

Sai Kiran Vepamani

LEAD ENGINEER - GEN AI · FULL STACK DEVELOPER

✉ +91 9493922218 | 📧 vepamanisaikiran@gmail.com | 🏷️ vepamanisaikiran.github.io/portfolio | 🌐 vepamanisaikiran | 💬 saikiranvepamani

Skills

Languages	Python · Java · C++/C · Rust · TypeScript · Node.js · MySQL · PostgreSQL · Bash
AI/GenAI	LLM Fine-Tuning (QLoRA/LoRA/PEFT/RLHF) · RAG/PathRAG · Prompt Engineering · Agentic AI · DSPy · Agno · LangChain · LangSmith · LangGraph · Hugging Face · Vector DBs (FAISS/ChromaDB)
ML/Data	PyTorch · Tensorflow · Transformers · Scikit-learn · OpenCV · Pandas · Numpy · Optuna · MLflow · MLOps
Frameworks	FastAPI · Django · Flask · Spring Boot · Angular 19 · Next.js · React · Flutter · Socket.IO · SQLAlchemy
Cloud/DevOps	AWS (Lambda/CDK/S3/Cognito) · Docker · Kubernetes · Helm · Jenkins · Git · CI/CD · SonarQube · Gradle
Protocols/Tools	MCP · A2A · gRPC · Protobuf · MQTT/mTLS · JWT · REST · WebSocket · VSS · CAN/DBC · Playwright · Postman · Jira

About me

Lead Engineer - Gen AI at Bosch Global Software Technologies with 6+ years of experience spanning Generative AI, LLM fine-tuning, functional safety (FuSA) AI, and cloud-native platform development. Led cross-functional teams building AI-powered products including RAG platforms, HARA (Hazard Analysis and Risk Assessment) AI for safety-critical automotive systems, voice-enabled developer tools, and multi-tenant SaaS infrastructure. Patent co-author, hackathon winner (Bosch AWS 2025, AppsForBharat 2025), and Bharat Mobility Expo 2025 presenter. Passionate about building complex safety-relevant AI systems that create meaningful impact in decision-making for functional requirements.

Work Experience

Bosch Global Software Technologies

LEAD ENGINEER - GEN AI

Bangalore, India

Mar. 2023 - Current

- Leading a cross-functional team of 8+ engineers across Gen AI, automotive middleware, and cloud platform initiatives.
- Built HARA AI, a safety-relevant AI system for Hazard Analysis and Risk Assessment that automates ASIL classifications per ISO 26262.
- Architected BRICK, an advanced document analysis platform featuring PathRAG and MinerU for high-precision retrieval from unstructured PDFs.
- Won Bosch AWS Hackathon 2025 for "Project IQ," achieving 90% accuracy in extracting competencies from SRS documents.
- Performed Parameter-Efficient Fine-Tuning (PEFT) on Qwen 7B+ models using QLoRA/LoRA with 4-bit quantization.
- Architected a Serverless Multi-Tenant SaaS platform using AWS CDK, reducing onboarding time by 90%.
- Developed CAN-to-VSS protocol mapper and optimized KUKSA.val databroker in C++/Rust for low-latency signal orchestration.
- Represented Bosch at Bharat Mobility Expo 2025, showcasing Vehicle Assistant AI for SDV apps on HMI and HPCs.

HashedIn by Deloitte

SDE - II

Bangalore, India

Aug. 2021 - Mar. 2023

- Built cloud-native content normalization platform for Thomson Reuters legal domain using AWS serverless architecture.
- Developed annotation component in Angular that reduced document content processing time from 7 days to 4 hours.
- Architected AWS serverless cloud infrastructure with relational and non-relational database integration.

Zapcom Solutions

SOFTWARE ENGINEER

Bangalore, India

Feb. 2020 - July 2021

- Built REST APIs using Python/Django DRF and implemented Celery-based async task processing.
- Created AWS instances for daily review extractions with NLP scoring using spaCy.
- Wrote automation test scripts in Python using Selenium, reducing manual testing effort by 35%.

Zapcom Solutions

FULL STACK INTERN

Bangalore, India

Jan. 2019 - Jan. 2020

- Developed web scraping pipeline in Java (Selenium) and Python (Scrapy) to extract 1000+ reviews from 540+ client hotels.
- Built scene classification model using Places365 dataset for image recognition.

Achievements & Awards

2025	Patent , Co-Author -- "A control unit for detunneling of data from an ethernet frame" (No: 202541072960)	India
2025	Winner , Bosch AWS Hackathon 2025 -- "Project IQ," AI-powered training & knowledge management system	Bosch
2025	Winner , AppsForBharat 2025 -- Developing impactful solutions for India	National
2025	Presenter , Bharat Mobility Expo 2025 -- Showcased Vehicle Assistant AI that generates SDV apps deployable on HMI and HPCs	Bosch

Education

JNTUA(Jawaharlal Nehru Technological University Anantapuramu)

B.TECH IN COMPUTER SCIENCE AND ENGINEERING WITH 7.74 GPA

Ananthapur, India

Aug. 2015 - May. 2019

Passion Projects

JobsChange.com -- AI Career Acceleration Platform

Co-FOUNDER

India

2025 - Current

- Built an AI-powered career platform with resume tailoring, cover letter generation, and mock interviews using OpenAI GPT-5.
- Engineered full-stack app with Next.js 15, React 19, TypeScript, and Firebase with ATS compatibility scoring.
- Developed Chrome extension supporting 50+ ATS platforms for one-click job application autofill.

MakeDemos.com -- Professional Screen Recording & Demo Software

Co-FOUNDER

India

2025 - Current

- Built desktop screen recording app with auto-zoom, 50+ visual effects, and 4K export capabilities.
- Engineered with Electron 30, React 18, PixiJS 8 with GPU-accelerated rendering via Web Codecs API.
- Developed real-time video effects engine and multi-track timeline editor with frame-accurate trimming.