

Mohana Satyanarayana Moganti

📍 San Jose, CA | 📞 +1 (669)-329-9412 | ✉ mohanmoganti2023@gmail.com | [in LinkedIn](#) | [GitHub](#) | [Portfolio](#)

TECHNICAL SKILLS

AI / ML: RAG, Agentic AI, LangChain, LlamaIndex, CrewAI, OpenAI, Gemini, HuggingFace, PyTorch, TensorFlow

Languages: Python, TypeScript, JavaScript, Golang, Scala, Java, SQL

Web & APIs: FastAPI, ReactJS, Next.js, Spring Boot, GraphQL, Django, Flask

Cloud & Infra: AWS (EC2, Lambda, EKS, Glue, DynamoDB), Azure, GCP, Docker, Kubernetes

Data & Vectors: PostgreSQL, MongoDB, Redis, Snowflake, Weaviate, Pinecone, ChromaDB, Neo4j, GitHub Actions, Harness, Langfuse, RAGAS

EXPERIENCE

AI Engineer

Apr 2026 – Present

NextPhase.ai

San Jose, CA

- Delivered a complete enterprise AI package covering infra provisioning, model deployment on AWS EKS with ArgoCD, LLM evaluation with RAGAS, observability via OpenTelemetry and Grafana, and per-client usage and cost analytics.
- Built a production RAG system for secure document interaction using LangChain, Weaviate, and OpenAI, integrating Microsoft Presidio, GLiNER, and Microsoft Purview for PII redaction, NER, and enterprise data governance before content entered the context window.
- Designed agentic workflows with CrewAI and LlamaIndex for multi-step document reasoning, automated report generation, and tool-augmented extraction pipelines across client datasets.
- Built an Ansible Playbook Sandbox, a browser-based IDE for storing, composing, and executing Ansible playbooks with a pre-built YAML block library covering common infra tasks, paired with an AI assistant for on-the-go playbook generation, explanation, and debugging in real time.

Founding Engineer

Aug 2025 – Apr 2026

Gembizz LLC

San Jose, CA

- Architected GemBizz from zero to MVP, a full-stack PWA connecting local business districts and merchants, built with a FastAPI backend and React/TypeScript frontend with Vite for hot-module reloading and optimized developer experience.
- Designed and implemented business support organization (BSO) management features including merchant onboarding flows, district dashboards, and role-based access control with JWT authentication.
- Built real-time communication features including announcements, merchant alerts, and chat using WebSockets and Redis Pub/Sub, increasing verified user engagement by 25%.
- Deployed scalable microservices on AWS EC2 with Docker and MongoDB Atlas; automated CI/CD pipelines via GitHub Actions reduced deployment time by 50%.
- Configured IAM roles, VPC security groups, and AWS API Gateway to enforce least-privilege networking and secure service-to-service communication.
- Integrated automated transactional and digest email systems and built admin dashboards for platform analytics, merchant activity tracking, and BSO reporting.

AI Engineer

Sept 2024 – Nov 2024

AstraNetix Corporation

San Jose, CA

- Developed multimodal Retrieval-Augmented Generation (RAG) systems using Python, Django, Weaviate, GraphQL, and OpenAI, deploying serverless architectures with AWS Lambda to boost scalability and accuracy.
- Integrated Apollo Router to optimize federated GraphQL query performance and improve cross-service communication by 40%.
- Improved response times using optimized recommendation algorithms, HNSW indexing, and efficient data pipelines.

Teaching Assistant

Aug 2024 – May 2025

San Jose State University

San Jose, CA

- Instructed three core computer science courses: CMPE 132 (Information Security), CMPE 148 (Networking), and CMPE 257 (Machine Learning), supporting 100+ students across multiple sections with lectures, labs, and tutoring sessions.
- Designed and implemented automated grading systems using Python and shell scripts, reducing manual grading time by 40% while maintaining rigorous evaluation standards across cryptographic protocols, network implementations, and ML model assessments.
- Developed hands-on lab assignments and practical exercises for information security (RSA encryption, TLS handshakes), networking (TCP/IP stack, routing algorithms), and machine learning (neural networks, model evaluation), fostering deeper technical understanding.
- Maintained comprehensive office hours and email support, providing detailed feedback on student submissions and troubleshooting technical issues, tracked student progress and identified struggling learners for early intervention and additional resources.
- Collaborated with faculty to review exam questions, standardize rubrics, and refine course content based on student performance metrics and feedback, contributed to curriculum improvements across all three courses.

AI Product Manager

May 2024 – Aug 2024

Flatirons AI LLC

San Jose, CA

- Led the design and deployment of scalable ETL pipelines with AWS Glue and PySpark for generative AI and recommendation systems, utilizing Neo4j and Microsoft GraphRAG for advanced graph analytics.
- Oversaw cross-functional teams, architecting model solutions and leveraging Azure and AWS for production AI model management and deployment.
- Coordinated sprint planning, stakeholder reviews, and QA cycles across engineering and data science teams to deliver AI features on schedule.

Associate Software Analyst

Aug 2021 – June 2023

Deloitte Touche Tohmatsu India LLP

Hyderabad, India

- Developed and optimized fintech web applications using React and TypeScript within a large-scale enterprise environment focused on security, performance, and user experience.
- Designed and implemented GraphQL API integrations to enhance data access efficiency and enable responsive real-time functionality across distributed systems.
- Implemented and maintained automated CI/CD pipelines using Harness, supporting continuous delivery and compliance across multiple client engagements, improving deployment velocity by 40%.
- Collaborated across global, cross-disciplinary Agile teams to align software development with complex business requirements, regulatory standards, and client objectives.
- Delivered high-quality solutions through automated testing frameworks and rigorous QA processes, reducing post-production defects by 25% and ensuring enterprise-grade software reliability.

Trainee Software Engineer

Jan 2020 – Jul 2021

Turito Pvt. Ltd. | YuppTV India Pvt. Ltd.

Hyderabad, India

- Designed and developed a RESTful API using Scala and SQL for a high-traffic educational platform, ensuring scalability to support 100,000+ monthly users.
- Optimized backend processes by integrating with AWS DynamoDB for low-latency data storage and retrieval.
- Crafted SQL queries, designed database schema, and programmed stored procedures to execute essential business functionalities.

PROJECTS

Agentic Grading System | *San Jose State University | Python, CrewAI, Weaviate, Gemini*

Dec 2024 – May 2025

- Built a multimodal agentic RAG system to automate student grading; agents decompose rubrics, retrieve relevant content chunks, and generate justified scores with citation tracing.

Remote Joystick | *Osmania University | ReactJS, Python, TCP*

Nov 2019 – Jan 2020

- Created a Google Stadia-like TCP system application to play high-end games on lightweight hardware with less than 10ms latency using ReactJS and Python.

EDUCATION

San Jose State University

May 2025

Master of Science in Software Engineering | GPA: 3.6 / 4.0

San Jose, CA

- Teaching Assistant: CMPE 132 (Information Security), CMPE 148 (Networking), CMPE 257 (Machine Learning)

Osmania University

Sept 2020

Bachelor of Engineering in Information Technology | GPA: 3.57 / 4.0

Hyderabad, India