MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT & ENGINEERING

SOHAM DALAL
B. Tech in Information Technology, 2027 Batch
Date of Birth: 10-06-2004
sohamdalal9481@gmail.com https://www.linkedin.com/in/sohamdalal01/
https://github.com/SohamDalal07 +91 9420367162



Professional Summary

Information Technology student with hands-on experience **applying LLMs, RAG**, and prompt engineering to build practical **AI-enabled solutions**, backed by strong backend fundamentals in **Python, Java, FastAPI, Spring Boot, SQL, and REST APIs**. Experienced in integrating **LLM APIs into production pipelines** and validating AI-generated outputs for accuracy.

Education

MPSTME, NMIMS University **2023 - 2027**
B. Tech in Information Technology
CGPA: 3.68/4.00

Moolji Jaitha College **2021 - 2023**
HSC(PCM): 81.17

St. Teresa's Convent Higher Secondary School **2010 - 2021**
SSC: 90.50

Experience

MavenArk, Mumbai — AI-based NBFC Application – AI/ML Developer Intern – Hybrid – June 2026 to August 2026

- Researched and extracted standardized financial and regulatory data **across 45+ NBFCs and 120+ annual filings**, covering profitability, balance sheet, and credit metrics. **Built an MCP-based Python pipeline integrating LLM APIs and web scraping** to automate **extraction across 150+ financial documents**, cutting manual effort by 85%. Validated extracted data against source PDFs using automated double-entry checks, **raising accuracy to 98%**

Projects

AI-Powered MahaRERA Compliance Verification Agent GitHub
Overview: AI-powered RAG application for verifying builder–buyer agreements against MahaRERA regulations.
Technologies: **Python, FastAPI, LangChain, FAISS, Ollama Qwen 2.5: 7B**

- Extracted clean text from **60+ scanned MahaRERA documents** (**713+ pages**) via a hybrid **EasyOCR/PyMuPDF** pipeline, handling corrupted **legacy Marathi fonts**
- Indexed **1,700+ regulation** chunks using **BAAI/bge-base-en-v1.5 embeddings** with **local FAISS search**
- Built **FastAPI REST APIs** for the compliance workflow and automated cited clause-level reports, **reducing review time by 95% (4 hours → <30 seconds)**.

CrowdVibe – Real-Time Crowd-Sourced Music Queuing Platform GitHub
Technologies: Java, Spring Boot, Spring Security, PostgreSQL, React JS

- Built a full-stack platform with secure **Google OAuth2/BCrypt authentication real-time queue updates** using **SSE.**
- Implemented concurrent voting with **Spring Data JPA** and transactional locking, and integrated the YouTube Data API for song search and metadata.

Satellite-Based Oil Tank Volume Estimation System GitHub
Overview: Computer vision pipeline that estimates crude oil volumes in floating-head storage tanks from satellite imagery by detecting tanks and extracting the crescent shadow that indicates the oil level.
Technologies: Python, OpenCV, YOLOv8, Streamlit, scikit-image

- Trained **YOLOv8 on 1,828** labeled satellite images, **achieving 94% mAP50** across 3 infrastructure classes
- Built a geometric shadow-analysis pipeline estimating **fill-percentage within 6% of ground truth** on 275 validation tanks
- **Exported PyTorch weights to ONNX**, cutting **memory footprint 80%** and enabling serverless FastAPI deployment

SKILLS

Technical Skills

- **Programming:** Python, Java, SQL
- **AI & GenAI:** LLMs, RAG, Prompt Engineering, Embeddings, Vector Search, LangChain, MCP, Machine Learning
- **Backend & APIs:** FastAPI, Spring Boot, Spring Security, REST APIs, Server-Sent Events (SSE)
- **Databases:** PostgreSQL, MySQL, MongoDB, FAISS
- **DevOps & Cloud:** Git, GitHub, Docker, AWS (EC2, S3), Postman
- **AI/ML:** Computer Vision, YOLOv8, OpenCV, scikit-image
- **Fundamentals:** OOP, DBMS, Data Structures & Algorithms, SDLC, Agile Methodology, SOLID Principles

Soft Skills: Analytical Thinking, Decision-Making Under Uncertainty, Attention to Detail, Adaptability, Peer Collaboration

Certifications & Publications

IEEE, ICIETSD 2026 - Designed AECL, a runtime governance layer enforcing organizational policies and hierarchical authority in multi-agent systems.
LangChain for LLM Application Development — DeepLearning.AI

Leadership&Extracurricular

Stuv.ai, Business Development Executive: Led lead-generation outreach via **100+ cold calls** for an AI-powered visual commerce startup, achieving an **~88% conversion rate** through the qualification funnel.
Volunteer – WASH & Community Development Program, CSRBOX, Nashik: Conducted community needs assessment and PRA sampling in Vani village for WASH and rural development initiatives.