

Sudhir

AI Engineer

☎ +91 8512854727 ✉ sudhirkr638@gmail.com in LinkedIn 🌐 GitHub 📍 Bengaluru, India

SKILLS

AI Engineering: Advanced RAG, AI Agents Multi-Agent Systems (LangChain, LangGraph), Fine-Tuning, Gen AI, Prompt Engineering, RAGAS, LangSmith, HuggingFace, Ollama

Backend: FastAPI, Node.js (v20), Nest.js (v10), Express.js, TypeScript

Frontend: React (v18), Next.js (v14), Redux, TypeScript, JavaScript, HTML/CSS

Databases: PostgreSQL, MongoDB, Redis, Vector DBs (pgvector, ChromaDB, Pinecone)

Tools & Platforms: AWS (EC2, S3, Amplify), Git, GitHub, Netlify, Postman, RESTful APIs

PROJECTS

RedAgent (RAG-as-a-Service Platform) — 🌐

- Architected a secure, **multi-tenant RAG-as-a-Service platform** using a **schema-per-tenant** model in **PostgreSQL** and **pgvector**, ensuring strict **data isolation** and supporting scalability to **100+ tenants without cross-tenant leakage**.
- Built an automated onboarding system that **dynamically provisioned unique subdomains** for clients and enforced fine-grained, query-level **RBAC**, reducing manual onboarding effort by **90%** and cutting setup time from **3 hours to under 10 minutes**.
- Deployed the architecture on **AWS**, engineering a multimodal ingestion pipeline from **S3** and leveraging **LangChain** to orchestrate the RAG workflow, boosting retrieval efficiency by **80%**, cutting latency to **<2 seconds**.
- Tech Stack:** FastAPI, **LangChain**, React, Next.js, PostgreSQL, pgvector.

DealSensei (AI-Powered CRM) — 🌐

- Developed a full-stack, **AI-CRM** platform to boost sales effectiveness, guaranteeing strict **data isolation** between client companies and scaling seamlessly.
- Engineered a suite of AI tools, including a **Deal Coach** that applies **sentiment analysis** on customer interactions to deliver actionable insights, improving sales win-rates by **25%**.
- Implemented a **RAG-based knowledge system** to surface contextual answers from historical lead data, reducing rep research time by **70%** and enabling faster decision-making.
- Tech Stack:** React, TypeScript, Redux, Node.js, Express, MongoDB (with Vector Search), Gemini SDK, AWS S3.

WORK/INTERNSHIP EXPERIENCE

Quicktouch Technologies Ltd., NSP, Delhi

May 2022 – Jun 2025

Sr. Software Engineer

- Mentored 3–4 junior developers, increasing team development velocity by 30% through technical coaching and rigorous code reviews.
- Led key architectural decisions for a modular platform, defining the tech stack and best practices to achieve 99.99% uptime.
- Drove the end-to-end delivery of core ERP/LMS features for web and mobile, engaging over 100,000 monthly active users.

QuickCampus (ERP)

- Architected and developed core modules from scratch (Results, Fees, Registration, User Permissions) for a large-scale school management ERP, successfully scaling the system to serve over 3,600 schools and 100,000+ active users.
- Integrated 5+ payment gateways, enabling 50K+ monthly transactions.
- Engineered the platform to reduce manual paperwork and migrate school data to the cloud, boosting data accuracy and cutting manual workload by 70%.

Partner Program

- Directed the project as a lead, managing the end-to-end React.js development lifecycle, including code reviews, blocker resolution, and pull request management to ensure high code quality and on-time delivery.

Code Inbound LLP, Janakpuri (Delhi)

Mar 2022 – May 2022

Software Developer Intern

- Maintained and enhanced a multi-tenant VoIP call-center platform using JavaScript, jQuery, PHP, and PostgreSQL.
- Integrated custom VoIP features (FreeSwitch, Asterisk, SBC, Kamailio) and supported backend modules such as CRM, network monitoring, and security tools.

EDUCATION

MisogiAI by Masai

14 Weeks AI Engineering Bootcamp

Jul 2025 – Sep 2025

Bengaluru, India

Guru Nanak Dev Institute of Technology

Diploma — Mechanical Engineering

Jul 2016 – Aug 2019

Delhi, India