

Md. Abdur Rakib Mollah

Machine Learning Engineer | Prospective PhD Student

 portfolio |  github |  linkedin |  rakib1703115@gmail.com |  +8801878099706 |  kaggle |  Bangladesh

Research Interests

My central research objective is advancing reasoning capabilities within Large Language Models and Generative AI systems. To achieve this overarching goal, my work explores: (a) refining internal model logic and alignment through contrastive signals; (b) grounding reasoning processes using Knowledge Graph-integrated Retrieval-Augmented Generation (RAG) frameworks; and (c) designing autonomous, multi-agent orchestration systems to execute multi-step reasoning tasks.

Experience

ResearchDen

Remote

Lead Researcher

Dec. 2024 – Present

- Lead collaborative AI and ML research groups, guiding first-time and seasoned researchers from ideation to publishing in peer-reviewed conferences and journals.
- Mentor community members on research methodology, literature review, and paper writing, fostering a collaborative culture of accountability.

AIML Professional Community Bangladesh

Bangladesh (Remote)

Mentor

Apr. 2025 – Present

- Mentor junior researchers on machine learning projects, providing guidance on research methodologies and technical implementations.
- Facilitate code reviews and conceptual discussions to improve model performance and adhere to industry best practices.

Euclido Inc.

Waterloo, Ontario, Canada (Remote)

Machine Learning Engineer-I

March 2024 – Present

- Design LangGraph-based RAG chatbots and implement knowledge graphs to map session relationships, enabling highly context-aware responses and deeper insights into user interactions.
- Develop autonomous AI agents to manage end-to-end ETL processes, successfully automating data ingestion, transformation, and loading to reduce manual intervention.

Department of CSE, RUET

Rajshahi, Bangladesh

Research Assistant

Jan. 2022 – July 2022

- Conducted extensive literature reviews and experimental analysis under faculty supervision to explore novel machine learning applications.
- Contributed to data processing and model evaluation, culminating in findings that were successfully published in a peer-reviewed research paper.

Publications

1. Moner Janala: A Curated Bengali Mental Health Advice Dataset from Newspaper Columns

- Curated a dataset of 625 Bengali mental health Q&A pairs from Prothom Alo's "Moner Janala" column, spanning 2009–2024.
- Provided a comprehensive resource for training and evaluating Bengali NLP models in psychological and counseling contexts.

Md. Abdur Rakib Mollah, Ann Naser Nabil

Mendeley Data, V1, 2026 (July 2026)

2. GRASP-ChoQ: Knowledge Graph-Based Retrieval Augmentation for Stance Detection in Political Texts with Chain-of-Questions Reasoning

- Addressed LLM struggles with identifying indirect political stances in dynamic, low-resource regional contexts.
- Introduced a Chain-of-Questions framework with Neo4j Knowledge Graph retrieval, boosting DeepSeek R1 stance detection F1 score to 93.33 (a 40.6% improvement).

3. **Adapting Contextual Embedding to Identify Sentiment of E-commerce Consumer Reviews and Addressing Class Imbalance Issues**

- Mitigated classification drops caused by severe class label imbalances in large-scale e-commerce datasets.
- Combined RoBERTa embeddings with XGBoost and class re-weighting, achieving 93.29% binary and 83.84% multi-class sentiment accuracy on 568k+ Amazon reviews.

Md. Abdur Rakib Mollah, Mir Md. Jahangir Kabir, Md. Sazid Reza, Monika Kabir

International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS), IEEE, 2024 (March 2024)

4. **Detection of Fake News with RoBERTa Based Embedding and Modified Deep Neural Network Architecture**

- Developed an advanced fake news detection system to capture subtle linguistic cues and evolving misinformation tactics.
- Achieved a state-of-the-art 99.76% accuracy on the 72,134-article WELFake dataset using RoBERTa embeddings and a custom deep neural network.

Md. Abdur Rakib Mollah, Mir Md. Jahangir Kabir, Monika Kabir, Md. Sazid Reza

26th International Conference on Computer and Information Technology (ICCIT), IEEE, 2023 (Dec. 2023)

5. **Exploring Deep Convolutional Neural Networks: A Grad-CAM Enhanced Comparative Study for Automated COVID-19 Diagnosis from Chest X-Ray Images**

- Designed transparent, cost-effective COVID-19 screening models for resource-limited clinical settings.
- Benchmarked five CNNs on 21,000+ X-rays using Grad-CAM heatmaps, reaching 99.13% diagnostic accuracy with a fine-tuned Inception V3 model.

Md. Sazid Reza, Mir Md. Jahangir Kabir, **Md. Abdur Rakib Mollah**, Nazmul Islam Nahin

3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), IEEE, 2023 (Dec. 2023)

6. **Improving Alzheimer's Disease Diagnosis on Brain MRI Scans with an Ensemble of Deep Learning Models**

- Improved early-stage Alzheimer's detection from MRI scans by mitigating overfitting issues common in single-model approaches.
- Constructed a CNN ensemble (ResNet-50, EfficientNet B7, DenseNet-121) that achieved peak multi-class accuracies of 99.96% across dual ADNI datasets.

Md. Sazid Reza, Mir Md. Jahangir Kabir, **Md. Abdur Rakib Mollah**

International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), IEEE, 2023 (Dec. 2023)

7. **An Ensemble Approach for Identification of Distracted Driver by Implementing Transfer Learned Deep CNN Architectures**

- Built a real-time computer vision framework to classify ten distinct driver distraction behaviors for road safety.
- Achieved 99.81% multi-class accuracy on the State Farm dataset using a parameter-efficient ensemble of fine-tuned MobileNet and DenseNet121 models.

Md. Abdur Rakib Mollah, Azmain Yakin Srizon, Mohiuddin Ahmed

25th International Conference on Computer and Information Technology (ICCIT), IEEE, 2022 (Dec. 2022)

Works Under Review & Undergraduate Thesis

1. **A Survey on Enhancing LLM Reasoning via Negative Supervision and Error Reflection**

Md. Abdur Rakib Mollah, Mir Md Jahangir Kabir

Submitted to ACL ARR (Under Review)

2. **Comparative Analysis of Tiny LLMs for Bangla Hate Speech and Emotion Detection Across Multiple Datasets**

Ann Naser Nabil, Saif Siddique, Rakib Ullah, **Md. Abdur Rakib Mollah**, Shamazda Islam, Md Sultanul Islam Ovi

Submitted to Journal (Under Review)

3. **Aligning Vision-Language Models for Financial Reasoning via DPO and Targeted Negative Sampling**
Md. Abdur Rakib Mollah, Mir Md. Jahangir Kabir
Ongoing Research
4. **Generating Synthetic Medical Data to Improve Clinical Reasoning in Monolingual and Code-Mixed Bengali**
Md. Abdur Rakib Mollah, Ann Naser Nabil, Md Sultanul Islam Ovi
Submitted to Journal (Under Review)
5. **Generation of Bangla Literature Texts by Fine-tuning Generative Pre-trained Transformer** 
Md. Abdur Rakib Mollah
B.Sc. Thesis, Department of CSE, RUET

Technical Skills

- **Programming & Scripting:** Python, C++, JavaScript/TypeScript, Cypher Query Language, LaTeX
- **Core ML & Vision Frameworks:** PyTorch, TensorFlow, Scikit-Learn, Pandas, NumPy, OpenCV, SSIM
- **Agentic AI & Orchestration:** LangGraph, LlamaIndex, CrewAI, AutoGen, LangChain
- **Methodologies & Techniques:** Retrieval-Augmented Generation (RAG), Parameter-Efficient Fine-Tuning (PEFT/LoRA), Model Quantization, Prompt Engineering, Explainable AI (Grad-CAM), Class Imbalance Mitigation, Entity Extraction (NER)
- **Data Engineering & Automation:** ETL Pipelines, Data Ingestion, Web Scraping, Browser Extension Development, DOM Manipulation
- **Databases, Infrastructure & Environments:** Neo4j, Pinecone, ChromaDB, Git, CI/CD, Docker, AWS

Education

Rajshahi University of Engineering and Technology (RUET)
 B.Sc. in Computer Science and Engineering (CGPA: 3.02/4)

Rajshahi, Bangladesh
 Jan. 2018 – Sep. 2023

Technical Projects

DetectVideoShotLength  *Python, SSIM, Video Processing*

- Developed a Python package for detecting scene cuts in videos using histogram and SSIM analysis.
- Implemented frame change identification, calculated shot durations, and visualized scene length distributions.

Bangla Text Generation using GPT-2  *Python, GPT-2, PyTorch, Hugging Face*

- Fine-tuned GPT-2 to generate Bangla text in the style of Satyajit Ray's writing.
- Created a text generation pipeline that replicates literary styles and compares language modeling strategies.

Bangtex  *Python, LaTeX, Transliteration*

- Created a Python package that automates the conversion of native Bangla into bangtex transliteration.
- Simplified document compilation for the pdfLaTeX compiler.

Low-Latency Retrieval Pipeline  *Python, RAG, Vector Search, E-commerce*

- Architected an optimized RAG system targeting sub-50ms query to retrieval latency.
- Implemented efficient vector indexing and custom search mechanisms to accelerate document retrieval.

Project Z: AI-Powered Socio-Economic Dashboard  *Python, Geospatial, Dashboard, AI Assistants*

- Created a unified geospatial dashboard overlaying hyper-local demographic data.
- Integrated an AI assistant to generate actionable development strategies for government policymaking.

Test Scores

- **IELTS Academic** (Overall 8.0 - L: 8.5, R: 8.5, W: 7.0, S: 7.0)  British Council, Dec. 2025