

Md Fahim

✉ fahimcse381@gmail.com | 🌐 md-fahim.github.io | 🎓 Google Scholar
🌐 LinkedIn | 🌐 md-fahim | 🌐 Kaggle | 🌐 Orcid Id

EDUCATION

• University of Dhaka

BSc in Computer Science & Engineering

Jan 2017 - Jan 2022

Dhaka, Bangladesh

◦ CGPA: 3.51/4.00 (CGPA 3.74 in last three semesters)

RESERACH INTEREST

NLP: Low Resource Language; LLM Reasoning & Concept in LLM; Cultural & Value Alignment in LLM

Multimodality: Modality Gap & Alignment; Multimodal Adapter; Multimodal Self-supervised Learning

EXPERIENCE

• Center for Computational & Data Sciences (CCDS Lab), IUB 🌐

Research Assistant

Sep 2021 - Present

Dhaka, Bangladesh

◦ **Current Project:** Concept and Value Alignment in LLM, Low Resource Language, Reasoning in LLM
Hate Speech & Misinformation Detection, Modality Gap & Alignment, Adapter-based Model

◦ **Publications:** Papers Accepted in ECAI, ICPR, NeurIPS-WS, ICML-WS, EMNLP-WS

• Penta Global Limited 🌐

AI/ML Researcher [Part-Time]

Feb 2024 - Present

Dhaka, Bangladesh

◦ **Working Project:** LLMs in Production, RAG and Prompt Designing, Social Media Text Analysis

◦ **Publications:** Papers Accepted in EMNLP, CLEF

PUBLICATIONS

C=CONFERENCE, J=JOURNAL, W=WORKSHOP

- [C6] Fabiha Haider*, Fariha Tanjim Shifat*, Md Farhan Ishmam*, Deeparghya Dutta Barua, Md Sakib Ul Rahman Sourove, **Md Fahim**, Farhad Alam Bhuiyan - **BanTH: A Multi-label Hate Speech Detection Dataset for Transliterated Bangla**. In *Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL) - Findings* | 2025
- [C5] **Md Fahim***, Fariha Tanjim Shifat*, Md Farhan Ishmam*, Deeparghya Dutta Barua, Fabiha Haider, Md Sakib Ul Rahman Sourove, Farhad Alam Bhuiyan - **BanglaTLit: A Benchmark Dataset for Back-Transliteration of Romanized Bangla**. In *Empirical Methods in Natural Language Processing (EMNLP) - Findings* | 2024
- [C4] Fahim Ahmed*, **Md Fahim***, Amin Ahsan Ali, Md Ashraful Amin, AKM Mahabubur Rahman - **Improving the Performance of Transformer-based Models Over Classical Baselines in Multiple Transliterated Languages [Full Talk Presentation]**. In *27th EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE (ECAI)* | 2024
- [C3] Farhan Noor Dehan*, **Md Fahim***, Amin Ahsan Ali, Md Ashraful Amin, AKM Mahabubur Rahman - **TinyLLM Efficacy in Low-Resource Language: An Experiment on Bangla Text Classification Task [Oral]**. In *27th International Conference on Pattern Recognition (ICPR)* | 2024
- [C2] Faria Sultana*, Md Tahmid Hasan Fuad*, **Md Fahim**, Rahat Rizvi Rahman, Meheraj Hossain, Md Ashraful Amin, AKM Mahabubur Rahman and Amin Ahsan Ali - **How Good are LM and LLMs in Bangla Newspaper Article Summarization?**. In *27th International Conference on Pattern Recognition (ICPR)* | 2024
- [C1] **Md Fahim**, Amin Ahsan Ali, Ashraful Amin, AKM Mahabubur Rahman - **EDAL: Entropy based Dynamic Attention Loss for HateSpeech Classification [Oral]**. In *37th Pacific Asia Conference on Language, Information and Computation (PACLIC)* | 2023
- [J1] Mohd. Sayemul Haque, Md. Fahim, Muhammad Ibrahim **An Exploratory Study on Simulated Annealing for Feature Selection in Learning-to-rank**. *International Journal of Intelligent Systems and Applications (IJISA)*, Vol. 16, Issue 4, Year 2024.
- [W9] Fariha Tanjim Shifat, Fabiha Haider, Md Sakib Ul Rahman Sourove, Deeparghya Dutta Barua, Md Farhan Ishmam, **Md Fahim**, Farhad Alam Bhuiyan - **Penta NLP at EXIST 2024 Task 1-3: Sexism Identification, Source Intention, Sexism Categorization In Tweets**. In *EXIST at Conference and Labs of the Evaluation Forum (CLEF)* | 2024
- [W8] Deeparghya Dutta Barua, Md Sakib Ul Rahman Sourove, Fabiha Haider, Fariha Tanjim Shifat, Md Farhan Ishmam, **Md Fahim**, Farhad Alam Bhuiyan - **Penta ML at EXIST 2024: Tagging Sexism in Online Multimodal Content With Attention-enhanced Modal Context**. In *EXIST at Conference and Labs of the Evaluation Forum (CLEF)* | 2024
- [W7] **Md Fahim**, Md Shihab Shahriar, Mohammad Ruhul Amin - **HateXplain Space Model: Fusing Robustness with Explainability in Hate Speech Analysis**. In *Efficient Natural Language and Speech Processing Workshop at NeurIPS* | 2023

- [W6] **Md Fahim**, Amin Ahsan Ali, Ashraful Amin, AKM Mahabubur Rahman - **Contextual Bangla Neural Stemmer: Finding Contextualized RootWord Representations for BanglaWords**. In *Bangla Language Processing(BLP) Workshop at EMNLP* | 2023
- [W5] **Md Fahim** - **Aambela at BLP-2023 Task 2: Enhancing BanglaBERT Performance for Bangla Sentiment Analysis Task with In Task Pretraining and Adversarial Weight Perturbation [Best Paper Award]**. In *Bangla Language Processing(BLP) Workshop at EMNLP* | 2023
- [W4] Dehan Noor, **Md Fahim**, Amin Ahsan Ali, Ashraful Amin, AKM Mahabubur Rahman - **Investigating the Effectiveness of Graph-based Algorithm for Bangla Text Classification**. In *Bangla Language Processing(BLP) Workshop at EMNLP* | 2023
- [W3] Nabilah Tabassum Oshin*, Syed Mohaiminul Hoque*, **Md Fahim**, Amin Ahsan Ali, Ashraful Amin, AKM Mahabubur Rahman - **BaTEClacor: A Novel Dataset for Bangla Text Error Classification and Correction** . In *Bangla Language Processing(BLP) Workshop at EMNLP* | 2023
- [W2] **Md Fahim** - **Aambela at BLP-2023 Task 1: Focus on [UNK] tokens: Analyzing Violence Inciting Bengali Text with Adding Dataset Specific New Word Tokens** . In *Bangla Language Processing(BLP) Workshop at EMNLP* | 2023
- [W1] **Md Fahim**, Md Shihab Shahriar, Mohammad Sabik Irbaz, Syed Ishtiaque Ahmed, Mohammad Ruhul Amin - **HateXplain2.0: An Explainable Hate Speech Detection Framework Utilizing Subjective Projection from Contextual Knowledge Space to Disjoint Concept Space**. In *AI & HCI Workshop at ICML* | 2023

UNDER REVIEW PAPERS

C=CONFERENCE, J=JOURNAL, W=WORKSHOP

- [C] **Md Fahim**, Meheraj Hossain, Shadman Rohan, Md Ashraful Amin, Akm Mahbubur Rahman, and Amin Ahsan Ali - **L-Context: Layer-wise Context Vectors for Better Text Classification Using Pre-trained Language Models**. [Manuscript]
- [C] Aldrin Kabya*, **Md Fahim***, AKM Mahabubur Rahman, M Ashraful Amin, and Amin Ahsan Ali - **Training-Free Layer Selection for Partial Fine-Tuning of Language Models**. [Manuscript]
- [C] **Md Fahim**, Md Shihab Shahriar, Mohammad Ruhul Amin - **Co-SpaM: Concept Space Modeling for Improving Interpretability and Performance: An Experiment on Hate Text Detection**. [Manuscript]
- [C] Deeparghya Dutta Barua*, Md Sakib Ul Rahman Sourove*, Md Farhan Ishmam*, Fabiha Haider, Fariha Tanjim Shifat, **Md Fahim**, Farhad Alam Bhuiyan - **ChitroJera: A Regionally Relevant Visual Question Answering Dataset for Bangla**. [Preprint]
- [C] Md Mubtasim Ahasan, **Md Fahim**, Tasnim Mohiuddin, AKMMAHBUBUR RAHMAN, Aman Chadha, Tariq Iqbal, M Ashraful Amin, Md Mofijul Islam, Amin Ahsan Ali - **DM-Codec: Distilling Multimodal Representations for Speech Tokenization**. [Preprint]
- [C] Md Mofijul Islam, Alexi Gladstone, Sujana Sarker, Ganesh Nanduru, **Md Fahim**, Keyan Du, Aman Chadha, Tariq Iqbal - **Embodied Referring Expression Comprehension Through Multimodal Residual Learning**
- [C] Sourav Saha, Jahedul Alam Junaed, **Md Fahim**, Maryam Saleki, Arnab Sen Sharma, Mohammad Rashidujaman Rifat, Mohamed Rahouti, Syed Ishtiaque Ahmed, Nabeel Mohammed, Mohammad Ruhul Amin - **Vio-Lens: A Novel Dataset for Detecting Violence-Inciting Texts with State-of-the-Art Evaluation**
- [J] Ayon Mia*, **Md Fahim***, Golam Mursalin, Jahid Basher, Hasan Murad, Rejaul Islam, Ibrahim Khan, Iqbal H. Sarker - **Code-Mixed Language Translation: An Integrated Framework with Enhanced Preprocessing and N-gram Optimization for Cross-Lingual Communication**. [Preprint]
- [J] Fahim Mahafuz Ruhad*, **Md Fahim***, Mir Sazzat Hossain, Md Fahad Monir, and Ashraful Islam - **Comparative Performance Analysis of YOLO Models For Tomato Leaf Disease Detection**.

MACHINE LEARNING ACHIEVEMENTS

- **Kaggle Expert (X3)** 2020 - Present
Participated in Several Competitions and Won 3 *silver* and 2 *bronze* Medals
- **Finalist in Robi Datathon 3.0** May 2024
In the top 20 teams among 3,500+ participants at Robi Datathon 3.0, Bangladesh's largest datathon
- **Best Paper Award** Dec 2023
1st Bangla Language Processing Workshop @ EMNLP
- **Winner @ Shared Task** Oct 2023
Achieved **1st Place** in Sentiment Analysis Shared Task at BLP Workshop
- **2nd Place @ Shared Task** Oct 2023
Achieved **2nd Place** in Violence Inciting Text Detection Shared Task at BLP Workshop
- **5th Place @ Datathon** March 2023
Achieved **5th Place** in Bhashabhrom Datathon: Bangla Grammatical Error Detection Challenge
- **Winner @ National NLP Competition** Nov 2022
Achieved **1st Place** in CUET ETE National NLP Competition
- **Finalist in Robi Datathon 2.0** July 2022
In the top 20 teams at Robi Datathon 2.0 among 384+ teams