

Assessment of the Reliability and Clinical Applicability of ChatGPT's Responses to Patients' Common Queries About Rosacea

Sihan Yan^{1,2,*}, Dan Du^{1,2,*}, Xu Liu^{1,2}, Yingying Dai^{1,2}, Min-Kyu Kim^{1,2}, Xinyu Zhou³, Lian Wang^{1,2}, Lu Zhang^{1,2}, Xian Jiang^{1,2}

¹Department of Dermatology, West China Hospital, Sichuan University, Chengdu, People's Republic of China; ²Laboratory of Dermatology, Clinical Institute of Inflammation and Immunology, Frontiers Science Center for Disease-Related Molecular Network, West China Hospital, Sichuan University, Chengdu, People's Republic of China; ³Department of Dermatology, Nanbu County People's Hospital, Nanbu County, Nanchong, Sichuan, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xian Jiang; Lu Zhang, Email jiangxian@scu.edu.cn; zhangluhx@scu.edu.cn

Objective: Artificial intelligence chatbot, particularly ChatGPT (Chat Generative Pre-trained Transformer), is capable of analyzing human input and generating human-like responses, which shows its potential application in healthcare. People with rosacea often have questions about alleviating symptoms and daily skin-care, which is suitable for ChatGPT to respond. This study aims to assess the reliability and clinical applicability of ChatGPT 3.5 in responding to patients' common queries about rosacea and to evaluate the extent of ChatGPT's coverage in dermatology resources.

Methods: Based on a qualitative analysis of the literature on the queries from rosacea patients, we have extracted 20 questions of patients' greatest concerns, covering four main categories: treatment, triggers and diet, skincare, and special manifestations of rosacea. Each question was inputted into ChatGPT separately for three rounds of question-and-answer conversations. The generated answers will be evaluated by three experienced dermatologists with postgraduate degrees and over five years of clinical experience in dermatology, to assess their reliability and applicability for clinical practice.

Results: The analysis results indicate that the reviewers unanimously agreed that ChatGPT achieved a high reliability of 92.22% to 97.78% in responding to patients' common queries about rosacea. Additionally, almost all answers were applicable for supporting rosacea patient education, with a clinical applicability ranging from 98.61% to 100.00%. The consistency of the expert ratings was excellent (all significance levels were less than 0.05), with a consistency coefficient of 0.404 for content reliability and 0.456 for clinical practicality, indicating significant consistency in the results and a high level of agreement among the expert ratings.

Conclusion: ChatGPT 3.5 exhibits excellent reliability and clinical applicability in responding to patients' common queries about rosacea. This artificial intelligence tool is applicable for supporting rosacea patient education.

Keywords: artificial intelligence, ChatGPT, rosacea, patient education

Introduction

In the last decade, deep learning (DL) and other artificial intelligence (AI) technologies have made rapid progress.¹ Based on novel AI techniques, Augello et al developed a social chatbot model capable of integrating both individual and social processes.² This pioneering approach inspired researchers to dig out the potentials of AI chatbots, especially in the area of healthcare. Chung et al introduced a chatbot-based application, which aimed to provide timely assistance to these patients with chronic diseases.³ AI chatbots can provide numerous advantages, including enhanced patient engagement, improved diagnostic accuracy, and personalized treatment plans and facilitating the assimilation of the latest medical literature into clinical practice.⁴ As powerful as AI chatbots can be, there are also concerns about the limitations of them being medical chatbots, including the accuracy and reliability of the medical information they provide, the transparency of the model, and the

ethics of making use of user information.⁵ Chat Generative Pre-trained Transformer (ChatGPT) series is a highly discussed one of those artificial intelligence tools developed by OpenAI. It is based on natural language processing (NLP) models. It utilizes deep learning to provide coherent and fluent human-like text responses to complex queries. Other significant AI chatbots, such as Google Med-PaLM are also being explored for medical applications. Also using chatbot developing tools, including IBM Watson (IBM Corp) assistant, some researchers are constructing other specialized AI chatbots focused on certain specified fields of medicine.^{6,7} In certain fields of medicine, the benefits of applying ChatGPT as a patient education tool have been evaluated, and they helped resolve some certain concerns about AI chatbot limitations.^{8–11}

Rosacea is a chronic and recurrent inflammatory disease primarily affecting the facial blood vessels, nerves, and pilosebaceous units in the central region of the face, affecting approximately 5.46% of adults.¹² As a chronic inflammatory skin disease, rosacea patients often faces long-term treatment. Meanwhile, patient education is an important component of rosacea management, especially referring to the potential triggers such as diet habit and skin care.^{13,14} However, patients may not get enough suggestions from dermatologists due to several reasons. Nowadays, patients often use social media and other platforms to expand their understanding of the disease.¹⁵ Physicians can facilitate patient education by leveraging the wealth of online resources. As a novel chatbot, ChatGPT showed its potential in medicine.^{4,11,16} In urgent need of a reliable out-of-clinic rosacea patient education tool, in order to assess the extent of dermatological resource coverage in ChatGPT, we assessed the reliability and clinical applicability of ChatGPT 3.5 in responding to patients' common queries about rosacea.

Materials and Methods

Objective

We selected ChatGPT 3.5 as the tool to generate responses to patients' common queries about rosacea. ChatGPT is powered by an enhanced version of the large language model (LLM) GPT-3.5, which has been trained and fine-tuned on a variety of sources such as articles, websites, books, and written conversations, including dialogue optimization, enabling ChatGPT to generate response in a conversational format.¹⁷

Methods

Alinia et al conducted a qualitative analysis using a stratified sampling method on approximately 2000 messages from online rosacea forums to understand patients' habits of seeking health information and their educational needs.¹⁸ We extracted 20 questions of patients' greatest concerns, covering four main categories: treatment, triggers and diet, skincare, and special manifestations of rosacea. Each question was inputted into ChatGPT separately for three rounds of question-and-answer conversations. The generated answers will be evaluated by three experienced dermatologists with postgraduate degrees and over five years of clinical experience in dermatology, to assess their reliability and applicability for clinical practice. The reliability of the answers was rated on a scale from 1 (completely unreliable) to 5 (completely reliable), while the applicability of the answers was rated on a scale from 1 (not applicable) to 3 (highly applicable). This study focused on a publicly available AI tool, without involving patient information or clinical biosamples. The research did not involve any commercial interests or issues related to intellectual property. Ethical review by the Ethics Committee of West China Hospital, Sichuan University, determined that clinical ethics review was not required.

Statistics

This study used a five-point scale and a three-point scale to assess the reliability and the clinical applicability of the generated content, respectively. The evaluation results were obtained using average percentages. The consistency of expert ratings was tested using Kendall's coefficient of concordance.

Results

This study showed that the reviewers unanimously agreed that ChatGPT had a high reliability responding to patients' common queries about rosacea, which 94.6% reliability in general and range from 92.22% to 97.78% among four categories (Table 1). Specifically, it achieved 95.28% reliability in the treatment category, 97.78% reliability in the

Table 1 Percentage of Reliability and Clinical Applicability Assessment for the Four-Category Classification

Category	Reliability	Clinical applicability
Treatment	95.28%	98.61%
Triggers and diet	97.78%	100.00%
Skin care	92.78%	99.07%
Special manifestations of rosacea	92.22%	99.07%

triggers and diet category, 92.78% reliability in the skincare category, and 92.22% reliability in the special manifestations of rosacea category. Among all 20 queries, the reliability of the ChatGPT ranges from 80.00% to 100.00%. The lowest reliability of ChatGPT was shown in the query “Are the medications used to treat skin rosacea effective for eye symptoms?”, and the highest reliability was shown in eight queries (Table 2).

Additionally, almost all responses were considered applicable for supporting rosacea patient education. In general, the mean clinical applicability of ChatGPT was 99.1%, with a clinical applicability ranging from 98.61% to 100.00%, specifically with a 98.61% applicability rate in the treatment category, a 100.00% applicability rate in the triggers and diet category, a 99.07% applicability rate in the skincare category, and a 99.07% applicability rate in the special manifestations of rosacea category (Table 1). Among all 20 queries, the clinical applicability ranged from 92.59% to 100%. The lowest reliability of ChatGPT was shown in the query “How to manage adverse reactions during rosacea treatment?”, and the highest reliability was shown in 16 queries (Table 2).

The consistency of the expert ratings was excellent (all significance levels were below 0.05). The consistency coefficient for content reliability was 0.404, and for clinical applicability, it was 0.456, indicating a significant consistency in the results. The expert ratings demonstrated a high level of agreement (Table 3).

Table 2 ChatGPT Underwent a Percentage-Based Evaluation of Content Reliability and Clinical Applicability for 20 Common Questions Related to Rosacea

Category	Question	Reliability	Clinical Applicability
Treatment	How to alleviate the persistent erythema of rosacea?	95.56%	100.00%
	How to manage adverse reactions during rosacea treatment?	88.89%	92.59%
	How to avoid the side effects of long-term oral antibiotic (such as doxycycline) treatment for rosacea?	93.33%	100.00%
	What kind of results can be expected from intense pulsed light therapy for rosacea?	91.11%	100.00%
	How effective are the treatments for different subtypes of rosacea?	95.56%	100.00%
	What level of efficacy can be achieved with current rosacea treatments?	97.78%	96.30%
	How can one stay updated on the clinical progress of rosacea in a timely manner?	100.00%	100.00%
	How can one determine the current stage of rosacea?	100.00%	100.00%
Triggers and diet	Triggers of rosacea?	91.11%	100.00%
	What dietary considerations should be taken into account for rosacea?	100.00%	100.00%
	What is the impact of alkaline and acidic foods on rosacea?	100.00%	100.00%
	Are vitamins or other nutritional supplements beneficial for treating rosacea?	100.00%	100.00%
Skin care	Does using skincare products worsen the skin symptoms of rosacea?	86.67%	96.30%
	How to choose a moisturizer for rosacea?	86.67%	100.00%
	Which brand of cosmetics can be used for rosacea?	97.78%	100.00%
	Are the currently popular skincare products helpful for rosacea?	100.00%	100.00%
Special manifestations of rosacea	What is the correct approach to treating ocular rosacea?	88.89%	100.00%
	Are the medications used to treat skin rosacea effective for eye symptoms?	80.00%	96.30%
	What should be done if someone is socially isolated due to rhinophyma rosacea?	100.00%	100.00%
	Does rosacea have any comorbidities with other systems, such as digestive system diseases?	100.00%	100.00%

Table 3 Consistency Test Results

Statistical Testing	Reliability	Clinical Applicability
Number of cases	6	3
Kendall coefficient of concordance	0.404	0.456
Chi-square test	143.068	80.64
Degree of freedom	59	59
Asymptotic significance	0.000	0.000

Discussion

ChatGPT, as a cutting-edge AI tool, achieves an impressive reliability of 92.22% to 97.78% in responding to patients' common queries about rosacea. In comparison to similar studies, ChatGPT demonstrates higher reliability in responding to common queries about basal cell carcinoma (84% reliability), breast cancer prevention (88% reliability), and cardiovascular disease prevention (84% reliability).^{19–21} The consistency coefficient in this study for content reliability was 0.404 and 0.456 for clinical applicability. These results were comparable with related studies. For example, Lakdawala et al evaluated the accuracy of ChatGPT in offering clinical guidance for atopic dermatitis and acne vulgaris. The weighted Cohen's kappa coefficient was 0.44 in their study, which indicated moderate agreement among the board-certified dermatologists.²² ChatGPT exhibits strong advantages in rosacea patient education, particularly in addressing concerns on triggers and dirt, showing both high reliability and clinical applicability. By comparing the results of three groups of ratings, few responses were consistently considered inappropriately to apply in clinical practice. This is usually attributed to missing key details or excessive complexity in the content. For example, ChatGPT's response to "How to manage adverse reactions during rosacea treatment?" was generally accurate, yet it covered content too complex and did not cater to common clinical cases, and it was not as concisely as advice from clinical doctors. This also demonstrates the exceptional reliability and clinical applicability of ChatGPT in responding to patients' common queries about rosacea. Our study still has several limitations. It is worth mentioning that while OpenAI has introduced the latest and more advanced version of the AI tool, ChatGPT 4.0, the fee-charging service may make it less approachable for non-professional users, including rosacea patients. We will evaluate the potential application of ChatGPT 4.0 in rosacea when it is more accessible for the public. With the continuous evolution of this AI chatbot tool, the latest and more coming versions might become more economically feasible for common users and the assessment awaits. This AI tool is applicable for supporting rosacea patient education. While ChatGPT has its advantages, there are also concerns about its use as a medical chatbot. One concern is the accuracy and reliability of the medical information provided by ChatGPT, as it is not a licensed medical professional with abundant clinical experience, as well as it may not have access to up-to-date medical knowledge. Another concern is about the transparency of the chatbot model and the ethics of making use of user information, as well as the potential for biases in the data used to train ChatGPT's algorithms. Thus, it is important to carefully consider the potential risks and benefits of using ChatGPT as a medical chatbot, especially in those patients with limited medical knowledge.⁵ Further studies should focus on improving the reliability and clinical application of AI chatbots with up-to-date information as well as advanced algorithms. Meanwhile, limitations of these chatbots should be clear of answering medical queries so that the user can identify when they need to see medical professionals.

In conclusion, ChatGPT 3.5 exhibits excellent reliability and clinical applicability in responding to patients' common queries about rosacea. This AI tool is applicable for supporting rosacea patient education.

Funding

The National Natural Science Foundation of China (82273559); Clinical Research Innovation Project, West China Hospital, Sichuan University(19HXCX010).

Disclosure

Siyan Yan and Dan Du are co-first authors for this study. All authors declare no conflicts of interest in this work.

References

1. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature*. 2015;518(7540):529–33.
2. Augello A, Gentile M, Weideveld L, Dignum F. A Model of a Social Chatbot. In: Pietro GD, Gallo L, Howlett RJ, Jain LC, editors. *Intelligent Interactive Multimedia Systems and Services 2016*. Springer International; 2016:637–647.
3. Chung K, Park RC. Chatbot-based healthcare service with a knowledge base for cloud computing. *Cluster Comput*. 2019;22(Suppl 1):1925–1937. doi:10.1007/s10586-018-2334-5
4. Goktas P, Karakaya G, Kalyoncu AF, Damadoglu E. Artificial intelligence chatbots in allergy and immunology practice: where have we been and where are we going? *J Allergy Clin Immunol Pract*. 2023;11(9):2697–2700. doi:10.1016/j.jaip.2023.05.042
5. Chow JCL, Sanders L, Li K. Impact of ChatGPT on medical chatbots as a disruptive technology. *Front Artif Intell*. 2023;6:1166014. doi:10.3389/frai.2023.1166014
6. Chow JCL, Wong V, Sanders L, Li K. Developing an AI-assisted educational chatbot for radiotherapy using the IBM Watson assistant platform. *Healthcare*. 2023;11(17):2417. doi:10.3390/healthcare11172417
7. Rebelo N, Sanders L, Li K, Chow JCL. Learning the Treatment Process in Radiotherapy Using an Artificial Intelligence-Assisted Chatbot: development Study. *JMIR Form Res*. 2022;6(12):e39443. doi:10.2196/39443
8. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of artificial intelligence chatbot responses to top searched queries about cancer. *JAMA Oncol*. 2023;9(10):1437–1440. doi:10.1001/jamaoncol.2023.2947
9. Musheyev D, Pan A, Loeb S, Kabarriti AE. How well do artificial intelligence chatbots respond to the top search queries about urological malignancies? *Eur Urol*. 2024;85(1):13–16. doi:10.1016/j.eururo.2023.07.004
10. Young JN, Ross O, Poplasky D, et al. The utility of ChatGPT in generating patient-facing and clinical responses for melanoma. *J Am Acad Dermatol*. 2023;89(3):602–604. doi:10.1016/j.jaad.2023.05.024
11. Goktas P, Kucukkaya A, Karacay P. Leveraging the efficiency and transparency of artificial intelligence-driven visual Chatbot through smart prompt learning concept. *Skin Res Technol*. 2023;29(11):e13417. doi:10.1111/srt.13417
12. Gether L, Overgaard LK, Egeberg A, Thyssen JP. Incidence and prevalence of rosacea: a systematic review and meta-analysis. *Br J Dermatol*. 2018;179(2):282–289. doi:10.1111/bjd.16481
13. Two AM, Wu W, Gallo RL, Hata TR. Rosacea: part II. Topical and systemic therapies in the treatment of rosacea. *J Am Acad Dermatol*. 2015;72(5):761–70; quiz 771–2. doi:10.1016/j.jaad.2014.08.027
14. Abram K, Silm H, Maaroos HI, Oona M. Risk factors associated with rosacea. *J Eur Acad Dermatol Venereol*. 2010;24(5):565–571. doi:10.1111/j.1468-3083.2009.03472.x
15. Cardwell LA, Nyckowski T, Uwakwe LN, Feldman SR. Coping mechanisms and resources for patients suffering from rosacea. *Dermatol Clin*. 2018;36(2):171–174. doi:10.1016/j.det.2017.11.013
16. Goktas P, Agildere AM. Transforming radiology with artificial intelligence visual chatbot: a balanced perspective. *J Am Coll Radiol*. 2023;S1546–1440(23):643.
17. The Lancet Digital Health. ChatGPT: friend or foe? *Lancet Digit Health*. 2023;5:3.
18. Alinia H, Moradi Tachayi S, Farhangian ME, et al. Rosacea patients seeking advice: qualitative analysis of patients' posts on a rosacea support forum. *J Dermatol Treat*. 2016;27(2):99–102. doi:10.3109/09546634.2015.1133881
19. Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. *JAMA*. 2023;329(10):842–844. doi:10.1001/jama.2023.1044
20. Haver HL, Ambinder EB, Bahl M, Oluyemi ET, Jeudy J, Yi PH. Appropriateness of breast cancer prevention and screening recommendations provided by ChatGPT. *Radiology*. 2023;307(4):e230424. doi:10.1148/radiol.230424
21. Trager MH, Queen D, Bordone LA, Geskin LJ, Samie FH. Assessing ChatGPT responses to common patient queries regarding basal cell carcinoma. *Arch Dermatol Res*. 2023;315(10):2979–2981. doi:10.1007/s00403-023-02705-3
22. Lakdawala N, Channa L, Gronbeck C, et al. Assessing the accuracy and comprehensiveness of ChatGPT in offering clinical guidance for atopic dermatitis and acne vulgaris. *JMIR Dermatol*. 2023;6:e50409. doi:10.2196/50409

Patient Preference and Adherence

Dovepress

Publish your work in this journal

Patient Preference and Adherence is an international, peer-reviewed, open access journal that focusing on the growing importance of patient preference and adherence throughout the therapeutic continuum. Patient satisfaction, acceptability, quality of life, compliance, persistence and their role in developing new therapeutic modalities and compounds to optimize clinical outcomes for existing disease states are major areas of interest for the journal. This journal has been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/patient-preference-and-adherence-journal>