

Mithin Sagar S

Chennai, India | +91 63817 90265 | mithinsagar@gmail.com | github.com/mithinsagar

EDUCATION

Vellore Institute of Technology, Chennai

B.Tech, Computer Science and Engineering; Specialization: AI and Machine Learning

Aug 2023 – May 2027

Chennai, India

EXPERIENCE

Machine Learning Intern — Deep Learning & Computer Vision

May 2025 – Jun 2025

Indira Gandhi Centre for Atomic Research (IGCAR)

Kalpakkam, India

- Weld inspection was manual and inconsistent between inspectors. Owned an automated defect-localization pipeline end to end, fine-tuning a **YOLOv8l** detector on high-resolution weld imagery.
- Curated and labelled the dataset, ran 100 GPU training epochs, and evaluated on precision, recall, F1 and mAP — **99.1% on training against 74.3% held-out**.
- Diagnosed the gap as overfitting, revised the dataset and splits, and presented the trade-offs and next steps directly to the scientific staff.

PROJECTS

FlashForensics AI — Agentic Recovery for Corrupted Storage | Python, LangGraph, LLM Agents | [GitHub](#)

2026

- Recovery tools return thousands of nameless fragments with no indication of what survived. Built a **multi-agent pipeline** on **LangGraph** that parses a corrupted FAT32/exFAT image, locates the damage and returns evidence-backed recovery verdicts.
- Each agent owns one stage: filesystem parsing, Shannon-entropy mapping to isolate corrupted sectors, signature-based carving of orphaned regions, then classification and scoring of each fragment.
- Ambiguous file signatures are resolved by **semantic retrieval** over an embedded file-type index rather than a lookup table, with an **LLM** producing the final recoverability verdict per fragment.

MediXplain — Calibrated Disease Prediction with Retrieved Evidence | Python, scikit-learn, FAISS, Flask | [GitHub](#)

2026

- In symptom checking a confidently wrong answer is worse than no answer. Predicts likely conditions from reported symptoms and pairs every prediction with evidence the user can independently verify.
- Seven **scikit-learn** classifiers **5-fold cross-validated** on ~190,000 patient records spanning **377 symptoms and 727 conditions**; probability calibration aligns reported confidence with observed accuracy (best **83.09%**).
- A **RAG** layer retrieves supporting medical evidence from a **FAISS** index of 384-dim embeddings; shipped with a **GitHub Actions** pipeline running linting and 12 tests across a Python 3.10–3.12 matrix on every change.

Society Maintenance Tracker — Full-Stack Complaint Platform | TypeScript, React, Node.js, PostgreSQL | [GitHub](#)

2026

- Housing societies triage maintenance complaints over paper and chat, so nothing is auditable. Designed and built the platform end to end: a **React/Next.js** interface over a typed **Node.js REST API** of 20+ handlers, with a normalised **PostgreSQL** schema and role-based authentication.
- Every status transition writes to an **append-only event log** kept immutable by a database trigger, and resident notifications commit through a **transactional outbox** so a failed third-party call can never roll one back.
- Documented the API and data model, and validated the whole flow end-to-end with **81 Vitest tests** before shipping.

Also on GitHub: [EXAI-ResumeIntel](#) (semantic matching over 1M+ postings, **SBERT** embeddings, FastAPI) | [XAI Attack & Defense](#) (PyTorch robustness study, 246K+ samples)

PUBLICATIONS

Automated AWS Resource Cleanup for Optimization of Cost and Security — Presented at the *2nd Intl. Conf. on Advanced Nexus of Data and Information Technology (ICANDIT 2026)*, INTI International University, Malaysia (*publication pending*). Idle cloud resources sit unused and forgotten, quietly driving up cost and widening the attack surface. Built an event-driven **Boto3** system that scans **EC2, EBS, S3, IAM and RDS** for waste, then reclaims it on a scheduled **Lambda** cron with dry-run simulation and protected-tag exclusions guarding against accidental deletes. [GitHub](#)

TECHNICAL SKILLS

Languages: Java, Python, C, JavaScript, TypeScript, SQL

Generative AI: LLM APIs, prompt engineering, Retrieval-Augmented Generation (RAG), agentic workflows, LangGraph, FAISS, embeddings

Machine Learning: scikit-learn, PyTorch, pandas, NumPy, cross-validation, model calibration

Web & Backend: React, Next.js, Node.js, FastAPI, Flask, REST APIs, PostgreSQL, MongoDB

Tools & Practices: Git/GitHub, GitHub Copilot, GitHub Actions (CI/CD), Docker, pytest, Vitest, OOP, Agile/Scrum

LEADERSHIP & CERTIFICATIONS

Student Welfare Outreach Head, TechnoVIT & Vibrance, VIT Chennai (2025–26). Led a 25+ member team across sponsorship, publicity and media for two flagship events.

Head of Visual Media, VOICE-IT, VIT Chennai's radio station club (2024–25). Led a 40+ member team.

Second Runner-Up, Figma × Apple Vision Pro Design Challenge, GDSC VIT Chennai (2023).

Certifications: Introduction to Generative AI (Google Cloud); Python Data Structures (Michigan); Google UX Design.