

Cohort builder in Tempus Lens: Querying a large oncology database with generative AI.

Matthew West, Samuel P. Heilbroner, Sai Prabhakar Pandi Selvaraj, Hannah Weiss, John Yazbek, Peter Halloran, Abigail Koorwinder, Gabriel Altay, Nick Pojman, Jesus Pedrosa, Victoria L. Chiou, Arpita Saha; Tempus AI, Inc, Chicago, IL; Tempus AI, Inc., Chicago, IL; Tempus, Chicago, IL; Tempus AI, Chicago, IL

Background: Tempus Lens is a hosted web application that allows users to query multimodal data from millions of de-identified patient records in the Tempus Database. Patient inclusion/exclusion criteria are represented by filters, a group of which constitute a query. Lens has demonstrated utility in a range of oncology research use-cases, but lack of familiarity with the filtering interface can be a barrier to new users. Generative AI has shown promise in facilitating database querying. Here, we introduce Lens Cohort Builder – an LLM-based text-to-query tool within Tempus Lens that allows users to query Tempus data using only natural language prompts. **Methods:** Each filter makes its own corresponding call to an LLM (GPT-4o), with a custom prompt tuned to return the most relevant matches for its associated filter concept. These are run in parallel and a set of filters is returned. A subsequent LLM call passes these returned filters together with the original prompt to group concepts together based on logical relationships. The resultant query is populated in the UI and the user is offered to apply or cancel the suggested inclusion/exclusion criteria. To test the accuracy of the tool, beta testing was run with users with scientific and product backgrounds. Users generated queries and evaluated the accuracy of the Cohort Builder responses based on their subject matter expertise. Expert users also provided gold-standard labels for a subset of queries (n = 105). **Results:** A total of 33 users evaluated responses for a total of 1596 queries. Of these, 1212 (75.9%) were deemed to be accurate or mostly accurate, and 384 (24.1%) were deemed to be inaccurate or mostly inaccurate. Overall and filter-specific experimental results from beta testing are summarized in Table 1. N refers to the number of times a given concept appears across all filters in the labelled set of queries. True positives are defined as the tool predicting a filter when the target query actually included that filter. Good performance is observed for most of the commonly used filters, particularly Primary Diagnosis, Somatic Variant Genes, and Drug Class. **Conclusions:** Lens Cohort Builder was able to generate responses to oncological queries that were mostly evaluated to be “accurate” or “mostly accurate”. Limitations were found in the lack of ability to match functionality otherwise possible in Lens, such as expressing temporal or complex logical relationships between filters in a query, as well as users not always following the accompanying testing guidelines. Future development efforts will focus on increased accuracy and usability to support more detailed biomedical and scientific analyses. Research Sponsor: None.

Results metrics overall and for most commonly used filters.

Filter Name	N	Precision	Recall	F1
Overall	395	0.78	0.82	0.80
Primary Diagnosis	71	0.89	0.99	0.93
Somatic Variant Genes	25	0.77	0.96	0.86
DNA Modality	25	0.46	0.88	0.60
Biopsy Modality	22	0.63	0.91	0.74
RNA Modality	22	0.83	0.86	0.84