

MEZBAUR RAHMAN

✉ mezbaur00797@gmail.com 🌐 [Website](#) 🐙 [Github](#) 🔗 [Linkedin](#) ☎ (312) 561-7899

RESEARCH INTERESTS

Large Language Models | Vision Language Models | Learning from Noisy Labels | Semi Supervised Learning

EDUCATION

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| University of Illinois Chicago Ph.D. in Computer Science and Engineering Current CGPA: 4.00 out of 4.00 | <i>Chicago, United States</i> Aug 2023 - Present |
| Islamic University of Technology M.Sc in Computer Science and Engineering CGPA: 3.96 out of 4.00 | <i>Gazipur, Bangladesh</i> Jan 2020 - June 2023 |
| Islamic University of Technology B.Sc in Computer Science and Engineering CGPA: 3.86 out of 4.00 (4 th in class) | <i>Gazipur, Bangladesh</i> Jan 2016 - Nov 2019 |

PROFESSIONAL EXPERIENCE

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| University of Illinois Chicago Graduate Research Assistant — advised by Prof. Cornelia Caragea <i>Department of Computer Science, dl4nlpspace Lab</i> | <i>Chicago, United States</i> Aug 2023 – Present |
| <ul style="list-style-type: none">• Conducting research at the intersection of Natural Language Processing and Machine Learning, focusing on semi-supervised learning and learning from noisy labels using guidance from Large Language Models (LLMs) and Vision-Language Models (VLMs).• Developed novel frameworks that leverage LLMs for label refinement, pseudo-labeling, and historical training dynamics-aware co-training, improving overall text classification performance. | |
| Islamic University of Technology Lecturer, Department of Computer Science and Engineering | <i>Gazipur, Bangladesh</i> Jan 2020 – Jul 2023 |
| <ul style="list-style-type: none">• Taught and designed undergraduate courses on Data Structures, Algorithms, and Database Systems; supervised research and capstone projects in Data Mining and NLP. | |
| Samsung R&D Institute Bangladesh Software Engineering Intern (Mobile Application Group) | <i>Dhaka, Bangladesh</i> Nov 2018 – Jan 2019 |
| <ul style="list-style-type: none">• Worked on developing an iOS app that simulated dynamic light projection and shading effects on 3D objects using OpenGL ES and native iOS rendering classes. | |

PUBLICATIONS

- Mohsinul Kabir, Tasnim Ahmed, **Md Mezbaur Rahman**, Polydoros Giannouris, and Sophia Ananiadou. 2026. [Semantic Label Drift in Cross-Cultural Translation](#). arXiv preprint. *To appear* in the **Language Resources and Evaluation Conference (LREC 2026)**.
- **Md Mezbaur Rahman** and Cornelia Caragea. 2025. [LLM-Guided Co-Training for Text Classification](#). In **Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP 2025)**, pages 31092–31109, Suzhou, China. Association for Computational Linguistics.
- Nishat Raihan, Mohammad Anas Jawad, **Md Mezbaur Rahman**, Noshin Ulfat, Pranav Gupta, Mehrab Mustafy Rahman, Shubhra Kanti Karmakar, and Marcos Zampieri. 2025. [Overview of BLP-2025 Task 2: Code Generation in Bangla](#). In **Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025)**, pages 373–387, December 23, 2025. Association for Computational Linguistics.

- **Md Mezbaur Rahman**, Mohammed Saidul Islam, Md Tahmid Rahman Laskar, Md Azam Hossain, and Abu Raihan Mostofa Kamal. 2023. *Multihop Factual Claim Verification Using Natural Language Prompts*. In **Proceedings of the Canadian Conference on Artificial Intelligence (Canadian AI 2023)**. Canadian Artificial Intelligence Association (CAIAC), June 2023.
- Syed Mohammed Sartaj Ekram, Adham Arik Rahman, Md Sajid Altaf, Mohammed Saidul Islam, Mehrab Mustafy Rahman, **Md Mezbaur Rahman**, Md Azam Hossain, and Abu Raihan Mostofa Kamal. 2022. *BanglaRQA: A Benchmark Dataset for Under-Resourced Bangla Language Reading Comprehension-Based Question Answering with Diverse Question-Answer Types*. In **Findings of the Association for Computational Linguistics: EMNLP 2022**, pages 2518–2532, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- **Md Mezbaur Rahman**, Saadman Malik, Mohammed Saidul Islam, Fardin Saad, Md Azam Hossain, and Abu Raihan Mostofa Kamal. 2022. *An Efficient Approach to Automatic Tag Prediction from Movie Plot Synopses Using Transformer-Based Language Model*. In **Proceedings of the 25th International Conference on Computer and Information Technology (ICCIT 2022)**, pages 501–505. IEEE.
- Mohammed Saidul Islam, Iqram Hussain, **Md Mezbaur Rahman**, Se Jin Park, and Md Azam Hossain. 2022. *Explainable Artificial Intelligence Model for Stroke Prediction Using EEG Signal*. *Sensors*, 22(24):9859. MDPI.

PREPRINTS

- Mohsinul Kabir, Tasnim Ahmed, **Md Mezbaur Rahman**, Shaoxiong Ji, Hassan Alhuzali, and Sophia Ananiadou. 2026. *XCR-Bench: A Multi-Task Benchmark for Evaluating Cultural Reasoning in LLMs*. arXiv preprint. Submitted to the **64th Annual Meeting of the Association for Computational Linguistics (ACL 2026)**.

SKILLS

| | |
|-------------------------------|------------------------------------------------------------------|
| ML Frameworks | PyTorch, Transformers, RLHF, DeepSpeed, vLLM, Hugging Face, LoRA |
| Programming Languages | Python, C/C++, Java, Bash/Shell |
| Data Science Libraries | NumPy, Pandas, Matplotlib, Scikit-learn |
| Tools & Systems | Git, Docker, Kubernetes, LaTeX/Overleaf, MySQL, Oracle SQL |

ONLINE CERTIFICATION

- **Neural Networks and Deep Learning** by DeepLearning.AI on Coursera
- **Natural Language Processing with Probabilistic Models** by DeepLearning.AI on Coursera
- **Natural Language Processing with Classification and Vector Spaces** by DeepLearning.AI on Coursera

SIGNIFICANT PROJECTS

- **Target Invariant Stance Detection Using Large Language Models** ([GitHub Link](#))
 - Explored the performance of Large Language Models (LLMs) in stance detection under zero-shot and fine-tuned settings.
 - Compared 7B-parameter models against smaller fine-tuned models to analyze trade-offs between model size, contextual reasoning, and generalization.
- **LLM-Guided Node Classification in Semi-Supervised Settings** ([GitHub Link](#))
 - Integrated LLM-based sentence encoders with Text-Attributed Graphs to enrich node representations.
 - Employed LLM-guided pseudo labels to initialize semi-supervised learning, improving performance with increasing unlabeled proportions.