

Manoj Kumar

✉ mojirao770@gmail.com ☎ +91 9462229627 in LinkedIn G Github e Portfolio

EDUCATION

Delhi Technological University

2020-2024

B. Tech in Mathematics and Computing Engineering

- CGPA: 8.6

WORK EXPERIENCE

Software Engineer

Bengaluru

Instead Tax India Pvt. Ltd.

Aug 2025 – Present

- Built **agentic AI infrastructure** across Claude, OpenAI, and Gemini with provider-agnostic abstractions, dynamic model routing, structured **tool calling**, depth-limited subagents, and permissioned agent execution for document operations, chat search, bash/text execution, and DOCX/XLSX/CSV generation using **python-docx** and **openpyxl**.
- Designed and shipped **Workflow Engine (V3)** in **Golang**, a markdown-defined multi-step AI orchestration system with persisted state, retries, reference-file binding, entity-scoped agent permissions, and live **WebSocket** progress streaming.
- Built a browser-based **Excel editor** by forking the Go **Excelize** library into an in-house engine compiled to **WebAssembly** and run fully client-side, and a native **DOCX editor** built feature by feature against the **ECMA-376** spec; shipped templates that extract styles from uploaded DOCX/PPTX files and render markdown into formatted documents.
- Worked on the **Instead Developer API** and Developer Apps (**REST APIs**, **OAuth2/PKCE**, **OIDC**, webhooks); moved Google Drive, OneDrive, Box, and Plaid onto a unified **Integration Platform** managing OAuth tokens.
- Architected a regulatory **RAG pipeline** by scraping IRS and Tax Court publications; implemented normalization, semantic chunking, embedding generation, vector indexing, and citation-grounded LLM responses for compliance and research workflows.
- Delivered enterprise **SSO (SAML 2.0 & OIDC)** with firm-scoped configs, multi-tenant routing, and token exchange; built Golang file activity tracking with **S3** object versioning, debounced writes, and **MongoDB** audit pipelines.

Full Stack Developer

Remote

TerraDX Technologies Inc.

Feb 2024 – Aug 2025

- Engineered **GAIA**, a RAG chatbot with a **Next.js/TypeScript** frontend and **FastAPI (Python)** backend on **OpenAI** and **LangChain**, achieving a 3-second average response time with user data in **CosmosDB**.
- Built a multi-agent system using **OpenAI** and **Anthropic** APIs for natural language-driven **code generation**, with a **React Flow** interface covering tasks from architecture to bash scripts.
- Worked on **QuantumDX**, a raster image compression platform for Geotiff files, using **Next.js** and **FastAPI**. Integrated **Mapbox API** for visualization and ran Geotiff processing with **Docker** on **Azure Kubernetes Service (AKS)**.
- Developed a data scraping pipeline for geological portals using **Selenium** and **BeautifulSoup**, landing ZIP, PDF, CSV, XLSX, and XML sources into **Azure Blob Storage** and **Databricks** Delta tables.

PROJECTS

I HATE CORRUPT FILES

[link](#) 

(Python, FastAPI, React, TypeScript, .NET Open XML SDK, Cloudflare Workers, GCP Cloud Run)

- Built solo a live service and API repairing corrupted **.xlsx/.docx/.pptx** files via a staged **FastAPI** pipeline with an Open XML SDK validation sidecar on **Cloud Run**, behind a **Cloudflare Workers** API gateway with API keys and quotas.
- Detects 133/133 faults vs 100 for OfficeCLI, with zero false positives on 5,342 real-world workbooks; shipped with **pytest/vitest** suites, **GitHub Actions** CI, staging smoke tests before production, and a billing kill switch.

SALON AI ASSISTANT

[link](#) 

(Python, LiveKit, Groq, Cartesia, Deepgram, WebSockets, FastAPI)

- Built a microservices voice AI system with LiveKit audio communication, Groq Deepseek model for NLP, Cartesia TTS, and Deepgram STT, implementing intelligent escalation workflow from AI to human supervisors via WebSocket notifications.
- Designed a knowledge base learning loop where human answers to unresolved queries are reused by the AI agent.

TECHNICAL SKILLS

- **Languages:** Golang, Python, TypeScript, JavaScript, SQL, C++
- **AI & LLM Systems:** Agentic Workflows, Tool Calling, Subagents, Multi-Model Orchestration (Claude, OpenAI, Gemini), RAG Pipelines, Vector Search, Embeddings, LangChain, Claude Code
- **Frameworks & APIs:** React, Next.js, Vue.js, FastAPI, Node.js, REST APIs, WebSockets
- **Databases & Storage:** PostgreSQL, MongoDB, CosmosDB, Cloudflare D1, Vector Stores, S3, Azure Blob Storage
- **Cloud & Infrastructure:** Docker, Kubernetes (AKS), Azure, GCP Cloud Run, Cloudflare Workers, GitHub Actions CI, Git