

VARADARAJ ACHARYA

FULL-STACK AI ENGINEER

varadaraj277@gmail.com | 8722593540 | www.varada.me | GitHub | LinkedIn

PROFESSIONAL SUMMARY

Full-Stack AI Engineer shipping production LLM systems: Retrieval-Augmented Generation (RAG), embeddings, semantic search, vector databases, prompt engineering, and agent-style tool calling - plus React/TypeScript and FastAPI fullstack delivery.

Built end-to-end products on React, TypeScript, FastAPI, PostgreSQL, MongoDB, and Redis, including local-first AI with on-device embeddings.

TECHNICAL SKILLS

Frontend: React, Next.js, TypeScript, Redux, RTK Query, React Native

Data: PostgreSQL, MongoDB, Redis, SQL Server, LanceDB, Qdrant

Backend: FastAPI, Node.js, Python, Django, Docker, Jenkins, CI/CD

Gen AI: LLMs, OpenAI, Claude, Gemini, Retrieval-Augmented Generation (RAG), Prompt Engineering, Embeddings, Semantic Search, Vector Search, Reranking, Structured Outputs, Multimodal

PROFESSIONAL EXPERIENCE

Software Engineer (Full-Stack AI) @ PrimeSophic Technologies

July 2024 – Present

- Shipped a production ERP and health platform serving 500+ active users across Client, Doctor, and Admin roles.
- Delivered role-based UIs in React, TypeScript, Redux Toolkit, and RTK Query for consistent multi-role product workflows.
- Owned FastAPI backends with JWT auth, PostgreSQL, and MongoDB for authenticated end-to-end product flows.
- Introduced Redis caching on hot read paths to cut database load and improve API responsiveness.
- Integrated Razorpay, Terra, and Oura so payments and wearable health data flow into the product without manual ops.

AI & LLM Work

- Shipped production Retrieval-Augmented Generation (RAG) chat systems with OpenAI: embed domain docs, retrieve top-k context, ground answers before generation, and harden prompt/retrieval pipelines to reduce hallucinations; fine-tuned models for JSON structured outputs.
- Designed multi-step LLM agent workflows with tool calling for task automation beyond single-shot prompts.
- Built a FastAPI semantic search API using OpenAI embeddings and Qdrant for catalog and customer search, with multi-stage retrieval (exact SKU/UPC/part match in SQL Server, dense vector recall, FlashRank cross-encoder reranking), purchase-history ranking boosts, indexing pipelines, and offline search eval/feedback loops.
- Shipped a natural-language cart API with GPT-4o-mini structured outputs to parse qty/UOM phrases and resolve reorder-from-last-invoice flows.
- Built a FastAPI GenAI document-generation service on Claude and OpenAI with prompt-controlled tone/length/strategy, multi-collection Qdrant RAG grounding, draft-then-critique refine loops, SQL Server context lookup, and multimodal image/PDF ingestion.

SELECTED PROJECTS

Aether - aether.varada.me

2026

- Shipped local semantic search for macOS files by meaning (not filename only), with no cloud upload of file contents.
- Indexed on-device embeddings into LanceDB for private, offline-first vector search on local indexes.
- Stack: SwiftUI, Python, embeddings, LanceDB.

Vamshio - vamshio.com

2025

- Shipped a Django full-stack platform for heritage records, finances, receipts, and member contributions used by Kerala families and organisations.
- Stack: Django, Tailwind.

EventURox - eventurox.in

2025

- Shipped a Next.js event brand site tuned for first impression, motion, SEO, and lead conversion.
- Stack: Next.js, Three.js.

Designer Portfolio - designport.vercel.app

2025

- Delivered a Next.js designer portfolio with WebGL accents under a tight performance budget.
- Stack: Next.js, Three.js.

EDUCATION

Bachelor of Computer Applications (BCA), MGM College, Udupi

2021 – 2024