Khondoker Ittehadul Islam

FINAL-YEAR EM-LCT MASTERS STUDENT

📱+49 163 3705859 | ✉ khondoker.islam9@gmail.com | 🏠 khondokerislam.github.io | ⬛ KhondokerIslam | in -kh | 🎓 Google Scholar

Education

University of Groningen, Saarland University   *Groningen & Saarbrücken*

DOUBLE DEGREE MASTER'S IN NATURAL LANGUAGE PROCESSING   *2024 - 2026 (expected)*

· Courses: **Advanced NLP** (9.2/10), **Deep Learning Practicals** (8.0/10).
· CGPA: 7.7/10

Shahjalal University of Science and Technology   *Sylhet, Bangladesh*

B.Sc. (ENGG.) IN COMPUTER SCIENCE AND ENGINEERING   *2017 - 2021*

· CGPA: 3.35/4.00
· Thesis: Multi-lingual BERT in Low-resource Language Tasks.

Publications

· **Khondoker Ittehadul Islam**, Gabriele Sarti; Reveal-Bangla: A Dataset for Cross-Lingual Multi-Step Reasoning Evaluation. **BLP workshop @ IJCNLP-AACL 2025**. ⤢

· **Khondoker Ittehadul Islam**, Tanvir Hossain Yuvraz, Md Saiful Islam, Enamul Hassan; EmoNoBa: A Dataset for Analyzing Fine-Grained Emotions on Noisy Bangla Texts. **AACL-IJCNLP 2022**. ⤢

· **Khondoker Ittehadul Islam**, Sudipta Kar, Md Saiful Islam, Mohammad Ruhul Amin; SentNoB: A Dataset for Analysing Sentiment on Noisy Bangla Texts. **Findings of EMNLP 2021**. ⤢

· **Khondoker Ittehadul Islam**, Md Saiful Islam, and Md Ruhul Amin; Sentiment analysis in Bengali via transfer learning using multi-lingual BERT. **ICCIT 2020**. ⤢

Professional Experience

2024 - 2025   **Research Intern**, InCLow Lab, University of Groningen

· Analyzed LLM reasoning across languages, generation strategies, and parameter scales.
· Utilized NLI, RAG, and attribution techniques in a cross-lingual evaluation.
· Observed competitive low-resource performance with provided reasoning steps.
· Publications: *BLP workshop @ IJCNLP-AACL 2025* ⤢ and on-going.

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

· Digitalized the administrative workflow process of a Transport Company.
· Incorporated Theory-of-Computation (ToC) concepts to automate the chain-of-decisions.
· Improved operational insights through various analytics with cross-functional team.
· Production: https://smart.dgt.gov.bd/login ⤢

2019 - 2021   **Research Intern**, SUST NLP Research Group, Shahjalal University of Science & Technology

· Evaluated NN models on low-resourced Bangla language.
· Developed a high-sparsity corpus and an annotation-quality metric.
· Reported mBERT's critical performance on two downstream tasks.
· Publications: *Findings of EMNLP 2021* ⤢, *AACL-IJCNLP 2022* ⤢

Student Projects

Attributing Multilingual LLM CoT Reasoning Answers   *2025*

University of Groningen   *Groningen, Netherlands*

· Attributed step-level reasoning in Qwen2.5-1.5B's CoT outputs.
· Identified excessive attribution score on the final reasoning step.
· Observed performance strongly tied to language resource availability.
· Received *Best NLP Student Project Award*.
· *Report*: https://arxiv.org/pdf/2511.15886 ⤢

OCR on Historical Documents   *2025*

University of Groningen   *Groningen, Netherlands*

- Improved linguistic capabilities of vLMs for OCR tasks.
- Composited image to augment ancient documents.
- Reduced image resolution through page-separation preprocessing.
- Heuristically pseudo-labeled multi-class unlabeled documents.
- *Report*: https://arxiv.org/pdf/2508.10356 ↗

Fallacy Detection On Presidential Speech   *2025*

University of Groningen   *Groningen, Netherlands*

- Leveraged emotional cues in debate speeches.
- Mapped paralinguistic features into text to enhance reasoning.
- *Report*: https://arxiv.org/pdf/2509.11127 ↗

Offensive Text Classification Leveraging Sentiment   *2024*

University of Groningen   *Groningen, Netherlands*

- Annotated texts with sentiment using DeBERTa V3.
- Observed no positive-sentiment effect on offensive classification.
- *Report*: https://arxiv.org/pdf/2412.17825 ↗

Affect of Word Embedding on Noisy Bangla Text   *2020*

Shahjalal University of Science & Technology   *Sylhet, Bangladesh*

- Evaluated diverse embeddings for low-resource text classification.
- Found BERT's BPE tokenization achieves competitive performance.
- *Publication*: ICCIT 2020 ↗

Skills

Technical Stack	Python, Java, R, PyTorch, HF's Transformers, vLMs, Git, LaTex
Interpretable ToolKit	Seaborn, t-SNE, inSeq, ContextCite, LogitLens

Awards

2025 **Best NLP Student Project Award**, @GroNLP, University of Groningen
Competed with 6 other student teams to produce impactful research.

2025 **Google & Zendesk Scholarship**, LxMLS 2025
Received expenses to participate in Lisbon ML Summer School.

2025 **Erasmus+ Scholarship**, Erasmus Mundus
Received stipendium to assist my research traineeship.

2024 **NL Scholarship**, Nuffic, Dutch Ministry of Education, Culture, and Science
Awarded stipendium to support my academic endeavors.

2023 **Outstanding Reviewer**, BLP Workshop @ EMNLP 2023
Assisted program chairs and authors through constructive feedback.

2021 **Student Travel Scholarship**, EMNLP 2021
Received financial coverage to present my proceedings paper.

Academia Services

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