Md. Ekramul Islam

✉ ei.ekramul.islam@gmail.com   in Ekramul Islam   🔗 ekram007.github.io   ⭘ Ekram007

🎓 Md. Ekramul Islam   📞 +8801521434732   📍 Dhaka, Bangladesh

Publications

[1] **Islam**, **M**., Chowdhury, L., Khan, F., Hossain, S., Hossain, M., Rashid, M., Mohammed, N., & Amin, M. (2023). SentiGOLD: A Large Bangla Gold Standard Multi-Domain Sentiment Analysis Dataset and Its Evaluation. In *Proceedings of the 29th ACM **SIGKDD** Conference on Knowledge Discovery and Data Mining* (pp. 4207–4218). Association for Computing Machinery. ⬈

[2] Afrin, S., Chowdhury, **M**., **Islam**, M., Khan, F., Chowdhury, L., Mahtab, M., Chowdhury, N., Forkan, M., Kundu, N., Arif, H., Rashid, M., Amin, M., & Mohammed, N. (2023). BanLemma: A Word Formation Dependent Rule and Dictionary Based Bangla Lemmatizer. In *Findings of the Association for Computational Linguistics: **EMNLP** 2023* (pp. 3695–3710). Association for Computational Linguistics. ⬈

[3] Mahtab, M., Khan, F., **Islam**, **M**., Chowdhury, M., Chowdhury, L., Afrin, S., Ali, H., Rashid, M., Mohammed, N., & Amin, M. (2025). BanNERD: A Benchmark Dataset and Context-Driven Approach for Bangla Named Entity Recognition. In *Findings of the Association for Computational Linguistics: **NAACL** 2025* (pp. 6807–6828). Association for Computational Linguistics. ⬈

[4] Sarker, S., **Islam**, **M**., Saurav, J., & Nahid, M. (2020). Word Completion and Sequence Prediction in Bangla Language Using Trie and a Hybrid Approach of Sequential LSTM and N-gram. In *2020 2nd International Conference on Advanced Information and Communication Technology (**ICAICT**)* (pp. 162-167). ⬈

- Research Impact: **30+** Citations, h-index: **3**, i10-index: **2**

Professional Experience

Machine Learning Engineer,   Aug 2024 – Present | Remote
IQVIA (A Fortune 500 Company), North Carolina, US ⬈
- Developed a conversational marketing co-planner agent with LangGraph **tool-calling** capabilities to **automate** complex **ML workflows**. This agent enables dynamic **clustering** of over 2,000 Healthcare Professionals by allowing users to **select features**, filter datasets, and generate personalized marketing campaigns while providing **answers to any query** from the dataset. Recognized with Impact Award. ⬈
- Deployed open-source LLMs (e.g., DeepSeek, Qwen) on **AWS SageMaker** DJL Serving with **vLLM** backends, achieving **first-token latency** as low as **90ms.**
- Collaborated with cross-functional teams to develop **data-exploration-agent** using **LangGraph**, where agent parse user queries, generate **SQL Query**, retrieved data and generate final answers with visualizations.

Machine Learning Engineer,   Jan 2022 – Jul 2024 | Hybrid
Giga Tech Limited, Dhaka, Bangladesh ⬈
- ***Shallow Parsing:*** Developed data preparation, training, inference pipelines with API (FastAPI), PyPI package, achieving SOTA results (**96.47** FScore) on Penn Treebank.
- ***Sentiment Analysis:*** Conducted transformer based model experimentation, cross-dataset analysis, statistical significance tests, leading to a publication at **KDD '23** (First Author).
- ***Named Entity Recognition:*** Achieved **SOTA** results on all Bangla datasets, published in NAACL 2025.
- ***Lemmatization:*** Developed a Bangla lemmatizer with **98.17%** accuracy, published in EMNLP 2023.

Academic Credentials

Bachelor of Science in Computer Science and Engineering,   CGPA: 3.18 / 4.00
Shahjalal University of Science and Technology, Sylhet, Bangladesh
January 2017 - July 2021

Academic Research

Handwritten Character Generation (Undergraduate Thesis)
Experimented with **GAN's** (Conditional GAN, Auxiliary GAN etc) to synthesize images of Bangla handwritten characters from textual prompts. GitHub ⬀

Bangla Next Sequence Prediction
Proposed a solution using Trie and a combination of LSTM and N- gram to predict the relevant next sequence list in Bangla published in **IEE (ICAICT '20 ⬀).** GitHub ⬀

Music Generation
Conducted Midi music files collection, notes preparation and utilized **BiLSTM GAN**. *GitHub* ⬀

Personal Project

CrossPress Digest   Jan 2025 – Present
- Led backend and ML development for a system that **identifies duplicate news stories from multiple sources** using semantic search, and summarizes with Gemini.
- Implemented **search** service using vector database (Chroma) and LLM-based RAG to answer user queries.
- **Tools:** Elasticsearch, Sentence Transformers, MongoDB, Gemini API, Chroma, Docker, FastAPI, Python
- Live Site ⬀

Technical Skills

Programming Language — Python

Libraries — *PyTorch, Tensorflow, Hugging Face, Scikit-learn, OpenCV, LangGraph, Pandas, Numpy*

Cloud & DevOps — *AWS SageMaker, Bedrock, Docker, PyPI*

Database — *MongoDB, MySQL, Chroma, Elasticsearch*

APIs & Tools — *FastAPI, Git/GitHub*

Linguistic Proficiency

English — IELTS: Overall **7.5** (Listening 8.5, Reading 8.0, Writing 6.0, Speaking 6.5)

Bangla — Native Language

Extra-curricular Activities

- 11th in National Junior **Chess** Championship-2013, Bangladesh
- Former Executive member (2017-2019) at **RIM Musical Club**, SUST, Bangladesh
- Active in Dhaka College Science Club (2014-2016) and NBPSC MUN & Debating Club (2010-2014).
- Volunteered in different events and performed as a **guitarist** in cultural events.

References

Nabeel Mohammed, *Associate Professor,* North South University, Dhaka, Bangladesh
nabeel.mohammed@northsouth.edu

Mohammad Ruhul Amin, *Assistant Professor,* Fordham University, New York, USA
mamin17@fordham.edu

Mahadi Hasan Nahid, *Assistant Professor,* Shahjalal University of Science and Technology, Sylhet, Bangladesh
nahid-cse@sust.edu

Faisal Ahamed Khan, *Head of R&D,* Giga Tech Limited, Dhaka, Bangladesh
faisal.cse06@gigatechltd.com