

Tharun Krishna

tharunkrishna1611@gmail.com | github.com/Tharun-bot | tharun-port.vercel.app | linkedin.com/in/tharun-krishna

EDUCATION

Indian Institute of Technology Madras

BS in Data Science and Applications

Chennai, India

07/2023 – 04/2027

National Institute of Technology Karnataka

B.Tech in Mining Engineering

Surathkal, India

07/2023 – 04/2027

EXPERIENCE

ML Research Intern

05/2026 – 07/2026

National Institute of Technology Karnataka

Surathkal, India

- Built an end-to-end **ML pipeline in Python** to predict blast-induced **Peak Particle Velocity (PPV)** from blast design parameters (charge weight, distance, delay), supporting safer blast-design decisions for mine sites
- Ran **exploratory data analysis** and feature engineering on field blast-monitoring data, and used **learning curves** to diagnose bias/variance trade-offs and select a model resilient to overfitting on a limited dataset

AI Research Intern

01/2025 – 03/2025

Scalar Labs

Surathkal, India

- Designed a **real-time time series forecasting system** that converts streaming numerical data into **natural language embeddings** and predicts future values using **pretrained, frozen LLMs (PyTorch, Transformers)**
- Implemented **embedding alignment** techniques to bridge numerical and language-based representations, enabling efficient **next-K-step forecasting** without fine-tuning the base LLM

Machine Learning Engineer Intern

05/2024 – 07/2024

Vaidsys Technologies

Remote, India

- Developed a **real-time anomaly detection pipeline** for streaming API and IoT sensor data, **cutting the false-positive alert rate by 35%** through feature engineering and adaptive threshold calibration
- Deployed **ML inference services on AWS EC2 with Docker**, achieving **sub-100ms prediction latency** with **S3-backed model artifact versioning**

PROJECTS

CodeLens | *Python, FastAPI, tree-sitter, sentence-transformers, FAISS, Supabase* | github.com/Tharun-bot/codelens

- Built a **semantic code search engine** that indexes codebases at the function/class level using **tree-sitter AST-aware chunking**, embeds each chunk with **sentence-transformers**, and serves lookups from a **FAISS** vector index with **Supabase (Postgres)** metadata, via a **CLI** and **REST API (FastAPI)**
- Implemented a **two-stage retrieve-then-rerank pipeline** (FAISS bi-encoder + cross-encoder reranking); **benchmarked end-to-end on the Flask codebase** (1,637 chunks) at **p50 291ms / p95 584ms / p99 683ms**, profiling per-stage latency to isolate cross-encoder inference and DB round-trips as the dominant cost

CyberOrch (AI SecOS) | *Python, FastAPI, Semgrep, Garak* | github.com/Tharun-bot/CyberOrch

- Architected an **AI-powered security orchestration platform** where a planner agent automatically selects and runs **specialized security agents** (Semgrep SAST, Garak LLM security testing), normalizing findings into a unified evidence model
- Built a **tamper-evident evidence lake (SHA-256 hash chain)** and a **deterministic policy engine** producing reproducible **Allow / Block / Human-Review** verdicts, plus a **grounded LLM explainability layer** that justifies decisions strictly from verified evidence

TECHNICAL SKILLS

Languages: Python, SQL, Go, C/C++, JavaScript

ML/AI: PyTorch, Transformers, sentence-transformers, scikit-learn, pandas, NumPy, LLM Orchestration, Embeddings

Infra & Retrieval: FastAPI, FAISS, Supabase/PostgreSQL, Docker, tree-sitter, REST APIs

Concepts: Semantic Search, Retrieval-Augmented Systems, Time Series Forecasting, Applied Security

Tools: Git, Linux, CI/CD

CERTIFICATIONS AND EXTRACURRICULAR

Solved **400+** problems on LeetCode

Oracle Cloud Infrastructure 2025 Certified Generative AI Professional | Certificate

Oracle Cloud Infrastructure 2025 Certified AI Foundations Associate | Certificate