

MD. EMON HASAN

AI/ML Engineer

📍 Dhaka, Bangladesh | ✉️ emon.mlengineer@gmail.com | ☎️ +8801834363533 | 🐙 Github | 🔗 LinkedIn | 🌐 Portfolio

SUMMARY

AI/ML Engineer specializing in Generative AI, NLP, and Machine Learning, with expertise in Deep Learning and developing production-grade multi-agent systems, RAG pipelines, and tool-augmented AI architectures. Skilled in fine-tuning LLMs and applying NLP techniques such as transformers and language modeling to build robust AI solutions. Experienced in FastAPI, Docker, Nginx, and CI/CD for designing and deploying scalable, end-to-end AI systems that improve decision-making, drive automation, and deliver measurable business outcomes.

EXPERIENCE

AutoMetaHQ

Machine Learning Engineer

Apr 2026 – Present
London, UK (Remote)

- Sketch to Scales for business workflows, covering data preprocessing, fine-tuning, validation, and optimization.
- Architected scalable ML pipelines aligned with business goals, improving production reliability and workflow efficiency.
- Integrated workflows, model monitoring, and feedback loops for faster iteration and seamless deployment.

Codixel

Junior Machine Learning Engineer

Jan 2026 – Present
Dhaka, Bangladesh

- Fine-tuned Gen AI and ML models using transformer architectures, improving performance and inference reliability.
- Built multi-agent RAG systems and multi-agent workflows, improving scalability and production readiness.
- Collaborated with senior engineers to deploy AI solutions using FastAPI, Docker, and CI/CD pipelines.

Hi-Tech Parks

Machine Learning Engineer Intern

Oct 2025 – Dec 2025
Dhaka, Bangladesh

- Developed RAG pipelines and supported multi-agent AI applications with model integration, testing, and deployment.
- Performed data preprocessing, feature engineering, and model optimization for reliable ML pipelines.

PROJECTS

MediGenius – Multi-Agent Medical AI Assistant

Github 🐙

- Built an AI-powered medical assistant using multi-agent orchestration to deliver context-aware healthcare consultations, achieving 82% clinical alignment, reducing misinformation by 60%, and improving query resolution rate by 75%.
- Parallel RAG retrieval over ChromaDB, Wikipedia, and Tavily with a LiteLLM model gateway for fallback, deterministic safety gating, SQLite session memory, and a FastAPI backend with React frontend deployed via Dockerized CI/CD.

InformaTruth – Explainable AI Fake News Detection System

Github 🐙

- Fine-tuned RoBERTa-base on the LIAR dataset (12.8K labeled statements) using LoRA/PEFT while preserving accuracy, achieving 70% accuracy and 0.69 F1 with recall-weighted evaluation to prioritize catching fake claims.
- Orchestrated a five-node LangGraph agent pipeline over various inputs with FLAN-T5 Llm, prompt-injection guardrails, heuristic hallucination-risk scoring, and concurrent fallback search on a Dockerized FastAPIReact stack.

AutoDocThinker – Agentic Document QA System

Github 🐙

- Architected a five-mode agentic RAG engine on supporting Naive, Advanced, CRAG, Self-RAG, and Deep workflows, where decomposes queries into sub-agents and reliable document intelligence across multiple business domains.
- Implemented Hybrid Search fusing ChromaDB dense retrieval with BM25 via RRF and CrossEncoder reranking and TTL caching to control LLM spend, improving query efficiency by 40% across 50+ documents with accurate answers.

FraudChurn Nexus – Unified Fraud and Churn Prediction

Github 

- Built two classical ML pipelines on imbalanced tabular data e-commerce fraud detection and telecom churn prediction using Voting Ensemble technique, applying SMOTE for class balancing and feature engineering on both datasets.
- Achieved 0.90 F1 (92% accuracy) on fraud detection and 0.82 F1 (86% accuracy) on churn, with sub-second inference; evaluated using precision/recall to handle the cost asymmetry of false negatives in fraud with full stack development.

SKILLS

- **Programming & Data:** Python, JavaScript, Pandas, NumPy, Matplotlib, Seaborn
- **MLOps & Backend:** FastAPI, Flask, REST APIs, GitHub, OOP, Nginx
- **Machine Learning & Deep Learning:** PyTorch, TensorFlow, Keras, Scikit-learn, Supervised & Unsupervised Learning, Classification, Regression, Hyperparameter Optimization, Model Evaluation
- **NLP & Large Language Models:** Transformers, Hugging Face, Fine-Tuning, PEFT, LoRA, QLoRA, Token Optimization, spaCy, NLTK, Embeddings, GPT, LLaMA, Mistral, Claude
- **AI Agents & Multi-Agent Systems:** LangChain, LangGraph, AgentOps, CrewAI, LangSmith, Agentic RAG, Tool-Augmented Agents, Context Engineering, Prompt Engineering
- **Vector Databases:** Qdrant, FAISS, ChromaDB
- **Databases:** MySQL, MongoDB, PostgreSQL, Redis
- **Cloud & Deployment:** Docker, CI/CD Pipelines, Render

EDUCATION

City University, Bangladesh

2022 – 2025

Bachelor of Computer Science and Engineering

Honors & Awards

- Received 4 consecutive scholarships for excellence in Machine Learning and project performance.
- Awarded Merit Scholarship for outstanding academic excellence.

CERTIFICATIONS

- **Machine Learning Specialization** Stanford University – (Coursera) 
- **Programming for Everybody** University of Michigan – (Coursera) 
- **Google IT Support Specialization** Google – (Coursera) 