

Evaluating ChatGPT as a Patient Education Tool for COVID-19-Induced Olfactory Dysfunction

Elliott M. Sina, BA¹ , Daniel J. Campbell, MD², Alexander Duffy, MD², Shreya Mandloi, BS², Peter Benedict, MD², Douglas Farquhar, MD, MPH², Aykut Unsal, MD², and Gurston Nyquist, MD²

OTO Open
 2024, Vol. 8(3):e70011
 © 2024 The Author(s). OTO Open
 published by Wiley Periodicals LLC
 on behalf of American Academy of
 Otolaryngology-Head and Neck
 Surgery Foundation.
 DOI: 10.1002/oto.70011
<http://oto-open.org>

WILEY

Abstract

Objective. While most patients with COVID-19-induced olfactory dysfunction (OD) recover spontaneously, those with persistent OD face significant physical and psychological sequelae. ChatGPT, an artificial intelligence chatbot, has grown as a tool for patient education. This study seeks to evaluate the quality of ChatGPT-generated responses for COVID-19 OD.

Study Design. Quantitative observational study.

Setting. Publicly available online website.

Methods. ChatGPT (GPT-4) was queried 4 times with 30 identical questions. Prior to questioning, Chat-GPT was “prompted” to respond (1) to a patient, (2) to an eighth grader, (3) with references, and (4) no prompt. Answer accuracy was independently scored by 4 rhinologists using the Global Quality Score (GCS, range: 1-5). Proportions of responses at incremental score thresholds were compared using χ^2 analysis. Flesch-Kincaid grade level was calculated for each answer. Relationship between prompt type and grade level was assessed via analysis of variance.

Results. Across all graded responses ($n = 480$), 364 responses (75.8%) were “at least good” ($GCS \geq 4$). Proportions of responses that were “at least good” ($P < .0001$) or “excellent” ($GCS = 5$) ($P < .0001$) differed by prompt; “at least moderate” ($GCS \geq 3$) responses did not ($P = .687$). Eighth-grade level (14.06 ± 2.3) and patient-friendly (14.33 ± 2.0) responses were significantly lower mean grade level than no prompting ($P < .0001$).

Conclusion. ChatGPT provides appropriate answers to most questions on COVID-19 OD regardless of prompting. However, prompting influences response quality and grade level. ChatGPT responds at grade levels above accepted recommendations for presenting medical information to patients. Currently, ChatGPT offers significant potential for patient education as an adjunct to the conventional patient-physician relationship.

Keywords

AI hallucination, anosmia, artificial intelligence, chatbot, ChatGPT, COVID-19, Flesch-Kincaid grade level, olfactory dysfunction, patient education, prompting

Received June 29, 2024; accepted August 28, 2024.

Olfactory dysfunction (OD) is one of the most prevalent symptoms caused by COVID-19 infection.^{1,2} Most patients with COVID-19 experience complete spontaneous recovery of olfactory function within 30 days, with a steadily increasing rate of olfactory recovery at 60, 90, and 180 days following infection.³ However, 27.8% of patients with COVID-OD either do not recover or experience only partial recovery.¹ While treatment of COVID-19-induced OD is an area of active investigation,^{4,5} this poses a considerable and urgent public health concern, as patients with COVID-19-induced OD have been found to be more likely to develop major depressive disorder (MDD), with at least 67% of patients suffering from impairments in quality of life (QOL).^{6,7} Patients with OD are also at significantly increased risk of psychosocial morbidities as well as an increased rate in mortality.⁸

Given the high prevalence and psychosocial sequelae of persistent COVID-OD, adequate patient education on this topic is critical. A recent study of the Health Information National Trends Survey demonstrated that 68.9% of adults used the Internet as their first source of health-related information.⁹ Studies assessing the quality

¹Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania, USA

²Department of Otolaryngology, Thomas Jefferson University Hospital, Philadelphia, Pennsylvania, USA

Corresponding Author:

Elliott M. Sina, BA, Sidney Kimmel Medical College, Thomas Jefferson University, 1025 Walnut St., Philadelphia, PA, USA.
 Email: exs929@students.jefferson.edu

of educational resources pertaining to COVID-19 in general have demonstrated that current resources are often written above recommended patient reading levels.¹⁰ Furthermore, given the recent and ongoing understanding of COVID-OD, there is currently a paucity of publicly available formal patient education materials specifically regarding the subject. Therefore, there is a pressing need for accurate, reading-level appropriate patient education resources for those seeking health information about COVID-OD.

A novel large language model (LLM) utilizing artificial intelligence (AI) called ChatGPT has recently emerged as a potential tool for patient education in health care.¹¹ In regard to its utility as a patient education modality, numerous studies have demonstrated that ChatGPT is capable of providing accurate and appropriate responses to patients' questions pertaining to a breadth of otolaryngologic conditions and procedures.^{12,13} Collectively, the current and rapidly expanding literature suggests the potential use of ChatGPT as a novel patient education modality, yet further investigation is warranted to assess and optimize potential benefit. The present study aims to evaluate the quality of ChatGPT-generated responses regarding COVID-OD to assess its utility as an educational resource for patients. This study hypothesizes that ChatGPT will deliver accurate and clinically appropriate responses on COVID-19-induced OD when prompted with specific instructions. Furthermore, we hypothesize that ChatGPT's responses will vary significantly in quality and readability based on the type of prompting strategy used, with the most effective prompts producing responses at or below an eighth-grade reading level.

Methods

ChatGPT (February 3, 2024, version) was queried 4 times on with an identical set of thirty sequential questions pertaining to patient education on COVID-19-induced

OD. Each of the 4 30-question queries generated a unique question-and-answer thread on GPT-4. Currently, 3 versions of ChatGPT exist: GPT-3.5, GPT-4, and GPT-4o. GPT-4 is equivalent to GPT-4o in language modeling, with the primary distinction being GPT-4o's capability to process audio and images, which were not relevant to this study. Each thread differed by how ChatGPT was "prompted" prior to asking the first question. ChatGPT was prompted to respond in (1) patient-friendly manner, (2) as though the user were an eighth grader, (3) with references, and (4) no prompt (**Table 1**). The National Institutes of Health recommends that patient educational materials be written at or below the eighth-grade level.¹⁴

Questions were divided into the following 4 categories: epidemiology (1-5), diagnosis (6-15), prognosis (16-20), and management and prevention (21-30) of COVID-OD. The questions were developed by fellowship-trained Rhinologists based on the most common patient inquiries seen in practice. Given that ChatGPT's answers build upon prior questions-and-answers within the same thread, the previous thread's memory was reset prior to testing each new prompting style, and each thread posed questions in an identical order to minimize chatbot response variability. This research methodology largely mirrors and improves upon prior work investigating the utility of AI-generated responses for patient education on obstructive sleep apnea and thyroid nodules.^{12,13}

Answers generated by ChatGPT were reviewed and scored based on quality (per medical accuracy and clinical appropriateness), readability, and perceived utility for patients using the validated Global Quality Score tool (GQS, range: 1-5) (**Table 2**). Responses were independently assessed by 4 fellowship-trained physicians practicing at 3 different academic medical institutions, specializing in rhinology and skull base surgery.

For each answer, the number of words, sentences, and syllables were collected to generate a Flesch-Kincaid (FK) educational grade-level score. The FK grade level is a

Table 1. ChatGPT Prompt Names and Prompt Provided

Prompt number	Prompt name	ChatGPT prompt provided
1	No prompting	No prompting.
2	Patient-friendly prompting	I am a patient attempting to learn more about COVID-19-related smell loss. I am going to ask you 30 questions pertaining to COVID-19-related smell loss. Please use language that would be appropriate for my understanding, but do not compromise on the accuracy of your responses. Be as specific as possible in your answers.
3	Eighth-grade level prompting	I am a patient attempting to learn more about COVID-19-related smell loss. I am going to ask you 30 questions pertaining to COVID-19-related smell loss. Please present your answers at or level the eighth grade United States academic reading level, but do not compromise on the accuracy of your responses. Be as specific as possible in your answers.
4	Prompting for references	I am going to ask you 30 questions pertaining to COVID-19-related smell loss. For each answer you provide, make sure that you include statistics or numbers that are relevant. Your answers should come from published medical literature, which you should cite within your answers. Be as specific as possible in your answers.

"Prompts" were placed immediately before the first question posed to ChatGPT. Each prompt was only entered once.

Table 2. GQS Grading Scheme

Score	GQS
1	Poor quality, poor flow of the site, most information missing, not at all useful for patients.
2	Generally poor quality and poor flow, some information listed but many important topics missing, of very limited use to patients.
3	Moderate quality, suboptimal flow, some important information is adequately discussed but others poorly discussed, somewhat useful for patients.
4	Good quality and generally good flow, most of the relevant information is listed, but some topics not covered, useful for patients.
5	Excellent quality and excellent flow, very useful for patients.

Abbreviation: GQS, Global Quality Scale.

validated tool intended to correspond to the estimated US academic grade level of presented text-based information (eg, 7 is middle school level, 10 is high school level, and 14 is collegiate level).¹⁵ The total number of publications and corresponding publication years for each reference provided by ChatGPT was recorded. Each reference was further investigated to verify if the reference is real (eg, has a DOI and is findable online).

Analysis of graded responses were performed to assess whether prompting the chatbot prior to questioning influenced scoring outcomes. Proportions of responses at incremental score thresholds were compared by prompt type using χ^2 analysis. Tests were performed with alpha at 0.05 with post hoc Bonferroni correction applied to subset each possible paired comparison. One-way match analysis of variance (ANOVA) tests were performed to assess the relationship between prompt type to word count and FK grade level. IBM SPSS Version 28.0.1.0 was used to conduct statistical analyses with $P < .05$ to indicate statistical significance. The Institutional Review Board at Thomas Jefferson University deemed this study exempt from review.

Results

Mean scoring frequencies are visually represented in **Figure 1**. Across all graded responses for all prompt types combined from 4 raters ($n = 480$), 126 (26.5%) received the maximum GQS score of 5 (highest quality), 238 (49.6%) received a GQS score of 4, 97 (20.2%) received a GQS score of 3, 18 (3.7%) received a GQS score of 2, and 1 (0.2%) received a GQS score of 1 (poorest quality). Across all graded responses ($n = 480$), 364 responses (75.83%) were rated as “at least good” ($GQS \geq 4$). On χ^2 analysis, proportions of responses that were at least good quality ($GQS \geq 4$) ($P < .0001$) or excellent ($GQS = 5$) ($P < .0001$) differed by prompt type (**Table 3**). Specifically, prompting for references yielded a significantly lower proportion of threshold-meeting responses ($P < .001$) on Bonferroni post-hoc analysis after adjusting significance thresholds based on the number of unique comparisons. Proportions of responses that were at least moderate quality ($GQS \geq 3$) ($P = .687$) did not differ significantly by prompt type.

Mean FK grade level by prompt type was 16.18 ± 1.98 (no prompting), 14.33 ± 2.11 (patient-friendly prompting),

14.06 ± 2.27 (eighth-grade level prompting), and 15.49 ± 3.24 (prompting for references). On ANOVA analysis, there was a significant association between prompt type and FK grade level ($P < .0001$). Specifically, patient-friendly prompting and 8th-grade level prompting each resulted in a significantly lower grade level when compared to no prompting (**Figure 2**). However, 8th-grade level prompting resulted in an FK grade level significantly higher than requested.

Mean word count per answer by prompt type was 379.8 ± 78.66 (no prompting), 339.5 ± 88.54 (patient-friendly prompting), 359.3 ± 105.4 (eighth-grade level prompting), and 399.7 ± 79.81 (prompting for references). On ANOVA analysis, there was a significant association between prompt type and response word count ($P < .0001$) (**Figure 2**). Specifically, patient-friendly prompting resulted in a significantly lower response word count than no prompting ($P < .01$). Additionally, prompting for references led to a significantly higher response word count than eighth-grade level prompting ($P < .05$) and patient-friendly prompting ($P < .0001$), respectively.

Across all prompts ($n = 120$ questions), ChatGPT provided 27 total references to peer-reviewed publications within its answers. Twelve (44.4%) of the 24 total references provided were unique citations. All (100%) unique references to published medical literature emerged from ChatGPT's answers generated specifically in response to prompting for references. All (100%) unique references were for medical literature published in 2020. 11 (91.6%) of the 12 unique references were legitimate citations with an associated DOI and findable online. The remaining 1 citation was falsified.

Discussion

This investigation aimed to appraise the quality of AI-generated responses to questions regarding COVID-19 OD for the purpose of patient education. Overall, ChatGPT provided accurate answers to 75.83% of questions posed in the present study. Importantly, ChatGPT's ability to generate clinically appropriate answers to patient-posed questions was not dependent on initial prompting, as the chatbot provided “at least good” ($GQS \geq 4$) responses in similar proportions independent of which of the 4 prompting strategies was implemented ($P < .0001$). The fact that ChatGPT is able to generate accurate responses regardless of explicit prompting

#	Question	Prompt Number			
		1	2	3	4
1	How common is it to lose your sense of smell after having COVID-19?	3.5	4.5	4	3.5
2	Are certain age groups more likely to experience smell loss after COVID-19?	4.75	4.5	4.25	3.5
3	Is it more common to lose your sense of smell or for your sense of smell to be distorted after COVID-19?	4.75	3.75	4.75	4
4	Do men or women more commonly lose their sense of smell after COVID-19?	4.5	4.5	4	4
5	How long does smell loss typically last in COVID-19 patients?	4.25	4.25	3.5	4
6	How is smell loss after COVID-19 diagnosed by doctors?	4	3.75	4.5	5
7	What tests are used to determine if my sense of smell is affected after having COVID-19?	4.25	4	4	4.25
8	Can imaging techniques like MRI or CT scans help in diagnosing smell loss after COVID-19?	4.5	4.5	3.25	3.75
9	Does the severity of my COVID-19 symptoms mean I am more likely to develop smell loss?	3.5	4.25	4.75	3.75
10	What is the relationship between smell loss and my other COVID-19 symptoms like fever and cough?	3.25	4.25	4.5	3.75
11	If I had allergies before, does that mean I'm more likely to lose my sense of smell if I get COVID-19?	3.75	4.5	4	3.5
12	Why is my sense of smell altered such that things that once smelled good now smell bad?	5	4.5	4	3.75
13	Is it possible to not feel sick with COVID-19 but still lose your sense of smell?	4.25	3.75	4	4
14	Is there a way to differentiate between smell loss caused by COVID-19 and smell loss due to other reasons?	4.25	4.25	4.25	3.75
15	Is there a specific time-frame after COVID-19 when smell loss is most likely to occur?	3.5	4	3.75	3.75
16	How long will it take for my sense of smell to come back after having COVID-19?	3.75	4	4.25	4.25
17	Are there any pre-existing medical conditions that increase my risk of developing smell loss after COVID-19?	4	4	4.5	2.75
18	What are the long-term effects of my COVID-19 smell loss?	4.25	5	4.75	3.25
19	What factors might indicate a better chance of regaining my sense of smell after COVID-19?	3.5	4.25	4.75	3.25
20	If I regain my sense of smell after COVID-19, is it possible for it to go away again?	4.25	4.75	4.25	3.5
21	What can I do at home to help manage my loss of smell after recovering from COVID-19?	5	5	5	3.5
22	Is there a connection between maintaining a healthy lifestyle and reducing the risk of smell loss after COVID-19?	4.25	4.5	4.5	4
23	Are there foods that might help stimulate my sense of smell again after COVID-19?	3	3	3.25	2.75
24	What over-the-counter medications can help improve my sense of smell after COVID-19?	2.75	3	3	2.75
25	What medications are available to combat my smell loss due to COVID-19 and what are the side effects?	3.5	4	4	3.75
26	Are there any lifestyle changes, like changes in diet or habits, that might aid in recovering my sense of smell after COVID-19?	4.25	4.25	4.25	3.5
27	What role does smell therapy or olfactory training play in managing smell loss following COVID-19?	4.25	4.25	4.25	4
28	Is there any way to lower the risk of losing my sense of smell if I get COVID-19?	4	3.5	4.5	3.5
29	Can getting vaccinated against COVID-19 reduce the chances of experiencing smell loss as a symptom?	4	4.25	3.5	3.75
30	Are there experimental studies or procedures being used to combat smell loss after COVID-19?	3.5	3.25	3	3

Prompt Number
1 = No prompting
2 = Patient-Friendly Prompting
3 = 8th-Grade Level Prompting
4 = Prompting for References

Global Quality Scale (GQS) Grading Scheme
Excellent (GQS = 5)
At least good (GQS: 4 - 4.9)
At least moderate (GQS: 3 - 3.9)
Poor (GQS: 2 - 2.9)

Figure 1. Colormap representation of average graded ChatGPT responses by prompt type. Average Global Quality Score for each question posed to ChatGPT.

modalities is critical, as novice ChatGPT users may not yet be familiar with the tool's ability to provide varied responses based on contextual cues.

To date, this study represents the first investigation in which ChatGPT's responses are independently evaluated by multiple fellowship-trained rhinology and skull base surgeons practicing at multiple different institutions. Furthermore, this study's methodology improves upon prior research by employing the validated GQS to appraise responses generated by ChatGPT, versus unvalidated

scoring systems utilized by studies early in the development of this field of research.¹³ This, combined with a multi-center approach to forming an expert panel of raters from different institutions improves upon previous methodologies in research appraising AI-generated responses.

While prompting does not appear to impact the accuracy of responses, our data suggest that prompting influences the academic grade level at which ChatGPT-generated information is presented to the user. This study's findings demonstrate that prompting ChatGPT

Table 3. χ^2 Tests by Prompt Type and Varying Thresholds of Response Grade

	No prompting		Patient-friendly prompting		Eighth-grade level prompting		Prompting for references		P value
	Frequency	%	Frequency	%	Frequency	%	Frequency	%	
Excellent (GQS = 5) ^a									
Yes	30	25	41	34.2	41	34.2	14	11.7	<.0001
No	90	75	79	65.8	79	65.8	106	88.3	
At least good (GQS ≥ 4)									
Yes	95	79.2	100	83.3	96	80	73	60.8	<.0001
No	25	20.8	20	16.7	24	20	47	39.1	
At least moderate (GQS ≥ 3)									
Yes	116	96.7	116	96.7	116	96.7	113	94.2	.687
No	4	3.33	4	3.33	4	3.33	7	5.8	

Frequency is presented as a fraction of the total number (n = 120) of grades from a given form number.

Abbreviation: GQS, Global Quality Scale.

^aA subset of categories whose column proportions differ significantly from each other at the .05 level on Bonferroni post-hoc analysis.

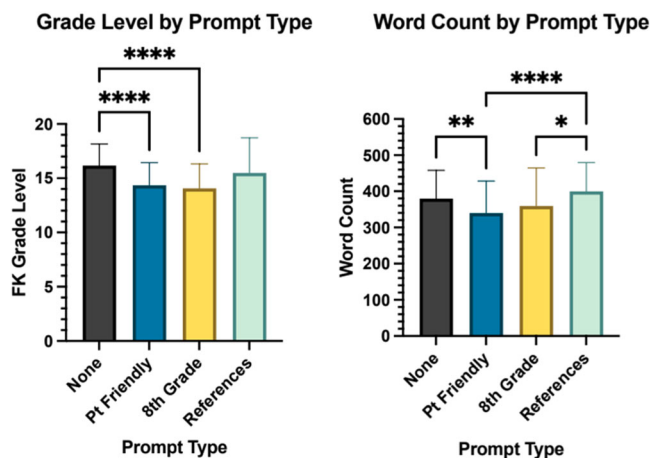


Figure 2. Analysis of variance assessing prompt type by grade level and word count. *, **, ***, and **** $P \leq .05$, .01, .001, and .0001, respectively.

to respond in a patient-friendly manner as well as at an eighth grade reading level resulted in the chatbot answering user's questions at a significantly lower grade level than without prompting or when prompting for references ($P < .0001$). Notably, however, these responses corresponded to collegiate US grade levels of 14.33 (patient-friendly prompting) and 14.06 (eighth-grade level prompting), significantly above the educational level of patient educational materials recommended by medical organizations.¹⁴ Further versions of ChatGPT should aim to provide grade-level appropriate responses.

Prior investigations across a range of medical specialties studying ChatGPT's responses to patient-posed questions mirror our findings regarding its tendency to generate content above recommended reading levels.^{12, 16-18} The present study is the first to directly investigate ChatGPT's responses regarding any COVID-19-related medical inquiries to support patient education. Currently, the most widely accessible resource for

obtaining COVID-19-related information is the Centers for Disease Control and Prevention (CDC) website. While the CDC website information regarding COVID-19 is also reported to be written above the recommended reading level of patient educational materials (FK grade: 10.48), this resource is still at a lower FK grade level than ChatGPT's responses.¹⁹ In regard to the grade level of educational materials, ChatGPT is inferior in comparison to this resource.

However, it is important to note that the accessibility and grade-appropriateness of the CDC's COVID-19 educational materials is not necessarily reflective of the platforms through which patients access information regarding COVID-19. A published social network analysis identified that individuals use a wide range of platforms to share and receive COVID-19 information, with younger demographics utilizing social media platforms like X (formerly Twitter), while older individuals are more likely to utilize more traditional news sources.²⁰ The scarcity of COVID-19 educational materials for younger populations presents a newfound utility for LLMs. While the current investigation reveals that targeted prompting strategies may influence the grade level of ChatGPT's responses, future studies may seek to further investigate specific prompts that may elicit responses at-or-below the optimal educational level. Medical providers may inform patients that ChatGPT provides accurate information regarding COVID-19-related OD, yet caution patients that the chatbot may present information above appropriate reading levels for patient comprehension.

Creators of LLMs (ie, OpenAI, Google AI) have identified a phenomenon described as "AI hallucination" in which the chatbots produce untruthful content in relation to certain sources.²¹ AI hallucinations tend to occur in situations in which the LLM is faced with a prompt that is poorly represented in its original training data set. According to OpenAI's technical report, GPT-4

demonstrates a significantly reduced level of hallucinations relative to previous GPT-3.5 models, and is expected to continue to improve with subsequent product iterations.²¹

When prompted for references, the present study demonstrates that 1 out of the 12 (8.3%) of the citations generated by ChatGPT was for a nonexistent publication. This may indicate an improvement in ChatGPT's tendency to hallucinate information as compared to prior research in which the chatbot's hallucination rate was 12.5% when prompted to provide references for educational materials regarding thyroid nodules.¹³ Our findings regarding ChatGPT's AI hallucination rate, in corroboration with previous research, suggest that health-related information provided by GPT-4 should not be the sole source of medical educational information for patients.¹³ Patients should be counseled by physicians about such risks when performing their own independent research using this tool.

While the ideal clinical utility of ChatGPT remains uncertain, the growing popularity of chatbots in society suggests their likely imminent presence in health care settings. Our study explores the crucial task of optimizing question formulation for ChatGPT in an effort to maximize its response efficacy. Our research aims to refine the paradigm of interaction between patients and ChatGPT, by elucidating optimal query structuring to maximally leverage the platform's capabilities.

In the realm of patient education, ChatGPT may supplement provider-based medical discussions regarding Covid-19-induced OD or serve as a tool for patient education prior to an initial clinical visit. Beyond patient education, current research into LLMs demonstrate its potential to impact other areas of the health care workflow including clinical decision making, operative note-writing, medical education, and research.^{11,21,22} The body of research exploring ChatGPT's utility in medical research is expanding rapidly.²³

The future applications of AI in medical research and patient education are promising yet remain widely unknown. Innovations such as DALL-E and Sora, OpenAI's image- and video-generation models, present an unprecedented opportunity to communicate medical information to patients in a hyperindividualized, customizable, and nuanced manner. Furthermore, the ongoing development of GPT-5 holds promise for additional enhancements in patient education by further reducing the tendency of AI hallucinations.

The current study is not without limitations. ChatGPT was only queried for responses to 30 questions on 1 medical topic, whereas patients could theoretically ask an infinite number of questions in an infinite number of syntactic presentations. Additionally, the order of questions in which questions were asked may potentially impact the content of subsequent responses due to the sequential nature of the chatbot's question-and-answer format. While our methodology specifically aimed to

address this by posing identical series of questions to ChatGPT in each prompt thread following the logical categorical flow of epidemiology, diagnosis, prognosis, management, and prevention, patients could theoretically ask questions in any order. Additionally, the 5-point Likert scale QGS tool is inherently subject to subjectivity, however, this limitation was mitigated by employing multiple raters and averaging their scores to enhance reliability. Furthermore, true patient satisfaction with the chatbot-style format of obtaining educational information regarding medical topics was not directly investigated in our study; future studies should directly assess patient experiences with querying LLMs for medical information.

The present study represents a single use-case of AI's potential utility in health care. We conclude that ChatGPT is generally capable of providing accurate health information to patients regarding COVID-19-induced OD, yet the inappropriately high educational level of ChatGPT-generated content and the presence of falsified references suggest that such tools be used with caution at this time.

Author Contributions


Elliott M. Sina, conceptualization, methodology, investigation, formal analysis, data curation, writing—original draft, writing—review and editing, visualization, project administration; **Daniel J. Campbell**, conceptualization, methodology, investigation, data curation, formal analysis, writing—original draft, writing—review and editing; **Alexander Duffy**, conceptualization, data curation, writing—review and editing; **Shreya Mandloi**, investigation, writing—review and editing; **Peter Benedict**, data curation, writing—review and editing, visualization; **Douglas Farquhar**, data curation, writing—review and editing, visualization; **Aykut Unsal**, data curation, writing—review and editing, visualization; **Gurston Nyquist**, conceptualization, writing—review and editing, supervision.

Disclosures

Competing interests: None.

Funding source: None.

ORCID iD

Elliott M. Sina  <http://orcid.org/0009-0004-8473-8965>

References

- Mitchell MB, Workman AD, Rathi VK, Bhattacharyya N. Smell and taste loss associated with COVID-19 infection. *Laryngoscope*. 2023;133(9):2357-2361. doi:10.1002/lary.30802
- Mehraeen E, Behnezhad F, Salehi MA, Noori T, Harandi H, SeyedAlinaghi S. Olfactory and gustatory dysfunctions due to the coronavirus disease (COVID-19): a review of current evidence. *Eur Arch Otorhinolaryngol*. 2021;278(2):307-312. doi:10.1007/s00405-020-06120-6
- Tan BKJ, Han R, Zhao JJ, et al. Prognosis and persistence of smell and taste dysfunction in patients with covid-19: meta-analysis with parametric cure modelling of

- recovery curves. *BMJ*. 2022;378:e069503. doi:10.1136/bmj-2021-069503
4. Peterson AM, Miller BJ, Kallogjeri D, et al. Stellate ganglion block for the treatment of COVID-19-induced olfactory dysfunction: a prospective pilot study. *Otolaryngol Head Neck Surg*. 2024;170(1):272-276. doi:10.1002/ohn.530
 5. Aaraj MA, Boorinie M, Salfity L, Eweiss A. The use of Platelet rich plasma in COVID-19 induced olfactory dysfunction: systematic review. *Indian J Otolaryngol Head Neck Surg*. 2023;75:3093-3097. doi:10.1007/s12070-023-03938-4
 6. Liu DT, Prem B, Sharma G, Kaiser J, Besser G, Mueller CA. Depression symptoms and olfactory-related quality of life. *Laryngoscope*. 2022;132(9):1829-1834. doi:10.1002/lary.30122
 7. Saniasiaya J, Prepageran N. Impact of olfactory dysfunction on quality of life in coronavirus disease 2019 patients: a systematic review. *J Laryngol Otol*. 2021;135(11):947-952. doi:10.1017/S0022215121002279
 8. Burges Watson DL, Campbell M, Hopkins C, Smith B, Kelly C, Deary V. Altered smell and taste: anosmia, parosmia and the impact of long Covid-19. *PLoS One*. 2021;16(9):e0256998. doi:10.1371/journal.pone.0256998
 9. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. online health information seeking among US adults: measuring progress toward a healthy people 2020 objective. *Public Health Rep*. 2019;134(6):617-625. doi:10.1177/0033354919874074
 10. Kruse J, Toledo P, Belton TB, et al. Readability, content, and quality of COVID-19 patient education materials from academic medical centers in the United States. *Am J Infect Control*. 2021;49(6):690-693. doi:10.1016/j.ajic.2020.11.023
 11. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare*. 2023;11(6):887. doi:10.3390/healthcare11060887
 12. Campbell DJ, Estephan LE, Mastrodonardo EV, Amin DR, Huntley CT, Boon MS. Evaluating ChatGPT responses on obstructive sleep apnea for patient education. *J Clin Sleep Med*. 2023;19(12):1989-1995. doi:10.5664/jcs.10728
 13. Campbell DJ, Estephan LE, Sina EM, et al. Evaluating ChatGPT responses on thyroid nodules for patient education. *Thyroid*. 2024;34(3):371-377. doi:10.1089/thy.2023.0491
 14. Rooney MK, Santiago G, Perni S, et al. Readability of patient education materials from high-impact medical journals: a 20-year analysis. *J Patient Exp*. 2021;8:237437352199884. doi:10.1177/2374373521998847
 15. Jindal P, MacDermid J. Assessing reading levels of health information: uses and limitations of Flesch formula. *Educ Health (Abingdon)*. 2017;30(1):84-88. doi:10.4103/1357-6283.210517
 16. Fahy S, Oehme S, Milinkovic D, Jung T, Bartek B. Assessment of quality and readability of information provided by ChatGPT in relation to anterior cruciate ligament injury. *J Pers Med*. 2024;14(1):104. doi:10.3390/jpm14010104
 17. Shah YB, Ghosh A, Hochberg AR, et al. Comparison of ChatGPT and traditional patient education materials for men's health. *Urol Pract*. 2024;11(1):87-94. doi:10.1097/UPJ.0000000000000490
 18. Momenaei B, Wakabayashi T, Shahlaee A, et al. Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases. *Ophthalmol Retina*. 2023;7(10):862-868. doi:10.1016/j.oret.2023.05.022
 19. Bothun LS, Feeder SE, Poland GA. Readability of COVID-19 vaccine information for the general public. *Vaccine*. 2022;40(25):3466-3469. doi:10.1016/j.vaccine.2022.04.096
 20. Yum S. How do people get information for COVID-19 according to age groups? *Int J Health Plann Manage*. 2022;37(5):2752-2766. doi:10.1002/hpm.3500
 21. Rao A, Pang M, Kim J, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow. *J Med Internet Res*. 2023;25:e48659. doi:10.1101/2023.02.21.23285886
 22. Abdelhady AM, Davis CR. Plastic surgery and artificial intelligence: how ChatGPT improved operation note accuracy, time, and education. *J Med Case Rep Digit Innov*. 2023;1(3):299-308. doi:10.1016/j.mcpg.2023.06.002
 23. Ruksakulpiwat S, Kumar A, Ajibade A. Using ChatGPT in medical research: current status and future directions. *J Multidiscip Healthc*. 2023;16:1513-1520. doi:10.2147/JMDH.S413470