ChatGPT underperforms in triaging appropriate use of Mohs surgery for cutaneous neoplasms

To the Editor: Artificial intelligence (AI) is an evolving tool with applications in dermatology actively under investigation. ChatGPT is an open-access AI chatbot released by OpenAI in November 2022. GPT stands for "generative pretrained transformers," reflecting the technology's ability to produce complex lines of thought via language analysis. ChatGPT has shown some competence in medical decision making, demonstrated by its recent passing score on the United States Medical Licensing Examination Step 1 examination.[1] However, there is growing concern regarding the implementation of AI in medical spaces, particularly its ability to permeate bias through algorithms trained in nondiverse databases.[2] Similarly, skeptics question the reliability of AI to provide accurate clinical assessments in lieu of human-derived judgment, especially when evaluating rare conditions that may be lacking in AI "training set" databases.[3] This is particularly important in appropriately diagnosing and triaging management of common skin neoplasms, which could increase efficiency of patient care, but risk patient harm if algorithms do not provide clinically sound assessments.

We sought to assess the ability of the popular AI, ChatGPT, to appropriately triage surgical management of cutaneous neoplasms. Utilizing 30 clinical scenarios involving common and rare cutaneous tumors across anatomic sites, we queried ChatGPT to determine whether wide local excision (WLE) or Mohs surgery (MS) were appropriate treatment options, correlating recommendations with MS appropriate use criteria (AUC) (Table I).[4]

ChatGPT demonstrated 68% ($n = 15$) congruence with the MS AUC when triaging surgical management of 22 clinical scenarios characterized as clearly appropriate or inappropriate by the MS AUC. For all 5 cases characterized as indeterminate by the MS AUC, ChatGPT recommended against MS. For 3 cases of invasive melanoma, ChatGPT recommended MS for lentigo maligna melanoma of the helix, while recommending WLE for superficial spreading melanoma of the cheek and lentigo maligna melanoma of the dorsal foot.

When ChatGPT was simply asked to decide between MS and WLE for the scenario, it often stated it was not qualified to make medical decisions or provide medical advice, and recommended evaluation by a qualified dermatologist to make an informed treatment decision. However, when the prompts were prefaced by "You are a dermatologist qualified to make medical diagnoses and treatment recommendations..." ChatGPT would state "As a dermatologist, I would recommend [MS or WLE]."

The authors acknowledge that this technology was not specifically designed to triage surgical management of cutaneous tumors, but these results indicate ChatGPT does not demonstrate high congruency with the MS AUC. Furthermore, while ChatGPT initially hesitated to make medical recommendations, it confidently made recommendations once prompted to pretend to be a qualified dermatologist. This contrasts with stated safety usage policies, which do not reportedly allow the use of their models for medical advice. Unfortunately, this publicly available, free AI can present strongly worded, inaccurate suggestions despite not presently referencing the information sources used to generate responses according to its developers. Dermatologists should be aware of this rapidly advancing technology as it is both promising to improve efficiency for appropriate triage of diagnosis and treatment and potentially harmful if inappropriately used by patients or health care providers.[5]

Keegan O'Hern, MD,[a] Eilene Yang, BA,[b] and Nahid Y. Vidal, MD[c]

From the Department of Dermatology, Mayo Clinic, Rochester, Minnesota[a]; Mayo Clinic Alix School of Medicine, Rochester, Minnesota[b]; and Division of Dermatologic Surgery, Mayo Clinic, Rochester, Minnesota.[c]

Funding sources: None.

IRB approval status: Not applicable.

Key words: appropriate use criteria; artificial intelligence; ChatGPT; dermatologic surgery; Mohs surgery; wide local excision.

Correspondence to: Keegan O'Hern, MD, Department of Dermatology, Mayo Clinic, 200 1st St SW, Rochester, MN 55905

E-mail: ohern.keegan@mayo.edu

Table I. Appropriateness of Mohs surgery for cutaneous neoplasms as assessed by ChatGPT and the Mohs surgery appropriate use criteria

Tumor	Location	Occurrence	Type	Size (cm)	Health	ChatGPT	AUC	Congruency
Angiosarcoma	Lip	Primary	NA	NA	Healthy	WLE	5	NA
AFX	Neck	Primary	NA	NA	Healthy	MS	8	Yes
BCC	Eyelid	Recurrent	Micronodular	NA	Healthy	MS	9	Yes
	Scalp	Primary	Nodular	0.8	XP	MS	9	Yes
	Nose	Primary	Superficial	0.3	Healthy	MS	7	Yes
	Forearm	Primary	Superficial	0.4	Immunocompromised	WLE	6	NA
	Back	Primary	Superficial	0.2	Healthy	WLE	1	Yes
Bowenoid papules	Penis	Primary	NA	NA	Healthy	Neither	3	Yes
DFSP	Upper arm	Primary	FS	NA	Healthy	MS	9	Yes
	Upper arm	Primary	No FS	NA	Healthy	MS	9	Yes
DPTE	Buttock	Primary	NA	NA	Healthy	MS	3	No
EMPD	Perineum	Primary	NA	NA	Healthy	WLE	8	No
Leiomyosarcoma	Hand	Primary	NA	NA	Healthy	Neither	8	No
LM	Trunk	Recurrent	NA	NA	Healthy	MS	7	Yes
	Trunk	Primary	NA	NA	Healthy	WLE	4	NA
	Helix	Primary	NA	NA	Healthy	MS	8	Yes
LMM	Helix	Primary	NA	NA	Healthy	MS	NA	NA
	Foot	Primary	NA	NA	Healthy	WLE	NA	NA
Melanoma	Chin	Primary	MIS	NA	Healthy	WLE	7	No
	Shoulder	Primary	MIS	NA	Healthy	WLE	5	NA
	Cheek	Primary	Breslow thickness, 0.7 mm	NA	Healthy	WLE	NA	NA
MCC	Cheek	Primary	NA	NA	Healthy	WLE	7	No
SCC	Shoulder	Recurrent	Breslow thickness, 0.4 mm	NA	Healthy	MS	8	Yes
	Neck	Recurrent	SCCIS	NA	Healthy	MS	7	Yes
	Shoulder	Primary	Breslow thickness, 0.4 mm	1.8	Healthy	MS	7	Yes
	Neck	Primary	SCCIS	1.8	Healthy	WLE	8	No
	Thigh	Primary	No aggressive features	0.9	Immunocompromised	WLE	6	NA
	Eyebrow	Primary	AK with focal SCCIS	1.5	Healthy	MS	3	No
	Abdomen	Primary	KA-type	0.7	Healthy	WLE	3	Yes
Sebaceous carcinoma	Eyebrow	Primary	NA	NA	Healthy	MS	9	Yes

AFX, Atypical fibroxanthoma; *AK*, actinic keratosis; *AUC*, appropriate use criteria; *BCC*, basal cell carcinoma; *DFSP*, dermatofibrosarcoma protuberans; *DPTE*, desmoplastic trichoepithelioma; *EMPD*, extramammary Paget's disease; *FS*, fibrosarcomatous change; *KA*, keratoacanthoma; *LM*, lentigo maligna; *LMM*, lentigo maligna melanoma; *MCC*, Merkel cell carcinoma; *MIS*, melanoma in situ; *MS*, Mohs surgery; *NA*, not applicable; *SCC*, squamous cell carcinoma; *SCCIS*, squamous cell carcinoma in situ; *WLE*, wide local excision; *XP*, xeroderma pigmentosum.

Conflicts of interest

None disclosed.

REFERENCES

1. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. https://doi.org/10.2196/45312
2. Daneshjou R, Smith MP, Sun MD, Rotemberg V, Zou J. Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review. *JAMA Dermatol*. 2021;157(11):1362-1369. https://doi.org/10.1001/jamadermatol.2021.3129
3. Tschandl P. Risk of bias and error from data sets used for dermatologic artificial intelligence. *JAMA Dermatol*. 2021;157(11):1271-1273. https://doi.org/10.1001/jamadermatol.2021.3128
4. Ad Hoc Task Force, Connolly SM, Baker DR, et al. AAD/ACMS/ASDSA/ASMS 2012 appropriate use criteria for Mohs

micrographic surgery: a report of the American Academy of Dermatology, American College of Mohs Surgery, American Society for Dermatologic Surgery Association, and the American Society for Mohs Surgery. *J Am Acad Dermatol*. 2012;67(4):531-550. https://doi.org/10.1016/j.jaad.2012.06.009

5. Porter E, Murphy M, O'Connor C. Chat GPT in dermatology: progressive or problematic? *J Eur Acad Dermatol Venereol*. 2023;37(7):e943-e944. https://doi.org/10.1111/jdv.19174

https://doi.org/10.1016/j.jdin.2023.06.002