

Saran Mahadev

Bangalore, India | +91-9360644776 | sarandevnet@gmail.com
saranmahadev.in | [LinkedIn](#) | [GitHub](#)

PROFESSIONAL SUMMARY

Generative AI Engineer with 3+ years of experience in Research & Development building enterprise Computer Vision solutions with AI platforms, Agentic AI systems, Retrieval-Augmented Generation (RAG), LLM orchestration, developer tooling, and scalable backend services.

EXPERIENCE

Associate Software Engineer (R&D) *Zentron Labs Private Limited*

Feb 2023 – Present
Bangalore, India

- Member of the Research & Development team responsible for designing and prototyping next-generation Computer Vision solutions with AI systems, developer platforms, and intelligent automation solutions.
- Architected enterprise Generative AI platforms using LLMs, Agentic AI, LangGraph, MCP, Retrieval-Augmented Generation (RAG), and specification-driven workflows.
- Designed AI-assisted software engineering pipelines for code generation, repository understanding, documentation automation, and process standardization.
- Researched, evaluated, and integrated emerging foundation models, prompting strategies, and orchestration frameworks to build the Model and Behavior Matrix.
- Collaborated with software engineers, QA teams, and product stakeholders to translate research prototypes into production-ready algorithm solutions.
- Developed deterministic Computer Vision algorithms and optimization systems for industrial inspection applications.

Cloud Computing Specialization Trainee *Centre for Development of Advanced Computing (C-DAC)*

Jul 2022 – Aug 2022
Chennai, India

- Gained hands-on experience with OpenStack and Docker, building a foundational understanding of cloud infrastructure, virtualization, and enterprise networking

PROJECTS

LabVIEW to Python AI Migration Pipeline *Python | LangGraph | Gemini | ADK | FastAPI | Docker*

- Designed a multi-stage AI transpilation Pipeline for converting legacy LabVIEW application source code into production-ready Python code.
- Implemented graph-based workflows, structured outputs, and validation loops for reliable code transformation.

Wiki.js AI Documentation Automation *Python | FastAPI | MCP | LangGraph | Wiki.js API | OpenAI SDK*

- Built an AI-powered documentation automation platform integrated with Wiki.js for engineering process standardization.
- Implemented intelligent documentation generation, repository analysis, and standards enforcement using MCP.

.agents Directory Design *OpenSpec | LLM Wiki | Graphify | Claude Code | Git | GitHub Actions*

- Designed an AI native software development template integrating specification-driven development with automated documentation and TDD workflows.
- Built repository-agnostic AI agents, skills and hooks for planning, implementation, testing and documentation.
- Integrated LLM Wiki, Graphify - repository graph generation, and GitHub Action workflows for Review.

Stride — Spec-Driven Agent Runtime

Python | Claude Code SDK

- Developing an open-source runtime for deterministic specification-driven AI agent execution with planning, validation, and execution workflows.
- Designing modular agent orchestration, context engineering, checkpointing, and execution pipelines.

Axon — Enterprise Memory SDK

Python | Pinecone | ChromaDB | Weaviate | FAISS | Qdrant

- Built a framework-agnostic memory abstraction layer for AI applications supporting semantic retrieval across multiple vector databases.
- Designed APIs with framework agnostic desing enabling seamless integration with LangChain, LlamaIndex, and enterprise RAG systems.

LifeOS — AI Productivity Platform

Golang | SQLite

- Designed a repository-based personal operating system powered by AI agents for knowledge management, productivity, and life management.
- Implemented native CLI tooling, structured data management, and AI-driven workflows for progressive discovery.

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript, C++ , Golang

Backend: FastAPI, Flask, REST APIs, Streamlit

Generative AI: Model Awareness (Claude, Gemini, OpenAI, GLM, Kimi, Qwen) Prompt Engineering, Context Engineering, Prompt Chaining

LLM Frameworks: LangChain, LangGraph, LlamaIndex, CrewAI, AutoGen, Semantic Kernel, Hugging Face Transformers, OpenAI SDK

Agentic AI: AI Agents, Multi-Agent Systems, MCP, Tool Calling, Workflow Orchestration, Memory Management

RAG: Enterprise RAG, Hybrid Search, Semantic Search, Embeddings, Chunking, Semantic Chunking, Cross Encoder Reranking

Vector Databases: Pinecone, ChromaDB, FAISS, Weaviate, Qdrant

Machine Learning: PyTorch, TensorFlow, Model Training, Fine-tuning, Transfer Learning, Model Evaluation

Cloud & DevOps: AWS and Google Cloud Platform, Docker, Kubernetes, Git, GitHub Actions, CI/CD

Databases: PostgreSQL, MySQL, SQLite, MongoDB, Redis

LLMOps: LangSmith, Langfuse, MLflow, Observability, Monitoring, Logging, Prompt Versioning, Model Versioning

Computer Vision: OpenCV and Image Processing Algorithm Design

EDUCATION

Government College of Engineering Srirangam

2019 – 2023

Bachelor of Engineering — Computer Science and Engineering

Trichy, India

- Department Representative for four consecutive years.
- Founded GCES Coderz Club and mentored over 150 engineering students.
- Organized technical workshops, hackathons, and national-level engineering events.

CERTIFICATIONS & ACHIEVEMENTS

- **Amazon:** AWS Certified Cloud Practitioner
- **DeepLearning.AI:** AI for Everyone
- **Python Foundation:** Certified Associate in Python Programming
- **NDG:** Linux Unhatched