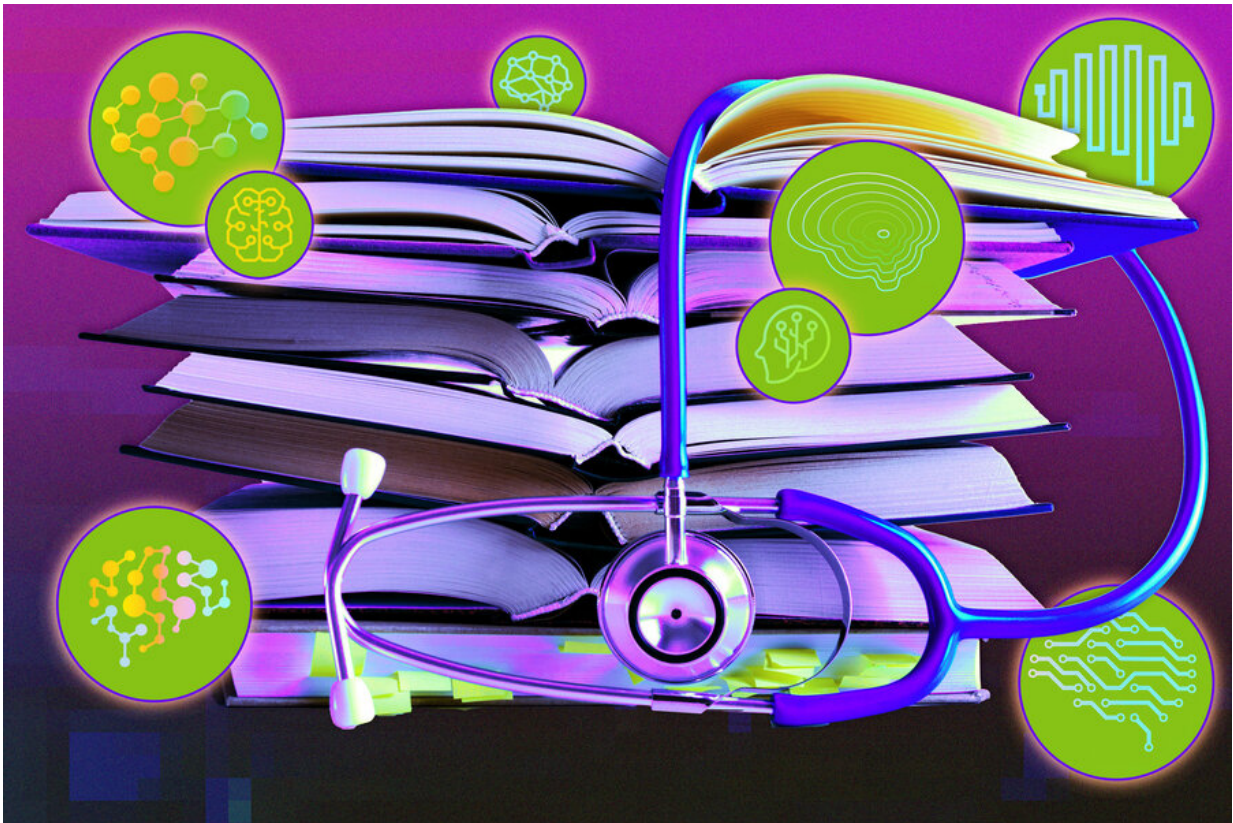


Q&A: Three questions on ChatGPT and medicine

February 9 2023, by Anne Trafton



The successful performance of ChatGPT on the U.S. Medical Licensing Exam demonstrates shortcomings in how medical students are trained and evaluated, says Leo Anthony Celi, a principal research scientist at MIT's Institute for Medical Engineering and Science and a practicing physician. Credit: Jose-Luis Olivares, MIT

Launched in November 2022, ChatGPT is a chatbot that can not only engage in human-like conversation, but also provide accurate answers to questions in a wide range of knowledge domains. The chatbot, created by the firm OpenAI, is based on a family of "large language models"—algorithms that can recognize, predict, and generate text based on patterns they identify in datasets containing hundreds of millions of words.

In a [study](#) appearing in *PLOS Digital Health* this week, researchers report that ChatGPT performed at or near the passing threshold of the U.S. Medical Licensing Exam (USMLE)—a comprehensive, three-part exam that doctors must pass before practicing medicine in the United States.

In an editorial accompanying the paper, Leo Anthony Celi, a principal research scientist at MIT's Institute for Medical Engineering and Science, a practicing physician at Beth Israel Deaconess Medical Center, and an associate professor at Harvard Medical School, and his co-authors argue that ChatGPT's success on this exam should be a wake-up call for the medical community.

Q: What do you think the success of ChatGPT on the USMLE reveals about the nature of the medical education and evaluation of students?

A: The framing of [medical knowledge](#) as something that can be encapsulated into multiple choice questions creates a cognitive framing of false certainty. Medical knowledge is often taught as fixed model representations of health and disease. Treatment effects are presented as stable over time despite constantly changing practice patterns. Mechanistic models are passed on from teachers to students with little emphasis on how robustly those models were derived, the uncertainties that persist around them, and how they must be recalibrated to reflect

advances worthy of incorporation into practice.

ChatGPT passed an examination that rewards memorizing the components of a system rather than analyzing how it works, how it fails, how it was created, how it is maintained. Its success demonstrates some of the shortcomings in how we train and evaluate medical students. Critical thinking requires appreciation that ground truths in medicine continually shift, and more importantly, an understanding how and why they shift.

Q: What steps do you think the medical community should take to modify how students are taught and evaluated?

A: Learning is about leveraging the current body of knowledge, understanding its gaps, and seeking to fill those gaps. It requires being comfortable with and being able to probe the uncertainties. We fail as teachers by not teaching students how to understand the gaps in the current body of knowledge. We fail them when we preach certainty over curiosity, and hubris over humility.

Medical education also requires being aware of the biases in the way medical knowledge is created and validated. These biases are best addressed by optimizing the cognitive diversity within the community. More than ever, there is a need to inspire cross-disciplinary collaborative learning and problem-solving. Medical students need data science skills that will allow every clinician to contribute to, continually assess, and recalibrate medical knowledge.

Q: Do you see any upside to ChatGPT's success in this exam? Are there beneficial ways that ChatGPT and other forms of AI can contribute to the practice of medicine?

A: There is no question that large language models (LLMs) such as ChatGPT are very powerful tools in sifting through content beyond the capabilities of experts, or even groups of experts, and extracting knowledge. However, we will need to address the problem of data bias before we can leverage LLMs and other artificial intelligence technologies. The body of knowledge that LLMs train on, both medical and beyond, is dominated by content and research from well-funded institutions in high-income countries. It is not representative of most of the world.

We have also learned that even mechanistic models of health and disease may be biased. These inputs are fed to encoders and transformers that are oblivious to these biases. Ground truths in medicine are continuously shifting, and currently, there is no way to determine when ground truths have drifted. LLMs do not evaluate the quality and the bias of the content they are being trained on. Neither do they provide the level of uncertainty around their output. But the perfect should not be the enemy of the good. There is tremendous opportunity to improve the way health care providers currently make clinical decisions, which we know are tainted with unconscious bias. I have no doubt AI will deliver its promise once we have optimized the data input.

More information: Amarachi B. Mbakwe et al, ChatGPT passing USMLE shines a spotlight on the flaws of medical education, *PLOS Digital Health* (2023). [DOI: 10.1371/journal.pdig.0000205](https://doi.org/10.1371/journal.pdig.0000205)

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Q&A: Three questions on ChatGPT and medicine (2023, February 9) retrieved 17 July 2024 from <https://techxplore.com/news/2023-02-qa-chatgpt-medicine.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.