MIDHUN NATH K R

midhunnathkr@gmail.com | +91-7907199758 | GitHub | LINKEDIN

Projects

AI-Powered Financial Analyst Chatbot

- Architected **Developed a Retrieval-Augmented Generation (RAG) pipeline** using **LangChain** to analyze complex financial documents; ingested and indexed over 100 SEC 10-K filings into a **Chroma** vector database, achieving a 90% accuracy in retrieving relevant context for user queries and enabling the system to provide source-cited answers.
- **Fine-tuned a Llama 3-8B model** on a curated dataset of financial Q&A pairs using **PyTorch** and Hugging Face. This process improved the model's understanding of financial jargon and numerical reasoning, **reducing inaccurate responses (hallucinations) by 40%** compared to the base model.
- **Engineered an end-to-end proof-of-concept** by integrating the fine-tuned LLM with the RAG system and deploying it as an interactive web application with Streamlit. The final agent could synthesize information from multiple reports to answer complex questions like "Compare the revenue growth of Company A and Company B over the last three years," demonstrating a practical solution for automating financial research.

RAG System for Document Analysis

- **Architected a scalable Retrieval-Augmented Generation (RAG) framework** using **LangChain** to enrich a base LLM with proprietary knowledge. The system was designed to seamlessly integrate custom data sources, creating a comprehensive proof-of-concept for an enterprise-grade, factual Q&A application.
- **Implemented a robust data ingestion and vectorization pipeline** for a large corpus of PDF documents. Utilized sentence transformers for high-quality embeddings and indexed them into a **ChromaDB** vector database, optimizing for low-latency similarity search that delivered highly relevant context to the model in real-time.
- **Enhanced model factuality and trustworthiness** by grounding LLM responses in verifiable documentation, successfully **reducing content hallucinations by over 90%**. This approach established a system for generating source-cited answers, dramatically increasing the model's reliability for critical business applications.

Skills

Programming Languages: Python, C++, JavaScript
Technologies:,TensorFlow, Pytorch, JAX, HuggingFace, LangChain
Other Skills: Pinecone, ChromaDB,Scikit,Docker, Kubernetes
CS Fundamentals: Data Structures and Algorithms,Operating Systems,Computer Networks,Database Management Systems
Soft Skills: Communication,Team Collaboration,Problem-Solving

Education

Vellore Institute of Technology **2022 - 2026**
- Computer Science spec in AI & ML | CGPA: **8.65** (expected)

St Antonys' Public School Kerala **2019 - 2021**
- *CBSE (Class XII), Percentage: 82%*

Arya Central School Pattom **2018-2019**
- *CBSE(Class X), Percentage: 95%*

Certifications

- **OCI Generative AI | Oracle**
- **Oracle AI Vector Search | Oracle**

Leadership & Achievements

- **Core Member** | *VIT Bhopal Android Club*
  - Led technical workshops on Android development for 50+ members, covering foundational and advanced concepts.
  - Collaborated in team settings to build and launch club projects and demonstration application.
- **Research Paper:** Authored a research paper on applying Information Theory and AI for agricultural crop prediction.
- **Hackathon Winner:** Awarded 1st Runner-Up in the VIT Bhopal Android Club hackathon.