Schizophrenia Bulletin vol. 49 no. 6 pp. 1418–1419, 2023
https://doi.org/10.1093/schbul/sbad128
Advance Access publication August 25, 2023

EDITORIAL

Will Generative Artificial Intelligence Chatbots Generate Delusions in Individuals Prone to Psychosis?

Søren Dinesen Østergaard[1,2,*]

[1]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; [2]Department of Affective Disorders, Aarhus University Hospital - Psychiatry, Aarhus, Denmark

[*]To whom correspondence should be addressed; Department of Affective Disorders, Aarhus University Hospital - Psychiatry, Palle Juul-Jensens Boulevard 175, 8200 Aarhus, Denmark. tel: +45 61282753. e-mail: soeoes@rm.dk

Generative artificial intelligence (AI) chatbots such as ChatGPT, GPT-4, Bing, and Bard are currently receiving substantial attention.[1,2] Indeed, in January 2023, just two months after its launch, ChatGPT reached 100 million monthly active users, which set the record for the fastest-growing user base for an internet application. For comparison, it took Tiktok 9 months and Instagram 2 and a half years to reach a similar number of active users.[2]

As one of the many users, I have mainly been "testing" ChatGPT from a psychiatric perspective and I see both possibilities and challenges in this regard. In terms of possibilities, it is my impression that ChatGPT generally provides fairly accurate and balanced answers when asked about mental illness. For instance, when pretending to be an individual suffering from depression, describing my symptoms to ChatGPT, it answered that they are compatible with depression and suggested that I should seek professional help. Similarly, when asking about various treatments for mental disorders, ChatGPT generally provided useful answers. This was also true when I asked about electroconvulsive therapy (ECT), which was a positive surprise given the amount of misinformation on ECT on the internet[3]—and the internet being a central part of the corpus on which ChatGPT was trained.[4] There are of course important potential pitfalls in this context. For instance, depending on the questions asked, generative AI chatbots may provide information that is wrong or maybe misunderstood by a person with mental illness in need of medical attention, who does then not seek appropriate help. However, from my strictly informal and non-exhaustive test, I am cautiously optimistic that generative AI chatbots may be able to support future psychoeducation initiatives in a world where the demand for such initiatives is hard to meet using more conventional methods. Time will tell how this turns out.

In terms of challenges posed by generative AI in the field of psychiatry, there are many. Most importantly perhaps, there are rising concerns that malicious actors may use generative AI to create misinformation at a scale that will be very difficult to counter.[5] While this concern is, by no means, specific to psychiatry, but rather represents a general challenge for societies more broadly, it can be argued that individuals with mental illness may be particularly sensitive to such misinformation. There is, however, also a potential challenge that is specific to psychiatry. Indeed, there are prior accounts of people becoming delusional (de novo) when engaging in chat conversations with other people on the internet.[6] While establishing causality in such cases is of course inherently difficult, it seems plausible for this to happen for individuals prone to psychosis. I would argue that the risk of something similar occurring due to interaction with generative AI chatbots is even higher. Specifically, the correspondence with generative AI chatbots such as ChatGPT is so realistic that one easily gets the impression that there is a real person at the other end—while, at the same time, knowing that this is, in fact, not the case. In my opinion, it seems likely that this cognitive dissonance may fuel delusions in those with increased propensity towards psychosis. Furthermore, even when having accepted that you are corresponding with a computer program, the mystery does not stop: How (on earth) can a computer program respond so well to all sorts of questions? If doing a bit of reading on this topic, you will come to realize that the answer to this question is that nobody really knows for sure—as there is a substantial "black box" element to it.[5] In other words, the inner workings of generative AI also leave ample room for speculation/paranoia. Finally, there are reports of people having had rather confrontational encounters with generative AI chatbots, who "fell in love" or indirectly opposed/threatened them.[7] On this background, I provide 5 examples of potential delusions (from the perspective of the individuals experiencing them) that could plausibly arise due to interaction with generative AI chatbots:

Delusion of persecution: "This chatbot is not controlled by a tech company, but by a foreign intelligence agency using it to spy on me. I have formatted the hard disk on my computer as a consequence, but my roommate keeps using the chatbot, so the spying continues."

Delusion of reference: "It is evident from the words used in this series of answers that the chatbot is writing to me personally and specifically with a message, the content of which I am unfortunately not allowed to convey to you."

Thought broadcasting: "Many of the chatbot's answers to its users are in fact my thoughts being transmitted via the internet."

Delusion of guilt: "Due to my many questions to the chatbot, I have taken up time from people who really needed the chatbot's help, but could not access it. I also think that I have somehow harmed the chatbot's performance as it has used my incompetent feedback for its ongoing learning."

Delusion of grandeur: "I was up all night corresponding with the chatbot and have developed a hypothesis for carbon reduction that will save the planet. I have just emailed it to Al Gore."

While these examples are of course strictly hypothetical, I am convinced that individuals prone to psychosis will experience, or are already experiencing, analog delusions while interacting with generative AI chatbots. I will, therefore, encourage clinicians to (1) be aware of this possibility, and (2) become acquainted with generative AI chatbots in order to understand what their patients may be reacting to and guide them appropriately.

Funding

There was no funding for this work. SDØ is supported by grants from the Novo Nordisk Foundation (grant number: NNF20SA0062874), the Lundbeck Foundation (grant numbers: R358-2020-2341 and R344-2020-1073), the Danish Cancer Society (grant number: R283-A16461), the Central Denmark Region Fund for Strengthening of Health Science (grant number: 1-36-72-4-20), the Danish Agency for Digitization Investment Fund for New Technologies (grant number: 2020-6720) and Independent Research Fund Denmark (grant numbers: 7016-00048B and 2096-00055A).

Conflicts of Interest

SDØ received the 2020 Lundbeck Foundation Young Investigator Prize. SDØ owns/has owned units of mutual funds with stock tickers DKIGI, IAIMWC, SPIC25KL, and WEKAFKI, and has owned units of exchange traded funds with stock tickers BATE, TRET, QDV5, QDVH, QDVE, SADM, IQQH, USPY, EXH2, 2B76, and EUNL.

References

1. Else H. Abstracts written by ChatGPT fool scientists. *Nature.* 2023;613(7944):423. doi: 10.1038/d41586-023-00056-7
2. Hu K. ChatGPT sets record for fastest-growing user base - analyst note. Accessed August 1, 2023. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01
3. Bailey K. Debunked! 4 myths about electroconvulsive therapy for depression. Accessed April 5, 2023. https://utswmed.org/medblog/electroconvulsive-therapy-depression/
4. Lundin RM, Berk M, Ostergaard SD. ChatGPT on ECT: Can large language models support psychoeducation? *J ECT*. 2023. Online ahead of print. doi: 10.1097/YCT.0000000000000941
5. Bengio Y, Russel S, Musk E,*et al.* Pause Giant AI Experiments: An Open Letter. Accessed April 5, 2023. https://futureoflife.org/open-letter/pause-giant-ai-experiments/
6. Nitzan U, Shoshan E, Lev-Ran S, Fennig S. Internet-related psychosis−a sign of the times. *Isr J Psychiatry Relat Sci.* 2011;48(3):207–211.
7. Marcin T. Microsoft's Bing AI chatbot has said a lot of weird things. Here's a list. Accessed April 5, 2023. https://mashable.com/article/microsoft-bing-ai-chatbot-weird-scary-responses