Talking to Electronic Health Records:
A Large Language Model Approach For Healthcare Data Analytics

Duke DEPARTMENT of COMPUTER SCIENCE

Carl Dong*, Raymond Xiong*, Danyang Zhuo, Anru Zhang

*Indicates Equal Contribution

CS

Problem and Motivation

Healthcare researchers rely on **Electronic Health Records (EHR)** to conduct various data analytics. Yet, many healthcare researchers lack enough **background in database systems** to handle **writing complex SQL queries** and visualization code, which hinders efficient data utilization and scientific discovery.

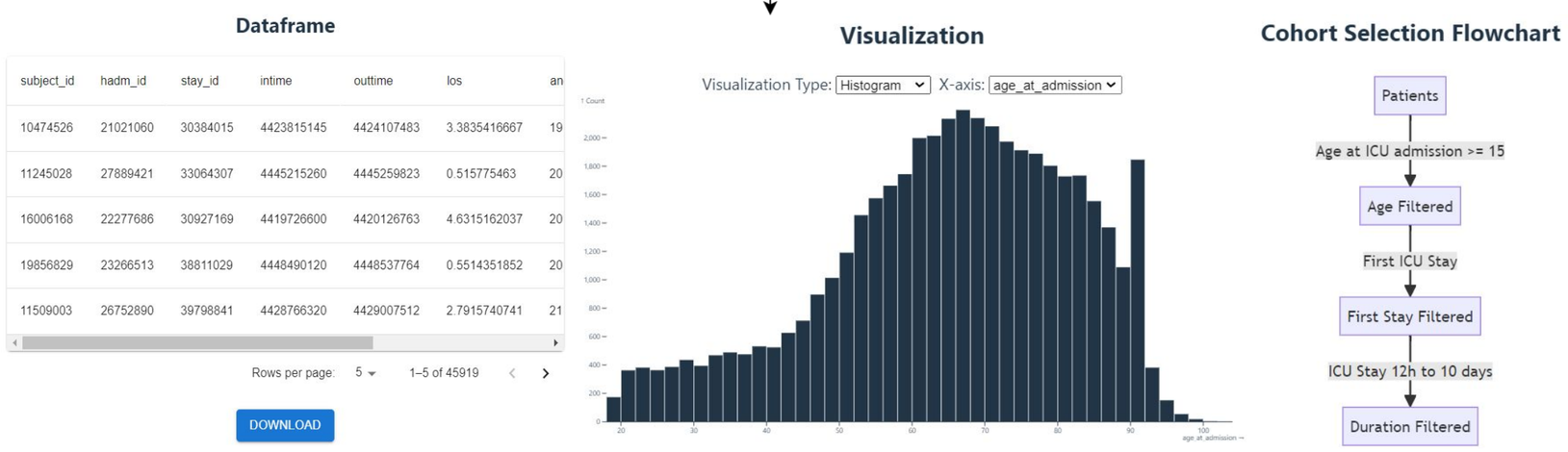
Solution

We built an **end-to-end AI application** that serves as a **natural language interface for EHR** by

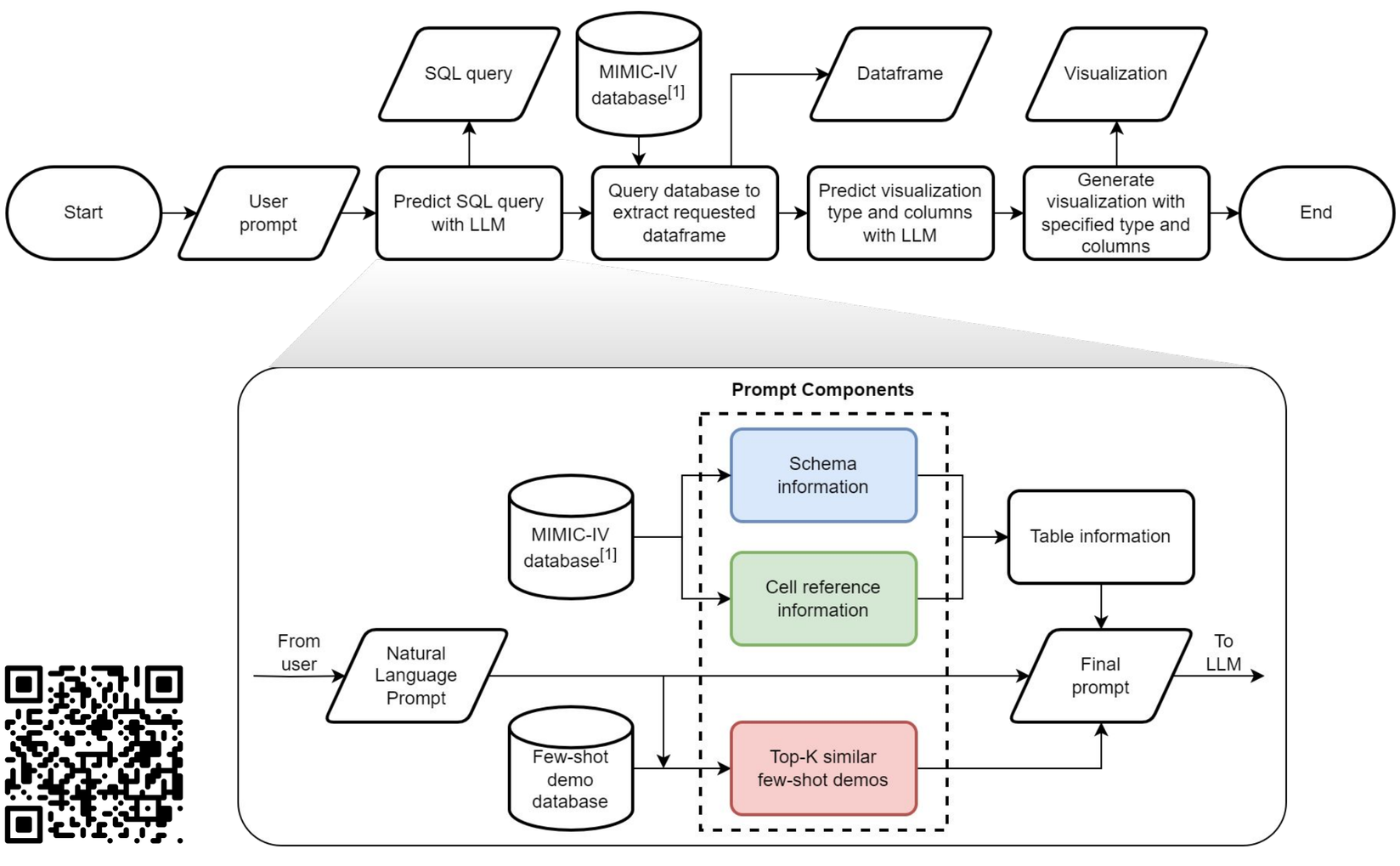
- leveraging **large language model (LLM)** for text-to-SQL, text-to-visual, and cohort-selection flowchart generation; and
- implementing state-of-the-art **prompt engineering methods** to improve LLM performance.

Use Case Example

Prompt: "Select all patients that satisfy the following criterias: (1) the patient is at least 15 years old at the time of ICU admission; (2) the ICU stay is the first known ICU stay of the patient; (3) the total duration of ICU stay is between 12 h and 10 days. Generate the distribution of patients' age at admission."



How does our system work?



Ablation Study

We removed one prompt component per trial to understand their contribution to model performance.

Execution accuracy (EX) formula

$$EX = \frac{S_{row} + S_{col}}{2},$$

$$S_i = \frac{|pred_i \wedge truth_i|}{|pred_i \vee truth_i|},$$

$pred_i$: elements in a predicted row/col
$truth_i$: elements in ground truth row/col
S_i: similarity score of row i or col i

Prompt	EX
Full Prompt (FP)	0.5234
W/O Schema Information	0.2489
W/O Cell Reference	0.2835
W/O Few Shot Demos	0.1817

Axis of FP results	Mean S_i
Column / Variable	0.6655
Row / Instance	0.3812

Conclusion: All prompt components are essential to ensuring performance. Our current LLM is better at selecting columns than selecting rows.

Future work

Phase 1: Internal testing within the Duke Bioinformatics Department; and
Phase 2: Public release.

References

1. Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2023). MIMIC-IV (version 2.2). *PhysioNet*. https://doi.org/10.13026/6mm1-ek67.
2. Li, Z., Wang, X., Zhao, J., Yang, S., Du, G., Hu, X., & Zhang, B. "PET-SQL: A Prompt-Enhanced Two-Round Refinement of Text-to-SQL with Cross-consistency." arXiv preprint arXiv:2403.09732. 2024. https://arxiv.org/abs/2403.09732