Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.JournalofSurgicalResearch.com

Association for Academic Surgery

Improving Readability and Automating Content Analysis of Plastic Surgery Webpages With ChatGPT

James E. Fanning, BS,[a] Maria J. Escobar-Domingo, MD,[a] Jose Foppiani, MD,[a] Daniela Lee, BS,[a] Amitai S. Miller, BA,[a] Jeffrey E. Janis, MD,[b] and Bernard T. Lee, MD, MBA, MPH[a,*]

[a] Division of Plastic and Reconstructive Surgery, Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts

[b] Department of Plastic and Reconstructive Surgery, Ohio State University Wexner Medical Center, Columbus, Ohio

ARTICLE INFO

Article history:
Received 4 December 2023
Received in revised form
10 February 2024
Accepted 14 April 2024
Available online xxx

Keywords:
Artificial intelligence
Machine learning
Patient education
Readability

ABSTRACT

Introduction: The quality and readability of online health information are sometimes suboptimal, reducing their usefulness to patients. Manual evaluation of online medical information is time-consuming and error-prone. This study automates content analysis and readability improvement of private-practice plastic surgery webpages using ChatGPT.

Methods: The first 70 Google search results of "breast implant size factors" and "breast implant size decision" were screened. ChatGPT 3.5 and 4.0 were utilized with two prompts (1: general, 2: specific) to automate content analysis and rewrite webpages with improved readability. ChatGPT content analysis outputs were classified as hallucination (false positive), accurate (true positive or true negative), or omission (false negative) using human-rated scores as a benchmark. Six readability metric scores of original and revised webpage texts were compared.

Results: Seventy-five webpages were included. Significant improvements were achieved from baseline in six readability metric scores using a specific-instruction prompt with ChatGPT 3.5 (all $P \leq 0.05$). No further improvements in readability scores were achieved with ChatGPT 4.0. Rates of hallucination, accuracy, and omission in ChatGPT content scoring varied widely between decision-making factors. Compared to ChatGPT 3.5, average accuracy rates increased while omission rates decreased with ChatGPT 4.0 content analysis output.

Conclusions: ChatGPT offers an innovative approach to enhancing the quality of online medical information and expanding the capabilities of plastic surgery research and practice. Automation of content analysis is limited by ChatGPT 3.5's high omission rates and ChatGPT 4.0's high hallucination rates. Our results also underscore the importance of iterative prompt design to optimize ChatGPT performance in research tasks.

* *Corresponding author*. Division of Plastic and Reconstructive Surgery, Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, 110 Francis Street, Boston, MA 02215.
E-mail address: blee3@bidmc.harvard.edu (B.T. Lee).

0022-4804/$ – see front matter © 2024 Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.jss.2024.04.006

Introduction

Variable quality and readability of online medical information related to esthetic and reconstructive procedures limits their usefulness to plastic surgery patients.[1-5] This difference creates a barrier to patients' access of high-quality information, which directly affects their quality of care. In fact, online medical information is a key element of patient planning, utilized by 95% of esthetic surgery patients before in-office consultations.[6] Despite the growing prevalence of plastic surgery content on social media platforms, practice websites remain a popular and heavily utilized information source for patients.[7-9] The expansion of online health information and ongoing struggles for quality control and accuracy require refined and efficient analytic tools. To improve patient access to quality information, numerous studies have conducted content and readability analyses of online health content regarding plastic and reconstructive surgery.[1-3,10-12] However, traditional methods for evaluating online posts are time-consuming and prone to human error, as each item must be individually reviewed then manually scored.[13-15] Automating this process would greatly expedite improvement of online patient information.

In November 2022, OpenAI launched ChatGPT 3.5, a sophisticated language learning model.[16] ChatGPT 3.5, commonly shortened to ChatGPT, is a free online chatbot that generates human-like responses to text input using the generative pretrained transformer (GPT) model. Most studies in literature that reference the application of ChatGPT in patient surgical education involve its use in generation of informative medical content.[17-26] However, there is also great potential for ChatGPT to be used to optimize and evaluate patient resources. Recently, a proof-of-concept study successfully used ChatGPT to improve the readability of scientific journal articles, although it was unable to reach the recommended sixth-grade reading level for online health information.[27] The chatbot has also been used to analyze and categorize tweets with #plasticsurgery.[28] While ChatGPT 3.5 has demonstrated great promise in these research tasks, the updated model, ChatGPT 4.0, was released in March 2023 and requires a paid subscription. ChatGPT 4.0 is useful for more complex instructions and tasks and its performance is more reliable than ChatGPT 3.5.[29] Specifically, when given instructions or tasks, its performance is more accurate and less affected by omissions or hallucinations. While omission represents performance marked by an absence or failure to meet requests, hallucination represents performance that contains false information or erroneous reasoning.[29]

Our goal was to determine if ChatGPT is an appropriate analytic tool to improve readability and automate content analysis of online health information. We manually completed content and readability analyses of 75 private practice plastic surgery webpages that detail decision-making factors for breast implant size selection. We utilized 12 established decision-making factors of breast implant size selection, including surgeon input, lifestyle, complications and adverse outcomes, fertility and aging, consultation with loved ones, conception of breast size, esthetic goals and motivations, breast asymmetry, body and tissue-based measurements, patient decision support devices, procedural considerations, and other implant features.[10] We utilized ChatGPT 3.5 and 4.0 with general prompt instruction (Prompt 1) and specific prompt instruction (Prompt 2) to perform these tasks. Using manual scores as a benchmark, we evaluated the performance of ChatGPT and instruction prompts.

Methods

Webpage identification

We utilized a previously published methodology for systematic collection of plastic surgery private practice webpages.[10] The terms "breast implant size factors" and "breast implant size decision" were searched using Google search engine in June 2023. To depersonalize search results, Startpage.com was utilized to send anonymous, depersonalized searches to Google's search engine that are not shaped by user settings, location, IP address, search history, or cookies.[30] The first 60 results from each search term were reviewed for inclusion. Duplicate results were removed. Inclusion criteria were United States plastic surgery practice webpages related to patient counseling specifically on breast implant size selection. Exclusion criteria were (1) advertisements, (2) webpages unrelated to breast implant size selection, (3) webpages belonging to practices outside of the United States, and (4) nonpractice webpages, such as informational sites, companies, academic institutions, or plastic surgery organizations.

Manual content analysis of webpage texts

Webpage texts were extracted and analyzed independently for stated decision-making factors of breast implant size selection by two authors (J.E.F. and M.J.E.). Decision-making factors were recorded as present or absent. Coding and identification of factors was completed to establish decision-making categories. Discrepancies between authors' coding were resolved by shared analysis and discussion with a third author (D.L.). Decision-making factors were tabulated by webpage for descriptive statistical analysis and used as reference scores for comparison with ChatGPT automated analyses.

ChatGPT-automated improvement of webpage text readability

To obtain baseline readability scores, webpage texts were batch imported into Readability Professional Studio software.[31] Six validated readability scales were utilized: Flesch–Kincaid, Flesch reading ease, Fry, Gunning fog, Raygor estimate, and Simple Measure of Gobbledygook (SMOG) (Table 1). We then utilized ChatGPT3.5 and ChatGPT4.0 with two instructions prompts to automate improvement of webpage readability by revising webpage transcripts.[32] Prompt 1 provided general instruction to rewrite webpage texts without altering the transcript structure. Prompt two provided greater specificity of instruction, listed established readability metrics for reference, and provided more examples of how to achieve improved readability scores (Supplementary Digital Content

Table 1 — Six readability measures utilized to assess plastic surgery webpage readability.

Readability measure	Scale	Text metrics utilized by readability metric
Flesch–Kincaid	Grade level	• Sentence length • Number of syllables
Flesch reading ease	Score range (0-100)	• Sentence length • Number of syllables
Fry	Grade level	• Sentence length • Number of syllables
Gunning fog	Grade level	• Sentence length • Percentage of difficult words
Raygor estimate	Grade level	• Sentence length • Word length
SMOG	Grade level	• Sentence length • Word complexity (polysyllables)

Readability measure scales and text metrics utilized for score calculation are described.

1). For each webpage, four inquiries were sent to ChatGPT to generate improved transcripts: ChatGPT3.5 with Prompt 1, ChatGPT 3.5 with Prompt 2, ChatGPT4.0 with Prompt 1, and ChatGPT4.0 with Prompt 2 (Fig. 1). Eight webpages were randomly selected as a validation dataset. ChatGPT-generated transcripts of these eight webpages were analyzed independently by two authors (J.E.F. and M.J.E.) to compare stated decision-making factors of breast implant size selection with those present in the original webpage to measure if ChatGPT altered webpage content.

ChatGPT-automated content analysis of webpage texts

We utilized ChatGPT3.5 and ChatGPT4.0 with two different instructions prompts to automate content analysis of decision-making factors of breast implant size selection reported in webpage texts. Prompt 1 provided general instructions for how to analyze webpage content and listed limited examples of each decision-making factors. Prompt two provided greater specificity for how to analyze webpage content and more examples of each factor (Supplementary Digital Content 2). For each webpage, four inquiries were sent to ChatGPT for content analysis: ChatGPT3.5 with Prompt 1, ChatGPT 3.5 with Prompt 2, ChatGPT4.0 with Prompt 1, and ChatGPT4.0 with Prompt 2 (Fig. 1). ChatGPT output was analyzed after each inquiry and decision-making factors were recorded as present or absent. Using the results of manual content analysis as a reference, we assigned the following three values to ChatGPT output: hallucination (decision-making factor absent in manual analysis, present in ChatGPT analysis), accurate (decision-making factor present in manual analysis and ChatGPT analysis), or omission (decision-making factor present in manual analysis, absent in ChatGPT analysis).

Statistical analysis

SPSS (IBM Corp., Armonk, NY, Version 28.0) was used to conduct all statistical analyses. A value of $P < 0.05$ was considered statistically significant. Univariate analysis was performed using Chi-square tests for categorical variables. One-way analysis of variance tests were utilized to determine if there was a significant difference between the quality of ranking (hallucination, accuracy, and omission) and the four prompts (ChatGPT 3.5 with Prompt 1 and Prompt 2; ChatGPT 4.0 with Prompt 1 and Prompt 2). The comparison between groups was performed for each of the 12 decision-making factors. Friedman's two-way analysis of variance by ranks was utilized to compare baseline readability scores and ChatGPT-generated transcript readability scored matched by each webpage.

Results

A total of 75 unique US private-practice plastic surgery webpages representing 73 US private practices related to breast implant size selection met the inclusion criteria. Private practices were located in 23 US states, with California, Texas, New York, Florida, and Virginia representing approximately half (52%, 38/73) of included practices (Fig. 2).

Readability

Average webpage readability scores of original webpages were obtained for Flesch–Kincaid (9.3 grade level), Flesch Reading Ease (60.5/100), Fry (9.7 grade level), Gunning Fog (11.1 grade level), Raygor Estimate (10.7 grade level), and SMOG (11.6 grade level) readability scales (Table 1). Webpage texts rewritten with ChatGPT 3.5 and Prompt 1 did not have significant changes in average readability scores from baseline, excluding SMOG scale which scored 10.7 ($P < 0.05$). Webpage texts generated with ChatGPT 3.5 and Prompt two were significantly improved from baseline in average readability scores for all six scales (Flesch Reading Ease $P < 0.05$, Flesch–Kincaid, Fry, Gunning Fog, Raygor Estimate, and SMOG $P < 0.001$). Webpage texts generated with ChatGPT 4.0 and both Prompt 1 and Prompt two were significantly improved from baseline in average readability scores for all six scales ($P < 0.001$) (Fig. 3). There were no differences in reported decision-making factors between original webpage texts and the four ChatGPT-rewritten texts for the eight webpages randomized to the validation dataset.

Content analysis

Significant variations in content scoring performance were observed for nine of 12 decision-making factors, including surgeon input, lifestyle, complications and adverse outcomes, conception of breast size, esthetic goals and motivations, correction of breast asymmetry, body and tissue-based measurements, patient decision support devices, and other implant features ($P < 0.05$). No significant differences in content scoring performance were observed for the remaining three patient decision-making factors, including fertility and aging, consultation with loved ones, and procedural considerations (Fig. 4).

Omission rates, accuracy rates, and hallucination rates varied between evaluators (Fig. 4). Average rates of omission in content scoring were lower with ChatGPT 4.0 compared to

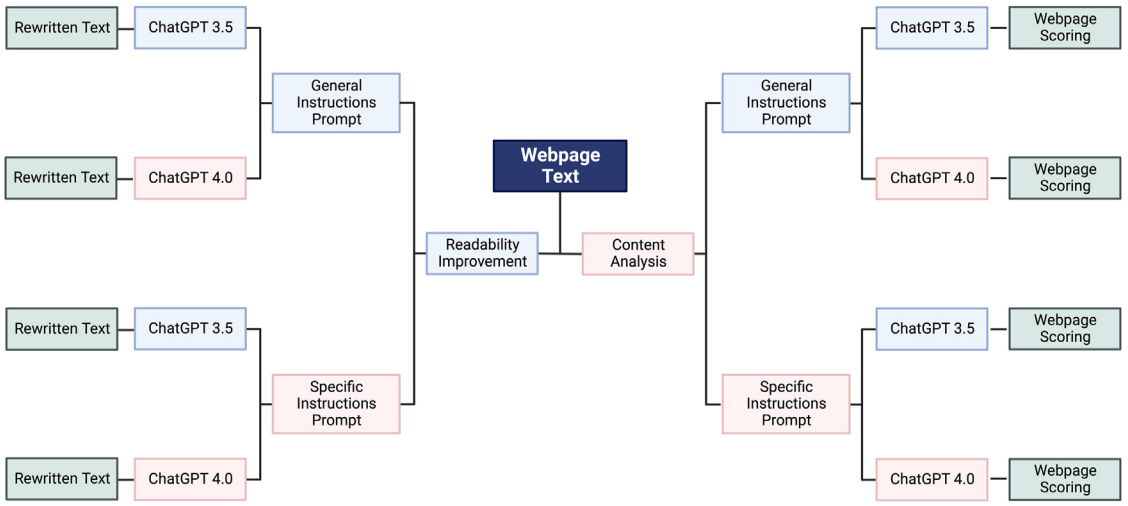

Fig. 1 – A flowchart of the study design utilizing ChatGPT 3.5 and ChatGPT 4.0 with general and specific instructions prompts to automate readability improvement and content analysis scoring of plastic surgery webpages.

ChatGPT 3.5, and decreased in scoring all decision-making factors, excluding esthetic goals and motivations. Average rates of accurate content scoring were higher with ChatGPT 4.0 compared to ChatGPT 3.5, and increased in scoring of six decision-making factors, including surgeon input, lifestyle, complications and adverse outcomes, fertility and aging, body and tissue-based measurements, and patient decision support devices. Average rates of hallucination in content scoring were comparable between all four models, and rates varied in scoring of individual patient decision-making factors (Fig. 4).

Discussion

The present study utilized ChatGPT to automate improvement of plastic surgery webpage readability and health information content analysis. We demonstrate that significant improvements in readability can be achieved with ChatGPT and that performance of ChatGPT 3.5 is comparable to ChatGPT 4.0 when utilized with a well-designed instruction prompt. Additionally, health information content analysis scoring was variable between ChatGPT raters. However, ChatGPT 4.0 demonstrated lower omission rates and higher accuracy rates than ChatGPT 3.5. Our findings demonstrate the utility of ChatGPT in automating important elements of plastic surgery research and clinical practice.

Suboptimal readability of online health information for plastic surgery patients is an ongoing challenge.[1-5,10-13] Prior recommendations for improving readability have focused on concrete guidelines to writing, including the use of simple words, short sentences, and various grammar/writing webtools.[1-5,10,12] In the present study, we automate this process to achieve improved readability more efficiently. It is notable that no significant improvements in readability were achieved

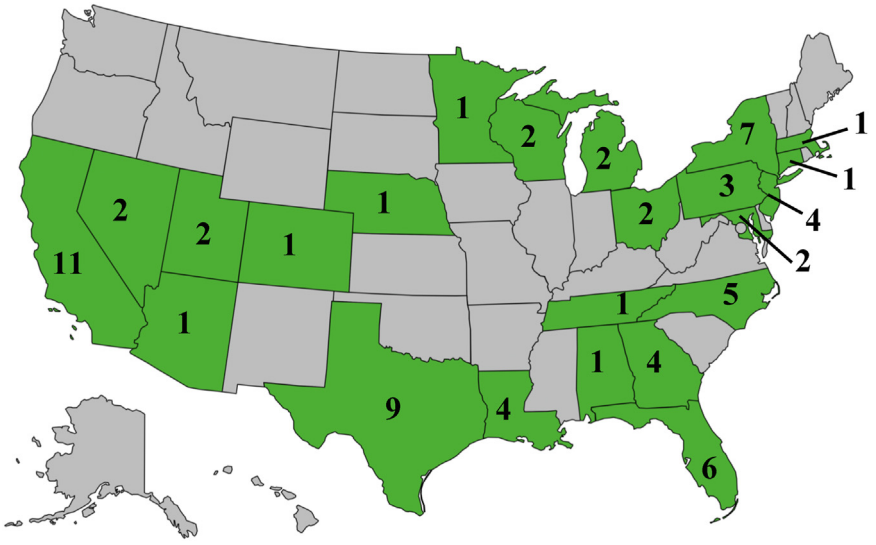

Fig. 2 – A map of the United States displaying the distribution by US state of 73 plastic surgery private practices represented by the 75 included webpages.

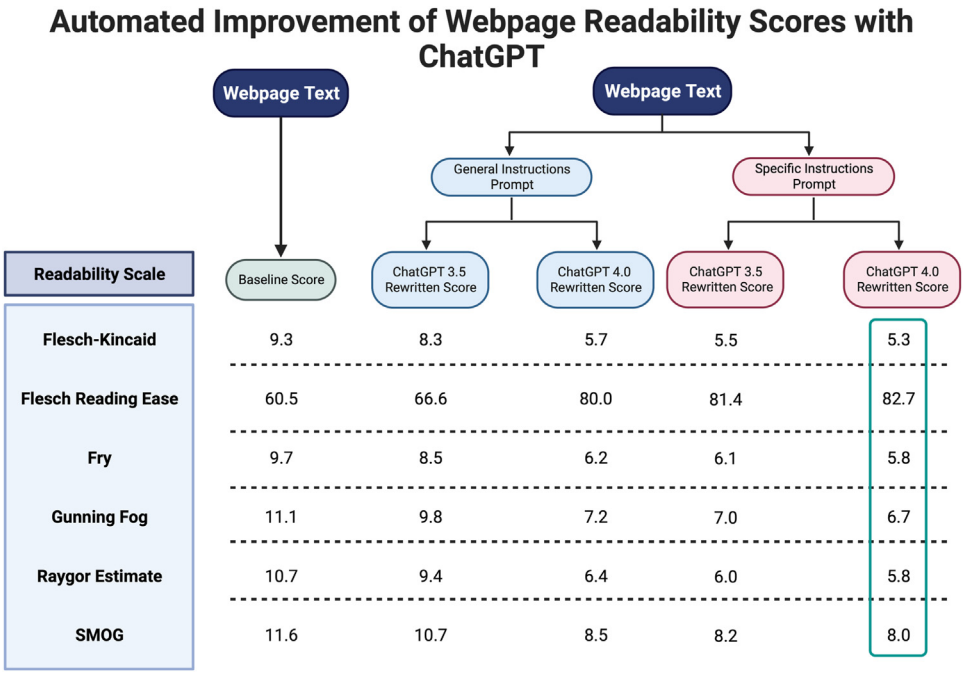

Fig. 3 — Average readability scores of original webpages and ChatGPT rewritten webpages using six different readability measures. The lowest readability scores were achieved with ChatGPT4.0 and a specific instructions prompt (circled), but scores were not significantly different from ChatGPT 3.5 with a specific instructions prompt or ChatGPT 4.0 with a general instructions prompt.

beyond ChatGPT 3.5 with a detailed instruction prompt. Our findings demonstrate that well-designed instruction prompts enable the baseline model to achieve the same results as the subscription model for readability improvement. This reflected that ChatGPT performance in research tasks can be optimized with iterative prompt design. We improved upon the general prompt by providing specific guidelines for rewriting webpage texts. These included references to readability measures, specific instructions for writing style (shortening sentences and simplifying words), and a goal readability level. Importantly, we specified that the rewritten text should maintain the original meaning and information to constrain the writing style. We utilized a validation dataset of eight webpages to confirm that ChatGPT-generated texts could improve readability without altering decision-making factors reported in the original webpages. Iteration of instruction prompts with language models such as ChatGPT for rewriting health online information offers several key advantages. These advantages include quickly improving the readability of one or more text samples with a single prompt (efficiency), generating content that consistently incorporates all stated guidelines (standardization), and enabling clinicians and health experts to write in their original style (feasibility).

While we demonstrated that text rewritten by ChatGPT to improve readability did not alter the medical information present in original webpage texts, creation and dissemination of medical misinformation by machine learning programs remains an unresolved concern with real potential for patient harm. Constraining machine learning performance with well-designed instructions prompts and mandating close review of information generated by machine learning programs are both necessary to ensure the safety of tools such as ChatGPT in research and clinical practice. Interestingly, clinicians can be further supported in this task through a strategy known as weak-to-strong generalization, in which simple machine learning programs are used to supervise the performance of larger, sophisticated machine learning programs.[33] Creating several checkpoints for medical information quality assurance can ensure that machine learning tools do not compromise patient education and decision-making.

We also employed ChatGPT to automate an analysis of health information contained in plastic surgery webpages. ChatGPT has previously been evaluated for identification of keywords in published medical research which prompted us to analyze its role in qualitative content scoring by use of multiple keywords and phrases.[34] Notably, the average rate of omission in content scoring decreased and the average rate of accurate content scoring increased with ChatGPT 4.0 compared to ChatGPT 3.5. In contrast, the average rate of hallucination in content scoring was stable between all four evaluators. The improved performance that we observed with ChatGPT 4.0 compared to ChatGPT 3.5 in content analysis, but not readability analysis, reflects that qualitative text scoring is a more subjective, analytical task than rewriting text with clear guidelines. For this reason, complex tasks in research and clinical practice are better suited to advanced language models or still outside the scope of their capabilities.

The capacity for medical misinformation and patient harm due to unacceptable levels of omission and hallucination by machine learning programs highlight the need for clinician oversight and continual improvement. The variable performance of content scoring for individual patient decision-making factors that we observed may reflect that some qualitative features of text are more identifiable by keywords,

Decision Making Factors	Performance	ChatGPT 3.5		ChatGPT 4.0		P-Value
		General Instructions	Specific Instructions	General Instructions	Specific Instructions	
Surgeon Input n (%)	Omission	22 (29.3)	41 (54.7)	**0 (0)**	2 (2.7)	**< 0.001**
	Accuracy	49 (65.3)	32 (42.7)	**66 (88.0)**	65 (86.7)	
	Hallucination	4 (5.3)	**2 (2.7)**	9 (12.0)	8 (10.7)	
Lifestyle n (%)	Omission	19 (25.3)	10 (13.3)	**5 (6.7)**	**5 (6.7)**	**0.001**
	Accuracy	51 (68.0)	45 (60.0)	65 (86.7)	**66 (88.0)**	
	Hallucination	5 (6.7)	20 (26.7)	5 (6.7)	**4 (5.3)**	
Complications and Adverse Outcomes n (%)	Omission	16 (21.3)	11 (14.7)	2 (2.7)	**1 (1.3)**	**0.010**
	Accuracy	49 (65.3)	52 (69.3)	**62 (82.7)**	60 (80.0)	
	Hallucination	**10 (13.3)**	12 (16.0)	11 (14.7)	14 (18.7)	
Fertility and Aging n (%)	Omission	9 (12.0)	10 (13.3)	**5 (6.7)**	6 (8.0)	0.893
	Accuracy	63 (84.0)	61 (81.3)	**70 (93.3)**	66 (88.0)	
	Hallucination	3 (4.0)	4 (5.3)	**0 (0)**	3 (4.0)	
Consultation with Loved Ones n (%)	Omission	6 (8.0)	6 (8.0)	**3 (4.0)**	4 (5.3)	0.486
	Accuracy	67 (89.3)	68 (90.7)	**69 (92.0)**	68 (90.7)	
	Hallucination	2 (2.7)	**1 (1.3)**	3 (4.0)	3 (4.0)	
Conception of Breast Size n (%)	Omission	18 (24.0)	3 (4.0)	**0 (0)**	1 (1.3)	**< 0.001**
	Accuracy	46 (61.3)	**55 (73.3)**	53 (70.7)	52 (69.3)	
	Hallucination	**11 (14.7)**	17 (22.7)	22 (29.3)	22 (29.3)	
Aesthetic Goals and Motivations n (%)	Omission	**5 (6.7)**	6 (8.0)	9 (12.0)	20 (26.7)	**< 0.001**
	Accuracy	45 (60.0)	49 (65.3)	**51 (68.0)**	47 (62.7)	
	Hallucination	25 (33.3)	20 (26.7)	15 (20.0)	**8 (10.7)**	
Correction of Breast Asymmetry n (%)	Omission	6 (8.0)	7 (9.3)	2 (2.7)	**1 (1.3)**	**0.049**
	Accuracy	64 (85.3)	66 (88.0)	65 (86.7)	**70 (93.3)**	
	Hallucination	5 (6.7)	**2 (2.7)**	8 (10.7)	4 (5.3)	
Body and Tissue-Based Measurements n (%)	Omission	25 (33.3)	18 (24.0)	4 (5.3)	**2 (2.7)**	**< 0.001**
	Accuracy	48 (64.0)	51 (68.0)	**66 (88.0)**	**66 (88.0)**	
	Hallucination	**2 (2.7)**	6 (8.0)	5 (6.7)	7 (9.3)	
Patient Decision Support Devices n (%)	Omission	17 (22.7)	29 (38.7)	**4 (5.3)**	7 (9.3)	**< 0.001**
	Accuracy	56 (74.7)	46 (61.3)	**68 (90.7)**	66 (88.0)	
	Hallucination	2 (2.7)	**0 (0)**	3 (4.0)	2 (2.7)	
Other Implant Features n (%)	Omission	9 (12.0)	12 (16.0)	**1 (1.3)**	3 (4.0)	**0.001**
	Accuracy	56 (74.7)	**57 (76.0)**	56 (74.1)	54 (72.0)	
	Hallucination	10 (13.3)	**6 (8.0)**	18 (24.0)	18 (24.0)	
Procedural Considerations n (%)	Omission	6 (8.0)	5 (6.7)	**1 (1.3)**	4 (5.3)	0.094
	Accuracy	52 (69.3)	48 (64.0)	54 (72.0)	**62 (82.7)**	
	Hallucination	17 (22.7)	22 (29.3)	20 (26.7)	**9 (12.0)**	

Fig. 4 – Content analysis scoring performance of ChatGPT measured by rates of omission, accuracy, and hallucination for detection of 12 different patient decision-making factors of breast implant size selection. Scoring performance was significantly different for nine factors.

phrases, and examples than others. We observed that some decision-making factors appeared more susceptible to omission or hallucination by ChatGPT-automated content analysis. This may be due to not only differences in text semantics of each decision-making factor, but also to how decision-making factors were represented or framed by our instructions

prompts. For this reason, we likely could have targeted these decision-making factors to ensure higher rates of accuracy in content scoring with continued interactions of instructions prompts. We performed one round of iteration in prompt design by expanding the number of text examples and keywords for decision-making factors. However, other strategies such as providing negative examples and keywords, typical locations of decision-making factors in the text, and refinement of existing examples may have achieved improved accuracy in content scoring.

Continued research validating the use of language learning models in research and clinical tasks will likely provide more strategies for improving performance for analytic tasks in research and clinical practice. As the capabilities of ChatGPT and other language learning models continue to expand, the scope and performance of these models will improve but must be continuously evaluated.[35] In recent time, OpenAI has introduced capabilities of "voice and vision" to ChatGPT that will further expand its potential in medical research and clinical practice.[36] These capabilities enable the use of ChatGPT to evaluate audio, image, and video samples. Importantly, video and reel-based content have become major sources of online medical information coupled to the rise of social media platforms such as YouTube, Instagram, and TikTok.[6-8,11] Future use of ChatGPT for evaluating patient resources in surgery should also include analysis of medical information from formal sources such as academic institutions and professional societies. Additionally, ChatGPT may prove useful in other forms of text-analyses, including the identification of medical information that is incomplete or inaccurate. These functions would empower both clinicians and patients to identify resources with an acceptable level of quality and accuracy.

Though we demonstrate the use of ChatGPT to accomplish two tasks relevant to plastic surgery research and clinical practice, this study is not without limitations. First, our evaluation of ChatGPT performance is limited to 75 webpage samples. A larger sample size or sample size of different online health information sources may change its performance in these tasks. Our analysis was limited to private practice webpages and did not include webpages authored by academic institutions or surgical societies. Additionally, we utilized a previously described set of patient decision-making factors for breast implant size selection. Use of a validated set of factors to guide content analysis enabled us to include more keywords, phrases, and examples into instruction prompts that could have inflated ChatGPT performance. Finally, in our evaluation of readability improvement and content analysis, we only performed one round of instruction prompt iteration and may have achieved improved, consistent performance with further attempts to improve prompts.

Conclusions

Continuous evaluation and improvement of online health information remains an important task for supporting patient education and shared decision making in plastic surgery. By utilizing ChatGPT to improve plastic surgery practice website readability and automate content analysis, we demonstrate a model for possible use of this technology to improve plastic surgery patient resources. Suboptimal performance in content analysis compared to readability improvement suggests that this emergent technology must be further refined to accomplish complex, analytic tasks for plastic surgeons. Still, its future capabilities continue to expand as ChatGPT assimilates additional functions. Continued studies of artificial intelligence and machine learning in plastic surgery will identify a growing number of roles to augment research and clinical practice.

Supplementary Materials

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jss.2024.04.006.

Disclosure

The authors have no relevant financial, professional, or personal disclosures to report.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of Data

Data available upon request.

Meeting Presentation

This manuscript was accepted for Quickshot Presentation at the 19th Annual Academic Surgical Congress to be held February 6-8, 2024 in Washington, DC.

CRediT authorship contribution statement

James E. Fanning: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Maria J. Escobar-Domingo:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jose Foppiani:** Writing – review & editing, Validation, Methodology, Data curation, Conceptualization. **Daniela Lee:** Writing – review & editing, Writing – original draft, Validation, Investigation, Data curation, Conceptualization. **Amitai S. Miller:** Writing – review & editing, Writing – original draft, Investigation, Data curation. **Jeffrey E. Janis:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization. **Bernard T. Lee:** Writing – review & editing, Validation, Supervision, Project administration, Investigation, Conceptualization.

REFERENCES

1. Patel AA, Joshi C, Varghese J, Hassan AM, Janis JE, Galiano RD. Do websites serve our patients well? A comparative analysis of online information on cosmetic injectables. *Plast Reconstr Surg*. 2022;149:655e–668e. https://doi.org/10.1097/PRS.0000000000008921.
2. Tiourin E, Barton N, Janis JE. Health literacy in plastic surgery: a scoping review. *Plast Reconstr Surg Glob Open*. 2022;10:e4247. https://doi.org/10.1097/GOX.0000000000004247.
3. Barton N, Janis JE. Missing the mark: the state of health care literacy in plastic surgery. *Plast Reconstr Surg Glob Open*. 2020;8:e2856. https://doi.org/10.1097/GOX.0000000000002856.
4. Powell LE, Andersen ES, Pozez AL. Assessing readability of patient education materials on breast reconstruction by major US academic hospitals as compared with nonacademic sites. *Ann Plast Surg*. 2021;86:610–614. https://doi.org/10.1097/SAP.0000000000002575.
5. Aliu O, Chung KC. Readability of ASPS and ASAPS educational web sites: an analysis of consumer impact. *Plast Reconstr Surg*. 2010;125:1271–1278. https://doi.org/10.1097/PRS.0b013e3181d0ab9e.
6. Montemurro P, Porcnik A, Hedén P, Otte M. The influence of social media and easily accessible online information on the aesthetic plastic surgery practice: literature review and our own experience. *Aesthet Plast Surg*. 2015;39:270–277. https://doi.org/10.1007/s00266-015-0454-3.
7. Sorice SC, Li AY, Gilstrap J, Canales FL, Furnas HJ. Social media and the plastic surgery patient. *Plast Reconstr Surg*. 2017;140:1047–1056. https://doi.org/10.1097/PRS.0000000000003769.
8. Janik PE, Charytonowicz M, Szczyt M, Miszczyk J. Internet and social media as a source of information about plastic surgery: comparison between public and private sector, A 2-center study. *Plast Reconstr Surg Glob Open*. 2019;7:e2127. https://doi.org/10.1097/GOX.0000000000002127.
9. Didie ER, Sarwer DB. Factors that influence the decision to undergo cosmetic breast augmentation surgery. *J Womens Health (Larchmt)*. 2003;12:241–253. https://doi.org/10.1089/154099003321667582.
10. Fanning JE, Okamoto LA, Levine EC, McGee SA, Janis JE. Content and readability of online recommendations for breast implant size selection. *Plast Reconstr Surg Glob Open*. 2023;11:e4787. https://doi.org/10.1097/GOX.0000000000004787.
11. Patel AA, Mulvihill L, Jin A, Patel A, Galiano RD. Websites or videos: which offer better information for patients? A comparative analysis of the quality of YouTube videos and websites for cosmetic injectables. *Plast Reconstr Surg*. 2022;149:596–606. https://doi.org/10.1097/PRS.0000000000008827.
12. Rayess H, Zuliani GF, Gupta A, et al. Critical analysis of the quality, readability, and technical aspects of online information provided for neck-Lifts. *JAMA Facial Plast Surg*. 2017;19:115–120. https://doi.org/10.1001/jamafacial.2016.1219.
13. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor—let the reader and viewer beware. *JAMA*. 1997;277:1244–1245.
14. DISCERN. The DISCERN instrument. Available at: http://www.discern.org.uk/discern_instrument.php. Accessed September 6, 2023.
15. Health On the Net (HON). Health on the net code of conduct (HONcode). Available at: https://www.hon.ch/HONcode/. Accessed September 6, 2023.
16. OpenAI. ChatGPT: optimizing language models for dialogue. 2022. Available at: https://openai.com/blog/chatgpt/.
17. Bellinger JR, De La Chapa JS, Kwak MW, Ramos GA, Morrison D, Kesser BW. BPPV Information on Google versus AI (ChatGPT) [e-pub ahead of print] *Otolaryngol Head Neck Surg*. 2023. https://doi.org/10.1002/ohn.506.
18. Liu HY, Alessandri Bonetti M, Jeong T, Pandya S, Nguyen VT, Egro FM. Dr. ChatGPT will see you now: how do Google and ChatGPT compare in answering patient questions on breast reconstruction? *J Plast Reconstr Aesthet Surg*. 2023;85:488–497. https://doi.org/10.1016/j.bjps.2023.07.039.
19. Shao CY, Li H, Liu XL, et al. Appropriateness and comprehensiveness of using ChatGPT for perioperative patient education in thoracic surgery in different language contexts: survey study. *Interact J Med Res*. 2023;12:e46900. https://doi.org/10.2196/46900.
20. Ayoub NF, Lee YJ, Grimm D, Divi V. Head-to-Head comparison of ChatGPT versus Google search for medical knowledge acquisition [e-pub ahead of print] *Otolaryngol Head Neck Surg*. 2023. https://doi.org/10.1002/ohn.465.
21. Gabriel J, Shafik L, Alanbuki A, Larner T. The utility of the ChatGPT artificial intelligence tool for patient education and enquiry in robotic radical prostatectomy. *Int Urol Nephrol*. 2023:2717–2732. https://doi.org/10.1007/s11255-023-03729-4.
22. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. *J Bone Joint Surg Am*. 2023;105:1519–1526. https://doi.org/10.2106/JBJS.23.00209.
23. Moazzam Z, Lima HA, Endo Y, Noria S, Needleman B, Pawlik TM. A paradigm shift: online artificial intelligence platforms as an informational resource in Bariatric surgery. *Obes Surg*. 2023;33:2611–2614. https://doi.org/10.1007/s11695-023-06675-3.
24. Jeha GM, Qiblawi S, Jairath N, et al. ChatGPT and generative artificial intelligence in mohs surgery: a New frontier of innovation. *J Invest Dermatol*. 2023;143:2105–2107. https://doi.org/10.1016/j.jid.2023.05.018.
25. Seth I, Cox A, Xie Y, et al. Evaluating chatbot efficacy for answering frequently asked questions in plastic surgery: a ChatGPT case study focused on breast augmentation. *Aesthet Surg J*. 2023;43:1126–1135. https://doi.org/10.1093/asj/sjad140.
26. ElHawary H, Gorgy A, Janis JE. Large language models in academic plastic surgery: the way forward. *Plast Reconstr Surg Glob Open*. 2023;11:e4949. https://doi.org/10.1097/GOX.0000000000004949.
27. Moons P, Van Bulck L. Using ChatGPT and Google Bard to improve the readability of written patient information: a proof-of-concept. *Eur J Cardiovasc Nurs*. 2023;23:122–126. https://doi.org/10.1093/eurjcn/zvad087.
28. Haman M, Školník M. Testing ChatGPT's capabilities for social media content analysis [e-pub ahead of print] *Aesthetic Plast Surg*. 2023. https://doi.org/10.1007/s00266-023-03607-5.
29. Research: ChatGPT-4. 2023. Available at: https://www.openai.com/research/gpt-4. Accessed November 1, 2023.
30. Startpage private search engine. 2022. Available at: http://www.startpage.com. Accessed September 1, 2023.
31. *Readability studio professional edition 2019.4, Oleander Software*; 2019. Available at: https://oleandersolutions.com/readabilitystudio.html. Accessed September 1, 2023.
32. Introducing ChatGPT plus. OpenAI Blog. 2023. Available at: https://openai.com/blog/chatgpt-plus. Accessed October 1, 2023.
33. Weak-to-Strong generalization. OpenAI Blog. 2023. Available at: https://openai.com/research/weak-to-strong-generalization. Accessed February 1, 2024.

34. Zhou J, Jia Y, Qiu Y, Lin L. The potential of applying ChatGPT to extract keywords of medical literature in plastic surgery. *Aesthet Surg J*. 2023;43:NP720—NP723. https://doi.org/10.1093/asj/sjad158.
35. Humar P, Asaad M, Bengur FB, Nguyen V. ChatGPT is equivalent to first year plastic surgery residents: evaluation of ChatGPT on the plastic surgery in-service exam. *Aesthet Surg J*. 2023;43:NP1085—NP1089. https://doi.org/10.1093/asj/sjad130.
36. ChatGPT can now see, hear, and speak. OpenAI Blog. 2023. Available at: https://openai.com/blog/chatgpt-can-now-see-hear-and-speak. Accessed October 1, 2023.