# SOWNDAPPAN S

**AI ENGINEER**

Salem, India | +91 7200339151 | sowndappan610@gmail.com | GitHub | LinkedIn | Portfolio

## PROFESSIONAL SUMMARY

AI Engineer specializing in agentic AI, RAG pipelines, and LLM fine-tuning, with hands-on experience building and deploying production AI systems using LangGraph, FastAPI, and AWS. Experienced in LLM orchestration, structured-output systems, and PEFT-based model fine-tuning.

## SKILLS

**Programming & Data:** Python, SQL, NumPy, Pandas, Scikit-learn, Data Structures & Algorithms, OOP

**GenAI:** LangChain, LangGraph, RAG, Prompt Engineering, Agentic AI

**LLM Fine-Tuning:** PyTorch, Transformers, PEFT, LoRA/QLoRA

**Backend & Infrastructure:** FastAPI, Flask, REST API Design, Pydantic, PostgreSQL, ChromaDB

**Cloud & DevOps:** AWS Lambda, CloudWatch, Git, Linux

## PROFESSIONAL EXPERIENCE

**Data Science Intern (AI/LLM Focus) — Logesys Solutions India Pvt. Ltd** *Sep 2025 – August 2026 | Bengaluru, India*

- Built LLM-backed question generation and explanation services using LangChain, LangGraph, and Gemini API, generating 1,000+ assessment questions and cutting manual curriculum effort by 40%
- Architected and deployed 10+ production AI APIs on AWS Lambda, handling 500+ daily requests across assessment, recommendation, and analytics workloads for 10,000+ learner interactions
- Designed normalized PostgreSQL schemas and modular FastAPI service layers enabling independent scaling of each AI-driven feature
- Implemented production monitoring and alerting via AWS CloudWatch, reducing time-to-detect for service degradation across the API layer

## KEY PROJECTS

**AdaptBot — Agentic AI Orchestration Backend** · *Python, LangGraph, LangChain, Pydantic, Groq API* [GitHub]

- Architected a self-extending AI agent with **50+ runtime-generated tools** using a static 4-node LangGraph pipeline and dynamic tool loading
- Achieved **93.2% intent classification** and **100% tool-execution success** through structured Pydantic extraction and runtime tool registration.

**Production RAG System — Document Intelligence API** · *Python, FastAPI, ChromaDB, Sentence Transformers, Groq API* [GitHub]

- Benchmarked **4 chunking strategies** across 100+ questions and 1,000+ indexed chunks, improving retrieval coverage from **58% to 68%** with sub-second latency.
- Built a FastAPI + ChromaDB evaluation system supporting **PDF, DOCX, and TXT** ingestion.

**Commit Message Generation API — Fine-Tuned LLM** · *Python, PyTorch, Transformers, PEFT, TRL, FastAPI* [GitHub]

- Fine-tuned **Qwen2.5-1.5B-Instruct** on 10,000 GitHub commit messages using **LoRA + 4-bit NF4 quantization**
- Reduced trainable parameters by **~98%** compared with full fine-tuning, enabling training on a **single T4 GPU**.

## OPEN-SOURCE CONTRIBUTIONS

Merged PRs to **DeepSpeed (Microsoft), LangChain, Caura AI**, and **Recursion Pharma** — covering quantization parsing, structured-output validation, memory-enrichment fallback, and graph-processing optimization.

## PUBLICATION

**Monocular Object Distance Estimation using Calibration-Based Perspective Mapping and Detection Integration**

*IRJET, Vol. 13, Issue 04, Apr 2026* · 98% interval accuracy, 0.36 m MAE, O(1) CPU-only real-time inference

## EDUCATION

**B.Tech in Information Technology** **CGPA: 8.93 / 10**

*Knowledge Institute of Technology, Salem · Nov 2022 – May 2026*

## CERTIFICATIONS

Microsoft AI & ML Engineering — Professional Certificate · [Credential]