Sapan Reddy Kanamathareddy

Atlanta, GA www.sapanreddy.com sapanreddy@usf.edu scodes73 /in/sapan-reddy

SUMMARY

Software Engineer with **5+** years of experience building and deploying production systems and scalable cloud-native applications. Experienced developing **LLM-powered** applications, **RAG** pipelines, intelligent agents, and high-performance inference APIs using **Python FastAPI**, while building intuitive user experiences with **React/TypeScript**. Proven ability to translate ambiguous business problems into production-ready AI solutions, architect distributed systems processing **500K+** daily transactions. Experiernced in leading cross functional teams of **10+** people.

SKILLS

AI/LLM: LangChain, LangGraph, RAG, MCP, Vector Databases, OpenAI APIs, Anthropic APIs

Frontend: ReactJS, TypeScript, JavaScript, NextJS, Angular, HTML5, CSS3, Redux, Responsive Design

Backend: Python, FastAPI, Java, Spring Boot, REST APIs, Apache Kafka, Redis

Databases: PostgreSQL, MySQL, MongoDB, DynamoDB

EXPERIENCE

SpaceXAI Jun. 2026 - Present
Software Engineer Specialist Remote

- Part of end-to-end **LLM evaluation** and **model training** for software engineering applications, developing coding **benchmarks**, implementing **automated evaluation pipelines**, analyzing model behavior, and driving iterative improvements that increased model quality and reliability in production. Developed and maintained **internal tooling** that streamlined evaluation and training workflows, reducing manual effort and accelerating experiment iteration.

Easley-Dunn Productions Feb. 2026 - Jun. 2026
Software Developer Los Angeles, CA

- Engineered large-scale **evaluation pipelines** for synthetic agent trajectories using **Python**, **asyncio**, and **Redis task queues**, implementing **batch simulation execution**, **parallel worker orchestration**, and **fault-tolerant metric aggregation pipelines** over **10K+ runs** with **Pandas** and **SQL-based analytics**, building automated **evaluation (evals)** for reward efficiency, progression stability, and economy balance, improving iteration speed for evaluation and tuning workflows by **40%**.

Camber, Inc Sep. 2025 - Nov. 2025
Software Developer Intern Piscataway, NJ

- Built a **Retrieval Augmented Generation (RAG) pipeline** using **Python**, **FastAPI**, **LangChain**, **OpenAI embeddings** to query over **50K+** internal documents, enabling grounded LLM responses and reducing manual document lookup time by **30%** for operations teams.
- Designed and deployed **MCP-based agent workflows** integrating shipment management systems, document repositories, and operational tooling, reducing manual workflow completion time by **40%** across key logistics processes.

Zemoso Technologies Mar. 2021 - Dec. 2023
Software Engineer III Hyderabad, India

- Engineered a **headless iframe-based SDK** using **ReactJS**, **Typescript** and **the postMessage API** for secure cross-domain communication, replacing a legacy event-driven system and improving scalability across client platforms.
- Developed **high-performance backend APIs** using **Java, Spring Boot, PostgreSQL, Kafka, and Redis**, supporting **thousands of concurrent requests** for ingestion, authorization, and asynchronous event processing.
- Improved frontend architecture by **migrating 100+ React components to TypeScript** and implementing **client-side caching and selective data hydration**, resulting in **10x faster UI workflows** and improved reliability.
- Built high-throughput **bulk-processing pipelines** in **Spring Boot** for **AWS S3-based ingestion**, enabling batch processing of **millions of records daily** and reliable third-party integrations.
- Reduced SDK **bundle size by 44%** and **startup latency by 30%**, improving load times, **Core Web Vitals**, and production observability via **structured logging with Graylog**.
- Established **automated testing** with **Cypress, Jest, Mockito, and JUnit**, achieving **90%+ coverage** for critical workflows and reducing regressions by **40%** through CI/CD integration.

Smart Campus Connect Jan. 2019 - Jan. 2021
Founding Software Engineer Hyderabad, India

- Designed a secure **multi-tenant backend architecture** using **Spring Boot, PostgreSQL, and JWT-based RBAC/ABAC**, reducing **privilege escalations by 60%** across institutions.
- Implemented **isolated-schema multi-tenancy** with **Liquibase migrations**, onboarding **25+ institutions** with strong data isolation and simplified tenant upgrades.
- Developed CRM data migration and operational automation tooling with validation and resumability, using **Python** and **Spring Batch pipelines**, reducing client onboarding time by **50%**.
- Automated **CI/CD pipelines** using **GitHub Actions**, enabling **zero-downtime deployments**, a **3x increase in release velocity**, and a **90% reduction in deployment failures**.
- Led backend **performance optimization** through query tuning and **Caffeine caching**, reducing **average API latency by 45%** under peak load.

EDUCATION

University of South Florida Jan. 2024 - Dec. 2025
*Master of Science in Computer Science and Engineering (**4.0**/4.0 CGPA)* Tampa, FL, US

PROJECTS

LocalImg-MLX

Native macOS app for fully offline Stable Diffusion inference on Apple Silicon using Swift, SwiftUI, SwiftData, and MLX, enabling privacy-preserving text-to-image generation. Engineered a modular ImageGenerationEngine protocol to support multiple models and backends, implementing memory-aware strategies with FP16 and quantized weights for devices with < 8GB RAM. Built real-time progressive denoising, zoomable previews, and generation history persistence while handling high-bandwidth GPU computations in a responsive SwiftUI interface.

Skyflow - Data Privacy Vault

Skyflow is a cloud-native data protection platform that tokenizes and securely governs access to sensitive PII in a multi-tenant environment. Contributed to high-availability Java, Spring Boot services for secure token resolution, policy enforcement, and tenant isolation. Implemented Kafka-based event workflows for audit logging and compliance tracking, and optimized PostgreSQL performance for low-latency access under high concurrency while improving observability in production systems.

Strateos - Cloud R&D Automation

Strateos is a distributed cloud platform that remotely orchestrates automated laboratory infrastructure and long-running scientific workflows. Built scalable Java/Spring Boot services supporting asynchronous experiment orchestration and concurrent scheduling. Designed Kafka-driven event pipelines for command dispatch and reliable state tracking, ensuring fault tolerance, idempotency, and secure multi-user access in a distributed environment.

HONORS AND ACHIEVEMENTS

- Candidate of **Dean's list** for having a **perfect 4.00 GPA** at University of South Florida.
- Ranked among **top 0.1% out of 150k participants**, Successfully cleared HackWithInfy, TCS Codevita coding challenges.
- Won **1st** place at **BioAsia 2019 Hackathon** for developing DripM, an innovative solution in healthcare technology.
- **Runners-up** at **Smart India Hackathon (India's largest hackathon)** for developing a chatbot solution for the Crime Department, Government of Maharashtra.