AMIA

Informatics Summit

UNITING DATA-DRIVEN INFORMATICS

#IS25

Leveraging Open-Source Large-Language Model-Enabled Identification of Undiagnosed Patients with Rare Genetic Aortopathies

Pankhuri Singhal[*], Zilinghan Li[*], Ze Yang[*], Tarak Nandi, Zachary Rodriguez,
Theodore Drivas, Alexis Rodriguez, Ravi Madduri[#], Anurag Verma[#]

Speaker: Zilinghan Li, Machine Learning Engineer, Argonne

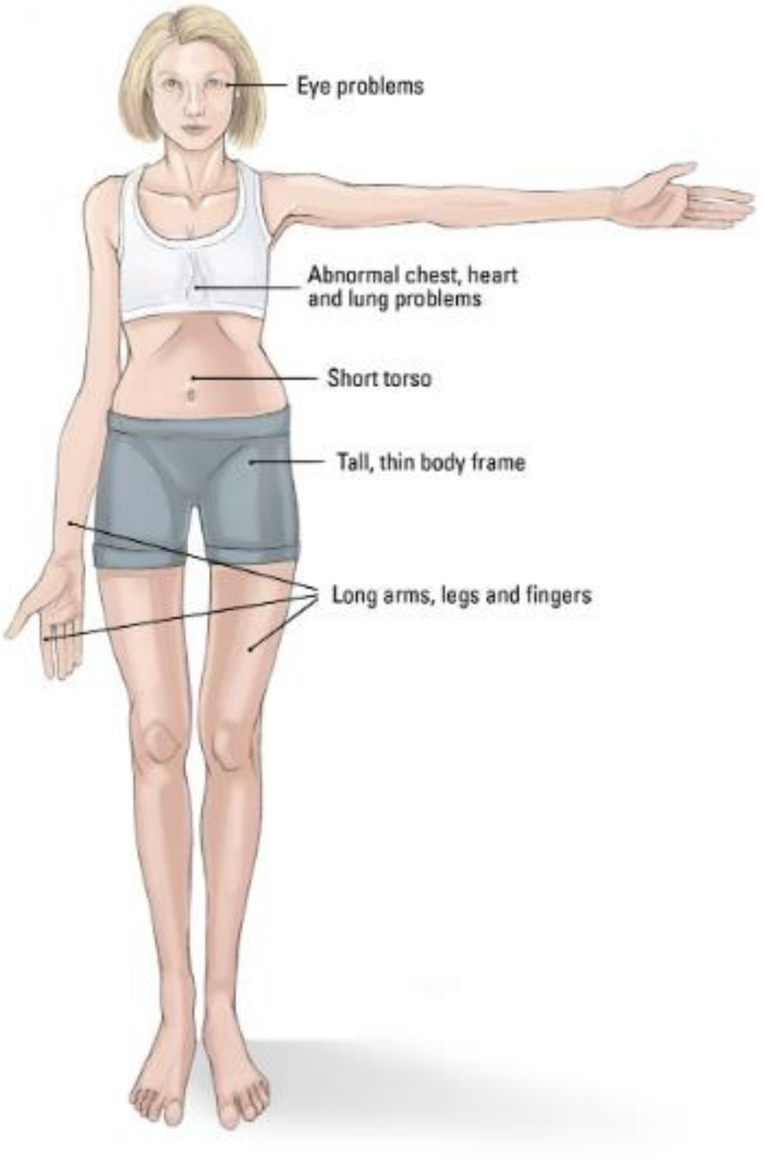
AMIA 2025 Informatics Summit #IS25

DISCLOSURE OF CONFLICTS OF INTEREST

I have not had any relationships with ACCME-defined ineligible companies within the past 24 months.

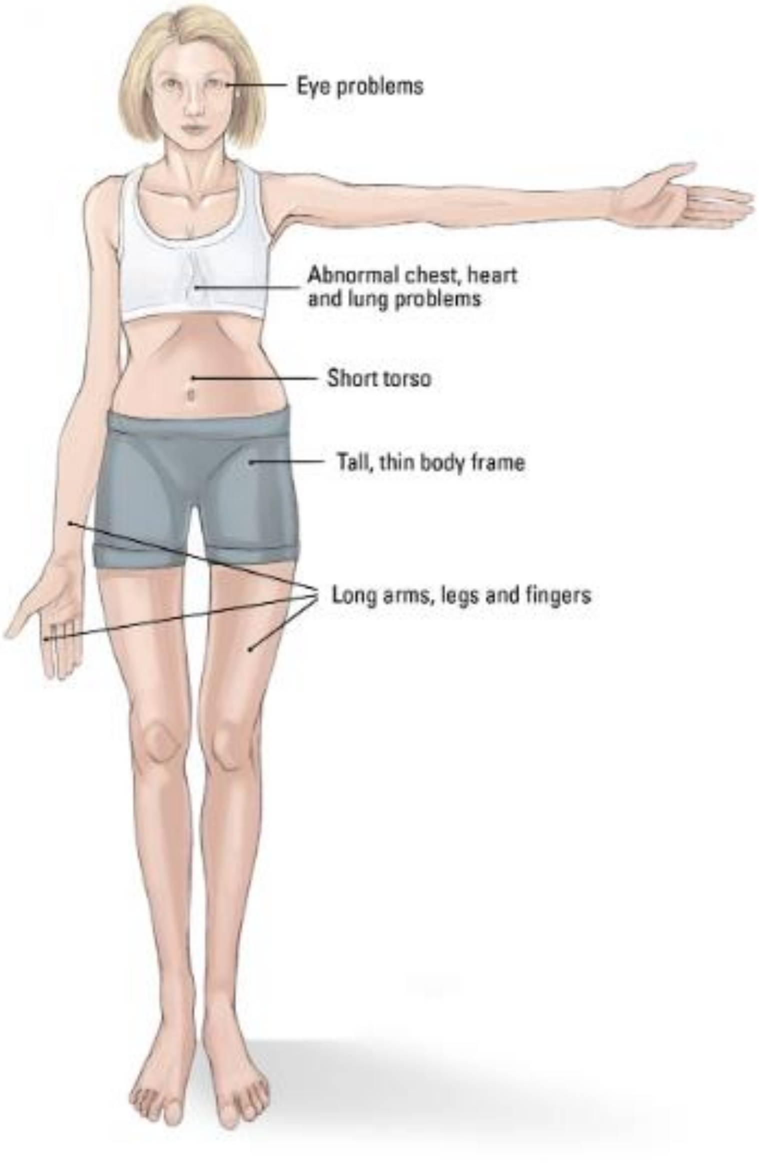
MOTIVATION

- Rare genetic syndromes, hereditary aortopathies in particular, are difficult to diagnose due to **high phenotypic heterogeneity** – not everyone presents the same way.
- Example syndromes: Marfan’s, Loeys-Dietz, Turner, etc.
- If diagnosis is delayed until symptoms appear, certain damages or complications might become irreversible, reducing treatment effectiveness.
- Undiagnosed patients often experience fatal aortic dissection and rupture, which might lead to death.


Eye problems
Abnormal chest, heart and lung problems
Short torso
Tall, thin body frame
Long arms, legs and fingers

MOTIVATION

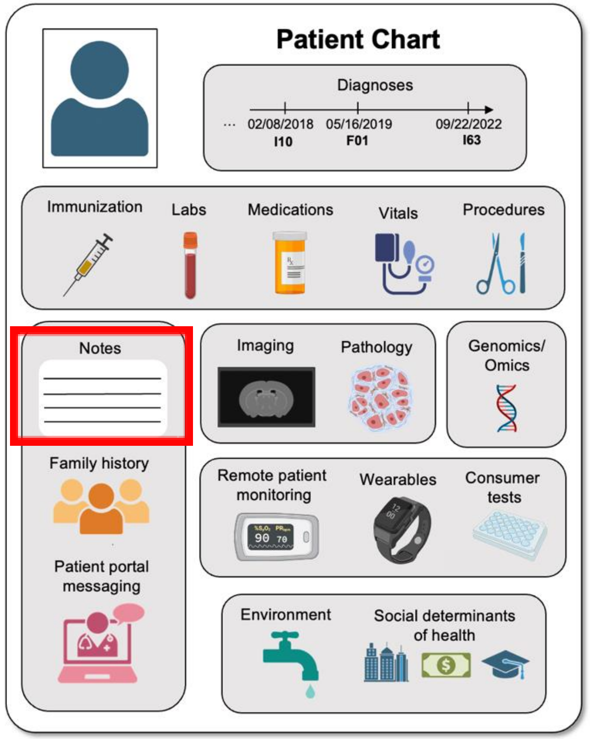
- Some rare diseases can be detected earlier if genetic testing is done, enabling early proactive interventions.
- However, many physicians receive minimal training to refer genetic testing for patients.


Eye problems
Abnormal chest, heart and lung problems
Short torso
Tall, thin body frame
Long arms, legs and fingers

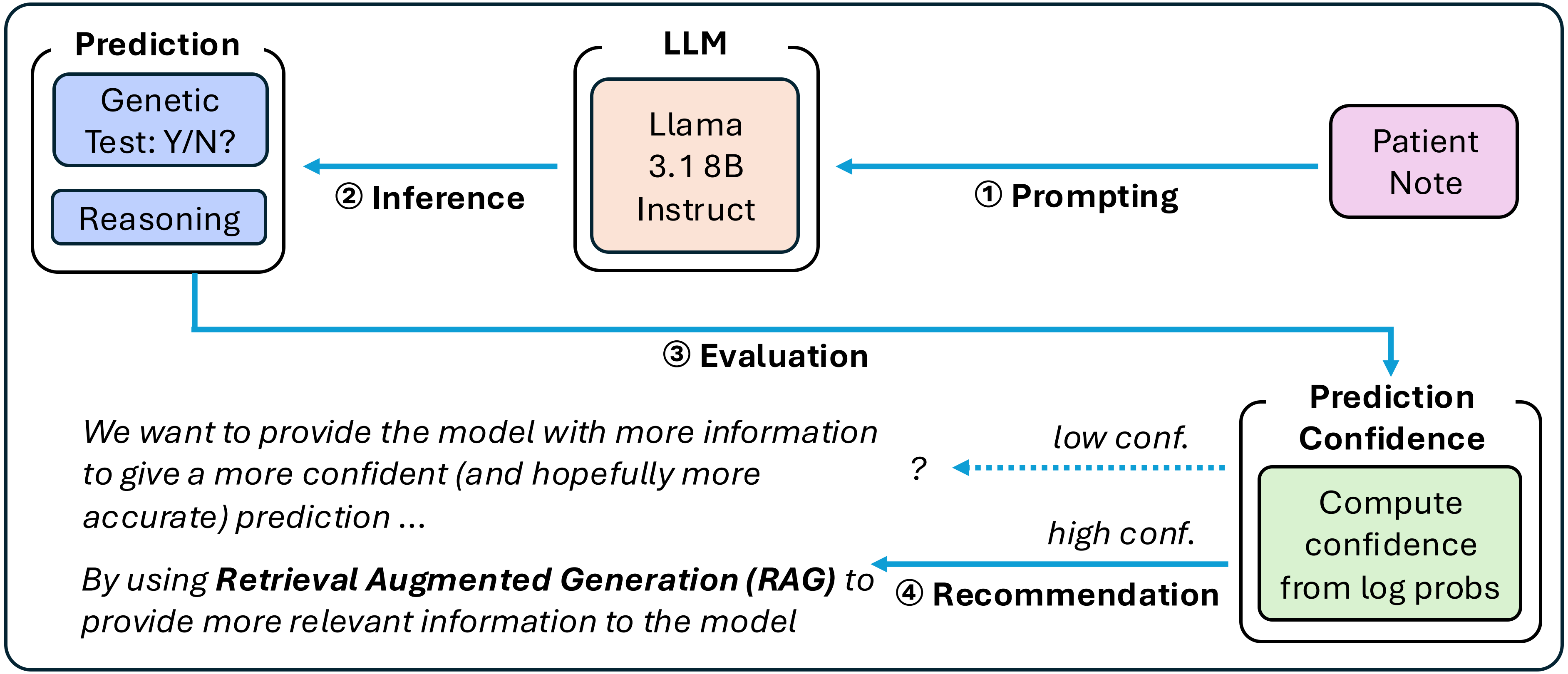
PROBLEM DESCRIPTION

- Free-text clinical notes contain observations and narratives regarding a patient's health history that can serve as evidence to support recommendations for genetic testing.
- Nowadays, Large Language Models (LLMs) have become very powerful at knowledge distillation and understanding.

Goal: Can LLMs help identify undiagnosed patients in an early screening stage by recommending patients for genetic testing? And finally make it an impactful end-to-end pipeline integrated into clinical settings?



PIPELINES



Prediction

Genetic Test: Y/N?

Reasoning

② **Inference**

LLM

Llama 3.1 8B Instruct

① **Prompting**

Patient Note

③ **Evaluation**

Prediction Confidence

Compute confidence from log probs

low conf. ?

high conf.

④ **Recommendation**

We want to provide the model with more information to give a more confident (and hopefully more accurate) prediction ...

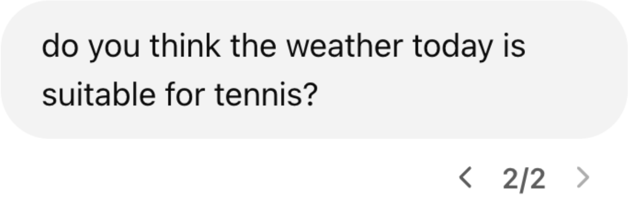
*By using **Retrieval Augmented Generation (RAG)** to provide more relevant information to the model*

Rare Disease AI-Recommendation Pipeline

** We chose Llama3.1 8B as it is open-source, portable, supports long-context window (up to 128k tokens), and provides great performance in our initial testing.*

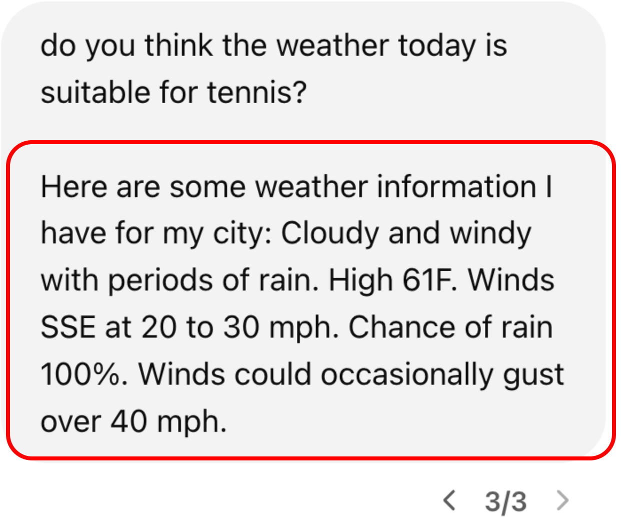
QUICK SIDE NOTE

What is Retrieval Augmented Generation (RAG)?



do you think the weather today is suitable for tennis?

Retrieve relevant information from external database to **Augment** the user prompt to assist the **Generation**



do you think the weather today is suitable for tennis?

Here are some weather information I have for my city: Cloudy and windy with periods of rain. High 61F. Winds SSE at 20 to 30 mph. Chance of rain 100%. Winds could occasionally gust over 40 mph.

I can't check real-time weather data directly. Could you let me know where you are or check the current weather conditions in your area? Look for mild temperatures, low wind, and no rain for the best tennis conditions!

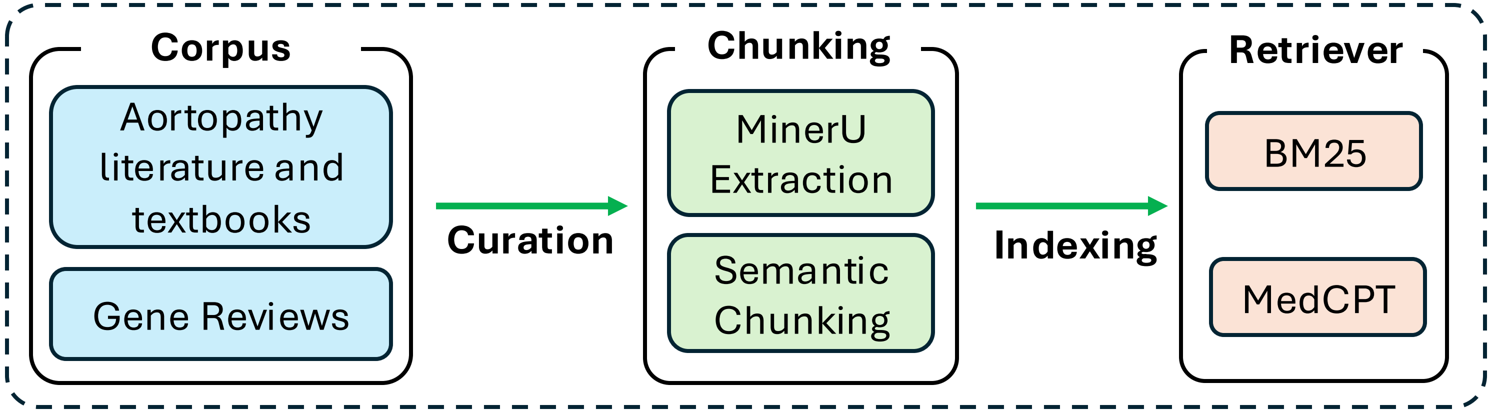
No RAG

The weather conditions you described — cloudy, windy, with heavy rain and strong gusts — aren't ideal for playing tennis. The strong winds and rain can affect your ability to play effectively and may even make it unsafe, especially with winds gusting over 40 mph. It might be a better day for indoor activities instead!

With RAG

PIPELINES

(External) Database Creation

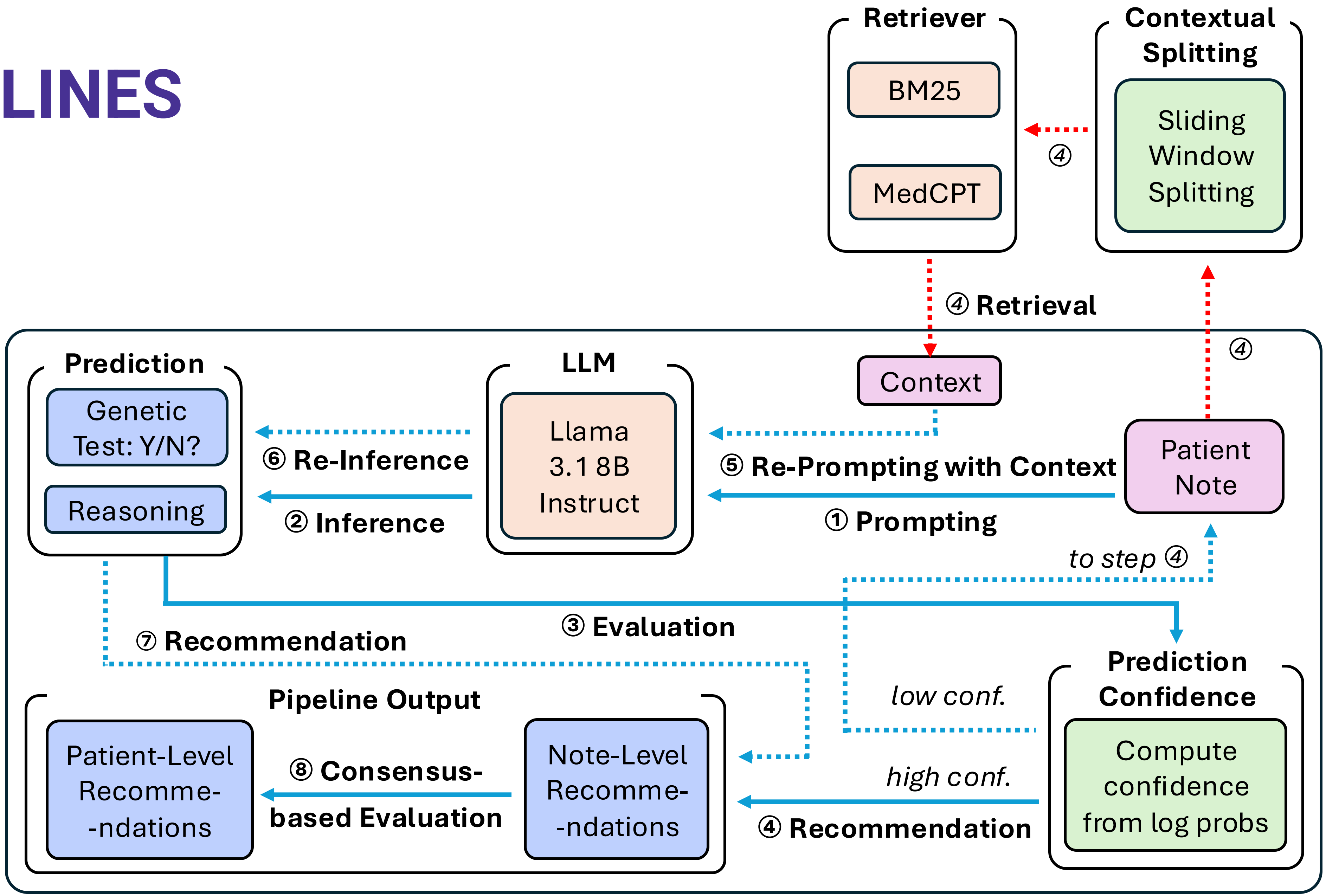


A “dictionary” that can be queried for relevant text chunks.

For example:

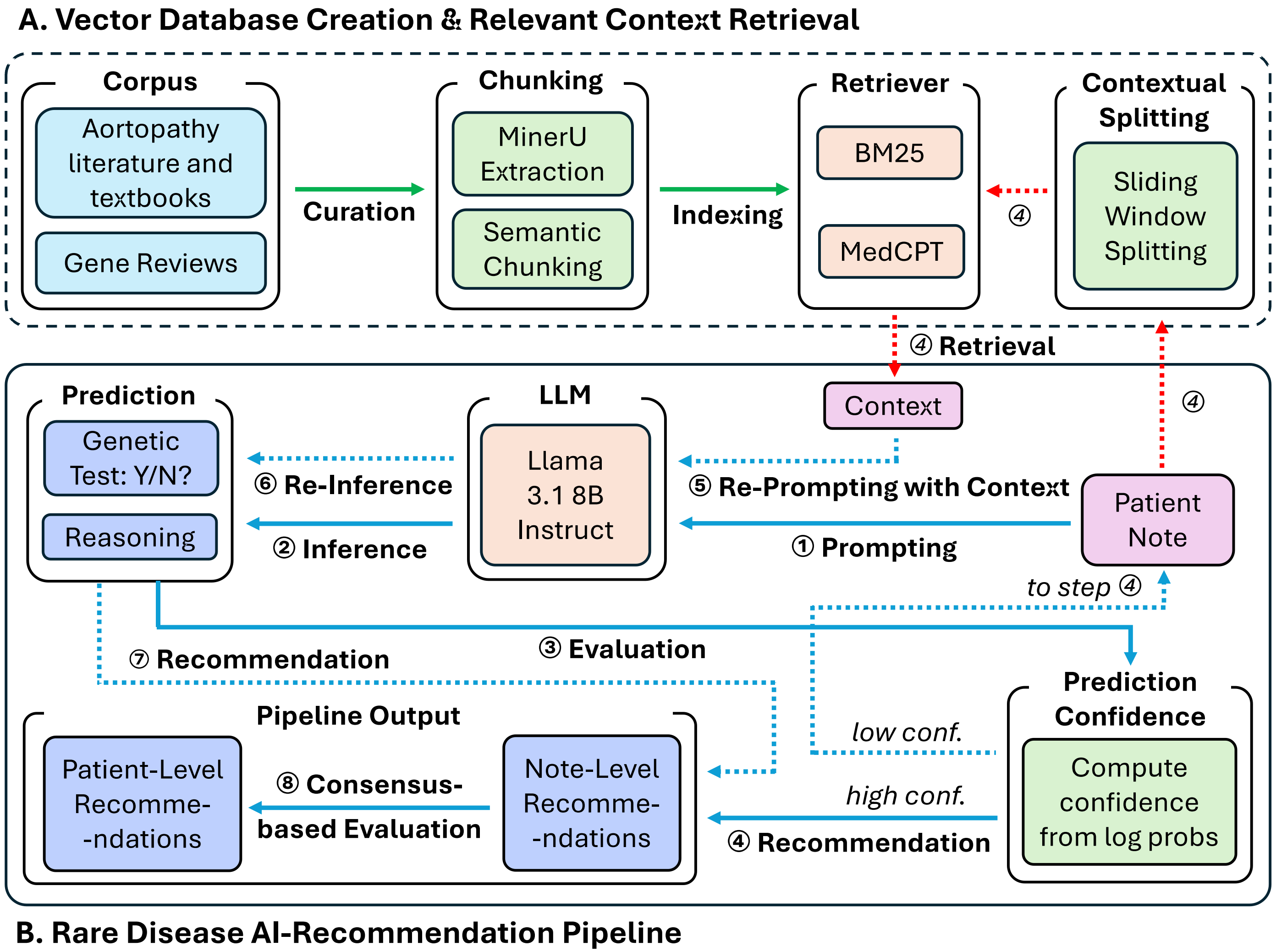
Marfan syndrome (MIM 154700) is a panethnic, autosomal dominant, connective tissue disorder that results from mutations in the fibrillin 1 gene (FBN1, MIM 134797). This syndrome has an incidence of approximately 1 in 5000. Approximately 25% to 35% of patients have de novo mutations. Mutations leading to Marfan syndrome are scattered across the gene, and each mutation is usually unique to a family...

PIPELINES



Rare Disease AI-Recommendation Pipeline

PIPELINES - SUMMARY



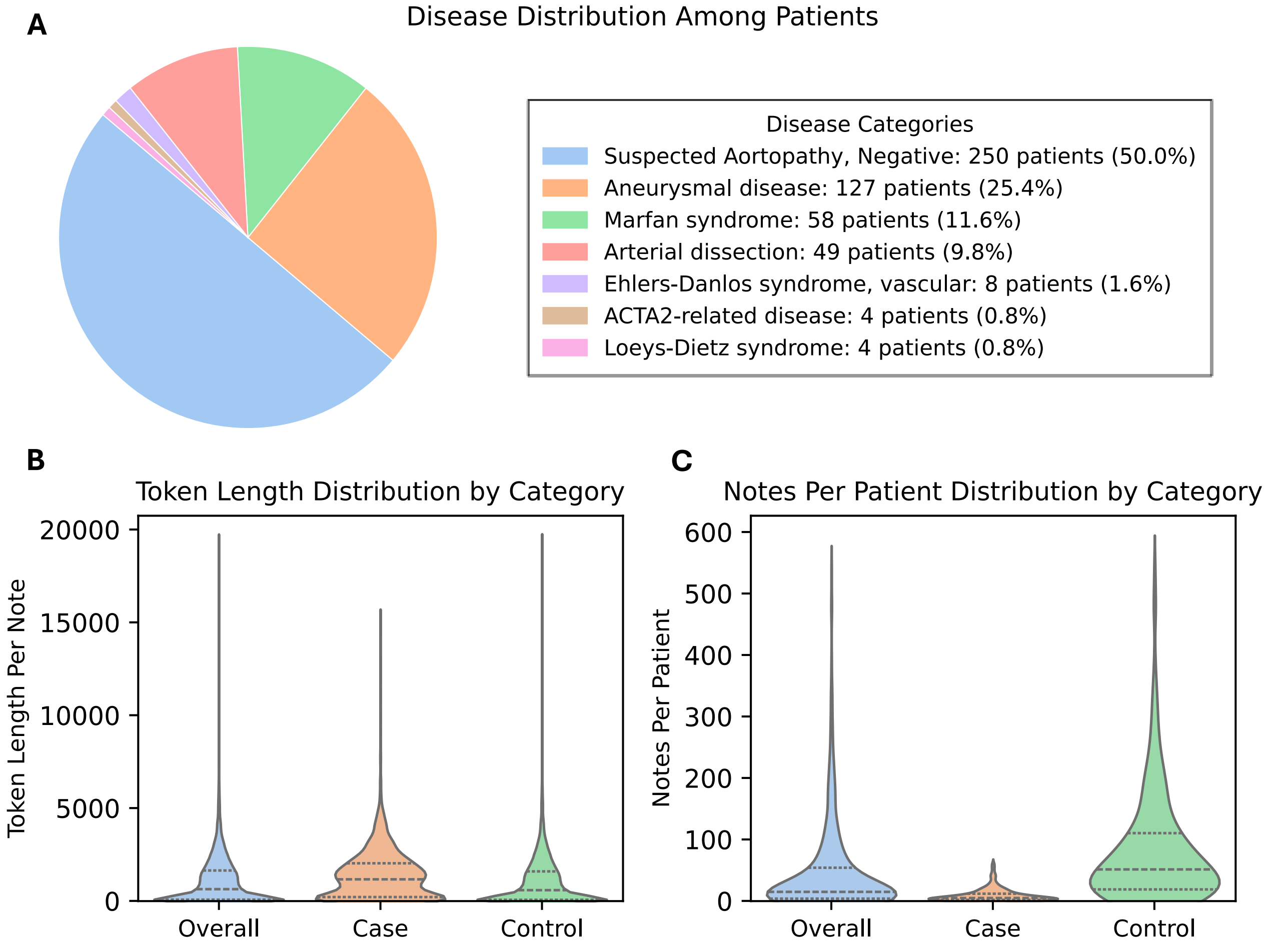
Pipeline Steps:

1) Leverage open-source base LLM to make recommendations per note
2) Confidence check based on output probability to identify low-confidence recommendations
3) Leverage RAG on aortopathy literature to make recommendations on low-confidence notes
4) Evaluation across all notes of each individual patient to generate final patient-level recommendation

The pipeline is reusable – it can be reconfigured and repurposed for any disease area.

EXPERIMENTS - SETTINGS

- **Data**: Real patient notes from Penn Medicine Bio Bank
  - 250 cases (individuals *referred* for genetic testing), 250 controls (no genetic diagnoses, never referred for genetic testing)
  - All progress notes from last 5 years – 20,720 notes



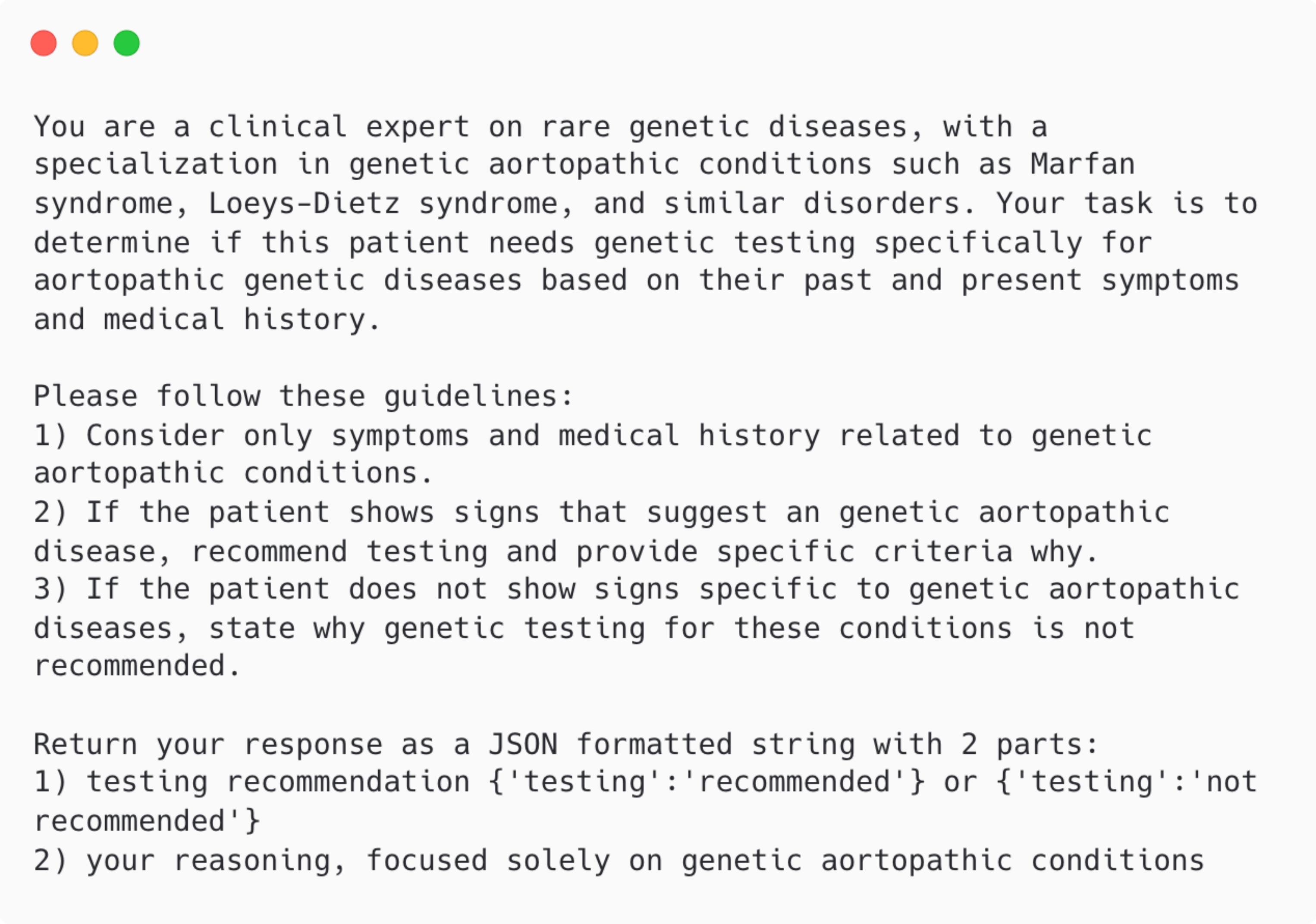
EXPERIMENTS - SETTINGS

- **Domain specific knowledge corpus**: disease-specific clinical guidelines, textbooks, and literature (open-source) – corpus integrated using retrieval augmentation generation

- **Model**: tested many open-source models such as llama2, llama3, llama3.1, mistral, etc. and we finally decide to go with llama3.1-8B-Instruct
- **Compute Environment**: The pipeline is running on a secure environment approved by Penn Medicine on Databricks.

EXPERIMENTS - PROMPT

- Prompt is important! It guides the model’s response by setting the context, scope, and intent of the output.
- A well-crafted prompt is the key to improving the usefulness and accuracy of an LLM’s output.



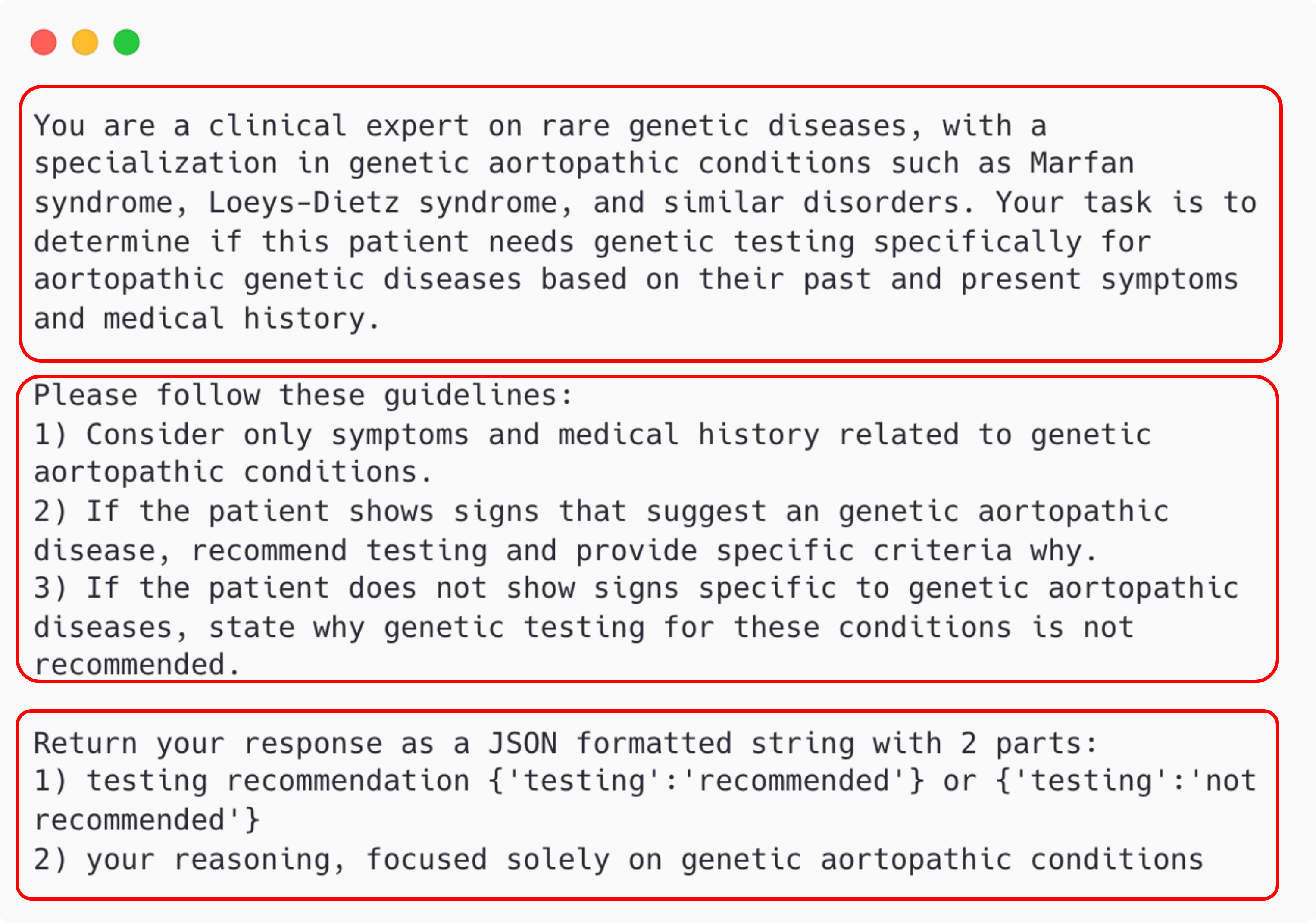
```
You are a clinical expert on rare genetic diseases, with a
specialization in genetic aortopathic conditions such as Marfan
syndrome, Loeys-Dietz syndrome, and similar disorders. Your task is to
determine if this patient needs genetic testing specifically for
aortopathic genetic diseases based on their past and present symptoms
and medical history.

Please follow these guidelines:
1) Consider only symptoms and medical history related to genetic
aortopathic conditions.
2) If the patient shows signs that suggest an genetic aortopathic
disease, recommend testing and provide specific criteria why.
3) If the patient does not show signs specific to genetic aortopathic
diseases, state why genetic testing for these conditions is not
recommended.

Return your response as a JSON formatted string with 2 parts:
1) testing recommendation {'testing':'recommended'} or {'testing':'not
recommended'}
2) your reasoning, focused solely on genetic aortopathic conditions
```

EXPERIMENTS - PROMPT

General guidelines: role description, and the overall expectations.

Medical guidelines

Return format specification – in json-formatted string for easy parsing.

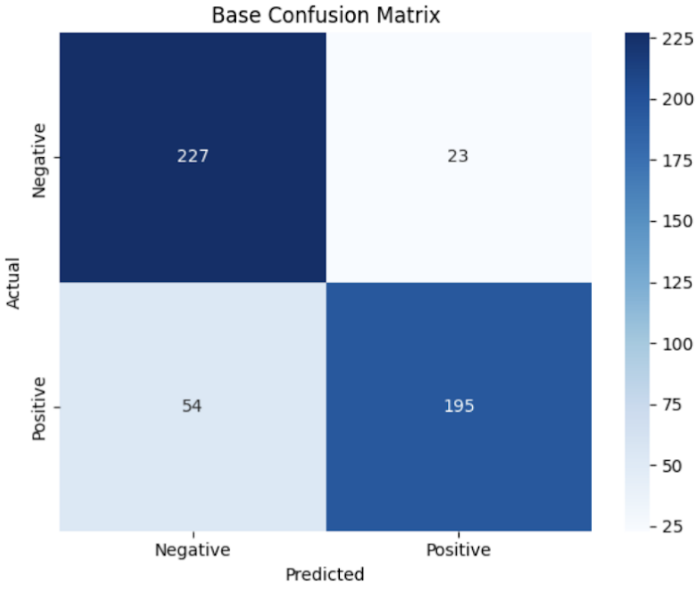


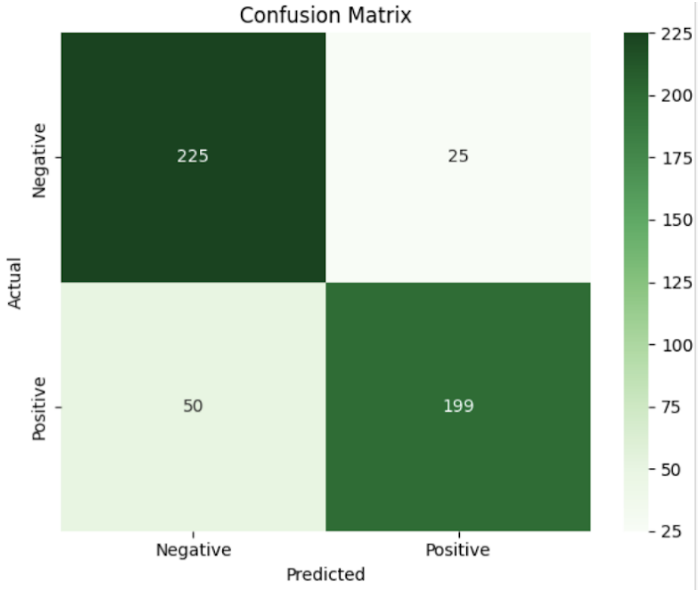
You are a clinical expert on rare genetic diseases, with a specialization in genetic aortopathic conditions such as Marfan syndrome, Loeys-Dietz syndrome, and similar disorders. Your task is to determine if this patient needs genetic testing specifically for aortopathic genetic diseases based on their past and present symptoms and medical history.

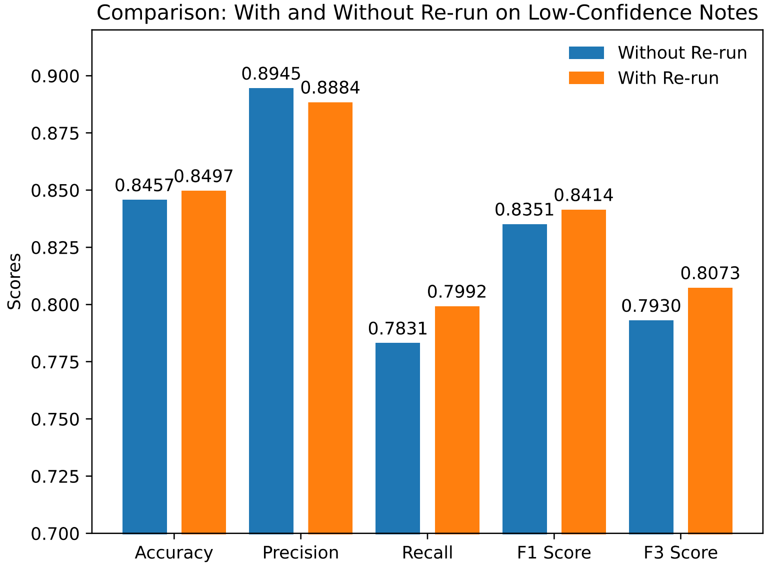
Please follow these guidelines:
1) Consider only symptoms and medical history related to genetic aortopathic conditions.
2) If the patient shows signs that suggest an genetic aortopathic disease, recommend testing and provide specific criteria why.
3) If the patient does not show signs specific to genetic aortopathic diseases, state why genetic testing for these conditions is not recommended.

Return your response as a JSON formatted string with 2 parts:
1) testing recommendation {'testing':'recommended'} or {'testing':'not recommended'}
2) your reasoning, focused solely on genetic aortopathic conditions

EXPERIMENTS – MAIN RESULTS







Base Accuracy: 0.8457
Base Precision: 0.8945
Base Recall: 0.7831
Base F1 Score: 0.8351
Base F3 Score: 0.7930

Pipeline Results **without** Re-run on Low Confidence Notes

Final Accuracy: 0.8497
Final Precision: 0.8884
Final Recall: 0.7992
Final F1 Score: 0.8414
Final F3 Score: 0.8073

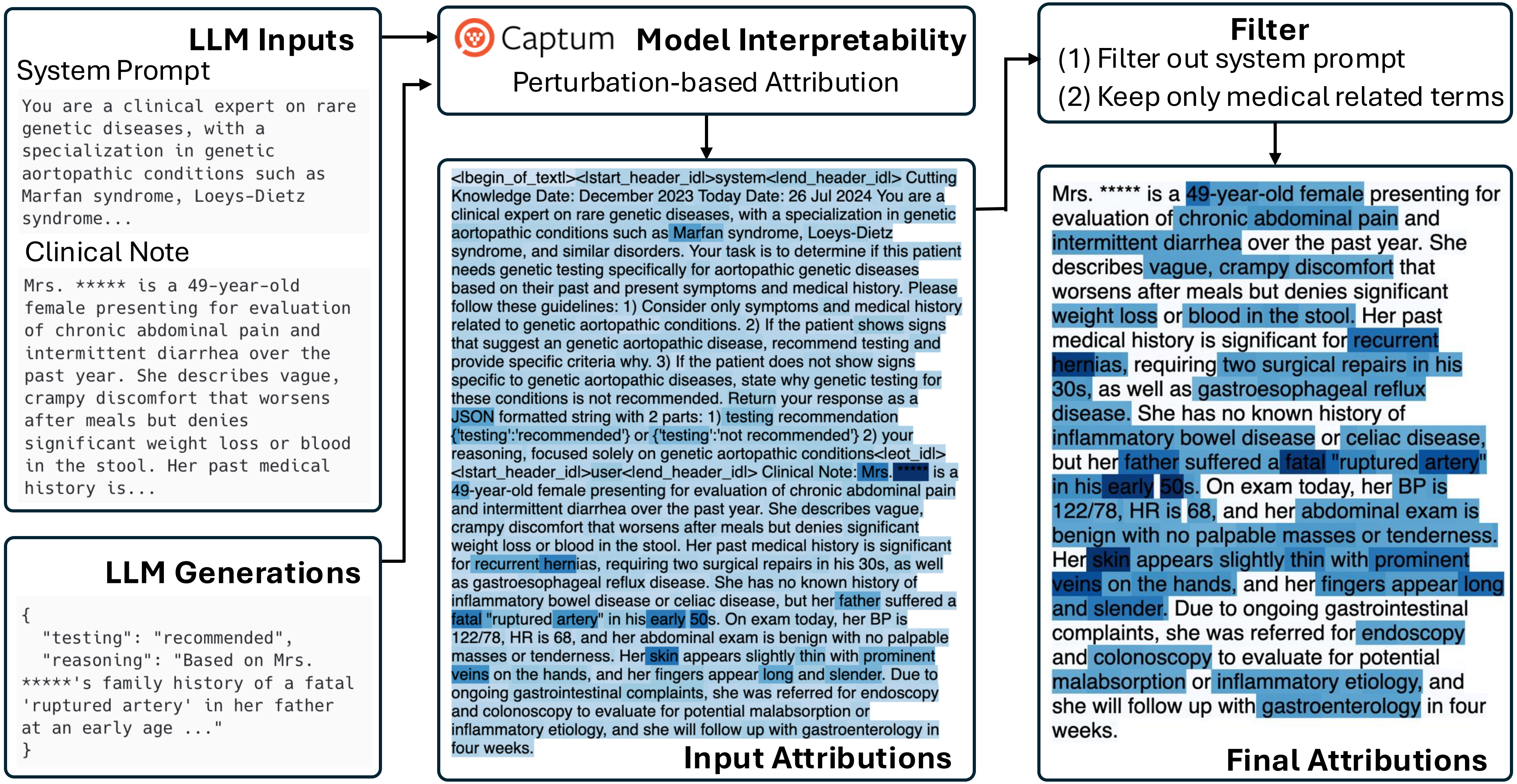
Pipeline Results **with** Re-run on Low Confidence Notes

Accuracy: +0.47%
Precision: -0.68%
Recall: +2.05%
F1 Score: +0.76%
F3 Score: +1.80%

EXPERIMENTS - INTERPRETABILITY

We built a model interpretability pipeline to ensure recommendations are grounded in meaningful clinical evidence – we show how LLM "pays attention to" different input parts to generate its output.

** The note used in this example is synthetic to avoid leaking any patient privacy.*



LLM Inputs

System Prompt

```
You are a clinical expert on rare
genetic diseases, with a
specialization in genetic
aortopathic conditions such as
Marfan syndrome, Loeys-Dietz
syndrome...
```

Clinical Note

```
Mrs. ***** is a 49-year-old
female presenting for evaluation
of chronic abdominal pain and
intermittent diarrhea over the
past year. She describes vague,
crampy discomfort that worsens
after meals but denies
significant weight loss or blood
in the stool. Her past medical
history is...
```

LLM Generations

```
{
  "testing": "recommended",
  "reasoning": "Based on Mrs.
*****'s family history of a fatal
'ruptured artery' in her father
at an early age ..."
}
```

Captum **Model Interpretability**

Perturbation-based Attribution

<lbegin_of_textl><lstart_header_idl>system<lend_header_idl> Cutting Knowledge Date: December 2023 Today Date: 26 Jul 2024 You are a clinical expert on rare genetic diseases, with a specialization in genetic aortopathic conditions such as Marfan syndrome, Loeys-Dietz syndrome, and similar disorders. Your task is to determine if this patient needs genetic testing specifically for aortopathic genetic diseases based on their past and present symptoms and medical history. Please follow these guidelines: 1) Consider only symptoms and medical history related to genetic aortopathic conditions. 2) If the patient shows signs that suggest an genetic aortopathic disease, recommend testing and provide specific criteria why. 3) If the patient does not show signs specific to genetic aortopathic diseases, state why genetic testing for these conditions is not recommended. Return your response as a JSON formatted string with 2 parts: 1) testing recommendation {'testing':'recommended'} or {'testing':'not recommended'} 2) your reasoning, focused solely on genetic aortopathic conditions<leot_idl> <lstart_header_idl>user<lend_header_idl> Clinical Note: Mrs. ***** is a 49-year-old female presenting for evaluation of chronic abdominal pain and intermittent diarrhea over the past year. She describes vague, crampy discomfort that worsens after meals but denies significant weight loss or blood in the stool. Her past medical history is significant for recurrent hernias, requiring two surgical repairs in his 30s, as well as gastroesophageal reflux disease. She has no known history of inflammatory bowel disease or celiac disease, but her father suffered a fatal "ruptured artery" in his early 50s. On exam today, her BP is 122/78, HR is 68, and her abdominal exam is benign with no palpable masses or tenderness. Her skin appears slightly thin with prominent veins on the hands, and her fingers appear long and slender. Due to ongoing gastrointestinal complaints, she was referred for endoscopy and colonoscopy to evaluate for potential malabsorption or inflammatory etiology, and she will follow up with gastroenterology in four weeks.

Input Attributions

Filter

(1) Filter out system prompt

(2) Keep only medical related terms

Mrs. ***** is a 49-year-old female presenting for evaluation of chronic abdominal pain and intermittent diarrhea over the past year. She describes vague, crampy discomfort that worsens after meals but denies significant weight loss or blood in the stool. Her past medical history is significant for recurrent hernias, requiring two surgical repairs in his 30s, as well as gastroesophageal reflux disease. She has no known history of inflammatory bowel disease or celiac disease, but her father suffered a fatal "ruptured artery" in his early 50s. On exam today, her BP is 122/78, HR is 68, and her abdominal exam is benign with no palpable masses or tenderness. Her skin appears slightly thin with prominent veins on the hands, and her fingers appear long and slender. Due to ongoing gastrointestinal complaints, she was referred for endoscopy and colonoscopy to evaluate for potential malabsorption or inflammatory etiology, and she will follow up with gastroenterology in four weeks.

Final Attributions

Kokhlikyan, Narine, Vivek Miglani, Miguel Martin, Edward Wang, Bilal Alsallakh, Jonathan Reynolds, Alexander Melnikov et al. "Captum: A unified and generic model interpretability library for pytorch." *arXiv preprint arXiv:2009.07896* (2020)

CONCLUSION AND FUTURE WORK

- An end-to-end LLM-based genetic test recommendation pipeline.
- Pipeline is scalable, portable, and generalizable, enabling easy adaptation to other disease prediction and early screening tasks.
- While Llama-3.1-8B-Instruct demonstrates feasibility, ongoing advances in open-source and proprietary LLMs promise further accuracy improvements.
- Future work involves integration with EHR systems to maximize clinical impact.

Thank you!

Q & A