

Aaron Jacob Sunil

Kuwait(Open to Relocation) — 📞 +965 66469426 — aaronjacobsunil@gmail.com

🌐 linkedin.com/in/aaron-jacob-sunil — 🐙 github.com/4aruu

Professional Summary

Full-Stack/AI Engineer with hands-on experience building LLM/RAG applications, deploying vector databases, and managing cloud infrastructure. Proven track record in developing end-to-end GenAI architectures, RESTful microservices, and optimizing system latency. Seeking to leverage expertise in machine learning workflows, Agentic AI, and cloud technology to deliver high-impact data science solutions.

Academics

APJ Abdul Kalam Technological University (KTU)

B.Tech in Information Technology

First Class, Principal's List- S1,S7,S8

Kottayam, India

2022 – 2026

CBSE Board

Higher Secondary Education

Kuwait

2022

Internship

Developer Intern - Sreenarayanaguru Open University (Cybercell)

May 2025 – Jun 2025

Kollam, Kerala

Built and deployed a bilingual (English/Malayalam) chatbot for prospective student inquiries, accessible via the university's public portal.

Improved backend response times by implementing query caching and optimizing database access patterns.

Integrated JWT-based authentication modules in close coordination with the university central IT support team.

Projects

Lumira – Interactive Multimodal AI Assistant

Nov 2025 – Apr 2026

ChromaDB, FastAPI, React, LangChain, LLaMA 3.2, Faster-Whisper

Engineered a local RAG pipeline using LangChain and ChromaDB vector store to deliver context-aware response generation across multi-domain knowledge bases.

Integrated Faster-Whisper for real-time speech recognition and Edge-TTS for audio synthesis, enabling low-latency voice-driven interaction.

Optimized chunking strategies, embedding selections, and async FastAPI endpoints to reduce retrieval latency and achieve 92% semantic relevance on evaluation test sets.

Nutriswap – Intelligent Food Alternative Finder | [food-alternate.vercel.app](#)

Dec 2025 – Apr 2026

React, FastAPI, Docker, PostgreSQL, Tailwind CSS, Render, Aiven

Built a full-stack health application with 231 curated Kerala-cuisine food alternatives, searchable by nutritional profile and allergen flags across Diabetic, Heart-Safe, and Weight Loss filters.

Designed client-side filtering architecture to eliminate per-filter API calls, enabling instant search responses across 231 items after a single 14.8kb payload fetch.

Fully containerized backend/frontend microservices using **Docker** and provisioned deployment pipelines across multi-cloud infrastructure (Vercel, Render, Aiven) with automated health checks.

InfoBot – AI Chat Assistant for Campus Management

Feb 2025 – Apr 2025

Python, React, FastAPI, MySQL

Built an automated timetable generation module that reduced multi-staff manual scheduling from weeks of coordination to an automated process completing in under 2 minutes.

Handled 50+ student queries on timetables and grades through a React frontend backed by FastAPI and managed MySQL instances.

Designed async REST API endpoints in FastAPI to support concurrent user sessions without server performance degradation.

Technical Expertise

AI/ML: RAG, LangChain, ChromaDB, Ollama (LLaMA 3.2), Faster-Whisper, Prompt Engineering, Vector Search

Languages/ Frameworks: Python, JavaScript, SQL, FastAPI, React, Node.js, Django, HTML/CSS

Databases/Cloud: PostgreSQL, MySQL, MongoDB, ChromaDB, Vercel, Render, Aiven, AWS (SAA-C03 in prep)

DevOps/Tools: Docker, Git, GitHub, REST APIs, Postman, JWT Authentication, System Administration

Spoken Languages: English (Professional), Malayalam (Native), Hindi.

Certifications

DeepLearning.AI: **ChatGPT Prompt Engineering for Developers.**

AWS: **AWS Solutions Architect – Associate (SAA-C03)** (In progress)