Canadian Medical Education Journal

Commentary and Opinions

Generative artificial intelligence in medical education: moving from potential to practice: when AI surpasses MDs

L'intelligence artificielle générative dans l'enseignement médical : du potentiel à la pratique : quand l'IA dépassera les médecins

Leo Morjaria,[1] Matthew Sibbald[1,2]

[1]Michael G. DeGroote School of Medicine, McMaster University, Ontario, Canada; [2]McMaster Education Research, Innovation and Theory (MERIT) Program, McMaster University, Ontario, Canada

Correspondence to: Leo Morjaria, 27 Bold Street, Hamilton Ontario; email: Leo.morjaria@medportal.ca

Published ahead of issue: May 20, 2025; published: Dec 22, 2025. CMEJ 2025, 16(6) Available at https://doi.org/10.36834/cmej.80607

Since the initial release of ChatGPT in 2022, promising early work has discussed the potential impact of generative AI on medical education. Despite this, over the course of the last two years, there has been limited real-world adoption of generative AI related to medical education curricula. This has been in part due to the limitations inherent in these tools which have led educators to exercise caution.[1]

However, we believe that recent advancements in generative AI warrant immediate reconsideration of both its role and speed of its institutional adoption in healthcare education. On September 12th, 2024, OpenAI released ChatGPT-o1, a new large language model (LLM) with the first commercial implementation of "chain-of-thought" reasoning.[2] By tuning the model to generate intermediate reasoning steps when solving complex problems, OpenAI significantly improved both the performance and reliability of their flagship LLM. Since then, several other models with comparable or even superior reasoning capabilities have been released, including DeepSeek-R1 and Claude 3.[3] These models signal a broader shift in LLM capabilities– from simple responders to capable reasoners. Preliminary testing conducted by OpenAI suggests that when tested on the Graduate-Level Google-Proof Q&A Benchmark (GPQA) PhD-level questions, ChatGPT-o1 is the first LLM able to consistently outperform human experts with PhDs.[2] Since its release, professionals in the fields of medicine, computing, and mathematics have stated that ChatGPT-o1 is able to perform at the level of a competent graduate student, and is able to accomplish many tasks that were previously impossible with generative AI.[4] However, it is important to note that these preliminary findings are based on internal evaluations and should be interpreted with caution.

Given the rapid pace of technological advancement, we believe that relying solely on exploratory research is no longer sufficient. Rapid cycle experimentation, facilitated by the establishment of sandbox environments, is urgently needed. However, it is important that in high stakes environments, such as board exams, sandbox use should be clearly framed as experimental and supplementary, not substitutive. These experimental environments can then allow both learners and educators to test these new tools in a controlled setting, encouraging an efficient cycle of iterative experimentation and feedback. Successful integrations could then help establish better practices and further inform curriculum design. We see three primary medical education implementations of generative AI that merit further study and pilot testing:

1. Automating Repetitive Tasks: By using generative AI to streamline time consuming tasks, educators will be able to reallocate their focus to the human-centered aspects of teaching and curriculum design. Examples of high time burden tasks that could be automated include question generation, grading, and providing feedback.

2. Expanding Learning Opportunities: By harnessing AI, educators can create new and novel educational experiences. Examples of such experiences include individualized testing environments, and dynamic learning scenarios that adapt to student performance and provide feedback in real-time.

3. Preparing Students for an AI-Enabled World: Many corporations, including some of the largest Electronic Health Record providers, such as Epic, have already begun investing large amounts of capital to create AI powered healthcare tools.[5] As the private sector begins to implement AI in the clinical setting, it is important that medical education is adapted to mirror this change. Curricula should include opportunities for students to learn how to effectively use AI tools, critically evaluate AI generated content and understand the inherent risks and ethical implications of AI use in medicine.

While these new reasoning LLMs show promise, implementation of any type of generative AI is not without risk. In the context of education, major concerns include hallucinations, data privacy, overreliance on AI generated outputs, as well as a lack of a standardized regulatory framework. With varying levels of institutional readiness and digital literacy, it is paramount that educators and technologists adopt a cautious and thoughtful approach to integrating these technologies.

Ultimately, we believe that the advent of reasoning-capable LLMs represents a significant development for medical education. Previous critiques and failures of generative AI must be revisited. Addressing the current gap in literature surrounding generative AI's application in real world medical education settings is paramount. Although we advocate moving beyond theoretical discussions and exploratory research towards practical integration, the uncertainties surrounding data management and privacy should be carefully addressed, and a robust framework for accountability must be established. In this way, we can cautiously harness AI technology to improve healthcare education delivery and better prepare learners for an AI-enabled future. We encourage both enthusiasts and skeptics to set aside their assumptions and experiment and report on the integration of these new AI models in medical education.

Conflicts of Interest: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding: The authors received no financial support for the research, authorship, and/or publication of this article.
Edited by: Marcel D'Eon (editor-in-chief)

References

1. ChatGPT: bullshit spewer or the end of traditional assessments in higher education*? J Appl Learn Teach*. 2023 Jan 25; 6(1). Available from: https://journals.sfu.ca/jalt/index.php/jalt/article/view/689. [Accessed on Sep 4, 2024].
2. OpenAI. *Learning to reason with LLMs*. OpenAI. 2024. Available from: https://openai.com/index/learning-to-reason-with-llms/ [Accessed on Sep 12, 2024].
3. DeepSeek-AI, Guo D, Yang D, et al. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. *arXiv*; 2025 Available from: https://arxiv.org/abs/2501.12948. [Accessed Apr 2, 2025]
4. Jones N. 'In awe': scientists impressed by latest ChatGPT model o1. *Nature.* 2024 Oct 10;634(8033):275–6.. https://doi.org/10.1038/d41586-024-03169-9
5. Jennings K. *Why $4.6 Billion health records giant epic is betting big on generative AI.* Forbes. 2024 Mar 18; Available from: https://www.forbes.com/sites/katiejennings/2024/03/18/why-46-billion-health-records-giant-epic-is-betting-big-on-generative-ai/