Review

Natural Language Processing in Plastic Surgery Patient Consultations

Ankoor Talwar [1], Chen Shen [1], Joseph H. Shin [1,*]

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

Correspondence to: Dr. Joseph H. Shin, Division of Plastic Surgery, Department of Surgery, Dartmouth-Hitchcock Medical Center, Lebanon, NH, 03766, USA. Email: joseph.h.shin@hitchcock.org

Received: 21 October 2024 | Approved: 22 October 2024 | Online: 23 October 2024

Abstract

Natural Language Processing (NLP) is the study of systems which allow machines to understand, interpret, and generate human language. With the advent of large language models (LLM), non-technical industries. Can also harness the power of NLP. This includes healthcare, specifically surgical care and plastic surgery. 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 towards building these applications and highlight their challenges. NLP has potential to personalize care, enhance patient satisfaction, and improve workflows for plastic surgeons. Altogether NLP can radically transform our current model of consultation to one that is more patient-centered.

DOI: 10.20517/scierxiv202410.09.v1

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 which allow machines to understand, interpret, and generate human language. The latter quality is important. While NLP has been around for decades, it has mostly existed in small, individual, and non-generative forms. Such models could understand text, but not create new text. Starting in 2021, with the release of Generative Pre-trained Transformer-3 (GPT-3), generative large language models (LLMs) have performed well on both generative and non-generative tasks. Newer LLMs (i.e. GPT-4, Gemini, Mistral) have only improved in these capacities. There is great potential for healthcare, specifically plastic surgery, to harness NLP and 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 our consultations can enhance the quality of care and streamline this process. In this manuscript, the authors will review the current state of clinical NLP integration and provide perspective for future growth. We discuss how NLP is revolutionizing tasks in the two overarching domains of documentation and communication. 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, translation, health literacy, and a patient-facing chatbot. We also discuss ethical considerations, limitations, and challenges of clinical NLP. While there have been notable advances, we are still in the early stages of development. Plastic surgeons must help guide towards clinically beneficial NLP programs.

DOCUMENTATION

Information Extraction and Summarization

Plastic surgery is a highly specialized discipline. Plastic surgeons operate across the whole body. It is no surprise that our experts require very specific knowledge about patients and their medical

history. This includes how well their comorbidities are controlled (i.e. diabetes) and specific information about their surgical history (i.e. history of abdominoplasty). We are fortunate that EHRs contain a wealth of this information. However, most patient EHRs contain hundreds of clinical documents. Manually searching for key information can be tedious and time-consuming. This is because the language in clinical documentation has unique vocabulary and syntax. Further, medical jargon can be obfuscating. For example, in a provider note, "PT" can mean "patient", "physical therapy", "posterior tibial artery", "posterior tibialis", "prothrombin time", or "part-time".

LLMs which comprehend clinical documentation, should be able to understand the context and easily extract necessary 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, what flaps were used in reconstruction.

Summarization, on the other hand, is a more complex and generative task. It requires a holistic understanding of a document or 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 or abdominal procedures), or overall health status. This would help plastic surgeons prepare before or during a consultation to guide the operative plan. A group from Stanford applied eight different LLMs for clinical summarization and found several untrained LLMs more completely summarized patient history than humans, and with 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 allow surgeons using Epic to perform instantaneous information extraction and summarization to aid in patient care. It would also facilitate powerful analytics and predictive modeling.

Ambient Transcription and Coding

Provider burnout is a well-described phenomenon in plastic surgery and healthcare, more broadly. While the problem is multifactorial, several studies have described administrative burden as a root cause, including documentation[4–6]. The ability for NLP systems to generate clinical documentation could ease this burden through ambient transcription. This function could be integrated into plastic surgery consultations, which involve extensive documentation, including patient history, treatment plans, and consent forms. Traditionally, this documentation process is time-consuming and prone to errors, as surgeons must manually record patient information during or after consultations. NLP can automate this process by transcribing spoken consultations into written records. Voice recognition systems, powered by NLP, can accurately capture and transcribe the conversation between the patient and surgeon, reducing the administrative burden on the surgeon and allowing for more focus on patient care. Indeed, several startups are building NLP models to perform these tasks, including Ambience ©, Nabla ©, and others[7,8]. While the potential is there, we have a long way to go before ambient transcription systems are in widespread use.

Once a patient's consultation is transcribed, clinicians should aspire to use NLP for one step further – by automatically assigning diagnosis and procedural codes. An editorial from Venkatesh et al., describes the current state of innovation in automated clinical coding (ACC), and the challenges companies are facing[9]. Briefly, while the need is apparent, it is difficult to integrate varied clinical data and several clinical documents into a few codes. Still, startups like AKASA © have made strides building an ACC tool that performs as well as human coders[10]. A practice-ready ACC tool will likely be several years in the making but 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 surgeon.

COMMUNICATION

Patient Goals, Patient-Reported Outcomes

Plastic surgery is uniquely focused on patient goals and understanding. This is because most of our services are to augment quality of life. It is no surprise our discipline has pioneered modern patient-reported outcomes (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 the narrative of 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. This way the patient is not burdened with doing so, and it may be more accurately completed. Having a preoperative BREAST-Q or Michigan Hand Questionnaire, for example, would help surgeons understand a patient's wellbeing in a more objective scale. 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 translation of clinical materials, both in conversation and in clinical documents. This would allow non-English speakers to more readily access care. This is particularly useful in plastic surgery, where a highly specialized surgeon may need to communicate surgical details to a patient who is more comfortable with another language. The scope of translation also encompasses a single language but across levels of health literacy. It has been suggested that translating clinical documents uniquely for geriatric patients may improve the patient-provider interaction[12]. More generally, translating plastic surgery jargon like "NAC" and "IMF" into terms that are more familiar to patients will help in their 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 level. 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 automatically 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 will enable patients to access care on-demand 24/7. After a consultation, patients can use such a platform to answer common questions, provide reminders for medication or follow-up appointments, and even alert the surgeon if the patient reports any concerning symptoms. These systems can help "prehabilitate" and optimize patients before their surgery, for example advising on food and activities before surgery.

Our discipline has a way to go before building a beneficial patient-facing chatbot. Recent otolaryngology and neurosurgery literature suggests we should temper our expectations. A study from 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 from Karimov et al., found untrained ChatGPT was inferior to UpToDate in answering several otolaryngology-specific patient questions[16]. However, both these studies used untrained versions of GPT, and future research should focus on creating a plastic surgery-specific chatbot to improve performance.

ETHICAL CONSIDERATIONS, LIMITATIONS, 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 this data is 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 is 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 nearly all large-

scale clinical datasets for training medical LLMs came from the Americas, Europe, and Asia, and covered nine languages total[17]. It leaves many people and languages underrepresented. It is crucial to develop and train NLP models using diverse datasets and to continuously monitor and address any biases that may arise. Another challenge for using clinical NLP will be integration into current systems and workflows. Many of the aforementioned technologies will have to integrate with electronic health records in order to access and document patient information.

While NLP has been purported to help in diagnosis and treatment plans in other specialties, NLP is limited in that capacity 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 to establish a diagnosis. Finally, the treatment plan also depends heavily on the physical exam, surgeon comfort, and patient goals. These are not qualities which NLP systems can detect. It reinforces the creative and patient-centered nature of plastic surgery.

CONCLUSION

The integration of NLP into plastic surgery patient consultations holds significant promise for improving both documentation and communication. These applications include information extraction, patient chart summarization, ambient transcription, coding, patient understanding, translation, and a patient-facing chatbot. In doing so, NLP has 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 must seek to create plastic surgery-specific models to ensure that NLP applications are effective. As NLP technology continues to advance, its role in plastic surgery consultations is likely to expand, offering new opportunities for innovation and improvement in patient care.

DECLARATIONS

Authors' contributions

Made substantial contributions to 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. 2020;[DOI: 10.48550/arXiv.1904.05342]

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.[PMID:37961377 DOI:10.21203/rs.3.rs-3483777/v1 PMCID:PMC10635391]

3. Epic and Microsoft Bring GPT-4 to EHRs | Epic [Internet]. [cited 2023 Sep 30];Available from: https://www.epic.com/epic/post/epic-and-microsoft-bring-gpt-4-to-ehrs

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.[PMID:32440400 DOI:10.1097/gox.0000000000002709 PMCID:PMC7209869]

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:10.7326/m16-0961]

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.[PMID:28893811 DOI:10.1370/afm.2121 PMCID:PMC5593724]

7. Ambience | Your documentation on autopilot [Internet]. [cited 2023 Apr 12];Available from: https://www.ambiencehealthcare.com/

8. Nabla Copilot · Enjoy care again [Internet]. [cited 2023 Sep 18];Available from: https://www.nabla.com/[DOI:10.51517/nabla.v8i1]

9. Venkatesh KP, Raza MM, Kvedar JC. Automating the overburdened clinical coding system: challenges and next steps. *NPJ Digit Med* 2023;6:16.[PMID:36737496 DOI:10.1038/s41746-023-00768-0 PMCID:PMC9898522]

10. Kim BH, Ganapathi V. Read, Attend, and Code: Pushing the Limits of Medical Codes Prediction from Clinical Notes by Machines.2021;[DOI:10.48550/arXiv.2107.10650]

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.[PMID:37688832 DOI:10.1016/j.bjps.2023.08.008]

12. Abdi S, Witte L, Hawley M. Exploring the Potential of Emerging Technologies to Meet the Care and Support Needs of Older People: A Delphi Survey. *Geriatrics (Basel)* 2021;6:19.[PMID:33668557 DOI:10.3390/geriatrics6010019 PMCID:PMC8006038]

13. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on Medical Challenge Problems. 2023;[DOI: 10.48550/arXiv.2303.13375]

14. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large Language Models Encode Clinical Knowledge. 2022;[DOI:10.48550/arXiv.2212.13138]

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:10.1227/neu.0000000000003070] Caution!

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.[PMID:38217726 DOI:10.1007/s00405-023-08423-w PMCID:PMC10942922]

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.[DOI:10.1056/aira2400012]