Tanvir Ahammed

Dhaka, Bangladesh +8801746345484 tanvir7535@gmail.com

tanvir-ahammedd.github.io itistanvir tanvir-ahammedd

Professional Summary

Computer Science and Engineering student specializing in AI/ML with hands-on experience building LLM-powered applications, RAG systems, and agentic AI solutions. Strong foundation in machine learning, NLP, and algorithms, with 700+ problems solved on Codeforces (1259 rating). Actively researching multilingual code generation and ML fairness, with proven competitive programming achievements and algorithmic foundations.

Technical Skills

Generative AI & LLMs: LangChain, OpenAI API, Groq API, RAG, ReAct/CodeAct Agents, Agentic AI Systems, Prompt Engineering, LoRA/QLoRA
Machine Learning & Deep Learning: PyTorch, Scikit-learn, Neural Networks, Model Training & Evaluation, Feature Engineering, Supervised Learning, Fine-tuning Strategies
Natural Language Processing (NLP): Hugging Face Transformers, Tokenization, Text Classification, Embeddings, Semantic Search, Information Retrieval, LLM-based NLP Pipelines
Data Science & Analytics: Exploratory Data Analysis, Statistical Analysis, Data Preprocessing, Predictive Modeling
Programming Languages: Python, C++, C, Java, JavaScript, SQL
Databases & Vector Stores: FAISS, Vector Databases, MySQL, PostgreSQL, SQLite
Web Frameworks & APIs: Streamlit, FastAPI, Django

Projects

Intelligent Document Search using RAG github.com/tanvir-ahammed/rag-pdf-qa | Live Demo

- Built a **RAG**-based PDF Q&A system using **LangChain**, **Groq LLM**, and **FAISS**, with **OCR** support for scanned PDFs via **Tesseract**, achieving sub-2s response times
- Developed a hybrid document processing pipeline with **automatic OCR fallback**, combining PyPDF extraction and pdf2image for 100% PDF compatibility
- Designed vector retrieval using **HuggingFace embeddings** (MiniLM) and integrated **Llama 3.1** retrieval chains
- Deployed a cloud-ready Streamlit app with multi-document ingestion, **session management**, and **source attribution**

Autonomous AI Research Agent github.com/tanvir-ahammed/search-engine-ai | Live Demo

- Developed AI research assistant using LangChain and Groq LLM with **multi-agent RAG architecture**, integrating real-time data retrieval from arXiv, Wikipedia, and DuckDuckGo with automated source citation tracking
- Built production-ready Streamlit web application featuring **conversation memory**, session state management, and **chat history export** functionality for enhanced user experience

Bengali Empathy LoRA github.com/tanvir-ahammedd/bengali-empathetic-llama-finetuning

- Fine-tuned **LLaMA 3.1-8B** on Bengali empathetic conversations using LoRA, implementing gradient checkpointing and mixed precision training on Kaggle free GPU while maintaining full sequence length
- Built modular fine-tuning pipeline with **OOP design patterns**, evaluation using perplexity, BLEU, and ROUGE
- Developed end-to-end **LoRA fine-tuning system** including dataset preprocessing, attention layer adaptation, automated evaluation, and human assessment protocol with comprehensive documentation

Student Performance Prediction & Analysis github.com/tanvir-ahammedd/student-performance-ml-analysis

- Developed ML regression model predicting student math scores with 88% accuracy using Random Forest and Gradient Boosting algorithms on dataset of 1000 students
- Performed **EDA**, **feature engineering**, and built scikit-learn **preprocessing pipeline** for handling categorical and numerical variables
- Implemented **modular code structure** with separate pipelines for data ingestion, transformation, model training, and prediction

Thesis & Research

Optimizing Bangla Code Generation (In Progress)

- Developing efficient code generation systems for low-resource languages using **LoRA/QLoRA**, reducing parameters by ∼90% while maintaining comparable performance.
- Implementing and evaluating reasoning-augmented methods **(ReAct/CodeAct)**
- Conducting benchmarking on multilingual code datasets (**500+ samples** from mHumanEval, MBPP), optimizing for efficiency and **cross-language generalization**

ML Fairness in Credit Scoring - Bias Auditing Framework (In Progress)

- Conducting comparative study of machine learning bias across **5+ model architectures** including Logistic Regression, Random Forest, and Gradient Boosting
- Developing open-source toolkit for **automated bias detection** in lending models, enabling compliance with regulatory fairness standards
- Applying **explainable AI** (SHAP, LIME) to identify root causes of discriminatory predictions, analyzing **10+ fairness metrics** across protected attributes

Achievements

Codeforces Rating: **1259** | **700+** Problems Solved — **codeforces.com/profile/tanviiir**
3rd Runner-Up: Inter-department Programming Contest, GUB CSE Carnival *(2024)*
2nd Place: Intra-university Programming Contest *(2024)*

Training & Certifications

AI / Machine Learning Course – Phitron *(Ongoing)*
5-Day In-Person AI/ML/IoT Bootcamp – Bondstein *(2025)*
Competitive Programming Course – CPS Academy *(2024)*
CSE Fundamentals – Phitron *(2022)*

Education

B.Sc. in Computer Science and Engineering *(Expected June 2026)*
Green University of Bangladesh

- CGPA: 3.58/4.00 (1 Semester remaining)

Higher Secondary Certificate (Science) *2019*
Chowgacha Government College

- GPA: 5.00/5.00

References

Professor Dr. Md. Ahsan Habib
Chairperson, Department of Computer Science and Engineering
Green University of Bangladesh
Email: mahabib@cse.green.edu.bd

Mr. Montaser Abdul Quader
Lecturer, Green University of Bangladesh
Green University of Bangladesh
Email: montaser@cse.green.edu.bd