

## PERSPECTIVE

# ChatGPT/GPT-4 in healthcare: Potential opportunities and limitations

Ruochen Wang\*

AbleTo Inc., New York 10018, United States

## INTRODUCTION

Artificial intelligence (AI) research has been fast-growing in recent decades due to computational hardware advancements supporting more and more complex and powerful machine learning models. AI has also demonstrated its ability to complete various medical-related tasks. In recent years, we have seen more use of AI in clinical practices to help improve healthcare service qualities and reduce healthcare professional workloads. A relatively more advanced AI could review healthcare data and monitor patients to give timely feedback on the prescribed treatments. Some other AI could also contribute to expanding human knowledge by discovering new proteins and drugs. More importantly, the release of Chat Generative Pre-trained Transformer (ChatGPT)/GPT-4 in March 2023 completely breakthrough the field of AI.<sup>[1]</sup>

ChatGPT/GPT-4 is the state-of-the-art large language model (LLM) developed by OpenAI, an AI research company. ChatGPT excels at its versatility compared to other models, maintaining high performances in various fields. The performance of ChatGPT on professional exams across many fields matches the levels of humans even though the model is not specifically trained to answer exam questions.<sup>[1]</sup> Moreover, researchers have found that ChatGPT could generate a suitable answer for the specific application if provided with a suitable prompt.<sup>[2]</sup> Therefore, the pre-train, prompt, and predicting paradigm is adopted to avoid fine-tuning training for ChatGPT, which generally requires large computation resources for large language models. The versatility of ChatGPT and the power of prompt

engineering granted machine learning practitioners and developers to derive variants of ChatGPT to perform the desired language tasks. These characteristics of ChatGPT revealed the potential that it could benefit healthcare and medical research depending on how we use it. Therefore, in this mini-review, we explore the existing use cases of GPT-4 in healthcare and outline the limits.

## CHATGPT/GPT-4 APPLICATIONS IN CLINICAL PRACTICE

ChatGPT is a strong candidate for relieving the workload of human experts in telemedicine for answering medical queries. With its intelligent question-answering features, patients could get the answers much faster than consulting medical doctors for preliminary information. Johnson *et al.*<sup>[3]</sup> evaluated ChatGPT with questions about common myths for cancers. The results suggest that there is no significant difference between the answers by ChatGPT and the answers on the website of the National Cancer Institute (NCI). This illustrates ChatGPT's ability for the common Q & A for the medical field, but there are no questions asked at the same level of complication as a real-life diagnosis in the hospital.

ChatGPT could also be used to generate medical documents, which is a necessary but time-consuming process. Formal clinical documents often require the information to be in a standardized format from existing templates. This standardization of the formats is also a workload that could be further reduced by ChatGPT. Patel *et al.*<sup>[4]</sup> used ChatGPT as an efficient tool to format the discharge summaries for the patients. Doctors'

### \*Corresponding Author:

Ruochen Wang, AbleTo Inc., 320 W. 37th Street, New York 10018, United States. Email: ruochen.wang@alumni.duke.edu; <https://orcid.org/0009-0000-5051-8746>

Received: 17 January 2024; Revised: 20 January 2024; Accepted: 22 April 2024

<https://doi.org/10.54844/prm.2024.0528>

necessary input is only the brief information to include, the concepts, and guidance to explain.

Since ChatGPT has been proven competent to answer these basic medical questions, some researchers looked into how it performs in clinical decision-making. In clinical practice, the decision-making process such as generating diagnoses of patients is complicated due to the complexity of human physiology. Hirosawa *et al.*<sup>[5]</sup> compared the performance of ChatGPT with internal medicine physicians and found high accuracy of the diagnosis in ChatGPT.

The language models are proven to be potentially beneficial to radiology doctors on multiple tasks, such as image captioning.<sup>[6]</sup> ChatGPT also has the potential to improve the explainability of the imaging diagnosis. ChatGPT could be finetuned with a dataset of simplified radiology reports to make ChatGPT provide a simplified version of the radiology report so that the patients could understand it more easily.<sup>[7]</sup> Prompt learning could also be applied to ChatGPT to translate magnetic resonance imaging (MRI) and computed tomography (CT) scan reports into language that could be much easier to understand by humans.<sup>[8]</sup> Their results showed that the task is best accomplished when ChatGPT is given a more detailed prompt. Both studies revealed the opportunity of ChatGPT to make the radiology reports more accessible to patients but also highlighted the potential risk that the performance is sensitive to the prompt.

## CHATGPT/GPT-4 APPLICATION IN BIOMEDICAL RESEARCH TEXTS

ChatGPT was trained on a large amount of text, making it a competent search engine to reply to the user with direct answers. Compared to traditional search engines which direct the users to other websites, ChatGPT offers a more convenient way for the authors to look up the information. As mentioned above, GPT-4 is a state-of-the-art LLM that can perform many natural language processing (NLP) tasks such as text-mining. In the most recent decade, more and more biomedical research has been published and expanded the total amount of text in the field, making text mining challenging. This work benchmarked ChatGPT on common NLP tasks with several open-source datasets in biomedical fields. However, the scores of ChatGPTs revealed poor performance overall. ChatGPT's performance on the question-answering task is the only one that is comparable to the baseline models.<sup>[9]</sup> This is not only because the performance of the ChatGPT is compared with biomedical field-specific models but also because the other tasks are designed for supervised models. This result showed that the current version of ChatGPT may

not be suitable for biomedical NLP tasks yet but it excels at extracting the relevant information from the text.

## DISCUSSION

In this study, we have reviewed the applications of ChatGPT in biomedical research and clinical practice. ChatGPT offers great convenience for users to look up information. The patients could benefit from the technology as they could look up basic medical information faster. Doctors could benefit from the generative feature of the technology as they could spend less time on the format of clinical documents. Biomedical researchers could benefit and enhancing their productivity in publishing their findings. However, such convenience comes at the cost of accuracy. The users of ChatGPT need to evaluate the answers with caution. Even though the reviewed publications showed the performance of the ChatGPT is near human, it does not suggest that it could always provide the correct information as human also makes mistakes. The challenge of applying ChatGPT to the biomedical field lies in the proper evaluation and the interpretability of the algorithm and the ultimate bottleneck is still dependent upon human review at the current stage.

Looking forward, the ethics problem remains challenging in fully automating healthcare as long as the ChatGPT remains a black box. Moreover, the proportion of medical-related text is much less compared to the text dataset used to train ChatGPT. This presents a limitation for ChatGPT to generate insightful medical documents. However, using ChatGPT as a professional assistant on formal medical documents, such as discharge reports and radiology reports, turns out to be a more reliable application that reduces the workload of the doctors. In the future, there will be more medical-related text datasets which will enhance the performance in the medical field.

## DECLARATION

### Author contributions

Wang RC developed the concept for the manuscript, reviewed the literature, formulated research questions, collected the data, conducted analyses and interpreted the data. The author read and approved the final manuscript.

### Source of funding

There is no funding for this work.

### Ethical approval

Not applicable.

**Informed consent**

Not applicable.

**Conflict of interest**

The author of the paper serves as Data Scientist at AbleTo, Inc. The article was subject to the journal's standard procedures.

**Data availability statement**

Data used to support the findings of this study are available from the corresponding author upon request.

**REFERENCES**

1. OpenAI, Josh A, Steven A, *et al.* GPT-4 Technical Report. *ArXiv*. 2023.
2. Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. *ACM Comput Surv*. 2023;55(9):1-35.
3. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. *JNCI Cancer Spectr*. 2023;7(2):1-9.
4. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. 2023;5(3):e107-e108.
5. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic accuracy of differential-diagnosis lists generated by generative pretrained transformer 3 chatbot for clinical vignettes with common chief complaints: a pilot study. *Int J Environ Res Public Health*. 2023;20(4):3378.
6. Chen J, Guo H, Yi K, Li B, Elhoseiny M. VisualGPT: data-efficient adaptation of pretrained language models for image captioning. *IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR)*. IEEE; 2022: 18009-18019.
7. Jeblick K, Schachtner B, Dexl J, *et al.* ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. *Eur Radiol*. 2024;34(5):2817-2825.
8. Lyu Q, Tan J, Zapadka ME, *et al.* Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: results, limitations, and potential. *Vis Comput Ind Biomed Art*. 2023;6(1):9.
9. Chen Q, Sun H, Liu H, *et al.* A Comprehensive Benchmark Study on Biomedical Text Generation and Mining with ChatGPT. *Cold Spring Harb Lab*. 2023.