SHREERAG NAMBOOTHIRI K H

+91 9188155396 |shreeragkh@gmail.com
Github |LinkedIn |www.shreerag.me

OBJECTIVE

MCA student at Vellore Institute of Technology with hands-on experience in prompt engineering, LLM application development, and AI agent design. Skilled in building Retrieval-Augmented Generation pipelines, multi-step conversational agents, and prompt evaluation workflows using tools such as LangChain, LangGraph, and the Claude API. Seeking an AI/ML internship to contribute to real-world LLM-powered products.

EDUCATION

Vellore Institute of Technology *May 2027 (expected)*
M.C.A
Naipunnya Institute of Management and Information Technology (Univeresity of Calicut) *2022-25*
B.sc in Computer Science

PROJECTS

Hybrid Search RAG Application *Python, FastAPI, LangChain, ChromaDB, BM25s, Redis, Streamlit*
https://github.com/shreeragkh/Hybrid-Search-RAG

- Built a hybrid RAG pipeline (BM25s + ChromaDB dense search, RRF fusion, cross-encoder reranking) with a FastAPI backend and Streamlit UI, exposing query, ingestion, auth, and cache-management endpoints.
- Integrated Firebase Google Authentication for role-based access control, gating admin dashboard and management endpoints behind an authorized account.
- Containerized with Docker and deployed the full stack (FastAPI, Streamlit, Redis) on AWS EC2, running end-to-end in a live cloud environment.

Repo Context Copilot *Python, FastAPI, LangChain, AstraDB, BM25s, Redis, React, Firebase, Groq/OpenAi*
https://github.com/shreeragkh/Repo-context-copilot

- Built a hybrid RAG pipeline for querying arbitrary GitHub repositories, combining BM25 sparse retrieval with AstraDB dense vector search fused via Reciprocal Rank Fusion and cross-encoder reranking.
- Reduced LLM context tokens by ˜29.9% while retaining ˜96%of baseline answer quality, using a query-complexity classifier that adaptively sets retrieval depth and an adaptive rerank-score cutoff, validated on a complexity-tiered eval set.
- Achieved ˜100% correct abstention on out-of-scope queries via an LLM-as-judge evaluation harness with a dedicated unanswerable-query test set.

Prompt Tester *Python, LangGraph, Streamlit, Claude API, LLM Evaluation*
https://github.com/shreeragkh/Promp-Testing

- Built a prompt-engineering benchmarking tool comparing zero-shot, few-shot, and chain-of-thought strategies via the Claude API, using Claude as an LLM-judge to score outputs and auto-select the winning strategy per run.
- Deployed as an interactive Streamlit app with session history, CSV export, and a win-count comparison chart across multiple benchmark runs.

AI to Detect Fake Open Source Projects

- Developed a machine learning model to detect inactive or fake open-source repositories using multi-dimensional GitHub data.
- Engineered features from repository metadata, contributor activity, bot contributions, code similarity, and contribution graphs using Github API.
- Evaluated multiple classification models (including XGBoost and Logistic Regression); Logistic Regression delivered the best performance for the task.

- Applied SHAP explainability to interpret model predictions and identify key influencing features.
- Currently preparing research work for publication.

TECHNICAL SKILLS

Programming	Python, C++
AI/ML	Generative AI, Large Language Models (LLMs), Prompt Engineering, RAG, AI Agents, Semantic Search, Intent Classification, LLM Evaluation, Machine Learning, Pandas, NumPy, Matplotlib
Frontend	React
Prompt Engineering	Zero-shot, Few-shot, Chain-of-Thought, Role-based Prompting, Prompt Optimization, Performance Benchmarking
Frameworks, Libraries	LangChain, LangGraph, FastAPI, Streamlit, Flask, React.js, Scikit-Learn
Retrieval & Search	Hybrid Search (BM25 + Dense), Reciprocal Rank Fusion (RRF), Cross-Encoder Reranking, Semantic Search, Embeddings, Vector Search
Vector Databases & Caching	FAISS, ChromaDB, AstraDB, Vector Embeddings, Redis
Tools, Platforms	Git, GitHub, Firebase Auth
Cloud	AWS EC2, Docker