

RAHUL SONAR

@ rahulsonarmain@gmail.com +91 7359637637 Surat, India

 [LinkedIn](#)
 [GitHub](#)
 [LeetCode](#)
 [Portfolio](#)

EDUCATION

Master of Technology in Information Technology

Parul University, Vadodara
July 2021 – May 2023
CGPA: 8.21 / 10

PUBLICATION

- i. Stock trade recommendations using knowledge graph and graph neural network - [Link](#)
- ii. Stock Trade Recommendations Using Knowledge Graph - [Link](#)

CERTIFICATION

- i. Oracle Cloud Infrastructure 2025 Certified Generative AI Professional (1Z0-1127-25), Oct 2025 - [Link](#)
- ii. Oracle Cloud Infrastructure 2025 Certified Data Science Professional (1Z0-1110-25), Oct 2025 - [Link](#)

SKILL

Languages

Python JavaScript SQL
NumPy Pandas Bash

Backend & Databases

FastAPI Django Node
REST APIs React MySQL
PostgreSQL Redis

AI/ML

PyTorch TensorFlow
Open-weight models SDK
HuggingFace Transformers

Deployment & Serving

Docker Kubernetes AWS
Cloudflare vLLM Ollama

EXPERIENCE

AI Engineer

Unitusk Technologies

Jun 2023 – Present
Vadodara, India

- Engineered and deployed, client-facing AI applications, covering backend architecture, database design, and production deployment across cloud environments.
- Engineered a full-duplex audio pipeline replacing sequential STT, LLM, TTS with simultaneous bidirectional inference, reducing response latency from 1.2s to 200ms for a conversational AI product.
- Tech stack: Python, SQL, API, GenAI, LLMs, AI Agents, FastAPI, Docker

Software Developer

Emptai IT Services

Jun 2019 – May 2021
Surat, India

- Worked on building key modules of a POS system, contributed to the billing flow, product listing, and inventory screens. and shipped features that went live in production.
- Contributed to building offline billing support, local SQLite stored transactions when internet dropped, synced back to PostgreSQL on reconnect, so the cashier never had to stop mid-rush because of a network issue.
- Tech stack: JavaScript, React.js, Redux Persist, SQLite, Service Workers, Node.js, PostgreSQL, REST API, Redis, Docker, AWS

PROJECTS

LLM Studio – Full-stack multimodal LLM chat platform

- Built end-to-end AI chat app with React frontend and Cloudflare Workers serverless backend, deployed globally via Cloudflare Pages with <50ms edge latency and zero cold starts.
- Integrated 25+ open-weight LLMs (Llama, Mistral, CodeLlama) with dynamic model switching and a multimodal STT pipeline delivering sub-100ms voice-to-response latency.
- Tech: React, Vite, JavaScript, Cloudflare Workers, REST API
- Live: <https://llmstudio.pages.dev> | [GitHub](#)

FormsMain – Production form submission and data collection platform

- Built serverless form platform on Cloudflare Workers + D1 SQLite with Google OAuth, role-based access control, multi-tenant architecture, and RESTful API for form CRUD, submission handling, and multi-channel embeds (HTML/iframe/AJAX).
- Deployed on Cloudflare edge with auto-scaling, email notifications via Resend API, admin dashboard, CSV export, and shareable public form links requiring no signup.
- Tech: Cloudflare Workers, D1 SQLite, Resend API, JavaScript
- Live: <https://tradesup.in> | [GitHub](#)