Academic Psychiatry
https://doi.org/10.1007/s40596-024-01972-0

COMMENTARY

From Bytes to Insights: The Promise and Peril of Artificial Intelligence–Powered Psychiatry

Alex Luna[1,2] · Steven Hyler[1]

Received: 25 September 2023 / Accepted: 16 April 2024

As of September 2023, ChatGPT has been reported to have passed multiple different board exams. ChatGPT 4.0 has passed the SAT, scored in the 90th percentile of the Uniform Bar Exam, and achieved a near passing grade on the United States Medical Licensing Exam (USMLE) [1, 2]. Such achievements have not gone unnoticed, and ChatGPT has sparked a plethora of discussion related to what impact it will have on various domains. In the case of psychiatry, the question is not a matter of if artificial intelligence (AI) tools will impact the field but, rather, a matter of how and when. In this commentary, we aim to discuss potential benefits and pitfalls of the eventual use of AI tools in psychiatry.

The path to a career in psychiatry is marked by several distinct milestones, which include passing the USMLE Step exams and the psychiatry board exam. ChatGPT has already been tested in multiple board exams across many different specialties, including radiology [3] and gastrointestinal medicine [4], with varying, but surprisingly impressive, levels of performance. Although no studies have yet examined ChatGPT's performance on the psychiatry board exam, its performance on yet another standardized exam would likely result in similar outcomes. The question that stands before psychiatry now, as a field, is how should ChatGPT and other future AI tools be implemented and what are the implications behind its use? Beginning to address this question requires a basic understanding of how ChatGPT and other similar AI tools are developed.

Large language models and similar AI approaches utilize advanced machine learning, natural language processing, neural networks, and deep learning models to generate predictions. The "GPT" in ChatGPT stands for Generative Pre-Trained Transformer, referring to its ability to generate text, inherent training process, and ability to identify relationships in a sequential manner. As with all advanced data science techniques in this domain, the models must undergo a training process. In much the same way as physicians utilize board preparation materials to practice and eventually take their standardized tests, the model is provided with a dataset to begin practicing making predictions, with its performance ultimately being evaluated in an independent test set. Once the parameters of the model are optimized and its performance is deemed acceptable, it can begin to be utilized for its intended purpose. In the case of ChatGPT, its training data was based on a subset of the internet and ultimately optimized to produce responses in a chat format.

AI in Medical Education

ChatGPT and similar advanced data analysis and AI tools, if effectively trained and optimized, have the potential for a revolutionary impact on many domains within medicine and even medical education. Leadership at some institutions has already begun to consider the use of AI to address the ever-increasing number of applicants to medical school and residency programs. Furthermore, changes in the structure of the medical admissions process, such as the change of board exams to pass/fail, have also added to the strain of medical school and residency application processes. An AI approach could potentially optimize such processes while simultaneously accounting for numerous different factors that encompass an application. Indeed, one such algorithm created as a proof-of-concept work was able to utilize over 60 elements of the Electronic Residency Application Service to accurately predict which applicants would receive an invitation to interview. More notably, it was also implemented during an actual interview cycle to identify 20 applicants who would have been otherwise overlooked using the typical admissions process [5].

✉ Alex Luna
Alex.luna@nyspi.columbia.edu

[1] New York Presbyterian Columbia University Irving Medical Center, New York, NY, USA

[2] New York State Psychiatric Institute, New York, NY, USA

Published online: 26 April 2024

Others have postulated that AI tools could help with the development of empathy among medical students, a crucial support, given reports of increasing compassion fatigue and burnout among medical students and residents that can ultimately affect patient care [6]. Such a feat would likely occur by offloading some of the clinical tasks students and residents are overworked with today, such as scheduling, note-writing, and chart reviewing/summarization. The use of AI can allow for a greater focus on the empathic components of the patient-doctor relationship, ironically leading to a situation where "AI allows doctors to be more human" [7]. Interestingly, a recent study seemed to indicate that ChatGPT was able to provide greater empathy than physicians to patient questions [8].

The evaluation of residents and medical students may also benefit from the incorporation of AI. One such example of how a large-language model such as ChatGPT could play a role in medical education was recently explored in the literature. In an interview with ChatGPT, it was able to simulate a patient with undiagnosed diabetes and responses the patient may make during a mock interview. Similarly, it was able to create an educational curriculum for training physicians on understanding the role of AI in health care [9]. Currently, evaluation of residents and students alike are limited by patient case availability, faculty limitations, and the inherent difficulty of curriculum revision. A well-validated series of psychiatric case summaries created by faculty, in combination with speech-to-text software and an interactive chatbot, such as ChatGPT, could be used to develop programs to assess clinical summaries and psychiatric formulations of standardized patients [10]. Residents could potentially work with simulations of patients demonstrating various psychiatric disorders and be evaluated on their interactions with such patients.

AI in Clinical Practice

The clinical domain of psychiatry would also stand to benefit from an increased use of AI tools. Multiple "artificial care providers" using machine learning methods and similar chatbot approaches have been used to provide and teach basic cognitive behavior therapy, dialectical behavior therapy, and motivational interviewing principles in substance use disorders and depression through the use of phone apps [11, 12]. ReachVet, an AI algorithm implemented by the National Institute of Mental Health, has been explored for its use in suicide prevention among veterans. Thus far, it has demonstrated improvement in outpatient appointments, safety plan creation, and reduction in inpatient admissions [13]. AI applications directed toward electrocardiogram interpretation have resulted in algorithms able to readily detect cardiac arrhythmias, at times even outperforming cardiologists [14, 15]. Given the known increased rates of comorbid cardiovascular dysfunction in patients with psychiatric disorders, as well as the known impact on QTc prolongation by antipsychotic medications, such tools can lead to improved cardiovascular risk reduction in psychiatric patients.

The role of AI will likely also extend beyond screening, diagnosis, and treatment. A study in 2013 determined that, on average, interns have only 12% of their time allocated to direct patient care, with the rest of their time divided among documentation, education, and obtaining collateral from other health care professionals [16]. Various AI tools currently in development are aimed at addressing this very issue. Through the combined use of dictation software and natural language processing, AI can produce documentation based on the conversation between a health care professional and a patient, resulting in less time spent on onerous paperwork, while also ideally emphasizing more evidence-based metrics, such as reminding health care professionals to implement scales such as the Patient Health Questionnaire to assess depression or the Generalized Anxiety Disorder 7 scale for anxiety. Similarly, ChatGPT has been found to be able to produce discharge summaries in a matter of seconds, as well as to write structured medical notes and patient clinic letters, all examples of how AI may reduce the administrative burden experienced by medical residents across all specialties [17–19].

Pitfalls of AI

Despite the potential benefits of ChatGPT and its AI cousins, the drawbacks to the use of AI in psychiatry may prove to be just as significant. While the use of a customizable training dataset to train an AI model is its greatest feature, it may also be its weakest point. Should the data provided to an AI algorithm not be sufficiently generalizable, it may have decreased accuracy and poor-quality output when implemented in real clinical situations, leading to dire consequences. For example, the National Eating Disorder Association recently attempted to implement a chatbot service to address deficiencies in its helpline, resulting in multiple dismissals of their human counterparts in favor of a chatbot. Subsequent interactions by patients seeking help resulted in the chatbot providing dieting advice and weight loss suggestions to individuals with eating disorders, prompting the removal of the chatbot from the organization [20]. A paper published in 2017 that purportedly used AI to identify suicide risk with 91% accuracy using only fMRI images was retracted after attempts to replicate the findings in the original paper failed. Yet, as a result of the study's impressive findings, clinical trials had been started, and the study had been cited 134 times within the first 3 years of publication [21].

Another disadvantage is perhaps the assumption that AI systems, due to their mechanical and allegedly objective nature, will not be subject to bias as humans are. Unfortunately, such an assumption would be a mistake. Using commercial algorithms implemented in the US health care system to determine health care needs, one study noted that Black patients placed in the same risk level as White patients through the algorithms were found to be much more ill than their White counterparts [22]. In doing so, Black patients that would have normally qualified for increased services did not receive them. Another study, using only X-rays of the wrist from a large public dataset, demonstrated that an AI algorithm could identify the race of an individual using only clinical imaging, even when "corrupted, cropped, and noised," putting into question whether such computer algorithms can truly be "race-blind" [23].

The ethics of privacy and the question of security with AI tools remains ever present, despite the novelty of ChatGPT. AI tools, as discussed earlier, rely on a large repository of data to make predictions successfully. The content of this data can range from basic demographics information to sensitive diagnostic information. In the realm of psychiatry, the nature of this information becomes even more sensitive, at times delving into more personal matters, such as trauma, abuse, and interpersonal relationships. In order to maintain psychiatric patients' autonomy, the use of AI must come with the patients' right to know what information was used, the risks of the use of AI, and the patients' right to refuse the use of AI in their care [24]. In addition, should these tools be accepted for implementation, the question of who owns the data collected by such AI tools will need to be explored and will likely require an interdisciplinary approach to ensure that transparency and accountability are maintained [25]. Yet, as of the writing of this commentary, Open AI has not disclosed the contents of its training data used for the development of ChatGPT4, a concerning development, given the ethical implications of AI in health care [26]. Furthermore, for AI to be used effectively, psychiatric patients' data must be safeguarded against the risk of data breaches, an especially notable barrier toward the implementation of AI in health care in light of OpenAI's first data breach in March 2023 [27].

The ability of ChatGPT to write human-like text naturally lends itself to other domains, which comes with many complex ethical questions. Regarding medical admissions, if a medical student applicant, whose first language is not English and who would be an excellent fit for the medical community, is discovered to have used ChatGPT to refine the application essay, should the student be denied admission? If a faculty member creates a letter of recommendation for a stellar resident for medical fellowship using ChatGPT, should that faculty member's contribution be summarily dismissed? Is a resident's work on a manuscript, created with ChatGPT but whose results are deemed to be 100% accurate and of high impact, rendered invalid because of the use of an AI tool?

From a technical standpoint, a full-scale ban may prove fruitless, given the limited efficacy by current tools to detect AI-generated material, and efforts to identify who used AI tools may prove to be less valuable than anticipated. Others have rightfully argued for assessing the situation from a more pragmatic perspective: if AI-assisted works are able to bypass the scrutiny of school admissions and peer review, perhaps the area of improvement lies not in finding ways to ban AI but, rather, in examining the way students and scientific research are assessed [28].

Future Directions

There is no doubt that ChatGPT and other similar advanced AI tools will leave an indelible impact on society as a whole and psychiatry more specifically. Discussions are already underway as to its use in medical education, admissions processes, and clinical decision-making. Labor disruption potential notwithstanding, ChatGPT and its relatives also hold the potential for causing significant damage if employed too hastily or without adequate safeguards and could result in direct patient harm and the perpetuation of existing racial inequalities in patient care and school admissions. Some organizations, such as the American Medical Association [29] and the American Psychiatric Association [30], have conceptualized these new tools as forms of augmenting, rather than replacing, clinician's efforts and have established principles to encourage transparency and oversight.

Similarly, extensive discussions should be held between psychiatrists and their patients on the patients' level of comfort with AI-driven clinical decision-making, especially given the crucial nature of the patient-provider relationship in psychiatry. As AI is in its infancy with regard to its implementation, psychiatry has the opportunity now to advocate for the appropriate legal frameworks necessary to minimize the risks of harm.

Current events in the rising labor movement of residents provide some guidance on how to address the potential drawbacks of AI implementation. Concerns regarding the use of AI in the entertainment industry led to the Screen Actors Guild and the American Federation of Television and Radio Artists unions making AI a critical negotiation point, given concerns for significant labor disruption and the ethics surrounding the use of a performer's likeness. Through collective action, residents may also be able to effectively bring about changes and safeguards surrounding the implementation of AI in health care systems.

Promoting the discussion and education of psychiatrists on how AI works through workshops and conferences will allow all psychiatrists to be more informed clinicians once such AI tools are more seriously implemented. A call to action by major medical organizations to better educate physicians on AI across all specialties will certainly aid in helping the field of psychiatry guide the use of AI and result in its successful and safe implementation. In its current state, ChatGPT and most AI tools, while promising, remain poorly understood, and few, if any, studies are currently available that robustly assess the risks and benefits of its use in psychiatric care. For the sake of its patients, psychiatry should take an active stance in the understanding of these tools and guide their development to ensure their proper use.

Declarations

Disclosures On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. 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(2):e0000198.
2. Koubaa A. GPT-4 vs. GPT-3.5: A Concise Showdown. Preprints. 2023. https://doi.org/10.20944/preprints202303.0422.v1.
3. Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations. Radiology. 2023;307(5):e230582.
4. Suchman K, Garg S, Trindade AJ. Chat generative pretrained transformer fails the multiple-choice American College of Gastroenterology self-assessment test. Am J Gastroenterol. 2023;118(12):2280–2.
5. Burk-Rafel J, Reinstein I, Feng J, Kim MB, Miller LH, Cocks PM, et al. Development and validation of a machine learning-based decision support tool for residency applicant screening and review. Acad Med. 2021;96(11S):S54–61.
6. Neumann M, Edelhäuser F, Tauschel D, Fischer MR, Wirtz M, Woopen C, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. Acad Med. 2011;86(8):996–1009.
7. Wartman SA, Combs CD. Reimagining medical education in the age of AI. AMA J Ethics. 2019;21(2):e146–52.
8. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023;183(6):589–96.
9. 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(1):e46885.
10. Kintsch W. The potential of latent semantic analysis for machine grading of clinical case summaries. J Biomed Inform. 2002;35(1):3–7.
11. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health. 2017;4(2):e19.
12. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. J Med Internet Res. 2021;23(3):e24850.
13. McCarthy JF, Cooper SA, Dent KR, Eagan AE, Matarazzo BB, Hannemann CM, et al. Evaluation of the recovery engagement and coordination for health–veterans enhanced treatment suicide risk modeling clinical program in the veterans health administration. JAMA Netw Open. 2021;4(10):e2129900.
14. Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Cardiologist-level arrhythmia detection with convolutional neural networks. 2017. https://doi.org/10.48550/arXiv1707.01836.
15. Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. Inf Sci. 2017;405:81–90.
16. Block L, Habicht R, Wu AW, Desai SV, Wang K, Silva KN, et al. In the wake of the 2003 and 2011 duty hours regulations, how do internal medicine interns spend their time? J Gen Intern Medicine. 2013;28:1042–7.
17. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digit Health. 2023;5(4):e179–81.
18. Patel SB, Lam K. ChatGPT: the future of discharge summaries? Lancet Digit Health. 2023;5(3):e107–8.
19 Lin SY, Shanafelt TD, Asch SM. Reimagining clinical documentation with artificial intelligence. Mayo Clin Proc. 2018;93(5):563–5.
20. Jargon J. A chatbot was designed to help prevent eating disorders. Then it gave dieting tips. Wall Street Journal. June 1, 2023. https://www.wsj.com/articles/eating-disorder-chatbot-ai-2aecb179. Accessed 17 Jan 2024.
21. Retraction Watch. How a now-retracted study got published in the first place, leading to a $3.8 million NIH grant. June 9 2023. https://retractionwatch.com/2023/06/09/how-a-now-retracted-study-got-published-in-the-first-place-leading-to-a-3-8-million-nih-grant/. Accessed 17 Jan 2024.
22. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53.
23. Gichoya JW, Banerjee I, Bhimireddy AR, Burns JL, Celi LA, Chen L-C, et al. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health. 2022;4(6):e406–14.
24. Farhud DD, Zokaei S. Ethical issues of artificial intelligence in medicine and healthcare. Iran J Public Health. 2021;50(11):i–v.
25. Martinez-Martin N, Luo Z, Kaushal A, Adeli E, Haque A, Kelly SS, et al. Ethical issues in using ambient intelligence in health-care settings. Lancet Digit Health. 2021;3(2):e115–23.
26. Josh A; Steven A SA, Lama A, Ilge A, Florencia A, et al. OpenAI. 2023. https://doi.org/10.48550/arXiv2303.08774.
27. Zakrzewski C. FTC investigates OpenAI over data leak and ChatGPT's inaccuracy. July 15, 2023. https://www.washingtonpost.com/technology/2023/07/13/ftc-openai-chatgpt-sam-altman-lina-khan/ Accessed 6 Aug 2023.
28. Lin Z. Why and how to embrace AI such as ChatGPT in your academic life. R Soc Open Sci. 2023;10:230658.
29. American Medical Association. Augmented intelligence in medicine. 28 Nov 2023. https://www.ama-assn.org/practice-management/digital/augmented-intelligence-medicine. Accessed 4 Feb 2024.
30. American Psychiatric Association. The basics of augmented intelligence: some factors psychiatrists need to know now. June 29, 2023. https://www.psychiatry.org/news-room/apa-blogs/the-basics-of-augmented-intelligence. Accessed 16 Apr 2024.