Fahad Siddique Faisal

Dhaka, Bangladesh

📞 +8801611346769 ✉️ Fahadsid1770@gmail.com in linkedin.com/in/fahadsiddique1770 github.com/fahadsid1770
🌐 fahad-portfolio-alpha.vercel.app

Research Interest

Human-Computer Interaction, Machine Learning, Deep Learning, Natural Language Processing, LLMs, AI

Education

Pabna Cadet College — **2012 – 2016**
Secondary School Certificate (SSC), Science Background — *GPA: 5.00*

Pabna Cadet College — **2016 – 2018**
Higher Secondary Certificate (HSC), Science Background — *GPA: 5.00*

Chittagong University of Engineering and Technology — **Feb 2019 – July 2024**
Bachelor of Science in Electronics and Telecommunication Engineering — *CGPA: 3.44*

Relevant Coursework

- C/C++ Programming
- Object-oriented Programming
- Numerical Analysis
- Algorithms and Data Structures
- Database Management
- Internet Programming
- Machine learning
- CCN
- Microcontroller
- Microprocessor
- VLSI Technology
- Multimedia programming

Experience

AI Solutions USA — **June 2025 – Present**
ML/AI Engineer — *New York, USA*

- Gained Proficiency in NoSQL (MongoDB) and Vector DBs (Lancedb, Qdrant, Milvus) and Production-Grade FastAPI development
- Designed a multi-agent orchestration framework with CrewAI, Instruct, and Agno to coordinate AI agents for collaborative task execution.
- Gained a strong understanding of Data-processing, Chunking techniques, Indexing, and vectorizing with embedding models and RAG-Techniques

Walton Hi-Tech Industries PLC — **Apr 2025 – Present**
ML/AI Engineer & Senior Deputy Assistant Director — *Kaliakoir, Chondra, Bangladesh*

- Developed a fully local AI-based knowledge-base and chat system using open-source AI models and Langchain and LangGraph with a Corrective-RAG pipeline, trained on internal data.
- Developed a Long-term memory system for any Conversational AI
- Working on an AI-based Real-time on-device Speech Recognition and Transcription system utilizing whisper.cpp and Open-sourced LLM models.
- Unlocked GPU inference capability of any LLM using PyTorch, CUDA, Llama.cpp, and containerized it using Docker for utilizing the local server's GPU rendering
- Gained proficiency in Finetuning AI models, Synthetic Data Generation, LLM Quantization

Outlier AI — **Oct 2024 – Feb 2025**
AI Data Curator

- Curated around 22GB of English to Bangla translated and Question Answering text
- Gained extensive knowledge regarding data curation and the importance of quality data for training AI models.
- Gained hands-on experience in developing datasets to train transformer-based and AI models

Riseup Labs | Jun 2023 – Jul 2023

Software Engineer Intern | *Uttara, Dhaka*

- Attained hands-on experience on Software development life-cycle and understanding the role of an expert programmer in SDLC.

Sheikh Kamal IT Business Incubator, CUET | Jun 2022 – Jul 2022

AI and Machine Learning Engineer Intern | *Chittagong*

- Developed Machine Learning algorithms like Regression and Classification, SVM, Neural Networks like CNN, TextCNN, RNN.
- Interacted with Computer Vision concepts like Model Fusion, Image Color Conversion, CNN with sequential API and functional API.
- Implemented Natural Language Processing models like LSTMs and Pretrained Word Embeddings like Word2Vec, GloVe, FastText.

Projects

Walton-AI: A Secure, In-House AI Assistant and AI-based Knowledge Management 🔗 | *Python, Langchain* **2025**

- Optimized a high-performance RAG pipeline utilizing the Llama-3.2-3B LLM, achieving a 70% reduction in memory requirements (from 12GB to 3.5GB) using 8-bit quantization (llama.cpp backed) while maintaining 99.5% of the original model's performance.
- Boosted organizational productivity by reducing employee knowledge search time from 20 minutes to seconds, saving thousands of work-hours monthly and accelerating customer issue resolution.
- A secure, 100% on local servers, thereby eliminating the critical risk of confidential data leakage

Real Time Image Pattern Detection and Recognition using OpenCV and CLIP Embeddings | *OpenCV, CLIP* **2025**

- Integrated Multi-Modal AI Pipeline: Developed an end-to-end computer vision system combining YOLOv8 for real-time refrigerator detection, CLIP embeddings for feature extraction, and Qdrant vector database for similarity search, enabling accurate pattern recognition across video streams with confidence scoring

Development of a 400M Parameter Small Language Model with Function-Calling and ONNX Deployment | **2025**

- Constructed from scratch a 400M-parameter language model incorporating function-calling capabilities, followed by ONNX-based model conversion and inference optimization to enable efficient experimentation and deployment in constrained environments.

PageWise - RAG based Question answering Chatbot | *Ollama,llama3.2,OpenAI,Pinecone,Langchain* **2024**

- Can import and save entire PDF data into a Pinecone or Faiss vector database.
- Both OpenAI and llama3.2 text embedding can be used for better Semantic Understanding and Efficient Similarity search into a vector database
- Implemeted using Langchain incorporation with OpenAI's GPT4o API and llama3.2 through Ollama locally

Multi-class Bangla News Classification using Early Fusion and Late Fusion | *Python, Bi-LSTM, CNN* **2022**

- Cleaned, Normalized, and Processed multiple Bangla News datasets by Bangla Tokenizer and Pad Sequence.
- Implemented Late Fusion by concatenating the output of BiLSTM, CNN, and BiLSTM+CNN
- Implemented Early Fusion with BiLSTM, CNN, and LSTM+CNN
- Tuned hyper-parameters like learning rate, batch size, and optimizer for finding better results.

Undergraduate Thesis

Investigation and Transforming Resource-Scarce Language Question Generation: A Study on Text-to-Text Transformer Models for Bengali | *Huggingface Transformers* **2024**

- Developed and improved an automatic FAQ generation system using text-to-text transformer models to enhance Question Generation (QG) in multilingual contexts.
- Fine-tuned custom multilingual transformer-based models in a low-resource setting

Researches and Publications

Optimizing Resource-Scarce Language Question Generation on Bengali: A Comparative Study of Transformer-Based Models 🔗 | *Huggingface, Kaggle* | **2024**

- Developed an automatic FAQ generation system using text-to-text transformer models to enhance Question Generation (QG) in multilingual contexts.
- Achieved Rouge 37.225, SacreBLEU 14.202, and METEOR 0.216 with a bn-TOK + mT5-base + Beam Search model, ensuring lexical accuracy, coherence, and semantic relevance.

Enhancing Bilingual Sarcasm Detection with an Advanced Ensemble Model: A Dual-Language Deep Learning Framework 🔗 | *Python, Colab* | **2024**

- Developed a Bilingual Sarcasm Detection model using an ensemble of BiLSTM and GRU to capture contextual dependencies in English and Bengali text.
- Achieved 97% and 89% accuracy on a news headline dataset and BanglaSARC dataset, leveraging pretrained word embeddings like Word2Vec and FastText.

Design and Comparative Analysis of 1-BIT ALU and Full Adder using MGDI Technique for High-Performance PLC Applications 🔗 | *Python* | **2024**

- Designed a programmable logic controller (PLC) circuit using Modified Gate Diffusion Input (MGDI) technique for high-speed, low-power applications in 90nm technology.
- Reduced transistor count by 58% (from 128 to 54 transistors) compared to conventional PLC units, optimizing chip area and power dissipation.
- Simulated and validated the design in Cadence Virtuoso, integrating 1-BIT ALU, 10-T Full Adder, and NAND logic with multiplexer-based selection.

Ongoing Research Projects

Distilled Neural Radiologist: Student–Teacher Light Vision-Language Models Leveraging LLaMA 3.2 11B for Explainable, High-Fidelity Chest X-Ray Interpretation and Autonomous Clinical Reporting 🔗 | *llama.cpp, Python* | **2025**

- The combination of structured visual classification via EfficientNetB7 (achieving 99.1% accuracy) and multimodal reasoning through the LLaMA 3.2 Vision-Instruct model successfully produces reliable, clinically coherent, and contextually fluent automated radiology reports, demonstrated by a strong METEOR score of 0.657 and a low Perplexity value of 2.31

A Federated and Explainable Machine Learning Framework for Robust Intrusion Detection and Network Security Enhancement in Industrial Internet of Things (IIoT) Environments. 🔗 | *Fedarated learning,XAI,IIOT* | **2025**

- The synergy of the proposed Federated and Explainable Hybrid XGB-LGBM framework resulted in remarkable IIoT intrusion detection performance, achieving the highest metrics (F1-score of 0.9374 and 93.71% accuracy) while simultaneously guaranteeing model transparency (via SHAP analysis) and data privacy across distributed industrial nodes

Conversational AI Model for the Medical Domain Adapted Using QLoRA on Qwen2-0.5B 🔗 | *SLM, AI, QLora* | **2025**

- The project adapts the model using the ruslanmv/ai-medical-chatbot dataset of patient–doctor conversations, optimizing training with cosine scheduling, gradient accumulation, and Hugging Face integration. Ongoing work involves evaluating domain accuracy and deploying the adapter for real-world medical assistance scenarios.

Technical Skills

Programming Languages: C, C++, Python, JavaScript, SQL, NoSQL, Matlab
Machine Learning: Scikit-learn, Keras, TensorFlow, PyTorch.
Data Science: Pandas, NumPy, Matplotlib
Electronics: Cadence, Logisim, CST, Livewire, Proteus.
Network Analysis: PacketTracer, HTML, CSS
Technologies/Frameworks: Linux, Git, GitHub, Tailwind CSS, React, Firebase
Developer Tools: Cursor VS Code, AWS, ONNX

Achievements

AI Creative Award **Sept, 2025**
Walton Hi-Tech Industries PLC *Dhaka*

- **Champion among 170 Engineering Team Participating**: AI-based Smart Recommendation System 🔗
- **7th-Position**: Walton-AI: Secure and Local Knowledge management system.🔗

Best Technical Paper in a session on Artificial Intelligence **Dec, 2024**
27th International Conference on Computer and Information Technology *Long Beach Hotel, Cox's Bazar*

Govt. Scholarship for Academic excellence **2018**
Higher Secondary Level *Rajshahi Board, Rajshahi*

Govt. Scholarship for Academic excellence **2016**
Secondary Level *Rajshahi Board, Rajshahi*

Leadership / Extracurricular

CUET Debating Society **July, 2023 – Sept, 2024**
Vice-President *Chittagong University of Engineering and Technology*

- Arranged a National Level Inter-University Debate Competition and Cultural Program

House of the Language and Literature **August,2023 – Sept,2024**
Organizing Secretary *Chittagong University of Engineering and Technology*

- Arranged a National Level Inter-University Debate Competition and Cultural Program

Infixion **2024**
Program Co-ordinator *Chittagong University of Engineering and Technology*

- Arranged a National Level Inter-University Programming and Machine Learning Contest, FIFA Gaming contest
- Performed anchoring in the Ceremonial and Cultural Show