ACS 2025 Surgeons and Engineers: A Dialogue on Surgical Simulation Meeting

P-A-09

Research Abstracts

A Proof of Concept Large Language Model (LLM) Assistant for Clinical Trial Screening in Surgical Oncology and Beyond

Samantha Lai; Alysala Malik; Tejas Sathe, MD; Caitlin Silvestri, MD; Gulam A. Manji, MD, PhD; and Michael D. Kluger, MD, MPH, FACS

Columbia University Vagelos College of Physicians and Surgeons, New York, NY; Columbia University Irving Medical Center, New York, NY

Introduction: Patient screening for clinical trials requires cross-referencing vast amounts of clinical information in patient charts against specific trial eligibility criteria. Given time constraints within a clinical encounter, screening often cannot be performed in real time, creating inefficiencies and lost opportunities for enrollment. We present initial validity evidence for the use of large language models (LLM) as a reliable screening tool to support clinical trial recruitment.

Methods: We developed an LLM application using LangChain and the GPT-4o model to assist in clinical trial screening. Patients who were screened for enrollment in 9 selected clinical trials at a single institution were used. Deidentified patient data from clinical notes and trial eligibility criteria from ClinicalTrials.gov was used as input. Model output included patient status (inclusion vs exclusion) with respect to each eligibility criterion as well as overall trial eligibility. Sensitivity and specificity were calculated for each criterion and overall trial eligibility.

Results: Twenty-four patients were used with a total of 105 individual eligibility criteria evaluated per patient. On average, the model correctly determined 101 out of 105 (96.2%) binary eligibility criteria. Sensitivity and specificity to overall trial eligibility ranged from 81.8% to 100% (mean = 94.2%) and 78.6% to 100% (mean = 95.1%), respectively (Figure 1). The mean cost for screening a patient was $0.63.

Conclusions: Overall, our model showed high efficiency and accuracy in selecting patients for appropriate clinical trials. Our results showed promise with a small cohort and future studies are needed to assess its accuracy with a larger sample of patients and trials. We present a framework for using this tool to widen the surface area covered by the screening process as well as adapting the model to any set of clinical trials, thereby opening the door for expansion to other surgical and medical disciplines.

Figure 1. Sensitivity and Specificity for Trial Eligibility.

Sensitivity was defined as identifying an individual by modeling to be eligible for a trial in alignment with a two-rater gold standard. Specificity was defined as identifying an individual by modeling to be not eligible in alignment with a two-rater gold standard.