

ADITYA RAVI RAJ

📍 Pune

✉️ adityame0010@gmail.com

📞 +91 7985155351

LinkedIn in adityadev30

Github StellarPhantasm30

LeetCode

aditya_dev30

EXPERIENCE

Senior AI Engineer

July 2025 - Present

Pune, India

Synechron

- Designed and developed Validate.AI, an enterprise-grade MLOps/LLMops platform serving as a central gateway for 100% of organisational LLM traffic with centralised **RBAC**, observability, **guardrails** and compliance control.
- Standardized the completion/chat endpoint interface for full LangChain compatibility, streamlining integration across multiple teams and tools.
- Enabled real-time observability and cost tracking by integrating Langfuse, ClickHouse, and PostgreSQL, providing visibility into token usage, cost savings, model cost, and caching efficiency.

AI Engineer

June 2022 - July 2025

Bhubaneswar, India

Infosys

- Automated the evaluation of Agentic RAG pipeline output using the **Ragas** framework and semantic similarity scoring, reducing manual effort by **20%**.
- Designed and developed an **Agentic RAG** back-end service using **Langchain & LangGraph** that leveraged **AWS Bedrock** API to generate test cases for close to **900** APIs in the client's environment, reducing manual effort by **50%**.
- Deployed Open source LLMs as a microservice using **Flask** and **Fast API** in Azure Virtual Machine.
- Developed and deployed **Generative AI** ChatBot by fine-tuning Large Language Models (LLMs) using domain-specific retail data, employing **QLORA** and **Parameter-Efficient Fine-Tuning (PEFT)** techniques resulting in over **40%** reduction in manual effort.
- Processed over **20,000** test cases using **Pandas**, enhancing compatibility for LLM fine-tuning and achieving more than **20%** increase in the accuracy of fine-tuned model for Test Case Generation.
- Developed a full-stack application with a **React** frontend and Express backend, integrating Dockerized microservices to host open-source LLMs via **Ollama**. Achieved **70%** text generation accuracy and **60%** code generation accuracy.
- Built and orchestrated autonomous AI agents to automate testing and reporting tasks in AIOps environments, using LLMs, prompt engineering, and agent orchestration techniques.
- Fine-tuned embedding models using **LlamaIndex**, increasing accuracy by **8%**.

SKILLS

Programming Languages and Database:

Java, Python, MySQL, FAISS, Chroma, Pinecone

Back End:

FastAPI, REST, MCP, Flask, Docker, Prompt Engineering, numpy, Streamlit

Cloud & DevOps:

AWS, Azure, Docker, Redis, Kafka, Git, serverless, litellm

GenAI:

LangChain, Tensorflow, CrewAI, Langgraph, PEFT, scikit, BedRock, OpenAI

LLMops & MLOps:

MLFlow, DVC, WandB, Langfuse, Langsmith

Architecture & Methodologies:

Microservices, Monolithic, System Design, Agile/SCRUM

Others:

Problem Solving, Data Structures and Algorithms

ACHIEVEMENTS

- Solved more than **200** questions on **LeetCode** with global rank under **60%**. [Supporting Documentation](#)
- Certified by **Google** in Generative AI. [Supporting Documentation](#)
- Certified by **GitHub** in GitHub Foundations. [Supporting Documentation](#)
- Won the **Infosys Topaz** AI Innovation Challenge, delivering the **best AI solution** for a client across all units.
- Won the IQE GenAI Hackathon sponsored by AWS for my solution on Automation framework modernization.
- Achieved **Finalist** position at the GFG Hack for Future Hackathon, out of **38,000** participants.
- Awarded **Employee of the month** for a proactive approach to solving production challenges in Gen AI applications.
- Technical blogger on Medium and LinkedIn with 1000+ views monthly.

EDUCATION

Bachelor of Technology, United College of Engineering and Research

2015 - 2019