

Patient records to timelines: A LLM-based approach for summarizing oncology patient history.

Sai Prabhakar Pandi Selvaraj, Sujoy Ganguly, Kunal Nagpal, Gabriel Altay, Hunter Lane, Victoria L. Chiou, Kathryn Mitchell, Guillem Garcia Gomez, Albert Coroleu Bonet, Arpita Saha; Tempus AI, Inc., Chicago, IL

Background: In the era of precision oncology, timely treatment decisions demand a clear view of a patient's clinical history, yet records are often fragmented, lengthy, and lack essential details in a structured format. Large Language Models (LLMs) can transform these notes into structured longitudinal timelines, helping providers quickly and accurately interpret a patient's journey and deliver optimal care. **Methods:** We synthesized a wide range of unstructured pan-cancer patient records into cohesive, easy-to-navigate timelines. The patient records range from dozens to hundreds of documents per patient (median number of documents = 139; total number of characters = 540k) across various types (e.g., pathology and progress reports, lab results, procedures, treatments). To obtain a timeline, we concatenate the documents and use Gemini-1.5-pro, a long-context LLM to identify key clinical events and details, (e.g., initial diagnosis, stages, symptom onset, treatment at all stages, side effects, and comorbidities) and enrich them with relevant excerpts. We generate timelines in a JSON format which also enables interactivity, e.g., filtering by event type. The clinical dataset used for evaluation contained 763 patients diagnosed with solid and hematologic cancers across various stages (e.g., lung cancer, leukemia, melanoma). To evaluate the quality of the timelines, we use the key medical information extracted manually (e.g., treatments, medications, diagnosis and stage, biopsy, and patient's general information like age and sex) by curators from the patient records, a total of 3839 datapoints across the dataset. For each curated data, a second LLM (GPT-4o) acts as a judge to assess if the data was accurately captured, missing, or misrepresented in the timeline (LLM-as-a-judge). Prior studies show LLM-as-a-judge aligns with human evaluations in clinical settings. **Results:** The timelines are 8976 ± 408 characters in length, achieving a $96.3 \pm 0.9\%$ reduction from the patient record length. Evaluation on the 763 patient dataset shows that $76.3 \pm 1.9\%$ of the curated data is accurately represented in the timeline, $17.3 \pm 1.6\%$ is missing, and $5.3 \pm 1\%$ is mismatched. Qualitative analysis indicates that missing information often stems from sensitivity to curation rules (e.g., pinpointing exact diagnosis dates in a complex diagnostic journey) rather than substantive differences, as these were not provided in the prompt. The generated timelines are also consistent across multiple runs with $< 2\%$ variation in the metrics. **Conclusions:** This study demonstrates LLMs' capability to transform unstructured oncology patient records into concise timelines, reducing the record length while retaining critical information. These timelines can help healthcare teams efficiently understand the patient journey and deliver cancer care. Research Sponsor: None.