

Viewpoint

From Hype to Implementation: Embedding GPT-4o in Medical Education

Sumaia Sabouni¹, PhD; Mohammad-Adel Moufti^{2,3}, PhD; Mohamed Hassan Taha^{3,4}, PhD

¹Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Al-Ahliyya Amman University, Amman, Jordan

²College of Dental Medicine, University of Sharjah, Sharjah, United Arab Emirates

³Medical Education Center, University of Sharjah, Sharjah, United Arab Emirates

⁴College of Medicine, University of Sharjah, Sharjah, United Arab Emirates

Corresponding Author:

Mohammad-Adel Moufti, PhD
College of Dental Medicine
University of Sharjah
Sharjah 27272
United Arab Emirates
Phone: 971 65057314
Email: mamoufti@sharjah.ac.ae

Abstract

The release of GPT-4 Omni (GPT-4o), an advanced multimodal generative artificial intelligence (AI) model, generated substantial enthusiasm in the field of higher education. However, one year later, medical education continues to face significant challenges, demonstrating the need to move from initial experimentation with the integration of multimodal AIs in medical education toward meaningful integration. In this Viewpoint, we argue that GPT-4o's true value lies not in novelty, but in its potential to enhance training in communication skills, clinical reasoning, and procedural skills by offering real-time simulations and adaptive learning experiences using text, audio, and visual inputs in a safe, immersive, and cost-effective environment. We explore how this innovation has made it possible to address key medical educational challenges by simulating realistic patient interactions, offering personalized feedback, and reducing educator workloads and costs, where traditional teaching methods struggle to replicate the complexity and dynamism of real-world clinical scenarios. However, we also address the critical challenges of this approach, including data accuracy, bias, and ethical decision-making. Rather than seeing GPT-4o as a replacement, we propose its use as a strategic supplement, scaffolded into curriculum frameworks and evaluated through ongoing research. As the focus shifts from AI novelty to sustainable implementation, we call on educators, policymakers, and curriculum designers to establish governance mechanisms, pilot evaluation strategies, and develop faculty training. The future of AI in medical education depends not on the next breakthrough, but on how we integrate today's tools with intention and rigor.

JMIR Med Educ 2025;11:e79309; doi: [10.2196/79309](https://doi.org/10.2196/79309)

Keywords: multimodal generative pretraining transformers; GPT-4o; artificial intelligence; interactive learning; medical education

Introduction

The impact of artificial intelligence (AI) within health care continues to expand, ranging from diagnostic imaging to medical training [1]. Its integration into medical education offers promising solutions to numerous challenges, particularly the need for substantial resources to effectively train students in competencies like communication, clinical reasoning, and procedural skills. AI tools offer scalable, low-resource alternatives that can help bridge these gaps.

In 2024, OpenAI released GPT-4 Omni (GPT-4o), a multimodal model combining text, image, and audio processing. Unlike text-only systems, GPT-4o supports more dynamic interactions, such as engaging in real-time, human-like conversations, interpreting visual content, and performing realistic voice-based simulations. GPT-4o received widespread attention within medical education due to its multimodal capabilities enabling realistic simulations that extend to procedural skill assessment and interactive case analyses. Early demonstrations highlight GPT-4o's ability to simulate patient interactions [2], visually identify medications [3], and

undertake medical examinations with higher accuracy than GPT-4 and GPT-3.5 [4,5].

However, one year after its release, there is clearly a slow pace of adoption and implementation remains limited [6]. Although some institutions have experimented with GPT-4o integration [7] and others analyzed medical students' attitudes toward this integration [8], sustained adoption within formal curricula has lagged.

This Viewpoint highlights how GPT-4o shows promise as a relevant tool to enhance medical education and address many of its challenges, despite the initial hype fading. We examine its strengths in simulating patient interactions, increasing equity due to low-resource deployment, and reducing educator workload. We also address concerns including accuracy, bias, ethical reasoning, and digital access. Our aim is not to present GPT-4o as a cure-all, but to advocate for a structured, research-informed pathway for integrating such tools meaningfully into medical training. Although this discussion focuses on general medical education, specific considerations for dental medicine and other health disciplines warrant future exploration.

Educational Challenges Addressed by GPT-4o

Here are some of the educational challenges addressed by GPT-4o:

1. **Communication skills:** Developing interpersonal communication is vital for effective clinical practice. Traditional role-playing methods, reliant on patients or actors, are resource-intensive and challenging to standardize and scale. GPT-4o-generated virtual patients simulate diverse and adaptive interactions, fostering essential skills such as delivering bad news and providing culturally sensitive care. By creating realistic, immersive environments, GPT-4o aligns with situated learning theory, enabling learners to acquire practical skills in context.
2. **Clinical reasoning:** Diagnostic and intervention decision-making requires iterative exposure to varied scenarios. Static case-based learning often fails to engage learners dynamically. GPT-4o generates interactive, context-rich cases tailored to learner actions, promoting experiential learning [9] and hypothetico-deductive reasoning [10]. Feedback is immediate and adaptive, allowing users to refine diagnostic strategies in real time. These features advance cognitive load management, enhancing learning efficiency.
3. **Essay-based assessments:** Essay writing fosters critical thinking and provides a platform for expressing individual ideas. Despite these benefits, educators often avoid using essay writing due to the time required for grading. GPT-4o automates the assessment of essay responses, providing detailed feedback on reasoning, structure, and language [11]. This automation reduces educator workloads, encourages higher-order assessments, and supports personalized student development.

However, overreliance on automated grading may overlook nuanced clinical reasoning or professional reflection, underscoring the need for educator oversight [12]. Additionally, some research suggests that students may perceive AI-generated feedback as impersonal, leading to reduced engagement or trust [11].

4. **Procedural skills:** Skill acquisition in procedures relies on hands-on practice, detailed feedback, and in-person, one-on-one observation and assessment. GPT-4o's ability to produce realistic simulations and analyze video inputs offers scalable solutions for performance evaluation. Future integration with augmented reality and virtual reality technologies could facilitate immersive training experiences, replicating clinical environments for enhanced procedural competence. However, the concept of fidelity, including how realistic and believable simulations appear to learners, is critical. Low-fidelity simulations risk invoking the "uncanny valley" effect, where interactions feel unnatural and may reduce engagement [13].

Potential Benefits

For Learners

GPT-4o supports ongoing skill development by offering, personalized instant performance evaluation and feedback. Due to its realistic clinical simulations, students are exposed to a wide range of medical cases and can practice clinical reasoning in a realistic and practical way. Learners can practice breaking bad news to virtual patients, interpret multimodal data such as radiological images or heart sounds, and receive adaptive feedback tailored to their level. As a result, students can develop their confidence in a secure, flexible, and nonjudgmental learning environment.

For Educators

GPT-4o helps ease the workload by performing time-consuming tasks such as writing patient scenarios, setting assessments, marking essays, and providing personalized feedback on student performance. This gives educators more time to focus on tasks such as student support and course development. Additionally, seeing as GPT-4o is already pretrained, there is no need for heavy investment in custom AI tools, making it easier and quicker for institutions to adopt. Nevertheless, institutions should address faculty resistance by providing hands-on workshops, co-designing AI use cases with educators, and sharing evidence from pilot studies to build confidence in its implementation.

Global Accessibility

Due to GPT-4o being cloud-based, it does not require costly infrastructure, making it accessible to both learners and educators in low-resource settings. This helps democratize medical education and training and results in a more even distribution of health care expertise globally [14].

Challenges and Considerations

Although AI holds promise in medical education, there are also some challenges and considerations:

1. **Accuracy:** AI programs like GPT-4o may generate errors or “hallucinations” [15]. To overcome this, the output of such tools requires continuous supervision and validation to ensure their reliability.
2. **Data privacy:** To uphold confidentiality and ethical standards, learner performance data must be protected. This may be achieved through the adoption of strict data security measures by educational establishments. However, implementing strict data security is complex, requiring closed systems, robust encryption, and governance frameworks. There are trade-offs between using real learner data, which raises confidentiality concerns, and synthetic data, which may lack realism.
3. **Bias:** Training data biases associated with gender, ethnicity, or socioeconomic status may influence AI outputs. As a result, initiatives to reduce these biases are essential, especially since AI tools are being used more and more to deliver and filter health care information, where transparency and trust are essential [16]. Emerging frameworks like Fairlearn and AI Fairness 360 offer tools to audit and mitigate bias in AI systems, supporting fairer educational outcomes [17].
4. **Ethics and professionalism:** Although GPT-4o performs very well in simulations, professionalism or complex ethical reasoning may be more challenging for it [12]. These limitations highlight the importance of human oversight and complementary teaching methods.
5. **Access:** Despite its general affordability, access to GPT-4o still depends on a stable internet connection, potentially limiting its reach in underserved regions with poor connectivity [18]. Addressing this challenge requires exploring offline or low-bandwidth alternatives to ensure a broader reach.

Next Steps

Emerging empirical studies are beginning to validate GPT-4o’s role in medical education. For example, a pilot by

Öncü et al [2] used GPT-4o as a virtual standardized patient to simulate complex communication and crisis scenarios, demonstrating high learner engagement and feasibility in training clinical reasoning. Bicknell et al [19] evaluated GPT-4o across 4 national licensing exams, where it achieved a 90.4% accuracy rate, surpassing GPT-4 and average medical student performance. Similarly, Zhong et al [20] benchmarked GPT-4o for rare disease diagnosis using multilingual clinical data, with the model achieving the highest diagnostic accuracy among tested large language models. These studies reinforce the potential of GPT-4o to enhance assessment, simulation, and clinical education, while also underscoring the need for rigorous longitudinal evaluation frameworks.

Future research should assess how effectively GPT-4o enhances learner competencies like critical thinking, decision-making, and procedural skills. Evaluation frameworks should include multimodal fidelity measures, learner satisfaction, performance metrics, and longitudinal tracking of outcomes such as clinical preparedness. These metrics will help assess the educational impact of GPT-4o in realistic, complex learning environments. Additionally, GPT-4o could support interprofessional education by simulating collaborative scenarios between medical, nursing, and dental students. Such integration may enhance communication, teamwork, and understanding across health disciplines, particularly in complex patient care settings.

To further expand its uses, GPT’s multimodal capabilities could be extended to interpret video and audio data, such as echo scans and heart sounds. To guarantee that implementation is ethical, scalable, and in line with curriculum objectives, cooperation between AI developers, educators, and legislators is equally vital. Addressing challenges such as accuracy, bias, and accessibility challenges will ensure equitable benefits for all learners. Even though GPT-4o is a significant step forward, its use in medical education should be guided by research and specific learning aims. To offer useful assistance for adoption, we propose a set of initial steps for institutions to implement GPT-4o (Textbox 1).

Textbox 1. Recommendations for institutions considering GPT-4o integration.

1. Begin with a small-scale pilot project focusing on a single competency (eg, communication).
2. Provide prompt engineering training for faculty members.
3. Develop ethical guidelines for student use.
4. Incorporate artificial intelligence literacy into medical curricula.
5. Collaborate with technology experts to conduct longitudinal assessments of performance.
6. Design simulated scenarios that support interprofessional education (eg, medical, dental, and nursing collaboration).
7. Address faculty concerns through workshops, co-designed implementation strategies, and sharing evidence from early adopters.

Conclusion

The integration of GPT-4o into medical education represents a promising shift. It not only tackles long-standing challenges

by offering scalable, immersive training solutions, but its multimodal nature also results in it offering advantages over prior models in certain contexts. Consequently, both educators and learners globally stand to benefit as access to quality education becomes more equitable. However, for

this integration to be effective, medical institutions must move past experimentation through pilot studies and toward embedding into formal curricula. Finally, GPT-4o must not be viewed as a quick fix but rather as a tool that

demands thoughtful design, evaluation, and governance, and its integration should be guided by rigorous research, ethical considerations, and faculty collaboration.

Acknowledgments

This study was partially funded by a research grant from the University of Sharjah (grant number 2301100278).

The authors alone are responsible for the content and writing of the article. Generative artificial intelligence tools, specifically OpenAI's ChatGPT-4, were used to proofread and improve the clarity of the manuscript. All original ideas, analyses, and conclusions were solely developed by the authors.

Conflicts of Interest

None declared.

References

1. Abu Owida H, R. Hassan M, Ali AM, et al. The performance of artificial intelligence in prostate magnetic resonance imaging screening. *IJECE*. 2024;14(2):2234. [doi: [10.11591/ijece.v14i2.pp2234-2241](https://doi.org/10.11591/ijece.v14i2.pp2234-2241)]
2. Öncü S, Torun F, Ülkü HH. AI-powered standardised patients: evaluating ChatGPT-4o's impact on clinical case management in intern physicians. *BMC Med Educ*. Feb 20, 2025;25(1):278. [doi: [10.1186/s12909-025-06877-6](https://doi.org/10.1186/s12909-025-06877-6)] [Medline: [39979969](https://pubmed.ncbi.nlm.nih.gov/39979969/)]
3. Bazzari AH, Bazzari FH. Assessing the ability of GPT-4o to visually recognize medications and provide patient education. *Sci Rep*. Nov 5, 2024;14(1):26749. [doi: [10.1038/s41598-024-78577-y](https://doi.org/10.1038/s41598-024-78577-y)] [Medline: [39501020](https://pubmed.ncbi.nlm.nih.gov/39501020/)]
4. Liu CL, Ho CT, Wu TC. Custom GPTs enhancing performance and evidence compared with GPT-3.5, GPT-4, and GPT-4o? A study on the emergency medicine specialist examination. *Healthcare (Basel)*. Aug 30, 2024;12(17):1726. [doi: [10.3390/healthcare12171726](https://doi.org/10.3390/healthcare12171726)] [Medline: [39273750](https://pubmed.ncbi.nlm.nih.gov/39273750/)]
5. Wu YC, Wu YC, Chang YC, Yu CY, Wu CL, Sung WW. Advancing medical AI: GPT-4 and GPT-4o surpass GPT-3.5 in Taiwanese medical licensing exams. *PLoS ONE*. 2025;20(6):e0324841. [doi: [10.1371/journal.pone.0324841](https://doi.org/10.1371/journal.pone.0324841)] [Medline: [40465748](https://pubmed.ncbi.nlm.nih.gov/40465748/)]
6. Perkins M, Pregowska A. The role of artificial intelligence in higher medical education and the ethical challenges of its implementation. *AIH*. Oct 21, 2024;2(1):1. [doi: [10.36922/aih.3276](https://doi.org/10.36922/aih.3276)]
7. Thomae AV, Witt CM, Barth J. Integration of ChatGPT into a course for medical students: explorative study on teaching scenarios, students' perception, and applications. *JMIR Med Educ*. Aug 22, 2024;10:e50545. [doi: [10.2196/50545](https://doi.org/10.2196/50545)] [Medline: [39177012](https://pubmed.ncbi.nlm.nih.gov/39177012/)]
8. Duan S, Liu C, Rong T, Zhao Y, Liu B. Integrating AI in medical education: a comprehensive study of medical students' attitudes, concerns, and behavioral intentions. *BMC Med Educ*. Apr 23, 2025;25(1):599. [doi: [10.1186/s12909-025-07177-9](https://doi.org/10.1186/s12909-025-07177-9)] [Medline: [40269824](https://pubmed.ncbi.nlm.nih.gov/40269824/)]
9. Salinas-Navarro DE, Vilalta-Perdomo E, Michel-Villarreal R, Montesinos L. Designing experiential learning activities with generative artificial intelligence tools for authentic assessment. *ITSE*. Oct 30, 2024;21(4):708-734. [doi: [10.1108/ITSE-12-2023-0236](https://doi.org/10.1108/ITSE-12-2023-0236)]
10. Khumrin P. The use of clinical decision support systems for the development of medical students' diagnostic reasoning skills [Dissertation]. URL: <https://minerva-access.unimelb.edu.au/handle/11343/227067> [Accessed 2025-10-07]
11. Banihashem SK, Kerman NT, Noroozi O, Moon J, Drachsler H. Feedback sources in essay writing: peer-generated or AI-generated feedback? *Int J Educ Technol High Educ*. 2024;21(1). [doi: [10.1186/s41239-024-00455-4](https://doi.org/10.1186/s41239-024-00455-4)]
12. Masters K. Artificial intelligence in medical education. *Med Teach*. Sep 2019;41(9):976-980. [doi: [10.1080/0142159X.2019.1595557](https://doi.org/10.1080/0142159X.2019.1595557)] [Medline: [31007106](https://pubmed.ncbi.nlm.nih.gov/31007106/)]
13. Johnston JL, Kearney GP, Gormley GJ, Reid H. Into the uncanny valley: simulation versus simulacrum? *Med Educ*. Oct 2020;54(10):903-907. [doi: [10.1111/medu.14184](https://doi.org/10.1111/medu.14184)] [Medline: [32314435](https://pubmed.ncbi.nlm.nih.gov/32314435/)]
14. Hamilton A. Artificial intelligence and healthcare simulation: the shifting landscape of medical education. *Cureus*. May 2024;16(5):e59747. [doi: [10.7759/cureus.59747](https://doi.org/10.7759/cureus.59747)] [Medline: [38840993](https://pubmed.ncbi.nlm.nih.gov/38840993/)]
15. Masters K. *Medical Teacher's* first ChatGPT's referencing hallucinations: lessons for editors, reviewers, and teachers. *Med Teach*. Jul 2023;45(7):673-675. [doi: [10.1080/0142159X.2023.2208731](https://doi.org/10.1080/0142159X.2023.2208731)] [Medline: [37183932](https://pubmed.ncbi.nlm.nih.gov/37183932/)]
16. Shambour QY, Abualhaj MM, Abu-Shareha A, Hussein AH, Kharma QM. Mitigating healthcare information overload: a trust-aware multi-criteria collaborative filtering model. *J Appl Data Sci*. 2024;5(3):1134-1146. [doi: [10.47738/jads.v5i3.297](https://doi.org/10.47738/jads.v5i3.297)]
17. Bellamy RKE, Dey K, Hind M, et al. AI Fairness 360: an extensible toolkit for detecting and mitigating algorithmic bias. *IBM J Res & Dev*. Jul 1, 2019;63(4/5):4. [doi: [10.1147/JRD.2019.2942287](https://doi.org/10.1147/JRD.2019.2942287)]

18. Abd-Alrazaq A, AlSaad R, Alhuwail D, et al. Large language models in medical education: opportunities, challenges, and future directions. *JMIR Med Educ*. Jun 1, 2023;9:e48291. [doi: [10.2196/48291](https://doi.org/10.2196/48291)] [Medline: [37261894](https://pubmed.ncbi.nlm.nih.gov/37261894/)]
19. Bicknell BT, Butler D, Whalen S, et al. ChatGPT-4 Omni performance in USMLE disciplines and clinical skills: comparative analysis. *JMIR Med Educ*. Nov 6, 2024;10:e63430. [doi: [10.2196/63430](https://doi.org/10.2196/63430)] [Medline: [39504445](https://pubmed.ncbi.nlm.nih.gov/39504445/)]
20. Zhong W, Liu Y, Liu Y, et al. Performance of ChatGPT-4o and four open-source large language models in generating diagnoses based on China's rare disease catalog: comparative study. *J Med Internet Res*. Jun 18, 2025;27:e69929. [doi: [10.2196/69929](https://doi.org/10.2196/69929)] [Medline: [40532199](https://pubmed.ncbi.nlm.nih.gov/40532199/)]

Abbreviations

AI: artificial intelligence

GPT-4o: GPT-4 Omni

Edited by Joshua Gentges; peer-reviewed by Claire Inglis, Mohamed El-Kishawi; submitted 18.06.2025; final revised version received 06.08.2025; accepted 10.09.2025; published 15.10.2025

Please cite as:

Sabouni S, Moufti MA, Taha MH

From Hype to Implementation: Embedding GPT-4o in Medical Education

JMIR Med Educ 2025;11:e79309

URL: <https://mededu.jmir.org/2025/1/e79309>

doi: [10.2196/79309](https://doi.org/10.2196/79309)

© Sumaia Sabouni, Mohammad-Adel Moufti, Mohamed Hassan Taha. Originally published in *JMIR Medical Education* (<https://mededu.jmir.org>), 15.10.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Medical Education*, is properly cited. The complete bibliographic information, a link to the original publication on <https://mededu.jmir.org/>, as well as this copyright and license information must be included.