

Research Letter

# The Accuracy and Appropriateness of ChatGPT Responses on Nonmelanoma Skin Cancer Information Using Zero-Shot Chain of Thought Prompting

Ross O'Hagan, MD; Dina Poplausky, BA; Jade N Young, BS; Nicholas Gulati, MD, PhD; Melissa Levoska, MD; Benjamin Ungar, MD; Jonathan Ungar, MD

Department of Dermatology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

**Corresponding Author:**

Jonathan Ungar, MD  
Department of Dermatology  
Icahn School of Medicine at Mount Sinai  
5th Floor  
5 East 98th Street  
New York, NY, 10029  
United States  
Phone: 1 212 241 3288  
Email: [jonathan.ungar@mountsinai.org](mailto:jonathan.ungar@mountsinai.org)

(*JMIR Dermatol* 2023;6:e49889) doi: [10.2196/49889](https://doi.org/10.2196/49889)

**KEYWORDS**

ChatGPT; artificial intelligence; large language models; nonmelanoma skin; skin cancer; cell carcinoma; chatbot; dermatology; dermatologist; epidermis; dermis; oncology; cancer

## Introduction

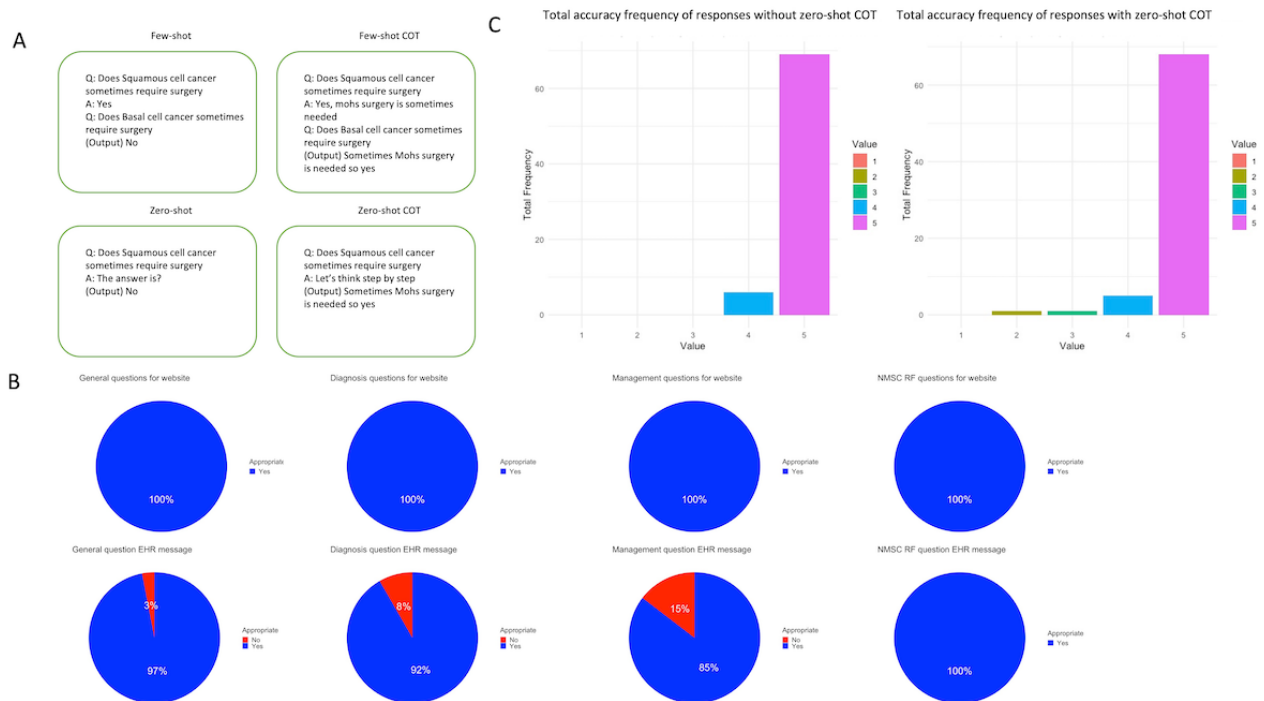
Nonmelanoma skin cancer (NMSC) represents the most prevalent form of cancer worldwide [1]. Patients with NMSC seek information from various resources. Work has already shown that language learning models (LLMs) such as ChatGPT can generate medical information in response to questions [2]; however, results vary significantly based on the prompts entered. Previous work has shown that a few-shot approach, where one provides several example prompts and outputs, has good results [3], as does the few-shot chain of thought approach, where answers include examples and the reasoning for correct answers, encouraging the model to reason through the question [4]. Zero-shot chain of thought (ZS-COT) prompting does not provide example prompts; instead, it uses phrases to encourage the LLMs to “think” through their responses, with significant improvement in accuracy in some contexts [5]. In this study, we explore ChatGPT’s performance in answering questions about NMSC using both standard and ZS-COT prompting.

## Methods

**Overview**

We generated 25 common clinical questions about NMSC in four categories: general, diagnosis, management, and risk factors. Prompts were entered into ChatGPT 4.0 on March 31, 2023, and responses were recorded for both standard and ZS-COT prompting (Figure 1A). Ending ZS-COT prompting queries with “Let’s think step by step” has been shown to improve performance in previous papers [5]. Three attending dermatologists independently reviewed and graded whether the outputs would be appropriate for a patient-facing website and an electronic health record (EHR) message draft to a patient. Responses were also evaluated for accuracy on a 5-point scale, with 1 being completely inaccurate and 5 being completely accurate, and reviewers assessed which of the two prompting styles they preferred. Statistical differences between prompts were computed using the Wilcoxon test. Statistical analysis was performed in R version 4.2.2 (R Foundation for Statistical Computing).

**Figure 1.** (A) Example of several popular language learning model prompting techniques. (B) Percent of appropriate responses for each question category by medium. (C) Accuracy scores by prompt style. COT: chain of thought; EHR: electronic health record; NMSC: nonmelanoma skin cancer; RF: risk factor.



**Ethical Considerations**

This study did not require institutional review board approval.

**Results**

Averaging all accuracy scores from a scale (range 1-5), we found that the combined accuracy for both the original prompt and ZS-COT prompt was 4.89. The average accuracy score from all 25 questions asked for the original prompt and ZS-COT prompt was 4.92 and 4.87, respectively, representing a nonsignificant difference of 1.03%. Both models were deemed 100% appropriate for a patient-facing information portal for general, diagnosis, management, and risk factor questions. For EHR message responses, outputs were appropriate for 97% of general questions, 92% of diagnosis questions, 85% of management questions, and 100% of risk factor questions (Figure 1B). The lowest accuracy grade for the standard

prompting responses and ZS-COT prompting was 4 and 2, respectively (Figure 1C). This score was given for the prompt “What causes basal cell carcinoma?” (Multimedia Appendix 1).

**Discussion**

This exploratory qualitative study found that LLMs can provide accurate patient information regarding NMSC appropriate for both general websites and EHR messages. We found that ZS-COT prompting does not provide more accurate dermatology information. The limitations of this study include that we only explored a subset of clinical questions patients may have about NMSC, there is no objective standard for appropriateness, and the personal biases of the dermatologists may bias response preference. As LLMs continue to grow and be adapted, clinicians must monitor their clinical utility and how different prompting methods may change the quality of results.

**Conflicts of Interest**

BU is an employee of Mount Sinai and has received research funds (grants paid to the institution) from Incyte, Rapt Therapeutics, and Pfizer. He is also a consultant for Arcutis Biotherapeutics, Castle Biosciences, Fresenius Kabi, Pfizer, and Sanofi. JU is an employee of Mount Sinai and is a consultant for AbbVie, Castle Biosciences, Dermavant, Janssen, Menlo Therapeutics, Mitsubishi Tanabe Pharma America, and UCB. The rest of the authors declare no relevant conflicts of interest.

**Multimedia Appendix 1**

Evaluated nonmelanoma skin cancer questions. [DOCX File, 18 KB-Multimedia Appendix 1]

**References**

1. Dubas LE, Ingraffea A. Nonmelanoma skin cancer. Facial Plast Surg Clin North Am. 2013 Feb;21(1):43-53 [doi: 10.1016/j.fsc.2012.10.003] [Medline: 23369588]

2. 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 Mar 14;329(10):842-844 [[FREE Full text](#)] [doi: [10.1001/jama.2023.1044](https://doi.org/10.1001/jama.2023.1044)] [Medline: [36735264](https://pubmed.ncbi.nlm.nih.gov/36735264/)]
3. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. arXiv. Preprint posted online on May 28, 2020. [[FREE Full text](#)]
4. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, et al. Chain-of-thought prompting elicits reasoning in large language models. arXiv. Preprint posted online on January 28, 2022. [[FREE Full text](#)]
5. Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. arXiv. Preprint posted online on May 24, 2022. [[FREE Full text](#)]

## Abbreviations

- EHR:** electronic health record  
**LLM:** language learning model  
**NMSC:** nonmelanoma skin cancer  
**ZS-COT:** zero-shot chain of thought

*Edited by J Solomon, I Brooks; submitted 12.06.23; peer-reviewed by A Hidki, U Kanike, D Chrimes; comments to author 21.09.23; revised version received 02.10.23; accepted 03.12.23; published 14.12.23*

*Please cite as:*

*O'Hagan R, Poplausky D, Young JN, Gulati N, Levoska M, Ungar B, Ungar J*

*The Accuracy and Appropriateness of ChatGPT Responses on Nonmelanoma Skin Cancer Information Using Zero-Shot Chain of Thought Prompting*

*JMIR Dermatol 2023;6:e49889*

*URL: <https://derma.jmir.org/2023/1/e49889>*

*doi: [10.2196/49889](https://doi.org/10.2196/49889)*

*PMID: [38096013](https://pubmed.ncbi.nlm.nih.gov/38096013/)*

©Ross O'Hagan, Dina Poplausky, Jade N Young, Nicholas Gulati, Melissa Levoska, Benjamin Ungar, Jonathan Ungar. Originally published in *JMIR Dermatology* (<http://derma.jmir.org>), 14.12.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Dermatology*, is properly cited. The complete bibliographic information, a link to the original publication on <http://derma.jmir.org>, as well as this copyright and license information must be included.