DocLens.ai

Intern Presentation

May 12th - July 12th 2025(Summer)

Yash Agrawal

A bit about me…👋

- In these 2 months
    - 🧑‍💻I got a chance to interact with almost everyone and work with Amit, Ramdoot, Shivam and a bit with Ankit
    - 🍵 Found a new liking for coffee
    - 💻Ran enough local models to justify the need of a graphics card in my laptop

- Scope of work: LLMs, LangChain(Embeddings, RAG), Knowledge Graphs, Synthetic Dataset Generation, NLP and ML
- Special Mention: backend test cases using Jest, learnt quite a bit about APIs, specifically all API calls our product(Claimlens) makes

My Work

Project #1: Medical Necessity Evaluation

Supervised by: Amit, Shivam

1. Web Scraper for the MayoClinic website, to scrape the medical information for all the diseases, to be used for generating the ‘Medical Necessity’ report
2. Implementation of 2 approaches for generating the ‘Medical Necessity’ report
   a. Embedding Vector search
   b. RAG Pipeline

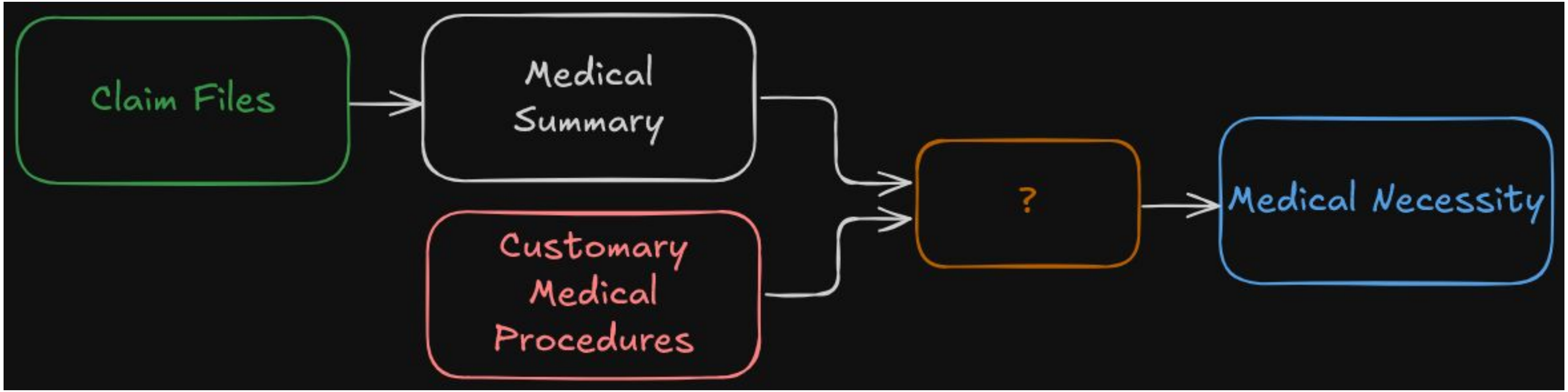
Project #2: Question Recommendation System for Chat

Supervised by: Ramdoot

1. Question clustering - Finding out the different types of questions that are asked in the chat
2. Knowledge graph based generation of QA(Question and Answer) dataset and RAGAs
3. LLM based generation of QA dataset, using DeepEval
4. Topic modelling of documents to generate questions

Problem Statement, Hypothesis

Project #1 - Medical Necessity Evaluation



Problem Statement:
The current LLM-based approach to generating Medical Necessity reports lacks reliability due to hallucinations, outdated information, and hard to verify outputs.

Hypothesis:
Providing relevant scraped ground truth information (MayoClinic) as context to the LLM, can significantly improve response accuracy and reduce hallucinations.

Methods(Ideas):

1. LLM - Knowledge Base(Current)
2. LLM + Tools(Internet Access)
3. LLM + Local database of MayoClinic's publicly available data
4. LLM + database passed as System Prompt

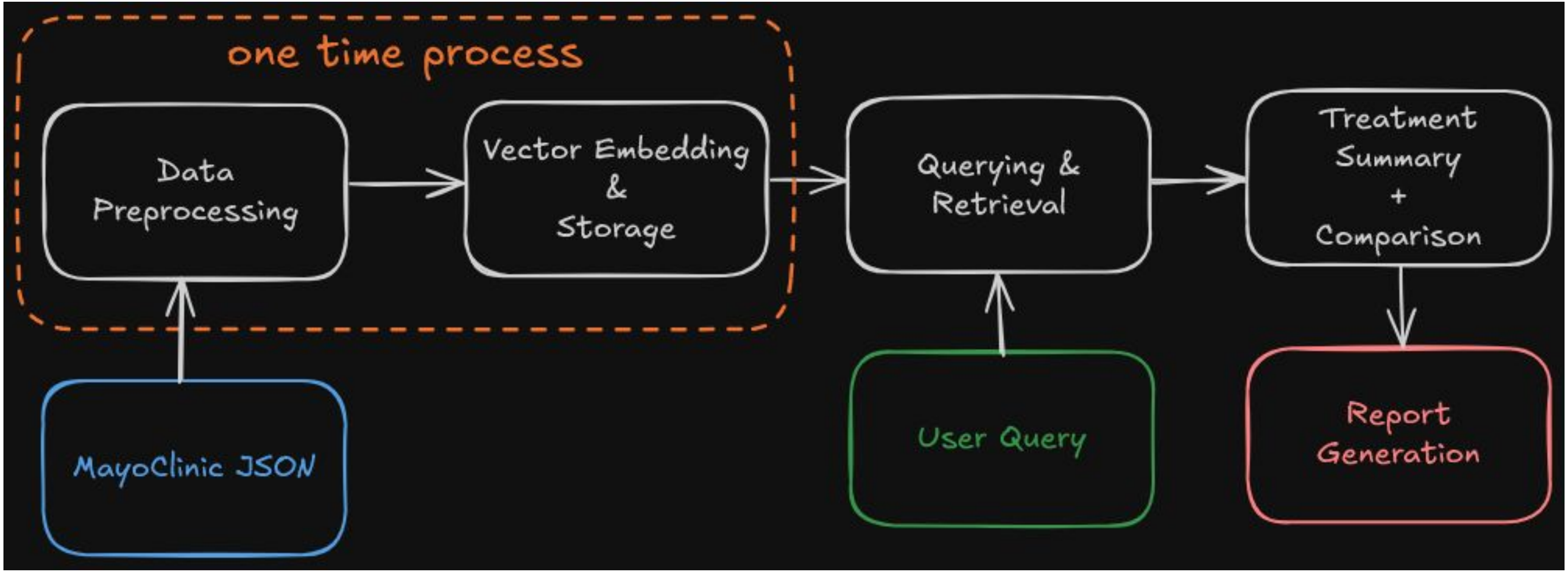
My work revolves around option 3.

Approaches

Project #1 - Medical Necessity Evaluation

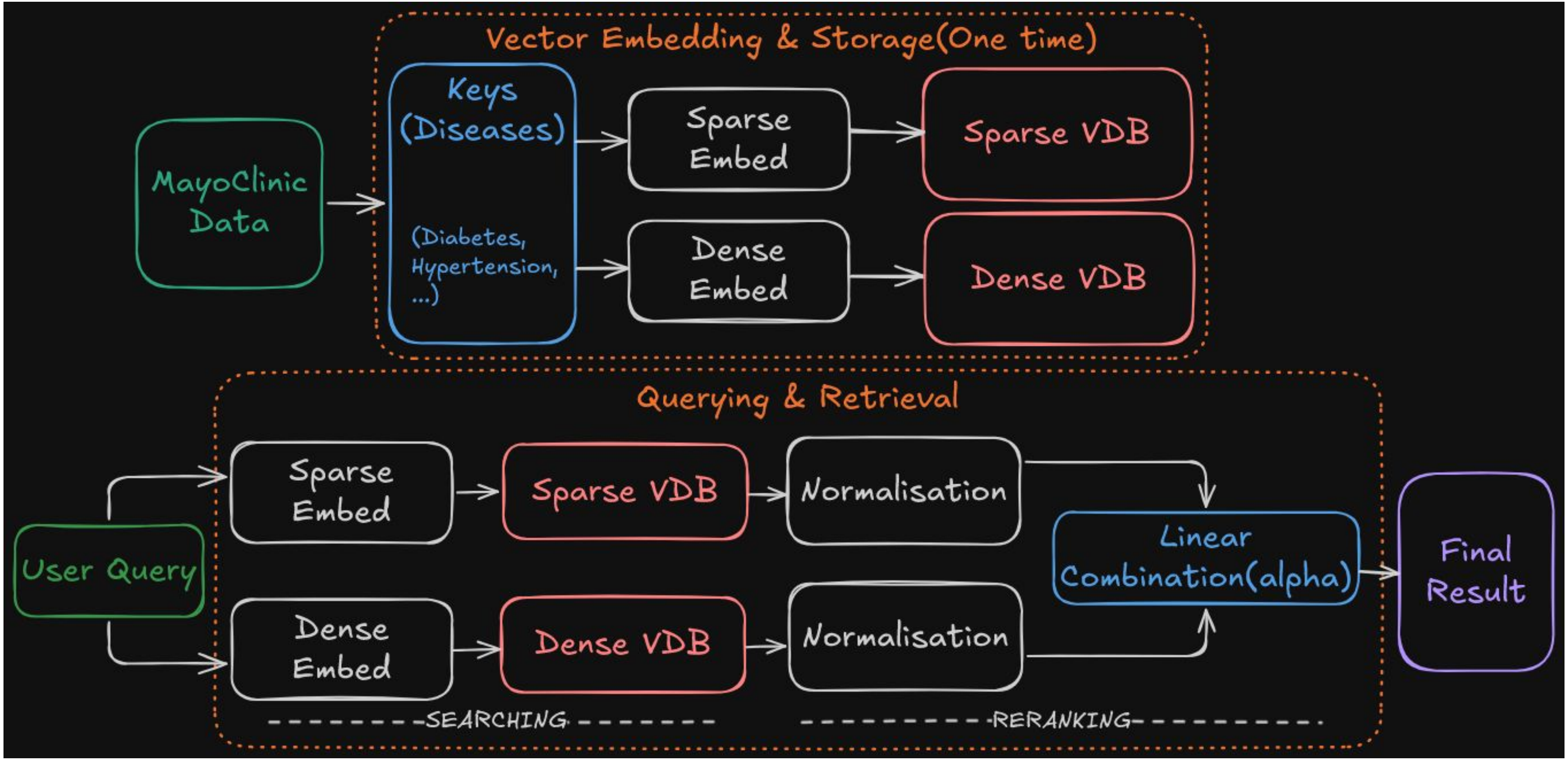
#1 - Semantic Search using Embedding Vectors

#2 - RAG Pipeline



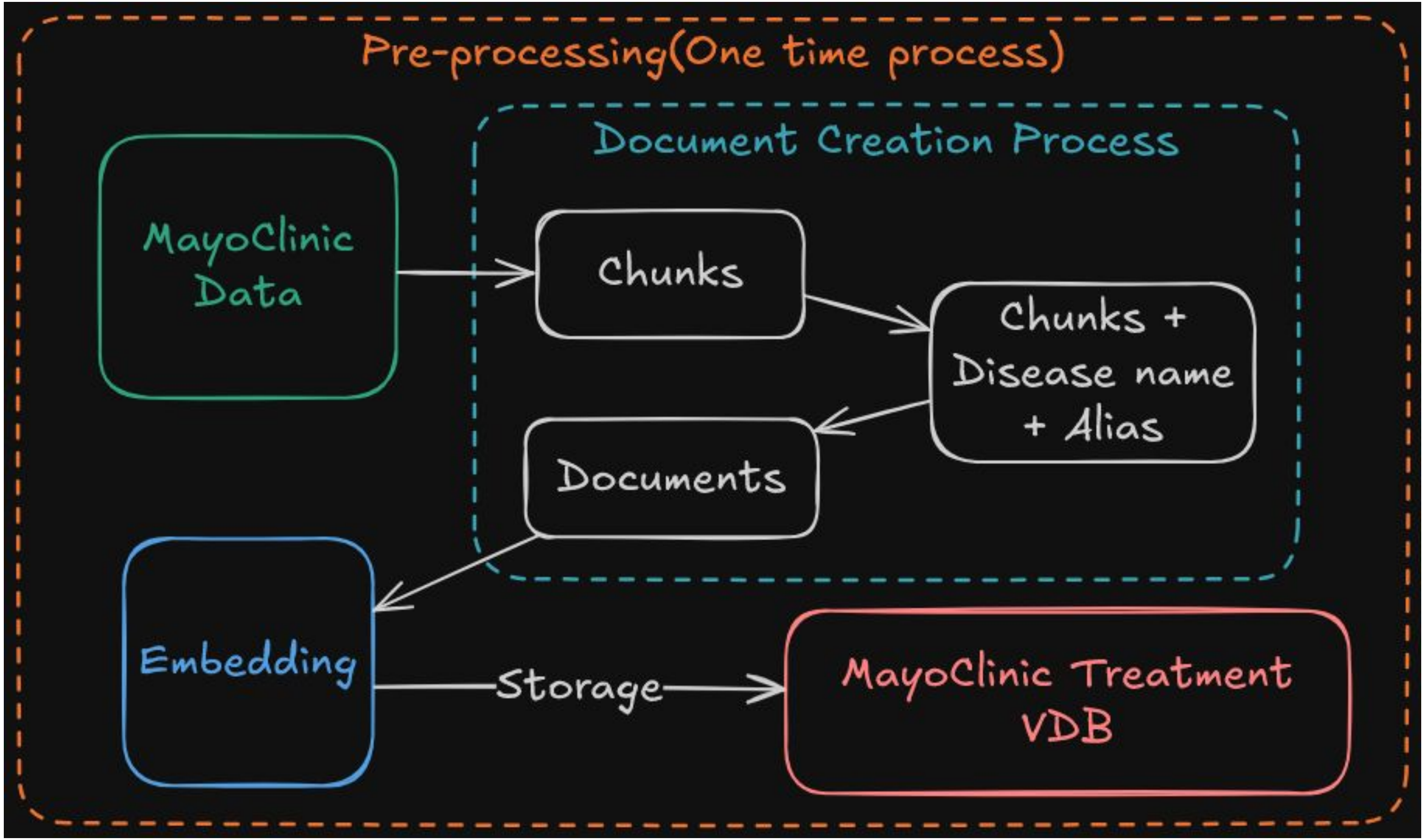
Approach #1 - Semantic Search using Embedding Vectors

Project #1 - Medical Necessity Evaluation

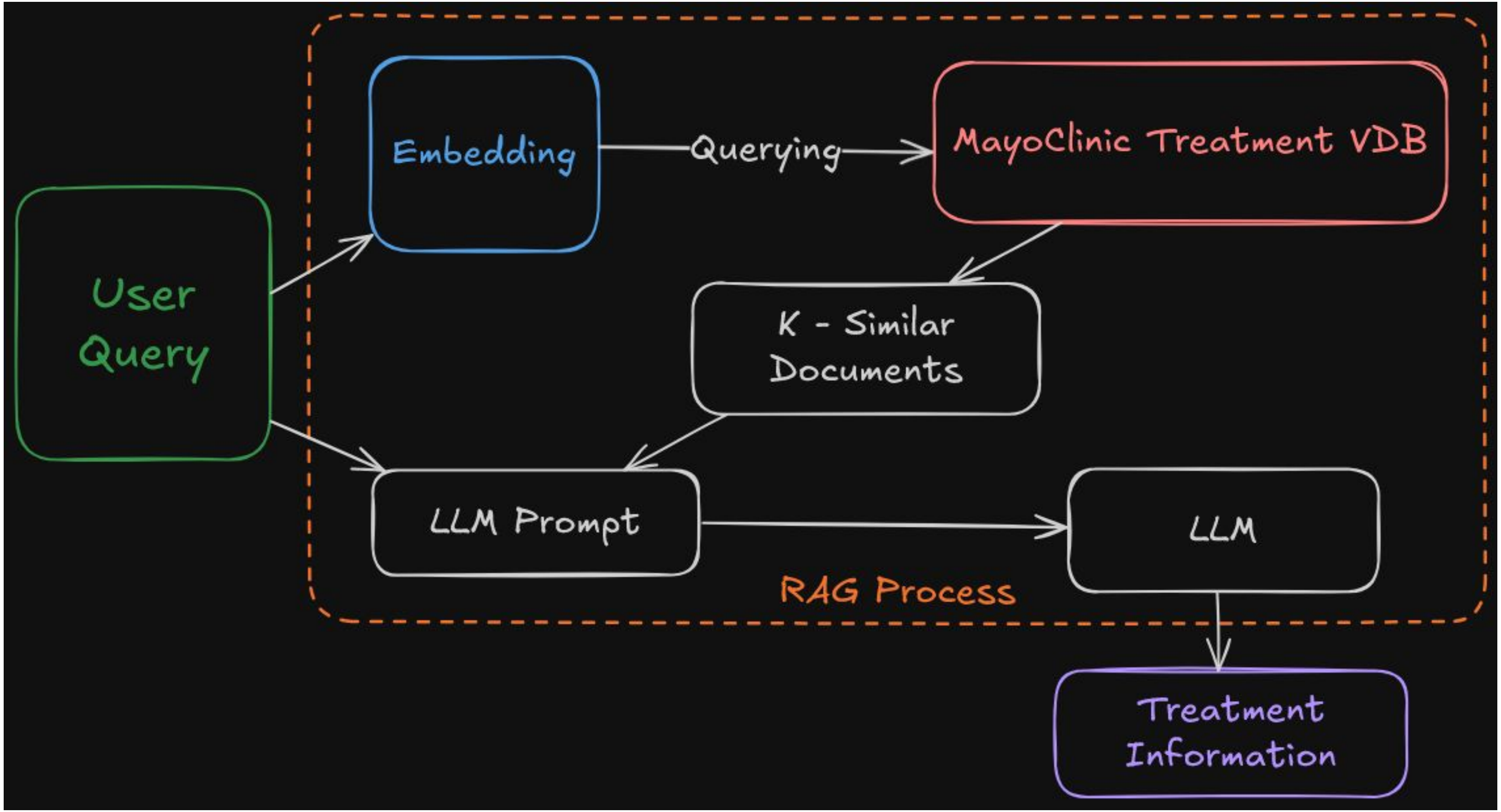


Approach #2 - RAG Pipeline

Project #1 - Medical Necessity Evaluation



Approach #2 - RAG Pipeline
Project #1 - Medical Necessity Evaluation



Implementation
Project #1 - Medical Necessity Evaluation

1. **MayoClinic Scraper**: Implemented a web-scraper for the MayoClinic website, extracts the links of every page, the overview, the symptoms, the risk-factors, the diagnosis and the treatment
2. **Embedding Vector Search**:
   a. Embedding Models:
      i. Google Text Embedding - text-embedding-004
      ii. AWS Titan Text Embeddings V2
      iii. Hugging Face Embeddings - neuml/pubmedbert-base-embeddings
      iv. Qdrant BM25
   b. LLMs:
      i. Google Gemini - models/gemini-2.5-flash
      ii. Amazon Nova Lite
      iii. Mistral 7B
3. **RAG Pipeline**:
   a. Embedding Models:
      i. Google Text Embedding model (text-embedding-004, used via Google AI Studio)
      ii. AWS Titan Text Embeddings V2 (amazon.titan-embed-text-v2:0, via AWS Bedrock)
      iii. Hugging Face Embedding - PubMedBert (neuml/pubmedbert-base-embeddings, Local)
      iv. Hugging Face Embedding - Bio_ClinicalBERT(emilyalsentzer/Bio_ClinicalBERT, Local)
   b. LLMs: same as Embedding Vector Search

Evaluation & Results

Project #1 - Medical Necessity Evaluation - Vector Embedding Search

✅Dialysis - CKD
✅Synovitis - Arthritis, rheumatoid
✅Left Shoulder Impingement - Frozen shoulder

⚠️❓ *Lumbar Fusion* - Spondylosis, cervical
⚠️ *Soft Tissue Injuries* - Sarcoma, soft tissue
⚠️ *Impingement Syndrome* - DiGeorge syndrome (22q11.2 deletion syndrome)
⚠️*Slip and Fall Accident* - Foot drop
⚠️ *Noninfectious Gastroenteritis* - Gastroenteritis, viral
⚠️ *Labral Tear* - Hip labral tear

- All obvious diseases and the ones present in the database get matched with little to no problems
- Too **generic**(Slip and Fall) or when they're too **niche**(non-infectious) and the dataset is lacking, it picks the closest, which we don't have a lot of control over
- **Keywords** like Syndrome, Abscess sometimes throw the model off
- **Acronyms** are a problem for the model to understand and match
- Goes too specific/too generic with few matches

Evaluation & Results

Project #1 - Medical Necessity Evaluation - RAG Pipeline

❓ *Lumbar Fusion*

- Fetches chunks that deal with lumbar fusion, from varied diseases, chances of incorrect information, but inclusive of the disease

❓ *Soft Tissue Injuries*

- Talks about using a brace, rest
- Lack of specific treatment(Medication, Processes)

❓ *Slip and Fall Accident*

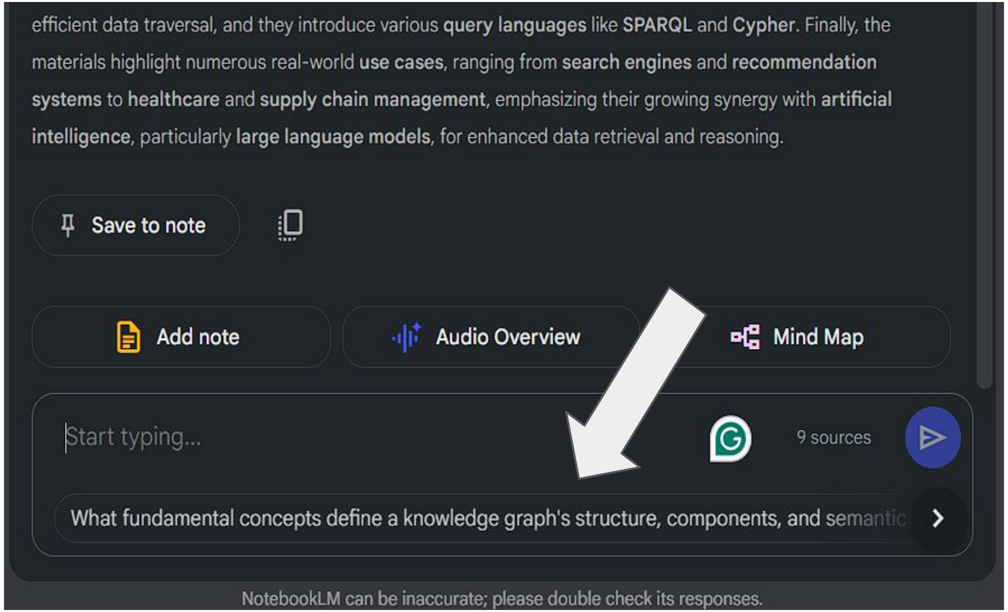
- Fetches all possible injuries and treatments(Broken Hip (Hip Fracture):
- Arm Fracture (Broken Arm), Leg Fracture (Broken Leg), Traumatic Brain Injury (TBI), Spinal Cord Injury (SCI), Anterior Cruciate Ligament (ACL) Injury…

Thoughts:

- Lack of specific treatment, generic solution to all problems, raises more flags than necessary.

Problem Statement, Hypothesis

Project #2 - Question Recommendation System



- **Problem Statement:** Process documents to generate and suggest questions in chat, relevant to the contexts of the documents
- **Hypothesis:** Using a knowledge graph to model relationships in the document would generate accurate and relevant questions

Approaches

Project #2 - Question Recommendation System

Approach #1 - Knowledge Graph(KG) based QA

- Attempted to create a KG and generate questions
- RAGAs - Framework that uses KGs for creation of QA Datasets

- Time taking, hard to implement, lack of resources

Approach #2 - DeepEval - LLM based QA

- Good customisation and filtering options for QA sets

- LLM Heavy, but very easy to implement

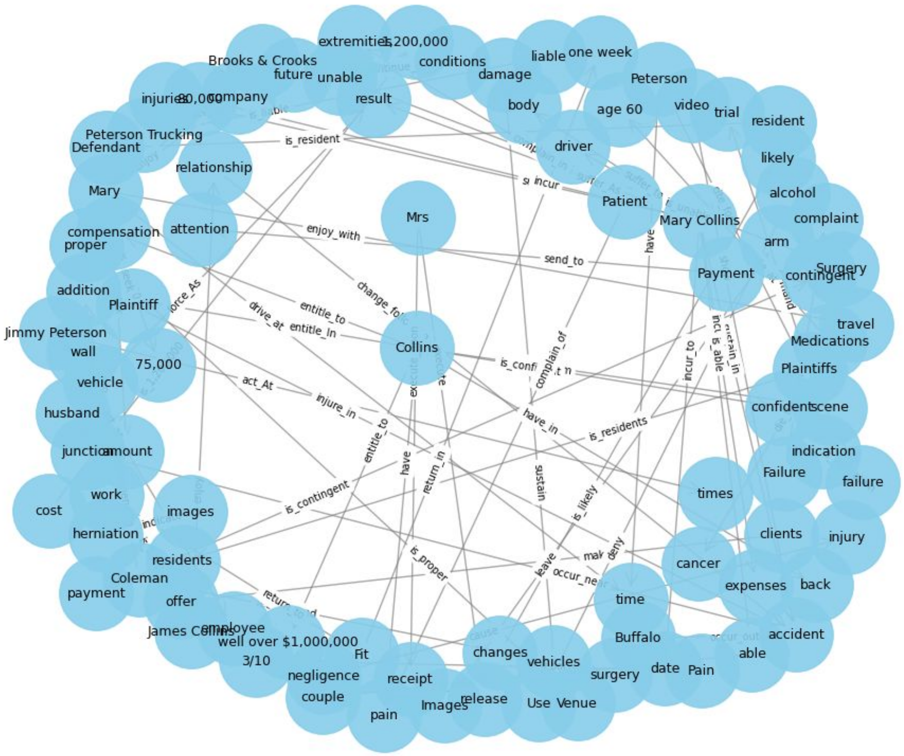
Approach #3 - Topic Modelling

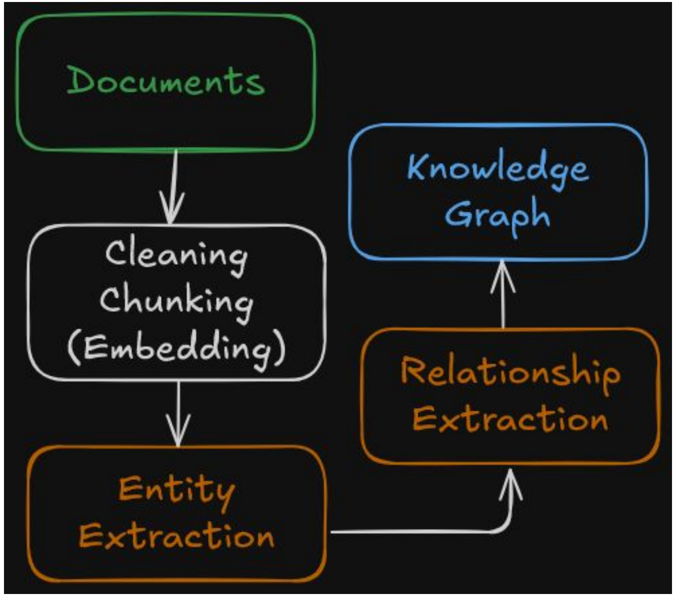
- Work in progress, so far, hallucination + good questions

- Simplest and easiest method
- Implementation from scratch - complete flexibility

Implementation - Knowledge Graph

Project #2 - Question Recommendation System



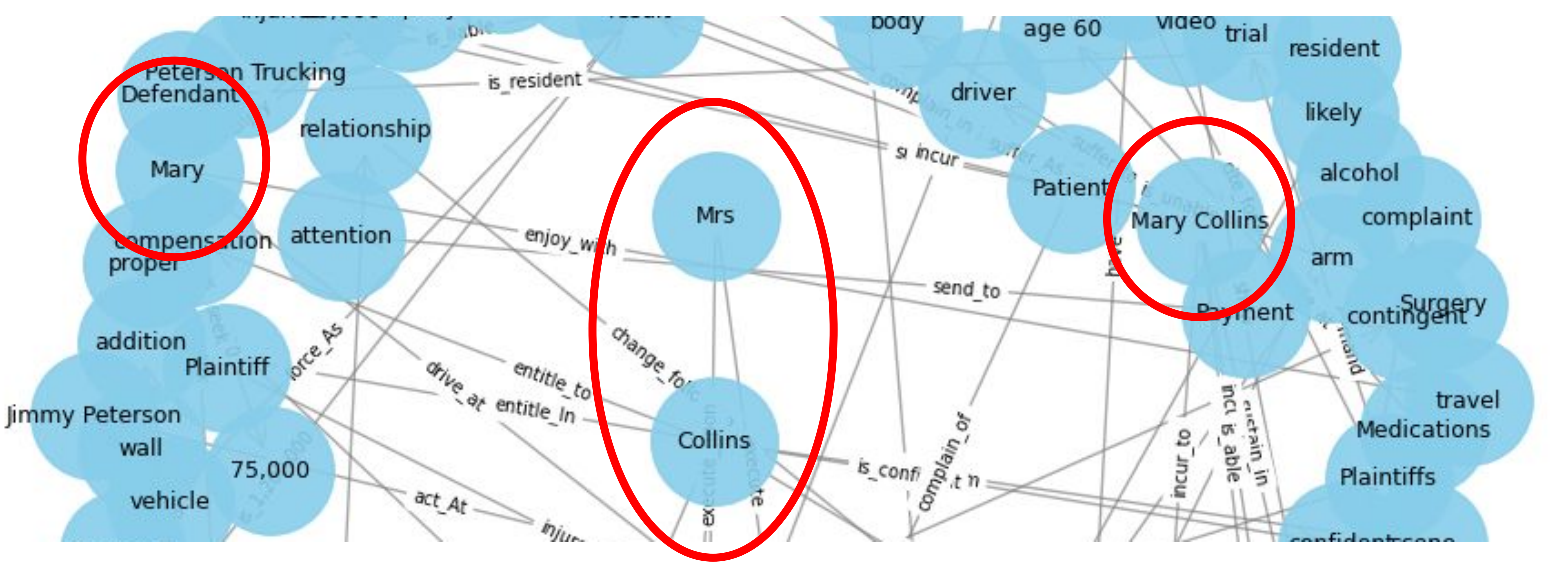


Entity Extraction:

- NER(NLP)
- LLMs

Relationship Extraction:

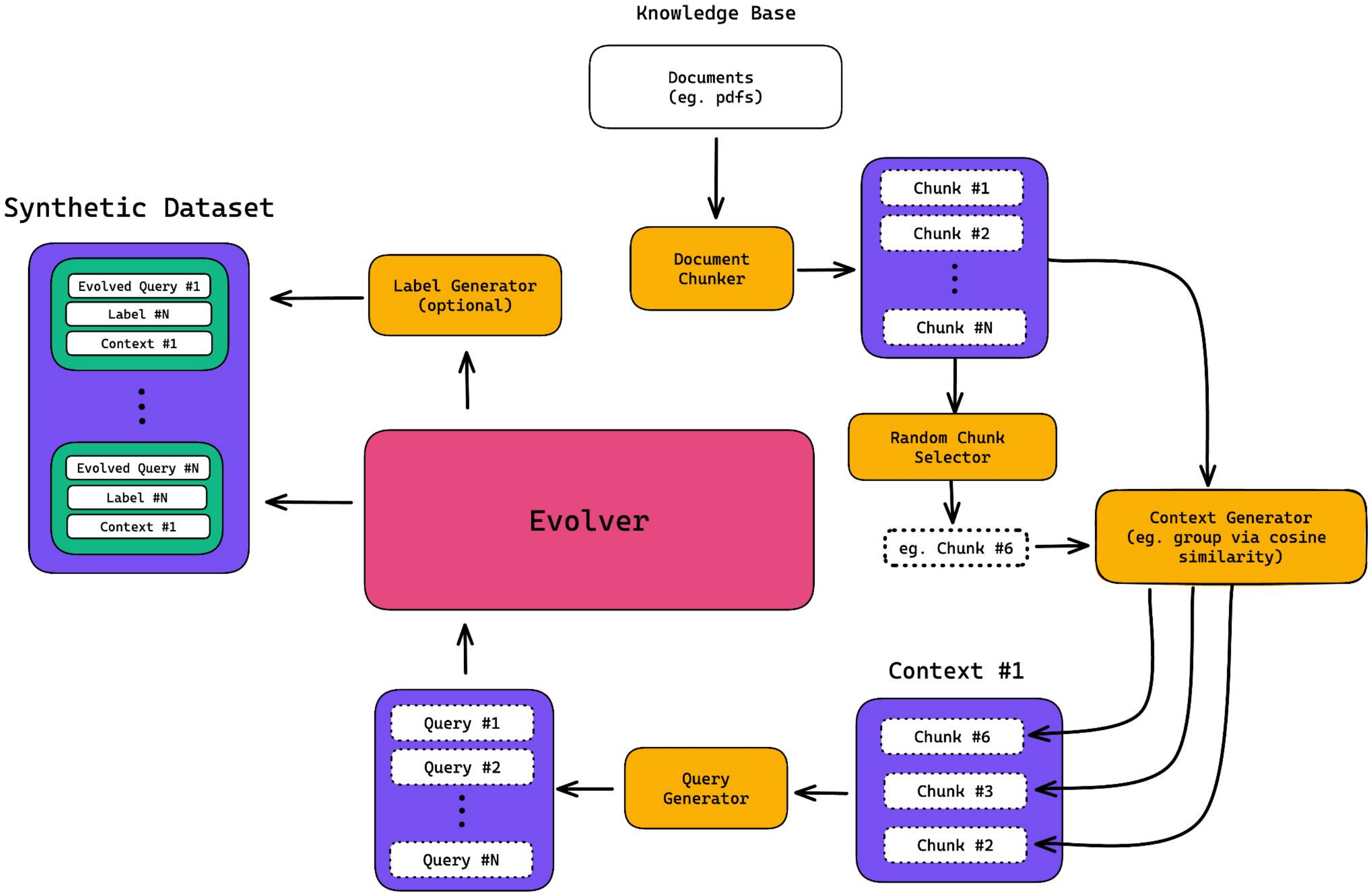
- Rule based(syntax)
- LLMs



- Relationship Extraction - relationships with high connectivity
- Ineffective rule based relationship extraction
- e.g., Ms. Alice and Mr. Bob travelled to Paris and to London, United Kingdom, for the purpose of attending a conference and attend meetings with prospective clients

- Moved onto RAGAs, KG based framework -> ran into similar issues
- Transforms - How it produces questions -> skipped half of the time

Implementation - DeepEval
Project #2 - Question Recommendation System



Implementation - DeepEval

Project #2 - Question Recommendation System

- **Evolutions**: Customise the type of QA sets:
  - REASONING: cause-effect
  - MULTICONTEXT: QA from >=1 contexts
  - CONCRETIZING: Only from context
  - CONSTRAINED: QA with limits to count/contexts
  - COMPARATIVE: Comparative QA
  - HYPOTHETICAL: Creative QA, what-if
  - IN_BREADTH: QA on the general theme, rogue
- **Filtering**: select a threshold - filtering out the responses & contexts below that threshold
- **Styling**: Influence the framing & wording of QA

Evaluation and Results

Project #2 - Question Recommendation System

RAGAs:

- Hallucinated, produced questions with slang, lost context

DeepEval:

- Why did Dr. Coleman recommend surgery for Mrs. Collins' lumbar pain?
- How do I calculate Mary Collins's total economic damages, including future care costs?
- What are the specifics of the April 3, 2024, Peterson Transport collision?
- Where was Mrs. Collins airlifted to after the accident, and who examined her in the emergency room?
- What is the process for assessing Mrs. Collins' compensation due to accident-induced injuries?

Topic Modelling:

- If Mrs. Peterson had been driving slower, could the accident have been avoided?
- Who was driving the truck in the incident?
- Is a person entitled to compensation for medical pain with an expense of $1,000,000?
- Is Dr. Collins related to Mrs. Coleman?
- Is James related to Collins or Mary?

Questions?

What’s Next?

Medical Necessity Evaluation

1. Trying different embedding models, better embeddings
2. Scraping different websites, cleveland clinic, better data -> lesser work for the models
3. Running these on more of our accounts, to get more testing data
4. Integrating into the workflow - independent of Medical Summary

Question Recommendation System

1. Topic Modelling seems promising in the sense that it’s easy, flexible, light on the LLM calls
2. DeepEval can produce decent results for the cost of time and LLM tokens
3. KG needs more work, time and effort to perfect and generalise that it works for any document

Recommendations made in previous presentation

Medical Necessity Evaluation

1. Creation of a knowledge graph for Medical Necessity
2. Classification of queries, to perform relevant searches

Question Recommendation System

1. Including ongoing chat context for question recommendation system

Thank You!