

Atharva Udavant

AI Engineer | Generative AI, Agentic Systems & Python Backend

Hyderabad, Telangana | 17.atharva@gmail.com | +91 9979849768 | LinkedIn | GitHub

Professional Summary

AI Engineer and Python backend professional building production LLM, agentic AI, RAG, document-processing, and intelligent-automation systems. Hands-on with Claude, Azure OpenAI, LangGraph, LangChain, MCP, embeddings, vector search, FastAPI, Kafka, and Airflow across Azure and AWS. Owns solutions from stakeholder requirements and architecture through integration, deployment, debugging, monitoring, and production support, delivering measurable improvements in processing effort, latency, and infrastructure efficiency.

Technical Skills

Programming & Data: Python, SQL, JavaScript, Pandas, NumPy, Data Analysis, Data Validation

Generative AI: Claude, Azure OpenAI, OpenAI API, Large Language Models, Prompt Engineering, LLM Fine-Tuning, Structured Outputs

Agentic AI: LangGraph, LangChain, Model Context Protocol, Tool Calling, Task Orchestration, Enterprise API Integrations

RAG & Retrieval: Embeddings, Semantic Search, Vector Search, Chunking, Knowledge Ingestion, ChromaDB, AstraDB

ML & NLP: Classification, Clustering, Text Classification, Question Answering, NER, Model Evaluation

Backend & Data Engineering: FastAPI, REST APIs, Python Microservices, Kafka, Airflow, CDC, ETL, Background Jobs

Cloud & Delivery: Azure, AWS, Docker, Linux, Serverless, Git/GitHub

AI Operations: OpenTelemetry, Phoenix, Log Analysis, Root-Cause Analysis, Monitoring, Error Handling, Reliability, Performance Optimization

Databases: SQL Server, ClickHouse, DynamoDB, MongoDB, Redis

Professional Experience

Software Engineer, AI/ML, Align Technology – Hyderabad, India

August 2025 – Present

- Owned the architecture, development, integration, and production rollout of an **AI Scrum Master agent** using **Claude, LangGraph, and MCP**, translating stakeholder requirements into agentic workflows and tool calls across ServiceNow and Microsoft Graph APIs.
- Led end-to-end delivery of the **Bulk KB Article Creator**, partnering with business users from requirements discovery and solution design through implementation, UAT, production deployment, and support; translated high-volume needs into validation rules, workflow logic, and API-driven automation.
- Designed and deployed an **LLM document-processing pipeline** that cleaned, chunked, classified, and validated more than **2,000 documents**, reducing manual effort by **95%** and saving approximately **750 hours**.
- Built a low-latency **RAG and data-synchronization pipeline** using embeddings, vector search, and CDC, reducing knowledge-base query-resolution time by **50%**.
- Developed an Azure-based **AI log-analysis workflow** to identify recurring production errors, accelerate root-cause analysis, and generate structured RCA reports for engineering teams.
- Built an **AI employee referral workflow** integrating Microsoft Forms, SharePoint, LLM resume parsing, skill classification, automated acknowledgements, and referral-status tracking.

Software Engineer, Locobuzz Solutions Pvt. Ltd. – Mumbai, India

April 2024 – July 2025

- Designed and supported high-throughput Python microservices and data pipelines for operational, analytics, and AI workloads, handling production debugging, error resolution, and performance optimization.
- Architected a highly available ETL pipeline using **Apache Kafka**, implementing scalable ingestion, transformation, and failure handling that reduced processing time by **80%**.
- Developed scheduled and on-demand **Apache Airflow** workflows with validation and error handling, reducing reporting latency by **80%** and improving pipeline reliability.
- Replaced inefficient Redis scans with a multi-level **DynamoDB caching strategy**, reducing CPU utilization and improving infrastructure efficiency.

Projects

Enterprise Action Assistant | Multi-Model Agentic AI Assistant — Python, FastAPI, React, Claude, Azure OpenAI, MCP

- Built a multimodal assistant with switchable Claude and Azure OpenAI models, SSE streaming, image/document processing, conversation persistence, and automatic context summarization.
- Developed provider-independent orchestration and MCP-based ServiceNow tool discovery; instrumented LLM workflows with Phoenix and OpenTelemetry to trace latency, token usage, errors, and execution behavior.

Snippety | Fine-Tuned Developer Assistant — Python, Hugging Face, LangChain, Vector Search

- Combined LLM fine-tuning with semantic retrieval and domain-specific training data to generate and evaluate context-aware, executable code snippets.

Education

BTech in Computer Science – GPA: **9.2/10.0**, SRM Institute of Science and Technology

Sept 2020 – May 2024