

Opportunities and risks of ChatGPT in medicine, science, and academic publishing: a modern Promethean dilemma

Jan Homolak^{1,2}

¹Department of Pharmacology, University of Zagreb School of Medicine, Zagreb, Croatia

²Croatian Institute for Brain Research, University of Zagreb School of Medicine, Zagreb, Croatia

Jan.homolak@mef.hr

The release of ChatGPT, the latest large (175-billion-parameter) language model by San Francisco-based company OpenAI, prompted many to think about the exciting (and troublesome) ways artificial intelligence (AI) might change our lives in the very near future. The OpenAI's chatbot allegedly gained more than 1 million users in the first few days after its launch and 100 million in the first 2 months, positioning itself as the fastest-growing consumer application in history (1). The hype surrounding ChatGPT is not unjustified: the model is (still) free, easy to use, and able to authentically converse on many subjects in a way that is almost indistinguishable from human communication. Furthermore, considering that ChatGPT was generated by fine-tuning the GPT-3.5 model from early 2022 with supervised and reinforcement learning (2), the quality of the chatbot-generated content can only be improved with additional training and optimization. As the inevitable implementation of this disruptive technology will have far-reaching consequences for medicine, science, and academic publishing, we need to discuss both the opportunities and risks of its use.

CAN CHATGPT REPLACE PHYSICIANS?

AI has a tremendous potential to revolutionize health care and make it more efficient by improving diagnostics, detecting medical errors, and reducing the burden of paperwork (3,4); however, chances are it will never replace physicians. Algorithms perform relatively well on

knowledge-based tests despite the lack of domain-specific training; ChatGPT achieved ~66% and ~72% on Basic Life Support and Advanced Cardiovascular Life Support tests, respectively (5), and performed at or near the passing threshold on the United States Medical Licensing Exam (6,7). However, they are notoriously bad at context and nuance (8) – two things critical for safe and effective patient care, which requires the implementation of medical knowledge, concepts, and principles in real-world settings. In their analysis of the future of employment, Frey and Osborne estimate that, while the probability of administrative health care jobs automation is relatively high (eg, 91% for health information technicians), the probability of automating the jobs of physicians and surgeons is 0.42% (9). While we might object as some evidence indicates that fully autonomous robotic systems might be “just around the corner” (10), the job of a surgeon goes far beyond performing a surgical procedure. The complexity of the physician's job lies in the ability to administer fully integrated care by providing treatment but also compassion. As medical students we were taught to always take care of patients and not of their medical records – a clinical skill that computer algorithms are still not able to comprehend. Therefore, the tremendous potential of AI in healthcare does not lie in the possibility of replacing physicians, but rather in the capacity to increase physicians' efficacy by redistributing workload and optimizing performance. In the words of Alvin Powell from *The Harvard Gazette*, “A properly developed and deployed AI, experts say, will be akin to

the cavalry riding in to help beleaguered physicians struggling with unrelenting workloads, high administrative burdens, and a tsunami of new clinical data." (11).

There are also some ethical issues to consider regarding conversational AI in medical practice. Training a model requires a tremendous amount of (high-quality) data, and current algorithms are often trained on biased data sets. In fact, the models are not only susceptible to availability, selection, and confirmation bias but are also reluctant to amplify it (12). For example, ChatGPT can provide biased outputs and perpetuate sexist stereotypes (13) – a challenge that has to be resolved before similar AI can be successfully and safely implemented in clinical practice (14–17). Other ethical issues are related to the legal framework. For example, it remains to be determined who is to blame when an AI physician makes an inevitable mistake.

A CHATBOT-SCIENTIST

ChatGPT already wrote essays, scholarly manuscripts, and computer code, summarized scientific literature, and performed statistical analyses (18,19). Furthermore, AI might soon be able to successfully perform more complex assignments such as designing experiments (20) or conducting a peer-review (18). In some of the mentioned tasks, ChatGPT performed alarmingly well. In a recent experiment, researchers used existing publications to generate 50 research abstracts that were able to pass the plagiarism check performed by a plagiarism checker, an AI-output detector, and human reviewers (21). On the one hand, the astounding ability of ChatGPT to write specialized texts suggests that similar tools might soon be able to write complete research manuscripts, which would enable scientists to focus on designing and performing the experiments rather than on writing manuscripts (18). The latter might promote quality and equity in research by shifting the focus from the presentation to the content and experimental results. On the other hand, conversational AIs are just language models trained to sound convincing, but without the ability to interpret and understand the content. Consequently, ChatGPT-generated manuscripts might be misleading, based on non-credible or completely made-up sources (18). The worst part is, the ability of ChatGPT to write a text of surprising quality might deceive reviewers and readers, with the final result being an accumulation of dangerous misinformation. StackOverflow, a popular forum for computer programming-related discussions, banned the use of ChatGPT-generated text *"because the average rate of*

getting correct answers from ChatGPT is too low, the posting of answers created by ChatGPT is substantially harmful to the site and to users who are asking and looking for correct answers" (22). ChatGPT seems to be equally unreliable when it comes to writing research articles. For example, Blanco-Gonzalez et al assessed the ability of ChatGPT to assist human authors in writing review articles and concluded that *"...ChatGPT is not a useful tool for writing reliable scientific texts without strong human intervention. It lacks the knowledge and expertise necessary to accurately and adequately convey complex scientific concepts and information."* (23). On top of that, the chatbot seems to have an alarming tendency to make up references with the goal of sounding convincing (18,24,25). In fact, the creators of ChatGPT openly disclosed that the fact that *"ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers"* will be a "challenging issue to fix" (2). A failure to acknowledge the limitations of conversational AI might pose an additional strain on the publishing system already flooded with meaningless data and low-quality manuscripts. Apart from the problem of unreliability, there are several additional ethical challenges (18,19,26). A chatbot cannot be held accountable for its work, and there is no legal framework to determine who owns the rights to the AI-generated work – the author of the manuscript, the author of the AI, or the (unknown) authors who contributed training data? Furthermore, since ChatGPT often fails to disclose the source of information, who is to blame for plagiarism if the chatbot decides to plagiarize? Until the ethical dilemmas are resolved, most publishers agree that the use of any kind of AI should be clearly acknowledged and that chatbots should not be listed as authors.

WHERE DO WE GO FROM HERE?

The powerful disruptive technology of conversational AIs is here to stay, and we can only expect them to improve with additional training and optimization. Banning or actively ignoring their use makes no sense – they can substantially improve many aspects of our lives by alleviating the burden of daunting and repetitive tasks. In medicine, AI might dramatically improve efficacy just by alleviating a fragment of the suffocating paperwork (27), and optimized chatbots (eg, Stanford's BioMedLM) (28) might dramatically speed up and improve literature search. Nevertheless, we should not be allured by the overwhelming potential of AI. For AI to realize its full potential in medicine and science, we should not implement it hastily but advocate its mindful introduction and an open debate about the risks and benefits.

References

- 1 Bartz D, Bartz D. 2023. As ChatGPT's popularity explodes, U.S. lawmakers take an interest. Available from: <https://www.reuters.com/technology/chatgpts-popularity-explodes-us-lawmakers-take-an-interest-2023-02-13/>. Accessed: February 24, 2022.
- 2 ChatGPT: Optimizing Language Models for Dialogue. Available from: <https://openai.com/blog/chatgpt/>. Accessed: February 24, 2022.
- 3 Ahn JS, Ebrahimi S, McDermott S, Lee L, Naccarato JF, Di Capua FY, et al. Association of artificial intelligence-aided chest radiograph interpretation with reader performance and efficiency. *JAMA Netw Open*. 2022;5:e2229289. [Medline:36044215](#) [doi:10.1001/jamanetworkopen.2022.29289](#)
- 4 Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. 2023;5:e145-6. [Medline:36754724](#)
- 5 Fijačko N, Gosak L, Štiglic G, Picard CT, John Douma M. Can ChatGPT pass the life support exams without entering the American Heart Association Course? *Resuscitation*. 2023;169:97-8. [Medline:36775020](#) [doi:10.1016/j.resuscitation.2023.109732](#)
- 6 Gilson A, Safranek C, Huang T, Socrates V, Chi L. How does ChatGPT perform on the United States Medical Licensing Examination? The Implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. [Medline:36753318](#) [doi:10.2196/45312](#)
- 7 Kung TH, Cheatham M, Medenilla A, Sillos C, Leon LD, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digit Health*. 2023;2:e0000198. [Medline:36812645](#) [doi:10.1371/journal.pdig.0000198](#)
- 8 Mbakwe AB, Lourentzou I, Celi LA, Mechanic OJ, Dagan A. ChatGPT passing USMLE shines a spotlight on the flaws of medical education. *PLoS Digit Health*. 2023;2:e0000205. [Medline:36812618](#) [doi:10.1371/journal.pdig.0000205](#)
- 9 Frey CB, Osborne MA. The future of employment: How susceptible are jobs to computerisation? *Technol Forecast Soc Change*. 2017;114:254-80. [doi:10.1016/j.techfore.2016.08.019](#)
- 10 Han J, Davids J, Ashrafian H, Darzi A, Elson DS, Sodergren M. A systematic review of robotic surgery: From supervised paradigms to fully autonomous robotic approaches. *Int J Med Robot*. 2022;18:e2358. [Medline:34953033](#) [doi:10.1002/rcs.2358](#)
- 11 Powell A. Risks and benefits of an AI revolution in medicine. Available from: <https://news.harvard.edu/gazette/story/2020/11/risks-and-benefits-of-an-ai-revolution-in-medicine/>. Accessed: February 24, 2023.
- 12 Rich AS, Gureckis TM. Lessons for artificial intelligence from the study of natural stupidity. *Nat Mach Intell*. 2019;1:174-80. [doi:10.1038/s42256-019-0038-z](#)
- 13 Chat GPT. Friend or foe? *Lancet Digit Health*. 2023;5:e1. [Medline:36754723](#)
- 14 Straw I. The automation of bias in medical Artificial Intelligence (AI): Decoding the past to create a better future. *Artif Intell Med*. 2020;110:101965. [Medline:33250145](#) [doi:10.1016/j.artmed.2020.101965](#)
- 15 Estiri H, Strasser ZH, Rashidian S, Klann JG, Wagholikar KB, McCoy TH, et al. An objective framework for evaluating unrecognized bias in medical AI models predicting COVID-19 outcomes. *J Am Med Inform Assoc*. 2022;29:1334-41. [Medline:35511151](#) [doi:10.1093/jamia/ocac070](#)
- 16 Adam H, Balagopalan A, Alsentzer E, Christia F, Ghassemi M. Mitigating the impact of biased artificial intelligence in emergency decision-making. *Commun Med (Lond)*. 2022;2:149. [Medline:36414774](#) [doi:10.1038/s43856-022-00214-4](#)
- 17 DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. *J Am Med Inform Assoc*. 2020;27:2020-3. [Medline:32574353](#) [doi:10.1093/jamia/ocaa094](#)
- 18 van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. *Nature*. 2023;614:224-6. [Medline:36737653](#) [doi:10.1038/d41586-023-00288-7](#)
- 19 Liebrez M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: Ethical challenges for medical publishing. *Lancet Digit Health*. 2023;5:e105-6. [Medline:36754725](#) [doi:10.1016/S2589-7500\(23\)00019-5](#)
- 20 Melnikov AA, Poulsen Nautrup H, Krenn M, Dunjko V, Tiersch M, Zeilinger A, et al. Active learning machine learns to create new quantum experiments. *Proc Natl Acad Sci U S A*. 2018;115:1221-6. [Medline:29348200](#) [doi:10.1073/pnas.1714936115](#)
- 21 Else H. Abstracts written by ChatGPT fool scientists. *Nature*. 2023;613:423-423. [Medline:36635510](#) [doi:10.1038/d41586-023-00056-7](#)
- 22 StackOverflow. 2022. Temporary policy: ChatGPT is banned. Available from: <https://meta.stackoverflow.com/questions/421831/temporary-policy-chatgpt-is-banned>. Accessed: February 25, 2023.
- 23 Blanco-Gonzalez A, Cabezon A, Seco-Gonzalez A, Conde-Torres D, Antelo-Riveiro P, Pineiro A, et al. The role of AI in drug discovery: challenges, opportunities, and strategies. 2022. arXiv:2212.08104.
- 24 Teresa Kubacka. Twitter. Available from: https://twitter.com/paniterka_ch/status/1599893718214901760. Accessed: February 25, 2023.
- 25 David Smerdon. Twitter. Available from: <https://twitter.com/dsmerdon/status/1618816703923912704>. Accessed: February 25, 2023.
- 26 Zohny H, McMillan J, King M. Ethics of generative AI. *J Med Ethics*. 2023;49:79-80. [Medline:36693706](#) [doi:10.1136/jme-2023-108909](#)
- 27 Siegler JE, Patel NN, Dine CJ. Prioritizing paperwork over patient care: why can't we do both? *J Grad Med Educ*. 2015;7:16-8. [Medline:26217415](#) [doi:10.4300/JGME-D-14-00494.1](#)
- 28 Bolton E, Hall D, Yasunaga M, Lee T, Manning C, Liang P. PubMedGPT 2.7B.158. Available from: <https://crfm.stanford.edu/2022/12/15/pubmedgpt.html>. Accessed: February 25, 2023.