

Will ChatGPT transform healthcare?



ChatGPT and other large language models may be able to enhance healthcare delivery and patients' quality of life. But they will need to be tailored to specific clinical needs first.

Large language models, such as ChatGPT, use deep learning (DL) to reproduce human language in a convincing and human-like way. They are becoming increasingly common and are already being used in content marketing, customer services and a variety of business applications. As a result, it is inevitable that language models will also soon debut in healthcare, an area where they hold tremendous potential to improve health and enhance patients' lives, but not without pitfalls.

ChatGPT's ability to engage people with human-like conversation serves as a reminder of how important language and communication are to the human experience and well-being. Effective communication through language helps people to forge relationships with others, including the relationships between patients and healthcare professionals. One way that language models could improve care is by learning and producing language to assist patients in communicating with healthcare workers and with each other. For instance, it could help to improve compliance with medical prescriptions, by making the language more accessible to the patient and by reducing the chance of miscommunication. In addition, given that the quality of patient–physician relationships affect patient outcomes in a variety of conditions ranging from mental health¹ to obesity² and cancer³, it is reasonable to assume that using language models to strengthen those relationships through better communication would have beneficial impact for patients.

Large language models could also help with health interventions that rely on communication between non-professional peers. In a recent study, a language model⁴ that was trained to rewrite text in a more empathic way made communication easier in a peer-to-peer mental health support system, which enhanced non-expert conversational abilities. This example highlights the potential of using

human–artificial intelligence collaboration to improve a variety of community-based health tasks that rely on peer- or self-administered therapy, such as in cognitive behavioral therapy. Against a background of limited healthcare resources coupled with a growing mental health crisis – as reported⁵ by the US Centers for Disease Control and Prevention – application of such tools could increase assistance coverage, especially in settings that bypass the need for delivery of care by specialized healthcare workers.

Language communication can be both a therapeutic intervention, as in psychotherapy, and the target of therapy, such as in speech impairments such as aphasia. There are various types of language impairment, with different causes and coexisting conditions. Language models could be useful tools for personalized medicine approaches. For example, patients with neurodegenerative conditions may lose their ability to communicate through spoken language and progressively lose their vocabulary, which can worsen their social isolation and accelerate the degenerative process. Given that individual patients often present with unique combinations of specific patterns of neurodegenerative phenotypes, they may benefit from personalized approaches facilitated by artificial intelligence. Language models could also help patients with neurodegenerative diseases to expand their vocabulary or comprehend information more easily. They can achieve this by complementing language with other media or by reducing the complexity of the input the patients receive. In each of these cases, the algorithm would be specifically tailored to needs of each individual patient. These models could also play an important part in the development of speech brain–computer interfaces, which are designed to decode brain signals and imagined speech into vocalized language in people with aphasia. Such technologies not only would enhance coherence, but also reproduce the patient's communication style and meaning more accurately.

While their potential is huge, most applications of DL-based language models in healthcare are not yet ready for primetime. Specific clinical applications of DL-based language models will require extensive training on expert annotations in order to achieve

acceptable standards of clinical performance and reproducibility. Early attempts⁶ at using these models as clinical diagnostic tools without additional training have shown limited success, with the algorithm performance remaining lower than that of practicing physicians. Therefore, while it is tempting to bypass this very expensive requirement by relying on large training datasets and the adaptive learning capabilities of these tools, the evidence accumulated so far highlights the need for extensive and formal evaluation of language models against standard clinical practices after they have been trained for specific clinical tasks, such as diagnostic advice and triaging.

Using ChatGPT or other advanced conversational models as sources of medical advice by the public should be also a source of concern. Part of the allure of these new tools stems from humans being innately drawn toward anthropomorphic entities. People tend to more naturally trust something that mimics human behaviors and responses, such as the responses generated by ChatGPT. Consequently, people could be tempted to use conversational models for applications for which they were not designed, and in lieu of professional medical advice, such as retrieving possible diagnoses from a list of symptoms or deriving treatment recommendations. Indeed, a survey⁷ reported that around one-third of the US adults sought medical advice on the internet for self-diagnoses, with only around half of these respondents subsequently consulting a physician about the web-based results.

This means that the use of ChatGPT and other language models in healthcare will require careful consideration to ensure that safeguards are in place to protect against potentially dangerous uses, such as bypassing expert medical advice. One such protection measure could be as simple as an automated warning that is triggered by queries about medical advice or terms to remind users that the model outputs do not constitute or replace expert clinical consultation. It is also important to note that these technologies are evolving at a much faster pace than the regulators, government and advocates can cope. Given their wide availability, and potential societal impact, it is critical that all stakeholders

– developers, scientists, ethicists, healthcare professionals, providers, patients, advocates, regulators and governmental agencies – get involved and are engaged in identifying the best way forward. Within a constructive and alert regulatory environment, DL-based language models could have a transformative impact in healthcare, augmenting

rather than replacing human expertise, and ultimately improving quality of life for many patients.

Published online: 14 March 2023

References

1. Horvath, A. O. & Luborsky, L. *J. Consult. Clin. Psychol.* **61**, 561–573 (1993).

2. Sturgiss, E. A., Sargent, G. M., Haesler, E., Rieger, E. & Douglas, K. *Clin. Obese.* **6**, 376–379 (2016).
3. Mack, J. W. et al. *Cancer* **115**, 3302–3311 (2009).
4. Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C. & Althoff, T. *Nat. Mach. Intell.* **5**, 46–57 (2023).
5. Centers for Disease Control and Prevention. *Youth Risk Behavior Survey. Data Summary & Trends Report* (CDC, 2021)
6. Levine, D. M. et al. Preprint at medRxiv <https://doi.org/10.1101/2023.01.30.23285067> (2023)
7. Kuehn, B. M. *JAMA* **309**, 756–757 (2013).