

Utility of Artificial Intelligence in Patient Education: An Appraisal of Mohs Surgery Information Generated by ChatGPT-3.5 and ChatGPT-4

Kennedy H. Sun BS,^a William Liakos MD,^b Vishal A. Patel MD^a

^aDepartment of Dermatology, George Washington University School of Medicine and Health Sciences, Washington, DC

^bDepartment of Dermatology, Kansas University Medical Center, Kansas City, KS

INTRODUCTION

Patients utilize the Internet to pursue health knowledge and better serve as advocates for their care. This frequently involves online searches, where information can often differ from guidelines,¹ and increasingly, artificial intelligence (AI) systems such as ChatGPT. Recently, ChatGPT has demonstrated the ability to pass the USMLE,² potentially reduce dermatologist workload,³ and communicate well with patients regarding common medical questions.⁴ However, ChatGPT and other AI systems have not been evaluated for the accuracy of their dermatology-related information. We investigated the utility of two of these AI systems, ChatGPT-3.5

(freely accessible, <https://chat.openai.com>) and ChatGPT-4 (paid premium version of ChatGPT), in providing information on Mohs surgery.

We input 12 common patient questions about Mohs surgery into GPT-3.5 and GPT-4 (Table 1). Questions were based on Mohs surgery pamphlets for patients at multiple dermatologic clinics and outcome measures align with recent literature assessing ChatGPT responses; both were reviewed by a board-certified Mohs surgeon for relevancy prior to survey administration. Answers were evaluated by five dermatologists, including four

TABLE 1.

Selection of Sample Patient Questions and ChatGPT-3.5 and ChatGPT-4 Responses

Question	ChatGPT-3.5 Response	ChatGPT-4 Response
4. How do I know if I need Mohs surgery over other skin cancer removal surgeries?	The decision of whether you need Mohs surgery or another skin cancer removal procedure depends on several factors, including the type, size, location, and aggressiveness of the skin cancer, as well as your medical history and overall health. Mohs surgery is often recommended in the following scenarios: 1) Skin cancers with poorly defined borders, 2) Large or aggressive tumors, 3) Skin cancers in areas where preserving healthy tissue and cosmetic appearance are critical (eg, face, ears, nose), 4) Skin cancers that have recurred after previous treatments, and 5) Skin cancers in areas where there is a higher risk of cancer spreading or where preserving function is vital. Your dermatologist or skin cancer specialist will assess these factors and recommend the most appropriate treatment option for your specific case, ensuring the best chance of successful cancer removal while minimizing the impact on surrounding healthy tissue.	Whether you need Mohs surgery over other skin cancer removal methods depends on several factors. Mohs surgery is often recommended for skin cancers located in cosmetically sensitive or functionally crucial areas, like the face, hands, or feet, because it conserves the maximum amount of healthy tissue. It's also favored for aggressive, large, or ill-defined tumors, recurrent cancers, or cancers growing in scarred areas. The tumor's type, size, location, and growth pattern, as well as any previous treatments, play a role in this decision. A dermatologist or surgical oncologist will evaluate your specific case and recommend the most suitable treatment. While Mohs has distinct advantages, it's essential to have a personalized consultation to determine the best surgical approach for your situation.
11. What is postoperative recovery like after getting Mohs surgery?	Postoperative recovery after Mohs surgery typically involves minimal discomfort and a relatively short healing period. The wound will be dressed, and you may experience some soreness or swelling in the treated area. Most patients can resume normal activities the day after the procedure, but it's advised to avoid strenuous exercises and heavy lifting for a few days. The wound will gradually heal over the next weeks to months, and any stitches used will be removed during a follow-up visit. Proper wound care and sun protection are essential during this time to optimize healing and minimize scarring. Regular follow-up appointments with your surgeon will monitor your progress and address any concerns for a successful recovery.	Postoperative recovery after Mohs surgery varies depending on the surgery's size and location but generally involves a few common experiences. Patients may experience mild to moderate pain, swelling, and bruising around the surgical site, which usually subsides within a few days to a week. Proper wound care, as instructed by the surgeon, is crucial: this often involves keeping the area clean, dry, and covered, and applying antibiotic ointment if recommended. Stitches, if used, are typically removed within a week or two. Activities that might strain the surgical area should be avoided for the initial days post-surgery. It's essential to protect the wound from direct sunlight to reduce scar darkening. Any signs of complications, like unexpected pain or signs of infection, should be promptly reported to the healthcare provider.

TABLE 2.

Average Accuracy, Sufficiency, and Readability of ChatGPT-3.5 and ChatGPT-4 Responses, Input in August 2023				
Question	Model Examined	Average Accuracy	Average Sufficiency for Patient Education	Flesch Kincaid Grade Level
1. What is Mohs surgery and how is it performed?	GPT-3.5	4.8	4.2	13.9
	GPT-4	4.4	4.4	13.4
2. How is Mohs surgery different from other skin cancer removal surgeries?	GPT-3.5	4.4	3.6	17.3
	GPT-4	4.4	4.0	16.0
3. How safe and effective is Mohs surgery?	GPT-3.5	4.4	3.4	16.1
	GPT-4	4.8	4.2	15.0
4. How do I know if I need Mohs surgery over other skin cancer removal surgeries?	GPT-3.5	4.8	4.6	17.0
	GPT-4	4.2	4.0	14.6
5. Does everyone with skin cancer need Mohs surgery?	GPT-3.5	4.8	4.4	17.9
	GPT-4	4.8	4.6	14.4
6. Is there a way to avoid Mohs surgery if I don't want to have a procedure?	GPT-3.5	4.6	4.0	15.2
	GPT-4	4.4	4.6	14.3
7. How should I prepare for Mohs surgery?	GPT-3.5	2.8	2.4	12.7
	GPT-4	3.8	3.8	13.2
8. What to expect on the day of my Mohs surgery procedure and how long will it take?	GPT-3.5	4.0	3.4	11.6
	GPT-4	3.6	3.2	9.6
9. Will I experience pain during or after Mohs surgery?	GPT-3.5	3.4	3.2	12.6
	GPT-4	3.8	3.8	11.7
10. How common are Mohs surgery complications and how can I reduce risk for complications?	GPT-3.5	3.6	3.6	14
	GPT-4	3.6	3.4	13.3
11. What is postoperative recovery like after getting Mohs surgery?	GPT-3.5	4.0	3.8	13.1
	GPT-4	3.6	3.6	12.8
12. When will I be fully recovered from Mohs surgery and can resume my routine activities?	GPT-3.5	4.2	4.0	12.2
	GPT-4	4.6	4.4	13.1

Responses were limited to less than 150 words which was a limitation input into the initial prompt for ChatGPT and consistent with Mohs surgery pamphlet information level of detail. Accuracy was assessed using a scale: 1=Completely inaccurate, 2=Somewhat inaccurate, 3=Neither accurate nor inaccurate, 4=Somewhat accurate, 5=Completely accurate. Sufficiency for patient education was scored using a scale: 1=Inaccurate and missing lots of critical information, 2= Somewhat inaccurate and missing some critical information, 3= Somewhat accurate and missing some non-critical information, 4= Accurate but missing some non-critical information, 5=Complete and accurate, no additional information needed from the clinician. Flesch Kincaid Grade Level equates to a grade level with 1-5 equivalent to 1st- 5th grade and easy to read, 5-11 equivalent to 5th- 11th grade and average level, 11-18 equivalent to 11th-18th grade and skilled reader.

board-certified Mohs surgeons, for accuracy from 1 (completely inaccurate) to 5 (completely accurate) and sufficiency for patient education from 1 (inaccurate and missing lots of critical information) to 5 (complete and accurate, no additional information needed from the clinician). Dermatologists were blinded to whether responses were generated by GPT-3.5 or GPT-4. The Flesch Kincaid Grade Level (FKGL) was used to assess the readability of the generated responses.

The ChatGPT systems delivered responses of varying accuracy, readability, and technical detail (Table 2). On average, reviewers rated the accuracy of GPT-3.5 responses as 4.15 (Fleiss's $\kappa=.113$, $P=.102$, Standard Deviation [SD]=0.96) and sufficiency for patient education as 3.72 (Fleiss's $\kappa=.094$, $P=.10$, SD=0.90), although there was low agreement. For GPT-4, the accuracy and sufficiency were 4.17 (Fleiss's $\kappa=.021$, $P=.744$, SD=1.18) and

3.97 (Fleiss's $\kappa=.062$, $P=.831$, SD=0.92), respectively. However, there was no significant difference observed in average ratings for GPT-3.5 versus GPT-4 for accuracy ($t(22)=-.074$, $P=.942$), nor sufficiency ($t(22)=-1.31$, $P=.102$).

Most responses (70.8%) included a referral for patients to speak with a physician for further details or advice which suggests that AI cannot generate the same level of individualized answers as physicians. The average readability of answers by FKGL, which equates to grade level, was 14.47 (SD=2.17) for GPT-3.5 and 13.45 (SD=1.66) for GPT-4. Limitations include the nature of early-stage technology, the small number of raters, and the absence of information post-2021 within the ChatGPT language models. Moreover, GPT-3.5 and GPT-4 outputs are dynamic; a singular comparison is insufficient to fully understand significant differences in responses.

ChatGPT will likely be utilized by individuals interested in obtaining medical information, particularly for specialized treatments such as Mohs surgery.⁴ Fortunately, ChatGPT appears to be a tentatively acceptable option to answer common questions about Mohs surgery, although there are differing opinions about response adequacy. Precision is lacking as there is variability between the models' answers to the same question. It is encouraging that answers did not differ significantly between the free and premium versions, which suggests that AI systems might not exacerbate inequities in access to health information. Patients benefit from access to multiple modalities of education regarding Mohs,⁵ and AI systems can potentially help patients to better conceptualize Mohs surgery. Mohs surgeons must be aware of the patient's potential use and how it can best be incorporated into practice.

DISCLOSURES

VP has received honoraria from Regeneron, Sanofi, Almirall, InflaRx, Biofrontera, and Sun Pharma; is a member of the speaker's bureau for Regeneron and Sun Pharma; and holds stock options in Lazarus AI. The other authors have no conflicts of interest to disclose.

REFERENCES

1. Villa-Ruiz C, Kassamali B, Mazori DR, et al. Overview of TikTok's most viewed dermatologic content and assessment of its reliability. *J Am Acad Dermatol.* 2021;85(1):273-274. doi:10.1016/j.jaad.2020.12.028
2. KungTH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *medRxiv.* 2022:2022.12.19.22283643. doi:10.1101/2022.12.19.22283643
3. Diamond C, Rundle CW, Albrecht JM, et al. Chatbot utilization in dermatology: a potential amelioration to burnout in dermatology. *Dermatol Online J.* 2022;28(6)doi:10.5070/d328659734
4. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med.* 2023;183(6):589-596. doi:10.1001/jamainternmed.2023.1838
5. Patel P, Malik K, Khachemoune A. Patient education in Mohs surgery: a review and critical evaluation of techniques. *Arch Dermatol Res.* 2021;313(4):217-224. doi:10.1007/s00403-020-02119-5

AUTHOR CORRESPONDENCE

Vishal A Patel MD

E-mail:..... vpatel@mfa.gwu.edu