# Legal Insights Engine

## DAMG 7374 - ST: Gen AI w/ LLM in Data Engineering

Mani Deepak Reddy Aila
Mitali Chouthai
Pranitha Poosa

Problem Statement

- In the legal domain, comprehending vast volumes of legal transcripts pose significant challenges.

- Who faces these challenges ?
  legal professionals, researchers, and normal people like us.

- The lack of tools to navigate and extract insights from these documents obstructs the accessibility and usability of crucial legal information.

- Traditional/Manual methods - time-consuming, error-prone, and often insufficient for handling the complexity and scale of legal data.
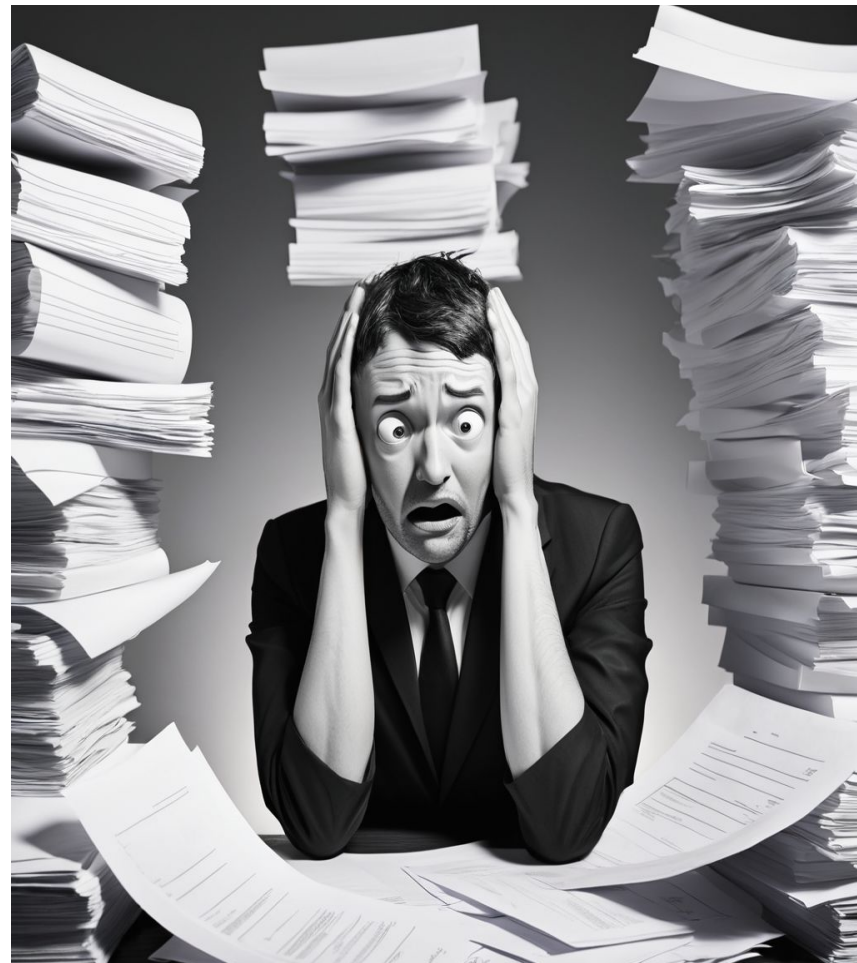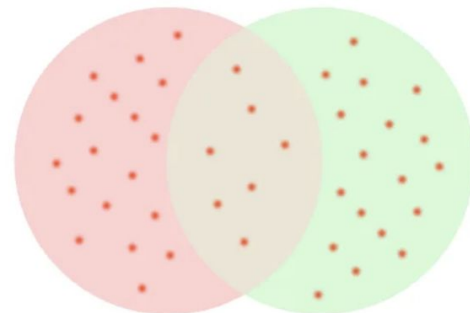
Solution
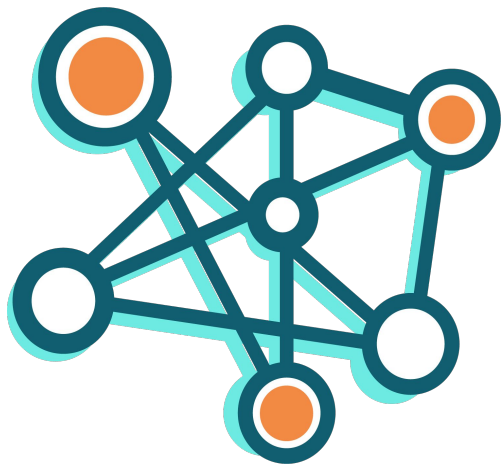
Graph Representation

Structured graph representation of legal cases to allow for efficient exploration, analysis, and interpretation.

KG based RAG Search

Predict case outcomes, identify relevant references and answer general questions by performing semantic search and graph search.

Similarity

Recommendations based on weighted similarity to identify multiple factors that might be important to establish relationship between two cases.

Dataset

- The [Case Law Access Project](#) (CLAP) - an initiative by the Harvard Library Innovation Lab

- Metadata of each case(JSON format) :

  - case ID

  - case name

  - decision date

  - the case it cites to

  - judges

  - opinions

- Single case view - https://api.case.law/v1/cases/435800/

- For this project, we are considering the jurisdiction of North Carolina

- Synthetic data with the help of LLM - case category

Raw Data Complexity

```
53958,11889181
53960,11913530
53962,11911479
53966,2511641,2529224,2550298,2571382,4757224,5314074,8564665
53968,2490196,4739612
53973,793157,793223,798956,2492599,2530367,2535184,2538239,25
53976,11888822
53978,11914395
53983,790147,790246,793180,793213,793223,795939,796105,796149
53986,2519615,2560558,8520186,8524041,8527889,8567219,8568007
53988,54098,790138,790211,790241,793083,793213,795906,795950,
53993,11868304
53997,796129,2402972,2498339,2504877,2509910,8521481,8525809,
54000,11889759
54001,11890040
54003,8561013,8562567
54005,11916400
54007,11889550
```

- **Adjacency matrix** in .csv format

- High complexity

- Difficulty in understanding

- Prone to error

Adjacency Matrix - First value represents the original case and the following are the cases the original case 'cites to'

Data Transformation : Graph Representation
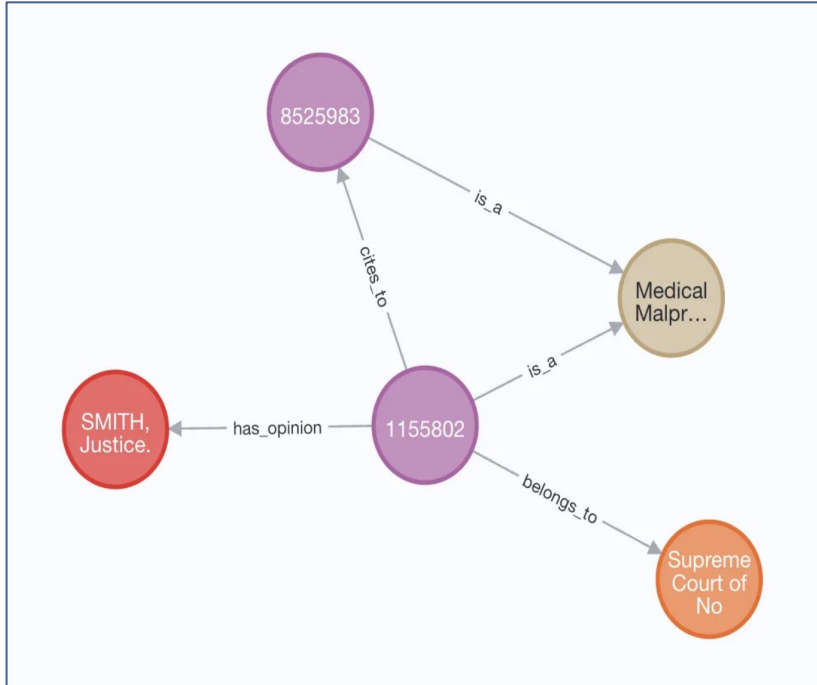
- Graph format emphasizes the visualisation of relationships between legal entities
- Intuitive navigation
- Scope for scalability
- Establish and explore connections
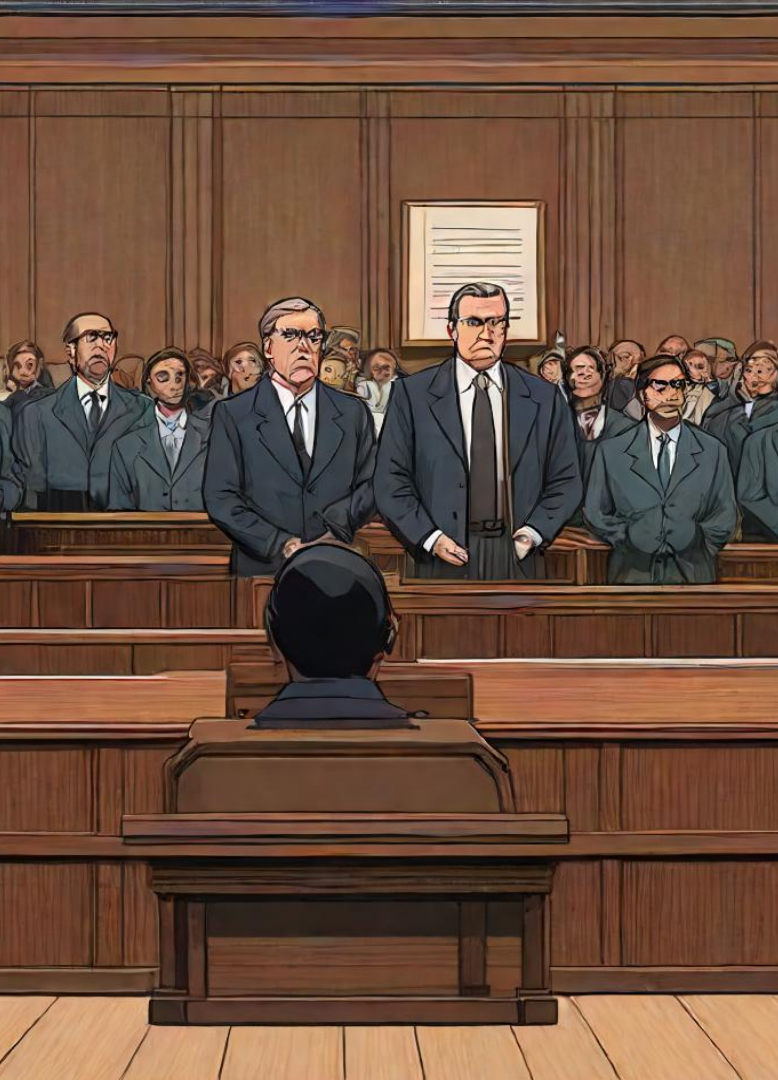
The Medical Malpractice Dilemma

What would you do if you found yourself stuck in a medical malpractice incident ?

How Legal Insights Engine Helps :

Graph Exploration

- Visualize information related to the case

- Explore more cases it cites to

- Easy to fetch references

- **Nodes:** Case, Category, Opinion, Court

- **Relationships:**

    - cites_to (Case - Case)

    - belongs_to (Case - Court)

    - is_a (Case - Category)

    - has_opinion (Case - Opinion)
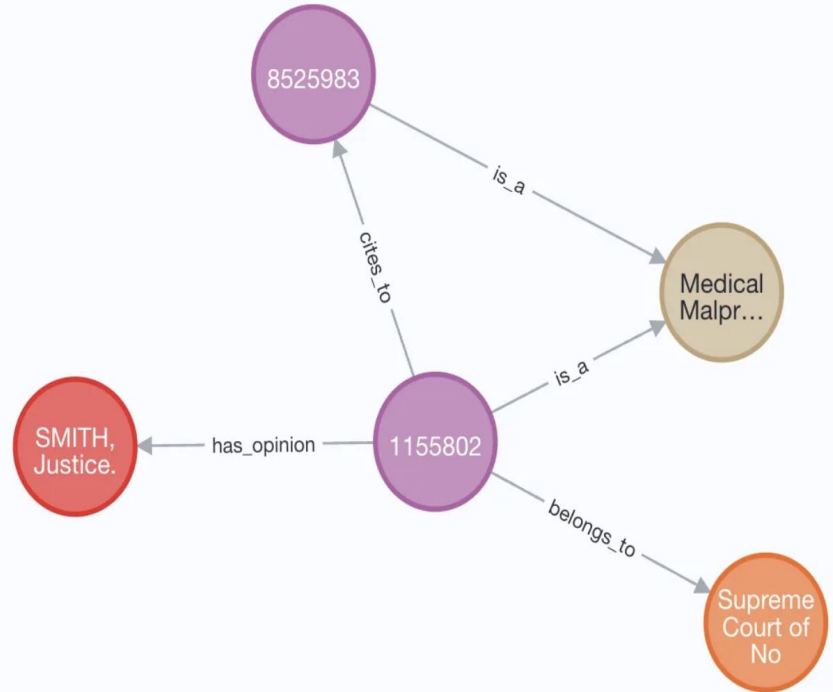
How Legal Insights Engine Helps :

Semantic search

- Semantic search vs. Keyword search
- Calculates similarity scores and retrieves the most similar opinion for a question by comparing word embeddings
- Retrieves additional information by graph traversal

Question:

?

What are the elements I need to prove to establish medical malpractice against my surgeon?

Why are we recommending similar cases?

Case-centric recommendations

- Recommend legal insights based on relationships with cited cases, court affiliations and case categories.
- Utilize graph-based analysis to offer recommendations tailored to specific legal cases.

Explainability

- Provides transparent insights into the reason behind recommendations.
- Enhances user understanding by demonstrating connections and criteria used in showing similar cases

What is Jaccard Similarity ?
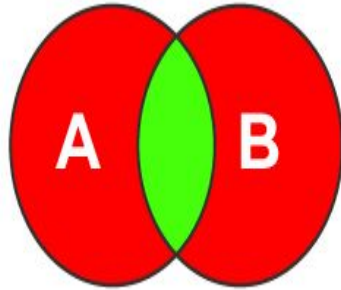
$$Jaccard = \frac{Intersection\ (A, B)}{Union\ (A, B)}$$

For a given relationship X between Case A and Case B:

Intersection (A,B) = Represents the number of common cases between A and B that share the relationship X

Union (A,B) = Denotes the total number of cases associated with either A or B, considering the relationship X

Weighted Jaccard Similarity - Heuristic Approach
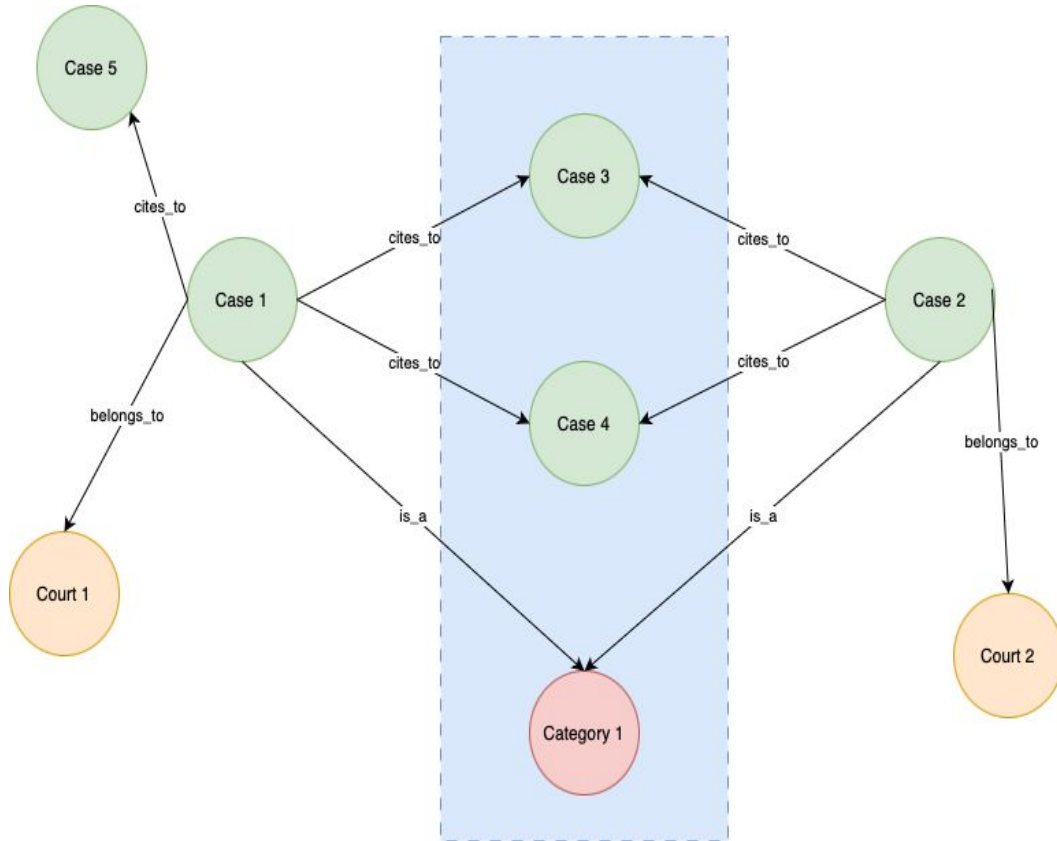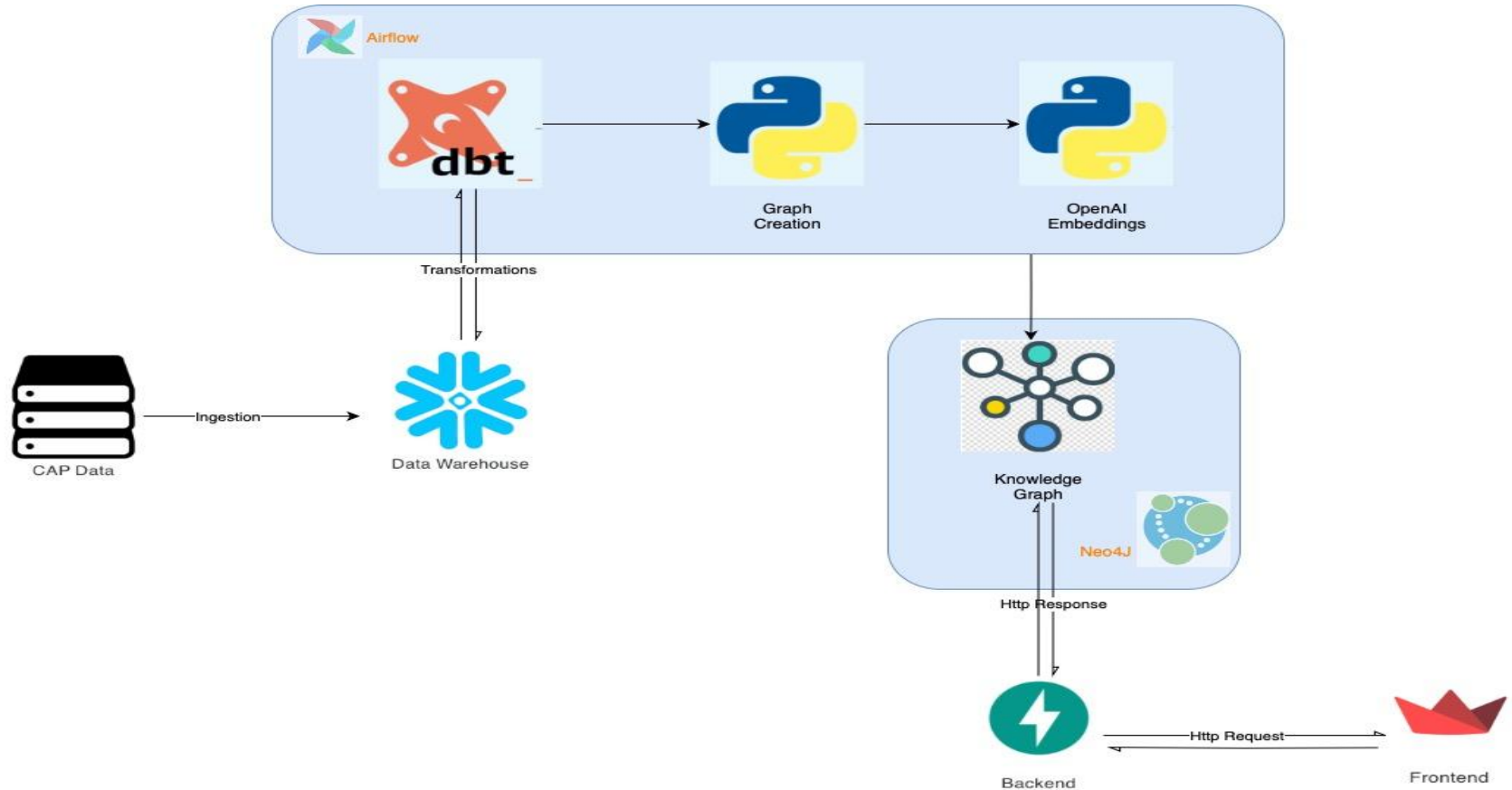

JS for relationship **'cites_to'** = 2/3

JS for relationship **'belongs_to'** = 0/2

JS for relationship **is_a = 1/1**

WJS = (w1 * js1) + (w2 * js2) + (w3 * js3)

where, w1 + w2 + w3 = 1

Architecture Diagram :

Business Value of the Product

- Easier to analyze large documents which would otherwise would be a lot of manual effort
- Visually explore complex relationships and nuances of legal cases
- Similarity based recommendations help in drawing parallels between the cases
- Explainability and transparency
- Semantic search and Graph search offer quick and accurate retrieval of information
- Users can do an informed decision-making

Cross-domain Applications

Healthcare -

- Analyze medical records, clinical trials, and drug discovery data
- Personalized treatment recommendations and decision support tools for healthcare professionals
- Improves patient outcomes and drives innovation in healthcare delivery

Finance Insights -

- Analyses market data, investment portfolios and economic indicators
- Provides investors and financial analysts with actionable insights for portfolio management and risk assessment
- Facilitates strategic planning and decision making in finance
- Panama Papers

Customer Support -

- Converting the large customer support manual into a graph.

Explore Further...

Source Code

Application Demo

Blog Post

Questions?