

MOHIT PRAVINKUMAR INGALE

mohitpraviningale13@gmail.com | +91-8118921925 | [Portfolio](#) | [GitHub](#) | [LinkedIn](#)

SUMMARY

Full-stack and AI engineer who ships complete products end-to-end — from database schema to production AI features to deployed UI. Independently built and deployed two live platforms with 50+ real users, including an autonomous agent system with tool-calling and multi-model failover, plus an open-source Chrome extension. Strong backend fundamentals (FastAPI, PostgreSQL) and hands-on production LLM experience spanning agentic orchestration, RAG, evaluation, and interpretability. B.E. in Information Technology, 2026.

TECHNICAL SKILLS

Languages: Python, JavaScript, TypeScript, SQL, C

Backend & Frameworks: FastAPI, Node.js, Express.js, React, React Native, Expo, REST APIs

AI/ML & LLM Systems: Agentic AI (multi-step tool-calling, autonomous orchestration loops), RAG (pgvector, hybrid search, RAGAS evaluation), Prompt Engineering, LLM Application Design, SHAP Interpretability, XGBoost, Random Forest, OpenCV, YOLOv8, PyTorch, Pandas, NumPy

Databases: PostgreSQL, MongoDB, Supabase, Firestore

Cloud & Tools: AWS (EC2), Docker, Git/GitHub, Linux, Postman, Android Studio

Core CS: OOP, DBMS, Operating Systems, Computer Networks, Data Structures & Algorithms

EXPERIENCE

AI/ML Intern — Future Labs Aug 2025 – Dec 2025

Multi-View Deep Learning System for Football Foul Detection — Python, PyTorch, YOLOv8, OpenCV, Streamlit

- Built a computer vision pipeline (YOLOv8, OpenCV, SoccerNet dataset) integrated into backend APIs for automated foul detection.
- Integrated a multi-view vision transformer with player-detection pipelines to classify foul events across camera angles and generate timestamped incident snapshots.
- Optimized inference pipelines and built a real-time visualization layer in Streamlit for field-level decision-making.

PROJECTS

SHG Digital Record Platform — Adopted by 5 Self Help Groups, 50+ active users ([GitHub](#) | [Live](#))

React Native, Python, TypeScript, Node.js, Express.js, PostgreSQL, AWS (EC2), Supabase, REST APIs

- Independently designed, built, and deployed an end-to-end **bilingual (Marathi/English) financial management** platform for rural Self Help Groups, digitizing member management, meetings, savings, and loan workflows.
- Built a loan management system with **reducing-balance calculations**, approval workflows, repayment schedules, passbooks, automated PDF generation, and audit-ready transaction history.
- Built a **voice-command workflow using speech recognition and LLM-based intent classification** to drive financial operations through natural-language commands.
- Implemented a **secure multi-tenant backend with role-based access control** and validation to protect financial record integrity.

Microflow — ML Experimentation & Observability Platform ([GitHub](#) | [Live](#))

Python, FastAPI, React, PostgreSQL, pgvector, scikit-learn, Gemini API, Docker

- Built and deployed an **ML experiment-tracking platform**: dataset management, reproducible training pipelines, versioned artifact storage, and real-time metrics dashboards.
- Designed a **repository-pattern backend** with clean service-layer separation, powering a “Compare Runs” analysis feature.
- Built an **autonomous “AI Investigator” agent**: given an open-ended question, it plans and executes **multi-step tool calls** (querying the database, pulling SHAP explanations, comparing experiment runs), evaluates the results, and loops until it reaches a conclusion — bounded by strict iteration limits to guarantee termination.
- Added production resilience to the agent layer: **cascading fallback to a secondary LLM** when the primary model call fails or times out, and response caching via query hashing to cut redundant LLM calls.
- Grounded the agent’s data access — Gemini is **sandboxed from the database** and reasons only over pre-fetched, strictly-typed (Pydantic) structured data, with output validated before deterministic SQLAlchemy ORM query construction, **eliminating hallucinated results**.
- Added **SHAP-based interpretability** to surface feature-level explanations alongside experiment metrics (“Explain This Run”).
- Built a **hybrid retrieval-augmented generation (RAG) pipeline** using pgvector for semantic + keyword search over experiment metadata and docs, with retrieval quality evaluated using the **RAGAS framework**.
- Covered the platform end-to-end with an **automated test suite (298+ tests)** spanning pipeline, API, and data-integrity paths.

SentinelFlow — Applied ML Traffic Intelligence Platform · Qualified, Flipkart Grid 7.0 “Gridlock Challenge,” Round 2

Python, XGBoost, Random Forest, Graph-Based Analytics

- Built a predictive traffic-operations platform modeling congestion patterns with XGBoost and Random Forest, and applying graph-based analytics to road/route networks to surface operational recommendations.

Other Shipped Work: WorthIt — an open-source, privacy-first Chrome extension (React 19, TypeScript, Vite/CRXJS) that overlays real-time “hours of work” price context on Amazon India, Flipkart, Myntra, and Meesho. All computation runs locally; **zero data collection or tracking**.

EDUCATION

Bachelor of Engineering (B.E.), Information Technology — SVPM College of Engineering, Pune 2022 – 2026

CGPA: 8.26

ACHIEVEMENTS

- Won 3 national-level hackathons, designing and demoing production-ready software under strict time constraints.
- Winner — National Prompt Engineering Competition (1st Place).