Mukaffi Bin Moin

mukaffi28@gmail.com Mukaffi Bin Moin mukaffi28.github.io mukaffi-bin-moin mukaffi28

Research Interests

Natural Language Processing (NLP), Social Media Analysis, Low-Resource Languages, Large Language Models, LLM Agents, Computer Vision, Vision-Language Models, Multimodal AI, Trustworthy AI, Multimodal Agents, AI For Healthcare

Education

B. Sc. **Ahsanullah University of Science and Technology**, *Computer Science and Engineering* July 2019 to Dec 2023

- GPA: 3.563/4.0 (Transcript)
- **Rank:** 39th Among 133 Students
- **Undergraduate Thesis Title:** *Generative Adversarial Networks for Crop Disease: A Case Study with Potato Disease Classification and Instance Segmentation*
- **Supervisor:** Dr. Mohammad Shafiul Alam, Professor, Dept. of CSE. AUST

Publications (* denotes equal contribution)

Conference Proceedings

- Fatema Tuj Johora Faria*, **Mukaffi Bin Moin***, Rabeya Islam Mumu, Md Mahabubul Alam Abir, Abrar Nawar Alfy, and Mohammad Shafiul Alam., **"Motamot: A Dataset for Revealing the Supremacy of Large Language Models Over Transformer Models in Bengali Political Sentiment Analysis,"** 2024 IEEE Region 10 Symposium (TENSYMP), New Delhi, India, 2024, pp. 1-8, doi: 10.1109/TENSYMP61132.2024.10752197
- **Mukaffi Bin Moin**, Pronay Debnath, Usafa Akther Rifa, Rijeet Bin Anis. **"Assessing the Level of Toxicity Against Distinct Groups in Bangla Social Media Comments: A Comprehensive Investigation."** In: Ullah, A., Anwar, S. (eds) Proceedings of International Conference on Information Technology and Applications. ICITA 2024. Lecture Notes in Networks and Systems, vol 1248. Springer, Singapore. 10.1007/978-981-96-1758-6_46
- Fatema Tuj Johora Faria, **Mukaffi Bin Moin**, Ahmed Al Wase, Md Rabius Sani, Khan Md Hasib, and Mohammad Shafiul Alam. **"Classification of potato disease with digital image processing technique: a hybrid deep learning framework."** In 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0820-0826, doi: 10.1109/CCWC57344.2023.10099162.
- Fatema Tuj Johora Faria, **Mukaffi Bin Moin**, Md Mahfuzur Rahman, Md Morshed Alam Shanto, Asif Iftekher Fahim, and Md Moinul Hoque. **"Uddessho: An Extensive Benchmark Dataset for Multimodal Author Intent Classification in Low-Resource Bangla Language."** In: Ullah, A., Anwar, S. (eds) Proceedings of International Conference on Information Technology and Applications. ICITA 2024. Lecture Notes in Networks and Systems, vol 1248. Springer, Singapore. 10.1007/978-981-96-1758-6_32
- **Mukaffi Bin Moin**, Fatema Tuj Johora Faria, Swarnajit Saha, Bushra Kamal Rafa, and Mohammad Shafiul Alam. **"Exploring Explainable AI Techniques for Improved Interpretability in Lung and Colon Cancer Classification,"** In: Kumar, A., Swaroop, A., Shukla, P. (eds) Proceedings of Fourth International Conference on Computing and Communication Networks. ICCCN 2024. Lecture Notes in Networks and Systems, vol 1396. Springer, Singapore. https://doi.org/10.1007/978-981-96-6124-4_1
- Fatema Tuj Johora Faria, **Mukaffi Bin Moin**, Asif Iftekher Fahim, Pronay Debnath, and Faisal Muhammad Shah. **"Unraveling the Dominance of Large Language Models Over Transformer Models for Bangla Natural Language Inference: A Comprehensive Study."** In: Kumar, A., Swaroop, A., Shukla, P. (eds) Proceedings of Fourth International Conference on Computing and Communication Networks. ICCCN 2024. Lecture Notes in Networks and Systems, vol 1396. Springer, Singapore. https://doi.org/10.1007/978-981-96-6124-4_2
- Fatema Tuj Johora Faria, **Mukaffi Bin Moin**, Asif Iftekher Fahim, Pronay Debnath, and Faisal Muhammad Shah. **"BanglaMemeEvidence: A Multimodal Benchmark Dataset for Explanatory Evidence Detection in Bengali Memes."** Under Review in 2025 9th International Conference on Vision, Image and Signal Processing
- Saidur Rahman Sujon, Ahmadul Karim Chowdhury, Fatema Tuj Johora Faria, **Mukaffi Bin Moin**, and Faisal Muhammad Shah. **"Tackling Hallucination in Bengali NLP: Enhancing Paraphrase Generation, Reading Comprehension, and Formal Application Writing Using LLMs with Few-Shot Learning, Fine-Tuning, and RAG."** [Under Review in 2025 29TH INTERNATIONAL COMPUTER SCIENCE AND ENGINEERING CONFERENCE
- Fatema Tuj Johora Faria*, **Mukaffi Bin Moin***,Mohammad Shafiul Alam*, Ahmed Al Wase, Md Rabius Sani, and Khan Md Hasib. **"PotatoGANs: Utilizing Generative Adversarial Networks, Instance Segmentation, and Explainable AI for Enhanced Potato Disease Identification and Classification."** arXiv preprint arXiv:2405.07332 (2024). [Under Review in IEEE i-COSTE 2025] [Preprint Link]

Journals

- Fatema Tuj Johora Faria, **Mukaffi Bin Moin**, Zayeed Hasan, Md. Arafat Alam Khandaker, Niful Islam, Khan Md Hasib, M.F. Mridha, **"MultiBanFakeDetect: Integrating advanced fusion techniques for multimodal detection of Bangla fake news in under-resourced contexts,"** International Journal of Information Management Data Insights, Volume 5, Issue 2, 2025, 100347, ISSN 2667-0968, https://doi.org/10.1016/j.jjimei.2025.100347 (Q1)
- Fatema Tuj Johora Faria, **Mukaffi Bin Moin**, Busra Kamal Rafa, Swarnajit Saha, Md. Mahfuzur Rahman, Khan Md Hasib, M.F. Mridha, **"BanglaCalamityMMD: A Comprehensive Benchmark Dataset for Multimodal Disaster Identification in the Low-Resource Bangla Language,"** International Journal of Disaster Risk Reduction, Volume 130, 2025, 105800, ISSN 2212-4209, https://doi.org/10.1016/j.jjimei.2025.100347 (Q1)
- Fatema Tuj Johora Faria, **Mukaffi Bin Moin**, Ahmed Al Wase, Mehidi Ahmmed, Md Rabius Sani, and Tashreef Muhammad. **"Vashantor: a large-scale multilingual benchmark dataset for automated translation of bangla regional dialects to bangla language."** arXiv preprint arXiv:2311.11142 (2023). [Under Review in Array (Q1)] [Preprint Link]
- Fatema Tuj Johora Faria, **Mukaffi Bin Moin**, Md. Mahfuzur Rahman, Khan Md Hasib, Md. Jakir Hossen, and M. F. Mridha. **"MindSpeak-Bangla: A HumanLLM Collaborative Dataset for Chain-of-Thought Adaptation in Bangla Mental Health Advice Generation"** [Under Review in IEEE Open Journal of the Computer Society (Q1)]
- Md. Arafat Alam Khandaker*, Ziyan Shirin Raha*, **Mukaffi Bin Moin***, Dipta Biswas, and Khan Md Hasib. **"FanSpeak: A Bangla Dataset for Multi-Class Toxicity Detection in Sports Discourse and a Comparative Evaluation of PLMs and LLMs"** [Under Review in Information Processing and Management (Q1)]
- Fatema Tuj Johora Faria, **Mukaffi Bin Moin**, Pronay Debnath, Asif Iftekher Fahim, and Faisal Muhammad Shah. **"Explainable Convolutional Neural Networks for Retinal Fundus Classification and Cutting-Edge Segmentation Models for Retinal Blood Vessels from Fundus Images."** arXiv preprint arXiv:2405.07338 (2024). [Under Review in Journal of Visual Communication and Image Representation (Q1)] [Preprint Link]

Ongoing Research Projects

- BanglaMedQA: A Dataset for Adapting Zero-Shot Chain-of-Thought Reasoning in Bengali Medical Question Answering
- Cross-Cultural Moral Bias Detection in Story Understanding: Analyzing Intentionality and Fairness in LLM Judgments
- Breaking Silence: A Jailbreaking Prompt Framework for Generating Sensitive and Controversial Narratives in Bangla
- Chain-of-Thought Enhanced Counter-Response Generation for Online Bullying Mitigation

Work Experience

Brain Station 23 PLC., Software Engineer (AI/ML) — *Dhaka, Bangladesh*
May 2025 – Present

- **Agentic Footfall Monitoring and Analytics Platform**
  - Led the design and development of a real-time footfall analytics platform powered by an agentic multi-VLM architecture (GPT-4, Gemini, Qwen-2.5, BLIP-2) for autonomous perception, contextual reasoning, and decision-making across multi-camera environments.
  - Integrated YOLOv11 and ByteTrack for real-time detection and tracking, enabling accurate multi-camera identity association with temporal consistency.
  - Applied Tree-of-Thought prompting to enable agents to collaboratively resolve ambiguities, interpret crowd dynamics, and derive coherent contextual insights.
  - Implemented hybrid identity matching combining visual (ReID) and semantic (text-based) embeddings, improving cross-frame and cross-camera identity consistency.
  - Introduced an LLM-based verification layer to confirm or reject identity matches, reducing false positives and enhancing re-identification reliability.
  - Optimized the embedding pipeline and similarity search using adaptive thresholds and intelligent LLM bypass logic for scalable, low-latency retrieval.
  - Developed a VLM-centric re-identification approach, achieving 0.962 accuracy versus 0.829 using traditional ReID methods.
  - Built a React-based visualization interface supporting text, image, and hybrid similarity search, enabling intuitive data exploration and analytics.
  - Engineered temporalspatial analytics modules to automatically generate natural language summaries of entries/exits, occupancy, and movement trends for actionable space utilization insights.
  - Led architecture design, scalability planning, and deployment strategy, mentoring junior developers and ensuring a modular, production-grade implementation.

Technology used: GPT-4, Gemini, Qwen-2.5, Tree-of-Thought Prompting, Multi-Agent Reasoning, Vector DB, Temporal-Spatial Analytics, YOLO11, ByteTrack

- **Bangla Text-to-Speech (TTS)**
  - Developed a Bangla Text-to-Speech system from scratch using a custom in-house speech corpus, enabling natural and intelligible speech synthesis for multiple speakers.
  - Preprocessed and curated the raw Bangla text and audio dataset for model training, including text normalization, phoneme extraction, and alignment with audio.
  - Implemented multi-speaker conditioning and phoneme-level modeling to improve clarity, adaptability, and speaker differentiation.
  - Enhanced prosody and naturalness in generated speech, ensuring expressive and human-like intonation.
  - Designed and optimized GPU-parallel training pipelines using PyTorch and NCCL for scalable and efficient model training.
  - Conducted systematic evaluation of pronunciation accuracy, speaker similarity, and intelligibility, achieving a robust and reliable TTS solution.
  - Produced a scalable architecture capable of handling multiple speakers and large datasets while maintaining high-quality synthesis performance.

    Technology used: xTTSv2, Orpheus, PyTorch, NCCL, CUDA, Speech Processing Libraries, Phoneme Modeling, Multi-Speaker TTS Techniques

- **BS23 Workplace Monitoring & Analytics Platform**
  - Led the development of a real-time workplace and desk monitoring platform using multi-camera input, integrating YOLOv11 for detection, ByteTrack/DeepSORT for tracking, and Qwen-2.5/BLIP-2 for vision-language analysis.
  - Implemented activity classification to differentiate active vs. idle time for employees, providing privacy-first insights without personal identifiers.
  - Monitored desk occupancy, entries/exits, seating duration, and meeting participation using non-intrusive signals such as headphone detection as proxies for engagement.
  - Developed consistent ID tracking without facial recognition, ensuring a privacy-conscious design while enabling longitudinal activity analysis.
  - Generated automated logs and visual playback, producing timestamped natural language summaries for actionable workspace and productivity insights.
  - Incorporated collaboration intensity mapping, stress-level estimation, and temporal-spatial analytics to inform resource utilization and operational efficiency.
  - Designed for scalability and future enhancements, including queue wait time estimation, role-based detection, peak hour analysis, and occupancy alerts.
  - Provided a full-stack solution using Python (FastAPI), OpenCV for vision processing, CSV logging, and AI/ML models for vision-language reasoning and activity classification.

    Technology used: YOLOv11, ByteTrack, DeepSORT, Qwen-2.5, BLIP-2, Chain-of-Thought Prompting, OpenCV, Python, FastAPI, Temporal-Spatial Analytics, Zone Mapping, CSV Logging

ANTT Robotics Ltd., Machine Learning Engineer — *Dhaka, Bangladesh*
March 2024 – April 2025

- **Gift Shop Intelligent Chatbot**
  - Developed an intelligent chatbot for gift shops to provide personalized gift recommendations and real-time product queries using Retrieval Augmented Generation (RAG).
  - Integrated GPT-4, ChromaDB, and LangChain with a ReactJS frontend and FastAPI backend to enhance customer engagement and streamline operations.
  - Enabled retrieval of product information and personalized suggestions, improving customer satisfaction and sales efficiency.

    Technology used: GPT-4, RAG, FastAPI, LangChain, ChromaDB, OpenAI Embeddings, ReactJS

- **ANTT AI Bot**
  - Designed ANTT AI Bot, a chatbot using fine-tuned LLMs (GPT-3.5, GPT-4o, LLaMA-2) to generate platform-specific code for Arduino, ESP32, and Raspberry Pi.

- Simplified embedded systems programming by producing code for sensor interfacing, actuator control, and device communication.
- Tested on real hardware to ensure reliability and correctness of generated code.
  Technology used: Prompt Engineering, LangChain, Python, GPT-3.5, GPT-4o, LLaMA-2

- **Predictive Maintenance Platform for Appliances**
  - Developed a web and mobile predictive maintenance system for appliances, using time series models and boosting algorithms for real-time health monitoring.
  - Implemented adaptive learning by retraining models weekly with new operational data.
  - Provided real-time updates on key metrics including temperature, vibration, energy usage, and shock events to optimize performance and reliability.
  - Deployed on AWS EC2 with FastAPI backend, enabling scalable and responsive monitoring across devices.
  - Oversaw full-cycle project development, managing workflows, and ensuring the timely delivery of high-quality outputs.
  - Provided technical mentorship and encouraged knowledge sharing within the AI team.

  Technology used: Python, Multiple Time Series Models, Boosting Algorithms, AWS EC2, Adaptive Learning Techniques, FastAPI, Data Analysis Tools
- **Real-Time ANPR and Dynamic Toll Pricing System**
  - Developed a real-time ANPR and toll pricing system using YOLOv8 and OCR.
  - Detected vehicle plates and calculated tolls dynamically based on vehicle type, time, and traffic conditions, deployed on cloud infrastructure.
  - Integrated high-definition cameras and a user-friendly interface, optimizing toll collection efficiency and reducing manual intervention.

  Technology used: Python, YOLOv8, OpenCV, Supervision, FastAPI, Ultralytics, PyQt5, ByteTrack

Technical skills

Programming Languages: Python (NumPy, SciPy, Matplotlib, Pandas, Seaborn), C++, Java

Deep Learning Frameworks: TensorFlow, Keras, PyTorch

LLM Application Frameworks: LangChain, LangGraph, LangSmith, LlamaIndex

Web Development: HTML5, CSS3, JavaScript, FastAPI, Flask, React JS, Streamlit

Database: MySQL, PostgreSQL, MongoDB

Cloud Services: AWS, Amazon SageMaker, Amazon EC2, Amazon Lambda

Vector Database: ChromaDB, FAISS, Milvus, Qdrant, Pinecone

Others: Prompt Engineering, Context Engineering, Hugging Face Transformers, Docker, Apache Airflow, OpenCV, Git, GitHub Actions, Github Copilot

Awards & Achievements

Poster Presentation

Dhaka, Bangladesh.
5th August 2023

- "**Classification of Potato Disease with Digital Image Processing Technique: A Hybrid Deep Learning Framework**", secured 1st position in **"RESEARCH SYMPOSIUM 2023: AN INTRA-AUST RESEARCH EXHIBITION"** organized by AUST Research and Publication Club. **[Poster Link]**