to clinical practice.

References

1. Albert Henry T. 2 in 3 physicians are using health AI—up 78% from 2023. American Medical Association; 2025.
2. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: The role of artificial intelligence in clinical practice. BMC Med Educ 2023;23(1):689. doi: 10.1186/s12909-023-04698-z.
3. Basnawi A, Koshak A. Application of artificial intelligence in advanced training and education of emergency medicine doctors: A narrative review. Emerg Care Med 2024;1(3):247–59. doi: 10.3390/ ecm1030026.
4. Busch F, Hoffmann L, Truhn D, et al. Global cross-sectional student survey on AI in medical, dental, and veterinary education and practice at 192 faculties. BMC Med Educ. 2024;24:1066. doi:10.1186/s12909-024-06035-4
5. General Medical Council. Artificial intelligence and innovative technologies. [cited January 12, 2025]. https://www.gmc-uk.org/ professional-standards/learning-materials/artificial-intelligence-and-innovative-technologies.
6. Gligorea I, Cioca M, Oancea R, Gorski AT, Gorski H, Tudorache P. Adaptive learning using artificial intelligence in e-learning: A literature review. Educ Sci (Basel) 2023;13(12):1216. doi: 10.3390/ educsci13121216.
7. Hamilton A. Artificial intelligence and healthcare simulation: The shifting landscape of medical education. Cureus 2024;16(5):e59747. doi: 10.7759/ cureus.59747.

EduVision: Journal of Innovations in Pedagogy and Educational Advancements
Volume 2, Issue 6, June 2026
brightmindpublishing.com
ISSN (E): 3061-6972

8. Lin SY, Shanafelt TD, Asch SM. Reimagining clinical documentation with artificial intelligence. Mayo Clin Proc. 2018;93(5):563–565. doi:10.1016/j.mayocp.2018.02.016
9. Mekki YM, Zughaier SM. Teaching artificial intelligence in medicine. Nat Rev Bioeng. 2024;2:450–1.
10. Omiye JA, Gui H, Rezaei SJ, Zou J, Daneshjou R. Large Language models in medicine:the potentials and pitfalls. Ann Intern Med. 2024;177:210–20.
11. Tariq M, Hayat Y, Hussain A, Tariq A, Rasool S. Principles and perspectives in medical diagnostic systems employing artificial intelligence (AI) algorithms. International Research Journal of Economics and Management Studies (IRJEMS) 2024;3(1). Artificial intelligence in medical education: Current scenario and future perspectives. J Adv Med Educ Prof 2023;11(3):133–40.
12. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large Language models in medicine. Nat Med. 2023;29:1930–40.
13. World Medical Association. WMA statement on augmented intelligence in medical care. Ferney-Voltaire: WMA; 2019
14. Xu X, Chen Y, Miao J. Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: A systematic scoping review. J Educ Eval Health Prof 2024;21:6. doi: 10.3352/jeehp.2024.21.6.
15. Yan M, Cerri GG, Moraes FY. ChatGPT and medicine: how AI Language models are shaping the future and health related careers. Nat Biotechnol. 2023;41:1657–8.