Sentinel AI: An LLM-Driven Framework for Real-Time Automated Outbreak Detection Using Doctor-Patient Conversations

Jain, Arjun (School: D. W. Daniel High School)

Accurate and timely detection of emerging infectious outbreaks is critical for an effective public health response. While current systems like the National Syndromic Surveillance Program have aided in outbreak detection and have reduced surveillance lag, they primarily rely on the rigid, structured data captured from patient electronic health records. However, the recent, widespread growth of clinical, speech-to-text AI scribes has unlocked a large amount of unstructured, conversational data. Using this new data source, I developed Sentinel: a large language model (LLM)-based outbreak surveillance system that extracts both known and new symptoms directly from patient-clinician conversations. My study involved four steps: 1) creation of 500 synthetic clinician-patient conversations across four patient types: novel virus cases, influenza cases, other medical conditions, and healthy controls, 2) symptom extraction, which tested three models, 3) clustering, which merged categorical symptoms with semantic novel symptom embeddings, and 4) an interpretable dashboard. The few-shot and zero-shot LLM extraction methods produced comparable symptom extraction success with Macro-F1 Scores of 0.923 and 0.918, respectively, but the few-shot proved to be superior at detecting novel symptoms with a recall of 0.967. The HDBSCAN clustering algorithm correctly isolated the novel virus patients with a normalized mutual information score of 0.880, indicating high accuracy in matching patients to their respective case types. This project addresses a critical need for earlier detection of new diseases by developing a novel framework that extracts symptoms directly from doctor-patient conversations and provides an interpretable dashboard for healthcare stakeholders.

Awards Won:

Fourth Award of $600

Non-Trivial Ventures: Non-Trivial Fellowship Scholarship