

Lohith Kovuri

lohithkovuri321@gmail.com | +1 (856) 652-4065 | LinkedIn | GitHub | San Francisco, CA

Summary

Software Engineer with 2+ years of experience building production backend systems and AI-powered applications. Has designed fault-tolerant distributed services on AWS, and separately built agentic AI systems with tool-call retry, evaluation loops, and cost-aware model routing. Works across Python, Java, and Rust, learns new languages and systems fast, and owns problems from design through production.

Skills

Backend & APIs: Java, Spring Boot, FastAPI, Node.js, REST

Agentic AI: LangGraph, LangChain, MCP, OpenAI API

Retrieval & Evaluation: FAISS, RAG, RAGAS

Cloud & Infra: AWS (ECS, S3, Lambda, CloudWatch), Docker, Kubernetes

Distributed Systems: Kafka, Redlock

Languages: Python, Java, Rust, TypeScript, SQL

Data: PostgreSQL, Redis, DynamoDB

Tools: Git, GitHub Actions, Pytest

Experience

Software Engineer | Mercury | Remote, USA

Oct 2025 – Jul 2026

- Designed a fault-tolerant, distributed event-ingestion service on AWS with idempotent writes and dead-letter replay, sustaining **99.9% write availability** across 8x traffic bursts.
- Reduced MTTR **by 40%** by redesigning failure-recovery logic with exponential backoff and documenting recovery procedures per failure mode.
- Built an AI-powered test-generation system in CI that synthesizes edge cases from schema definitions across 3 services, cutting development time **by 60%**.

Backend Engineer (Contract) | Rowan ACM Engineering | Remote, USA

Jun 2025 – Oct 2025

- Architected a multi-tenant AI backend serving **200+ concurrent tenants**, designing per-tenant isolation that eliminated cross-tenant data exposure by construction.
- Deployed a live evaluation system that scores AI output quality and automatically rolls back on regression, catching failures **15 to 30 minutes before user impact**.
- Cut p95 latency **from 120s to under 60s** by converting a synchronous service to async with connection pooling and backpressure control.

Software Developer | Roothoot Pvt Ltd | Hyderabad, India

Sep 2022 – Aug 2023

- Engineered real-time data pipelines with Kafka, improving throughput **by 30%** through tuned checkpointing and backpressure handling.
- Optimized Spring Boot REST endpoints with composite indexing and caching, dropping P95 API latency **by 25%**.
- Created a unified monitoring dashboard consolidating consumer lag, error rates, and retry counts into a single view, cutting incident triage time **by 50%**.

Projects

AI Code Review Agent | Python, LangGraph, FastAPI, OpenAI API, GitHub Actions

- Built a multi-tool agent system from scratch across 4 tools (diff retriever, static analyzer, test generator, PR commenter), implementing tool-call retry and state checkpointing so a mid-chain failure recovers instead of restarting.
- Kept the orchestration loop model-agnostic and wired it into GitHub Actions to run on every pull request, catching **85%+ of seeded defects** in benchmark diffs at under \$0.05 per review.

High-Throughput Data Pipeline | Rust, PostgreSQL, Prometheus

- Learned Rust from scratch specifically to get zero-copy memory control a garbage-collected language couldn't offer, then shipped a distributed ingestion service processing **1M+ records/hour** at under 1% failure rate.
- Instrumented query latency, retry rates, and buffer utilization with Prometheus, catching and resolving 2 connection-pool saturation events under load before they breached write SLAs.

VendorOps AI Full-Stack Agentic Platform | React, TypeScript, Python, FastAPI, MCP

- Orchestrated a 3-agent pipeline with confidence-gated human handoff, reaching **90%+ extraction accuracy** across 100+ documents, and cut inference cost **by 60%** via model-size routing.
- Exposed the platform as an MCP server, standardizing how external agents and tools discover and call its capabilities instead of relying on hardcoded integrations.

Education

Rowan University | M.S. Data Science

Sep 2023 – May 2025

Dr. MGR Educational and Research Institute | B.Tech Information Technology

Sep 2019 – May 2023