REVIEW

Can ChatGPT Be Used as a Research Assistant and a Patient Consultant in Plastic Surgery? A Review of 3 Key Information Domains

Joseph A. Campolo, NREMT; Daniel Y. Kwon, BS; Peter W. Henderson, MD, MBA, FACS

Abstract

Background. Chat Generative Pretrained Transformer (ChatGPT), a newly developed pretrained artificial intelligence (AI) chatbot, is able to interpret and respond to user-generated questions. As such, many questions have been raised about its potential uses and limitations. While preliminary literature suggests that ChatGPT can be used in medicine as a research assistant and patient consultant, its reliability in providing original and accurate information is still unknown. Therefore, the purpose of this project was to conduct a review on the utility of ChatGPT in plastic surgery.

Methods. On August 25, 2023, a thorough literature search was conducted on PubMed. Papers involving ChatGPT and medical research were included. Papers that were not written in English were excluded. Related papers were evaluated and synthesized into 3 information domains: generating original research topics, summarizing and extracting information from medical literature and databases, and conducting patient consultation.

Results. Out of 57 initial papers, 8 met inclusion criteria. An additional 2 were added based on the references of relevant papers, bringing the total number to 10. ChatGPT can be useful in helping clinicians brainstorm and gain a general understanding of the literature landscape. However, its inability to give patient-specific information and act as a reliable source of information limit its use in patient consultation.

Conclusion. ChatGPT can be a useful tool in the conception of and execution of literature searches and research information retrieval (with increased reliability when queries are specific); however, the technology is currently not reliable enough to be implemented in a clinical setting.

Introduction

Chat Generative Pretrained Transformer (ChatGPT; OpenAI) has gained significant attention across various industries, including medicine, since ChatGPT-3.5 was released in November 2022. An advanced version, ChatGPT-4, was developed in March 2023, which was more capable in handling detailed prompts and providing responses with increased accuracy. ChatGPT utilizes recent developments in machine learning and language processing that enable users to ask questions in a

conversational format and receive prompt and pertinent responses. ChatGPT was trained on multiple data sets, exposing it to a variety of texts, including books and scientific journals.[1,2] As a result, many medical professionals have sought to determine the capabilities of ChatGPT in clinical, educational, and research settings.[3,4,5]

It is speculated that the field of plastic surgery can benefit from advances in artificial intelligence (AI) and machine learning technology. As such, preliminary research has been conducted on ChatGPT's utility; however, it is unclear the degree to which it can be trusted in all information realms. The objective of this study is to aggregate and synthesize all relevant current literature on ChatGPT's ability to generate novel ideas, provide accurate surgical information in a patient consultation, and summarize as well as compare plastic surgery literature. By synthesizing the included papers, this review will illustrate potential uses of ChatGPT in plastic surgery research and patient consultation as well as its efficacy in those areas.

Methods and Materials

On August 25, 2023, a thorough literature search was conducted on PubMed using keywords *ChatGPT*, *blepharoplasty*, *breast augmentation*, *systematic review*, *utility*, *plastic surgery*, *cosmetic surgery*, *plastic and reconstructive surgery*, *rhinoplasty*, and *liposuction*. Papers having to do with the use of ChatGPT in plastic surgery and medical research were included, while others were excluded. The literature search was limited to journals written in English.

The literature regarding ChatGPT was organized into 3 domains: ability to generate original research topics (domain 1), ability to summarize and extract information from medical literature and databases (domain 2), and ability to conduct patient consultation (domain 3) (**Figure 1**). The papers were assessed in each of these domains, and based on this an overall conclusion was made on ChatGPT's utility in plastic surgery.
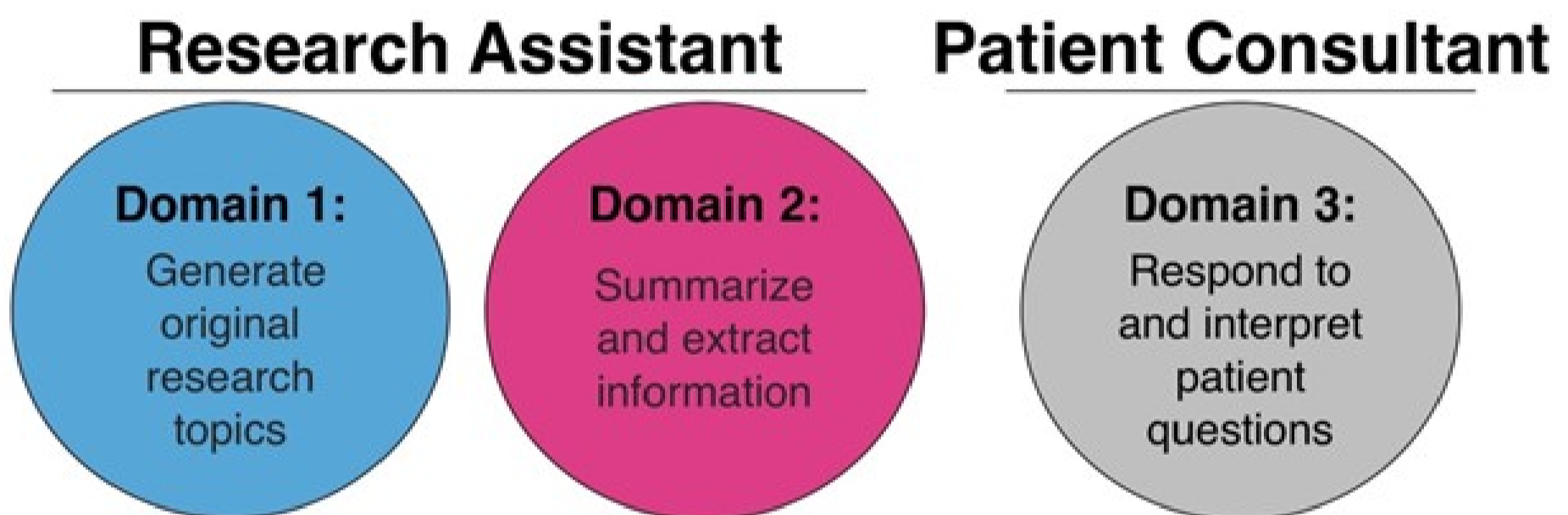

Research Assistant

Domain 1: Generate original research topics

Domain 2: Summarize and extract information

Patient Consultant

Domain 3: Respond to and interpret patient questions

Figure 1. The 3 information domains into which the results of the literature search for ChatGPT were organized.

Results

A total of 57 papers were initially identified. All were screened, and 30 were excluded based on irrelevance to topic. The remaining 27 papers were further reviewed, and this yielded 8 papers that were relevant in both topic and content. Review of the references of the 8 papers led to the identification of 2 additional sources, which brought the total number of papers for this review to 10 (**Figure 2**). Three papers were relevant to domain 1 (generate research topics), 4 papers were relevant to domain 2 (summarize medical information), and 4 papers were relevant to domain 3 (conduct patient consultation). Of note, 1 paper was shared in both domains 2 and 3 (**Figure 3**).
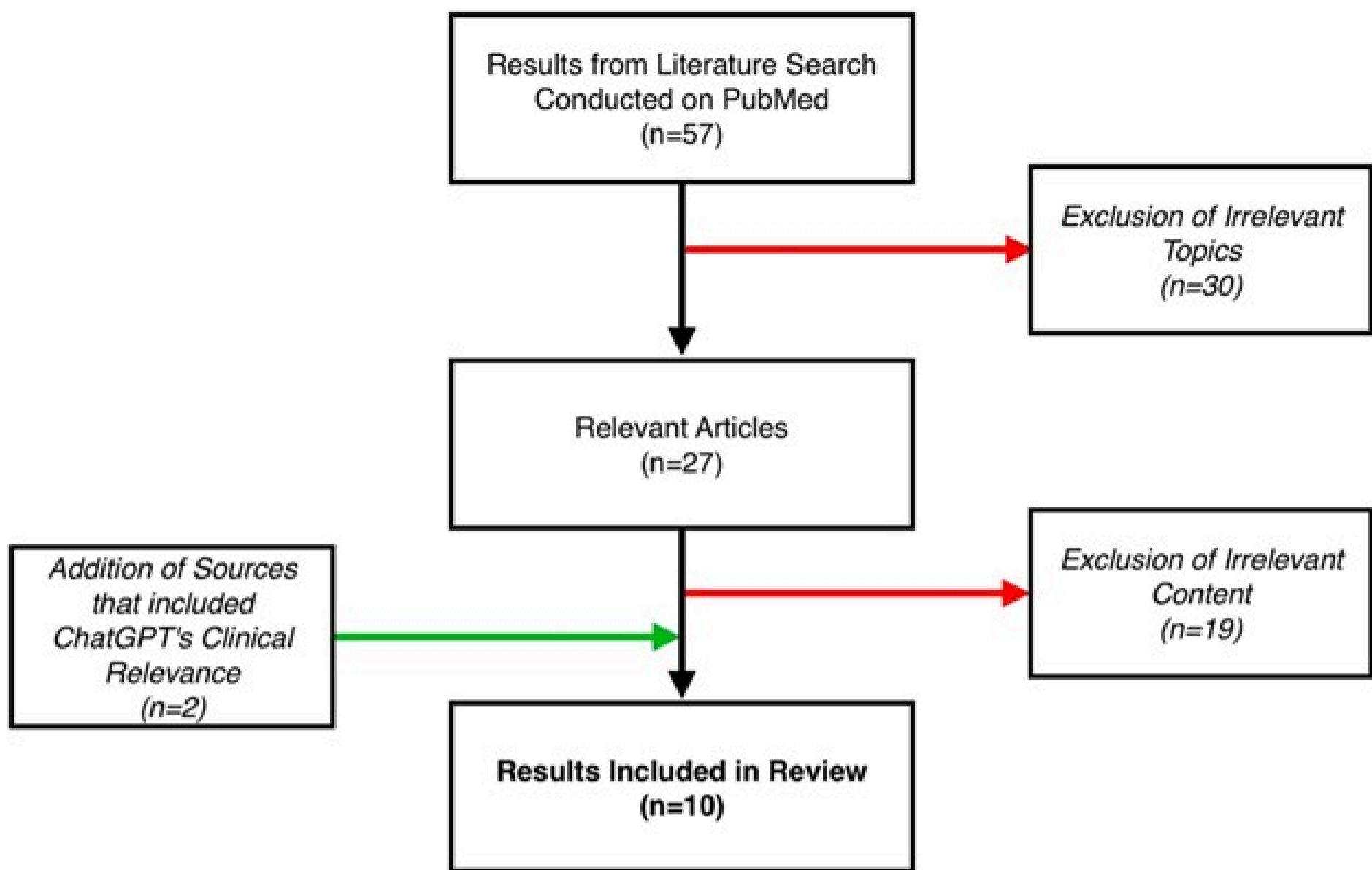
Figure 2. Literature search results.

Domains	Authors	Title	Journal	Institution	Year
Domain 1: Generate original research topics	Gupta, Park et al.	**Expanding cosmetic plastic surgery research with ChatGPT.**	Aesthetic Surgery Journal	Saint Louis University School of Medicine	2023
	Gupta, Pande et al.	**Application of ChatGPT in cosmetic plastic surgery: ally or antagonist?**	Aesthetic Surgery Journal	Saint Louis University School of Medicine	2023
	Gupta, Herzog et al.	**Utilization of ChatGPT for plastic surgery research: friend or foe?**	Journal of Plastic, Reconstructive and Asthetic Surgery	Saint Louis University School of Medicine	2023
Domain 2: Summarize and extract information	Salvagno et al.	**Can artificial intelligence help for scientific writing?**	Critical Care Journal	Université Libre de Bruxelles	2023
	Wu et al.	**ChatGPT in head and neck scientific writing: a precautionary anecdote.**	American Journal of Otolaryngology	Stanford University Hospital	2023
	Zhou et al.	**The potential of applying ChatGPT to extract keywords of medical literature in plastic surgery.**	Aesthetic Surgery Journal	Peking Union Medical College	2023
	Chevernak et al.	**The promise and peril of using a large language model to obtain clinical information: ChatGPT performs strongly as a fertility counseling tool with limitations.**	Fertility and Sterility Journal	Albert Einstein College of Medicine	2023
Domain 3: Respond to and interpret patient questions	Chevernak et al.	*As above*	*As above*	*As above*	*As above*
	Seth et al.	**Evaluating Chatbot efficacy for answering frequently asked questions in plastic surgery: A ChatGPT case study focused on breast augmentation.**	Aesthetic Surgery Journal	Monash University School of Medicine	2023
	Xie et al.	**Aesthetic Surgery Advice and Counseling from Artificial Intelligence: A Rhinoplasty Consultation with ChatGPT.**	Aesthetic Plastic Surgery	Monash University School of Medicine	2023
	Cox et al.	**Utilizing ChatGPT-4 for Providing Medical Information on Blepharoplasties to Patients.**	Aesthetic Surgery Journal	Monash University School of Medicine	2023

Figure 3. Included papers, organized by information domain.

Domain 1: Ability to Generate Original Systematic Review Topics (3 papers)

A significant challenge in plastic surgery is keeping up with current literature and the new and emerging techniques. Current research has evaluated the language processing performance of ChatGPT in generating topics for unpublished systematic reviews, permitting surgeons to readily identify gaps in current literature as well as areas of needed interest. This review identified 3 papers that addressed the use of ChatGPT-3.5 in this capacity.

In order to assess the reliability of ChatGPT-3.5 in producing novel topics, Gupta et al prompted ChatGPT-3.5 with generating 240 unpublished systematic review topics.[6,7,8] There were 12 different topics for the prompts: 6 of the most common surgical and 6 of the most common nonsurgical plastic surgery procedures. For each topic, 20 queries were given: 10 of which were general, and the other 10 more specific. For example, blepharoplasty was a topic with 10 queries having to do with general blepharoplasty and 10 others having to do with retrobulbar hematoma and fat augmentation in blepharoplasty. The novelty of responses was evaluated with literature searches on PubMed, EMBASE, Cochrane, and CINAHL.[6] Topics given by ChatGPT-3.5 that were subsequently found to have published literature were considered an inaccuracy, whereas topics that had no record of published systematic reviews were considered accurate. When compared with current published literature, ChatGPT-3.5 had 55% overall accuracy in producing a topic that had not been published. When queries were more specific (ie, asking for retrobulbar hematoma in blepharoplasty instead of blepharoplasty) ChatGPT-3.5 had 75% accuracy, indicating a greater reliability in topic generation when queries are narrowed.

Gupta et al published other studies evaluating novel topic generation and produced similar results finding that there was a greater accuracy in more developed and specific queries.[7,8] As such, ChatGPT-3.5 demonstrated a rudimentary ability to evaluate the current plastic surgery research landscape and use it to generate unpublished systematic review concepts. Potential is limited in the recency of results as ChatGPT-3.5 is pretrained with literature up to a certain date.[8] This can lead to outputs considered to be novel when in fact there have been publications after the pretrained date. Furthermore, due diligence should be exercised when considering ChatGPT-3.5 topics as some information provided can be incorrect and illogical. Although results are promising, literature evaluating the ability of ChatGPT-3.5 to generate novel systematic review topics is limited. Further research conducted in this area would be beneficial to corroborate such findings.

Domain 2: Ability to Summarize and Extract Information From Medical Literature and Databases (4 papers)

Keeping up with current research and developments is a critical responsibility for surgeons. However, because of the vast and extensive nature of medical literature, investigating and quantifying journal articles is a time-consuming and arduous task. ChatGPT language processing and data set training gives it the ability to extract information from these journals and provide a comprehensive summary of its findings and results. Four papers were identified that addressed the use of ChatGPT-3.5 in this capacity.

Salvagno et al tasked ChatGPT-3.5 with summarizing and comparing 3 journal articles having to do with different techniques for cardiopulmonary resuscitation.[9] ChatGPT-3.5 gave an accurate and prompt summary as well as a general description for each article. It also gave a statement laying out the conclusions and findings for all 3 papers.[9] Because of the lack of specificity in responses, Salvagno et al indicate that ChatGPT-3.5 can be used as a good tool for giving surgeons a quick and general understanding of the literature. While studies like the one conducted by Wu et al have raised concern on the ability of ChatGPT-3.5 to accurately cite papers,[10] Salvagno et al establish its reliability in reviewing and summarizing them. When used as a research assistant, ChatGPT-3.5 can save time in literature searches as well as create a clear outlook of the research landscape.

Zhou et al[11] prompted ChatGPT-3.5 with identifying 10 frequent keywords in papers across 5 different plastic surgery procedures: mammoplasty, blepharoplasty, rhinoplasty, anti-wrinkle, and fat grafting. Responses were compared with keywords that authors of the published journals provided. This was done by using a digital database called Biblioshiny, a bibliometric analysis tool. Outputs from ChatGPT-3.5 were evaluated based on their utility to surgeons and researchers and their similarity to the database's keywords. Answers from ChatGPT-3.5 deviated from that of the bibliometric analysis. However, compared with the keywords provided by bibliometric analysis, Zhou et al determine that the keywords produced by ChatGPT-3.5 were more apropos for scientific research.[11]

Although outside the scope of plastic surgery, a comprehensive look at the uses for ChatGPT-3.5 in extracting clinical data was conducted.[12] Chervenak et al tested its literacy in clinical data in 3 domains: patient consultation, performance on Fertility and Infertility Treatment Knowledge Score (FIT-KS) and Cardiff Fertility Knowledge Scale (CFKS), and ability to recite American Society of Reproductive Medicine (ASRM) facts[12]. The FIT-KS and CFKS are questionnaires given to patients to assess patients' and medical students' general knowledge on fertility statistics based on data from The Society for

Assisted Reproductive Technology. The Chervenak et al study directly assesses the ability of ChatGPT-3.5 to extract and purvey clinical data and information. For performance on FIT-KS and CFKS, ChatGPT-3.5 scored in the 95th and 86th percentiles, respectively. In a similar fashion, ChatGPT-3.5 had a 100% accuracy in providing ASRM facts.[12]

This ability to relay accurate clinical data on fertility indicates the great potential ChatGPT has to become a clinician's tool for fast facts, mitigating searching time. It is worth mentioning that ChatGPT is trained on a large data set, and having a frequently updated medical equivalent could potentially increase reliability and clinical usage.

Domain 3: Ability to Conduct Patient Consultation (4 papers)

The importance of providing patients with information in the preoperative setting can contribute to patient satisfaction as well as the procedure's overall success. As such, being able to purvey readily accessible and accurate medical information is an asset to any surgery. Because of the language processing abilities and substantial data set training of ChatGPT, it has the ability to take patient questions and respond with accurate and pertinent information. Plastic surgeons have evaluated the extent to which ChatGPT can be used in preoperative patient consultation. This review identified 4 papers that addressed the use of ChatGPT in this capacity.

In addition to evaluating the ability of ChatGPT-3.5 to pull facts from health institutions, Chervenak et al also looked at the performance of ChatGPT-3.5 in patient consultation. ChatGPT-3.5 was prompted with frequently asked questions from the Centers for Disease Control and Prevention (CDC) and was compared with the responses written by the CDC. Similar to the results obtained below, Chevernak et al found that responses were mostly accurate with the highest accuracy in responses having to do with medical statistics.[12]

Seth et al assessed the ability of ChatGPT-4 to create and coherently respond to 6 frequently asked questions that patients might have for breast augmentation in both the preoperative and postoperative settings. The relevance and veracity of the responses were assessed by a panel of 3 experienced plastic surgeons. A subsequent literature search was conducted on PubMed and Cochrane to verify response accuracy.[13]

The generated questions were relevant as they were concerned with breast augmentation. ChatGPT responded to all questions accurately while indicating that answers were dependent on case details. The responses also emphasized the importance of following up with the patient's surgeon about the questions.[13] Responses were also explained in a way that patients would understand, increasing the accessibility of medical information as well as taking the burden off surgeons. However, because ChatGPT-4 was unable to provide personalized and case-based answers, the responses were vague and superficial. Seth et al emphasized that ChatGPT-4 cannot be used as a replacement for the operating surgeon's consultation and requires improvement in order to be brought into clinical use.

A similar study conducted by Xie et al assessed the ability of ChatGPT-3.5 to respond to frequently asked questions for rhinoplasty. Xie prompted ChatGPT-3.5 with 9 frequently asked questions about rhinoplasty developed from an American Society of Plastic Surgery rhinoplasty consultation checklist. Answers were evaluated by 4 experienced plastic surgeons in the field of rhinoplasty.[14]

The generated responses were accurate and addressed important concepts. However, Xie et al established that responses were vague, lacking the imperative ability to provide patient-specific information.[14] With this being said, Xie et al concluded that while it can be a useful source of information, further research is required to assess the capacity to which ChatGPT-3.5 can be brought into the clinical setting.

In addition to the previous studies discussed, Cox et al conducted a study evaluating the ability of ChatGPT-4 to respond to frequently asked questions having to do with blepharoplasty.[15] Cox et al presented ChatGPT-4 with 3 frequently asked questions for blepharoplasty in the preoperative and postoperative setting. Responses were evaluated by the study contributors. Similar to the previous studies mentioned, Cox et al contend that ChatGPT-4 can be a beneficial tool for providing mostly accurate and vague information for blepharoplasty. Authors emphasized its limitations as the program is only relevant to data up till the pretrained date and has the potential to provide invalid answers.[15] Cox et al also suggest that ChatGPT-4 should be updated and reviewed regularly to increase its reliability and permit regular use in the clinical setting.

Discussion

Plastic surgery is a dynamic and continuously changing field. Staying up to date on new advancements and developing technology is imperative. Recent developments in machine learning technology have given AI the potential to aid surgeons in both the research process and the clinical setting.

On a basic level, ChatGPT can evaluate current plastic surgery literature. It can then take that information to generate new ideas for systematic reviews, with increased reliability when queries are more specific. This can give a picture of the publication landscape, allowing surgeons to easily identify areas of greater attention and gaps in literature. As a result, the

overall diversity and efficiency of newer publications can increase if implemented.

When prompted, ChatGPT also gave an accurate yet broad summary and comparison for different medical journals. Because ChatGPT does not have direct access to the internet, it has to rely on pretrained data. As such, the data can be confounded during recall and output, causing hallucination. Surgeons must be cautious, exercise due diligence, and substantiate all data and sources provided by ChatGPT when using it as a research assistant.

In answering frequently asked questions, ChatGPT showed an impressive ability to provide accurate medical data. Utilizing ChatGPT as a potential patient consultation tool to answer patients' preoperative and postoperative questions would not only lessen the burden for the surgeon but would also improve patient knowledge and overall satisfaction for procedures. However, the evidence provided by the current literature shows that ChatGPT still needs improvement to be implemented in patient consultation. Despite impressive accuracy rates in answering frequently asked patient questions, there have been accounts of ChatGPT providing wrong and misleading information.[10] If given to patients, this can lead to a spread of skewed and misinformation. This is due, in part, to its lack of recency and the lack of regular fact checking and improvement of responses. While ChatGPT shows promise in this capacity, the literature indicates that it is not performing at a level that would be acceptable in the clinical setting. Further advancements are required to implement ChatGPT as a trusted outlet for patient consultation. On the other hand, ChatGPT showed an almost perfect ability to provide medical statistics and data. In a setting where surgeons are utilizing ChatGPT as a consultant for medical data and statistics and due diligence is being exercised, it can be a useful and time-effective tool for surgeons getting information.

Because of the overall novelty of ChatGPT, it has gained a lot of attention, especially in medicine. Surgeons and physicians can find many ways to implement this technology into their practice, mostly in the research process. ChatGPT is limited in its uses in clinical settings because of its lack of overall reliability to provide true and current information. Surgeons should exercise due diligence and substantiate the outputs from ChatGPT. Developments in AI technology, regular data set updating, and regular fact checking can increase ChatGPT's reliability and lead to uses in the clinical setting.

To date multiple other chatbots have been developed. While this review only focused on ChatGPT, ChatGPT serves as a representative benchmark for the potential of current chatbot usage in plastic surgery. As developments and advancements are made to chatbot technology, further review will be conducted.

Conclusions

While ChatGPT can be a useful tool in the conception of literature searches and the execution of information retrieval and synthesis for research purposes, it is currently not reliable enough to be implemented in the clinical setting. As ChatGPT and other chatbots further develop, further research on its uses as a patient consultation tool should be conducted and reevaluated.

References

1. https://openai.com/blog/chatgpt

2. Adamopoulou E, Moussiades L. An overview of chatbot technology. *Artif Intell Applications Innov*. 2020;584:373-383.

3. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312.

4. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023 Mar 19;11(6):887.

5. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus*. 2023 Feb 19;15(2):e35179.

6. Gupta R, Park JB, Bisht C, et al. Expanding cosmetic plastic surgery research with ChatGPT. *Aesthet Surg J*. 2023;43(8):930-937. doi:10.1093/asj/sjad069

7. Gupta R, Pande P, Herzog I, et al. Application of ChatGPT in cosmetic plastic surgery: ally or antagonist? *Aesthet Surg J*. 2023;43(7):NP587-NP590. doi:10.1093/asj/sjad042

8. Gupta R, Herzog I, Weisberger J, Chao J, Chaiyasate K, Lee ES. Utilization of ChatGPT for plastic surgery research: friend or foe? *J Plast Reconstr Aesthet Surg*. 2023;80:145-147. doi:10.1016/j.bjps.2023.03.004

9. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? [published correction appears in Crit Care. 2023 Mar 8;27(1):99]. *Crit Care*. 2023;27(1):75.

10. Wu RT, Dang RR. ChatGPT in head and neck scientific writing: A precautionary anecdote. *Am J Otolaryngol*. 2023;44(6):103980. doi:10.1016/j.amjoto.2023.103980

11. 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(9):NP720-NP723. doi:10.1093/asj/sjad158

12. Chervenak J, Lieman H, Blanco-Breindel M, Jindal S. The promise and peril of using a large language model to obtain clinical information: ChatGPT performs strongly as a fertility counseling tool with limitations [published online ahead of print, 2023 May 20]. *Fertil Steril*. 2023;S0015-0282(23)00522-8. doi:10.1016/j.fertnstert.2023.05.151

13. 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 [published online ahead of print, 2023 May 9]. *Aesthet Surg J*. 2023;sjad140. doi:10.1093/asj/sjad140

14. Seth I, Xie Y, Cox A, et al. Aesthetic surgery advice and counseling from artificial intelligence: a rhinoplasty consultation with ChatGPT [published online ahead of print, 2023 Apr 24]. *Aesthetic Plast Surg*. 2023 Sep 14;43(10): 1126-1135.

15. Cox A, Seth I, Xie Y, Hunter-Smith DJ, Rozen WM. Utilizing ChatGPT-4 for providing medical information on blepharoplasties to patients. *Aesthet Surg J*. 2023;43(8):NP658-NP662. doi:10.1093/asj/sjad096