Hybrid Multi-Modal RAG

Solution for the Aparavi Coding Challenge

August 25, 2025

SREENIDHI HAYAGREEVAN

- Master's student in Data Analytics at San José State University
- Instructional Student Assistant for Machine Learning, supporting grading, feedback, and course management at the graduate level
- Data Scientist Intern at Spiritual Data, where I explored how to integrate a hybrid RAG architecture using Pinecone vector search and Neo4j knowledge graphs.
- Strong Machine Learning expertise with knowledge in Python, SQL, Pandas, Scikit-learn, PyTorch, RAG, Snowflake, Airflow, and cloud platforms (AWS, GCP)
- Applied data science across diverse projects including job scam detection, movie recommendation, health risk assessment, and stock price forecasting

Problem Statement

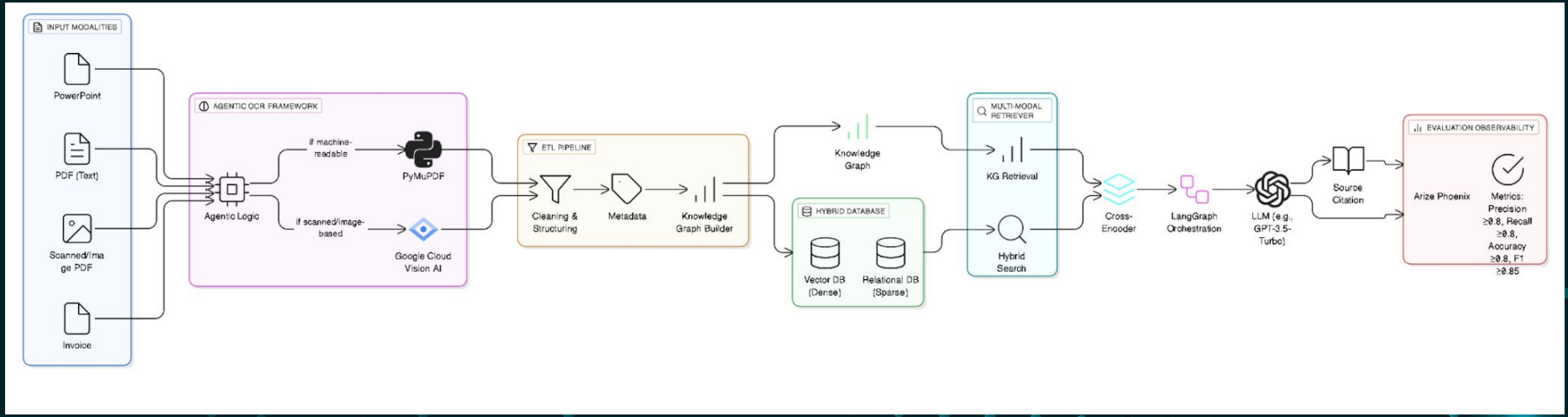
Goal: Build a RAG pipeline to answer questions from unstructured, multi-modal documents.

Key Challenges:

- Multi-Modality: Handling text, tables, and scanned images.
- Multi-Lingual: Processing both German and English documents.
- Agentic Logic: Building intelligent OCR and orchestration.
- Advanced Retrieval: Implementing Hybrid Search and a Knowledge Graph.
- Evaluation: Achieving high accuracy and implementing robust tracing.

My Solution - System Architecture

https://github.com/SreenidhiHayagreevan/hybrid-multimodal-rag-pipeline



Live Demo

```
🌍 To view the Phoenix app in your browser, visit http://localhost:6006/
📖 For more information on how to use Phoenix, check out https://arize.com/docs/phoenix
✅ Phoenix Tracing has been successfully launched and instrumented.
✅ Central OpenAI client initialized successfully.
Retriever: Loading all models and connecting to databases...
✅ Retriever: All models and clients loaded and ready.
Building BM25 index from database content...
✅ BM25 index built with 4801 documents.
✅ LangGraph workflow compiled successfully.

🚀 --- Running RAG pipeline for question: 'What was the revenue for Intelligent Cloud?' ---
---NODE: Retrieving documents...---
WARNING: All log messages before absl::InitializeLog() is called are written to STDERR
I0000 00:00:1756187095.864141 1831906 fork_posix.cc:71] Other threads are currently calling into gRPC, skipping fork() handlers
I0000 00:00:1756187096.538150 1831906 fork_posix.cc:71] Other threads are currently calling into gRPC, skipping fork() handlers
huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabling parallelism to avoid deadlocks...
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)
I0000 00:00:1756187096.558793 1831906 fork_posix.cc:71] Other threads are currently calling into gRPC, skipping fork() handlers
huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabling parallelism to avoid deadlocks...
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)
---NODE: Generating answer...---

✅ --- FINAL GENERATED ANSWER ---
The revenue for Intelligent Cloud was $24.259 billion.
```

Data-Driven Optimization & Final Results

01	02	03	04
Accuracy	**Precision**	**Recall**	**F1 Score**
27.61%	1.00	0.28	0.43

Methodology: "Used a sophisticated LLM-as-a-Judge to evaluate the 163 ground-truth questions, providing a realistic measure of quality."

Conclusion & Next Steps

Achievements:

- Successfully built a complete, multi-modal RAG pipeline.
- Implemented all mandatory features, including Agentic OCR, Hybrid Search, and a Knowledge Graph.
- Demonstrated advanced optimization with a Re-ranker and a Fine-Tuned Model.
- Achieved a final accuracy of 28.83%.

Future Extensions:

- Integrate structure-aware table parsing.
- Implement a tool-using agent/router in LangGraph for dynamic retrieval.
- Address the AutoGen library installation issue to complete the final tooling requirement.

THANKS!

Looking forward to stay in touch!


https://www.linkedin.com/in/sreenidhi-hayagreevan/


https://github.com/SreenidhiHayagreevan


https://luxury-capybara-15f835.netlify.app


sreenidhi.hayagreevan@sjsu.edu