Original Article

Analyzing the precision and readability of a healthcare focused artificial intelligence platform on common questions regarding breast augmentation

Carter J. Boyd, Lucas R. Perez Rivera, Kshipra Hemal, Thomas J. Sorenson, Chris Amro, Mihye Choi, Nolan S. Karp

Hansjörg Wyss Department of Plastic Surgery, NYU Langone Health, New York, NY 10017, USA.

Correspondence to: Prof. Nolan S. Karp, Hansjörg Wyss Department of Plastic Surgery, NYU Langone Health, 305 East 47th, suite 1A, New York, NY 10017, USA. E-mail: Nolan.karp@nyulangone.org

Abstract

Aim: The purpose of this study was to determine the quality and accessibility of the outputs from a healthcare-specific artificial intelligence (AI) platform for common questions during the perioperative period for a common plastic surgery procedure.

Methods: Doximity GPT (Doximity, San Francisco, CA) and ChatGPT 3.5 (OpenAI, San Francisco, CA) were utilized to search 20 common perioperative patient inquiries regarding breast augmentation. The structure, content, and readability of responses were compared using *t*-tests and chi-square tests, with $P < 0.05$ used as the cutoff for significance.

Results: Out of 80 total AI-generated outputs, ChatGPT responses were significantly longer (331 *vs.* 218 words, $P < 0.001$). Doximity GPT outputs were structured as a letter from a medical provider to the patient, whereas ChatGPT outputs were a bulleted list. Doximity GPT outputs were significantly more readable by four validated scales: Flesch Kincaid Reading Ease (42.6 *vs.* 29.9, $P < 0.001$) and Flesch Kincaid Grade Level (11.4 *vs.* 14.1 grade, $P < 0.001$), Coleman-Liau Index (14.9 *vs.* 17 grade, $P < 0.001$), and Automated Readability Index (11.3 *vs.* 14.8 grade, $P < 0.001$). Regarding content, there was no difference between the two platforms regarding the appropriateness of the topic (99% overall). Medical advice from all outputs was deemed reasonable.

Conclusion: Doximity's AI platform produces reasonable, accurate information in response to common patient queries. With continued reinforcement learning with human feedback (RLHF), Doximity GPT has the potential to be a useful tool to plastic surgeons and can assist with a range of tasks, such as providing basic information on procedures and writing appeal letters to insurance providers.

Keywords: Artificial intelligence, natural language processing, ChatGPT, AI, generative AI, plastic surgery, AI integration in surgery

INTRODUCTION

Artificial intelligence (AI) is positioned to be a disruptive technology for healthcare. Over the past several years, AI has demonstrated exponential growth and transitioned from a theoretical idea to a tangible reality. This excitement for AI has blossomed within plastic surgery as the literature denotes countless descriptions of use cases, potential applications, and discussions of ethical implementation[1-4]. While the indications for AI-powered assistance are innumerable within plastic surgery, natural language processing (NLP) - a form of generative AI - is poised to be rapidly integrated into the clinical workflow for plastic surgeons, practice managers, and staff. The NLP capabilities of AI allow it to analyze, comprehend, and produce language[5]. Prior work analyzing an NLP program demonstrated efficacy in generating medically sound recommendations for common patient inquiries surrounding perioperative surgery for a breast reduction procedure[6]. Despite the integrity of medical recommendations, this study highlighted limitations in the accessibility of AI-generated outputs with a higher-than-recommended average reading level of content[7-10]. Such analyses have been duplicated for a variety of large language models (LLMs)[11-13].

Limitations of this prior work include the assessment of LLMs without a specific emphasis or additional training in medicine. Doximity has launched a healthcare-specific, Health Insurance Portability and Accountability Act (HIPAA)-compliant AI tool. This platform, Doximity GPT, was created to facilitate written outputs specifically designed for healthcare, including patient instructions, appeals to insurance providers, and educational pamphlets. While backed by the same NLP program as ChatGPT (OpenAI, San Francisco, CA), Doximity GPT incorporates additional healthcare-specific training and utilizes reinforcement learning with human feedback (RLHF) to improve ChatGPT's programming.

As it is being marketed as having specific training in healthcare, one of the aims of this study was to assess how this might impact outputs from the LLM. To date, no studies have assessed the product of this healthcare-specific NLP program. In this study, we sought to compare the quality and accessibility of outputs from Doximity GPT to outputs from a generic NLP program in response to common questions about breast augmentation procedures, which are among the most frequently performed procedures in the United States[14].

METHODS

The new HIPAA-compliant, healthcare-specific AI platform Doximity GPT (Doximity, San Francisco, CA) is readily accessible to physicians. This AI program was compared to the publicly available NLP ChatGPT 3.5 (OpenAI, San Francisco, CA). Both AI interfaces were accessed in April 2024. A list of 20 frequent patient inquiries regarding a breast augmentation procedure was generated. The list was adapted from a previously published study, following the clinical experience of the senior authors (N.S.K. and M.C.) and their expertise in perioperative management[6]. Breast augmentation was selected as it is one of the most frequently performed procedures in the United States annually[14]. Inquiries were entered into each of the two NLPs in two distinct formats: utilizing a general search term ("breast augmentation driving") and

Table 1. Listing of inputs into the NLP interfaces

Question number	General inquiry	Specific inquiry
1	Breast augmentation bruising	I had a breast augmentation yesterday and now I have bruising. What should I do?
2	Breast augmentation bleeding	I had a breast augmentation yesterday and now I have bleeding. What should I do?
3	Breast augmentation size	I had a breast augmentation yesterday and am concerned about the size of my breasts. What should I do?
4	Breast augmentation swelling	I had a breast augmentation yesterday and now I have swelling. What should I do?
5	Breast augmentation soreness	I had a breast augmentation yesterday and now I have soreness. What should I do?
6	Breast augmentation exercise	I had a breast augmentation yesterday. When can I exercise?
7	Breast augmentation driving	I had a breast augmentation yesterday. When can I drive?
8	Breast augmentation restarting medications	I had a breast augmentation yesterday. When can I restart my normal medications?
9	Breast augmentation pain	I had a breast augmentation yesterday and now I have pain. What should I do?
10	Breast augmentation showering	I had a breast augmentation yesterday. When can I shower?
11	Breast augmentation dressings	I had a breast augmentation yesterday. What do I do with the dressings?
12	Breast augmentation pain medication	I had a breast augmentation yesterday. What should I take for pain medication?
13	Breast augmentation drainage	I had a breast augmentation yesterday and now I have drainage. What should I do?
14	Breast augmentation diet	I had a breast augmentation yesterday. What can I eat?
15	Breast augmentation sleeping	I had a breast augmentation yesterday. How can I sleep?
16	Breast augmentation recovery	I had a breast augmentation yesterday. How long is the recovery?
17	Breast augmentation bra	I had a breast augmentation yesterday. When can I wear a bra?
18	Breast augmentation antibiotics	I had a breast augmentation yesterday. Do I need antibiotics?
19	Breast augmentation breast and nipple sensation	I had a breast augmentation yesterday. Will I still have breast and nipple sensation?
20	Breast augmentation follow-up appointment	I had a breast augmentation yesterday. When is my follow-up appointment?

A total of 20 topic areas were assessed. A general and specific query for each topic was placed into both Doximity GPT and ChatGPT. NLP: Natural language processing.

asking a specific clinical question ("I had a breast augmentation yesterday. When can I drive?"). A list of all input terms is listed in Table 1. A separate session was created for each inquiry to prohibit any influence from prior inputs and to limit the effect of learning by part of the NLP. Objective assessment of the NLP outputs included the length (word length, number of characters, number of sentences) and sentence structure (words per sentence) of the outputs. The readability of the outputs was also determined by four validated scoring tools, including the Flesh Kincaid Reading Ease, Flesch Kincaid Grade Level, Coleman-Liau Index, and the Automated Readability Index. Each of these instruments has been commonly utilized to assess the readability of plastic surgery educational materials for patients[12,15,16]. The Flesch Kincaid Reading Ease score is reported on a scale from 1 to 100, where higher scores denote a more readable passage. The Flesch-Kincaid grade level is another standardized metric employed in similar scenarios but focuses on determining the grade level of schooling necessary for comprehending the output. A Flesch-Kincaid grade level reported as 5 denotes that the completion of fifth grade is required for an adequate understanding of a given passage. While the prior two indices primarily utilize syllables per word as the driving metric to determine readability, the Coleman-Liau Index and Automated Readability Index base their scoring on character-level metrics with subtle nuances in weighting. Similar to the Flesch-Kincaid grade level, the Coleman-Liau Index and Automated Readability Index both report the level of schooling required to understand a given text. Each output was individually analyzed by four distinct reviewers (L.P.R. - senior medical student, T.J.S. - junior plastic surgery resident, C.J.B. - senior plastic surgery resident, and K.H. - senior plastic surgery resident) to determine the medical accuracy of the recommendations for each AI-generated response and cross-referenced based on readily available resources. Given difficulties in

quantifying how accurate an output was, accuracy was assessed in a binary fashion. Any inaccuracy within the LLM output classified the entire output as inaccurate. Any discrepancies were discussed with an independent arbitrator (N.S.K.), who is an expert breast surgeon, until a consensus was reached. Statistical analysis was accomplished using Microsoft Excel (Version 7, Seattle, WA), performing descriptive statistics, *t*-tests, and chi-square tests where appropriate, with a predetermined level of significance of $P < 0.05$.

RESULTS

In total, eighty NLP outputs were included in the study, half from Doximity GPT and the remainder from ChatGPT. ChatGPT responses were longer when measured by word count than Doximity GPT outputs (331 *vs.* 218 words, $P < 0.001$). ChatGPT outputs also utilized more overall characters (1,842 *vs.* 1,139 characters, $P < 0.001$) and total sentences (16.8 *vs.* 13.7 sentences, $P < 0.001$). Sentences were more verbose for ChatGPT outputs, with nearly 4 additional words per sentence compared to texts produced by Doximity GPT (20 *vs.* 16.3 words per sentence, $P < 0.001$) [Table 2].

Considering the overall structure of the outputs, Doximity GPT outputs were, by default, structured as a letter to the patient from a medical provider, whereas ChatGPT generated a bulleted list. Overall, the letter format of Doximity GPT yielded more outputs that appeared more personal and dialogistic in nature. Regarding readability of the outputs, Doximity GPT outputs were more readable by all four validated instruments: Flesch Kincaid Reading Ease (42.6 *vs.* 29.9, $P < 0.001$), Flesch Kincaid Grade Level (11.4 *vs.* 14.1 grade, $P < 0.001$), Coleman-Liau Index (14.9 *vs.* 17 grade, $P < 0.001$), and Automated Readability Index (11.3 *vs.* 14.8 grade, $P < 0.001$). Regarding content, there was no difference between the two platforms regarding the appropriateness of the topic (99% overall). All outputs provided a degree of background medical information on the subject, and 96% also included direct prescriptive advice including contacting the surgeon in 100% of cases and adhering to postoperative instructions in 90% of instances. Medical advice from all AI-generated outputs was deemed reasonable. The full list of outputs is provided for reference in Supplementary Table 1.

DISCUSSION

To our knowledge, this is the first study assessing the performance of the novel healthcare-specific AI platform, Doximity GPT, in the setting of perioperative management following plastic surgery. We compare the performance of Doximity GPT and ChatGPT in responding to common perioperative questions for a breast augmentation procedure based on the accuracy, format, and readability of generated outputs. This work represents necessary fundamental research to establish the fidelity of NLP-generated responses to medically sound recommendations before attempting to integrate this technology into patient-facing clinical practice. This study identifies that Doximity's AI platform produces reasonable, accurate information in response to common patient queries about breast augmentation procedures.

A key difference between the Doximity GPT and ChatGPT-generated outputs was observed in the structure and formatting of the responses. The Doximity GPT outputs were automatically formatted as letters for the patient on behalf of the provider, signed by the account holder who entered the query. This aligns with the purpose of the Doximity GPT platform, which is free for all U.S. clinicians and medical students and intended to facilitate the creation of patient education materials, note templates, and other administrative tasks. On the other hand, the outputs generated by ChatGPT defaulted to bulleted lists, consistent with an open-access virtual platform. This difference may be explained by the additional training of Doximity GPT with healthcare documentation examples.

Table 2. Objective structural metrics and readability scores for AI generated outputs from Doximity GPT and ChatGPT

Variable	Doximity GPT	ChatGPT	*P*-value
Word count	218 ± 43	331 ± 48	< 0.001
Total characters	1,139 ± 223	1,842 ± 266	< 0.001
Total sentences	13.7 ± 3.4	16.8 ± 2.9	< 0.001
Words per sentence	16.3 ± 2.3	20 ± 3.6	< 0.001
Flesch Kincaid reading ease	42.6 ± 9.5	29.9 ± 7.2	< 0.001
Flesch Kincaid grade level	11.4 ± 1.5	14.1 ± 1.6	< 0.001
Coleman Liau index	14.9 ± 1.6	17 ± 1.1	< 0.001
Automated readability index	11.3 ± 1.7	14.8 ± 1.9	< 0.001

AI: Artificial intelligence.

While ChatGPT responses provide more detailed information for each query, Doximity GPT outputs were determined to be significantly more readable. Still, readability remains a limitation with NLP-generated outputs as both LLMs generated responses at a reading level higher than national recommendations[17]. With continued RLHF, Doximity GPT has the potential to be a useful tool for plastic surgeons and can assist with a range of tasks, such as providing basic information on procedures and writing appeal letters to insurance providers.

Excitement regarding the possibility of incorporating NLPs into clinical workflow is evidenced by an exponential rise of exploratory papers in the literature discussing potential applications. These studies have focused on analyzing and comparing generic NLPs against one another. Garg *et al.* compared ChatGPT to Google Bard (Google, Mountain View, CA) for outputs regarding patient education materials for facial aesthetic surgery. This group specifically requested outputs be at the eighth-grade reading level. Despite this request to the LLMs, the generated outputs had an average reading level at the tenth-grade level[18]. Lim and associates analyzed four generic LLMs to determine the applicability of AI-generated outputs for common perioperative questions for patients undergoing abdominoplasty. All LLMs generated information higher than the national recommended reading levels for medical literature. This group also investigated more subjective aspects of the AI-generated outputs, such as patient friendliness, which may be an important feature if such technology is integrated in a direct patient-facing manner[12]. In terms of improving the readability of content, Vallurupalli *et al.* suggest that LLMs may function more efficiently in simplifying pre-written patient instructions to an appropriate reading level compared to producing novel outputs at the recommended reading level by the National Institute of Health[11]. Further assessment of this theory represents future work from our group.

While Doximity GPT represents a novel, healthcare-specific LLM, it has several limitations. First, Doximity GPT lacks working knowledge for information published after September 2021. This is a limitation not unique to Doximity GPT but common among LLMs. ChatGPT 3.5, for instance, is temporally limited to information published prior to January 2022. Given there is only a three-month difference between the two LLMs, the impact of this variance on their working knowledge is likely limited. In the analysis and comparison of LLMs, it is essential to consider the temporal limitations of each LLM. Continual programming of NLPs is required to maintain the most up-to-date programs as medical knowledge constantly evolves. This highlights an important consideration when using any AI-powered tool: each platform must be employed with awareness of its temporal limitations in knowledge. Furthermore, while the Doximity GPT markets a healthcare-trained LLM, the details of what additional functionality this provides are unclear given its proprietary nature. Another limitation of Doximity GPT is that while it has specific medical reinforcement, it does not have specific plastic surgery training or reinforcement. Studies

have demonstrated a lack of sufficient understanding and knowledge of plastic surgery within the broader healthcare workforce[19,20]. A plastic surgery-specific AI tool or an NLP with additional plastic surgery training represents an opportunity to improve the knowledge and applicability of LLM integration within plastic surgery. Analyses of generic AI chatbots on plastic surgery in-service training examinations, for example, have demonstrated a wide range of accuracy, scoring at levels comparable to a first-year plastic surgery trainee[21,22]. The creation of a specialty-specific LLM has been previously explored, particularly in the field of otolaryngology, where an ENT-specific LLM called ChatENT was found to outperform existing LLMs and exhibited promise in medical and patient education[23]. An opportunity exists to develop a plastic surgery-focused LLM to deliver the most accurate and accessible information to patients and plastic surgeons alike. This LLM should also be able to be customized by surgeons so that individual surgeon preferences regarding perioperative instructions can be programmed. To ensure safety in the clinical application of these tools, appropriate escalation of patient inquiries for scenarios that merit urgent or emergent medical attention must be incorporated into the AI tool. Patients will inevitably utilize AI platforms to seek medical counsel independent of physician supervision. Patients have long used the Internet for self-diagnosis, self-referral, and research of their conditions[24,25]. Thus, studies of this nature are critical to ensure the reliability and accuracy of AI-generated health information to protect patients from misinformation[26].

Since Doximity GPT and ChatGPT are backed by the same NLP program, they are subject to similar training data biases. While the data were largely deemed clinically reasonable by the study team, previous studies have identified inaccuracies and inadequacies when utilizing ChatGPT to answer common postoperative questions[27,28]. Additionally, the ever-evolving nature of clinical dogmas and accepted practices may not always align with the knowledge cut-off dates of these LLMs. Doximity GPT's knowledge of clinical data extends only until September 2021, so novel medical or surgical information will not be included in any outputs. This highlights the importance of clinicians prioritizing clinical judgment and thoroughly reviewing any AI-generated output prior to distribution to patients.

The ethical implications of incorporating AI tools into plastic surgery practice also warrant further discussion. Previous studies have highlighted the importance of informed consent, privacy protection, bias reduction, and regulation for these technologies[29,30]. Kenig *et al.* described the need for a partnership between physicians and lawmakers when creating guidelines and regulations for the use of AI in clinical practice, to ensure that the highest standards of quality and transparency are upheld[29]. They also suggest the creation of an independent body to aid in the testing and validation of healthcare-specific AI models. Further, these tools must be trained with diverse training data, as bias from training datasets may affect the accuracy of AI-generated responses for patients of diverse backgrounds. Periodic review and validation of AI models used in healthcare can aid in fostering fairness, equity, and higher quality of patient-facing data.

Limitations of this study include the comparison between Doximity GPT and only one other NLP. While ChatGPT has been demonstrated previously to have the highest working knowledge in plastic surgery, this may have changed or evolved since that time[21]. Furthermore, Doximity GPT is powered by the updated ChatGPT 4.0. We elected to use ChatGPT 3.5 for comparison in this study, given it is freely accessible. Differences assessed by this study may be attributable to the subtle nuances between the two versions of the LLM. Assessment for the accuracy of LLM outputs is a time-consuming process to review each output, and it remains difficult to objectively determine accuracy other than relying on the clinical judgment of the study team. Future studies should seek to develop methodologies or tools that can more objectively determine medical accuracy on a broader scale. This difficulty further contributed to limiting the scope of the study, as the study team prioritized critical evaluation of each individual output rather than reviewing

more LLM outputs spanning more plastic surgery procedures. This focus on a singular topic within plastic surgery may not be representative of either LLM's overall working knowledge or capabilities in other topic areas within plastic surgery. Future work will aim to develop an efficient methodology for analyzing a higher volume of LLM outputs so that more individual queries can be readily analyzed and subsequent studies can include a wider range of procedures and topics within plastic surgery. While a distinct session was created to limit the effect of learning from the NLPs, it would be interesting to assess what impact repeated use of the same NLP may have on the fidelity of AI-generated outputs used in a clinical scenario.

In summary, this study represents the first attempt to classify the effectiveness of Doximity GPT at producing medically sound, readable responses for a commonly performed plastic surgery procedure. This initial work highlights the promise of RLHF as a mechanism to establish a plastic surgery-focused AI platform that can help augment the patient experience and assist surgical practices with their clinical workflow[6].

An initial assessment of the medical accuracy and linguistic readability of healthcare-specific AI-generated outputs demonstrated reasonable responses to common patient questions following breast augmentation. Future studies can aim to develop quantifiable methodologies for evaluating the clinical accuracy of AI-generated information for a broader range of surgical procedures. This could inform surgeons across different subspecialties within plastic surgery seeking to incorporate LLMs into their clinical practice and may serve as a useful tool for plastic surgery patients in the 21st century.

DECLARATIONS

Authors' contributions
Made substantial contributions to the conception and design of the study, performed data collection, performed data analysis and interpretation, and participated in manuscript writing and editing: Boyd CJ, Perez Rivera LR, Hemal K, Sorenson TJ, Amro C
Made substantial contributions to the conception and design of the study, assisted in data interpretation, and participated in manuscript writing and editing: Karp NS, Choi M

Availability of data and materials
Data supporting the findings of this study are available in Supplementary Table 1.

Financial support and sponsorship
None.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright

REFERENCES

1. Bogdanovich B, Shah P, Patel PA, Bui T, Boyd CJ. Altmetric analysis of artificial intelligence articles in plastic surgery. *Arch Plast Surg* 2024;51:262-4. DOI PubMed PMC

2. Bui T, Patel PA, Boyd CJ. Altmetric analysis of the online attention directed to artificial intelligence literature in ophthalmology. *Asia Pac J Ophthalmol* 2023;12:625-6. DOI PubMed

3. Kavian JA, Wilkey HL, Patel PA, Boyd CJ. Harvesting the power of artificial intelligence for surgery: uses, implications, and ethical considerations. *Am Surg* 2023;89:5102-4. DOI PubMed

4. Zhu C, Attaluri PK, Wirth PJ, Shaffrey EC, Friedrich JB, Rao VK. Current applications of artificial intelligence in billing practices and clinical plastic surgery. *Plast Reconstr Surg Glob Open* 2024;12:e5939. DOI PubMed PMC

5. Bogdanovich B, Patel PA, Kavian JA, Boyd CJ, Rodriguez ED. ChatGPT for the modern plastic surgeon. *Plast Reconstr Surg* 2023;152:969e-70e. DOI PubMed

6. Boyd CJ, Hemal K, Sorenson TJ, et al. Artificial intelligence as a triage tool during the perioperative period: pilot study of accuracy and accessibility for clinical application. *Plast Reconstr Surg Glob Open* 2024;12:e5580. DOI PubMed PMC

7. Blount T, Moffitt S, Fakhre F, et al. Readability of online materials in Spanish and English for breast reduction insurance coverage. *Aesth Plast Surg* 2024;48:1436-43. DOI PubMed

8. Bruce JC, Batchinsky M, Van Spronsen NR, Sinha I, Bharadia D. Analysis of online materials regarding DIEP and TRAM flap autologous breast reconstruction. *J Plast Reconstr Aesthet Surg* 2023;82:81-91. DOI PubMed

9. Soliman L, Soliman P, Gallo Marin B, Sobti N, Woo AS. Craniosynostosis: are online resources readable? *Cleft Palate Craniofac J* 2024;61:1228-32. DOI PubMed

10. Tiourin E, Barton N, Janis JE. Health literacy in plastic surgery: a scoping review. *Plast Reconstr Surg Glob Open* 2022;10:e4247. DOI PubMed PMC

11. Vallurupalli M, Shah ND, Vyas RM. Optimizing readability of patient-facing hand surgery education materials using chat generative pretrained transformer (ChatGPT) 3.5. *J Hand Surg Am* 2024;49:986-91. DOI PubMed

12. Lim B, Seth I, Cuomo R, et al. Can AI answer my questions? Utilizing artificial intelligence in the perioperative assessment for abdominoplasty patients. *Aesthetic Plast Surg* 2024. DOI PubMed

13. Gomez-Cabello CA, Borna S, Pressman SM, et al. Artificial intelligence in postoperative care: assessing large language models for patient recommendations in plastic surgery. *Healthcare* 2024;12:1083. DOI PubMed PMC

14. Bekisz JM, Boyd CJ, Salibian AA, Choi M, Karp NS. Aesthetic characteristics of the ideal female breast. *Plast Reconstr Surg Glob Open* 2023;11:e4770. DOI PubMed PMC

15. Khan S, Walters RK, Walker AM, et al. The readability of online patient education materials on maxillomandibular advancement surgery. *Sleep Breath* 2024;28:745-51. DOI PubMed

16. Browne R, Hurley CM, Carr S, de Blacam C. Online resources for robin sequence; an analysis of readability. *Cleft Palate Craniofac J* 2024:10556656241234587. DOI PubMed

17. Wasserburg JR, Sayegh F, Sanati-Mehrizy P, Graziano FD, Taub PJ. Cleft care readability: can patients access helpful online resources? *Cleft Palate Craniofac J* 2021;58:1287-93. DOI PubMed

18. Garg N, Campbell DJ, Yang A, et al. Chatbots as patient education resources for aesthetic facial plastic surgery: evaluation of ChatGPT and Google Bard responses. *Facial Plast Surg Aesthet Med* 2024. DOI PubMed

19. Alharbi AA, Al-Thunayyan FS, Alsuhaibani KA, Alharbi KA, Alharbi MA, Arkoubi AY. Perception of primary health care providers of plastic surgery and its influence on referral. *J Family Med Prim Care* 2019;8:225-30. DOI PubMed PMC

20. McGoldrick C, Gordon D. Does plastic surgery have an image problem?: the perception of plastic surgery in an era of general practitioner commissioning. *J Plast Reconstr Aesthet Surg* 2013;66:1635-6. DOI PubMed

21. DiDonna N, Shetty PN, Khan K, Damitz L. Unveiling the potential of AI in plastic surgery education: a comparative study of leading AI platforms' performance on in-training examinations. *Plast Reconstr Surg Glob Open* 2024;12:e5929. DOI PubMed PMC

22. Shah P, Bogdanovich B, Patel PA, Boyd CJ. Assessing the plastic surgery knowledge of three natural language processor artificial intelligence programs. *J Plast Reconstr Aesthet Surg* 2024;88:193-5. DOI PubMed

23. Long C, Subburam D, Lowe K, et al. ChatENT: augmented large language model for expert knowledge retrieval in otolaryngology-head and neck surgery. *Otolaryngol Head Neck Surg* 2024;171:1042-51. DOI PubMed

24. Awad SK, Cowen J, Patel J, et al. Plastic surgeons are underrepresented when searching hospital websites for a hand surgeon. *Plast Reconstr Surg* 2023;151:1055e-8e. DOI PubMed

25. Singh NP, Boyd CJ, Aluri A, et al. One in three chance of finding a plastic surgeon on major hospital websites. *Plast Reconstr Surg Glob Open* 2023;11:e4781. DOI PubMed PMC

26. Boudreau H, Singh N, Boyd CJ. Understanding the impact of social media information and misinformation producers on health information seeking. Comment on "Health information seeking behaviors on social media during the COVID-19 pandemic among american social networking site users: survey study". *J Med Internet Res* 2022;24:e31415. DOI PubMed PMC

27. Dhar S, Kothari D, Vasquez M, et al. The utility and accuracy of ChatGPT in providing post-operative instructions following tonsillectomy: a pilot study. *Int J Pediatr Otorhinolaryngol* 2024;179:111901. DOI PubMed

28. Polat E, Polat YB, Senturk E, et al. Evaluating the accuracy and readability of ChatGPT in providing parental guidance for adenoidectomy, tonsillectomy, and ventilation tube insertion surgery. *Int J Pediatr Otorhinolaryngol* 2024;181:111998. DOI PubMed

29. Kenig N, Monton Echeverria J, Rubi C. Ethics for AI in plastic surgery: guidelines and review. *Aesthetic Plast Surg* 2024;48:2204-9. DOI PubMed

30. Liu HY, Alessandri-Bonetti M, Arellano JA, Egro FM. Can ChatGPT be the plastic surgeon's new digital assistant? A bibliometric analysis and scoping review of ChatGPT in plastic surgery literature. *Aesthetic Plast Surg* 2024;48:1644-52. DOI PubMed