

## DHARANI ESWARAMURTHI

AI Engineer | Generative AI | RAG | Agentic Systems | Cloud ML

Coimbatore, Tamil Nadu, India | +91-9655612306 | dharani96556@gmail.com

linkedin.com/in/dharani-eswaramurthi | github.com/Dharani-Eswaramurthi | medium.com/@wordsbydharani

### PROFESSIONAL SUMMARY

AI Engineer with 2+ years of professional experience building and shipping production GenAI, RAG, speech AI, and cloud ML systems across AWS, Azure, and GCP. Currently Lead AI Engineer at aXtrLabs after joining as Founding AI Engineer, with hands-on ownership across LLM orchestration, retrieval, backend APIs, cost/latency optimization, and team delivery.

### TECHNICAL SKILLS

**AI / LLM:** Python, GPT-4/4o, Claude, Gemini, Mistral, LangChain, LangGraph, RAG, agentic workflows, prompt engineering, structured outputs,

**Retrieval / Speech:** Pinecone, FAISS, pgvector, embeddings, hybrid search, re-ranking; Whisper, Soniox STT, ElevenLabs TTS

**Backend / Cloud:** FastAPI, Django, Flask, React, TypeScript, REST APIs, WebSockets; AWS Bedrock/SageMaker/EC2/Lambda/S3, Azure OpenAI, GCP Vertex AI, Docker

**ML / Data:** PyTorch, TensorFlow, Hugging Face Transformers, Pandas, NumPy; PostgreSQL, MySQL, MongoDB, Redis

### PROFESSIONAL EXPERIENCE

**aXtrLabs** | Coimbatore, India

**Lead AI Engineer** | Jan 2026 - Present

- Lead architecture and delivery of enterprise GenAI products, including an RFP proposal automation platform using LLMs, RAG, vector search, and automated Word/PDF/PPT generation.
- Set technical roadmap and architecture decisions; standardized prompt-engineering and AI-assisted development workflows, and mentor engineers on production RAG, LLM evaluation, and MLOps.

**Founding AI Engineer** | May 2025 - Dec 2025

- Built a real-time pre-sales and sales call evaluation system using Soniox STT and AWS SageMaker, increasing audit coverage from 30% manual review to 100% automated, evidence-backed scoring.
- Engineered enterprise RAG pipelines with hybrid retrieval and LLM re-ranking, improving retrieval precision by ~35%; reduced LLM API cost by 40% using prompt caching, token compression, and model routing.
- Developed and deployed a trade payload generation service with mathematical optimization, reducing processing time 30% and error rate 20% while maintaining 99.9% uptime on AWS EC2.

**Heuro** | Remote

**Freelance Full-Stack Developer** | Sep 2024 - Apr 2025

- Built and deployed a React + FastAPI conversational AI product using RAG and OpenAI APIs; integrated Slack workflows and context-aware recommendation logic.

**goML** | Coimbatore, India

**ML Engineer** | Jun 2024 - Sep 2024

**AI/ML Engineer Intern** | Jun 2023 - Jun 2024

- Built classification and regression pipelines with PyTorch and scikit-learn and deployed models on AWS SageMaker; contributed prompt engineering and output-validation layers for LLM-backed services.
- Developed a real-time voice feedback application using Whisper STT, emotion detection, and Google TTS; prototyped sales AI and digital-twin systems.

### SELECTED VENTURE

**Greenvy.store - Founder & Full-Stack Builder** | 2025

- Built a custom React + FastAPI e-commerce application with MongoDB, Google Cloud Storage, and Razorpay integration for Greenvy; registered the venture as a Udyam micro enterprise and launched the store publicly.

### EDUCATION & CERTIFICATIONS

**B.Tech, Artificial Intelligence & Data Science** - Karpagam College of Engineering, Coimbatore | 2020 - 2024 | CGPA: 8.58/10

**Certifications:** AWS Certified Machine Learning - Specialty (2024) | AWS Certified Cloud Practitioner (2023)