

Kirtan Tank

AI & Backend Engineer

Surat, India | tankkirtan21@gmail.com | +91 82009 85811
kirtantank.in | linkedin.com/in/kirtan-tank | github.com/Kirtan-Tank | huggingface.co/knt21

AI engineer who takes retrieval systems, agents and guardrails into production, and builds the backends that hold them up. Two years of **end-to-end ownership, from prototype to deployment.**

EXPERIENCE

AI & Full-Stack Engineer at RedSoft Solutions Pvt. Ltd. Apr 2024 – Present
Promoted from intern to engineer · Remote

- **MosaicAI** US REAL ESTATE · Human-in-the-loop **agentic RAG** over live property APIs and documents, giving ~25% better query accuracy than keyword search. LangChain, LlamaIndex, FAISS, GPT-4o
- **Svaaya** SKIN & HAIR CLINIC · **HIPAA-compliant** clinical image analysis with consent tokens, async job tracking and prompt guardrails; benchmarked GPT-4o, Gemini, Groq and Claude to pick the production model. FastAPI
- **Candore IVF** FERTILITY CLINIC · Summarises staged IVF lab reports for clinicians, cutting manual review ~70% while blocking unsupported conclusions. React, FastAPI, Claude API
- **Decode Age** LONGEVITY E-COMMERCE · Frappe/ERPNext back office syncing orders from **Amazon, Flipkart, Myntra, Tata Cliq and Shopify**, removing ~15 hrs/week of reconciliation and ~40% of sync errors.
- **Watlow** MALAYSIA · Claude-powered invoice extraction into Microsoft SharePoint and Excel, cutting manual entry ~80%. n8n, Claude API
- **CarDecode** WORKSHOP SAAS · Led end-to-end delivery: invoicing, approval workflow and owner analytics, used daily across multiple locations at ~99.5% uptime. Next.js 14, Prisma, PostgreSQL, Docker
- **Quality Asia, Prime Pneumatics, Maharana Foods** ISO certification portals, a field-service app with RLS-based RBAC and WhatsApp reminders (~50% faster reports), and a Frappe back office.
- **Platform** Own deployments across **Vercel, Frappe Cloud and Linux VPS** and run weekly disaster-recovery backups with **zero data-loss incidents to date.**

Data Science Intern at Nektar P.L. Jan – Apr 2024
Mumbai, Maharashtra

- Sole developer of an educational **RAG chatbot**, improving answer relevance ~30% over a naive retrieval baseline. Hugging Face Transformers, LangChain, LlamaIndex
- Ran a **quantized Zephyr-7B** after evaluating GPTQ, AWQ and QAT (~65% less inference memory) and benchmarked FAISS, Milvus and Chroma for latency and recall; shipped behind FastAPI and Flask endpoints.

SKILLS

Generative AI	RAG, agentic workflows, human-in-the-loop, prompt engineering, guardrails, evals, quantization
LLM Tooling	LangChain, LlamaIndex, smolagents, HF Transformers, n8n; FAISS, Pinecone, Chroma, BGE/GTE
Backend & Web	Python, TypeScript, SQL, FastAPI, Node.js, Next.js, React, PostgreSQL, Prisma, Frappe/ERPNext
DevOps	Docker, Nginx, PM2, Vercel, Frappe Cloud, VPS/Linux, GitHub Actions, disaster-recovery backups

SELECTED PROJECTS

GAIA Agent, multimodal agent for the GAIA benchmark: reasoning, web search, transcription. **LIVE DEMO**
Event Planner AI, a smolagents CodeAgent for event planning, traced with OpenTelemetry and LangFuse. **GITHUB**

EDUCATION

B.E. Computer Science & Engineering, Gujarat Technological University (RNGPIT) 2020 – 2024
CGPA 9.37/10 · **University Gold Medal** (AI/ML minor topper) · 1st Rank, Semester 5

CERTIFICATIONS

LLM Engineering, Udemy (*ongoing*) · Hugging Face Agentic AI & MCP · Python for Data Science, NPTEL (*Top 5%*)