Talwar *et al. Art Int Surg.* 2025;5:46-52
DOI: 10.20517/ais.2024.81

Artificial Intelligence Surgery

Perspective

Open Access

Check for updates

Natural language processing in plastic surgery patient consultations

Ankoor Talwar, Chen Shen, Joseph H. Shin

Division of Plastic Surgery, Department of Surgery, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03766, USA.

Correspondence to: Dr. Ankoor Talwar, Division of Plastic Surgery, Department of Surgery, Dartmouth-Hitchcock Medical Center, 1 Medical Center Drive, Lebanon, NH 03766, USA. E-mail: ankoor.a.talwar@hitchcock.org

How to cite this article: Talwar A, Shen C, Shin JH. Natural language processing in plastic surgery patient consultations. *Art Int Surg.* 2025;5:46-52. https://dx.doi.org/10.20517/ais.2024.81

Received: 19 Sep 2024 **First Decision:** 9 Dec 2024 **Revised:** 21 Dec 2024 **Accepted:** 25 Dec 2024 **Published:** 10 Jan 2025

Academic Editor: Andrew Gumbs **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Natural language processing (NLP) is the study of systems that allow machines to understand, interpret, and generate human language. With the advent of large language models (LLMs), non-technical industries can also harness the power of NLP. This includes healthcare, specifically surgical care and plastic surgery. This manuscript is an introductory review for plastic surgeons to understand the current state and future potential of NLP in patient consultations. The integration of NLP into plastic surgery patient consultations can transform both documentation and communication. These applications include information extraction, patient chart summarization, ambient transcription, coding, enhancing patient understanding, translation, and a patient-facing chatbot. We discuss the current progress toward building these applications and highlight their challenges. NLP has the potential to personalize care, enhance patient satisfaction, and improve workflows for plastic surgeons. Altogether, NLP can radically transform our current model of consultation into one that is more patient-centered.

Keywords: NLP, Large language models, GPT, transcription, coding, translation, literacy, Chatbot

INTRODUCTION

Natural language processing (NLP), a subfield of artificial intelligence (AI), focuses on the systems that allow machines to understand, interpret, and generate human language - a capability that is particularly significant. While NLP has been around for decades, it has mostly existed in small, non-generative forms.

Such models could understand text, but not create new text. With the release of Generative Pre-trained Transformer-3 (GPT-3) in 2021, generative large language models (LLMs) have performed well on both generative and non-generative tasks. Newer LLMs (i.e., GPT-4, Mistral, Claude, Bard, Perplexity) have only improved in these capacities. Plastic surgeons must understand how NLP tools can be harnessed to improve our workflows.

In plastic surgery, patient consultations are a critical component of the care process. Consultation requires effective communication and thorough documentation. The integration of NLP into consultations can enhance the quality of care and streamline this process. In this manuscript, we review the current state of clinical NLP integration and provide a perspective for future growth.

NLP is revolutionizing the two overarching domains of documentation and communication, summarized in Table 1. Examples of NLP tasks related to documentation include information extraction and summarization, ambient transcription, and coding. NLP tasks related to communication include understanding patient goals, patient-reported outcomes (PRO), translation, health literacy, and a patient-facing chatbot. We also discuss ethical considerations, limitations, and challenges of clinical NLP. We are still in the early stages of clinical NLP development. Plastic surgeons must help guide toward beneficial applications.

DOCUMENTATION

Information extraction and summarization

Plastic surgery is a highly specialized discipline. Plastic surgeons operate across the whole body. Unsurprisingly, our experts require very specific knowledge about patients and their medical history. This includes how well their comorbidities are controlled (i.e., diabetes) and information about their surgical history (i.e., history of abdominoplasty). Electronic health records (EHRs) contain a wealth of this information. However, most patient EHRs contain hundreds of clinical documents. Manually searching for information can be tedious due to the volume and jargon. For example, in a provider note, "PT" can mean "patient", "physical therapy", "posterior tibial artery", "posterior tibialis", "prothrombin time", or "part-time".

LLMs that comprehend clinical documentation should be able to understand the context and easily extract information. The most basic task is "named entity recognition" - deriving the names of patients, medical procedures, and medications directly written in a document[1]. Plastic surgeons might use named entity recognition to identify key surgical information, such as what tissue was resected, what type of mesh/implant was used, and what flaps were used in the reconstruction.

Summarization, on the other hand, is a more complex and generative task. It requires a holistic understanding of a text. Here, the generative LLMs could help plastic surgeons by processing several clinical documents in the EHR to summarize a patient's surgical history (i.e., all previous breast procedures) or overall health status. This would help plastic surgeons prepare before or during a consultation to guide operative planning. A group from Stanford applied eight different LLMs for clinical summarization and found several untrained LLMs more completely summarized patient history than humans and had fewer errors[2]. Importantly, their study highlighted the importance of "prompt engineering", that is, phrasing questions so the model can generate useful summaries. Their study also found that GPT-4 had the best performance compared to other models. Indeed, industry leader Epic © has already announced a forthcoming integration of GPT-4 into its platform[3]. This would enable surgeons to use Epic to instantaneously extract and summarize information for patient care. It would also facilitate powerful

Table 1. A summary of applications, potential, and the most prominent challenges/limitations of NLP in plastic surgery patient consultations

Application		Potential	Challenges and limitations
Documentation applications	Information extraction and summarization	NLP can find relevant information from a patient chart and succinctly summarize it	Models need to be integrated into EHRs
	Ambient transcription and coding	NLP can automate clinical documentation to help providers focus on the patient. They can even apply the correct codes for billing	Integration into EHRs. Additionally, inaccuracies or hallucinations are legal risks
Communication applications	Patient goals, PRO	NLP can enable surgeons to better understand goals to personalize treatment	Models need to be integrated into EHRs
	Translation and health literacy	NLP can translate patient conversations and documents across languages and literacy levels	Bias may lead to inaccuracies
	Chatbot	NLP chatbots can act as personal health assistants to engage patients in real time	Platforms must provide accurate information and deference to providers. Potential legal risks. Integration challenges and the need to use protected health information

NLP: Natural language processing; EHRs: electronic health records; PRO: patient-reported outcomes.

analytics and predictive modeling.

Ambient transcription and coding

Provider burnout is a well-described phenomenon in plastic surgery and healthcare. While the problem is multifactorial, several studies have described administrative burden as a root cause, including documentation[4-6]. The ability of NLP systems to generate clinical documentation could ease this burden through ambient transcription. This function could be integrated into plastic surgery consultations, which involve documentation of patient history, treatment plans, and consent forms. Traditionally, this process is time-consuming as surgeons must manually record patient information during or after consultations. Additionally, the current medico-legal milieu holds our discipline to the highest standard of accurate documentation. NLP can automate this process by transcribing spoken consultations into written records. Voice recognition systems can capture and transcribe the conversation between the patient and surgeon, reducing the administrative burden on the surgeon and allowing for greater attention toward the patient. Indeed, several startups are building and field-testing models to perform these tasks, including Ambience ©, Nabla ©, and others[7,8].

Once a patient's consultation is transcribed, clinicians should aspire to use NLP to automatically assign diagnosis and procedural codes. An editorial from Venkatesh *et al.* describes the current state of innovation in automated clinical coding (ACC), and the challenges facing companies[9]. While the need is apparent, it is difficult to integrate clinical data and documents into a few codes. Startups like AKASA © have made strides in building an ACC tool that performs as well as human coders[10]. In the future, an ACC tool could radically streamline the workflow for plastic surgeons.

In summary, the promise of NLP in plastic surgery documentation is widespread: improved accuracy and consistency of patient records, minimizing the risk of errors from manual entry, and decreased administrative burden on plastic surgeons.

COMMUNICATION

Patient goals, PRO

Plastic surgery is uniquely focused on patient goals and understanding. This focus arises from the nature of our services, which primarily aim to augment quality of life. It is no surprise our discipline has pioneered

modern PRO measures[11]. To this end, NLP systems should help us better understand patients and their goals. Ambient transcription systems will be able to identify important themes within a consultation and highlight what the patient perceives as important. For example, if a patient consistently expresses concerns about scarring, the surgeon can address this issue more thoroughly during and after the consultation.

NLP systems may also complete a PRO questionnaire during the consultation, alleviating the burden on patients. Having a preoperative BREAST-Q or Michigan Hand Questionnaire, for example, would help surgeons understand a patient's well-being in a more objective way. This would also facilitate postoperative patient tracking to measure a patient's progress using the same PRO instruments.

Translation and health literacy

LLMs should also assist in translating clinical materials, both in conversation and clinical documents. This would allow non-English speakers to more readily access care. This is particularly useful in plastic surgery, where a surgeon may need to communicate surgical details to a patient who is more comfortable with another language. Beyond language translation, LLMs can also address varying levels of health literacy within the same language. Studies have found translating clinical documents for geriatric patients may improve patient-provider interaction[12]. More generally, translating plastic surgery jargon like "NAC" and "IMF" into more familiar terms will aid patient understanding. For example, NLP can analyze the complexity of the language used in consent forms and provide suggestions for simplifying the text to match the patient's comprehension. Furthermore, NLP can be used to generate personalized educational materials tailored to the patient.

Chatbot

While NLP systems can facilitate conversations between plastic surgeons and patients, they may also conduct their own conversations to provide patients with real-time support and information. "Question answering" is a discipline in NLP involving systems that answer questions posed by humans, using a relevant context and in a natural language. Several foundation models, including GPT-4 and Med-PaLM 2, have shown promise in the medical question answering space[13,14].

Question answering chatbots will enable patients to access care on demand 24/7. After a consultation, patients can use a chatbot to answer common questions (i.e., about postoperative showering or dressings), provide reminders for medication or follow-up appointments, and even alert the surgeon if the patient reports any concerning symptoms. These systems can help optimize patients' conditions before and after their surgery, such as by advising on food and activities.

Still, recent otolaryngology and neurosurgery literature suggests we should temper our expectations. A study by Gajjar *et al.* compared versions of GPT in answering neurosurgery-specific patient questions. They found that, while factually accurate, the responses lacked readability and were not rated as highly beneficial[15]. Another study by Karimov *et al.* found ChatGPT was inferior to UpToDate in answering several otolaryngology-specific patient questions[16]. However, both these studies used untrained versions of GPT. Future investigations should attempt to train plastic surgery-specific chatbots to improve performance.

ETHICAL CONSIDERATIONS, LIMITATIONS, AND CHALLENGES

While the applications of NLP in plastic surgery consultations offer numerous benefits in both documentation and communication, there are also ethical considerations, limitations, and challenges that must be addressed. One major concern is the privacy and security of patient data. NLP systems rely on large

amounts of data, including sensitive patient information, to perform tasks effectively. Ensuring that these data are stored and processed in a HIPAA-secured manner is essential to protect patient confidentiality.

Another challenge is the potential for bias, a well-known limitation of LLMs. If the data used to train NLP models are not diverse or representative, the algorithms may produce biased or inaccurate results. This could lead to disparities in patient care, particularly for patients from underrepresented groups. Indeed, a recent study published in NEJM AI found that nearly all large-scale clinical datasets for training medical LLMs came from the Americas, Europe, and Asia, and covered nine languages in total[17]. This leaves people and languages underrepresented. To address this disparity, it is crucial to develop and train NLP models using diverse datasets and to continuously monitor and address bias. Another challenge for using clinical NLP will be integration into current systems and workflows. Many of the aforementioned technologies will have to integrate with EHRs and access protected health information. In addition, future research should compare multiple LLMs (GPT-4, Mistral, Claude, Bard, Perplexity, Claude, Bard, Perplexity, *etc.*) to find which ones are most appropriate for each clinical application.

A critical challenge is that LLMs are prone to "hallucination" - the generation of false content due to an LLM's extrapolation of its training data[18]. There are several examples in the literature[19,20]. Hallucinations and inaccuracies with LLMs will be a barrier to clinical adoption for medico-legal reasons, as the uncertain legal status of these models puts providers at risk. This must be addressed by the medical community before such technologies can be deployed. The key seems to be extensive prompt engineering, including the identification of best practices and close human oversight, as suggested by Shah[18]. Kwong *et al.* also suggest LLMs should indicate uncertainty when it is most appropriate[21].

While NLP has been purported to help in diagnosis and treatment plans in other specialties[22-24], this capacity may be limited in plastic surgery. This is because patients usually come in for a plastic surgery consultation with a known or apparent diagnosis (i.e., breast cancer, lipodystrophy, facial aging). Even if the diagnosis is unclear (i.e., hand pain), an astute plastic surgeon will rely on the physical exam and imaging to establish a diagnosis. A study by Pressman *et al.* reinforced this notion, as they purported to use LLMs to diagnose hand injuries based on clinical vignettes, but these vignettes included detailed physical exam and imaging findings[25]. Additionally, treatment planning across all plastic surgery also depends heavily on the physical exam, surgeon comfort, and patient goals - factors that NLP systems cannot assess.

CONCLUSION

This manuscript is an introduction for plastic surgeons to understand how NLP can be integrated into plastic surgery patient consultations to improve both documentation and communication. These applications include information extraction, summarization, ambient transcription, coding, patient understanding, translation, and a patient-facing chatbot. In doing so, NLP has the potential to personalize care, enhance patient satisfaction, and improve workflows for plastic surgeons.

However, there are ethical considerations and challenges associated with NLP development. Plastic surgeons should seek to create plastic surgery-specific models to maximize effectiveness. As NLP technology continues to advance, its role in plastic surgery consultations is likely to expand, offering new opportunities for innovation and patient-centered care.

DECLARATIONS

Authors' contributions

Made substantial contributions to the conception and design of the review: Talwar A, Shen C, Shin JH

Availability of data and materials
Not applicable.

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.

REFERENCES
1. Huang K, Altosaar J, Ranganath R. ClinicalBERT: modeling clinical notes and predicting hospital readmission. *arXiv*. 2019;arXiv:1904.05342. Available from: http://arxiv.org/abs/1904.05342. [accessed 6 Jan 2025]
2. Van Veen D, Van Uden C, Blankemeier L, et al. Clinical text summarization: adapting large language models can outperform human experts. *Res Sq*. 2023:rs.3.rs-3483777. DOI PubMed PMC
3. Epic and Microsoft bring GPT-4 to EHRs. Available from: https://www.epic.com/epic/post/epic-and-microsoft-bring-gpt-4-to-ehrs/. [Last accessed on 6 Jan 2025].
4. Seu M, Cho BH, Pigott R, et al. Trends and perceptions of electronic health record usage among plastic surgeons. *Plast Reconstr Surg Glob Open*. 2020;8:e2709. DOI PubMed PMC
5. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med*. 2016;165:753-60. DOI PubMed
6. Arndt BG, Beasley JW, Watkinson MD, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *Ann Fam Med*. 2017;15:419-26. DOI PubMed PMC
7. Ambience. Available from: https://www.ambiencehealthcare.com/. [Last accessed on 6 Jan 2025].
8. Nabla Copilot. Enjoy care again. Available from: https://www.nabla.com/. [Last accessed on 6 Jan 2025].
9. Venkatesh KP, Raza MM, Kvedar JC. Automating the overburdened clinical coding system: challenges and next steps. *NPJ Digit Med*. 2023;6:16. DOI PubMed PMC
10. Kim BH, Ganapathi V. Read, attend, and code: pushing the limits of medical codes prediction from clinical notes by machines. *arXiv*. 2021;arXiv:2107.10650. Available from: https://doi.org/10.48550/arXiv.2107.10650. [accessed 6 Jan 2025]
11. Talwar AA, Niu EF, Broach RB, Nelson JA, Fischer JP. Patient-reported outcomes: a primer for plastic surgeons. *J Plast Reconstr Aesthet Surg*. 2023;86:35-47. DOI PubMed
12. Abdi S, de Witte L, Hawley M. Exploring the potential of emerging technologies to meet the care and support needs of older people: a Delphi survey. *Geriatrics*. 2021;6:19. DOI PubMed PMC
13. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. *arXiv*. 2023;arXiv:2303.13375. Available from: https://doi.org/10.48550/arXiv.2303.13375. [accessed 6 Jan 2025]
14. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *arXiv*. 2022;arXiv:2212.13138. Available from: https://doi.org/10.48550/arXiv.2212.13138. [accessed 6 Jan 2025]
15. Gajjar AA, Kumar RP, Paliwoda ED, et al. Usefulness and accuracy of artificial intelligence chatbot responses to patient questions for neurosurgical procedures. *Neurosurgery*. 2024. DOI PubMed
16. Karimov Z, Allahverdiyev I, Agayarov OY, Demir D, Almuradova E. ChatGPT vs UpToDate: comparative study of usefulness and reliability of Chatbot in common clinical presentations of otorhinolaryngology-head and neck surgery. *Eur Arch Otorhinolaryngol*. 2024;281:2145-51. DOI PubMed PMC
17. Wu J, Liu X, Li M, et al. Clinical text datasets for medical artificial intelligence and large language models - a systematic review. *NEJM AI*. 2024;1:AIra2400012. DOI
18. Shah SV. Accuracy, consistency, and hallucination of large language models when analyzing unstructured clinical notes in electronic medical records. *JAMA Network Open*. 2024;7:e2425953. DOI PubMed
19. Beutel G, Geerits E, Kielstein JT. Artificial hallucination: GPT on LSD? *Crit Care*. 2023;27:148. DOI PubMed PMC

20. Zaretsky J, Kim JM, Baskharoun S, et al. Generative artificial intelligence to transform inpatient discharge summaries to patient-friendly language and format. *JAMA Netw Open.* 2024;7:e240357. DOI PubMed PMC
21. Kwong JCC, Wang SCY, Nickel GC, Cacciamani GE, Kvedar JC. The long but necessary road to responsible use of large language models in healthcare research. *NPJ Digit Med.* 2024;7:177. DOI PubMed PMC
22. Haider SA, Pressman SM, Borna S, et al. Evaluating large language model (LLM) performance on established breast classification systems. *Diagnostics.* 2024;14:1491. DOI PubMed PMC
23. Schmidl B, Hütten T, Pigorsch S, et al. Assessing the use of the novel tool Claude 3 in comparison to ChatGPT 4.0 as an artificial intelligence tool in the diagnosis and therapy of primary head and neck cancer cases. *Eur Arch Otorhinolaryngol.* 2024;281:6099-109. DOI PubMed PMC
24. Kunze KN, Varady NH, Mazzucco M, et al. The large language model ChatGPT-4 exhibits excellent triage capabilities and diagnostic performance for patients presenting with various causes of knee pain. *Arthroscopy.* 2024;In Press. DOI PubMed
25. Pressman SM, Borna S, Gomez-Cabello CA, Haider SA, Forte AJ. AI in hand surgery: assessing large language models in the classification and management of hand injuries. *J Clin Med.* 2024;13:2832. DOI PubMed PMC