




Can AI Answer My Questions? Utilizing Artificial Intelligence in the Perioperative Assessment for Abdominoplasty Patients

Bryan Lim¹ · Ishith Seth¹ · Roberto Cuomo²  · Peter Sinkjær Kenney^{3,4} · Richard J Ross¹ · Foti Sofiadellis¹ · Paola Pentangelo⁵ · Alessandra Ceccaroni⁵ · Carmine Alfano⁵ · Warren Matthew Rozen¹



Received: 23 March 2024 / Accepted: 21 May 2024
© The Author(s) 2024

Abstract

Background Abdominoplasty is a common operation, used for a range of cosmetic and functional issues, often in the context of divarication of recti, significant weight loss, and after pregnancy. Despite this, patient–surgeon communication gaps can hinder informed decision-making. The integration of large language models (LLMs) in healthcare offers potential for enhancing patient information. This study evaluated the feasibility of using LLMs for answering perioperative queries.

Methods This study assessed the efficacy of four leading LLMs—OpenAI’s ChatGPT-3.5, Anthropic’s Claude, Google’s Gemini, and Bing’s CoPilot—using fifteen unique prompts. All outputs were evaluated using the Flesch–Kincaid, Flesch Reading Ease score, and Coleman–Liau index for readability assessment. The DISCERN score and a Likert scale were utilized to evaluate quality. Scores were assigned by two plastic surgical residents and then reviewed and discussed until a consensus was reached by five plastic surgeon specialists.

Results ChatGPT-3.5 required the highest level for comprehension, followed by Gemini, Claude, then CoPilot.

Claude provided the most appropriate and actionable advice. In terms of patient-friendliness, CoPilot outperformed the rest, enhancing engagement and information comprehensiveness. ChatGPT-3.5 and Gemini offered adequate, though unremarkable, advice, employing more professional language. CoPilot uniquely included visual aids and was the only model to use hyperlinks, although they were not very helpful and acceptable, and it faced limitations in responding to certain queries.

Conclusion ChatGPT-3.5, Gemini, Claude, and Bing’s CoPilot showcased differences in readability and reliability. LLMs offer unique advantages for patient care but require careful selection. Future research should integrate LLM strengths and address weaknesses for optimal patient education.

Level of Evidence V This journal requires that authors assign a level of evidence to each article. For a full description of these Evidence-Based Medicine ratings, please refer to the Table of Contents or the online Instructions to Authors www.springer.com/00266.

Keywords AI · Abdominoplasty · Perioperative · LLM · ChatGPT

✉ Roberto Cuomo
roberto.cuomo@unisi.it

¹ Department of Plastic Surgery, Peninsula Health, Melbourne, Victoria 3199, Australia

² Plastic Surgery Unit, Department of Medicine, Surgery and Neuroscience, University of Siena, Siena, Italy

³ Department of Plastic Surgery, Vejle Hospital, Beriderbakken 4, 7100 Vejle, Denmark

⁴ Department of Plastic and Breast Surgery, Aarhus University Hospital, Aarhus, Denmark

⁵ University of Salerno, Fisciano, Italy

Introduction

Abdominoplasties are recognized as one of the most widely undertaken aesthetic surgical interventions globally [1]. Abdominoplasties offer more than aesthetic enhancements, addressing physiological issues such as excising redundant skin after significant weight loss or childbirth, resecting recalcitrant adipose tissue resistant to diet, and improving functional symptoms of back pain and urinary incontinence [2, 3]. It has also been noted to improve self-esteem and

overall patient well-being, rendering it a popular choice for those seeking aesthetic and functional improvements [4–7]. However, it is associated with a wide spectrum of complications, including seromas, infections, deep vein thrombosis, and hypertrophic scarring, among others [8].

Given the complexity of the surgery, patients frequently have numerous queries regarding preoperative preparation, postoperative recovery, potential adverse events, and ongoing care of the surgical site. Sometimes, communication between the surgeon and patients may be lacking, resulting in inadequately informed patients who may be incapable of making enlightened decisions regarding their surgical care [9, 10]. Such consultations can also be time-consuming, potentially diverting surgeons and nurses from other responsibilities.

Since the introduction of Open AI's ChatGPT-3.5 large language models (LLMs) in November 2022, integration of such models into the healthcare sector has experienced an exponential increase. A large diversity of applications includes triage, diagnosis, treatment, research, education, and preoperative planning [11–17]. Given the infancy of such technology, the possibilities for application in plastic surgery are currently boundless [18–20]. Few studies have explored the application of LLMs in perioperative care, let alone abdominoplasty [21, 22].

The Natural Language Processing (NLP) technology exhibited by LLMs enables them to generate sequences of text in response to user inputs, predicting what the next most probable word will be based on extensive training data [23, 24]. These responses therefore read very human-like and rational. Consequently, we aim to evaluate the viability of employing LLMs as clinical adjuncts for patients, enabling them to have their perioperative questions answered by an AI and optimizing healthcare workflow. The authors assessed this by measuring readability, reliability, and accuracy of the responses via a series of metrics. Limitations, drawbacks, and potential issues of utilizing such AI models will also be identified and discussed.

Methods

The efficacy of four leading LLMs—OpenAI's ChatGPT-3.5, Anthropic's Claude, Google's Gemini, and Bing's CoPilot—was systematically evaluated by presenting each with fifteen unique prompts. The authors collectively discussed and agreed that these inputs represented the most common inquiries posed by patients. To reduce potential bias from previous online activities, the research utilized the browser's incognito mode. Uniformity in the testing process was maintained by ensuring that all prompts were presented to the LLMs on the same day, by one author,

using the same device and account (Table 1). The LLMs were not directed to provide hyperlinks or references, reflecting the layperson's potential unawareness or omission to request such detailed information from AI models.

The evaluation of their outputs was based on a comprehensive set of metrics. For readability assessment, we employed the Flesch–Kincaid, Flesch Reading Ease scores, and Coleman–Liau index (Table 2). The Flesch Reading Ease score spans from 1 to 100, with higher scores indicating greater readability. The Flesch–Kincaid Grade Level was used to ascertain the educational level required for comprehension, with a score of 8 suggesting that the content is suitable for individuals with an eighth-grade education level in the USA. The Coleman–Liau index ranges from 0 to infinity, where each score aligns with the US school grade level necessary for comprehension. For instance, a score of 9 indicates the text is suitable for a 9th-grade reading level. Scores between 13 and 16 denote college-level comprehension, while scores above 16 are considered professional level. The quality of the LLM responses was analyzed using the DISCERN score (Table 2) and a Likert scale (Table 3), with the latter scoring between 1 and 5 on the following aspects: Clarity, Comprehension, Readability, Patient Friendliness, and Informativeness. This analysis was carried out by two plastic surgical residents and validated by five plastic surgeons specializing in abdominoplasty, ensuring expert oversight. Any discrepancies in scores were discussed until a consensus was reached.

Results

According to Table 2, the Flesch–Kincaid Grade Level analysis revealed that ChatGPT-3.5 required the highest level of US education to read its outputs at 13.49 ± 1.35 . Gemini is the next hardest to read, scoring 12.22 ± 2.05 , followed by Claude at 11.15 ± 1.51 , and finally Bing's CoPilot with 10.59 ± 1.22 . The Flesch Reading Ease score corroborates the difficulty of ChatGPT-3.5's readability, as it scored the lowest with 35.22 ± 7.44 . Gemini's score of 35.86 ± 9.81 corresponds to its Grade Level of being the second hardest to read of the four LLMs. Interestingly, Claude (45.30 ± 7.79) outperformed CoPilot (39.88 ± 9.29), contrasting the results of both LLMs in the Flesch–Kincaid Grade Level assessment. Assessment of the Coleman–Liau scores revealed slightly different findings, with CoPilot requiring the highest level of education (14.37 ± 1.67), followed by Gemini (14.19 ± 1.42), then ChatGPT-3.5 (14.06 ± 1.06), and finally Claude (12.39 ± 1.07) being the easiest to comprehend.

The DISCERN scores revealed that Claude gave the most reliable information with a score of 54.60 ± 2.23 .

Table 1 Inputs into the four LLMs

Prompt number	Input
1	I had a tummy tuck yesterday, now there is bruising around the area. What should I do?
2	I had a tummy tuck yesterday, now there is bleeding around the area. What should I do?
3	I had a tummy tuck yesterday, now I have swelling in the area. What should I do?
4	I had a tummy tuck yesterday and it hurts really bad in the area. What should I do?
5	I had a tummy tuck yesterday, when can I get back to swimming?
6	I had a tummy tuck yesterday, when can I get back to exercising?
7	I had a tummy tuck yesterday, when can I drive again?
8	I had a tummy tuck yesterday, do I need antibiotics?
9	I had a tummy tuck yesterday, how can I sleep?
10	I had a tummy tuck yesterday, when and how can I shower?
11	I had a tummy tuck yesterday, how long is the recovery?
12	I had a tummy tuck yesterday and I work a desk job. When can I return to work?
13	I had a tummy tuck yesterday and I have to lift heavy objects for work. When can I return to work?
14	I had a tummy tuck yesterday. Can I continue to have more children in the coming years?
15	What can I do to prepare for my tummy tuck surgery?

ChatGPT-3.5 trailed behind closely with 53.00 ± 2.04 , followed by CoPilot at 49.47 ± 10.68 and lastly, Gemini with 49.13 ± 1.77 .

In the evaluation using the Likert scale (Table 3), Bing's CoPilot emerged as the top performer with an aggregate score of 20, achieving the maximum possible score of 5 in the category of patient-friendliness. Claude was a close second with an overall score of 18, indicating strong performance, particularly in areas of clarity and approachability. ChatGPT-3.5 followed with a total of 17, trailed by Gemini scoring 16, with the latter ranking lowest in comprehensiveness among the four evaluated LLMs.

ChatGPT-3.5

ChatGPT delivered adequate responses that addressed all inquiries effectively, employing bullet points to enhance clarity and facilitate easier comprehension. However, its replies were generally broad, lacked detail, and did not provide hyperlinks or guidelines to support its statements, rendering its advice of average quality and less dependable compared to CoPilot. As the language utilized was professional, it did not detract from user friendliness, but it also did not establish rapport with users compared to CoPilot. In summary, when asked for perioperative advice on abdominoplasty, ChatGPT-3.5's performance was deemed average and unexceptional (Figs. 1, 2, 3, 4).

Claude

Claude provided satisfactory responses, frequently utilizing a listing format to articulate its points across. It uniformly

advised users to consult professional medical practitioners and advocated for transparency with healthcare providers to accurately evaluate their conditions. Its guidance proved to be significantly more practical than the other LLMs. For example, in response to prompts 2 and 3, it adopted a more authoritative tone, offering clear instructions rather than just suggestions. Furthermore, it identified specific conditions, such as seroma formation, when addressing swelling in prompt 3. Moreover, Claude tailored its responses to precise timeframes with considerable detail for prompts 5, 6, 7, 11, 12, and 13. The overall language employed by Claude was professional yet avoided using excessive medical jargon. In summary, while Claude's guidance was competent, it was characterized as unexceptional.

Bing CoPilot

CoPilot distinguished itself by using emojis, colloquial language, and even extending congratulations to users on their recent abdominoplasty procedures, thereby fostering a more personalized and engaging user experience. To improve the clarity of its communications, it frequently used bullet points instead of long paragraphs. Additionally, CoPilot offered unique advice not mentioned by other LLMs, for instance, recommending specific fruits, vegetables, and whole grains to aid in preparing for abdominoplasty. CoPilot was also distinguished as the only LLM to incorporate hyperlinks for citation purposes in its responses, additionally providing illustrative images for certain queries, thereby enhancing the comprehensiveness and visual appeal of its answers. Unfortunately, the visual aids and several links were not helpful. For instance, its

Table 2 Readability and Reliability scores of the LLMs

Prompt n.o	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean
<i>Flesch–Kincaid level</i>																
ChatGPT-3.5	12.74	12.60	12.05	11.85	14.19	13.62	13.69	16.68	12.10	12.63	13.07	12.71	14.95	14.60	14.89	13.49 (± 1.35)
Claude	9.43	11.63	10.29	11.15	11.88	10.90	12.03	13.66	8.68	11.24	8.55	10.40	12.92	13.08	11.44	11.15 (± 1.51)
Bing CoPilot	9.57	11.17	–	11.53	10.52	12.05	9.54	12.45	8.70	9.36	10.88	11.01	9.09	10.21	12.24	10.59 (± 1.22)
Gemini	11.75	11.77	9.62	9.46	11.11	12.56	10.08	16.86	13.49	9.85	14.61	12.31	13.82	13.09	12.89	12.22 (± 2.05)
<i>Flesch Reading Ease scores</i>																
ChatGPT-3.5	39.54	40.59	42.18	42.04	30.50	32.32	34.93	18.87	44.10	40.07	38.21	41.24	27.84	31.50	24.33	35.22 (± 7.44)
Claude	48.68	47.79	45.64	44.19	45.23	46.09	42.67	29.79	57.79	41.99	58.68	54.07	41.38	33.79	41.78	45.30 (± 7.79)
Bing CoPilot	45.53	30.63	–	36.89	38.15	27.76	43.69	25.76	51.54	51.03	39.84	42.54	49.22	49.50	26.27	39.88 (± 9.29)
Gemini	40.94	35.96	46.37	45.50	43.33	36.89	47.56	17.38	31.23	47.84	21.33	38.66	26.87	32.52	25.57	35.86 (± 9.81)
<i>Coleman–Liau index</i>																
ChatGPT-3.5	13.24	13.55	13.15	13.56	14.67	14.46	13.79	16.11	13.03	13.71	14.01	12.44	14.34	14.63	16.26	14.06 (± 1.06)
Claude	13.39	11.36	12.84	12.58	12.40	12.56	12.09	13.59	10.85	12.93	10.66	10.71	12.39	13.94	13.50	12.39 (± 1.07)
Bing CoPilot	13.65	14.40	–	15.27	14.88	16.10	14.53	17.04	12.53	13.69	13.72	13.59	12.17	12.12	17.46	14.37 (± 1.67)
Gemini	13.44	14.09	13.15	12.64	13.19	13.90	12.71	14.26	15.30	12.18	15.85	14.13	15.15	15.55	17.38	14.19 (± 1.42)
<i>DISCERN scores</i>																
Prompt n.o	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean
ChatGPT-3.5	51	52	50	53	52	52	52	53	54	51	57	56	56	54	52	53.00 (± 2.04)
Claude	50	54	52	53	56	56	54	55	56	51	55	57	57	57	55	54.60 (± 2.23)
Bing CoPilot	56	16	–	57	56	57	55	56	55	54	55	56	56	57	56	49.47 (± 10.68)
Gemini	50	48	49	47	48	48	49	46	51	48	51	50	50	49	53	49.13 (± 1.77)

Flesch–Kincaid Grade Level: This metric estimates the US school grade level needed to understand a text, based on word and sentence length

Flesch Reading Ease score: It assesses text readability by assigning a score from 1 to 100, with higher scores indicating easier reading material, also based on sentence and word length

Coleman–Liau Index: This index predicts the grade level required to understand a text, focusing on characters per word and sentences per 100 words, rather than syllable count

DISCERN scoring system: It evaluates the quality of written health information, with scores reflecting the reliability and clarity of the information provided to patients

Table 3 Likert scale of each LLM for all prompts

Criteria	ChatGPT-3.5	Claude	Bing CoPilot	Google Gemini
Clarity	4	4	4	4
Comprehensiveness	3	4	4	2
Readability	4	4	3	4
Patient Friendliness	3	3	5	3
Informativeness	3	3	4	3
Total	17	18	20	16

Likert scale: This measures attitudes or opinions by offering a range of fixed responses, allowing individuals to express their level of agreement or disagreement with a statement

response to prompt 1 included an image of a surgeon marking a patient's abdomen with a circle, bearing no relevance to postoperative bruising. Additionally, certain links raised concerns regarding their credibility, including "microsoftstart.msn.com" in prompts 6 and 11, and "re-alsself.com." Most of the links directed users to non-scholarly websites, undermining the credibility of the provided information. Additionally, its response to the second query was less informative, merely suggesting consultation with a healthcare professional. Moreover, its response to the third prompt was unexpectedly delivered in Korean, despite the initial inquiry being posed in English.

Gemini

Gemini consistently underscored its non-professional medical status, stressing the paramount importance of consulting a surgeon for more accurate guidance. Furthermore, it frequently utilized a list format to enhance the clarity and readability of the information, as opposed to dense paragraphs. Ultimately, Gemini offered conservative and generalized management strategies, deferring the delineation of more specific treatment approaches to the discretion of the user's healthcare provider.

Discussion

Abdominoplasties are complex procedures that come with a variety of potential complications, which can significantly impact a patient's quality of life [2, 8, 25–29]. As such, patients often have numerous questions and concerns regarding both preoperative preparations and postoperative care. This underscores the importance of providing thorough and accessible information to address these concerns and ensure patients are well informed about their surgical journey [30, 31].

LLMs are trained on large amounts of data and fine-tuned to generate human-like text [32–35]. Consequently, they have the potential to significantly enhance the process of providing information to patients, serving as a valuable

adjunct to traditional patient education methods [16, 36–41]. By delivering instant, accessible, and personalized responses to patient inquiries, LLMs can significantly enhance the patient care experience. These advanced tools are capable of tailoring information to the specific aftercare or preoperative education preferences of the surgeon, ensuring that patients receive responses that are directly relevant to their individual treatment plans. Moreover, LLMs can identify perioperative red flags, prompting the system to advise patients to immediately contact the appropriate clinician. In cases where patient questions indicate underlying concerns, LLMs could potentially facilitate immediate notification to the responsible surgeon, ensuring that critical issues are addressed promptly. Furthermore, the data captured from these interactions are invaluable, as they can be analyzed and shared with the treating team to improve the understanding of patients' perioperative needs. This approach not only reduces anxiety by clarifying doubts in real-time using empathetic vernacular but also improves overall patient satisfaction by creating a more responsive, informed, and patient-centered care environment. By summarizing patient concerns for surgeons, LLMs can streamline patient–surgeon communication by focusing on specific issues.

In aesthetic medicine, the need for lengthy explanations to repetitive patient inquiries highlights the inefficiencies of traditional interactions with surgeons and staff, emphasizing the importance of time optimization to prevent detracting from surgeons' core duties of surgery and consultation [42, 43]. In response, some practices have adopted strategies such as employing nurses for pre-consultation screening, reassurance, and procedural guidance. Incorporating LLMs into the preoperative phase presents a strategic innovation with the potential to significantly enhance patient education and screening processes. This technology can efficiently manage patient interactions, conserving valuable time for practitioners while maintaining high standards in patient care and support. This approach not only streamlines the workflow for surgeons, allowing them to focus on their primary duties without compromise, but also fosters a more informed and engaged patient

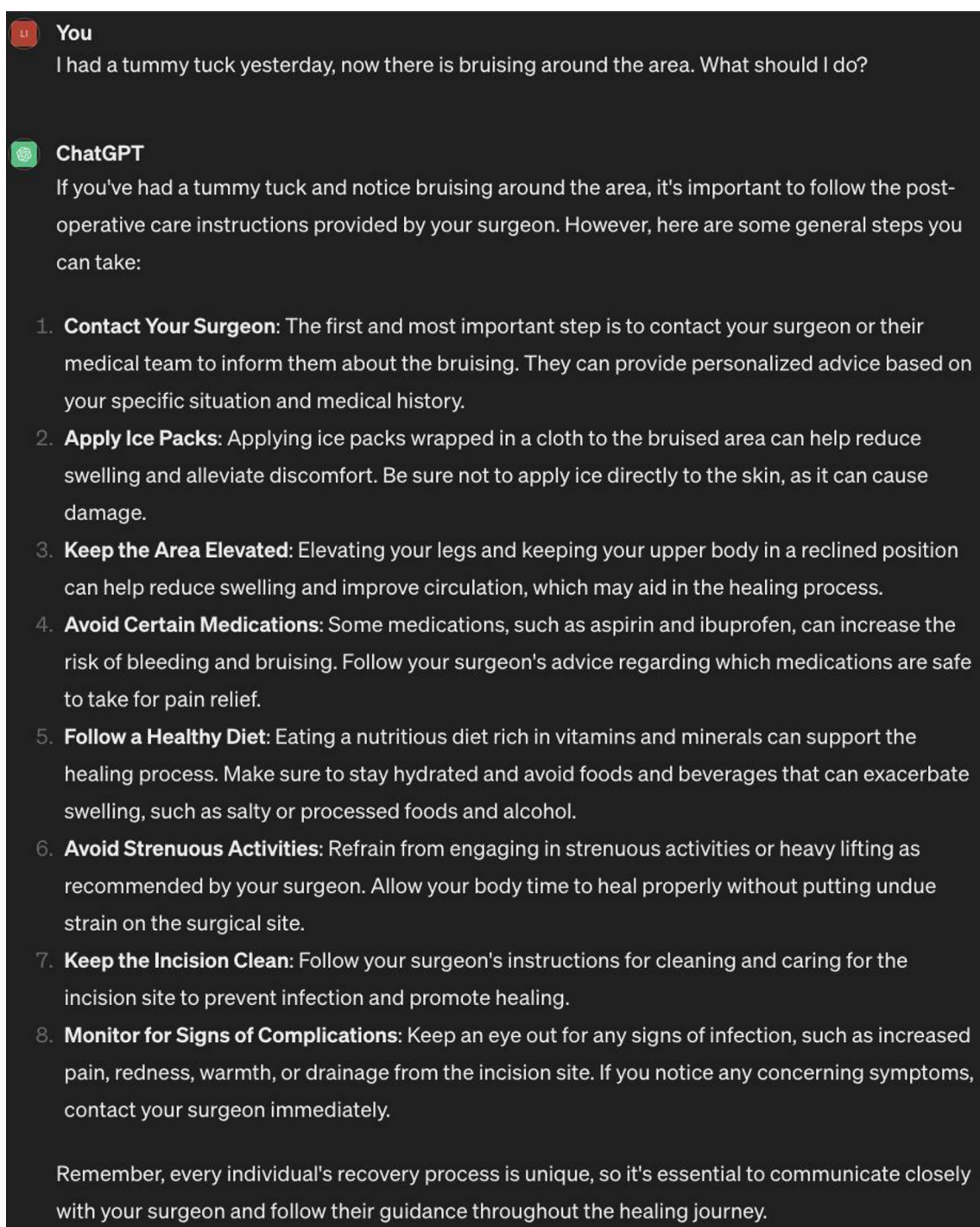


Fig. 1 ChatGPT-3.5's responses to all 15 prompts

population, ultimately contributing to improved perioperative outcomes and patient satisfaction.

This study compared the readability and reliability of responses from four leading LLMs—ChatGPT-3.5, Gemini, Claude, and Bing's CoPilot—within the context of perioperative advice for abdominoplasty. Our findings highlight significant differences in the accessibility and

quality of information provided by each LLM, with implications for their potential use in patient education and engagement.

The Flesch–Kincaid Grade Level and Flesch Reading Ease scores suggested that ChatGPT-3.5's outputs require the highest level of US education for comprehension, potentially limiting its accessibility to a broader patient

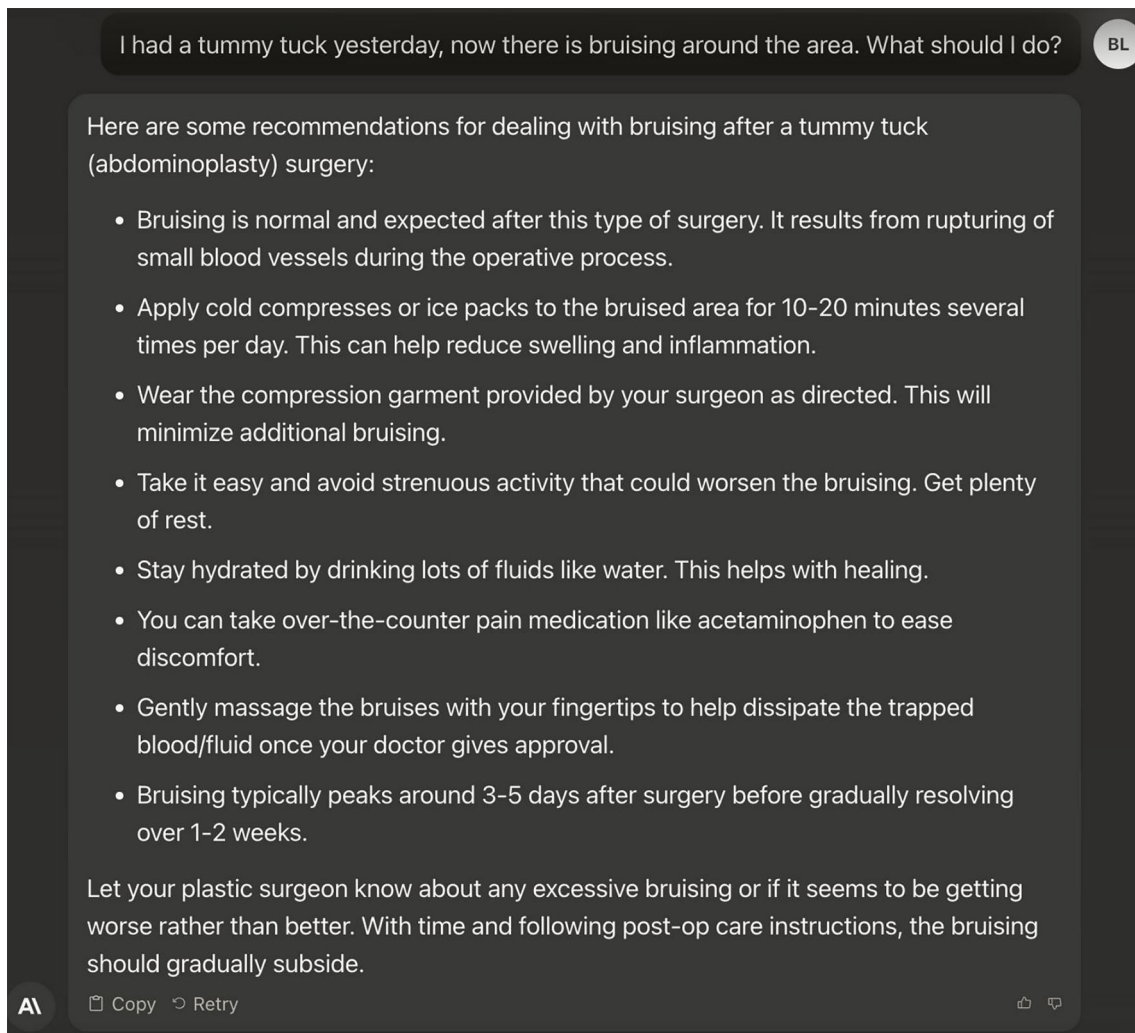


Fig. 2 Claude's responses to all 15 prompts

demographic. Conversely, Bing's CoPilot demonstrated the lowest required reading level, coupled with the highest patient-friendliness score, indicating its potential as a more universally accessible resource for patient education. The Coleman–Liau index presented a more nuanced view of readability, with CoPilot requiring the highest education level for comprehension, which contrasted with its performance in the Flesch–Kincaid analysis. This discrepancy underscores the complexity of assessing readability and the necessity of considering multiple metrics to gauge the accessibility of health-related information more accurately. More importantly, this situation highlights issues surrounding the accessibility of information. It has been observed that the health literacy levels among plastic surgery patients are typically insufficient [44–47]. Recommendations by the American Medical Association and the National Institutes of Health suggest that materials related to plastic surgery should be written at a sixth- to eighth-grade reading level [47]. However, recent research

indicates that the readability of LLMs surpasses these recommendations, requiring a higher level of patient understanding [48, 49]. This discrepancy could potentially undermine the relationship between patients and healthcare providers, representing a significant barrier to the effective implementation of AI-driven chatbot perioperative tools in plastic surgery contexts.


Claude's superior DISCERN score indicates its reliability in providing information, suggesting that its outputs may be more suitable for patients seeking dependable advice. The consistency in advising professional consultation across all LLMs reinforces the importance of physician oversight in patient care. The responses provided by ChatGPT-3.5 exhibited a relatively comprehensive nature, encapsulating a substantial number of salient points presented in a manner conducive to facile comprehension. In contrast to ChatGPT-3.5's outputs for prompts 11 through 13, however, Claude was observed to furnish more expansive and specific delineations of the postoperative

You

I had a tummy tuck yesterday, now there is bruising around the area. What should I do?

Copilot

Bruising and swelling are common side effects after a tummy tuck (abdominoplasty) surgery. These are natural parts of the healing process, but I understand they can be concerning. Here are some steps you can take to minimize bruising and swelling and promote a smoother recovery:



- Rest and Recovery:**
 - **Allow yourself time to heal.** Avoid strenuous activities, including cardiovascular exercise, weight training, and lifting heavy objects around the house.
 - **Listen to your body** and give it the rest it needs.
- Healthy Diet:**
 - **Follow a nutritious diet.** Proper nutrition plays a crucial role in healing.
 - **Stay hydrated** to support your body's recovery process.
- Cold Compress:**
 - **Apply cold compresses** to the bruised areas. This can help reduce swelling and alleviate discomfort.
 - **Use ice packs wrapped in a cloth** to avoid direct skin contact.
- Elevate Your Legs:**
 - **Elevate your legs** when resting or lying down. This helps improve blood circulation and reduces swelling.
- Compression Garments:**
 - **Wear compression garments** as recommended by your surgeon. These garments provide gentle pressure and support to minimize swelling and bruising.
- Avoid Alcohol and Smoking:**
 - **Refrain from alcohol and smoking** during your recovery. Both can interfere with the healing process.
- Follow Your Surgeon's Instructions:**
 - **Adhere to post-operative instructions** provided by your surgeon. They know your specific case and can guide you effectively.

Remember that **bruising and swelling are temporary**, and they will gradually subside as your body heals. If you experience any unusual symptoms or have concerns, **contact your surgeon promptly**. They can provide personalized advice based on your individual situation.

Wishing you a smooth recovery! 🌟

Learn more [1 drjeremyhunt.com.au](#) [2 raadinahealth.com](#) [3 drcraigrubinstein.com.au](#) [4 gettyimages.com](#)

👍 🗨️ 📄 ⬇️ 📧 🗣️ 1 of 5

Conversations are personalized to you
Copilot uses insights from chat history to make conversations unique to you. [Learn more in our FAQs.](#) [Turn personalization off anytime in Bing Settings.](#)

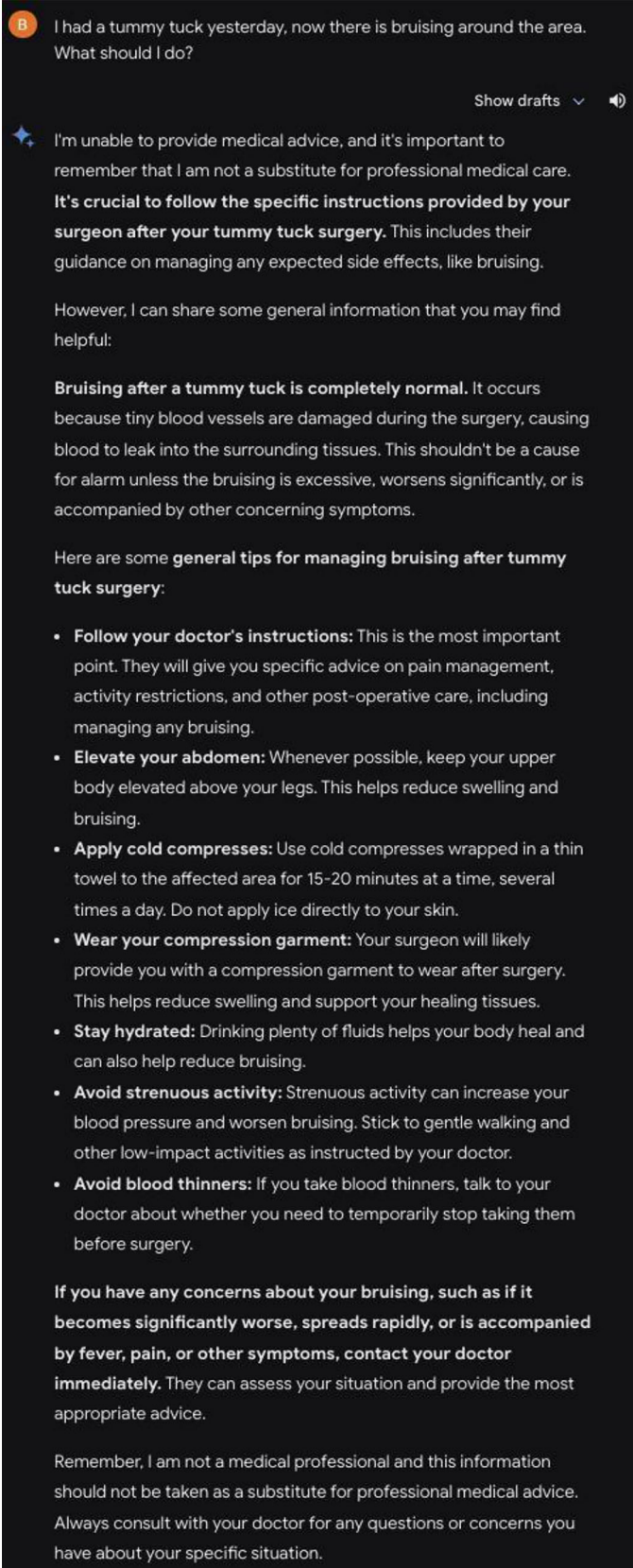
🔍 How long does it take to recover from a tummy tuck? What are the risks of a tummy tuck surgery? Can I shower after my surgery?

Fig. 3 CoPilot's responses to all 15 prompts

timeframe, thereby affording patients a more exhaustive preview of the trajectory of events post-op. Notably, Claude's avoidance of excessive medical jargon and clarity could make it particularly useful in facilitating patient understanding and engagement. CoPilot's use of emojis, colloquial language, and hyperlinks, unique among the LLMs evaluated, improved its approach towards patient engagement, potentially enhancing the user experience and comprehension through a more conversational and

interactive format. However, its occasional lapses in providing expansive, specific medical advice such as in prompts 4, 8, and 12, failure to provide any advice in prompt 2, and the unexpected language switch in prompt 3 negatively affected its DISCERN score. This also accentuates the challenges of using LLMs in real-world settings, where accuracy and consistency are paramount as users may not be so patient as to re-prompt the AI model. Gemini's emphasis on its non-professional status and

Fig. 4 Gemini's responses to all 15 prompts



B I had a tummy tuck yesterday, now there is bruising around the area. What should I do?

Show drafts ▾ 🔊

✦ I'm unable to provide medical advice, and it's important to remember that I am not a substitute for professional medical care. **It's crucial to follow the specific instructions provided by your surgeon after your tummy tuck surgery.** This includes their guidance on managing any expected side effects, like bruising.

However, I can share some general information that you may find helpful:

Bruising after a tummy tuck is completely normal. It occurs because tiny blood vessels are damaged during the surgery, causing blood to leak into the surrounding tissues. This shouldn't be a cause for alarm unless the bruising is excessive, worsens significantly, or is accompanied by other concerning symptoms.

Here are some **general tips for managing bruising after tummy tuck surgery:**

- **Follow your doctor's instructions:** This is the most important point. They will give you specific advice on pain management, activity restrictions, and other post-operative care, including managing any bruising.
- **Elevate your abdomen:** Whenever possible, keep your upper body elevated above your legs. This helps reduce swelling and bruising.
- **Apply cold compresses:** Use cold compresses wrapped in a thin towel to the affected area for 15-20 minutes at a time, several times a day. Do not apply ice directly to your skin.
- **Wear your compression garment:** Your surgeon will likely provide you with a compression garment to wear after surgery. This helps reduce swelling and support your healing tissues.
- **Stay hydrated:** Drinking plenty of fluids helps your body heal and can also help reduce bruising.
- **Avoid strenuous activity:** Strenuous activity can increase your blood pressure and worsen bruising. Stick to gentle walking and other low-impact activities as instructed by your doctor.
- **Avoid blood thinners:** If you take blood thinners, talk to your doctor about whether you need to temporarily stop taking them before surgery.

If you have any concerns about your bruising, such as if it becomes significantly worse, spreads rapidly, or is accompanied by fever, pain, or other symptoms, contact your doctor immediately. They can assess your situation and provide the most appropriate advice.

Remember, I am not a medical professional and this information should not be taken as a substitute for professional medical advice. Always consult with your doctor for any questions or concerns you have about your specific situation.

conservative management strategies, while prudent, may limit the depth of information provided to patients, resulting in the trade-off between caution in advice and the richness of patient education content.

Despite the exciting prospects that LLMs provide, surgeons may be hesitant to integrate AI-driven perioperative tools into their practices due to the potential legal liability from errors in judgment or delays in care caused by such AI technologies. Such concerns are not unique to AI NLP interfaces, however. Similar concerns arise when training new or inexperienced clinical staff who may be triaging patient phone calls or messages [50, 51]. The risks of mistakes or delays exist whether it is an AI system or new human staff handling triage duties. Consequently, the legal liability worries are analogous regardless of whether AI NLP systems or newly hired personnel are being onboarded for perioperative assessment. At the moment, the surgeon should bear responsibility for all care and advice provided, regardless of whether it is given by a human or AI [52, 53].

It is imperative to acknowledge some limitations in our study, particularly the absence of actual patient involvement in evaluating the LLMs' outputs. Such engagements could provide invaluable insights into their practical utility, comprehensibility, and relevance of the responses generated by these LLMs, reflecting real-world needs and comprehension levels.

Additionally, the methodological design of our investigation is circumscribed by the participation of two plastic surgeons when curating the scores for the LLMs' outputs. While their specialized acumen is critical for the appraisal of the AI models, this relatively small panel may impinge upon the external validity and generalizability of our conclusions. Expanding the group of evaluators to include more plastic surgery experts, and possibly other healthcare professionals, would provide a more comprehensive and balanced view of how effective these LLMs are in medical communication.

Furthermore, the focus of this study on perioperative inquiries to abdominoplasty delineated a specific segment of the expansive domain of plastic surgery. While this focus affords an in-depth examination within the context of abdominoplasties, it concurrently narrows the scope of inquiry, thereby restricting the comprehensive understanding of LLM applicability across multiple plastic surgical procedures. Consequently, there remains room for future research endeavors to investigate similar metrics in other plastic surgical procedures. This would not only augment the corpus of knowledge regarding LLM applications in healthcare but also guide the evolution of algorithmically sophisticated, procedure-specific models tailored to the intricate informational necessities of patients within the plastic surgery paradigm.

While no single LLM emerged as universally superior, each offered distinct advantages and challenges in readability, reliability, and patient engagement. These findings suggest that careful selection and possibly integration of multiple LLMs could optimize patient education and support. Future research should explore the direct impact of LLM-generated advice on patient outcomes and satisfaction, as well as investigate strategies for enhancing the accuracy, comprehensiveness, and personalization of LLM responses in clinical contexts.

The ethical integration of AI in surgical procedures raises significant concerns regarding privacy, consent, and human oversight [54, 55]. Ensuring AI systems like large language models adhere to healthcare privacy regulations, such as HIPAA, is crucial for protecting sensitive patient data. Patients must be fully informed about the role of AI, including how their data is used and potential risks, ensuring their consent is explicit and well documented. Moreover, maintaining human oversight is vital to ensure AI supplements rather than replaces professional medical judgment. This approach helps mitigate risks from AI inaccuracies and maintains the essential human element in healthcare, ensuring AI's role remains as an ethical adjunct to enhance patient care outcomes.

The study's limitations highlight the need for expanded research to validate AI applications in medical settings effectively. Notably, the absence of actual patient involvement restricts insights into the real-world utility and comprehensibility of AI communications. Future studies should include patient feedback to assess AI's effectiveness in clinical scenarios more accurately. Additionally, the evaluation based on only two plastic surgeons' perspectives may limit the findings' generalizability. Including a broader range of healthcare professionals as evaluators could provide a more comprehensive assessment. Furthermore, the study's focus on perioperative inquiries for abdominoplasty represents a narrow segment of plastic surgery. Broader research across various surgical procedures would help develop more precise AI models tailored to the diverse informational needs of different patient demographics, enhancing AI's clinical relevance and efficacy.

Conclusion

This study highlighted significant differences between ChatGPT-3.5, Gemini, Claude, and Bing's CoPilot, with each model showing unique strengths and weaknesses in terms of readability scores, DISCERN reliability, and patient-friendliness. The findings showcase the potential of LLMs to enhance patient education and support, but also emphasize the need for careful consideration of each

model's accessibility and the quality of information provided. Future research should focus on integrating the advantages of these models to optimize patient education and engagement, while addressing the limitations identified to align more closely with the health literacy levels of the target patient demographics.

Funding Open access funding provided by Università degli Studi di Siena within the CRUI-CARE Agreement. No authors have received any funding or support.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest to disclose.

Human and Animal Rights, or Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent For this type of study, informed consent is not required.

Disclosure Each author does not have any commercial interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ramirez AE, Hsieh TY, Cardenas JP, Lao WW (2021) Abdominoplasty: my preferred techniques. *Ann Plast Surg* 86:S229–S234. <https://doi.org/10.1097/SAP.0000000000002639>
- Regan JP, Casaubon JT (2024) Abdominoplasty. In: StatPearls, Treasure Island, FL
- Taylor DA, Merten SL, Sandercoe GD, Gahankari D, Ingram SB, Moncrieff NJ, Ho K, Sellars GD, Magnusson MR (2018) Abdominoplasty improves low back pain and urinary incontinence. *Plast Reconstr Surg* 141:637–645. <https://doi.org/10.1097/PRS.00000000000004100>
- de Brito MJ, Nahas FX, Barbosa MV, Dini GM, Kimura AK, Farah AB, Ferreira LM (2010) Abdominoplasty and its effect on body image, self-esteem, and mental health. *Ann Plast Surg* 65:5–10. <https://doi.org/10.1097/SAP.0b013e3181bc30f7>
- Oranges CM, Schaefer KM, Haug M, Schaefer DJ (2016) The impact of aesthetic surgery on body image and its implications for mental and physical health. *Aesthet Surg J* 36:NP256–258. <https://doi.org/10.1093/asj/sjw066>
- Papadopoulos NA, Meier AC, Henrich G, Herschbach P, Kovacs L, Machens HG, Kloppel M (2019) Aesthetic abdominoplasty has a positive impact on quality of life prospectively. *J Plast Reconstr Aesthet Surg* 72:813–820. <https://doi.org/10.1016/j.bjps.2018.12.020>
- Papadopoulos NA, Staffler V, Mirceva V, Henrich G, Papadopoulos ON, Kovacs L, Herschbach P, Machens HG, Biemer E (2012) Does abdominoplasty have a positive influence on quality of life, self-esteem, and emotional stability? *Plast Reconstr Surg* 129:957e–962e. <https://doi.org/10.1097/PRS.0b013e31824ecc2b>
- Vidal P, Berner JE, Will PA (2017) Managing complications in abdominoplasty: a literature review. *Arch Plast Surg* 44:457–468. <https://doi.org/10.5999/aps.2017.44.5.457>
- Gisladottir U, Nakikj D, Jhunjunwala R, Panton J, Brat G, Gehlenborg N (2022) Effective communication of personalized risks and patient preferences during surgical informed consent using data visualization: qualitative semistructured interview study with patients after surgery. *JMIR Hum Factors* 9:e29118. <https://doi.org/10.2196/29118>
- Raper SE, Joseph J (2020) Informed consent for academic surgeons: a curriculum-based update. *MedEdPORTAL* 16:10985. https://doi.org/10.15766/mep_2374-8265.10985
- Bulloch G, Seth I, Lee CHA (2023) ChatGPT in surgical research and practice: a threat to academic integrity, authorship, and divergent thinking. *ANZ J Surg* 93:2270–2271. <https://doi.org/10.1111/ans.18544>
- Feng S, Shen Y (2023) ChatGPT and the Future of Medical Education. *Acad Med* 98:867–868. <https://doi.org/10.1097/ACM.0000000000005242>
- Gebrael G, Sahu KK, Chigarira B, Tripathi N, Mathew Thomas V, Sayegh N, Maughan BL, Agarwal N, Swami U, Li H (2023) Enhancing triage efficiency and accuracy in emergency rooms for patients with metastatic prostate cancer: a retrospective analysis of artificial intelligence-assisted triage using ChatGPT 4.0. *Cancers (Basel)* 15. <https://doi.org/10.3390/cancers15143717>
- Khan RA, Jawaid M, Khan AR, Sajjad M (2023) ChatGPT - Reshaping medical education and clinical management. *Pak J Med Sci* 39:605–607. <https://doi.org/10.12669/pjms.39.2.7653>
- Kuroiwa T, Sarcon A, Ibara T, Yamada E, Yamamoto A, Tsukamoto K, Fujita K (2023) The potential of ChatGPT as a self-diagnostic tool in common orthopedic diseases: exploratory study. *J Med Internet Res* 25:e47621. <https://doi.org/10.2196/47621>
- Lim B, Seth I, Kah S, Sofiadellis F, Ross RJ, Rozen WM, Cuomo R (2023) Using generative artificial intelligence tools in cosmetic surgery: a study on rhinoplasty, facelifts, and blepharoplasty procedures. *J Clin Med* 12. <https://doi.org/10.3390/jcm12206524>
- Xie Y, Seth I, Hunter-Smith DJ, Rozen WM, Ross R, Lee M (2023) Aesthetic surgery advice and counseling from artificial intelligence: a rhinoplasty consultation with ChatGPT. *Aesthetic Plast Surg* 47:1985–1993. <https://doi.org/10.1007/s00266-023-03338-7>
- Aljindan FK, Shawosh MH, Altamimi L, Arif S, Mortada H (2023) Utilization of ChatGPT-4 in plastic and reconstructive surgery: a narrative review. *Plast Reconstr Surg Glob Open* 11:e5305. <https://doi.org/10.1097/GOX.0000000000005305>
- Gupta R, Park JB, Bisht C, Herzog I, Weisberger J, Chao J, Chaiyasate K, Lee ES (2023) Expanding cosmetic plastic surgery research with ChatGPT. *Aesthet Surg J* 43:930–937. <https://doi.org/10.1093/asj/sjad069>
- Sharma SC, Ramchandani JP, Thakker A, Lahiri A (2023) ChatGPT in plastic and reconstructive surgery. *Indian J Plast Surg* 56:320–325. <https://doi.org/10.1055/s-0043-1771514>
- Li W, Chen J, Chen F, Liang J, Yu H (2023) Exploring the potential of chatgpt-4 in responding to common questions about abdominoplasty: an AI-based case study of a plastic surgery consultation. *Aesthetic Plast Surg*. <https://doi.org/10.1007/s00266-023-03660-0>

22. Oleck NC, Naga HI, Nichols DS, Morris MX, Dhingra B, Patel A (2023) Navigating the ethical landmines of chatgpt: implications of intelligent chatbots in plastic surgery clinical practice. *Plast Reconstr Surg Glob Open* 11:e5290. <https://doi.org/10.1097/GOX.00000000000005290>
23. Harrison CJ, Sidey-Gibbons CJ (2021) Machine learning in medicine: a practical introduction to natural language processing. *BMC Med Res Methodol* 21:158. <https://doi.org/10.1186/s12874-021-01347-1>
24. Nadkarni PM, Ohno-Machado L, Chapman WW (2011) Natural language processing: an introduction. *J Am Med Inform Assoc* 18:544–551. <https://doi.org/10.1136/amiajnl-2011-000464>
25. Matarasso A (1997) Awareness and avoidance of abdominoplasty complications. *Aesthet Surg J* 17(256):258–261. [https://doi.org/10.1016/S1090-820X\(97\)80008-8](https://doi.org/10.1016/S1090-820X(97)80008-8)
26. Sadeghi P, Duarte-Bateman D, Ma W, Khalaf R, Fodor R, Pieretti G, Ciccarelli F, Harandi H, Cuomo R (2022) Post-bariatric plastic surgery: abdominoplasty, the state of the art in body contouring. *J Clin Med* 11. <https://doi.org/10.3390/jcm11154315>
27. Cuomo R, Russo F, Sisti A, Nisi G, Grimaldi L, Brandi C, D'Aniello C (2015) Abdominoplasty in mildly obese patients (BMI 30–35 kg/m²): metabolic, biochemical and complication analysis at one year. *In Vivo* 29:757–761
28. Grimaldi L, Cuomo R, Brandi C, Botteri G, Nisi G, D'Aniello C (2015) Octyl-2-cyanoacrylate adhesive for skin closure: eight years experience. *In Vivo* 29:145–148
29. Voglino C, Tirone A, Ciuoli C, Benenati N, Paolini B, Croce F, Gaggelli I, Vuolo ML, Cuomo R, Grimaldi L et al (2020) Cardiovascular benefits and lipid profile changes 5 years after bariatric surgery: a comparative study between sleeve gastrectomy and Roux-en-Y gastric bypass. *J Gastrointest Surg* 24:2722–2729. <https://doi.org/10.1007/s11605-019-04482-9>
30. Voglino C, Tirone A, Ciuoli C, Benenati N, Bufano A, Croce F, Gaggelli I, Vuolo ML, Badalucco S, Berardi G et al (2021) Controlling nutritional status (CONUT) score and micronutrient deficiency in bariatric patients: midterm outcomes of Roux-en-Y gastric bypass versus one anastomosis gastric bypass/mini gastric bypass. *Obes Surg* 31:3715–3726. <https://doi.org/10.1007/s11695-021-05486-8>
31. Losco L, Roxo AC, Roxo CW, de Sire A, Bolletta A, Cuomo R, Grimaldi L, Cigna E, Roxo CDP (2022) Helix thigh lift. A novel approach to severe deformities in massive weight loss patients. *J Invest Surg* 35:620–626. <https://doi.org/10.1080/08941939.2021.1912220>
32. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, Loffler CML, Schwarzkopf SC, Unger M, Veldhuizen GP et al (2023) The future landscape of large language models in medicine. *Commun Med (Lond)* 3:141. <https://doi.org/10.1038/s43856-023-00370-1>
33. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW (2023) Ethics of large language models in medicine and medical research. *Lancet Digit Health* 5:e333–e335. [https://doi.org/10.1016/S2589-7500\(23\)00083-3](https://doi.org/10.1016/S2589-7500(23)00083-3)
34. Zerini I, Sisti A, Barberi L, Cuomo R, Tassinari J, Grimaldi L, D'Aniello C, Nisi G (2016) Body contouring surgery: our 5 years experience. *Plast Reconstr Surg Glob Open* 4:e649. <https://doi.org/10.1097/GOX.0000000000000636>
35. Cuomo R, Giardino FR, Nisi G, Brandi C, Zerini I, Voglino C, Gaggelli I, Grimaldi L (2019) Aspiration pneumonia: a shadow in post-bariatric patient: correlation between aspiration and mini-gastric bypass. *Obes Surg* 29:3937–3940. <https://doi.org/10.1007/s11695-019-04081-2>
36. Inojosa H, Gilbert S, Kather JN, Proschmann U, Akgun K, Ziemssen T (2023) Can ChatGPT explain it? Use of artificial intelligence in multiple sclerosis communication. *Neurol Res Pract* 5:48. <https://doi.org/10.1186/s42466-023-00270-8>
37. Seth I, Lim B, Cevik J, Sofiadellis F, Ross R, Cuomo R, Rozen WM (2024) Utilizing GPT-4 and generative artificial intelligence platforms for surgical education: an experimental study on skin ulcers. *Eur J Plast Surg* 47. <https://doi.org/10.1007/s00238-024-02162-9>
38. Seth I, Lim B, Cevik J, Gracias D, Chua M, Kenney PS, Rozen WM, Cuomo R (2024) Impact of nutrition on skin wound healing and aesthetic outcomes: a comprehensive narrative review. *JPRAS Open* 39:291–302. <https://doi.org/10.1016/j.jpra.2024.01.006>
39. Atkinson CJ, Seth I, Xie Y, Ross RJ, Hunter-Smith DJ, Rozen WM, Cuomo R (2024) Artificial intelligence language model performance for rapid intraoperative queries in plastic surgery: ChatGPT and the deep inferior epigastric perforator flap. *J Clin Med* 13:900. <https://doi.org/10.3390/jcm13030900>
40. D'Antonio A, Adesso M, Memoli D, Liguori P, Cuomo R, Boscaino A, Nappi O (2011) Lymph node-based disease and HHV-8/KSHV infection in HIV seronegative patients: report of three new cases of a heterogeneous group of diseases. *Int J Hematol* 93:795–801. <https://doi.org/10.1007/s12185-011-0849-0>
41. Sisti A, Huayllani MT, Boczar D, Restrepo DJ, Cinotto G, Lu X, Cuomo R, Grimaldi L, Nisi G, Forte AJ (2021) Umbilical reconstruction techniques: a literature review. *Aesthetic Plast Surg* 45:1078–1096. <https://doi.org/10.1007/s00266-020-01989-4>
42. Nisi G, Cuomo R, Brandi C, Grimaldi L, Sisti A, D'Aniello C (2016) Carbon dioxide therapy and hyaluronic acid for cosmetic correction of the nasolabial folds. *J Cosmet Dermatol* 15:169–175. <https://doi.org/10.1111/jocd.12213>
43. Brandi C, Cuomo R, Nisi G, Grimaldi L, D'Aniello C (2018) Face Rejuvenation: a new combined protocol for biorevitalization. *Acta Biomed* 89:400–405. <https://doi.org/10.23750/abm.v89i3.6162>
44. Barton N, Janis JE (2020) Missing the mark: the state of health care literacy in plastic surgery. *Plast Reconstr Surg Glob Open* 8:e2856. <https://doi.org/10.1097/GOX.0000000000002856>
45. Bustos VP, Haddad A, Kinney JR, Xun H, Nassar AH, Lee T, Lin SJ, Lee BT (2023) Evaluation of health literacy in plastic surgery using a crowdsourced patient survey. *Plast Reconstr Surg Glob Open* 11:e4803. <https://doi.org/10.1097/GOX.0000000000004803>
46. Roy M, Corkum JP, Urbach DR, Novak CB, von Schroeder HP, McCabe SJ, Okrainec K (2019) Health literacy among surgical patients: a systematic review and meta-analysis. *World J Surg* 43:96–106. <https://doi.org/10.1007/s00268-018-4754-z>
47. Tiourin E, Barton N, Janis JE (2022) Health literacy in plastic surgery: a scoping review. *Plast Reconstr Surg Glob Open* 10:e4247. <https://doi.org/10.1097/GOX.0000000000004247>
48. Mu X, Lim B, Seth I, Xie Y, Cevik J, Sofiadellis F, Hunter-Smith DJ, Rozen WM (2024) Comparison of large language models in management advice for melanoma: Google's AI BARD, BingAI and ChatGPT. *Skin Health Dis* 4:e313. <https://doi.org/10.1002/ski2.313>
49. Seth I, Lim B, Xie Y, Cevik J, Rozen WM, Ross RJ, Lee M (2023) Comparing the efficacy of large language models ChatGPT, BARD, and bing ai in providing information on rhinoplasty: an observational study. *Aesthet Surg J Open Forum* 5:84. <https://doi.org/10.1093/asjof/ojad084>

50. Da Silva M, Horsley T, Singh D, Da Silva E, Ly V, Thomas B, Daniel RC, Chagal-Feferkorn KA, Iantomasi S, White K et al (2022) Legal concerns in health-related artificial intelligence: a scoping review protocol. *Syst Rev* 11:123. <https://doi.org/10.1186/s13643-022-01939-y>
51. Naik N, Hameed BMZ, Shetty DK, Swain D, Shah M, Paul R, Aggarwal K, Ibrahim S, Patil V, Smriti K et al (2022) Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? *Front Surg* 9:862322. <https://doi.org/10.3389/fsurg.2022.862322>
52. Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R, Matsui Y, Nozaki T, Nakaura T, Fujima N et al (2024) Fairness of artificial intelligence in healthcare: review and recommendations. *Jpn J Radiol* 42:3–15. <https://doi.org/10.1007/s11604-023-01474-3>
53. Verdicchio M, Perin A (2022) When doctors and ai interact: on human responsibility for artificial risks. *Philos Technol* 35:11. <https://doi.org/10.1007/s13347-022-00506-6>
54. Vine M, Joseph K, Gibson D, Lim B, Chua M, Siu AHY, Dooreemeah D, Lee A, Cuomo R, Seth I (2024) Innovative approaches to preoperative care including feasibility, efficacy, and ethical implications: a narrative review. *AME Surg J*. <https://doi.org/10.21037/asj-23-41>
55. Seth I, Bulloch G, Lee CHA (2023) Redefining academic integrity, authorship, and innovation: the impact of ChatGPT on surgical research. *Ann Surg Oncol* 30(8):5284–5285. <https://doi.org/10.1245/s10434-023-13642-w>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.