

Chandras | Generative AI Engineer

 chandrasreddyv@gmail.com  6362617962

 <https://www.linkedin.com/in/chandras-reddy-75b0a6172>

PROFESSIONAL SUMMARY

AI Engineer with 2.9+ years of experience building scalable Generative AI solutions using Large Language Models (LLMs), Retrieval -Augmented Generation (RAG), and Agentic AI architectures. Experienced in designing multi-agent systems with planner, executor, and verifier roles. Proficient in building and integrating end-to-end AI applications using LangGraph, LangChain, Hugging Face, FastAPI, and Streamlit, with a focus on performance, efficiency, and cost optimization through fine-tuning and model optimization.

TECHNICAL SKILLS

- **Programming Languages:** Python
- **Tools:** Fast API, VS Code, Streamlit, Git, Excel VBA
- **Skills:** LLMs, Generative AI, Agentic AI, Machine Learning, NLP, Model Fine-tuning.
- **Frameworks & Libraries:** LangGraph, LangChain, LangSmith, Scikit-Learn, NLTK, Hugging Face Transformers, Vectore Databases(FAISS, Pinecone, ChromaDB)

EXPERIENCE

Cognizant Technology Solutions
AI Engineer

Feb 2024 - Present
Bengaluru, Karnataka

- Built a natural language **SQL query assistant** enabling users to query databases conversationally, export results to Excel, and auto-generate reports via email improving data accessibility for non-technical users by reducing manual reporting effort by 50%.
- Developed an AI-powered system leveraging large language models (**LLMs**) to extract and summarize content from websites and generating clear, concise, and context-aware summaries.
- Built a chatbot with message history using large language models and vector databases (**FAISS, ChromaDB, Pinecone**) to support context-aware and dynamic user interactions.
- Implemented end-to-end workflows combining **LangChain, Hugging Face, Vector DB, Fast API** and **Streamlit** to deliver interactive and user-friendly AI tools with real-time data processing.
- Built an **Agentic AI orchestration framework** where multiple autonomous agents collaboratively planned, reasoned, and executed tasks using role-based responsibilities (planner, executor, verifier), enabling reliable multi-step decisionmaking and automated task execution.
- Implemented an **Agentic RAG** system using multiple vector databases to improve retrieval quality and answer accuracy.

Cognizant Technology Solutions
Performance Engineer

May 2023 – Feb 2024
Bengaluru, Karnataka

- Developed and optimized LoadRunner scripts using VuGen for comprehensive performance testing of web and clientserver applications.
- Built API performance test scripts using AI to automate script generation and validation.
- Conducted rigorous load testing and performance tuning using JMeter, identifying system bottlenecks and optimizing scalability.
- Collaborated with development and operations teams to enhance overall system efficiency and reliability based on performance testing insights.

EDUCATION

Bachelor of Technology – Computer Science and Engineering
Hindustan Institute of Technology and Science – Chennai

2022
CGPA: 8.16/10

PROJECTS

GitHub: <https://github.com/Chandrasreddy1999?tab=repositories>