

Vatsa Joshi

📍 Vadodara, Gujarat 📞 +91 97375 57799 ✉ Mail

🌐 LinkedIn 🌐 vatsajoshi.in 🌐 Github

Technical Skills

Languages	Python, JavaScript/TypeScript, Rust, Java, C
AI/ML Frameworks	PyTorch, TensorFlow, Keras, Scikit-learn, HuggingFace Transformers, Unsloth
LLM & Agents	LangChain, LangGraph, CrewAI, RAG, CAG, LLM Fine-tuning (LoRA/QLoRA), FAISS, Groq
Computer Vision	OpenCV, YOLO, SSD, PaddleOCR, Diffusers
Backend / Cloud	FastAPI, NestJS, Node.js, Docker, AWS (S3, CloudFormation), PostgreSQL, MongoDB, Elasticsearch
Frontend	Next.js, React.js, React Native, Tailwind CSS

Work Experience

Software Engineer Trainee (AI)

Jan 2026 – Present

Promact · Vadodara

- Contributing to **BharatLaw.ai**, an AI-powered legal technology product serving Indian law firms.
- Designing and developing end-to-end AI-driven Document Management Systems enabling intelligent document translation, parsing, indexing, and retrieval.
- Developed a multi-agent Petition Writer system that drafts structured legal petitions from case inputs and documents using coordinated AI agents, with human-in-the-loop validation through iterative clarification before producing the final draft.
- Developed a multi-agent legal intelligence platform featuring judge profiling, case brief-driven litigation strategy with lawyer recommendations, and lawyer compatibility analysis — with coordinated agents handling vector similarity, LLM reasoning, and live web search.

Full Stack Developer — Internship

Jun 2025 – Dec 2025

Hiraeth Tech · Vadodara

- Built scalable products using NestJS, Next.js, and PostgreSQL with full-stack API integration and version-controlled deployments.

Deep Learning Engineer — Internship

Mar 2025 – May 2025

Prodigal AI · Remote

- Trained CNNs for image classification achieving 92% accuracy; fine-tuned transformer models for NLP tasks.
- Developed real-time object detection pipelines with YOLO and SSD; contributed to LLM architecture for a finance firm.

AI/ML Engineer — Internship

Feb 2025 – Apr 2025

Sparks To Ideas · Remote

- Trained XGBoost model for customer churn prediction with 95% accuracy.
- Built an AI content intelligence pipeline with LLM-based parsing and a FAISS (HNSW) deduplication system using multilingual embeddings.

Projects

Agentic Systems & Multi-Agent Frameworks:

Booky | 🌐 Multi-Agent · LLM · Email — Conversational multi-agent appointment booking system with real-time availability checks and automated email confirmations.

FloWork | 🌐 LangGraph · Groq · Next.js · Flask — N8N-inspired agentic workflow automation platform with a visual node builder and AI-powered conditional routing.

Hiring-Assistant | 🌐 CrewAI · Gemini 2.0 Flash · Streamlit — Agentic recruitment chatbot performing screening, dynamic technical assessment generation, and sentiment-based evaluation.

PersonAI | 🌐 Groq · LangChain · FastMCP · Next.js — AI agent automating Gmail, Google Calendar, and Tasks via Workspace APIs with OAuth 2.1 and real-time chat interface.

Ecommerce Agent | 🌐 LangGraph · LLM · WhatsApp — Multi-agent WhatsApp shopping assistant for product discovery, price comparison, and purchase decision support.

RAG, Document Intelligence & Memory:

LegalOCR | 🌐 PaddleOCR · FastAPI · Elasticsearch · Docker — Enterprise legal document parser with 22+ language OCR, RAG-ready semantic chunking, and Elasticsearch indexing.

MultiModal-RAG | 🌐 RAG · Multimodal · LLM — RAG pipeline supporting text, image, and document modalities for

unified cross-modal information retrieval.

Email Voice AI Agent | [LangGraph](#) · [Deepgram STT](#) · [Gmail API](#) · [Next.js](#) — Voice-first AI email assistant enabling hands-free inbox management — read, summarize, star, and compose via conversation.

Synapse | [Next.js](#) · [Qdrant](#) · [Upstash Redis](#) · [OpenAI Embeddings](#) — Multi-channel AI memory system unifying conversations across web, WhatsApp, email, and phone with probabilistic identity linking.

Multi-Agent Finance Advisor | [RAG](#) · [Multi-Agent](#) · [LLM](#) — Multi-agent system converting financial documents (10-K/10-Q, SEC filings, earnings transcripts) into actionable intelligence.

LLM Fine-tuning & Model Research:

Hindi2Guj-Qwen3.5 | [Qwen3.5](#) · [DeepSpeed ZeRO](#) · [Accelerate](#) — Fine-tuned Qwen3.5-0.8B on 1.5k Hindi→Gujarati pairs using ZeRO-2/3 multi-GPU distributed training.

Gemma 3 Fine-tune | [Unsloth](#) · [LoRA](#) · [HuggingFace](#) — Fine-tuned Gemma 3 (4B) on FineTome-100k for enhanced instruction-following via LoRA/QLoRA on Google Colab.

Rustformer | [Rust](#) · [Transformer Architecture](#) — Complete transformer architecture implemented from scratch in Rust for performance-focused inference research.

Computer Vision, Deep Learning & Mobile AI:

Deepfake Video Detection | [Python](#) · [Deep Learning](#) · [Computer Vision](#) — **2nd place** — **Tinkerthon 3.0 + 1-year SSIP Gujarat grant**; fine-tuned DL models for deepfake video classification.

AarogyaKosh | [React Native](#) · [FastAPI](#) · [HuggingFace](#) · [Docker](#) — Mobile health app translating lab reports to plain language with fully on-device medical transformer inference.

TradeXpert | [Next.js 15](#) · [Agentic AI](#) · [Inngest](#) · [MongoDB](#) · [Tailwind CSS](#) — Full-stack Agentic AI platform orchestrating automated financial intelligence workflows, featuring real-time market data integration, event-driven background analysis pipelines via Inngest, and institutional-grade technical visualizations.

Flash-Banana | [Next.js 15](#) · [Firebase](#) · [Genkit](#) · [Multimodal AI](#) · [TypeScript](#) — Advanced Multimodal AI suite for creative automation, orchestrating complex Genkit-driven workflows for natural language image manipulation, contextually-aware advertisement generation, and platform-optimized social media content synchronization.

Education

B.Tech in Computer Science

Navrachana University, Vadodara

Aug 2022 – Sep 2026

Awards

- **2nd Place** — **Tinkerthon 3.0 Hackathon** (Feb 2025) · Deepfake Video Detection System
- **1-Year Research Grant** — **SSIP Gujarat** · Awarded to further develop the Deepfake Detection project
- **Top 5 out of 3300 teams in Odoo x NMIT Hackathon**

Publications

- **Report & Analysis of Rabin-Karp Algorithm** — ResearchGate
DOI: 10.13140/RG.2.2.24592.01283
- **Blockchain Technology** — ResearchGate
DOI: 10.13140/RG.2.2.32465.75365