

RITHWIK SAI POSANI

FULL STACK DEVELOPER

Hyderabad, India • +91 80965 34845 • posanirithwik@gmail.com

linkedin.com/in/rithwikposani • github.com/rithwik1510

SUMMARY

Full-stack developer focused on building production-grade web applications with strong UX and AI-assisted workflows. Experienced in shipping end-to-end products using React/Next.js and Fast API, integrating LLMs responsibly via backend services, and iterating post deployment. Comfortable using AI tools to accelerate development while owning architecture, reliability, and final quality.

EDUCATION

Bachelor of Technology in Computer Science

SRM Institute of Science and Technology, Chennai | Aug 2021 – Jun 2025 | CGPA: 8.13 / 10

TECHNICAL SKILLS

AI / ML: RAG Pipelines, Large Language Models (LLMs), Natural Language Processing (NLP), Computer Vision, Vector Search (FAISS), Prompt Engineering, Model Evaluation, Hybrid Retrieval (BM25 + Semantic)

Languages: Python, TypeScript, C, HTML, CSS

Frameworks & Libraries: FastAPI, React, Next.js, Tailwind CSS, HTMX, Pandas, NumPy, Scikit-learn

Databases & Tools: PostgreSQL, MongoDB, SQLAlchemy, Prisma, Docker, Git, GitHub Actions, Ollama

APIs & Platforms: OpenAI API, Google Gemini API, REST API Design, Vercel, CI/CD Pipelines

PROJECTS

NestAi — RAG-Powered Document Intelligence System

Python, FastAPI, PostgreSQL, Ollama, BM25, Vector Embeddings, HTMX

Engineered a hybrid retrieval pipeline (BM25 + vector similarity) with strict RAG citations, achieving sub-second latency across 100+ documents.

Improved response accuracy by ~40% by designing a cite-or-abstain policy that eliminates hallucinations and enables full traceability.

Integrated local LLM inference via Ollama for privacy-first document Q&A with zero data leaving the device.

FridgeChef — Computer Vision + Generative AI Platform

Next.js 14, FastAPI, SQLAlchemy 2.0, PostgreSQL, Google Gemini 2.5 Flash, React Query

Built a multi-modal AI pipeline using Gemini 2.5 Flash to detect, categorize, and estimate quantities from fridge images in real time.

Automated recipe generation by feeding detected ingredients into a generative model with dietary and preference constraints.

Deployed full-stack app to Vercel with 78% backend test coverage, rate-limiting, and request logging for observability.

AI Financial Advisor — LLM-Powered Analytics & Coaching

Python, FastAPI, React, Tailwind CSS, OpenAI API

Developed an analytics engine that parses CSV/XLSX bank statements and computes 10+ financial metrics (savings rate, DTI, emergency fund).

Integrated an OpenAI-compatible chat endpoint that converts computed analytics into personalized recommendations and what-if scenarios.

Exposed REST APIs with health checks, enabling local and cloud deployment with automated testing via Pytest.

CERTIFICATIONS

AWS Academy Machine Learning Foundations — Amazon Web Services

Supervised Machine Learning: Regression & Classification — Coursera (Stanford University)

Networking Basics — Cisco Networking Academy