

Jeeva M

Chennai, Tamil Nadu, India • +91 9445737120 • jeeva4772@gmail.com • [LinkedIn](#) • [GitHub](#)

Education

Vellore Institute of Technology, Chennai

Expected June 2028

Integrated M.Tech in Software Engineering – CGPA: **9.07/10.00**

Technical Skills

Languages: Python, Java, C++, JavaScript, TypeScript, SQL

Backend: FastAPI, REST APIs, WebSockets, Microservices

Databases: PostgreSQL, pgvector, Redis, ChromaDB, SQLite

Cloud & DevOps: Docker, Kubernetes, Terraform, AWS, AWS Bedrock, GCP, Git, Linux

Frontend: React, Next.js, Tailwind CSS

ML / AI: PyTorch, YOLOv8, LangChain, LangGraph, OpenCV, Scikit-learn

Experience

Zevaras

May 2026 – July 2026

Software Development Engineer Intern

Chennai, India

- Built and deployed a FastAPI-based RAG platform serving an internal legal knowledge base; supported ingestion across 4 document formats (PDF, DOCX, TXT, Markdown) with per-conversation retrieval isolation.
- Combined dense vector search, BM25 keyword search, Reciprocal Rank Fusion, and a BGE cross-encoder reranker to surface more relevant passages to the LLM before generation.
- Designed a provider interface that allowed switching between 3 LLM providers (**Amazon Nova, Claude, DeepSeek**), embedding models, and vector stores without modifying application logic.
- Prototyped on ChromaDB, then migrated to **PostgreSQL (pgvector)** for the production deployment, cutting the number of separate data services to operate.

Indian Institute of Science (IISc), Bangalore

May 2025 – June 2025

AI/ML Research Intern, M.S. Bobji Force Microscopy Lab

Bangalore, India

- Built and trained a YOLOv8 Nano segmentation model (3.1M parameters) to detect and outline individual leaves in field images for edge-deployed precision agriculture; collected 358 field images and annotated 4,812 leaf polygons via Roboflow.
- Trained the model for 100 epochs, reaching **0.674 mAP@0.5** and **0.543 mAP@0.5:0.95**, and evaluated robustness under leaf overlap, occlusion, and varying field illumination.
- Reduced inference time to **78ms** on a Qualcomm Snapdragon 855, enabling near real-time deployment for drone-based crop monitoring.

Projects

ResearchOS – AI-Powered Research Platform

[GitHub](#)

- Built a platform for managing experiments, literature, notebooks, and AI-assisted paper writing, unifying research workflows behind **FastAPI** and **Next.js**.
- Built a hybrid search engine combining pgvector, BM25, and trigram similarity with Reciprocal Rank Fusion to support semantic, keyword, and fuzzy search across research documents.
- Implemented asynchronous document indexing using **Redis Streams** and PostgreSQL, with JWT-based authentication for multi-tenant isolation.
- Deployed via Docker Compose with **Kubernetes** and **Terraform**-managed infrastructure, monitored using Prometheus and Grafana.

ForgeMCU Studio – Firmware Generation Platform

[GitHub](#)

- Generated firmware projects for **STM32, ESP32, and RP2040** from natural-language specifications using **LangGraph**-based agents that handle architecture, code generation, testing, and build verification.
- Grounded generation in embedded systems documentation through a RAG pipeline to reduce incorrect or outdated API usage; built a React/FastAPI web app with WebSockets for live build status.

Achievements & Leadership

- **First Prize**, HCLTech Embedded Systems Hackathon
- **Merit Student Award**, Vellore Institute of Technology – Awarded Twice for Academic Excellence
- Coordinator, Frontend Kickstart – Office of Student Welfare, ACM VIT Chapter