Strategic Data Evolution

Unlocking Domain Intelligence for a Competitive Edge

Steven Keith Platt

Director of Analytics and Executive Lecturer of Applied AI
Director, Lab for Applied AI and AI Business Consortium
LUC Site Director, IDEAL
splatt1@luc.edu

LOYOLA UNIVERSITY CHICAGO

The Evolving Data Landscape & The Need for Specialized AI

The Data Explosion Challenge

- Exponential growth in data volume and complexity.
- Proliferation of general-purpose LLMs (some with limited domain expertise).

Key Challenges

- **Domain knowledge gaps** - Limited understanding of industry-specific terminology, regulations, and practices.
- **Format inconsistency** - Inability to consistently follow specific output formats or templates.
- **Hallucinations** - Generation of plausible but factually incorrect information, especially for niche topics.
- **Security risks** - Potential exposure of sensitive information or compliance violations.

Solutions:

Fine-tuned LLM's and Knowledge Graphs (training local small LLM's are another option).

Why General-Purpose LLM's Fall Short

Key Limitations

Domain Knowledge Gaps

Lack of specialized knowledge in industry-specific terminology, regulations, security risks, and practices. General-purpose models have breadth but lack depth in specialized domains.

Hallucinations

Generation of plausible but factually incorrect information, especially for niche topics. This creates significant risks for business-critical applications requiring factual accuracy.

Format Inconsistency

Inability to consistently follow specific output formats, templates, non-deterministic behavior, or maintain corporate voice. This creates additional review overhead and compliance risks.

Explainability Challenges

Difficulty in tracing reasoning paths and providing transparent decision rationales. This creates barriers to trust, adoption, and regulatory compliance in high-stakes domains.

Unlocking Domain-Specific Intelligence & Relationship-Aware Analytics

> *General-purpose LLMs are like encyclopedias: vast, impressive, and instantly available.*
>
> *They don't know the latest research in your lab, the nuances of your regulatory environment, or the subtle relationships hidden in your data.*

- To unlock **domain-specific intelligence,** we need to fine-tune models with your organization's proprietary knowledge, terminology, and context.
- For **relationship-aware analytics,** knowledge graphs connect disparate data points into a unified semantic layer that reveals hidden patterns and insights.
- Both approaches must be **refreshed with real-time information** to maintain accuracy and relevance in rapidly changing environments.
- That's when AI stops guessing and starts **knowing** — delivering actionable intelligence tailored to your specific business context.

The AI Competitive Imperative

Strategic Value Proposition

Strategic imperative:

Organizations that fail to implement domain-specific AI risk significant competitive disadvantage as general-purpose AI commoditizes.

Domain-specific AI delivers measurable competitive advantage

through precision intelligence tailored to your business context.

Organizations implementing domain intelligence

create sustainable competitive moats through proprietary knowledge assets.

Business Impact

Data integration cost reduction through unified semantic layer.

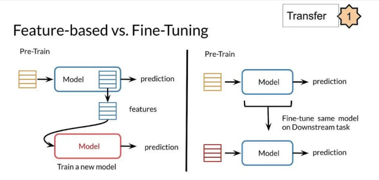
Industry benchmark estimate across enterprise implementations

Faster time-to-insight on complex business questions.

Source: "Knowledge graph-driven data processing for business intelligence" (Dey, 2024)

Strategic Solution Framework: Three-Pillar Approach

Fine-Tuning



Precision intelligence tailored to your specific business domain and output requirements.

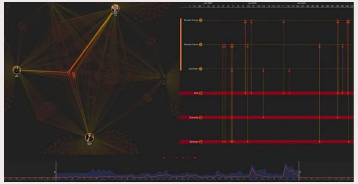
- Domain-specific accuracy and relevance.
- Consistent brand voice and output formats.
- Reduced review cycles and operational costs.

Knowledge Graphs

Relationship-aware intelligence that connects disparate data sources into a unified semantic layer.

- Hidden pattern discovery and strategic insights.
- Breaking down data silos across the enterprise.
- Industry benchmark: 40-60% data integration cost reduction.

GraphRAG



Enhanced retrieval-augmented generation that leverages relationship context for superior insights.

- Accuracy improvements of 6-12% for complex multi-hop queries.
- Significant reduction in hallucinations and factual errors.
- Enhanced explainability through transparent reasoning paths.

Fine-Tuning

Precision intelligence tailored to your specific business domain

Fine-Tuning of Large Language Models

When Fine-Tuning Makes Sense

Consistent Format & Voice

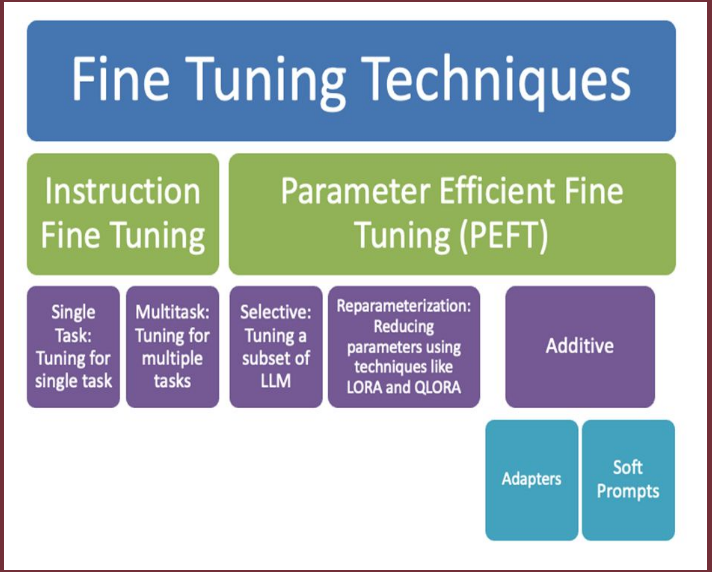
When you need task-specific behavior, structured outputs, or consistent brand voice. Fine-tuning excels at teaching models to follow specific patterns and styles.

Stable Domain Knowledge

For encoding domain expertise that rarely changes (SOPs, brand guidelines, regulations). Fine-tuning embeds this knowledge directly into model weights.

Efficiency & Performance

Creating smaller, specialized models with lower latency and cost via Parameter Efficient Fine-Tuning (PEFT). Techniques like LoRA and QLoRA enable efficient adaptation with minimal resources.



Fine-Tune Process Overview

Fine-tuning adapts pre-trained models with domain-specific data, transferring learned patterns and features to improve performance with:

1. **Define goal & metrics:** task scope, baseline, success criteria.

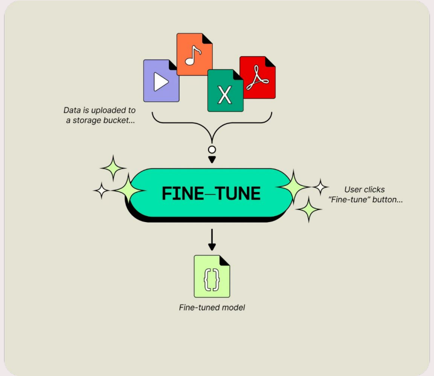
2. Data curation: collect/clean/label, split (train/val/test).

3. Model & method: pick base model + approach (LoRA/adapters) and confirm compute/budget.

4. Training plan: hyperparameters, early stopping, experiment tracking; watch for catastrophic forgetting.

5. Evaluate and align: automatic + human eval, robustness/safety/bias checks.

6. Deploy & monitor: optimize (quantize/distill), guardrails, drift monitoring.



Key Benefits

- **Reduced data requirements** compared to training from scratch.
- **Lower computational needs** through efficient adaptation Techniques.
- **Faster time to production** with accelerated training cycles.
- **Domain-specific accuracy** tailored to your business context.

When Fine-Tuning Isn't the Answer

Avoid Fine-Tuning When:

Rapidly Changing Facts

When information updates frequently, retrieval-based approaches are more efficient than constant retraining.

Limited Labeled Data

Fine-tuning requires substantial high-quality examples; insufficient data leads to poor performance.

Simple Prompting Would Suffice

When the need is primarily prompting discipline or tool invocation (function calling, workflows).

Decision Flow

Is knowledge stable?

Yes → Fine-tune for behavior/format

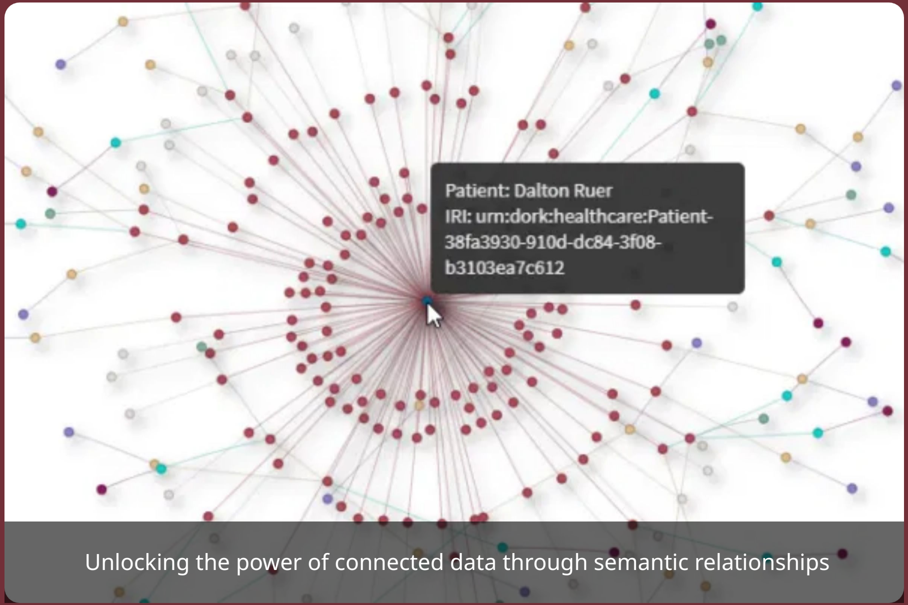
No → RAG for facts/freshness

Consider alternatives like:

Retrieval-Augmented Generation (RAG)
Few-shot prompting
Chain-of-thought techniques

Knowledge Graphs and GraphRAG

Connecting data into relationship-aware intelligence for deeper insights and context-driven understanding



Unlocking the power of connected data through semantic relationships

The Power of Knowledge Graphs

Knowledge graphs are structured representations of information that models entities, their attributes, and the relationships between them in a connected network structure.

Semantic Connections

Explicitly capture meaningful relationships between entities, enabling context-aware data analysis.

Inference Capabilities

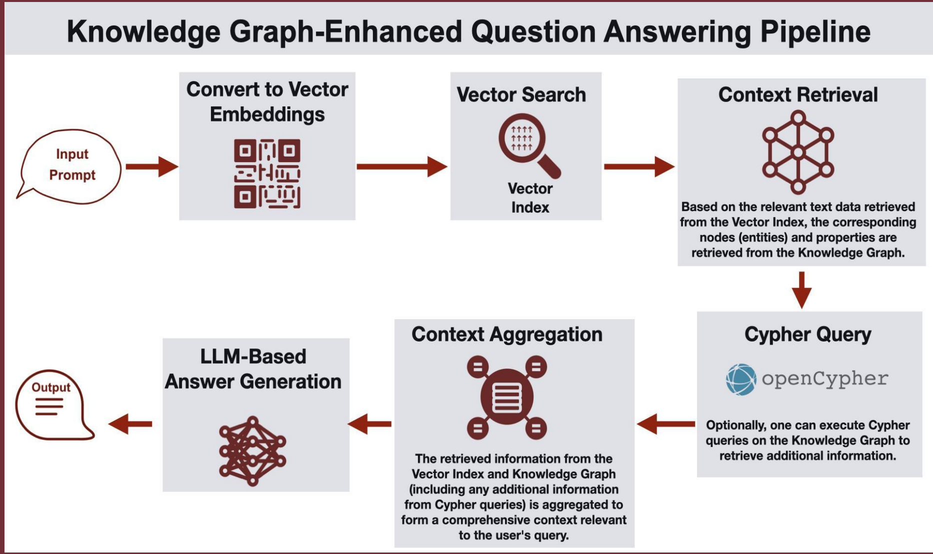
Derive new insights through traversal patterns and relationship analysis not explicitly stored.

Flexible Schema

Adapt to evolving data needs without rigid structure constraints, unlike traditional relational databases.

Integration Hub

Connect disparate data sources through a unified semantic layer, breaking down data silos.



Business Advantages of Knowledge Graphs

Knowledge graphs translate complex data relationships into tangible business outcomes across multiple dimensions.

Enhanced Decision Making

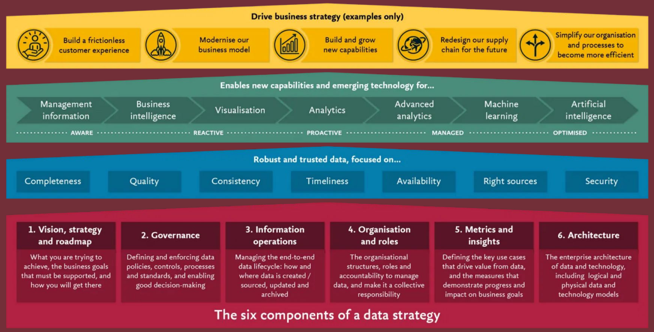
Reveal hidden patterns and connections that drive more informed strategic decisions. Knowledge graphs enable executives to see relationships that would otherwise remain obscured in siloed data.

Regulatory Compliance

Improve auditability and explainability of AI systems with transparent data lineage. This is particularly valuable in highly regulated industries where decision traceability is mandatory.

Operational Efficiency

Reduce data integration costs through unified data access and simplified architecture. Knowledge graphs create a semantic layer that abstracts away underlying data complexity.



Knowledge Graphs: Relationship Intelligence for Strategic Advantage

Business Value Proposition

Transform disconnected data into relationship-aware intelligence

that reveals hidden patterns and strategic opportunities.

Enable multi-hop queries and inference capabilities

that traditional databases cannot support, delivering deeper business insights.

Break down data silos through semantic integration,

creating a unified business intelligence layer across the enterprise.

Competitive Advantage

Create proprietary knowledge assets that form sustainable competitive moats as AI commoditizes.

Quantified Business Impact

Knowledge graphs facilitate significant data integration cost reduction by unifying diverse and fragmented data sources.

Faster time-to-insight on complex business questions.

Proven Business Impact: Industry Success Stories

BenevolentAI

Built a knowledge graph connecting biomedical literature, clinical trials, and molecular data to accelerate drug discovery.

48hrs

To identify baricitinib as COVID-19 treatment.

Source: Richardson, P.J., et al. (2022)

JPMorgan Chase

Implemented a knowledge graph to identify complex fraud patterns across transaction networks.

Significant

Improvements in fraud detection capabilities.

Note: Mastercard has reported using graphs to double compromised-card detection rates.

Mayo Clinic

Implemented a knowledge graph connecting patient data, treatments, and outcomes across departments.

Reported

Leverage KG to improve the patient journey and support clinical decision-making in clinical decision-making and resource allocation.

Source: Mayo Clinic Platform (2023) - External case study

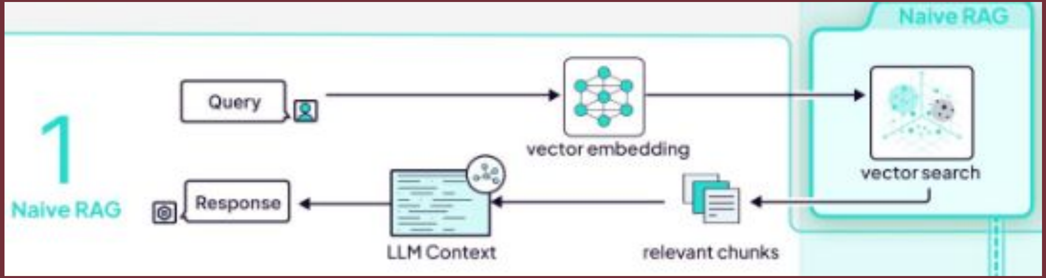
Alibaba

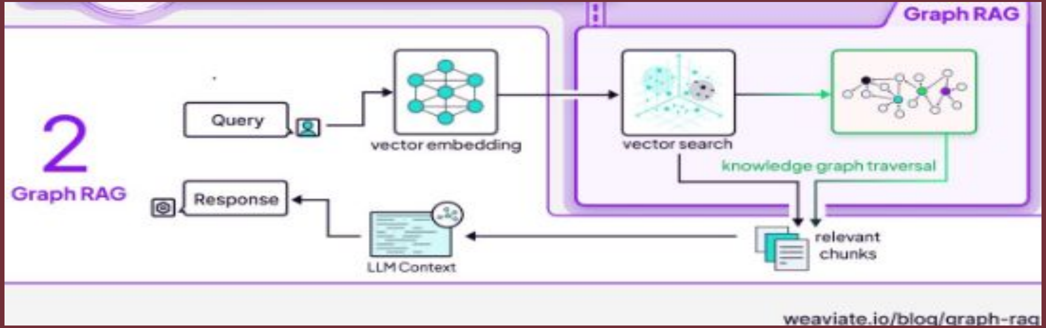
Built a product knowledge graph connecting items, attributes, user behaviors, and contextual data.

Demonstrated

External case study - specific metrics not verified in provided research

Rag and GraphRag





Naïve RAG

Treats every piece of data as an island. Each entry is independent, with no understanding of relationships beyond semantic similarity.

Limited to keyword matching and vector similarity, missing the contextual connections between entities.

GraphRAG

Leverages knowledge graphs to understand not just content, but connections.

Enables traversal across related entities to answer complex questions requiring multiple reasoning steps.

Preserves relationship context for more nuanced understanding and explainable answers.

The GraphRAG Advantage

How Knowledge Graphs Enhance RAG

Structural Context

Preserves relationships between entities, enabling more nuanced understanding than flat document retrieval. This context-aware approach captures the interconnected nature of information.

Enhanced Relevance

Improves retrieval precision by considering semantic relationships, not just keyword matching. This leads to more accurate and contextually appropriate results.

Multi-hop Reasoning

Enables traversal across connected nodes to answer complex questions requiring multiple reasoning steps. This allows for solving problems that would be impossible with traditional retrieval methods.

Explainable Pathways

Provides transparent reasoning paths that show how conclusions were reached, increasing trust. Users can follow the chain of evidence that led to a particular answer or recommendation.

Performance Impact

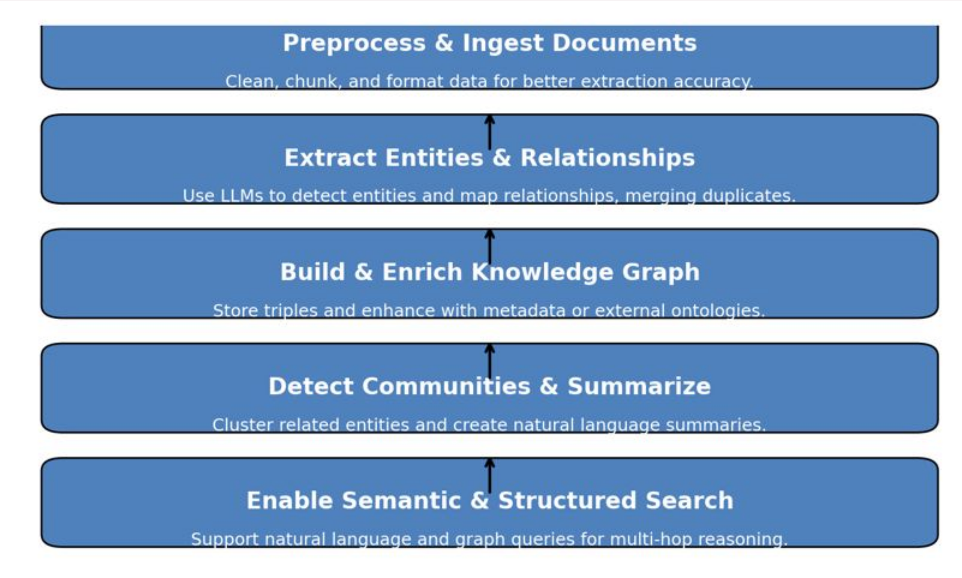
GraphRAG systems show 25-40% improvement in complex query accuracy compared to traditional RAG.

Key Insight: Graph-RAG's greatest advantage is in complex queries requiring multi-step reasoning and contextual understanding.

Enhancing Knowledge Graphs with RAG

GraphRAG implementation:

- Preprocess and ingest documents.
- Extract entities and relationships.
- Build and enrich the knowledge graph.
- Detect communities and generate summaries.
- Enable semantic and structured search.



CURRENT USE CASE- HEALTH INFORMATICS

Goal

Clinicians and researchers struggle to ask natural language questions and retrieve precise answers from complex clinical data.

Problem

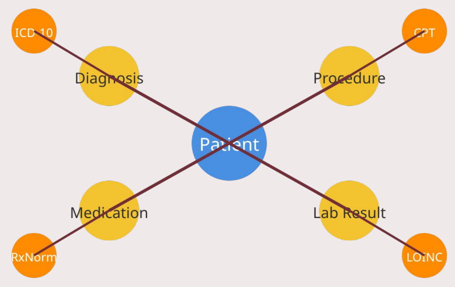
Clinical datasets (MIMIC-IV) provide structured representations, but nested and resource-centric architecture do not optimize for semantic reasoning or complex information retrieval. Also, we need to introduce external medical ontologies for comprehensive clinical understanding.

Proces
s

Transformed data to a knowledge graph, which was enriched and aligned with medical terminologies to create a unified semantic layer for clinical information.

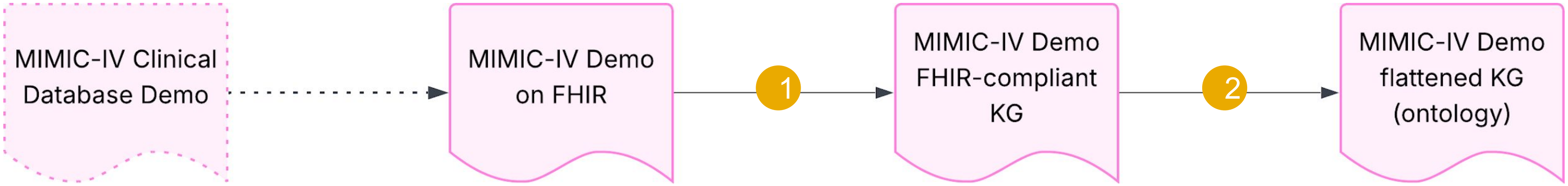
Solution

MIMIC-IV FHIR data transformed into **FHIR-compliant knowledge graph** , then an **ontology integration** of ICD-9 and ICD-10. This enables a retrieval system using LLM's and SPARQL queries (data retrieval and manipulation for graphs) for clinicians to get precise answers from complex clinical data.



MIMIC-IV clinical data transformed into a knowledge graph with medical ontology integration

CONSTRUCTING THE KNOWLEDGE GRAPH



- 100 deidentified Patients
- 26 tables (each contains from 5-15 columns)
- time frame: 2011 - 2013 or 2014 - 2016

12 resource type, each contain 17-98 fields:
- Observation, Count: 762633 documents
- MedicationAdministration, Count: 49635
- Medication, Count: 26536
- Specimen, Count: 23634
- MedicationRequest, Count: 16366
- MedicationDispense, Count: 13455
- Condition, Count: 4181
- Procedure, Count: 2105
- Encounter, Count: 401
- Patient, Count: 100
- Location, Count: 39
- Organization, Count: 1

Total: 898,996 documents

- Classes: 12 (from resource type)
- Object Properties: 13 (connection between 2 classes)
- Datatype Properties: 0 (predefined by fhir)
- **Entities: 887,390**
- **Edges: 33,133,045**
- **Nodes: 20,782,214**

- Classes: 15 (12 from before, and add DosageInstruction, MedicationMix, LocationEncounter)
- Object Properties: 16
- Datatype Properties: 103 (a value string or number or date)
- **Entities: 1,228,742**
- **Edges: 15,582,681**
- **Nodes: 4,252,079**

Knowledge Graph Construction Implementation

Defining Nodes and Relationships

- **Nodes (Entities):** Patient, Disease, Medication, Observation
  - Resource types become classes/nodes
- **Properties:** Patient demographics, disease descriptions, medication dosages
- **Relationships (Edges):** Patient HAS_DIAGNOSIS Disease
  - Intra-issue: hierarchical connections within tickets
  - Inter-issue: connections across different tickets

Ontology Integration

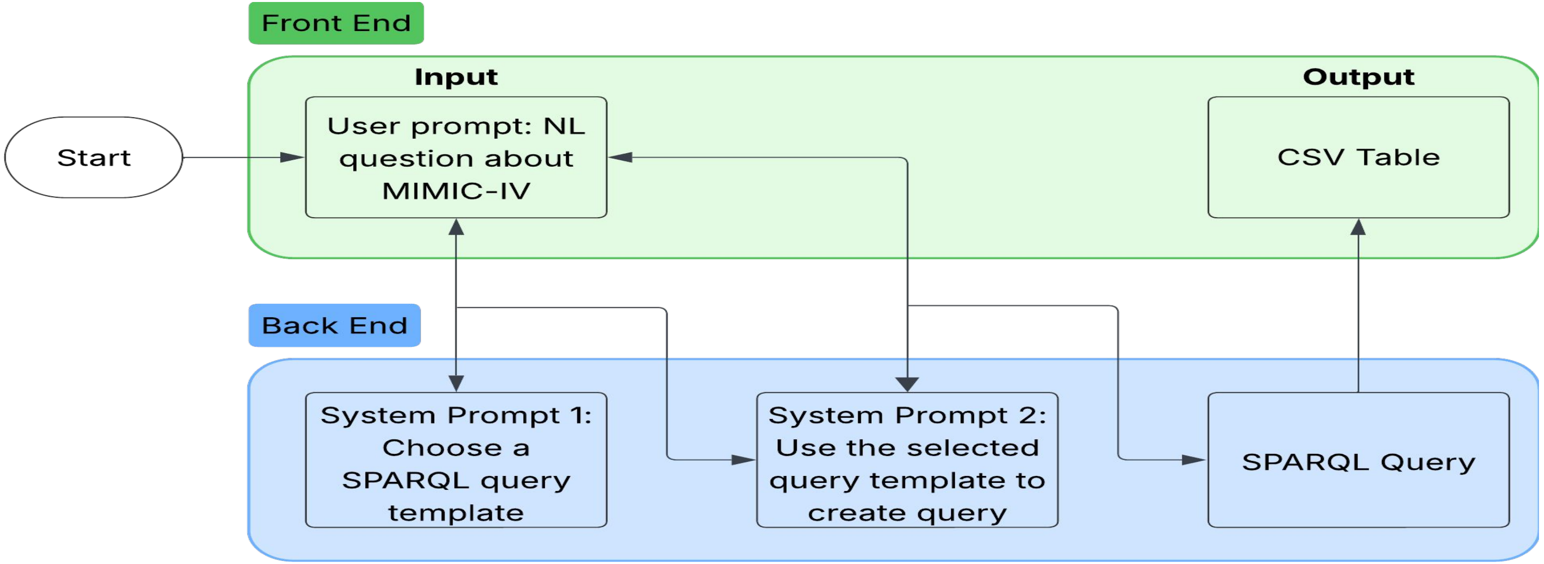
- **Medical Terminologies:** ICD-9, ICD-10, SNOMED CT
- **Unified Semantic Layer:** Consistent clinical information understanding
- **Vector Embeddings:** Pre-trained text-embedding models
  - Stored in vector database for semantic search

Scale and Complexity

1,228,742	15,582,681	4,252,079
Entities	Edges	Nodes

- **Enhanced Domain Modeling:** Added specialized classes like Dosage Instruction
- **Integration Flexibility:** Seamless new data integration without disrupting existing relationships

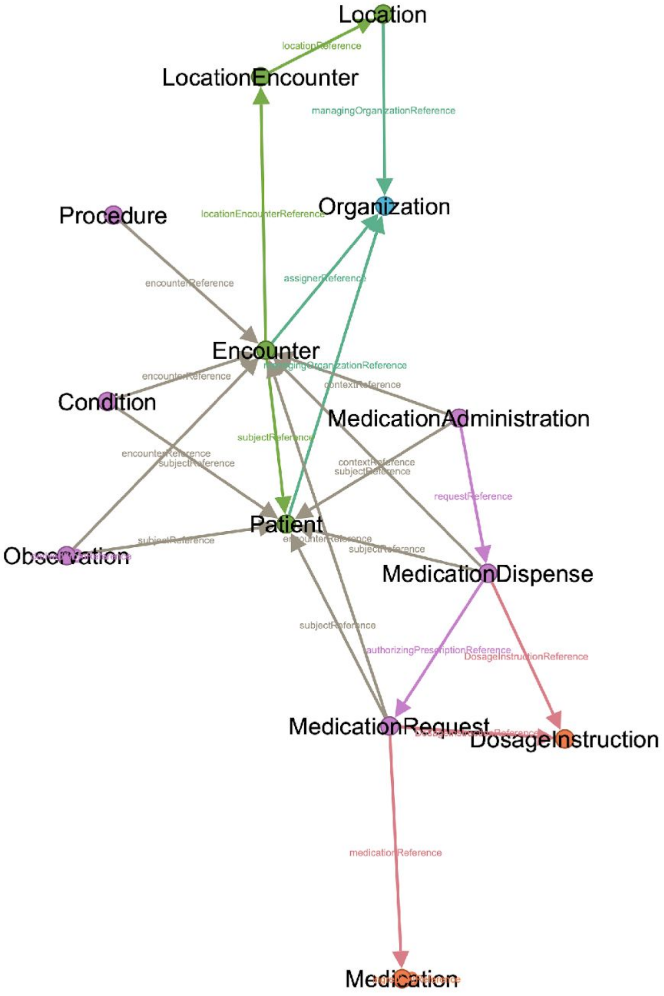
RETRIEVING INFORMATION PIPELINE



This hybrid system combines the reasoning power of LLM's with the accuracy and structure of SPARQL-based querying, facilitating access to complex clinical data.

OUTCOME

- A fully constructed FHIR-compliant Knowledge Graph of MIMIC-IV demo data.
- A flattened ontology structure with reduced nesting complexity.
- Successful integration with ICD-9, ICD-10, and SNOMED CT terminologies.
- A retrieval system with with high accuracy and faster query performance.



CLASSES AND EDGES OF

MIMIC-IV FLATTENED KNOWLEDGE GRAPH

Thank You!

NOTE: we are always interested in collaborating!! splatt1@luc.edu