Artificial Intelligence in Medical Education

Tıp Eğitiminde Yapay Zeka

● Sezen Milli Avtan

University of Health Sciences Türkiye, Hamidiye Faculty of Medicine, Department of Medical Education, İstanbul, Türkiye

ABSTRACT

The concept of artificial intelligence (AI) encompasses systems capable of exhibiting intelligent behaviors such as perception, reasoning, learning, and communication, and performing tasks traditionally requiring human cognition. In contemporary practice, AI paradigms are increasingly employed across various domains, most notably in the automotive, finance, economics, healthcare, and education sectors. Within the context of medical education, the integration of AI-based systems, often referred to as "teacher bots," has emerged as a noteworthy innovation. These systems serve as content providers, feedback generators, and instructional modulators, enhancing the educational process. AI modules have found particular relevance in widely adopted instructional strategies such as problem-based learning (PBL) and Objective Structured Clinical Examinations (OSCEs) in undergraduate medical education. In both PBL and OSCE settings, AI technologies effectively support clinical decision-making, the generation of virtual patient cases, real-time feedback delivery, and the simulation of patient interactions. The incorporation of deep learning (DL) techniques and artificial neural networks into AI platforms has further enabled the execution of more complex and nuanced tasks. In postgraduate medical training, especially in visually intensive specialties such as radiology, dermatology, and pathology, AI has facilitated the development of machine learning models for diagnostic verification, the generation of synthetic patient imagery, and the instruction of key diagnostic features, independent of real patient data. These applications have been successfully implemented and demonstrate significant educational potential. This study aims to present selected examples of AI applications currently utilized in medical education. Furthermore, it will explore the challenges encountered, or likely to be encountered, in the implementation of AI, as well as the potential contributions AI may offer to the future of medical training.

Keywords: Artificial intelligence, medical education, undergraduate medical education

ÖZ

Yapay zeka (YZ) kavramı, algı, akıl yürütme, öğrenme veya iletişim becerileri gibi zeki davranışlar sergileyebilen ve insani görevleri yerine getirme yetisine sahip sistemleri tanımlar. Günümüzde YZ paradigmaları en çok otomotiv, finans, ekonomi, tıp ve eğitim alanlarında kullanılmaktadır. Tıp eğitiminde, "teacher bots" olarak tanımlanan; eğitim sürecine içerik sağlayıcı, geribildirim veren ve eğitimi modüle edici özellikleri olan sistemlerin yaygınlaşması önemli bir yenilik olarak kabul edilmektedir. Mezuniyet öncesi tıp eğitiminde yaygın olarak kullanılan probleme dayalı öğrenme (PBL) ve nesnel yapılandırılmış klinik sınavlar (OSCE)'da da YZ modülleri kullanıma girmiştir. PBL eğitimleri ve OSCE sınavlarında klinik karar verme, sanal vaka oluşturma, anında geribildirim ve simüle hasta desteği etkili ve verimli bir şekilde YZ tarafından sağlanmaktadır. Derin öğrenme (DL) ve yapay sinir ağlarının YZ modüllerine entegre edilmesiyle daha komplike görevlerin gerçekleştirilmesi mümkün olmuştur. Özellikle mezuniyet sonrası uzmanlık eğitiminde; radyoloji, dermatoloji ve patoloji gibi görsel figürlerin ve tanımlamaların değerli olduğu branşlarda, klinik tanıyı doğrulamaya yönelik makine öğrenmesi, yapay hasta görüntülerinin oluşturulması ve bazı spesifik tanıların olmazsa olmaz anahtar figürlerinin hasta görüntüsü olmaksızın yapay olarak eğitim alanlara öğretilmesi gibi didaktik çalışmalar yapılmış ve başarıyla uygulanmıştır. Çalışmamızda tıp eğitiminde kullanılan bazı YZ çalışmalarından örnekler sunulacaktır. Bunun yanı sıra YZ'nin uygulamasında yaşanan veya yaşanabilecek güçlükler ve ileride tıp eğitimine sağlayabileceği katkılar üzerinde durulacaktır.

Anahtar Kelimeler: Yapay zeka, tıp eğitimi, mezuniyet öncesi tıp eğitimi

Introduction

Although artificial intelligence (AI) has gradually become integrated into our lives over the years, it is still not a concept widely understood or commonly used by many people. With the emergence of deep learning and artificial neural networks (ANNs), the significance of these technologies has steadily increased. AI refers to machines capable of exhibiting intelligent behavior such as perception, reasoning, learning, or communication, and performing human-like tasks (1). AI paradigms target various problem areas such as perception, reasoning, knowledge, planning, and communication. Today, AI applications are intensively utilized in fields such as automotive, finance, economics, medicine, and education. The use of AI in the medical field continues to grow rapidly (1). One notable advancement is that machines can achieve diagnostic success rates in radiology comparable to, or even exceeding, expert consultants (2). In addition to AI's widely publicized role in radiological diagnosis, it is also being employed as a tool for the optimal management of chronic diseases such as cancer and persistent mental health disorders (2).

In education, notable applications of AI include "teacher bots," which are teaching assistants responsible for delivering content, providing feedback, and monitoring educational progress. This growing use in education holds the potential to offer individualized support to students and identify knowledge gaps. In this way, educators can be relieved from routine tasks and offer more effective support to students, thereby enhancing the personalized and adaptive teaching process. Students also gain time to develop unique and individualized learning techniques. This aligns with a rising trend in medical education that emphasizes student autonomy in customizing their learning experience to best suit their comprehension (3,4). Medical education is a lifelong learning process spanning undergraduate, graduate, residency, and postgraduate stages, which includes "continuing medical education" (5). This process involves not only interactions with physicians but also with other healthcare professionals, including nurses and allied health personnel. Currently, there are a limited number of studies reviewing or discussing the existing applications of AI in medical education (6,7). The aim of this study is to comprehensively review the current academic literature on the use of AI in medical education. The study addresses the following questions:

• How is AI currently being used in medical education?
• What are the key challenges in implementing AI in medical education?
• How might the relationship between AI and medical education evolve in the future?

Medical Education and AI

The integration of AI and ANNs into the evaluation of medical students' curricula represents a significant advancement. One article discussed the use of AI, ANN and support vector machines (SVM) for assessing the curriculum of medical students. Chen et al. (8) described the advantage of ANN and SVM over logistic regression in the data analysis: they are more adept models for solving nonlinear problems and establishing relationships between variables.

The use of AI in evaluating medical education curricula is highly valuable, as it provides a comprehensive overview of program effectiveness and student satisfaction. Accurate evaluation is essential for assessing various educational components, ranging from entire curricula to the efficiency of small-group instruction, through diverse methodologies commonly referred to as program evaluation in medical education (8).

In curriculum development, the collection of digital data and the analysis of printed and visual materials are of great importance. While AI performs well in collecting digital data, it has not yet reached sufficient maturity in analyzing visual materials. Nonetheless, AI has found extensive use in curriculum evaluation in medical schools across Canada and the United Kingdom (9,10). AI has also been utilized in student assessments to identify knowledge gaps following examinations and to determine whether learning objectives and assessment strategies are aligned (11). Additionally, AI tools that provide feedback as a critical component for fostering lasting learning strongly support educational processes. Studies indicate that, particularly at the postgraduate level, many residents report receiving little feedback and feeling unsupported in this area, whereas AI-based tools offering instant, formative feedback have shown promising results (11,12).

However, certain observations exist regarding the nature of feedback. Effective feedback should be performance-focused and structured to support the attainment of learning objectives. While AI is capable of providing immediate and prompt feedback, its feedback pool is limited to existing data, and it lacks depth in experiential and emotional inference (13,14).

AI has also proven effective in problem-based learning (PBL) and Objective Structured Clinical Examinations (OSCE), both of which have been shown to offer more effective and efficient training compared to traditional methods (15,16). The use of AI is expanding in OSCEs and similar exams to ensure exam security, evaluate programs based on exam outcomes, and align exam results with learning objectives. AI is also playing a growing role in designing new educational programs (17,18).

System models developed using virtual reality (VR) simulation programs such as Touch, Lahystotrain, and EchoComJ, in combination with intelligent tutoring systems, have reinforced the use of AI in surgical specialties. These systems provide the benefits of both immersive virtual environments and smart instructional systems. Immersive, interactive, and safe VR settings are particularly effective in eliminating the risks associated with lengthy, strenuous, and potentially unsafe training scenarios for learners (19).

In healthcare, generative AI (GenAI) tools are increasingly being utilized in clinical settings, particularly in areas such as clinical documentation and physician-patient communication. These tools have shown promise in addressing inefficiencies in electronic health records, challenges with big data, and healthcare worker burnout (20-22).

Within the context of medical education, GenAI offers several potential advantages, including facilitating personalized learning experiences, simulating real-life scenarios and patient interactions, and enhancing communication skills training (23). However, these benefits also carry significant risks, such as concerns regarding the reliability of AI-generated content and threats to academic integrity (24,25).

Graduate medical education (GME) shares many characteristics with both underGME and other forms of health education. Research has shown that adult learners achieve better learning outcomes when motivated and autonomous, particularly when using AI-assisted learning methods focused on practical applications (26).

Historically, medical education has equated time spent in educational environments with learning success. However, in recent years, a shift toward competency-based medical education (CBME), which prioritizes the acquisition of specific competencies over time spent, has gained renewed attention (27,28). CBME forms the foundation of the Accreditation Council for GME's (ACGMEs) accreditation model. ACGME programs use "Milestones," a system designed to assess and enhance educational progress based on competency-based learning (29).

After receiving foundational training in medical sciences and basic clinical skills, residents spend little time in traditional classrooms. Most learning occurs in real clinical environments as part of a healthcare team. One of the core principles of GME, "graduated responsibility," allows learners to develop increasing levels of autonomy, ultimately achieving readiness for independent practice.

Moreover, residents are expected to become "physician-scholars." Participants in ACGME-accredited GME programs engage in academic activities such as research, scholarly writing, quality improvement initiatives, and curriculum development (30). AI modules, when aligned with defined competencies and regularly updated, provide outcome-based assessment opportunities in line with educational objectives (10-12).

In both undergraduate and postGME, AI is also used for objective assessment of students' work, including the evaluation of their portfolios and dossiers. The key benefit of this system is the ability to provide instant feedback and quickly correct errors through machine learning-based operations (31).

Numerous publications have explored the potential of GenAI in the context of residency training. For example, VR-based simulations of children with rare genetic conditions have been used in pediatric residency training instead of live patient interactions (32). Large language models (LLMs) have also been effectively used to enhance clinical decision-making skills during pediatric education (33,34). In surgical training, AI models are employed in case-based learning focused on ethical dilemmas and in the evaluation of various surgical scenarios (35,36). In studies where LLMs simulate patient conversations for specific anesthesia procedures, AI modules have demonstrated near-realistic accuracy in modeling patient reactions and behaviors (37).

One of the earliest examples of innovative AI-based teaching methods involved enhancing emergency physicians' communication skills, particularly in delivering bad news. This model simulated patient responses and dialogue during the disclosure of difficult diagnoses such as cancer (38).

Challenges in the Application of AI in Medical Education

When introducing a new model or method in medical education, the most critical factor is clearly demonstrating the benefit it provides. To evaluate this, two key aspects must be considered: how easily can the system's effectiveness be assessed? What are the difficulties and limitations in measuring the system? Studies have shown that the most reliable approach involves comparisons with traditional teaching methods. In evaluating both the model and its educational effectiveness, pre-test and post-test results should be examined, and it must be ensured that the baseline knowledge levels of comparison groups are approximately equal before any educational intervention is introduced (39). There are numerous studies showing that AI modules outperform or underperform human-based models depending on the context.

GenAI models have been found to be significantly more effective than traditional methods in various applications. They are actively used in clinical decision support, medical education, clinical documentation, research assistance, and as communication tools (40). Even though models like

ChatGPT have not been trained on specialized medical datasets, they have demonstrated near-passing or passing performance on all three stages of the United States Medical Licensing Examination (41). In some medical exams using LLMs, LLMs have achieved performance comparable to that of final-year medical students (42-45).

In one study involving AI-assisted diagnostic simulators, a statistically significant 22% improvement in diagnostic accuracy was reported compared to traditional methods. However, in another study, a web- and multimedia-based AI educational model showed an 8% improvement in student success, although this was not found to be statistically significant (46,47). The broad specialization within the field of medicine limits the applicability of any single AI model, which in turn narrows the research sample needed for reliable measurement. Other notable challenges include a shortage of experts who can design curricula compatible with machine learning and the temporal, spatial, and interpersonal difficulties in fostering collaboration between physicians and engineers. Developing AI models necessitates a multidisciplinary team comprising data scientists (to collect and process large datasets), medical professionals (to validate clinical applicability), and, ideally, biomedical engineers with dual expertise in both domains (46,47).

Some studies also emphasize the ethical challenges associated with developing and deploying AI models. These include concerns about preserving patient privacy during data collection and safeguarding user data confidentiality (46,47).

Although AI has gained considerable traction in medical diagnosis, clinical reasoning remains inherently complex. Clinical reasoning involves deep learning, deductive thinking, and substantial emotional input, which makes it unrealistic, at least for now, to expect AI to match the diagnostic capabilities of highly experienced clinicians. However, with robust implementation of machine learning and deep learning frameworks, and with active support from domain experts, AI's diagnostic capabilities can be significantly enhanced (48,49).

Several authors have discussed the potential of GenAI as a supportive tool in academic writing and research processes (50-55). This technology can be particularly helpful for non-native English speakers in improving writing skills and translating foreign language content. Many studies have highlighted the usefulness of GenAI in literature reviews and summarization tasks (56-59). However, these models have also been noted to exhibit a phenomenon known as "hallucination," where they fabricate references or present non-existent information.

This issue was clearly demonstrated in an editorial from Medical Teacher that exposed fabricated citations in a manuscript submitted for publication (60).

Such incidents and similar studies have drawn attention to unethical practices, such as presenting AI-generated content as human-authored work. They have emphasized the importance of maintaining awareness and adhering strictly to principles of academic integrity when using these tools (55-59).

Some publications have also warned of the potential negative impact of GenAI on learning. Overreliance on this technology may hinder the development of learners' critical thinking and complex problem-solving skills (56,57,59,61).

The widespread adoption of AI also poses challenges to the validity of current assessment and evaluation methods, necessitating adaptations in assessment strategies (62,63). Moreover, an excessive focus on AI-driven learning opportunities may impair human interaction and communication skills, fundamental components of medical education (51,64). Relying on AI as a primary source of knowledge could lead to the dissemination of inaccurate medical information. Therefore, the integration of AI into learning processes must be carried out in a balanced and well-regulated manner.

Future Directions

Research indicates a critical need for evidence-based studies that clearly demonstrate the superiority of AI over traditional methods. Future investigations should focus on evaluating the effectiveness of AI in medical education. To accurately assess the success of AI systems relative to conventional approaches, extensive and time-intensive research remains necessary across various medical subspecialties.

As medical curricula become increasingly digital and collaboration between data scientists and healthcare professionals intensifies, the use of AI systems is expected to expand. Consequently, data protection is emerging as an important area of inquiry. Specifically, there is a growing need for studies that explore how to enhance data security and bolster user trust in AI applications.

With the continuous advancement of technology, the potential applications of AI in medical education are expected to increase. One such development is the integration of AI with immersive technologies such as VR and augmented reality. These combinations promise to revolutionize educational experiences through simulation-based learning environments.

It is widely acknowledged that GenAI will have a broad societal impact and will be increasingly integrated into daily life. GenAI holds the potential to transform multiple sectors, including healthcare and education. Already, these systems are being used for document summarization, translation, conversation, language support, and image generation.

In the near future, they are expected to expand their capabilities to include emotion analysis and multilingual interaction.

As AI becomes more embedded in healthcare delivery processes, its integration into medical education is seen as both inevitable and transformative. This convergence has sparked intense debate regarding the potential roles, advantages, and limitations of AI in medical training.

Integrating such a transformative technology into existing educational systems requires a careful, evidence-based approach. Medical education experts must not only understand the technical capabilities and limitations of GenAI but also develop forward-looking strategies to guide its educational applications.

The articles reviewed in this study further emphasize the urgent need for research. Most current publications are speculative in nature or consist of opinion pieces. There is a significant gap in research that directly implements and evaluates this technology in student populations. To generate meaningful and applicable outcomes, future studies must be guided by carefully formulated research questions. Enhancing students' AI literacy, evaluating the impact of AI on assessment processes, identifying technical and ethical risks, and investigating the dynamics of human-AI interaction will all be essential steps.

Conclusion

Particularly when using AI-based technologies, providing students with face-to-face patient management experience is crucial. The use of AI-based applications, their role in medical education, and their advantages, disadvantages, and limitations is discussed. The active use of AI in medical education provides an innovative approach to student-centered learning. The integration of AI into education will enable the effective use of innovative technologies in future clinical practice, both in undergraduate education and in lifelong learning.

Footnotes

Conflict of Interest: No conflict of interest was declared by the author.

Financial Disclosure: The author declared that this study received no financial support.

REFERENCES

1. Russell S, Norvig P. Artificial intelligence: a modern approach, global edition. Harlow, United Kingdom: Pearson Education Limited; Nov 28, 2018. [Crossref]

2. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017;69S:S36-S40. [Crossref]

3. Bayne S. Teacherbot: interventions in automated teaching. Teaching in Higher Education. 2015;20:455-467. [Crossref]

4. Botrel L, Holz E, Kübler A. Brain painting V2: evaluation of P300-based brain-computer interface for creative expression by an end-user following the user-centered design. Brain-Computer Interfaces. 2015;2:135-149. [Crossref]

5. Bashook PG, Parboosingh J. Recertification and the maintenance of competence. BMJ. 1998;316:545-548. [Crossref]

6. Lillehaug SI, Lajoie SP. AI in medical education-another grand challenge for medical informatics. Artif Intell Med. 1998;12:197-225. [Crossref]

7. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. Acad Med. 2018;93:1107-1109. [Crossref]

8. Chen CK. Curriculum assessment using artificial neural network and support vector machine modeling approaches: a case study. In: IR Applications. Tallahassee, FL: Association for Institutional Research; 2010. [Crossref]

9. Frehywot S, Vovides Y, Talib Z, Mikhail N, Ross H, Wohltjen H, et al. E-learning in medical education in resource constrained low- and middle-income countries. Hum Resour Health. 2013;11:4. [Crossref]

10. Willett TG. Current status of curriculum mapping in Canada and the UK. Med Educ. 2008;42:786-793. [Crossref]

11. Hattie J, Timperley H. The power of feedback. Review of Educational Research. 2016;77:81-112. [Crossref]

12. Hewson MG, Little ML. Giving feedback in medical education: verification of recommended techniques. J Gen Intern Med. 1998;13:111-116. [Crossref]

13. Mason BJ, Bruning RH. Providing feedback in computer-based instruction: What the research tells us. CLASS Research Report 2001;9. [Crossref]

14. Stasiu RK, De Britto J, Da Silva Dias J, Scalabrin E. Teaching of electrocardiogram interpretation guided by a tutorial expert. In: Proceedings of the 14th IEEE Symposium on Computer-Based Medical Systems; 2001 Jul 26-27; Bethesda, MD. p. 487-492. [Crossref]

15. Shin JH, Haynes RB, Johnston ME. Effect of problem-based, self-directed undergraduate education on life-long learning. CMAJ. 1993;148:969-976. [Crossref]

16. Epstein RM. Assessment in medical education. N Engl J Med. 2007;356:387-396. [Crossref]

17. Ward JP, Gordon J, Field MJ, Lehmann HP. Communication and information technology in medical education. Lancet. 2001;357:792-796. [Crossref]

18. Van Der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJB, Heeneman S. Twelve tips for programmatic assessment. Med Teach. 2015;37:641-646. [Crossref]

19. Caudell TP, Summers KL, Holten J 4th, Hakamata T, Mowafi M, Jacobs J, et al. Virtual patient simulator for distributed collaborative medical education. Anat Rec B New Anat. 2003;270:23-29. [Crossref]

20. Liu TL, Hetherington TC, Stephens C, McWilliams A, Dharod A, Carroll T, et al. AI-powered clinical documentation and clinicians' electronic health record experience: a nonrandomized clinical trial. JAMA Netw Open. 2024;7:e2432460. [Crossref]

21. Garcia P, Ma SP, Shah S, Smith M, Jeong Y, Devon-Sand A, et al. Artificial intelligence-generated draft replies to patient inbox messages. JAMA Netw Open. 2024;7:e243201. [Crossref]

22. Small WR, Wiesenfeld B, Brandfield-Harvey B, Jonassen Z, Mandal S, Stevens ER, et al. Large Language model-based responses to patients' in-basket messages. JAMA Netw Open. 2024;7:e2422399. [Crossref]

23. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med. 2024;99:22-27. [Crossref]

24. Bhardwaj P, Bookey L, Ibironke J, Kelly N, Sevik IS. A Meta-analysis of the economic, social, legal, and cultural impacts of widespread adoption of large language models such as ChatGPT|OxJournal. (2023). [Crossref]

25. Knopp MI, Warm EJ, Weber D, Kelleher M, Kinnear B, Schumacher DJ, et al. AI-enabled medical education: threads of change, promising futures,

and risky realities across four potential future worlds. JMIR Med Educ. 2023;9:e50373. [Crossref]

26. Knowles MS III, Swanson RA. The adult learner: The definitive classic in adult Education and human resource development. 7th ed. London New York: Butterworth- Heinemann (2011). 424 p. [Crossref]

27. Carraccio C, Englander R, Van Melle E, Ten Cate O, Lockyer J, Chan MK, et al. Advancing competency-based medical education: a charter for clinician-educators. Acad Med. 2016;91:645-649. [Crossref]

28. Cecchini MJ, Borowitz MJ, Glassy EF, Gullapalli RR, Hart SN, Hassell LA, et al. Harnessing the power of generative artificial intelligence in pathology education: opportunities, challenges, and future directions. Arch Pathol Lab Med. 2025;149:142-151. [Crossref]

29. Edgar L, McLean S, Hogan S, Hamstra S, Holmboe E. The Milestones Guidebook. Accreditation Council for Graduate Medical Education. (2020). [Crossref]

30. Accreditation Council for Graduate Medical Education. ACGME common program requirements (residency). 2023. [Crossref]

31. Walkowski S, Lundin M, Szymas J, Lundin J. Exploring viewing behavior data from whole slide images to predict correctness of students' answers during practical exams in oral pathology. J Pathol Inform. 2015;6:28. [Crossref]

32. Waikel RL, Othman AA, Patel T, Hanchard SL, Hu P, Tekendo-Ngongang C, et al. Generative methods for pediatric genetics education. medRxiv [Preprint]. 2023:2023.08.01.23293506. [Crossref]

33. Ba H, Zhang L, Yi Z. Enhancing clinical skills in pediatric trainees: a comparative study of ChatGPT-assisted and traditional teaching methods. BMC Med Educ. 2024;24:558. [Crossref]

34. Suresh S, Misra SM. Large language models in pediatric education: current uses and future potential. Pediatrics. 2024;154:e2023064683. [Crossref]

35. Lia H, Atkinson AG, Navarro SM. Cross-industry thematic analysis of generative AI best practices: applications and implications for surgical education and training. Global Surg Educ. 2024;3:61. https://doi.org/10.1007/s44186-024-00263-4

36. Sathe TS, Roshal J, Naaseh A, L'Huillier JC, Navarro SM, Silvestri C. How I GPT It: Development of custom artificial intelligence (AI) chatbots for surgical education. J Surg Educ. 2024;81:772-775. [Crossref]

37. Sardesai N, Russo P, Martin J, Sardesai A. Utilizing generative conversational artificial intelligence to create simulated patient encounters: a pilot study for anaesthesia training. Postgrad Med J. 2024;100:237-241. [Crossref]

38. Webb JJ. Proof of Concept: Using ChatGPT to Teach Emergency Physicians How to Break Bad News. Cureus. 2023;15:e38755. [Crossref]

39. Suebnukarn S, Haddawy P. A Bayesian approach to generating tutorial hints in a collaborative medical problem-based learning system. Artif Intell Med. 2006;38:5-24. [Crossref]

40. Moulaei K, Yadegari A, Baharestani M, Farzanbakhsh S, Sabet B, Reza Afrash M. Generative artificial intelligence in healthcare: A scoping review on benefits, challenges and applications. Int J Med Inform. 2024;188:105474. [Crossref]

41. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2:e0000198. [Crossref]

42. Kung JE, Marshall C, Gauthier C, Gonzalez TA, Jackson JB 3rd. Evaluating ChatGPT Performance on the Orthopaedic In-Training Examination. JB JS Open Access. 2023;8:e23.00056. [Crossref]

43. Lum ZC. Can Artificial Intelligence Pass the American Board of Orthopaedic Surgery Examination? Orthopaedic Residents Versus ChatGPT. Clin Orthop Relat Res. 2023;481:1623-1630. [Crossref]

44. Cheong RCT, Pang KP, Unadkat S, Mcneillis V, Williamson A, Joseph J, et al. Performance of artificial intelligence chatbots in sleep medicine certification board exams: ChatGPT versus Google Bard. Eur Arch Otorhinolaryngol. 2024;281:2137-2143. [Crossref]

45. Khan AA, Yunus R, Sohail M, Rehman TA, Saeed S, Bu Y, et al. Artificial Intelligence for anesthesiology board-style examination questions: Role of large language models. J Cardiothorac Vasc Anesth. 2024;38:1251-1259. [Crossref]

46. Islam G. Informatics approach to improving surgical skills training. Ann Arbor, MI: ProQuest LLC; 2013. [Crossref]

47. McFadden P, Crim A. Comparison of the effectiveness of interactive didactic lecture versus online simulation-based CME programs directed at improving the diagnostic capabilities of primary care practitioners. J Contin Educ Health Prof. 2016;36:32-37. [Crossref]

48. Samek W, Wiegand T, Müller KR. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv Preprint. 2017. [Crossref]

49. Barrows H, Tamblyn R. Problem-based learning: An approach to medical education. New York: Springer Publishing Company; Mar 15, 2025. [Crossref]

50. Ellaway RH, Tolsgaard M. Artificial scholarship: LLMs in health professions education research. Adv Health Sci Educ Theory Pract. 2023;28:659-664. [Crossref]

51. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. JMIR Med Educ. 2023;9:e48291. [Crossref]

52. Zumsteg JM, Junn C. Will ChatGPT Match to Your Program? Am J Phys Med Rehabil. 2023;102:545-547. [Crossref]

53. Chen HL, Chen HH. Have you chatted today?-Medical education surfing with artificial intelligence. Journal of Medical Education. 2023;27:1-4. [Crossref]

54. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus. 2023;15:e35179. [Crossref]

55. Arif TB, Munaf U, Ul-Haque I. The future of medical education and research: Is ChatGPT a blessing or blight in disguise? Med Educ Online. 2023;28:2181052. [Crossref]

56. Eysenbach G. The Role of ChatGPT, Generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. JMIR Med Educ. 2023;9:e46885. [Crossref]

57. Gandhi Periaysamy A, Satapathy P, Neyazi A, Padhi BK. ChatGPT: roles and boundaries of the new artificial intelligence tool in medical education and health research - correspondence. Ann Med Surg (Lond). 2023;85:1317-1318. [Crossref]

58. Karabacak M, Ozkara BB, Margetis K, Wintermark M, Bisdas S. The advent of generative language models in medical education. JMIR Med Educ. 2023;9:e48163. [Crossref]

59. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. Pak J Med Sci. 2023;39:605-607. [Crossref]

60. Masters K. Medical Teacher's first ChatGPT's referencing hallucinations: Lessons for editors, reviewers, and teachers. Med Teach. 2023;45:673-675. [Crossref]

61. Friederichs H, Friederichs WJ, März M. ChatGPT in medical school: how successful is AI in progress testing? Med Educ Online. 2023;28:2220920. [Crossref]

62. Pearce J, Chiavaroli N. Rethinking assessment in response to generative artificial intelligence. Med Educ. 2023;57:889-891. [Crossref]

63. Masters K. Response to: Aye, AI! ChatGPT passes multiple-choice family medicine exam. Med Teach. 2023;45:666. [Crossref]

64. Wang LK, Paidisetty PS, Cano AM. The next paradigm shift? ChatGPT, artificial intelligence, and medical education. Med Teach. 2023;45:925. [Crossref]