

Md. Mohsinul Kabir

Github: github.com/mohsinulkabir14

Email: mohsinulkabir@iut-dhaka.edu

Mobile: +447494846125

RESEARCH INTEREST

Natural Language Processing

- Large Language Models (LLM), Multilinguality, Low Resource Languages, Social Network Analysis & Mining, Computational Social Science

EDUCATION

Doctor of Philosophy (PhD) - Computer Science

- University of Manchester, United Kingdom

Manchester, United Kingdom

October 2024 - Present

Master of Science - Computer Science & Engineering (CSE)

- Islamic University of Technology, Bangladesh; CGPA: **4.0** (out of 4.0)

Gazipur, Dhaka

July 2019 - April 2022

Bachelor of Science - Computer Science & Engineering (CSE)

- Islamic University of Technology, Bangladesh; CGPA: **3.75** (out of 4.0)

Gazipur, Dhaka

January 2015 - October 2018

Higher Secondary Certificate - Science

- Rajshahi College, Bangladesh; GPA: 5.0 (out of 5.0)

Rajshahi

April 2012 - May 2014

Secondary School Certificate - Science

- Govt. Laboratory High School, Bangladesh; GPA: 5.0 (out of 5.0)

Rajshahi

January 2004 - February 2012

EXPERIENCE

AI Coding Trainer

- Outlier.ai, Remote

November 2024 - Present

Teaching Staff

- Islamic University of Technology (IUT), Bangladesh

January 2019 - September 2024

Research Intern

- Samsung R&D Institute Bangladesh (Full-time)

November 2017 - January 2018

SKILLS

- Programming:** Proficient in C, C++, Python and \LaTeX ; Capable of C#, SQL
- Toolkits & Frameworks:** Git, PyTorch, TensorFlow, Keras, numpy, pandas

PUBLICATIONS

- Mohsinul Kabir**, Saidul Islam, Md Tahmid Rahman Laskar, Mir Tafseer Nayeem. BenLLMEval: A Comprehensive Evaluation into the Potentials and Pitfalls of Large Language Models on Bengali NLP. *LREC-COLING*, 2024.
- Mohsinul Kabir**, Faria Binte Kader, Tasmia Binte Sogir, Nafisa Hossain Nujhat, Fatin Abrar Shams. Unveiling Depression on Social Media: Active Learning with Human-in-the-Loop Labeling for Mental Health Data Annotation and Analysis. In *29th International Conference on Natural Language & Information Systems (NLDB 2024)*.
- Md Tariquzzaman, Wasif Kader, Audwit Anam, Naimul Haque, **Mohsinul Kabir**, Hasan Mahmud, and Md Kamrul Hasan. the_linguists at BLP-2023 Task 1: A Novel Informal Bangla Fasttext Embedding for Violence Inciting Text Detection. In *Proceedings of the 1st International Workshop on Bangla Language Processing (BLP-2023)*, Association for Computational Linguistics (ACL), 2023. (Best Shared Task Paper)
- Mohsinul Kabir**, Obayed Bin Mahfuz, Syed Rifat Raiyan, Hasan Mahmud, and Md Kamrul Hasan. BANGLABOOK: A Large-scale Bangla Dataset for Sentiment Analysis from Book Reviews. In *Findings of the Association for Computational Linguistics (ACL)*, pp. 1237–1247, 2023.
- Faria Binte Kader, Nafisa Hossain Nujat, Tasmia Binte Sogir, **Mohsinul Kabir**, Hasan Mahmud, and Md Kamrul Hasan. “When Words Fail, Emojis Prevail”: A Novel Architecture for Generating Sarcastic Sentences With Emoji Using Valence Reversal and Semantic Incongruity. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)* (Volume 4: **ACL Student Research Workshop**), pp. 334–351, 2023.
- Syed Rifat Raiyan, Md Nafis Faiyaz, Shah Md. Jawad Kabir, **Mohsinul Kabir**, Hasan Mahmud, and Md Kamrul Hasan. 2023. Math Word Problem Solving by Generating Linguistic Variants of Problem Statements. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)* (Volume 4: **ACL Student Research Workshop**), pp. 362–378, 2023.
- Mohsinul Kabir**, Tasnim Ahmed, Md Bakhtiar Hasan, Md Tahmid Rahman Laskar, Tarun Kumar Joarder, Hasan Mahmud, and Kamrul Hasan. DEPTWEET: A typology for social media texts to detect depression severities. *Computers in Human Behavior* 139: 107503, 2023. (Q1, Impact Factor: 9.9)

- Tasnim Ahmed, Shahriar Ivan, **Mohsinul Kabir**, Hasan Mahmud, and Kamrul Hasan. Performance analysis of transformer-based architectures and their ensembles to detect trait-based cyberbullying. *Social Network Analysis and Mining*, 12 (1) : 99, 2023. **(Q1, Impact Factor: 2.8)**
- Tasnim Ahmed, **Mohsinul Kabir**, Shahriar Ivan, Hasan Mahmud, and Kamrul Hasan. Am I being bullied on social media? an ensemble approach to categorize cyberbullying. In 2021 *IEEE international conference on big data (IEEE Big data)*, pp. 2442-2453. IEEE, 2021.
- **Mohsinul Kabir**, Quazi Fahim Faisal Dhruba, Hasan Mahmud, Md Kamrul Hasan, and Ahsan Rejwan Zaman. Gaming Insight: Conversion of Popular Sedentary Games into Motion-Based Form. *International Journal of Human-Computer Interaction*, 36 (13) : 1205-1215, 2023. **(Q1, Impact Factor: 4.7)**
- Nazmus Sakib, G. M. Shahariar, **Mohsinul Kabir**, Md Kamrul Hasan, and Hasan Mahmud. Assorted, Archetypal and Annotated Two Million (3A2M) Cooking Recipes Dataset Based on Active Learning. In *International Conference on Machine Intelligence and Emerging Technologies*, pp. 188-203. Cham: Springer Nature Switzerland, 2022.

UNDER REVIEW MANUSCRIPTS

- Md Tariquzzaman, Audwit Anam, Naimul Haque, **Mohsinul Kabir**, Hasan Mahmud, and Md Kamrul Hasan. BDA: Bangla Text Data Augmentation Framework *The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)*. **(Under Review)**
- Ajwad Abrar, Nafisa Tabassum Oeshy, **Mohsinul Kabir**, Sophia Ananiadou. Religious Bias Landscape in Language and Text-to-Image Models: Analysis, Detection, and Debiasing Strategies *The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)*. **(Under Review)**
- Faria Binte Kader, Nafisa Hossain Nujat, Tasmia Binte Sogir, **Mohsinul Kabir**, Hasan Mahmud, Kamrul Hasan, Computational Sarcasm Analysis on Social Media: A Systematic Review. *Knowledge and Information Systems*. **(Under Review)**
- Nazmus Sakib, G. M. Shahariar, **Mohsinul Kabir**, Md Kamrul Hasan, and Hasan Mahmud. Towards Automated Recipe Genre Classification using Semi-Supervised Learning. *Knowledge and Information Systems*. **(Under Review)**

PROJECTS

- **Evaluations of Large Language Models (LLM)** : This project aims to assess if ChatGPT and other LLMs can be effectively applied to languages other than English through comprehensive evaluations across various tasks and languages, using large-scale datasets. (2023-Ongoing).
- **Bias Mitigation of LLMs**: Despite excellent advancements, LLMs exhibit various biases that can cause representational harm. The purpose of this project is to detect, quantify, and mitigate the inherent bias of LLMs using various means. (2024-Ongoing).

HONORS AND AWARDS

- Bronze Medal in **Asian Pacific Astronomy Olympiad (APAO)**, Bangladesh. (December 2012)
- Participation in **International Olympiad on Astronomy & Astrophysics (IOAA)**, Greece. (August 2013)
- **Best Shared Task Paper** award in EMNLP Bangla Language Processing Workshop. (December 2023)
- Top 5 in the UNDP Earthquake Shake-It-App Challenge. (December 2016)
- Champion in AUW National IT Fest in Mobile Application Category. (April 2017)
- Champion in IUT Esonance in Project Showcasing Category. (April 2017)
- Finalist in IUT Lawn Tennis Singles and Doubles Tournament. (October 2018)
- Passed Undergrad with a First class with Honors. (November 2018)
- Received OIC Partial Scholarship for Completing Undergrad Study. (January 2015- October 2018)

SERVICE

- **Conference Reviewing**
- EMNLP 2023, EMNLP 2024

TEACHING EXPERIENCE

- **Data Engineering (University of Manchester COMP 60711)** Fall 2024
- Graduate Teaching Assistant
- **Natural Language Processing (IUT SWE 4841)** Summer 2023
- Instructor
- **Linear Algebra (IUT Math 4341)** Winter 2021, Winter 2022, Winter 2023
- Instructor
- **Integral Calculus and Differential Equations (IUT Math 4241)** Summer 2020, Summer 2021
- Instructor
- **Discrete Mathematics (IUT CSE 4203)** Summer 2019
- Instructor