@GitHub=sanketshinde3001

# SANKET RAJENDRA SHINDE

8625912150 • sanketshinde3123@gmail.com • linkedin.com/in/sanketshinde04 • github.com/{GITHUB}

## EXPERIENCE

**AI Engineer | BuildFastWithAI** **Apr 2026 – Present**

- Working on **AI/ML-powered client projects**, building and deploying production-ready solutions using **LLMs, RAG pipelines, AI agents, ML models, and modern full-stack technologies**.
- Designed and implemented **GenAI workflows, ML-powered automation systems, and intelligent AI agents** tailored to client requirements, integrating **LLM APIs, vector databases, ML inference pipelines, model evaluation, and external services**.
- Collaborated across multiple client projects to develop, optimize, and ship **scalable AI/ML features from development to production**, including **LLM integration, model inference, prompt engineering, RAG pipelines, and AI workflow optimization**.

**SDE I | Neurolaw AI** **May 2025 – Apr 2026**

- Engineered a custom **RAG and ML-powered retrieval pipeline** on **1 lakh+ documents with Realtime custom web search**, using **vector embeddings, LLM-based retrieval, and semantic search**, achieving **95% Accuracy** for legal document retrieval and legal research chatbot.
- Integrated Cloud OCR with proprietary **ML-assisted document processing and information extraction algorithms**, achieving **99% accuracy** with legal formatting; optimized batch processing to handle **500+ documents/hour**.
- Built and deployed **20+ real-time court scrapers** with residential proxy rotation for data collection, successfully extracting **2M+ public legal documents** from multiple court systems to create a comprehensive **ML/LLM-ready legal dataset** for downstream AI applications.
- Optimized backend services with **async Node.js and Python architecture, ML/LLM inference workflows, caching, and fallback mechanisms**, reducing API response time by **60%** while enabling efficient horizontal scaling across distributed AI systems.

**Open Source Contributor | Hacktoberfest, DevfestAI, GSSOC** **Oct 2024 – Present**

- Improved functionalities across **15+ repositories** involving **Next.js, TypeScript, LLM APIs, AI/ML workflows, and developer tools**, built demo applications showcasing AI capabilities and fixed backend-specific issues for CopilotKit, Tolgee, and MindsDB.

## PROJECTS

**Quickmed |** **User Portal** **Doctor and Admin Portal**

ReactJs, NodeJs, ExpressJs, MongoDB, Cloudinary, GEMINI API, Redis, Python, ML, RAG

- Built a 3-panel healthcare platform (Patient/Doctor/Admin) with **RAG-based and ML-powered symptom-to-doctor matching**, using **LLM reasoning, semantic retrieval, and vector-based knowledge matching** for intelligent nearby specialist recommendations.
- Implemented digital prescription generation, **LLM/ML-powered medical report analysis**, appointment scheduling with **73% faster response (1.5s → 0.3s via Redis)**, and **automated Telegram reminders** reducing no-shows by 30%.

**InterviewAce |** **Link**

Next.js, TypeScript, Supabase, PostgreSQL, WebRTC, AI/ML, LLMs

- Built an interview prep portal with **LLM-powered question generation, AI-based topic planning, personalized practice workflows, and progress tracking**.
- Implemented a real-time **LLM-powered AI mock interview system** using WebRTC peer connections with custom timer logic, **AI-driven interview evaluation, response analysis, and intelligent feedback generation**.

**GenCodex |** **Link**

Gemini API, AI-Powered IDE, ReactJs, NodeJs, Python, ML, LLMs

- Built an **AI/ML-powered online IDE** with **LLM-based code explanation, optimization, debugging assistance, and UI redesign**, supporting web development and C++.
- Integrated **LLM inference workflows and AI coding capabilities** to analyze source code, generate improvements, and provide contextual development assistance.

## TECHNICAL SKILLS

**Full Stack** - ReactJS, NextJS, TypeScript, NodeJS, ExpressJS, MongoDB, Tailwind CSS, PostgreSQL, Prisma, Redis, Docker, Jest
**Cloud** - Azure Document Intelligence, Cosmos Vector DB, Azure VM, Azure Services, AWS EC2, AWS S3, Posthog
**AI/ML & GenAI** - Python, **Machine Learning, Deep Learning, Reinforcement Learning (RL), ML Models, Model Training, Model Evaluation, Model Inference**, Streamlit, Langchain, Langraph, **LLMs, LLM APIs, AI Agents, RAG Systems, Vector Databases, Embeddings, Semantic Search, Prompt Engineering, Fine-tuning**

## ACHIEVEMENTS

- Secured **Global 6th Rank** in DevfestAI.
- Runner-up in i2i 2025 in HHH Domain by COEP.
- First Rank in **KIMO AI Competition**
- Won Base Bounty in **Unfold 2024 (Web3 Hackathon)**.
- 3rd Rank in **GenAI Hackathon by BuildFastWithAI**.
- Runners-up in **Medecro Medical Hackathon**.

## CERTIFICATIONS

**Generative AI with Large Language Models | Coursera**

- **Skills:** Large Language Models (LLMs), Generative AI Principles, **ML Model Training, Fine-tuning, Prompt Engineering, Model Evaluation**

## EDUCATION

**Government College of Engineering & Research, Avasari Khurd** **Dec 2021 - Present**

BE in computer Engineering; **CGPA - 8.71**