Nafis Irtiza Tripto

Contact Information

Room No: 413, Eric J. Barron Innovation Hub, University Park, PA-16802
E-mail: nit5154@psu.edu , **Phone:** (+1)814-769-0720
Personal Website, **LinkedIn**, **Google Scholar**

Education

Ph.D. in Informatics Fall, 2022 - Present (expected graduation May, 2026)
College of Information Science and Technology (IST), Pennsylvania State University (PSU)
Track: Data Science & Artificial Intelligence, Privacy & Security
GPA: 4.00/4.00 (Now)
Advisor: Dongwon Lee (dongwon@psu.edu)

M.Sc. in Computer Science and Engineering, 2017 - 2021
Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh.
GPA: 4.00/4.00
Advisor: Mohammed Eunus Ali (eunus@cse.buet.ac.bd)

B.Sc. in Computer Science and Engineering, 2013 - 2017
Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh.
CGPA: 3.97/4.00 (Ranked 2nd in a class of 150 students)

Technical Skills

- **Programming Language:** Java, Python, C, C++, Assembly, PHP
- **Scripting Language & Database:** SQL, MySQL, Oracle, SQLite, MATLAB, LaTeX
- **Deep Learning Frameworks:** PyTorch, Keras, Tensorflow, LLM finetuning, Prompt Engineering
- **Other Frameworks/Tools:** ZAProxy, MOBSF, Androbugs, Dialogflow. VertexAI

Fellowship and Awards (Selected)

- **NSF LinDiv Fellowship** (Linguistic Diversity Across the Lifespan) Fall 2024-Spring 2026
- **University Graduate Fellowship** , Pennsylvania State University, Fall 2022 – Spring 2023
- **Regional Winner (Asia)**, *Global Undergraduate Awards*, Computer Sciences 2018

Relevant Experiences (selected)

***Data science intern*, *Contact Center Data Science team*, Home Depot** Summer, 2024

- **Predictive models for customer call intent:** I Utilized in-store purchase information, website browsing behavior, and historical call records to create comprehensive customer journey profiles and later utilized ML algorithms and LLM to predict potential reasons for customer calls. The process involved extensive data cleaning, preprocessing, and pattern recognition within 5 billion rows from various sources, alongside advanced LLM prompting techniques.
- **Enhanced Customer Query Resolution with Retrieval-Augmented Generation (RAG):** I contributed to the project utilizing RAG combined with LLMs to address customer queries by leveraging knowledge base documents and evaluated the effectiveness of the solution using various automated metrics.
- **Tools:** Python with Pandas, Google Big Query, Adobe Clickstreams, Transformers, LLM APIs, LLM prompt engineering, Dialogflow

***Research Assistant*, PIKE Lab, Pennsylvania State University** Fall,2022-Present

- Conducted advanced Natural Language Processing (NLP) research with a focus on Large Language Models (LLM), including ChatGPT, to explore their generative capabilities across diverse spoken text contexts, encompassing speech, interviews, and conversational interactions (**ACCEPTED in EMNLP'23** *Findings*).
- Conducting research to investigate the influence of paraphrasing on authorship style transformation within the domain of LLMs. This research holds significant relevance in the contemporary landscape of text generation, particularly in the context of collaborative human-machine interactions. (**ACCEPTED in ACL'24 Main**)
- Engaged in collaborative multi-author authorship analysis. (*manuscript under review*)
- Contributed to the project about how multi-lingual AI text detectors perform in case of authorship obfuscation. (*manuscript under review*)
- Leading the project for finding which part of the text (introduction /body/ conclusion) differs most between human and AI text, motivated by the famous analogy in the Chess game (manuscript in preperation)

Lecturer at Bangladesh University of Engineering and Technology (BUET) (2018-22)

- Instructed relevant courses while working as a lecturer, such as CSE 301: Mathematical Analysis, CSE 405: Computer Security, and Introduction to Machine Learning with Python.

***Research Assistant*, Data Lab, BUET** (2017-2021)

- Participated in the development and deployment of an AI-based medical support Chatbot focusing on adolescent sexual and reproductive health, including feasibility studies and prototype creation, subsequently deployed as an FB messenger Plugin. (*published in* **CHI'21**)
- Conducted research on anomaly detection and forecasting in time series gene expression data, utilizing deep learning+statistical techniques. (published in **PLOS One Journal**, 2021)
- Developing a Bengali currency recognizer for visually impaired people. (published **ICTD'19**). **Tools:** Android for app development, MobileNet for image classification.
- Conducted Bengali literature analysis using innovative word2vec graph methodology to address author attribution, genre detection, and stylochronometry challenges. Additionally, employed character interaction graphs to explore societal influence on literary fiction. *(under review)*
- Sentiment analysis from Bengali noisy online data (YouTube comments) (published in, ICBSLP'18).

***Intern*, Dingi Technologies Ltd** (2017)

- Participated in a research project focused on developing an automated system for travel time estimation utilizing historical GPS data from OpenStreetMap. **Tools:** Python with GeoPandas, Fiona, Shapely.

Publications

Journal Publications:

1. **Nafis Irtiza Tripto**, Mohimenul Kabir, Md. Shamsuzzoha Bayzid, Atif Rahman, Evaluation of classification and forecasting methods on time series gene expression data. Plos one, 15(11), e0241686.
2. **Nafis Irtiza Tripto**, Mahjabin Nahar, Mohammed Eunus Ali, Farhana Murtaza Choudhury, J. Shane Culpepper, Timos Sellis, Top-k trajectories with the best view. GeoInformatica, 23, 621-661.

Conference Publications:

1. **Nafis Irtiza Tripto**, Saranya Venkatraman, Dominik Macko, Robert Moro, Ivan Srba, Adaku Uchendu, Thai Le, Dongwon Lee "A Ship of Theseus: Curious Cases of Paraphrasing in LLM-Generated Texts", In Proceedings of the 2024 Association of Computational Linguistics (ACL'24)
2. **Nafis Irtiza Tripto**, Adaku Uchendu, Thai Le, Mattia Setzu, Fosca Giannotti, Dongwon Lee. HANSEN: Human and AI Spoken Text Benchmark for Authorship Analysis. In Proceedings of the 2023 Findings of Empirical Methods in Natural Language Processing
3. Rifat Rahman, Md. Rishadur Rahman, **Nafis Irtiza Tripto**, Mohammed Eunus Ali, Sajid Hasan Apon, and Rifat Shahriyar. AdolescentBot: Understanding opportunities for chatbots in combating adolescent sexual and reproductive health problems in Bangladesh. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-15).
4. Hasan Murad, **Nafis Irtiza Tripto**, Mohammed Eunus Ali, Developing a bangla currency recognizer for visually impaired people. In Proceedings of the Tenth International Conference on Information and Communication Technologies and Development (pp. 1-5).
5. **Nafis Irtiza Tripto**, Mohammed Eunus Ali (2018, September). Detecting multilabel sentiment and emotions from bangla youtube comments. In 2018 International Conference on Bangla Speech and Language Processing (ICBSLP) (pp. 1-6). IEEE.

Other Manuscripts:

1. Macko, Dominik, Robert Moro, Adaku Uchendu, Ivan Srba, Jason Samuel Lucas, Michiharu Yamashita, **Nafis Irtiza Tripto**, Dongwon Lee, Jakub Simko, and Maria Bielikova. "Authorship obfuscation in multilingual machine-generated text detection." arXiv preprint arXiv:2401.07867 (2024).
2. Venkatraman, Saranya, **Nafis Irtiza Tripto**, and Dongwon Lee. "CollabStory: Multi-LLM Collaborative Story Generation and Authorship Analysis." arXiv preprint arXiv:2406.12665 (2024).
3. **Nafis Irtiza Tripto** and Mohammed Eunus Ali "Writing Style Analysis in Bengali Literature using W2V Graphs", 2022 (Arxived).
4. **Nafis Irtiza Tripto** and Mohammed Eunus Ali "Understanding Social Structures from Contemporary Literary Fiction using Character Interaction Graph - Half Century Chronology of Influential Bengali Writers", 2022 (Arxived).

References

Dr. Dongwon Lee, Professor, College of Information Science and Technology,
Pennsylvania State University
E-mail: dongwon@psu.edu, **Profile:** https://pike.psu.edu/