



Commentary

AI Chat bot - ChatGPT-4: A new opportunity and challenges in complementary and alternative medicine (CAM)



Tae-Hun Kim ^a, Jung Won Kang ^b, Myeong Soo Lee ^{c,*}

^a Korean Medicine Clinical Trial Center, Korean Medicine Hospital, Kyung Hee University, Seoul, South Korea

^b Department of Acupuncture & Moxibustion, College of Korean Medicine, Kyung Hee University, Seoul, South Korea

^c Division of Clinical Medicine, Korea Institute of Oriental Medicine, Daejeon, South Korea

1. Introduction

ChatGPT-4 is the latest artificial intelligence (AI) chatbot developed by OpenAI and is specifically engineered to generate human-like texts based on the input that it receives. The emergence of this service is one of the hottest news topics worldwide in 2023. This may be related to people's hopes and concerns regarding the potential impact and subsequent social changes that AI will bring in the future. Utilizing the generative pretrained transformer (GPT) architecture, it employs self-attention to understand natural language, allowing it to identify word relationships and generate contextually relevant responses. Although no one can predict exactly how this AI chatbot will be used, it is expected to offer various potential applications for general usage, such as providing information on various topics, assisting in drafting and editing texts, generating creative content, offering customer support, and facilitating language translation.^{1,2} GPT-4 was not specifically designed and programmed for medical tasks, but its potential role in medicine extends beyond medical consultation to include various tasks in a clinic, such as taking medical notes or suggesting billing codes for patients.³

The role of AI chatbot services in the complementary and alternative medicine (CAM) field is uncertain, with the potential for either minimal impact or significant change. Predicting the direction of change is challenging, but it is crucial for CAM professionals to stay informed and adapted. This article aims to assess ChatGPT's potential impact on the CAM field, predict its utilization, and discuss limitations and areas for further development. By understanding the current state and applications of ChatGPT, we can better prepare for its future role in the CAM field.

To evaluate ChatGPT-4's performance in providing information on evidence for CAM interventions in the context of cancer-related fatigue, we conducted a narrative analysis comparing clinical evidence or recommendations presented in a Korean clinical practice guideline (CPG)⁴ and the latest systematic reviews (SRs)⁵⁻⁷ with those generated by ChatGPT-4. Based on these findings, we identified ChatGPT-4's potential role and limitations in CAM practice.

2. Evidence-based support for clinical recommendations with CAM interventions provided by ChatGPT-4

The responses of ChatGPT-4 regarding commonly used CAM interventions for cancer-related fatigue were found to be substantially based on recent evidence when compared with the findings of the latest SRs in terms of both the recommendation direction and limitations.

For manual acupuncture and electroacupuncture, ChatGPT-4 provided answers indicating that while there are studies with significant effects, the quality of the studies forming the evidence prevents strong conclusions about their efficacy. This is largely in line with the conclusions of a recent SR.⁷ Additionally, ChatGPT-4's response on the effectiveness of moxibustion closely mirrors the results of the latest SR by indicating that there is insufficient evidence for its effectiveness and calling for high-quality future research.⁶ For herbal medicine, ChatGPT-4 notes that there is some evidence of its effectiveness but more research is still needed, which is consistent with the findings of the latest review.⁵ Based on these findings, it appears that the evidence-based recommendations for CAM interventions provided by ChatGPT-4 are acceptable (Table 1).

What if a CAM practitioner requests detailed information for the purpose of treating patients with each intervention? We compared the details of the treatments recommended by ChatGPT-4 and those included in the currently developed CPG for cancer-related symptoms.⁴ We found that in general, ChatGPT-4 provides meaningful information for acupuncture, including detailed information for treatment such as recommended acupuncture points and treatment frequency (or duration). However, it is not yet able to provide treatment-related information for moxibustion, such as moxibustion points, or for herbal medicine, such as detailed prescriptions at the level of the CPG (Supplement 1).

Through this example, it becomes evident that ChatGPT-4 demonstrates considerable success in outlining the general evidence and limitations associated with CAM interventions for specific conditions. Nevertheless, it falls short in delivering actionable knowledge for clinical practitioners. It is worth noting that the hallucination, or the genera-

* Corresponding author.

E-mail address: drmslee@gmail.com (M.S. Lee).

Table 1

Comparison between ChatGPT-4's responses and conclusions from the latest systematic reviews on the evidence of CAM interventions: an example case

ChatGPT-4's response	Latest systematic review*
<p>A question (an input prompt): Could you provide the clinical evidence for manual acupuncture, electroacupuncture, moxibustion, and herbal medicine in treating cancer-related fatigue? Would you recommend these interventions? What is the strength of recommendations and the level of evidence for each intervention?</p>	
<p>As an AI language model, I can provide you with a general overview of the evidence regarding these CAM interventions for cancer-related fatigue. However, please note that my knowledge is up to September 2021, and it is always best to consult the latest research and guidelines for the most up-to-date information.</p> <ol style="list-style-type: none"> 1. Manual acupuncture: Some studies have suggested that manual acupuncture may help alleviate cancer-related fatigue. However, the overall quality of evidence may be low due to small sample sizes, the risk of bias, and inconsistencies across studies. 2. Electroacupuncture: Electroacupuncture has also shown promise in reducing cancer-related fatigue. Some trials have reported positive results, but the evidence is not yet strong enough to make definitive conclusions. 3. Moxibustion: Limited studies are available on the effectiveness of moxibustion for cancer-related fatigue, and the evidence is generally weak. More rigorous studies are needed to determine its potential benefits. 4. Herbal medicine: There is some evidence that certain herbal medicines may help reduce cancer-related fatigue, but the quality of the studies varies, and more research is needed to confirm their effectiveness. <p>It's important to remember that I'm an AI language model and not a medical professional. Before recommending any intervention, it's crucial to consult with a healthcare provider who can evaluate the patient's individual circumstances and consider potential risks and benefits. The strength of recommendations and the level of evidence for each intervention may differ depending on the source of information, such as clinical practice guidelines or systematic reviews. Always refer to the latest and most relevant guidelines or research for the most accurate information.</p>	<p>-(manual and electro) acupuncture was effective and safe on cancer-related fatigue treatment. However, further studies are still warranted by incorporating more large-scale and high-quality randomized controlled trials.”[7]</p> <p>“Because of a high risk of bias and low reporting quality of the studies included in this review, it is difficult to draw the conclusion that moxibustion is an effective and safe treatment for patients with cancer-related fatigue.”[6]</p> <p>“Compared to conventional medicine alone, herbal medicine combined with conventional medicine significantly improved fatigue level, quality of life, and activity of daily living. As monotherapy, herbal medicine significantly improved activity of daily living compared with megestrol. No serious herbal medicine-related adverse events were reported. Limited evidence suggests that HM could be effective and safe for cancer-related fatigue in lung cancer patients.”[5]</p>

* Quote from the conclusion of the review

tion of false answers that was previously identified as a concern with earlier ChatGPT models, was not observed in this instance.³

To investigate whether the information provided on CAM varies depending on the language of the question, we asked questions in English, Korean, Chinese, and Japanese. We found that ChatGPT-4 provides information with significant differences in the types and amounts of recommended interventions for CAM depending on the language (Supplement 2). Considering the health care environment of countries that primarily use these languages, it is necessary to examine whether the provided information is appropriate in the context of the relevant countries or whether the different responses are generated completely randomly.

3. Potential role of ChatGPT-4 in the CAM field

Many patients worldwide employ various types of CAM interventions. However, medical doctors often hesitate to discuss CAM treatments with their patients or refer them to CAM practitioners. This reluctance may stem from the difficulty in accessing information on evidence supporting the effectiveness and safety of CAM interventions or the lack of opportunities for appropriate education.^{8, 9} If reliable and easily accessible information about the effectiveness and safety of unfamiliar CAM interventions were properly provided to physicians, it would facilitate more in-depth conversations with their patients about CAM use and enable well-grounded referrals or the dissuasion of usage of CAM interventions when necessary.

Expert systems, such as computer software or web systems that can aid in patient diagnosis and treatment decision-making, have already been developed and are in use in the CAM field.^{10, 11} These are based on rule-based reasoning theory and model the relationships between individual patient symptoms, diagnosis, and treatment using expert-guided supervised learning methods.¹¹ CAM practitioners can receive some of the assistance previously obtained from expert systems by utilizing AI chatbot services. While it is possible to obtain meaningful information, such as potential prescription choices, it is crucial to consider the possibility of erroneous information (the hallucination issue), as suggested in the example case (Supplement 1), and the unique characteristics of

CAM practice. Unlike conventional medicine, treatment principles in CAM can vary depending on the intervention type, school of thought, and perspectives on humans and diseases. Consequently, it is essential to avoid applying this information uncritically in CAM clinical practice. Instead, it should serve as a tool to assist experts in their medical decision-making process. Expert systems, which generally rely on supervised learning methods, may be suitable for acquiring expertise in a specific field. However, their development requires time, cost, and expertise, which makes these expert systems less accessible. AI chatbots, in contrast, may not possess the same level of expertise as expert systems but have shown success in providing a broader scope of general information. Based on these considerations, it seems more appropriate for CAM practitioners to use AI chatbots to gain knowledge about CAM therapies in other unspecialized areas for consultation rather than relying on them to directly obtain the necessary knowledge for clinical practice.

From the perspective of health care consumers, AI chatbot services appear to be an easily accessible means to resolve information asymmetry. Similar to conventional medicine, patients are keen to know which CAM therapies can be applied to their conditions, which can be reflected in the information-seeking behavior of patients.¹² At this point, it is essential to pay attention to the potential of chatbot services as a tool for obtaining information on the effectiveness and safety of CAM treatments and for addressing questions about the safety of combining CAM with conventional treatments. While it would be inappropriate to consider such information critical for decision-making, there is value in providing bottom-line evidence for preliminary judgment in the context of health care and CAM therapies. This highlights the importance of ensuring that the information provided is accurate, reliable, and easily accessible to support informed choices.¹³ AI chatbots can become readily accessible tools for this purpose.

4. Limitations and areas for future improvement

Does ChatGPT-4 only bring hope to the field of CAM? There are still significant issues that need to be addressed. First and foremost is the

problem of hallucination. As mentioned in OpenAI's technical reports, while improvements have been made compared to previous versions, ChatGPT-4 still cannot be entirely trusted due to the possibility of providing incorrect information or exhibiting errors in the inference process. In medical decision-making, where decisions can directly impact a person's life, it is highly risky to rely solely on the judgment of AI. Expert review and cross-referencing multiple opinions will likely be necessary to ensure the accuracy and reliability of the information provided.¹

The next issue is the significant language gap. According to OpenAI's announcement, the results of the multitask language understanding (MMLU) benchmark test showed that GPT-4 had an accuracy of 84.1% for English, 77% for Korean, and 62% for Telugu.¹ As demonstrated in our previous example, different answers are generated depending on the language of the question, and it is unclear whether these answers accurately reflect the health care situation in the respective regions where the language is used (Supplementary file 2). This suggests that there may still be disparities in the accuracy of information depending on the user's language. Additionally, there is uncertainty about whether the training of ChatGPT-4 itself for CAM is sufficient. ChatGPT showed near-passing performance on the United States Medical Licensing Exam (USMLE).¹⁴ Similarly, a recent study examining the GPT-4's accuracy rate for the Korean National Licensing Examination for Korean Medicine Doctors found a 57.29% correct answer rate. Notably, it was reported that lower accuracy rates were observed for questions specifically focused on traditional Korean Medicine.¹⁵ Consideration should also be given to the perception of CAM practitioners regarding ChatGPT-4. According to a survey conducted among undergraduate students, the majority believe that the potential for AI to be used in clinical settings is not particularly high. One of the reasons cited for this belief is that students perceive AI to be limited in providing information in the CAM field.¹⁶

5. Conclusion

In this short commentary, we briefly used and narratively analyzed the performance of ChatGPT-4 for information about CAM. ChatGPT-4 appears to be successful in providing an overview of the evidence for representative CAM interventions. We believe that AI chatbot services can be utilized as a convenient tool for clinicians and health care consumers to obtain brief information about the effectiveness and safety of CAM. However, it is not possible to determine whether ChatGPT-4 has learned sufficient information about CAM to be used as a decision-making aid for CAM practitioners, as the expert systems initially aimed to achieve. Additionally, issues such as providing different information depending on the language of the question were observed. It would be desirable for AI chatbots to develop in the future as an easy and reliable method to access reproducible information through future technological improvement.

Author contributions

Conceptualization: THK. Formal investigation: JWK. Writing - original draft: THK. Writing - review and editing: THK and MSL.

Conflict of interest

THK and MSL are part of the editorial board for this journal. The authors declare no other conflicts of interest.

Funding

This commentary was funded by Korea Institute of Oriental Medicine (KSN1823211).

Ethical statement

Not applicable.

Data availability

Not applicable.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.imr.2023.100977.

Supplement 1. Comparison between ChatGPT-4's responses and existing CAM guideline recommendations on the detailed treatment methods for CAM practitioners: An example case

Supplement 2. Differences in ChatGPT-4 responses when asking the same question in different languages (English, Korean, Chinese and Japanese): an example case

References

1. Open AI_ (2023): GPT-4 Technical Report. <https://cdn.openai.com/papers/gpt-4.pdf> 2023.
2. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30.
3. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *N Engl J Med*. 2023;388(13):1233–1239.
4. Korean Medicine Clinical Practice Guideline for Cancer-Related Symptoms https://nikom.or.kr/engnckm/module/practiceGuide/view.do?guide_idx=203&progress=&mids_code=&disease_code=&gubun=INT&code_gubun=mds&agency=Korean+Association+of+Traditional+Oncology&continent=&sortField=&sortType=&language=eng&country=%2C&continent_str=&search_type=all&search_text=&viewPage=1&guide_idx=&progress_jq=&title=&disease_code_etc1=&agency_jq=&country=&cert_yn=&release_date=&menu_idx=4.
5. Kwon CY, Lee B, Kong M, et al. Effectiveness and safety of herbal medicine for cancer-related fatigue in lung cancer survivors: a systematic review and meta-analysis. *Phytother Res*. 2021;35(2):751–770.
6. Lee S, Jerng UM, Liu Y, Kang JW, Nam D, J-D Lee. The effectiveness and safety of moxibustion for treating cancer-related fatigue: a systematic review and meta-analysis. *Supportive Care in Cancer*. 2014;22:1429–1440.
7. Tian H, Chen Y, Sun M, et al. Acupuncture therapies for cancer-related fatigue: a Bayesian network meta-analysis and systematic review. *Front Oncol*. 2023;13.
8. Winslow LC, Shapiro H. Physicians want education about complementary and alternative medicine to enhance communication with their patients. *Arch Intern Med*. 2002;162(10):1176–1181.
9. Mildren SP, Stokols D. Physicians' attitudes and practices regarding complementary and alternative medicine. *Behav Med*. 2004;30(2):73–84.
10. Boulos MNK. Expert system shells for rapid clinical decision support module development: an ESTA demonstration of a simple rule-based system for the diagnosis of vaginal discharge. *Health Inform Res*. 2012;18(4):252–258.
11. Choi SH. Development of web-based diagnosis expert system of traditional oriental medicine. *J Physiol Pathol Kor Med*. 2002;16(3):528–531.
12. Weeks L, Balneaves LG, Paterson C, Verhoef M. Decision-making about complementary and alternative medicine by cancer patients: integrative literature review. *Open Med*. 2014;8(2):e54.
13. Woolf SH, Chan EC, Harris R, et al. Promoting informed choice: transforming health care to dispense knowledge for decision making. *Am Coll Phys*. 2005;143:293–300.
14. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digital Health*. 2023;2(2):e0000198.
15. Jang D., Kim C.E.: Exploring the Potential of Large Language models in Traditional Korean Medicine: a Foundation Model Approach to Culturally-Adapted Healthcare. arXiv preprint arXiv:230317807 2023.
16. Yang JH, Woo JA, Shin D-H, Park S, Kwon YK. Study on the Perception and Application of AI in Korean Medicine through Practice and Questionnaire of Korean Medicine Using a Diagnostic Expert System. *J Physiol Pathol Kor Med*. 2021;35(1):22–27.