MANJUSHA RAAVI

manjushaa.raavi@gmail.com — +91 9390759782 — Hyderabad, India
LinkedIn — GitHub

Summary

Backend and AI Engineer with 3+ years of global experience building cloud-native AI systems and scalable microservices. Specialized in AI automation, LLM fine-tuning, RAG pipelines, and deploying production-grade models. US-educated with real-world startup-paced experience delivering high-impact AI tools. Passionate about solving business problems through AI-driven innovation in fast-paced, iterative environments.

Experience

AI Engineer
May 2024 – Present
Brillius Pvt Ltd, Hyderabad
– Designed secure FastAPI microservices with OAuth2 and JWT, enabling scalable access to AI model APIs.
– Fine-tuned LLMs (Mistral, GPT, LLaMA2) on client-data using LoRA/ QLoRA, boosting NLP task accuracy by 40%.
– Architected and productionized Retrieval-Augmented Generation (RAG) pipelines using LangChain, LlamaIndex, and HuggingFace Transformers.
– Built high-performance semantic search systems with FAISS, Qdrant, and Redis, reducing query latency by 30%.
– Containerized and scaled services on AWS, automating model tracking and deployments via Docker, CI/CD pipelines, and MLflow.
– Worked closely with leadership to drive core AI initiatives, contributing to product roadmaps and fast feature delivery.

Production Support Analyst
Aug 2022 – Feb 2024
Tata Consultancy Services, Irving, TX (USA)
– Handled critical production incidents (P1/P2) across large-scale distributed systems, reducing the mean-time-to-resolution (MTTR) by 20%.
– Automated incident triage and health checks, increasing operational efficiency by 60%.
– Collaborated cross-functionally with development teams to improve system reliability and minimize downtime.
– Conducted root cause analysis to proactively eliminate recurring system failures.

Software Development Engineer in Test
Jun 2018 – Nov 2019
Planful Inc, Hyderabad
– Developed robust Java/TestNG automation frameworks, increasing test coverage to over 80%.
– Integrated static analysis and code coverage tools into CI/CD pipelines to maintain code quality standards.
– Validated backend data integrity and API responses using Postman and SQL-based testing.

Projects

PDF PII Redaction Tool
GitHub
– Built an AI-driven tool to automate the redaction of sensitive data in PDFs, achieving 95%+ precision, and reducing manual review time by 50%.

Document QA & Summarizer with RAG
GitHub
– Created a Streamlit app using Llamaindex and ChromaDB to perform context-aware QA and summarization, reducing manual document reading time by 70%.

AI Content Generator with Web Scraping
GitHub
– Built an automated system that combines reddit/ Wikepedia scraping, GPT-4, and ModelsLab API for fast dynamic content and image generation, reducing manual effort by 85%.

Technical Skills

– **Languages:** Python, Java, SQL
– **Frameworks/Tools:** FastAPI, REST, LangChain, LlamaIndex, Spring, Docker, MLflow, Jenkins, Git, AWS (EC2, S3, Lambda), WebSockets
– **Databases:** MongoDB, Vector DBs (FAISS, Qdrant, Chroma), MySQL
– **AI/ML**: LLM Fine-tuning (LoRA, QLoRA), RAG, NLP (Summarization, NER), PyTorch, TensorFlow, HuggingFace.

Education

M.S. in Computer Science
May 2022
New Mexico Institute of Mining and Technology, USA

Certifications

Coursera – Generative AI Engineering with LLMs (2025)
Amazon Web Services – AWS Certified Cloud Practitioner (2022)