Original Article

Discussion of the ability to use chatGPT to answer questions related to esophageal cancer of patient concern

Fengxia Yu, Mingyu Lei, Shiyu Wang, Miao Liu, Xiao Fu, Yuan Yu

Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Abstract

Background: Chat Generation Pre-Trained Converter (ChatGPT) is a language processing model based on artificial intelligence (AI). It covers a wide range of topics, including medicine, and can provide patients with knowledge about esophageal cancer. **Objective:** Based on its risk, this study aimed to assess ChatGPT's accuracy in answering patients' questions about esophageal cancer. **Methods:** By referring to professional association websites, social software and the author's clinical experience, 55 questions concerned by Chinese patients and their families were generated and scored by two deputy chief physicians of esophageal cancer. The answers were:(1) comprehensive/correct, (2) incomplete/partially correct, (3) partially accurate, partially inaccurate, and (4) completely inaccurate/irrelevant. Score differences are resolved by a third reviewer. **Results:** Out of 55 questions, 24 (43.6%) of the answers provided by ChatGPT were complete and correct, 13 (23.6%) were correct but incomplete, 18 (32.7%) were partially wrong, and no answers were completely wrong. Comprehensive and correct answers were highest in the field of prevention (50 percent), while partially incorrect answers were highest in the field of treatment (77.8 percent). **Conclusion:** ChatGPT can accurately answer the questions about the prevention and diagnosis of esophageal cancer, but it cannot accurately answer the questions about the treatment and prognosis of esophageal cancer. Further investigation and refinement of this widely used large-scale language model are needed before it can be recommended to patients with esophageal cancer, and ongoing research is still needed to verify the safety and accuracy of these tools and their medical applications.

Keywords: ChatGPT, discussion, disease management, esophageal cancer, telemedicine, patient concern

Introduction

As humanity enters a new era marked by significant advances in artificial intelligence (AI), the integration of AI with the field of bioinformatics offers great potential to improve the healthcare industry. Chat Generation Pre-Trained Converter (ChatGPT) is one of the latest AI models that is considered to have great potential in answering medical questions of concern to patients. It uses human feedback for reinforcement learning, fine-tuning conversational tasks to improve the accuracy and consistency of the information it provides.[1] Since its initial public launch, ChatGPT has exploded in popularity on the web, largely thanks to its ability to skillfully and accurately handle a large number of questions and tasks through an intuitive user interface.[2] Esophageal cancer is a common malignant tumor of the digestive system, according to the latest global cancer data released by the International Agency for Research on Cancer, the number of global esophageal cancer deaths in 2022 is as high as 544,000,

Address for correspondence: Dr. Yuan Yu, Department of Thoracic surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing - 100 021, China.
E-mail: zlyyyuyuan@163.com

Received: 18-07-2024 **Revised:** 29-08-2024
Accepted: 01-11-2024 **Published:** 25-04-2025

Access this article online

Quick Response Code:

Website: http://journals.lww.com/JFMPC

DOI: 10.4103/jfmpc.jfmpc_1236_24

How to cite this article: Yu F, Lei M, Wang S, Liu M, Fu X, Yu Y. Discussion of the ability to use chatGPT to answer questions related to esophageal cancer of patient concern. J Family Med Prim Care 2025;14:1384-8.

which is the seventh largest cancer-related cause of death in the world.[3] China is a country with a high incidence of esophageal cancer, accounting for more than half of the world's incidence.[4]

In the current surge of esophageal cancer patients, medical resources are relatively scarce. In an increasingly digital communication environment, patients, including those diagnosed with esophageal cancer, need to learn about the disease when they are not in the hospital, and they have begun to collect medical information through search engines and AI chat software such as ChatGPT. However, the reliability and accuracy of the information provided by ChatGPT have yet to be fully validated, especially when answering patient questions related to specific medical conditions such as esophageal cancer. Therefore, the purpose of this study was to investigate the accuracy and reliability of ChatGPT's answers to questions related to esophageal cancer, with a view to exploring the potential effects of ChatGPT on the medical industry.

Methods

By referring to professional association websites, social software, and the author's clinical experience, 55 questions concerned by Chinese patients and their families were generated and scored by two attending physicians of esophageal cancer tumors. The answers were: (1) complete/correct, (2) incomplete/partially correct, (3) partially accurate, partially inaccurate, and (4) completely inaccurate/irrelevant. Score differences are resolved by a third reviewer. Reproducibility scores were assessed by repeating questions and analyzing grading consistency as summarized by a third reviewer. If the answers do not agree, the answers will be scored by another esophageal cancer specialist to resolve the differences. If the answer is still inconsistent after the third reviewer, the answer will be sent to the fourth expert. The final score for each question is a score agreed upon by the two experts. The study was a back-to-back survey, and none of the experts knew each other's scores. For further analysis, scores 1 and 2 are classified as "correct" scores, while scores 3 and 4 are classified as "incorrect" scores. All statistical analyses were performed using Microsoft Excel.

Result

Of the 55 questions [Figure 1], ChatGPT provided 24 complete and correct answers (43.6%), 13 correct but incomplete answers (23.6%), 18 partially wrong answers (32.7%), and no completely wrong answers. Of the 55 questions [Figure 2], in the field of prevention, the proportion of comprehensive and correct is highest (50%). The proportion of partially incorrect answers was highest in the field of therapy (77.8%).

Discussion

With the popularization of contemporary technology, the use of AI is becoming more common, and its use in healthcare is also increasing. Many patients are using AI to learn about their disease. Although ChatGPT is now widely used in various fields. Its ability to accurately answer questions related to diseases that are of concern to non-medical personnel remains to be verified. In this study, we evaluated the reliability and accuracy of ChatGPT by determining the correctness of the answers provided by AI when answering questions had some research. However, the research on esophageal cancer is still blank, and the literature in other fields shows that ChatGPT's results in answering questions related to other diseases are similar to our investigation results; that is, ChatGPT is reasonable in answering questions related to some aspects, but the accuracy of different questions is different.

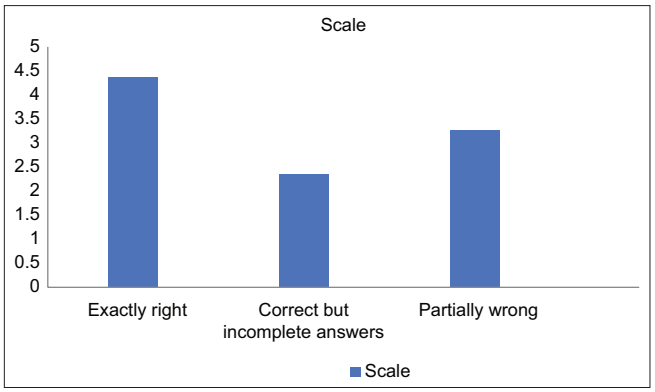

Figure 1: Category score

The results of our investigation show that ChatGPT can provide relatively accurate answers to questions about esophageal cancer prevention, and ChatGPT can provide easy-to-understand advice and guidance on esophageal cancer prevention, such as "If I have reflux esophagitis, how can I prevent developing esophageal cancer?" ChatGPT answers these questions with different recommendations based on the patient's current symptoms, such as "medications that the doctor may give, lifestyle changes, weight management, etc." ChatGPT can explain complex ideas in straightforward language, providing patients with knowledge that enables them to make informed decisions about their health. Our findings showed that questions involving medical-related words were given explanations that could be understood by non-specialists, such as when referring to drugs, ChatGPT responded that "some drugs, such as omeprazole, esomeprazole, lansoprazole, pantoprazole, and rabeprazole, work by reducing the production of stomach acid and promoting healing of the lining of the esophagus." As a tumor of the digestive system, esophageal cancer may not be understood by non-professionals, but ChatGPT can fully explain the action, method of action, side effects, etc., while doctors' explanations are often not very comprehensive. In addition to the ability to communicate simply and easily, ChatGPT also has the ability to participate in casual conversations, ChatGPT can communicate anytime, anywhere, and patients can use various channels (such as mobile apps or websites) to communicate with ChatGPT and get information and advice on how to manage their disease anytime, anywhere. And get information and advice on how to manage the disease.[5] This means that patients can learn about the disease anytime, anywhere, rather than entering the hospital to talk to their own

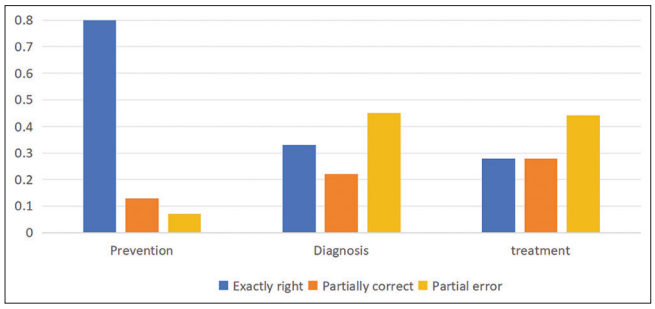

Figure 2: Percentage of each category score

doctors, which provides a great convenience for patients. The transition from the old generation of AI chatbots to ChatGPT represents that the content produced by human-computer interaction can realize the flow from mutual isolation to each other. In popular terms, ChatGPT has a "memory system" similar to that of humans. Therefore, when a user re-establishes a conversation with ChatGPT, they can recall their previous interaction and the annotations made by the user.[6] This is the advantage of AI, although it does not have the emotional support capabilities of humans, its database is quite large. Although AI has the above advantages in medicine and can provide convenience to patients, we still need to consider its accuracy and reliability, which is the primary consideration when we investigate whether AI can be applied to clinical medicine.

Yeo *et al*.[7] studied the expression of ChatGPT in cirrhosis and hepatocellular carcinoma (HCC). The correct rate was 79% for the cirrhosis question and 74% for the HCC question, but the correct rate for the combined answer was much lower, at 47.3% and 41.4%, respectively. Similar inconsistencies can be seen in the obesity literature. ChatGPT answered 86% of the questions across all study areas, but there was a large variation within the question subgroups (66.7% to 93.3%).[8] This shows that ChatGPT's answers to some medical questions are still different from clinical ones and are not mature. When we investigated the use of ChatGPT to answer questions related to esophageal cancer, there were also huge differences between the overall and different subgroups, with the overall accuracy of 0.67 and the accuracy of 0.93 for prevention, but only 0.55 and 0.56 for diagnosis and treatment. Although ChatGPT can give more accurate answers in terms of prevention, its results are not satisfactory in answering the questions concerning treatment and prognosis of patients with esophageal cancer, which also indicates that ChatGPT cannot completely replace doctors to give targeted plans. In addition, the wrong answers in the diagnosis and treatment are as high as 0.45 and 0.46, respectively, which indicates that the answers given by ChatGPT in answering relevant questions have no reference significance, which will not only make patients make inappropriate decisions, but also cause a crisis of trust between doctors and patients. The knowledge of different medical fields varies greatly, and the reliability and security of information provided by AI should be paid primary attention.[9] Park *et al*.'s research also highlights the limitations of different AI tools in understanding problems and the importance of careful use. Priority should be given to the security of the information received by the patient. In addition, the complexities of surgery and/or chemoradiotherapy, including their associated risks and benefits, are often not well understood by these patients. Therefore, it is important for them to provide comprehensive, accurate, and easy-to-understand information, and our findings show that ChatGPT does not do a good job in esophageal cancer. On the other hand, ChatGPT answers questions that require an understanding of established facts (e.g. "What causes esophageal cancer?") better than questions that require an understanding of the nuances that result from specific patient situations or controversial changes in practice patterns (e.g. "After a diagnosis of esophageal cancer, is a cure possible?"). When performing better. This indicates that ChatGPT lacks individualization and individualization in answering questions, and it is difficult to give personalized answers for different patients, which is consistent with Catherine and other studies.[10]

We found that ChatGPT's answers to some questions were vague and incomplete, often including the words "possible," "best," and "appropriate," such as "What are the complications after esophageal cancer surgery?" Among the answers ChatGPT gave was that "esophageal surgery can affect nutrient intake and absorption, leading to nutritional deficiencies." Patients may need nutritional support, dietary modifications, or supplements to address these deficits and promote healing and recovery. In addition, the complications listed are not comprehensive. When patients have unknown complications after surgery, it will greatly increase the psychological burden of patients and produce anxiety. Patients who receive a cancer diagnosis often show high emotional sensitivity and anxiety, especially in terms of their disease prognosis and survival. Once they are given inaccurate information about the disease, this often causes them to reject some treatments. Research shows that patients primarily seek and trust healthcare providers who are competent, compassionate, and have good interpersonal skills.[10-13] The researchers also found that empathic and compassionate doctor-patient interactions can improve patient satisfaction and patient clinical adherence.[14] In contrast, AI does not capture the emotional changes of patients when the service is provided, which may create a communication barrier between patients and AI applications. In addition, a survey shows that patients' reliance on the clinical application of AI may reduce the interaction and dialogue between doctors and patients, resulting in estrangement and mistrust between doctors and patients, and increasing the psychological burden of patient treatment.[15] In addition, ChatGPT lacks emotional support when answering questions, and it is difficult to provide psychological comfort and guidance to patients who ask questions, which also indicates that although ChatGPT can be used as a medical aid for patients, it cannot replace the medical care and psychological support provided by licensed healthcare providers, which is a difficult difficulty for current AI to break through.

Although ChatGPT can facilitate the healthcare system, care is still an irreplaceable role. Although AI is still immature in some areas of medicine, in the rapid spread of networks

and the increasing maturity of AI, the advantages of AI are also emerging, which can free doctors from the laborious work of medical records and patient education, so that they can focus on more complex tasks that require more human expertise and flexibility. At present, ChatGPT cannot completely replace doctors in managing the health status of patients with esophageal cancer, and it still needs to expand the database, improve the model and continue to train for continuous optimization. We are only in the early stages of the era of AI healthcare delivery, and it is not recommended to rely entirely on existing versions of ChatGPT as the sole source of healthcare information at this time. ChatGPT still has a good medical future, it can answer patients' questions anytime, anywhere, especially in the current era of medical resources, we need to continue to pay attention to the development of AI.

The main strength of this study is its novelty, which is the first quantitative analysis of ChatGPT in the field of health effects of esophageal cancer. The design of our questions is mainly based on the summary of the clinical patient's concerns, and the wording of the questions is as close as possible to the patient's way of asking. The selected reviewers are all Associate chief physicians in the Department of Esophageal Oncology, all of whom have 20 years of clinical experience, and our scores are highly trustworthy. Limitations of our study include differences between reviewers, with 15% of questions requiring additional reviewers to assign final scores. The fact that multiple participants are required to assess the accuracy of answers is in itself a reflection that ChatGPT needs improvement. In addition, we chose a scoring system that has been used in multiple other studies, all of which have some degree of subjectivity.

Conclusion

With the development of AI and ChatGPT, ongoing research is still needed to verify the safety and accuracy of these intelligent tools and their medical applications. In conclusion, ChatGPT accurately answered questions about the prevention and diagnosis of esophageal cancer, but it did not accurately answer 7 can be recommended for patients with esophageal cancer diagnosis.

The primary strength of our study is its novelty; this is the first quantitative analysis of ChatGPT's performance in a high-impact area of esophageal cancer health. Our question panel was designed using questions from well-regarded patient resources, and questions were phrased in a way to best approximate a patient entry. The reviewers are all attending esophageal oncologists with a cumulative 30 years of attending experience between them, lending credibility to our scoring assessments.

The limitations of our study include inter-reviewer variability, with 29% of questions requiring additional reviewers to assign a final score. Additionally, we chose a scoring system utilized in multiple other studies, all scoring systems have some degree of subjectivity, and there were no discrepancies that were not able to be resolved by the additional reviewers.

CRediT authorship contribution statement

Fengxia Yu: Methodology, Project, administration, Data curation, Formal analysis, Writing. Yuan Yu: review and editing. Mingyu Lei/Shiyu Wang/Miao Liu/Xiao Fu/Jianxin Liu: Conceptualization, Methodology, Writing.

Financial support and sponsorship

This study was supported by the Beijing Hope Run Special Fund of Cancer Foundation of China [Project Number: LC2022C02] and CAMS Innovation Fund for Medical Sciences(CIFMS) [Project Number: 2022-I2M-C and T-B-084].

Conflicts of interest

There are no conflicts of interest.

References

1. Mohapatra DP, Thiruvoth FM, Tripathy S, Rajan S, Vathulya M, Lakshmi P, *et al.* Leveraging large language models (llm) for the plastic surgery resident training: Do they have a role? Indian J Plast Surg 2023;56:413-20.
2. Kuscu O, Pamuk AE, Sutay SN, Hosal S. Is chatgpt accurate and reliable in answering questions regarding head and neck cancer? Front Oncol 2023;13:1256459. doi: 10.3389/fonc.2023.1256459.
3. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, *et al.* Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74:229-63.
4. GBD 2017 Oesophageal Cancer Collaborators. The global, regional, and national burden of oesophageal cancer and its attributable risk factors in 195 countries and territories, 1990-2017: A systematic analysis for the global burden of disease study 2017. Lancet Gastroenterol Hepatol 2020;5:582-97.
5. Khan I, Agarwal R. Can chatgpt help in the awareness of diabetes? Ann Biomed Eng 2023;51:2125-29.
6. Chatterjee J, Dethlefs N. This new conversational ai model can be your friend, philosopher, and guide. And even your worst enemy. Patterns (N Y) 2023;4:100676. doi: 10.1016/j.patter.2022.100676.
7. Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, *et al.* Assessing the performance of chatgpt in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol 2023;29:721-32.
8. Samaan JS, Yeo YH, Rajeev N, Hawley L, Abel S, Ng WH, *et al.* Assessing the accuracy of responses by the language model chatgpt to questions regarding bariatric surgery. Obes Surg 2023;33:1790-96.
9. Park I, Joshi AS, Javan R. Potential role of chatgpt in clinical otolaryngology explained by chatgpt. Am J Otolaryngol 2023;44:103873. doi: 10.1016/j.amjoto.2023.103873.
10. Hermann CE, Patel JM, Boyd L, Growdon WB, Aviki E, Stasenko M. Let's chat about cervical cancer: Assessing the accuracy of chatgpt responses to cervical cancer questions. Gynecol Oncol 2023;179:164-68.

11. Kerasidou A. Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. Bull World Health Organ 2020;98:245-50.
12. Singh P, King-Shier K, Sinclair S. South asian patients' perceptions and experiences of compassion in healthcare. Ethn Health 2020;25:606-24.
13. Charon R. The patient-physician relationship. Narrative medicine: A model for empathy, reflection, profession, and trust. JAMA 2001;286:1897-902.
14. Kelley JM, Kraft-Todd G, Schapira L, Kossowsky J, Riess H. The influence of the patient-clinician relationship on healthcare outcomes: A systematic review and meta-analysis of randomized controlled trials. Plos One 2014;9:e94207. doi: 10.1371/journal.pone.0094207.
15. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. Plos Med 2018;15:e1002689. doi: 10.1371/journal.pmed.1002689.