KARAN RAWAT

karanrawat.cse@gmail.com +91-9672618163 LinkedIn Github Coding Profile

Experience

Think41 Feb 2025 – Jun 2026

Software Engineer

- **Built 10+ production backend modules** in 5 months for a **healthcare B2B/B2C application**, delivering **gamification**, **notifications**, **deal-of-the-day**, **combo offers**, **REST APIs**, **database schemas**, and **React Native/PWA** frontends **at scale**.
- **Implemented** backend integrations including **AWS S3**, **OCR-based license extraction**, **QR-code onboarding**, **OTP authentication**, and **WhatsApp Business API** notifications, improving **application workflows**, **user onboarding**, and **system reliability**.
- **Developed** features for an **internal multi-tenant AI platform** used by **70+ employees**, building a **CRM proof of concept**, **leave and timesheet modules**, **Google Calendar MCP** integration, and **real-time AI chat** with **streaming LLM responses**.
- Engineered **real-time AI chat** using **WebSockets** and **streaming LLMs**, built **MCP server integrations** for **Google Drive, Gmail, and Chat** that let **AI agents** act on **Google Workspace** data, plus a **PDF-to-LLM** document conversation pipeline.
- **Shipped** a production **real-time voice AI agent** using **Pipecat**, **Deepgram STT**, **ElevenLabs TTS**, and **AWS EC2**, reducing **end-to-end latency by 40% (5s→3s)** through **EarlyFlush**, **VAD tuning**, **sentence splitting**, and **parallel STT processing**.

Skills

- **Languages:** Python, TypeScript, C++, SQL, JavaScript
- **Frameworks:** FastAPI, Django, React, SQLAlchemy, Node.js
- **Tools:** Git, GitHub, Docker, Postman, GitHub Actions, pytest
- **Platforms:** AWS (EC2, S3) Linux
- **AI / GenAI:** LLMs, Agentic AI, MCP Protocol, RAG, Prompt Engineering, STT/TTS, Voice Pipelines
- **Backend:** REST APIs, WebSockets, Server-Sent Events (SSE), Microservices, PostgreSQL, MongoDB, SQLite, System Design

Education

SRM Institute of Science and Technology, Kattankulathur Campus Aug 2021 – Jun 2025

B.Tech in Computer Science; *CGPA: 9.21* *Chennai*

Projects

RefundAI — AI Refund Support Agent Jun 2026 – Jun 2026

- **Built** an **AI agent** that approves, denies, or escalates refunds against a strict policy, with all hard rules enforced in **deterministic code** (not the LLM) as the single source of truth, plus **prompt injection defense** at both layers.

Tech Stack: Python, FastAPI, PostgreSQL, React, TypeScript, LLMs, Agentic AI, SSE

Real-Time Multiplayer Game Sep 2025 – Oct 2025

- **Built** a **full-stack real-time multiplayer game** with **WebSocket** game-state synchronization, in-match chat, and a live leaderboard, deployed to the cloud with low-latency updates synchronized across connected players in real time.

Tech Stack: React, Node.js, Socket.io

StockPilot — AI Inventory Dashboard Feb 2025 – Mar 2025

- **Developed** an **LLM-powered inventory dashboard** that transforms uploaded sales data (CSV) into **AI-generated analytics**, demand forecasts, and restock recommendations, visualized through interactive charts for smarter inventory decisions.

Certificate

- Oracle Cloud Infrastructure 2024 Generative AI Certified Professional
- Oracle Cloud Infrastructure Foundations Associate

Achievements

- IP India Patent & IEEE Publication — Real-Time Driver Alertness Monitoring
- Runner-up — Dizijest 1.0 Hackathon