Academic Psychiatry
https://doi.org/10.1007/s40596-024-01996-6

COMMENTARY

# Large Language Models and Artificial Intelligence in Psychiatry Medical Education: Augmenting But Not Replacing Best Practices

John Torous[1] · William Greenberg[1]

Received: 17 February 2024 / Accepted: 5 June 2024


> "From three to eight years we will have a machine with the general intelligence of an average human being."—Marvin Minsky, Ph.D.

In 2023, large language models epitomized by the prominence of ChatGPT raised both interest and fear about the future of medical student and resident education in general and more specifically in psychiatry. Reports that these technologies could pass medical exams, draft clinical notes, and even converse with patients promoted a wave of responses from the medical education community [1–3]. However, considering the potential of large language models from a more technical perspective offers a less discussed perspective that reveals the more likely and more positive implications for educators.

Large language models are not magical. For years we have all seen spell-check programs advance in their recommendations and email programs offering to complete sentences or even draft brief responses. Advances in technology, pioneered by Google's 2017 research into transformers [4] (the "T" in ChatGPT), allowed computers to scale up the already utilized technology behind search-engine autocomplete functions to vastly larger and more complex tasks. Tasks like autocompletion of an email phrase have now advanced to that of a clinical note.

Understanding how these models work is relevant to understanding their role in medical education. The technology underlying large language models is based on pattern matching. Typing "Mary had a little" into a search engine will result in the search engine offering the next word "lamb." The models responsible for this selection of the next word are not sentient but instead based on showing the model millions of examples of what word comes next when a user types "Mary had a little." Perhaps Mary is a medical student who had a little DSM-5 case series book. No large language model will offer this response as this sentence has likely never been shown to such a model. The key is that pattern matching is only as good as the patterns that the model has been trained on. This pattern matching powers today's large language models. These models do not "think" or "reason."

Building off this knowledge of large language models, the implications for psychiatry education become clearer. Rather than suggesting a future where psychiatrists become mere operators of these models and the prospects for the field are bleak, instead, trainees may spend less time writing notes and be able to devote more time to focus on comprehensive and improved interviewing, more advanced training on clinical foundations and differential diagnosis, and more expertise in interpreting results and honing treatment plans. Below, an outline of each point explains why psychiatry education will become more, not less, valuable in this era of large language models.

Trainees will need improved interviewing skills. Building off the "garbage-in, garbage-out" adage in computer science, any model including large language models is only as good as the data it is given. More specifically, if a psychiatrist seeks to use these models to help formulate a differential diagnosis, the output will only be as good as the information about the case provided. A careful and comprehensive interview and exam will provide the optimal input that a large language model requires to return a more thorough and focused result. Given that psychiatry remains a field without any validated biomarkers of illness and where illness interacts with cultural, developmental, and social factors to create a myriad of diverse presentations for even the same condition, the importance of accurate and precise information gathered will be more important than in any other field of healthcare. Large language models do not change the need for expert examination, but only amplify the need for ever more skilled capture of data.

Likewise, trainees will need more advanced training on clinical foundations. As new models seek to match inputs about exam and symptom findings into diagnosis and

✉ John Torous
jtorous@bidmc.harvard.edu

[1] Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

Published online: 06 August 2024

treatment plans, psychiatrists need to be trained to critically assess these "outputs." Are they valid? While research remains nascent in psychiatry, reports from oncology note that when ChatGPT 3.5 was asked to provide cancer treatment recommendations, it was most likely to mix incorrect recommendations with correct ones, making errors difficult to detect even for experts [5]. Given that many of the data sources for training these pattern matching programs are currently derived from internet social media sites (e.g., Reddit, Wikipedia, etc.), it is clear that the models will have errors and bias. In December 2023, one of the largest data sets powering AI images was removed from the internet after it was found to be offering images (and thus patterns to match) of child sexual abuse [6]. Sadly, the harm caused by improper training data sets was already witnessed in 2022 when researchers found that entering the word "schizophrenia" in a large language model AI image program returned only photos of horror movie genre clowns [7]. While this case of harm is obvious for all to see, it underlies the same challenge of subtle harm, errors, and bias already embedded into models. Only with advanced training on clinical foundation will psychiatrists be able to identify and correct that harm. Large language models do not change the need for psychiatric expertise, but only amplify the need for even more advanced knowledge and capacity for critical review.

Finally, trainees will now require more expertise in interpreting and communicating results and treatment plans to patients. The output of large language models, whether a list of differential diagnoses or a draft of a progress note, will need to be carefully reviewed and edited. Beyond assessing the validating of the outputs as discussed above, trainees need to develop the skills to select the right information and deliver it in a clinically and culturally appropriate manner to the patient. Unlike ChatGPT passing a medical test in which each question has only one single correct answer, the actual application of large language models will be in care situations in which there are numerous correct choices. This mirrors the need to increase emphasis in curricula from information acquisition to knowledge management and communication which others have already suggested [8]. Fortunately, the move to problem-based learning in response to the constant access to information from the internet has already prepared medical educators for training in this area. Large language models do not change the need for similar competencies and skills, but only amplify the need for mastery.

Each of these areas of change in medical education will require the careful attention of students, residents, and faculty. One can imagine major changes to the current ACGME Milestones [9] for psychiatry to both reflect and mandate these changes. For example, the Milestones for Patient Care 1: Psychiatric Evaluation and Patient Care 2: Psychiatric Formulation and Differential Diagnosis will need to include upper-level competencies in reviewing AI-generated patient notes to identify and address where "garbage in" (inadequate data from the patient interview) led to "garbage out" (misleading or inadequate psychiatric formulation and differential diagnosis). It might also include a higher-level competency in which residents master interviewing and perhaps even train against a standardized library of AI simulations. A new competency will likely need to be developed with its own milestones for treatment planning that addresses discussions with patients about AI-generated differential diagnosis and treatment options and their relative strengths and limitations. Where today, patients who read their medical records may want to talk with their physicians about what a particular test result might reflect or why a particular diagnostic term was employed, with AI-generated notes, our patients will want to know how their physicians distinguish between meaningful AI-generated diagnoses and treatment recommendations and "garbage" outputs in these areas.

The assistance that AI tools can bring to psychiatry could not come at a better time. As the incidence of mental illness continues to rise, especially in youth, the need for innovation in care delivery is patent. While it is not fair to ascribe these rising rates of illness to the impact of technology, the irony of therapy chatbots trained on pattern matching from internet data sources that have potentially already caused harm highlights the red-herring that AI-driven care represents. The 2023 example of the Tessa-chatbot for eating disorders where the program caused harm to people seeking help for this condition [10] underscores the challenges this technology faces in the simple psychoeducation space. While AI-powered bots may one day be able to offer wellness-focused help to people, such a primary prevention approach should be welcomed by the field and will be a necessary tool for medical students and residents to learn about. However wellness-based tools do not abnegate the need for a trained and skilled psychiatric workforce.

The quote at the beginning of this article, from 1970, by Marvin Minsky PhD of MIT to Life Magazine illustrates that while the potential of AI is tremendous, transformation is gradual [11]. Medical educators must prepare for the role of AI in education but do not need to panic. There have been prior AI hype and bust cycles as the above quote suggests. But assuming the current boom is real, there is little for psychiatry educators to fear as the need for high-quality teaching will be even more necessary. Just as the personal computer, then the internet, and finally, the smartphone did not alter the fundamental need for superior psychiatry training, neither will AI.

**Data Availability** No datasets were generated or analyzed during the current study.

## Declarations

**Disclosures** Dr. Greenberg is an editorial board member for the journal but did not participate in the peer review process other than as an author.

## References


1. Pak TK, Montelongo Hernandez CE, Do CN. Artificial intelligence in psychiatry: threat or blessing? Acad Psychiatry. 2023;47(6):587–8.
2. Cooper A, Rodman A. AI and medical education—a 21st-century Pandora's box. N Engl J Med. 2023;389(5):385–7.
3. Chang BS. Transformation of undergraduate medical education in 2023. JAMA. 2023;330(16):1521–2.
4. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Advances in neural information processing systems. 2017;30. https://dl.acm.org/doi/10.5555/3295222.3295349
5. Chen S, Kann BH, Foote MB, Aerts HJWL, Savova GK, Mak RH, et al. Use of artificial intelligence chatbots for cancer treatment information. JAMA Oncol. 2023;9(10):1459–62.
6. Thiel D. Identifying and Eliminating CSAM in Generative ML Training data and models. Technical report, Stanford University, Palo Alto, CA, 2023. https://purl.stanford.edu/kh752sm9123. Accessed 20 Dec 2023.
7. King M. Harmful biases in artificial intelligence. Lancet Psychiatry. 2022;9(11): e48.
8. Wartman SA, Combs CD. Reimagining medical education in the age of AI. AMA J Ethics. 2019;21(2):E146-152.
9. Accreditation Council for Graduate Medical Education. Psychiatry Milestones. https://www.acgme.org/globalassets/pdfs/milestones/psychiatrymilestones.pdf. Accessed 20 Dec 2023.
10. Sharp G, Torous J, West ML. Ethical challenges in AI approaches to eating disorders. J Med Internet Res. 2023;25(1): e50696.
11. Heaven WD. Artificial general intelligence: are we close, and does it even make sense to try? MIT Technol Rev. 2020. https://www.technologyreview.com/2020/10/15/1010461/artificial-general-intelligence-robots-ai-agi-deepmind-google-openai/