Harsha Vardhan Doppalapudi

International Institute of Information Technology, Bangalore

+91 9014216288 | itsmeharsha081459@gmail.com | linkedin | github.com/Harsha081459 | Portfolio

Education

International Institute of Information Technology Bangalore **Jul 2024 – May 2028**

Bachelor of Technology in Data Science and Artificial Intelligence *GPA: 3.34 / 4.0*

Experience

NLP Research Intern (SRIP) **Jun 2026 – Aug 2026**

IIIT Bangalore – Natural Language Processing Lab | Advisor: Prof. Tulika Saha *Bangalore, India*

- Extracted **4,785** structured cultural rules (29 Indian states × 11 domains, **95%+** precision) via an automated data pipeline using PyMuPDF and the **OpenAI API** to process **60,000+** words of text.
- Scaled LLM evaluation to **95,700** judged dialogues by benchmarking 10 open-source models (Llama 3.1, Qwen3.6, DeepSeek-R1, Gemma) via an **LLM-as-a-judge** pipeline. Co-authored a **research paper (under review)**.

Projects

Traffic Violation Detection System | Live Demo | *YOLOv11, ONNX, FastAPI, Docker* **May – Jun 2026**

- **Designed an ingest–process–transform pipeline** over **17,670** annotated images, converting raw video frames into structured violation records through a 3-stage cascade (**mAP50** 0.935 plate / 0.838 helmet).
- **Deployed a production FastAPI microservice** with **Docker** and **GitHub Actions** CI/CD, raising throughput **1.8**× via **ONNX Runtime** and adding per-IP rate limiting plus magic-byte input validation.

SynthoGen AI | Dashboard | *PyTorch, Opacus, Streamlit* **Mar – Apr 2026**

- **Built a privacy-first data pipeline** over **258K+** patient records across **4** generative architectures, retaining **94.6%** downstream ML utility with zero privacy breaches via **Opacus** DP-SGD.
- **Documented and published** the resulting synthetic EHR datasets on **IEEE DataPort** (DOI: 10.21227/64c7-vj34) and shipped a **Streamlit** dashboard for fidelity and privacy-audit review.

Event Ticketing Platform | *C, POSIX Threads, Sockets, IPC, 2PL, WAL* **Jun – Jul 2025**

- **Engineered a multi-threaded TCP server in C** using bounded thread pools and connection-limiting semaphores to process concurrent bookings, with a 4-role **role-based access control** (RBAC) matrix.
- **Implemented a page-based storage engine** with Write-Ahead Log (WAL), **Two-Phase Locking** (2PL), an LRU buffer pool, and B+ tree indexes to guarantee **ACID** durability, plus concurrent stress tests to debug lock and recovery paths.

Air-Travel Cargo Database Engine | *MySQL 8.0, FastAPI, JWT Auth, Docker* **Apr – May 2025**

- **Designed and shipped a production logistics REST API** on a normalized 5-table **MySQL 8.0** schema, ingesting and joining cargo, booking, and tracking data via multi-table SQL queries, containerized with **Docker Compose**.
- **Enforced end-to-end data integrity** via **ACID** transactions with commit/rollback safety, `ON DELETE CASCADE` foreign keys, and signed-**JWT RBAC** (3 roles, PBKDF2-hashed credentials), covered by a 10-case test suite.

Technical Skills

Languages: Python, C/C++, SQL, Bash
Data & Systems: MySQL, Pandas, NumPy, REST APIs, POSIX Threads, Sockets, IPC, JWT
ML & DL: scikit-learn, LightGBM, PyTorch, Hugging Face Transformers, ONNX Runtime, OpenAI API
Deployment & Tools: Docker, Docker Compose, FastAPI, Streamlit, GitHub Actions (CI/CD), Git, Linux

Relevant Coursework

- Machine Learning
- Deep Learning
- Natural Language Processing
- Data Structures & Algorithms
- Database Management Systems
- Operating Systems
- Probability & Statistics
- Linear Algebra
- Computer Vision

Achievements

- **Published** synthetic EHR datasets to **IEEE DataPort** (DOI: 10.21227/64c7-vj34) by building a generative AI framework evaluating 4 architectures across 258K+ records.
- Engineered a novel data-driven policy analysis solution, advancing to the **finals** of the Data Driven Policy Innovation Datathon (IIIT-Bangalore), surpassing **333** registered teams.
- Ranked in the **Top 16%** of the Kaggle Stellar Classification challenge by engineering a robust cross-validated meta-stacker with a defensive consensus architecture.