KARMA: A Knowledge Graph Framework for Explainable and Personalized AI in Healthcare

Udiptaman Das
Computer Science, School of Science and Engineering

UMKC School of Science and Engineering

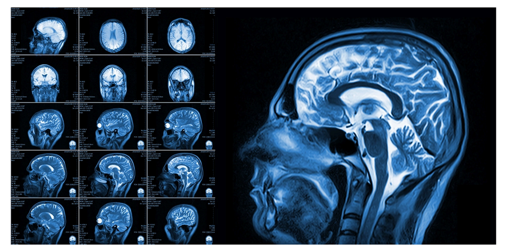
Introduction

Data Complexity in Healthcare: Clinical, genomic, imaging, and real-world data in healthcare are often siloed, heterogeneous, and lack semantic interoperability.
Challenges in Data Utilization: Fragmented data formats and inconsistent terminologies hinder integration, analysis, and reuse across systems.
AI Advancements: Deep learning and NLP models have improved diagnostics, treatment recommendations, and predictive analytics in oncology and beyond.
Limitations of Current AI:
- Often function as "black boxes" with limited explainability.
- CORAL [1] demonstrates the value of structured data extraction in oncology. However, it lacks ***robust ontology integration***, ***validations*** and does not support community-level patient insights.
- Struggle with incorporating domain knowledge and context.

Potential of Knowledge Graphs (KGs):
- Enable ***semantic integration*** of diverse data using standardized ontologies[2].
- Support ***explainable*** and ***context-aware*** reasoning.
- Facilitate personalized, evidence-driven clinical decision-making.

Domain

Struggles in Healthcare:
Healthcare grapples with integrating complex data (EHR, MRI scans, X-rays) for decision-making. Without AI, manual processes are slow and error-prone, often leading to generic treatments. With AI, lack of explainability hinders trust, and semantic data relationships are often missed, limiting personalization [3][4]



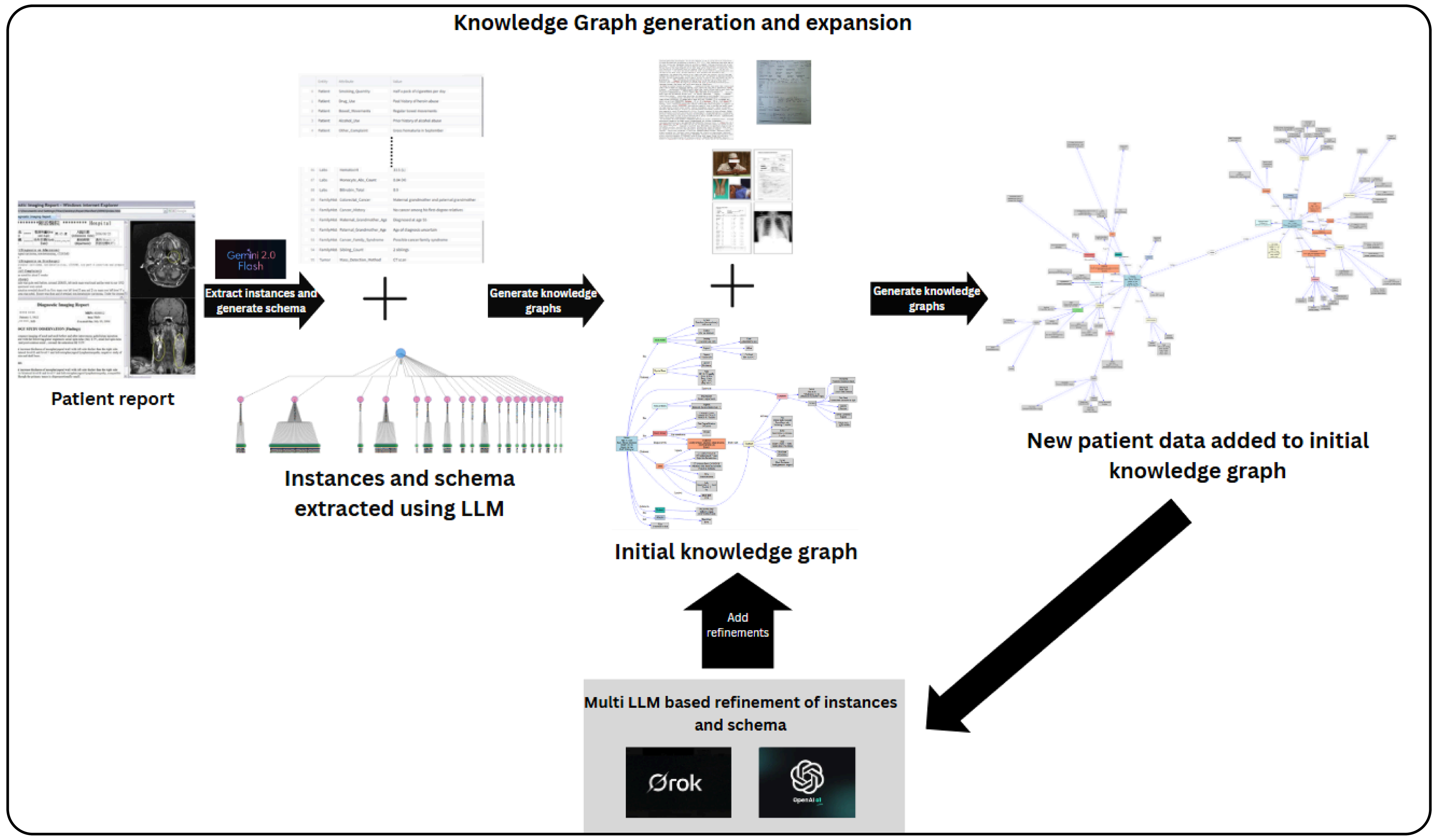
Relevance of KARMA:
KARMA introduces a knowledge graph framework to unify diverse healthcare data, making AI recommendations both explainable and personalized. We tested KARMA using the CORAL dataset (Expert-Curated Oncology Reports), which includes real-world oncology reports—detailed medical summaries of cancer patients, such as their diagnoses, treatments, and genetic information. For example, CORAL data might describe a pancreatic cancer patient's tumor mutations and past chemotherapy responses. KARMA uses this data to connect a patient's genetic profile to specific drug mechanisms, helping doctors understand why a particular treatment is recommended (e.g., suggesting a chemotherapy drug tailored to a patient's unique cancer markers). By making these connections clear and actionable, KARMA bridges the gap between complex data and personalized, trustworthy care, addressing both manual and AI-related challenges.

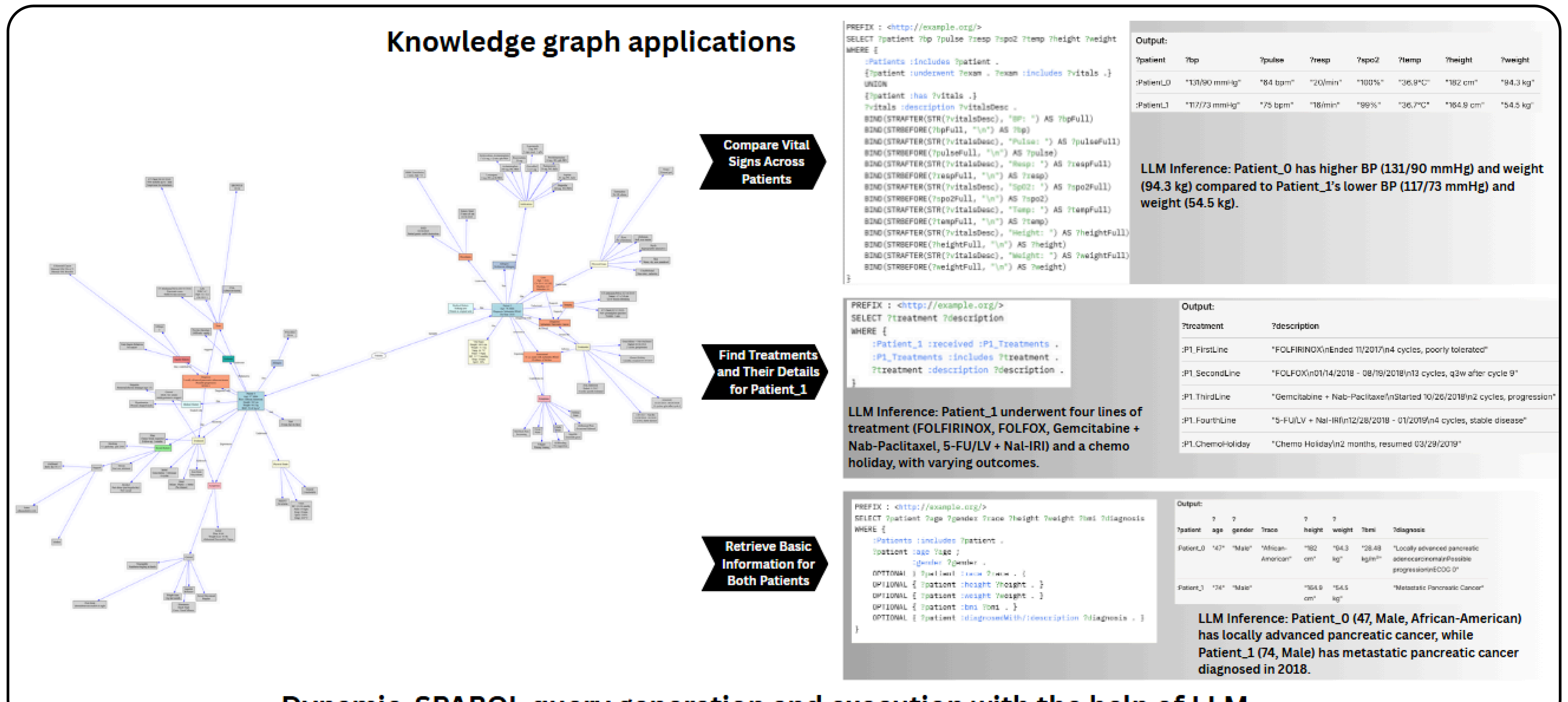
Acknowledgements

1. Dr. Yugyung Lee, supervised the research.
2. Dr. Duy H Ho, helped with frontend design and web UI implementation.
3. Atmakuri Krishnasai helped with data cleanup and management.

Methodology

- **Data Normalization**: Extract and standardize clinical data with metadata annotation for consistency.
- **Entity Extraction**: Use ***Gemini Flash 2.0*** to convert clinical entities into RDF instances.
- **Relationship & Schema Creation**: Define semantic relationships and construct initial ontology-aligned schemas.
- **Knowledge Graph Construction**: Build ontology-enriched patient KGs using structured RDF and standard ontologies.
- **Graph Expansion**: Incrementally update schemas and RDF with new patient data for KG growth.
- **Community Detection & Reasoning**: Apply Louvain clustering and semantic rules (SPARQL, SWRL) for personalization.
- **Dynamic Query Generation**: Convert clinician keywords into precise SPARQL queries via backend AI (Gemini Flash 2.0).
- **Variable Discovery Refinement**: Use Grok and GPT-o1 for second-pass entity-attribute refinement via model agreement.



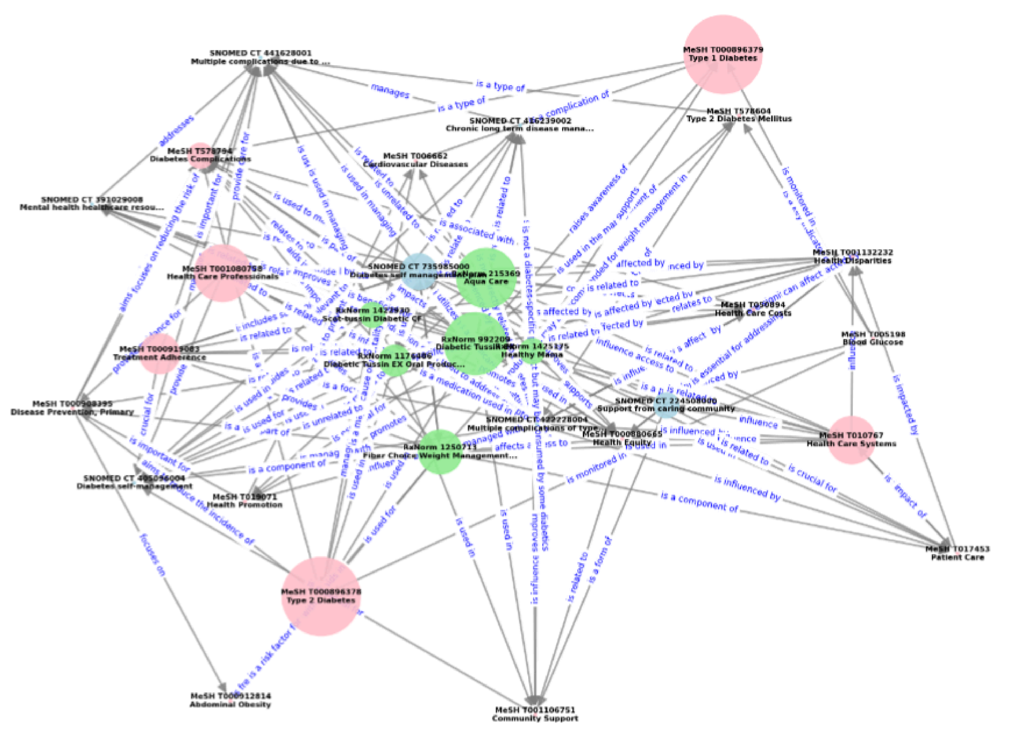


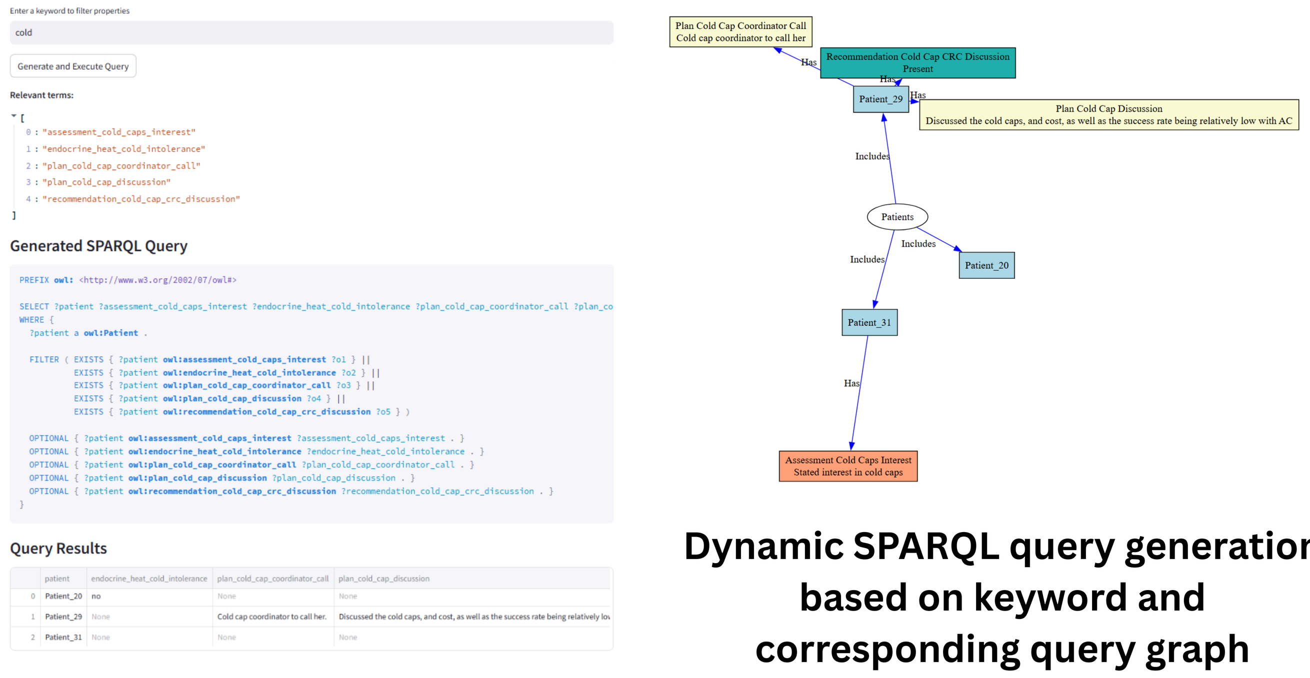
Dynamic SPARQL query generation and execution with the help of LLM

Results

A case study with two patient groups—**Breast Cancer (BRCA)** and **Pancreatic Cancer (PDAC)**, with 20 patients in each group—demonstrated KARMA Web's effectiveness of generating a knowledge graph for patient 0 and extending it by adding data from other patients subsequently.

Knowledge graph generated in application





Dynamic SPARQL query generation based on keyword and corresponding query graph

Future works and conclusion

KARMA enables scalable, ontology-enriched clinical knowledge graphs using LLMs.
Future work includes refining variable extraction with **GPT-o1** and **Grok**, evaluated via **ontology coverage, semantic consistency, hallucination rate, and information gain**—enhancing reliability and reducing model hallucinations.

References

[1] M. Sushil, V. E. Kennedy, D. Mandair, B. Y. Miao, T. Zack, and A. J. Butte, "Coral: expert-curated oncology reports to advance language model inference," NEJM AI, vol. 1, no. 4, p. AIdbp2300110, 2024.
[2] D.H. Ho, U.Das, R.Ho, and Y.Lee, "Leveraging Multi-Agent Systems and Large Language Models for Diabetes Knowledge Graphs," in \emph{2024 IEEE International Conference on Big Data (BigData)}, Washington, DC, USA, 2024, pp.3401--3410, IEEE.
[3] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. Briefings in Bioinformatics, 19(5), 1236-1246
[4] Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2019). What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923.