Indian Journal of Surgical Oncology (September 2023) 14(3):537–539
https://doi.org/10.1007/s13193-023-01836-3

REVIEW ARTICLE

From ChatGPT to Treatment: the Future of AI and Large Language Models in Surgical Oncology

Adhitya Ramamurthi[1] · Chandrakanth Are[2] · Anai N. Kothari[1]

Received: 28 September 2023 / Accepted: 4 October 2023 / Published online: 17 October 2023

Abstract
This paper explores the transformative potential of Large Language Models (LLMs) within the context of surgical oncology and outlines the foundational mechanisms behind these models. LLMs, such as GPT-4, have rapidly evolved in terms of scale and capabilities, with profound implications for their applications in healthcare. These models, rooted in the Generative Pretrained Transformer architecture, exhibit advanced natural language understanding and generation skills. Within surgical oncology, LLMs, when integrated into a Generalist Medical AI (GMAI) framework, hold great promise in offering real-time support throughout the cancer journey. However, alongside these opportunities, this paper underscores the importance of ethical, privacy, and efficacy considerations, especially in light of issues like data drift and potential biases. Collaborative efforts among healthcare providers, AI developers, and regulatory bodies are pivotal in ensuring responsible and effective use of LLMs in surgical oncology, thereby contributing to enhanced patient care and safety. As LLMs continue to advance, they are poised to become indispensable tools in the delivery of high-quality, efficient care in this specialized medical field.

Keywords LLMs · Surgical oncology · AI

Editorial

Since the introduction of deep learning and neural networks in the early 2000s, there has been substantial interest in the application of artificial intelligence (AI) to help solve a multitude of problems in the healthcare industry [1–4]. With the increased digitization of health records and electronically available health data, the field of medicine, at first glance, appears to be an ideal candidate for the integration of AI to improve outcomes, cost, and efficiency. Recently, generative AI technologies, defined by their capability to create new content, have received a tremendous amount of recognition for their potential to improve human-AI system interactions. This includes large language models (LLMs), which are designed to comprehend and generate human-like text. These models have exhibited exceptional capabilities in areas such as language translation, content generation, and even programming assistance. With their ability to analyze, interpret, and generate text as well as learn from user input, LLMs will be an increasingly important tool across several industries. Within medicine, surgical oncology, given its complex patient population and recent advances in precision medicine and multimodal therapy, is an ideal area to consider how the application of LLMs could aid in the development of novel and innovative solutions to accelerate progress and improve patient care [5–8].

Prior to considering the ways LLMs could impact surgical oncology, it is crucial to understand the machinery that allows these models to behave the way they do. LLMs are fundamentally machine learning models that require rigorous training prior to any deployment. Broadly speaking, the main methods for training machine learning models are through supervised, semi-supervised, and unsupervised approaches. LLMs most commonly "learn" through unsupervised learning, where they are exposed to large amounts of unlabeled training data and identify patterns and associations. Typically, the training corpus is a large volume of text with billions of words from what is ideally a diverse set of

Disclosures Dr. Are serves as an editor for the Indian Journal of Surgical Oncology (IJSO)

✉ Anai N. Kothari
akothari@mcw.edu

[1] Department of Surgical Oncology, Medical College of Wisconsin, Milwaukee, WI, USA

[2] Department of Surgery, University of Nebraska Medical Center, Omaha, NE, USA

sources. During the training process, the LLM is attempting to predict subsequent words in a sentence. This is like other machine learning modeling tasks where the goal is to use input data (in this case, text) to predict an outcome (the next word). The computational needs for training LLMs can be quite demanding, typically requiring powerful high-performance clusters and infrastructure, and still often require weeks to months to process the high-volume of data. Availability of computational power and time have been a major limitation to advancements in the field of language modeling; however, recent technological advances are enabling researchers to develop more robust LLMs at previously unseen pace. OpenAI, the organization responsible for ChatGPT, released their first large language model in June 2018 which was trained using 117 million parameters. Less than 5 years later, they released GPT-4 which was trained using 175 billion parameters.

Diving deeper, GPT is an acronym that stands for generative pretrained transformer, a type of artificial neural network designed for natural language processing tasks. Pretraining occurs via masked language modeling, a process during which some words in a sentence are masked and the model must predict the missing words. By providing a large collection of text data, the model starts to acquire a broad understanding of language, including grammar, syntax, and semantics. The term transformer refers to a type of neural network architecture that is at the core of this LLM. The key components of a transformer include a self-attention mechanism, multi-head attention, positional encoding, and feedforward neural networks. The self-attention mechanism allows the model to figure out which words in a sentence are most important in relation to a longer piece of text while multi-head attention is like having many expert eyes reviewing a single piece of text, allowing the model to focus on different aspects of the text simultaneously. Transformers do not naturally know the order of words in a sentence, which is a problem that positional encoding solves. Beyond GPT, other common LLMs include BERT, transformer-XL, T5, RoBERTa, XLNet, and ERNIE, each developed by researchers at other major organizations such as Meta, Google, and Hugging Face [9]. It is important to emphasize that LLMs alone, regardless of the underlying model type, require some additional customization to accomplish specific tasks. This often involves fine-tuning on task-specific data, integrating with domain-specific knowledge bases, or other techniques to ensure the LLM is performing the pre-specified language task.

We are just beginning to see the ways LLMs might be used to impact the delivery of surgical oncology care through a generalist medical AI (GMAI) framework. GMAI is defined by key model capabilities that include flexible combinations of input data and dynamic task specification for quick adoption to new tasks by providing natural language descriptions [10]. For example, LLMs have shown capabilities in the generation of radiology reports that describe abnormalities and relevant findings from input images [11]. In surgical oncology, a similar approach could be used to assess preoperative images and generate annotated reports that include tumor location, predictions on resectability, or highlight key anatomic abnormalities. Another possible application is using GMAI chatbots, supported by LLMs, to assist patients through their cancer journey. GMAI chatbots could provide interactive summaries of diagnoses and treatment utilizing the patient's health data in real time [12]. They could assist with scheduling appointments, medication management, and treatment navigation. With further development, GMAI chatbots could serve as always-available resources to educate, monitor, and guide surgical oncology patients. As this field continues to rapidly change, new opportunities and applications will undoubtedly arise.

While the potential for improving patient care with LLMs is significant, the integration of these tools into clinical settings necessitates careful and thoughtful consideration of ethical, privacy, and efficacy concerns. The data used for training and model output interpretation need to be well understood, especially given the sensitive nature of health information. LLMs, like other AI models, are subject to the phenomenon of drift, where the model's performance deteriorates with time as the data it was trained on becomes less representative of current data due to shifts in user behavior or evolving language usage. This can result in LLMs becoming less accurate, outdated with respect to current trends, or even exhibiting biases that were not present before, which can be devastating in a clinical setting [13]. Addressing drift poses a significant challenge and robust strategies for model monitoring, maintenance, and updates are needed. The ongoing collaboration between healthcare providers, AI developers, and regulatory bodies will be instrumental in ensuring and maintaining high standards of patient care and safety.

LLMs equipped with their natural language understanding and generation capabilities, have the potential to transform numerous aspects of surgical oncology. However, any clinical implementation must include a cautious approach, with careful attention to ethical, privacy, and safety considerations to ensure responsible use. Importantly, as LLMs evolve and improve, so will the opportunities and applications to surgical oncology. Multi-disciplinary efforts will be essential in harnessing the full potential of these AI models. With the right safeguards in place, LLMs will be important adjuncts in the provision of high quality and efficient care in surgical oncology.

References

1. Ramgopal S, Sanchez-Pinto LN, Horvat CM, Carroll MS, Luo Y, Florin TA (2023) Artificial intelligence-based clinical decision support in pediatrics. Pediatr Res 93(2):334–341. https://doi.org/10.1038/s41390-022-02226-1
2. Mascagni P, Alapatt D, Sestini L et al (2022) Computer vision in surgery: from potential to clinical value. NPJ Digit Med 5(1):163. https://doi.org/10.1038/s41746-022-00707-5
3. Zhang Y, Jiang B, Zhang L et al (2022) Lung Nodule detectability of artificial intelligence-assisted CT image reading in lung cancer screening. Curr Med Imaging 18(3):327–334. https://doi.org/10.2174/1573405617666210806125953
4. Park CW, Seo SW, Kang N et al (2020) Artificial intelligence in health care: current applications and issues. J Korean Med Sci 35(42):e379. https://doi.org/10.3346/jkms.2020.35.e379
5. Kothari AN (2023) ChatGPT, large language models, and generative AI as future augments of surgical cancer care. Ann Surg Oncol 30(6):3174–3176. https://doi.org/10.1245/s10434-023-13442-2
6. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW (2023) Large language models in medicine. Nat Med 29(8):1930–1940. https://doi.org/10.1038/s41591-023-02448-8
7. Temsah O, Khan SA, Chaiah Y et al (2023) Overview of early ChatGPT's presence in medical literature: insights from a hybrid literature review by ChatGPT and human experts. Cureus. 15(4):e37281. https://doi.org/10.7759/cureus.37281
8. Gala D, Makaryus AN (2023) The utility of language models in cardiology: a narrative review of the benefits and concerns of ChatGPT-4. Int J Environ Res Public Health 20(15):6438. https://doi.org/10.3390/ijerph20156438
9. Egli A (2023) ChatGPT, GPT-4, and other large language models: the next revolution for clinical microbiology? Clin Infect Dis. https://doi.org/10.1093/cid/ciad407
10. Moor M, Banerjee O, Abad ZSH et al (2023) Foundation models for generalist medical artificial intelligence. Nature. 616(7956):259–265. https://doi.org/10.1038/s41586-023-05881-4
11. Tiu E, Talius E, Patel P, Langlotz CP, Ng AY, Rajpurkar P (2022) Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. Nat Biomed Eng 6(12):1399–1406. https://doi.org/10.1038/s41551-022-00936-9
12. Xu L, Sanders L, Li K, Chow JCL (2021) Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer 7(4):e27850. https://doi.org/10.2196/27850
13. Sahiner B, Chen W, Samala RK, Petrick N (2023) Data drift in medical machine learning: implications and potential remedies. Br J Radiol 96:20220878. https://doi.org/10.1259/bjr.20220878