

Khondoker Ittehadul Islam

NLP MASTERS STUDENT

✉ khondoker.islam9@gmail.com | 🏠 khondokerislam.github.io | 📺 KhondokerIslam | 🌐 -kh

Education

University of Groningen

Groningen, Netherlands

MASTER'S IN NATURAL LANGUAGE PROCESSING

2024 - Now

- Courses: **Advanced NLP** (9.2/10), **Deep Learning Practicals** (8.0/10), **Learning from Data** (8.0/10).
- Collaborated with CLCG's (<https://www.rug.nl/research/clcg/>) research group to evaluate decoder-only LLMs' reasoning.
- Researched on the QA domain leveraging NLI, RAG, translation & attribution techniques on a cross-lingual experiment.
- Discovered that for the same ground-reasoning chain, different steps influence the model's responses across various languages.
- Submitted at: IJCNLP-AACL 2025 (<https://arxiv.org/pdf/2508.08933>).
- CGPA: **7.7/10**

Shahjalal University of Science and Technology

Sylhet, Bangladesh

B.SC. (ENGG.) IN COMPUTER SCIENCE AND ENGINEERING

2017 - 2021

- In my undergraduate final year, I evaluated the mBERT's performance on low-resource languages on multiple downstream classification tasks. Developed a corpus generalization metric. Formulated an algorithm to evaluate the annotation quality of 80+ annotators.
- **Presented at:** Findings of EMNLP 2021, AACL-IJCNLP 2022.
- CGPA in Major Courses: 3.52/4.00

Professional Experience

2024 - 2025 **Research Assistant**, University of Groningen

- Analyzed the reasoning capabilities of LLM models across languages, generation strategies, and parameter sizes.
- These works are done with InCLow research group: <https://inc-low-lm.github.io/>.
- Publications: Submitted at IJCNLP-AACL 2025 and ongoing.

2025 - 2025 **Research Scholar**, University of Groningen

- Improved linguistic capabilities of vLMs for OCR task.
- Applied page-separation preprocessing to reduce the image resolution by half for better content analysis.
- Report: <https://arxiv.org/pdf/2508.10356>

2025 - 2025 **Research Scholar**, University of Groningen

- Leveraged emotional cues in debate speech for fallacy detection.
- Mapped tonal emotion values into textual representations to improve comprehension of audio features.
- Report: <https://arxiv.org/pdf/2509.11127>

2022 - 2022 **Research Scholar**, Fordham University

- Created a Bangla fake news dataset by collecting source articles.
- Performed manual annotation with explanatory justifications.

2019 - 2021 **Research Scholar**, SUST

- Evaluated the performance of encoder-only models on low-resourced Bangla language.
- These works were done with SUST NLP Research Group. It resulted in two publications.
- Findings of EMNLP 2021: <https://aclanthology.org/2021.findings-emnlp.278>
- AACL-IJCNLP 2022: <https://aclanthology.org/2022.aacl-short.17/>

2021 - 2024 **Software Engineer**, REVE Systems

- **Project Details:**
- Automating the administrative workflow process of Bangladesh Transport Ministry.
- Modules include Vehicle Maintenance, Human Resource Management, Bill and Budget Management, and more.
- **Roles & Responsibility:**
- Communicating with clients and incorporating their requirements.
- **Notable Implementations:**
- Developed administrative benchmark reports for business modules using data structures and algorithms.
- Improved operational insights related to inventory and incident through various analytics.
- The project is in production: <https://smart.dgt.gov.bd>.

Publications

- **Khondoker Ittehadul Islam**, Sudipta Kar, Md Saiful Islam, and Mohammad Ruhul Amin; SentNoB: A Dataset for Analysing Sentiment on Noisy Bangla Texts. **Findings of EMNLP 2021**.
- **Khondoker Ittehadul Islam***, Tanvir Hossain Yuvraz*, Md Saiful Islam, and Enamul Hassan; EmoNoBa: A Dataset for Analyzing Fine-Grained Emotions on Noisy Bangla Texts. **AACL-IJCNLP 2022**.

Skills

Technical Stack Python, PyTorch, HF's Transformers, vLLM, Git, Linux

Research Stack Paper Reviewing, Academic Writing, Grant Writing, Project Management

Awards

- 2025 **Best NLP Student Project Award**, @GroNLP, CLCG
Competed with 7 student teams in conducting and producing quality research.
- 2025 **LxMLS Scholarship**, Google & Zendesk
Supported expenses to participate in Lisbon Summer School.
- 2025 **Erasmus+ Scholarship**, Erasmus Mundus
Stipend to support exchange students.
- 2024 **NL Scholarship**, Nuffic, Dutch Ministry of Education, Culture, and Science
Grant to support researchers in Master's program.
- 2024 **Consortium Scholarship**, Erasmus Mundus
Reduced annual program fee for Joint Master's Degree students.
- 2023 **Outstanding Reviewer**, BNLP @ EMNLP 2023
Provided feedback to assist program chairs and to help authors in improving their work.
- 2021 **Student Travel Scholarship**, EMNLP 2021
Supported expenses to attend the conference.

Academia Services

- 2024-2025 **Programme Committee Member**, ReMa Linguistics, CLCG, University of Groningen
Represented student cohort and provided insights on course structure to the Dutch Education Ministry.

Complementary Education

- 2025 **15th LxMLS Summer School**, Instituto Superior Técnico (IST)
Courses: Learning with Linear Models, Introduction to Sequence Models, Pretrained Transformer Models, Causality
- 2025 **Erasmus Mundus LCT Summer School 2025**, DFKI & Saarland University
Workshops: Low-Resource Machine Translation, Interpreting Transformer Internals with Sparse Auto-encoders
- 2025 **ASML Student Excursion Day**, ASML HQ
Day-long hands-on event at the ASML Headquarters.
- 2020 **Fundamental of Reinforcement Learning**, Alberta Machine Intelligence Institute via Coursera (University of Alberta)
An online non-credit course authorized by University of Alberta, Alberta Machine Intelligence Institute and offered through Coursera. CREDENTIAL ID: VYBJSWD5UKCH