JEEDIPALLI RITHISH KUMAR REDDY

GenAI Engineer | RAG & LLM Applications | LangChain & Vector Search | Python

Hyderabad, Telangana, India | 9346422829 | jeedipallirithish@gmail.com | github.com/Jeedipallirithish

PROFESSIONAL SUMMARY

B.Tech Computer Science (Data Science) graduate (2025) with hands-on experience building a Retrieval-Augmented Generation (RAG) application end-to-end — document ingestion, chunking, embeddings, vector search, and LLM-based response generation — using LangChain, ChromaDB, HuggingFace embeddings, and the Google Gemini API. Comfortable with prompt engineering, retrieval evaluation, and exposing GenAI pipelines through FastAPI and Streamlit. Seeking an entry-level GenAI/RAG Engineer role.

TECHNICAL SKILLS

Programming: Python, SQL

Generative AI & LLMs: LangChain, Prompt Engineering, Google Gemini 2.5 Flash (API), HuggingFace

RAG & Vector Search: Retrieval-Augmented Generation (RAG), Embeddings, Semantic Search, ChromaDB, FAISS, PyPDFLoader, RecursiveCharacterTextSplitter

Deployment & APIs: FastAPI, Streamlit, REST API Development, SQLite (query/analytics logging)

Supporting ML: Scikit-learn, XGBoost, CatBoost, Feature Engineering (from applied ML projects)

Version Control: Git, GitHub

WORK EXPERIENCE

Team Lead – Agentic AI Internship, Innomatics Research Labs **Nov 2025 – Mar 2026**

- Built and validated 3+ Agentic AI/RAG automation solutions using LangChain, HuggingFace embeddings, and vector databases (FAISS, ChromaDB), following structured workflows and coding standards.
- Automated document processing across 100+ pages of unstructured content — cleaning, chunking, semantic search, and retrieval — for an AI-powered knowledge assistant.
- Ran prompt engineering and output validation cycles, checking retrieval accuracy and response relevance against defined evaluation criteria.
- Debugged retrieval and generation pipelines, flagging unreliable outputs before delivery to ensure accurate, reproducible results.
- Documented experiments and presented pipeline retrieval performance to leadership, translating technical metrics into stakeholder-facing insights.

Team Lead – Data Analysis Internship, Upto Skills **Nov 2024 – Feb 2025**

- Led a team of 5 interns to deliver a data analytics project on schedule, coordinating priorities, timelines, and stakeholder reporting.
- Cleaned and analyzed 10,000+ records using Python (Pandas, NumPy) and Excel to support structured hypothesis testing.

PROJECTS

AWS Customer Agreement RAG Chatbot | *Python, LangChain, ChromaDB, HuggingFace, Google Gemini 2.5 Flash, FastAPI, Streamlit, SQLite*

- Built an end-to-end RAG application to answer natural-language queries over AWS Customer Agreement document content.
- Pipeline: document ingestion and chunking with PyPDFLoader and RecursiveCharacterTextSplitter → HuggingFace embeddings → semantic search over ChromaDB → context-aware retrieval → Google Gemini 2.5 Flash for source-grounded response generation.
- Applied iterative prompt engineering to improve answer relevance and reduce inaccurate responses, validating retrieval quality through structured testing.
- Developed FastAPI REST endpoints for ingestion, question-answering, and analytics, paired with a Streamlit frontend that displays answers alongside retrieved source chunks.
- Added SQLite-based analytics logging to track query volume, response latency, and unanswered questions for ongoing monitoring. Built and tested locally; not deployed to cloud infrastructure.

Employee Attrition & Income Prediction Systems | *Python, Scikit-learn, FastAPI, Streamlit, SQLite*

- Built two end-to-end ML services (classification and regression) with FastAPI backends and Streamlit interfaces, reflecting the same API-first, application-oriented approach used in GenAI project delivery.

EDUCATION

B.Tech, Computer Science (Data Science) — Malla Reddy College of Engineering, 2025

Intermediate (MPC) — Sri Chaitanya Junior College, 2021

CERTIFICATIONS

Data Science with Advanced GenAI — Innomatics Research Labs | Python Programming — Innomatics Research Labs | HTML, CSS & JavaScript — NSDC