Tej Kiran Yenugunti

(408) 581-5806 | tejkiran.yenugunti@sjsu.edu | linkedin.com/in/tejkiran-yenugunti | github.com/tejkiran-yenugunti

Education

Master of Science in Software Engineering **Aug 2025 – May 2027**
San Jose State University, San Jose, CA
Relevant Coursework: Distributed Systems, Data Mining & Machine Learning, Software Testing, Algorithms & Data Structures

Bachelor of Engineering in Computer Science **Jul 2019 – May 2023**
New Horizon College of Engineering, Bengaluru, India

Technical Skills

Languages: Python, Java, Go, JavaScript/TypeScript, SQL, C++, Bash/Shell

Backend & APIs: FastAPI, Flask, Node.js, REST APIs, microservices, PostgreSQL, MongoDB, Redis, async I/O

AI/ML: LLMs, PyTorch, TensorFlow, scikit-learn, LangChain, LangGraph, OpenAI API, RAG pipelines, prompt engineering, AI agents

Cloud & DevOps: AWS (EC2, S3, Lambda), Azure, Docker, Kubernetes, Apache Kafka, RabbitMQ, CI/CD (Jenkins), Prometheus, Grafana, Git

Systems: Distributed systems, concurrency, event-driven architecture, caching, fault tolerance

Experience

Software Engineer, **Deloitte USI** **Apr 2025 – Jun 2025**
- Built and optimized **Python/FastAPI** backend services on a platform serving **500K+ users**; eliminated I/O bottlenecks reducing **p95 latency from 450ms to 270ms (40%)** through systematic debugging and performance optimization.
- Designed and shipped production-grade services with **85%+ test coverage** processing **10K+ requests/day**; maintained reliability through Jenkins CI/CD and automated validation workflows.
- Investigated and resolved **25+ production incidents** using **Prometheus/Grafana**; built automated Python tooling for root-cause detection reducing resolution time by **30%**.

Software Engineer, **Carelon Global Solutions (Elevance Health)** **Jul 2023 – Mar 2025**
- Built **Python/FastAPI** microservices with **Kafka** event-driven pipelines processing **50K+ transactions/day**; optimized **PostgreSQL** query paths by **35%**, enabling scalable low-latency data access at sub-**200ms p95 latency**.
- Engineered **ETL integrations** between Python microservices and legacy IBM DB2 systems; maintained schema consistency and data integrity across **3M+ records** under concurrent production workloads.
- Developed Python-based **data validation and ranking systems** with statistical checks and regression testing; achieved **99.9% data accuracy** while reducing operational errors by **35%**.
- Led the deployment of an internal **AI chatbot** for on-call failure troubleshooting, enabling teams to query solutions and summarize production error logs; decreased incident diagnosis time by **25%**.
- Shipped features through **Jenkins CI/CD** pipelines with post-deployment monitoring; collaborated with product and operations teams in agile sprints to deliver scalable backend services.

Software Engineer Intern, **Secure Machines** **Feb 2023 – May 2023**
- Built **REST API** backend services with **AES-256/SHA-3** encryption serving **20K+ users**; developed real-time ingestion and anomaly detection modules improving responsiveness by **40%**.
- Optimized edge-device data pipelines for continuous sensor stream processing using Python and asynchronous workflows; improved real-time event detection reliability under high-throughput workloads.

Projects

Distributed AI Task Orchestrator **Go, Python, Gemini AI, Kafka, RabbitMQ, Docker, Kubernetes**
- Built a **distributed task execution system** in Go that accepts natural language instructions, uses **Gemini AI** to decompose workflows into ordered subtasks, and dispatches execution through Kafka/RabbitMQ to Python worker nodes on Kubernetes.
- Engineered workload-aware scheduling and self-healing recovery mechanisms with automated health checks and failover; reduced average job completion time by **55%** and rerouted failed tasks within **10 seconds**.
- Instrumented the platform using **Prometheus/Grafana** dashboards for queue depth, worker utilization, and failure monitoring; deployed containerized services through Docker Compose with full observability.

AccessPath AI, Multi-Agent AI Navigation System **LangGraph, FastAPI, NVIDIA Nemotron, Next.js, Vercel**
- Built and shipped to production a **multi-agent LLM system** using **LangGraph** orchestrating **3 NVIDIA Nemotron models** for reasoning, speech processing, and safety validation in real-time accessible route planning.
- Designed a **FastAPI backend** integrating LLM inference, speech-to-text, and safety guardrails; engineered prompt chaining and contextual state management reducing query resolution time by **60%** versus a single-model baseline.

Stratification of Garbage using Deep Learning **CNN, TensorFlow, Keras, scikit-learn, Flask, Python**
- Published peer-reviewed research on **CNN-based multi-class image classification**; designed and trained a custom deep learning model using preprocessing, feature engineering, regularization, and hyperparameter tuning.
- Built an end-to-end **computer vision pipeline** from data ingestion to multi-class inference behind a Flask REST API; demonstrated measurable accuracy improvements over traditional classification approaches.