THARANIKA R

kavintharanika@gmail.com — LinkedIn — GitHub — LeetCode — HuggingFace — Portfolio

OBJECTIVE

Engineering student specializing in **Generative AI**, **NLP**, and **Full-Stack Development**, with hands-on experience building **LLM-powered applications**, **RAG pipelines**, **agentic AI systems**, and **transformer-based** ML models for real-world deployment.

EDUCATION

B.E – Computer Science and Design, Bannari Amman Institute of Technology **2023–2027 — CGPA: 8.12**

TECHNICAL SKILLS

Programming & Problem Solving: C, Java, Python, SQL
Full-Stack Development: React.js, Node.js, Express.js, Next.js, REST APIs, WebSockets, Fast API
AI & LLM: RAG, Agentic AI, MCP, LangChain, LangGraph, FAISS, Prompt Engineering, Fine-Tuning(LoRA, QLoRA), NLP, ML, DL
LLM Serving & Observability: Ollama, HuggingFace, LangFuse
Libraries & Frameworks: Pandas, NumPy, PyTorch, TensorFlow, HuggingFace Transformers, NLTK, spaCy, Gradio
Database: MySQL, PostgreSQL, Prisma ORM, Vector DB
Cloud: AWS (S3, EC2, RDS, VPC)
Big Data: Hadoop, Spark, Databricks
DevOps: Docker, Kubernetes, Vercel, Render, HuggingFace Spaces
Tools: Git, GitHub, Postman, JupyterNotebook, Google Colab, VS Code

PROJECTS

MediGuard — Healthcare Fraud Intelligence System *2 months* *Live Demo*

- Built **end-to-end ML pipeline** with **EDA**, **feature engineering**, and **binary classification** to detect fraudulent Medicare providers using **XGBoost+SHAP explainability** on 125,841 outpatient claims.
- Engineered **15+features** billing patterns, diagnosis code frequency, claim duration, zero-deductible fraud signals handled **10:1 class imbalance** via `scale_pos_weight`
- Achieved **93% accuracy** and **ROC-AUC 0.9658**; served model via **HuggingFace Spaces** with **Gradio** UI supporting real-time CSV fraud prediction and **SHAP** visualization

Stack: Python, XGBoost, SHAP, Pandas, Scikit-learn, Gradio, HuggingFace

AI Token Monitor — Open-Source Python Package *Published* *PyPi*

- Published production-ready package to **monitor LLM inference token usage, latency, throughput, and cost** across **OpenAI, Groq, Anthropic, Gemini, OpenRouter** with unified response normalization
- Built **FastAPI middleware integration** for plug-and-play **MLOps observability** designed **scalable chat session tracking** for multi-turn conversations and real-time **inference metrics** aggregation

Stack: Python, FastAPI, LLM APIs, Middleware Design

Agentic RAG-Based Student Assistance System *2 months* *Live Demo*

- Built **agentic RAG chatbot** with **tokenization**, **embedding workflows**, and **semantic search** using **FAISS+ MiniLM** and **Azure OpenAI** for grounded, low-hallucination responses
- Designed **query routing pipeline** with **multi-step reasoning** and **tool augmented retrieval** across knowledge base and web search tracked **inference metrics** via **LangFuse** observability

Stack: FAISS, MiniLM, LangChain, LangFuse, Azure OpenAI

ACHIEVEMENTS

- **CDAC National Urban IoT Challenge (Smart City 2.0), 2025** — **Runner-Up nationally**; built end-to-end IoT dashboard for real-time urban monitoring; awarded **Rs.25,000**
- **IntelliMobility Ideathon, ARAI 2024** — **2nd Runner-Up**; developed *"Active Safety for Two-Wheelers using Advanced Rider Assistance"*; awarded **Rs.25,000**
- **Neonexus 36.0 Hackathon, 2025** — **2nd & 5th place** across tracks; built Edge AI solution for pedestrian safety in 36-hour challenge
- **Hackzilla 24-Hour Hackathon, 2025** — **Special Recognition**; built **EduLearn**, offline-first PWA for low-connectivity education using service workers and caching

CERTIFICATIONS

- AWS Cloud Practitioner Essentials **View Certificate**
- Machine Learning with Python — freeCodeCamp **View Certificate**
- Introduction to Model Context Protocol — Anthropic **View Certificate**

AREAS OF INTEREST

Generative AI & LLMs — NLP & Transformer Models — Data Engineering & MLOps — Full-Stack & DevOps