Asian Journal of Technology and Applied Sciences
ISSN: 2301-xxxx (online) | IF: 10.01 | Vol: 1, issue 1
Available: http://asiansciencejournal.org

THE ROLE OF ARTIFICIAL INTELLIGENCE AND OPPORTUNITIES FOR ITS USE IN MEDICAL STUDENT EDUCATION

Zuhra Kamoliddin Sattorova
Faculty of Medicine, Profi University
General Medicine Programme, Group 02
E-mail: zuhrasattorova474@gmail.com
Academic supervisor
Munavvar Shavkat Ibodullayeva
Lecturer, Department of Medicine
Navoi Branch of Profi University

Abstract: This article examines the opportunities, pedagogical conditions, and limitations associated with the use of artificial intelligence in medical student education. The study aims to identify the functions of generative artificial intelligence in knowledge acquisition, the discussion of clinical scenarios, and independent learning, and to develop recommendations for responsible use. The article is based on a narrative analysis of six selected scholarly and institutional sources. These sources indicate that model performance on examination questions, student learning outcomes, and effects on clinical reasoning cannot be treated as equivalent measures. Based on the analysis, a learning procedure is proposed that combines an independent response, discussion with artificial intelligence, source-based verification, and instructor assessment. The article concludes that the pedagogical value of the technology should be assessed through students' ability to justify their reasoning, identify errors, and make independent decisions, rather than through response speed.

Keywords: medical education, artificial intelligence, generative model, clinical reasoning, independent learning, digital literacy.

Introduction

Medical education requires students not only to retain information but also to apply it to clinical scenarios, compare alternative explanations, and demonstrate the evidence supporting a decision. The introduction of digital tools that provide rapid explanations therefore calls for a reconsideration of educational content and assessment methods. Kung et al. found that, under certain conditions, ChatGPT performed at or near the passing threshold on questions from the United States Medical Licensing Examination. However, this was an assessment of model knowledge, rather than an educational experiment testing whether students learned more effectively with the tool. This distinction defines the central research problem: under what conditions does using a system capable of generating correct answers lead to independently acquired knowledge and skills?

The randomized study involving medical students conducted by Gan et al. directly examined the effects of an educational tool. By contrast, Goh et al. demonstrated in a trial involving physicians that access to a large language model did not automatically improve diagnostic reasoning. These studies differ in their participants, tasks, and outcome measures. Presenting them as a single, consistent pattern of success would weaken the scientific interpretation. The World Health Organization highlights misinformation, bias, and data security concerns alongside the potential use of generative models in medical education. UNESCO guidance prioritizes human agency, pedagogical objectives, and data protection.

Surveys of students' experiences and attitudes constitute a separate type of evidence that highlights the need to distinguish technology acceptance from educational effectiveness.

The aim of this article is to distinguish the pedagogical opportunities offered by artificial intelligence in medical education according to the level of evidence and to connect them to a practical learning procedure. Its objectives are to differentiate technological performance from learning outcomes, compare existing studies, interpret risks in relation to educational activities, and develop recommendations that can be tested in a local setting. The object of the study is medical student education; its specific focus is the didactic opportunities and organizational conditions associated with the use of generative artificial intelligence.

Materials and Methods

The article was prepared as a narrative literature analysis based on purposively selected sources. It is not presented as a systematic review or meta-analysis. The analytical corpus comprised four original studies and two international guidance documents published in 2023–2024. The original studies included an assessment of model capabilities, an educational trial involving students, a clinical reasoning trial involving physicians, and a survey of student attitudes. The sources were checked on 18 October 2026 using publicly accessible publisher and international organization webpages. The search terms included "medical education," "medical students," "ChatGPT," "randomized controlled trial," "diagnostic reasoning," and "AI ethics." Selection criteria were direct relevance to the topic, an explicitly described research design, verifiable bibliographic information, and the potential to inform educational conclusions. Promotional materials and texts with unidentified authors or publication outlets were not used as primary evidence. Each source was compared across five characteristics: participants or unit of analysis, the role of artificial intelligence, outcome measure, principal finding, and limits of applicability. International guidance documents were interpreted as establishing conditions for use rather than demonstrating experimental effects. Because the studies used different measures and populations, their findings were not converted into a pooled percentage or a single effect size.

No survey, observation, or educational experiment was conducted among students at Profi University. Accordingly, the numerical findings reported in this article refer exclusively to the cited publications, while the practical model is the author's proposal. No empirical claims are made about students' actual knowledge levels, frequency of technology use, or local educational effectiveness. No primary personal or clinical data were collected.

Results

The first finding concerns the need to distinguish three levels of evidence. Model capability describes the response generated by a system for a given task; an educational outcome measures a change in student learning; and professional competence refers to the transfer of knowledge to independent clinical practice. In this article, this distinction is proposed as an analytical framework for comparing the sources. Success at the first level does not, by itself, establish success at the second or third level.

The findings of Kung et al. demonstrate a language model's capacity to generate explanatory answers to medical questions [1]. Their educational relevance lies in the possibility of providing additional material for discussion. However, performance on examination questions does not measure patient communication, practical skills, or long-term knowledge retention. This evidence therefore cannot be directly translated into a conclusion about students' professional readiness. The study by Gan et al. initially enrolled 129 students, of whom 110 completed follow-up. Following a two-week intervention, benefits were observed in some theoretical tasks. However, the difference in overall end-of-semester scores

was not statistically significant [2]. It is therefore unjustified to interpret an advantage in a particular subject or task as a sustained improvement in medical training as a whole.

In a trial involving 50 physicians, Goh et al. reported median diagnostic reasoning scores of 76% in the group with access to the model and 74% in the conventional-resources group. The adjusted difference was 2 percentage points, with a 95% confidence interval from −4 to 8 and $p = 0.60$ [3]. Although this result cannot be directly generalized to students, it shows that access to a tool and effective use of that tool are separate issues.

Table 1. Selected evidence and the limits of its interpretation

Source and design	Outcome measured	Limits of interpretation
[1] Model evaluation	Responses to USMLE questions	Student learning was not measured
[2] Student trial	Theory tests and semester grades	No advantage across all outcomes
[3] Physician trial	Diagnostic reasoning	Participants were not students
[6] Cross-sectional survey	Experiences and subjective perceptions	Does not establish causality
[4], [5] Guidance documents	Ethical and pedagogical conditions	No experimental effect was measured

The table was compiled by the author using the cited sources. The principal conclusion from this comparison is that model accuracy, user satisfaction, and independent professional competence represent different outcomes. Evaluations of educational programmes must specify in advance which outcome they are intended to achieve.

Discussion

Based on the comparison of the sources, we propose assessing the pedagogical value of artificial intelligence through the actions students perform with knowledge. For example, a student who copies a prepared explanation and a student who verifies its causal relationships may use the same digital tool while engaging in different learning activities. Assessment should therefore consider both the final product and the evidence underlying it. Suitable independent learning tasks may include explaining an unfamiliar concept at different levels of complexity, comparing one's own response with an alternative explanation, and generating self-assessment questions. These are didactic approaches recommended in the present article. For example, a student studying physiology first explains a process in their own words and then asks the model to identify missing connections in that explanation. The final revision is checked against the relevant textbook content. When working with clinical scenarios, students should formulate the initial hypothesis themselves. The student defines the problem in a scenario developed for educational purposes, offers alternative explanations, and identifies missing information. The model may then be asked to provide arguments against the student's hypothesis. Finally, the instructor evaluates the logical consistency of the response. The advantages of this procedure have not yet been confirmed in a local trial; it is an approach proposed for testing. The experiences and perceptions examined by Alkhaaldi et al. help clarify users' views [6]. Such information is useful for identifying learning needs, but preference or convenience cannot be equated with learning outcomes. In our interpretation, future research should record subjective assessments separately from performance on independently completed tasks. This distinction would help identify cases in which confidence increases without a corresponding improvement in knowledge.

The risks of inaccurate content and excessive reliance on automated responses identified by WHO are particularly relevant to medical education [4]. Students need to verify the evidential accuracy of fluently written text. We propose recording the textbook section or approved guidance document used to support each important claim. Even when a bibliographic reference exists, whether it supports the specific statement in question must be checked separately.

At a medical education institution such as Profi University, implementation could begin within a single subject and with a clearly defined learning objective. The instructor first specifies the knowledge students should acquire, the tools they are permitted to use, and the required components of the submitted work. To ensure equal access, paid services should not become a mandatory requirement; where necessary, students may use an identical response prepared by the instructor. The proposed learning procedure consists of four stages. First, the student writes an initial response without artificial intelligence. Second, the student discusses that response with the model and identifies areas of uncertainty. Third, the explanations received are compared with an approved source, and accepted or rejected statements are marked. Fourth, a brief oral defense is conducted with the instructor. These stages help distinguish independent thinking, external assistance, and final knowledge. The task sheet provides space for the initial response, the prompt submitted to the model, the verified source, and the student's final conclusion. A sample prompt is: "Check the causal relationships in my explanation, identify unclear points, and generate three questions to test my understanding of the answer." This prompt alone does not guarantee accuracy. Verification against a source recommended by the instructor remains necessary. An author-developed assessment rubric is proposed: accuracy of knowledge, justification of reasoning, error identification, and use of sources are each scored from 0 to 2. A score of 0 indicates that the criterion has not been met, 1 that it has been partly met, and 2 that it has been met with adequate justification. The resulting eight-point rubric is not a validated instrument. Before using it, instructors should independently assess several sample submissions to check whether they interpret the criteria consistently. Patient names, contact information, copies of medical records, or identifying details must not be entered into a publicly accessible model. Synthetic scenarios created specifically for teaching are preferable. The principles of data protection and pedagogical oversight in UNESCO guidance support this organizational approach [5]. A brief statement specifying how artificial intelligence was used in the submitted work also promotes academic transparency.

The first limitation of this analysis concerns the small number of sources and their purposive selection. No comprehensive systematic search, screening by two independent reviewers, or formal risk-of-bias assessment was conducted. The conclusions should therefore not be interpreted as a definitive judgment covering all available research. The second limitation is the variation in educational settings, participants, and tasks across the sources. Results from English-language tasks do not establish the effectiveness of Uzbek-language teaching activities. The third limitation concerns changes in model versions and operating conditions. Findings from 2023–2024 do not imply that all current services have the same level of accuracy. Local trials should therefore retain the service name, model version where available, date of use, and task prompt. This is recommended to enable replication and distinguish educational outcomes from technical changes. A future study could conduct a controlled educational trial within a single subject. Groups would follow the same curriculum and have the same time allocation; one group would use conventional resources, while the other would use artificial intelligence within the verification procedure described above. Baseline knowledge, independently completed post-session tasks, and

knowledge retention several weeks later would be assessed. The required sample size should be calculated in advance, and assessors should, where possible, be unaware of group allocation. This is a proposal for future research, not a completed experiment.

Conclusion

The analysis of the selected sources indicates that artificial intelligence is a promising supplementary tool for explanation, question generation, and the discussion of educational scenarios in medical education. However, a model's performance in answering questions, students' attitudes toward the technology, and their independent professional readiness are different measures. Treating them as equivalent leads to overstated conclusions about effectiveness.

The author's proposal combines discussion of an independently prepared response with the model, source-based verification, and an oral defense. The practical value of this approach lies in making explicit which claims the student accepted and why. Local educational evaluation is required before it can be described as a method with demonstrated effectiveness. In medical education, the primary outcome is students' ability to justify their knowledge and apply it without assistance, rather than the volume of generated text or the speed of a response.

REFERENCES

1. Kung T. H., Cheatham M., Medenilla A., Sillos C., De Leon L., Elepaño C. etal. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health. 2023. Vol. 2, No. 2. e0000198. DOI: 10.1371/journal.pdig.0000198.
2. Gan W., Ouyang J., Li H., Xue Z., Zhang Y., Dong Q., Huang J., Zheng X., Zhang Y. Integrating ChatGPT in Orthopedic Education for Medical Undergraduates: Randomized Controlled Trial. Journal of Medical Internet Research. 2024. Vol. 26. e57037. DOI: 10.2196/57037.
3. Goh E., Gallo R., Hom J. etal. Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial. JAMA Network Open. 2024. Vol. 7, No. 10. e2440969. DOI: 10.1001/jamanetworkopen.2024.40969.
4. World Health Organization. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. Geneva: World Health Organization, 2024. 98 p. ISBN: 978-92-4-008475-9. URL: https://www.who.int/publications/i/item/9789240084759 (accessed 18 October 2026).
5. Miao F., Holmes W. Guidance for generative AI in education and research. Paris: UNESCO, 2023. 48 p. ISBN: 978-92-3-100612-8. DOI: 10.54675/EWZM9535.
6. Alkhaaldi S. M. I., Kassab C. H., Dimassi Z., Oyoun Alsoud L., Al Fahim M., Al Hageh C., Ibrahim H. Medical Student Experiences and Perceptions of ChatGPT and Artificial Intelligence: Cross-Sectional Study. JMIR Medical Education. 2023. Vol. 9. e51302. DOI: 10.2196/51302.