Mir Tafseer Nayeem

PhD Candidate in Computing Science, University of Alberta

mnayeem@ualberta.ca Website Google Scholar GitHub LinkedIn

Education

University of Alberta **Edmonton, Canada**
Ph.D. in Computing Science Sept 2021 – *Current*
Thesis Title: *"User-Centric Modeling and Synthesis in Language Technologies"*
Supervisor: Prof. Davood Rafiei
Committee members: Prof. Osmar Zaiane, Prof. Lili Mou

University of Lethbridge **Lethbridge, Canada**
M.Sc. in Computer Science
Thesis Title: *"Methods of Sentence Extraction, Abstraction and Ordering for Automatic Text Summarization"*
Supervisor: Prof. Yllias Chali
Committee members: Prof. Wendy Osborn, Prof. John Anvik

Islamic University of Technology **Dhaka, Bangladesh**
B.Sc. in Computer Science & IT

Honors & Fellowships

- **Best Short Paper Award**, IEEE VIS 2025
- **Best Resource Paper Award**, EMNLP 2024
- **Huawei PhD Fellowship**, 2022–2027
  Nominated for PhD Early Achievement Award
  University of Alberta Graduate Recruitment Scholarship
- **Area Chair Favorite Paper Award**, COLING 2018
  ACM SIGIR Student Travel Grant
  S.G.S Research Dissemination Travel Award
  Graduate Student's Association (GSA) Travel Award
  School of Graduate Studies (S.G.S) Dean's Scholarship
  School of Graduate Studies (S.G.S) International Tuition Award
  IUT-OIC Full Free Scholarship

Publications

Preprints and Working Papers

[1] **Mir Tafseer Nayeem** and Davood Rafiei. *"Whose Data, Whose Benchmark? A User-Centric Framework for Auditing Evaluation Resources."* Preprint, 2026. Paper.

[2] **Mir Tafseer Nayeem** and Davood Rafiei. *"How Does English (US) Become the Default? Triangulating Structural Bias Towards American English Across the LLM Pipeline."* Preprint, 2026. Paper.

[3] **Mir Tafseer Nayeem** and Davood Rafiei. *"Who Gets a Token, and What Does It Carry? Unequal Name Support and Concept Access in Large Language Models."* Preprint, 2026. Paper.

[4] **Mir Tafseer Nayeem**, Susmoy Chakraborty, and Davood Rafiei. *"CineSubBench: Evaluating LLMs on Long-Form Narrative and Cultural Understanding from Multilingual Movie Subtitles."* Preprint, 2026. Paper.

[5] Ridwan Mahbub, Mohammed Saidul Islam, Md Tahmid Rahman Laskar*, Mizanur Rahman*, **Mir Tafseer Nayeem***, and Enamul Hoque. *"Chart Deception in Vision-Language Models: From Vulnerability to Mitigation."* Preprint, 2026. Paper.

Selected Peer-Reviewed Publications

* Equal contribution. See also 🎓 Google Scholar.

[1] **Mir Tafseer Nayeem***, Sawsan Alqahtani*, Md Tahmid Rahman Laskar, Tasnim Mohiuddin, and M Saiful Bari. *"Stop Taking Tokenizers for Granted: They Are Core Design Decisions in Large Language Models."* **EACL 2026** (Main, Oral). Paper.

[2] Mohammed Saidul Islam, Md Tahmid Rahman Laskar, **Mir Tafseer Nayeem**, Shumaita Binte Burhan, Raian Rahman, Morsalina, Shafiq Joty, Enamul Hoque, and Jimmy Huang. *"What Happens When the Font Changes? A Controlled Study of Typographic Robustness in Vision-Language Models."* **NeurIPS 2026** Workshop. Paper.

[3] Md Tahmid Rahman Laskar, Mohammed Saidul Islam, **Mir Tafseer Nayeem**, Amran Bhuiyan, Mizanur Rahman, Shafiq Joty, Enamul Hoque, and Jimmy Huang. *"Lost in Translation: Do LVLM Judges Generalize Across Languages?"* **ACL 2026** (Findings). Paper.

[4] Awal Ahmed Fime*, Mohammed Saidul Islam*, Md Tahmid Rahman Laskar, Md Zarif Hossain, Ridwan Mahbub, **Mir Tafseer Nayeem**, Mizanur Rahman, Ahmed Imteaj, and Enamul Hoque Prince. *"From Benchmarks to Real-World Use: Evaluating the Robustness of Vision-Language Models for Chart Understanding."* **EMNLP 2026** (Industry Track). Paper.

[5] Tong Bao, **Mir Tafseer Nayeem**, Yi Zhao, Davood Rafiei, and Chengzhi Zhang. *"SurveyAgent-HKA: A Multi-Agent Framework for Scientific Survey Generation with LLMs and Human Knowledge Augmentation."* **Knowledge-Based Systems**, 2026. Paper.

[6] Saddam Mukta, Md. Ahashan Habib, **Mir Tafseer Nayeem**, and Md Adnanul Islam. *"RL-MR: Reinforcement Learning based Budget-Aware Sequential Pretrained Model Ranking."* **AACL 2026** (Main, Oral). Paper.

[7] **Mir Tafseer Nayeem** and Davood Rafiei. *"OpinioRAG: Towards Generating User-Centric Opinion Highlights from Large-scale Online Reviews."* **COLM 2025**. Paper.

[8] **Mir Tafseer Nayeem***, Sawsan Alqahtani*, Md Tahmid Rahman Laskar, Tasnim Mohiuddin, and M Saiful Bari. *"Beyond Fertility: Analyzing STRR as a Metric for Multilingual Tokenization Evaluation."* **NeurIPS 2025** Workshop. Paper.

[9] Tong Bao, **Mir Tafseer Nayeem**, Davood Rafiei, and Chengzhi Zhang. *"SurveyGen: Quality-Aware Scientific Survey Generation with Large Language Models."* **EMNLP 2025** (Main, Oral). Paper.

[10] Ridwan Mahbub*, Mohammed Saidul Islam*, **Mir Tafseer Nayeem**, Md Tahmid Rahman Laskar, Mizanur Rahman, Shafiq Joty, and Enamul Hoque. *"From Charts to Fair Narratives: Uncovering and Mitigating Geo-Economic Biases in Chart-to-Text."* **EMNLP 2025** (Main). Paper.

[11] Md Tahmid Rahman Laskar, Mohammed Saidul Islam, Ridwan Mahbub, Mizanur Rahman, Amran Bhuiyan, Israt Jahan, **Mir Tafseer Nayeem**, Shafiq Joty, Enamul Hoque, and Jimmy Huang. *"Deploying Tiny LVLM Judges for Real-World Evaluation of Chart Models: Lessons Learned and Best Practices."* **EMNLP 2025** (Industry Track). Paper.

[12] Ridwan Mahbub, Mohammed Saidul Islam, Md Tahmid Rahman Laskar, Mizanur Rahman, **Mir Tafseer Nayeem**, and Enamul Hoque. *"The Perils of Chart Deception: How Misleading Visualizations Affect Vision-Language Models."* **IEEE VIS 2025** (Oral). Paper. 🏆 **Best Short Paper Award**.

[13] Md Tahmid Rahman Laskar, Mohammed Saidul Islam, Ridwan Mahbub, Ahmed Masry, Mizanur Rahman, Amran Bhuiyan, **Mir Tafseer Nayeem**, Shafiq Joty, Enamul Hoque, and Jimmy Huang. *"Judging the Judges: Can Large Vision-Language Models Fairly Evaluate Chart Comprehension and Reasoning?"* **ACL 2025** (Industry Track). Paper.

[14] Luis Antonio Gutiérrez Guanilo, **Mir Tafseer Nayeem**, Cristian López, and Davood Rafiei. *"eC-Tab2Text: Aspect-Based Text Generation from e-Commerce Product Tables."* **NAACL 2025** (Industry Track). Paper.

[15] **Mir Tafseer Nayeem** and Davood Rafiei. *"KidLM: Advancing Language Models for Children – Early Insights and Future Directions."* **EMNLP 2024** (Main). Paper. 🏆 **Best Resource Paper Award**.

[16] Mohammed Saidul Islam, Raian Rahman, Ahmed Masry*, Md Tahmid Rahman Laskar*, **Mir Tafseer Nayeem***, and Enamul Hoque. *"Are Large Vision Language Models up to the Challenge of Chart Comprehension and Reasoning."* **EMNLP 2024** (Findings). Paper.

[17] **Mir Tafseer Nayeem***, Faisal Tareque Shohan*, Samsul Islam, Abu Ubaida Akash, and Shafiq Joty. *"XL-HeadTags: Leveraging Multimodal Retrieval Augmentation for the Multilingual Generation of News Headlines and Tags."* **ACL 2024** (Findings). Paper.

[18] Mohsinul Kabir*, Mohammed Saidul Islam*, Md Tahmid Rahman Laskar, **Mir Tafseer Nayeem**, M Saiful Bari, and Enamul Hoque. *"BenLLMEval: A Comprehensive Evaluation into the Potentials and Pitfalls of Large Language Models on Bengali NLP."* **COLING 2024**. Paper.

[19] Ali Naeim abadi, **Mir Tafseer Nayeem**, and Davood Rafiei. *"Product Entity Matching via Tabular Data."* **CIKM 2023**. Paper.

[20] **Mir Tafseer Nayeem** and Davood Rafiei. *"On the Role of Reviewer Expertise in Temporal Review Helpfulness Prediction."* **EACL 2023** (Findings). Paper.

[21] **Mir Tafseer Nayeem***, Abu Ubaida Akash*, Faisal Tareque Shohan, and Tanvir Islam. *"Shironaam: Bengali News Headline Generation using Auxiliary Information."* **EACL 2023** (Main). Paper.

[22] **Mir Tafseer Nayeem***, Susmoy Chakraborty*, and Wasi Ahmad. *"Simple or Complex? Learning to Predict Readability of Bengali Texts."* **AAAI 2021** (Main). Paper.

[23] **Mir Tafseer Nayeem***, Radia Rayan Chowdhury*, Tahsin Tasnim Mim, Md. Saifur Rahman Chowdhury, and Taufiqul Jannat. *"Unsupervised Abstractive Summarization of Bengali Text Documents."* **EACL 2021** (Main). Paper.

[24] **Mir Tafseer Nayeem**, Tanvir Ahmed Fuad, and Yllias Chali. *"Neural Diverse Abstractive Sentence Compression Generation."* **ECIR 2019**. Paper.

[25] **Mir Tafseer Nayeem***, Tanvir Ahmed Fuad*, Asif Mahmud, and Yllias Chali. *"Neural Sentence Fusion for Diversity Driven Abstractive Multi-Document Summarization."* **Computer Speech & Language**, 58:216–230, 2019. Paper.

[26] **Mir Tafseer Nayeem**, Tanvir Ahmed Fuad, and Yllias Chali. *"Abstractive Unsupervised Multi-Document Summarization using Paraphrastic Sentence Fusion."* **COLING 2018** (Oral). Paper. 🏆 **Area Chair Favorite Paper Award**.

[27] **Mir Tafseer Nayeem** and Yllias Chali. *"Paraphrastic Fusion for Abstractive Multi-Sentence Compression Generation."* **CIKM 2017**. Paper.

[28] Yllias Chali, Moin Mahmud Tanvee, and **Mir Tafseer Nayeem**. *"Towards Abstractive Multi-Document Summarization Using Submodular Function-Based Framework, Sentence Compression and Merging."* **IJCNLP 2017**. Paper.

[29] **Mir Tafseer Nayeem** and Yllias Chali. *"Extract with Order for Coherent Multi-Document Summarization."* **ACL 2017** Workshop (Oral). Paper.

Mentoring

Research mentoring across institutions and countries through collaborative projects and publications.

Tong Bao, PhD student, Nanjing University of Science and Technology, China
Joint publications: *SurveyGen*, **EMNLP 2025**; *SurveyAgent-HKA*, **Knowledge-Based Systems** (2026).

Luis Antonio Gutiérrez Guanilo, undergraduate student, University of Engineering and Technology (UTEC), Peru
Joint publication: *eC-Tab2Text*, **NAACL 2025** (Industry Track).

Ridwan Mahbub, MSc student, York University, Canada
Joint publication: *From Charts to Fair Narratives*, **EMNLP 2025** (Main).

Ali Naeimabadi, Senior AI Engineer, RBC Borealis, Canada
Joint publication: *Product Entity Matching via Tabular Data*, **CIKM 2023**.

Faisal Tareque Shohan, MSc student, Université de Sherbrooke, Canada
Joint publications: *Shironaam*, **EACL 2023** (Main); *XL-HeadTags*, **ACL 2024** (Findings).

Susmoy Chakraborty, Senior Software Engineer, ICT Center, AUST, Bangladesh
Joint publication: *Simple or Complex? Learning to Predict Readability of Bengali Texts*, **AAAI 2021**.

Radia Rayan Chowdhury, Data Scientist, Teesside University, UK
Joint publication: *Unsupervised Bengali Summarization*, **EACL 2021** (Main).

Abanti Chakraborty Shruti, PhD student, UNC Charlotte, USA
Joint project: *Collaborative Filtering by Finding the Experts*.

Professional Activities

Academic Service

Area Chair

- **2027:** NAACL, ARR
- **2026:** ACL, EMNLP, EACL, ARR
- **2025:** ACL, EMNLP, ARR

Reviewer

- **2027:** ICLR
- **2026:** NeurIPS, COLM
- **2025:** NAACL
- **2024:** ACL, EMNLP, NAACL, EACL, SIGIR, WSDM
- **2023:** SIGIR, KDD, WSDM, CIKM
- **2022:** SIGIR, KDD, CIKM, ICDE

Outstanding Reviews

Reviews recognized as **"Great Reviews"** by ACL-family venues

- **1/3 reviews** recognized at NAACL 2025
- **2/4 reviews** recognized at ACL 2024

Invited Talks

- Talk on KidLM at Honors Seminar, University of Alberta (UofA), AB, Canada. [2024]
- Talk on Introduction to NLTK at University of Lethbridge (UofL), AB, Canada. [2017]
- Guest Lecturer in CPSC 3750 - Artificial Intelligence course, UofL, AB, Canada. [2017]

Leadership

- Lab Manager, Data & Language Intelligence (DaLI) Lab, University of Alberta. [2022 – *Current*]
- Founding admin of a Discord community with 5,000+ members supporting graduate-school applicants. [2021 – *Current*]
- Member, Logistics & IT Support Committee, ICCIT 2020 Conference.
- Organizing committee, Commonwealth of Learning (COL) TEL AUST Project [report].
- Organizer and Host, Intel Nervana AI Academy - Workshop.
- Convener, Project Showcasing, AUST CSE WEEK.
- Organizing committee member, AUST CSE Fest; BUBT CSE Fiesta; IUT ICT Fest.

Teaching & Professional Experience

Graduate Teaching Assistant **Sept 2021 – *Current***
University of Alberta, Computing Science Department Edmonton, Canada

Teaching assistant for:

- CMPUT 101 - Introduction to Computing
- CMPUT 291 - Introduction to File and Database Management

Faculty Member, CSE Department **Oct 2018 – Jul 2021**
Ahsanullah University of Science and Technology (AUST) Dhaka, Bangladesh

Courses Taught:

- CSE 4237 & 4238 - Soft Computing with Lab
- CSE 4129 & CSE 4130 - Formal Languages & Compilers with Lab
- CSE 3223 & 3224 - Information System Design and Software Engineering with Lab
- CSE 3104 - Database Lab
- CSE 1108 - Introduction to Computer Systems

Graduate Teaching Assistant **Sept 2015 – Sept 2017**
University of Lethbridge, Computer Science Department Lethbridge, Canada

Teaching assistant for:

- CPSC 1620 & 2620 - Fundamentals of Programming I & II
- CPSC 2720 - Practical Software Development