MD. ZAHIDUL HAQUE

1330 S Mount Vernon Ave., Unit G
Williamsburg, VA, 23185

Email : mhaque@wm.edu
Mobile : +1 (757) 696-4223
Linkedin : md-zahidul-haque-6298781a7

EDUCATION

- **William & Mary** Williamsburg, VA, USA
  PhD in Computer Science; GPA : 3.92 *August 2023 - Present*
  - **Key Courses**: Software Mntnce/Evolution, Deep Learning, Auton Driving & Con Mobility, Deep Transfer Learning, Data Mining & Decision Making, Human Centered Computin, AI for S/W Eng, Systems for Neural Networks.
- **Sylhet Engineering College (SEC)** Sylhet, Bangladesh
  Bachelor of Science in Computer Science & Engineering; GPA : 3.46 *March 2016 - March 2021*
  - **Key Courses**: Structured Programming, Data Structures & Algorithms, OOP, Database, Operating System, Machine Learning.

SKILLS SUMMARY

- **Languages**: Python, C++, C, SQL, CUDA
- **Framework**: Tensorflow, Keras, Pytorch, Pandas, Scikit-learn
- **Tools**: Slack, Microsoft Teams & Office365, Notion, Airtable

PROFESSIONAL EXPERIENCE

- **Graaho** Reston, VA
  Software Engineer (Machine Learning) *March 2022 - September 2022*
  - **Qesfera**: Developed a financial machine learning application on credit risk by cleaning the dataset, testing with pytest, and baseline modeling.
- **Mayalogy** Dhaka, Bangladesh
  Research Engineer (Machine Learning) *February 2021 - February 2022*
  - **MayaNERC**: Developed a System named "MayaNERC: A Pipeline to Biomedical NER from Patient Label Text" which can identify 17 entities from patent label text and can validate 111 diseases by only checking those entities out of 272 categories (covering five major categories Men's health, Menstruation, Pregnancy, Pain, & Others).
  - **Others:**: Developed Auto Tag, Group Referrer, and Pakistani Spam Filter (Maya Pakistan) where the dataset was Romanized Bangla and Pakistani Text. Macro F1 score was respectively 85%, 89%, and 88%.
  - **Overall**: Leading an AI annotation team of 20 engineers and 15 psychologists for developing deep learning datasets. Closely working with a medical team of 150 doctors to provide better AI assisted healthcare.

PUBLICATIONS

- **TOSEM**
  *Saima Afrin[*], **Md. Zahidul Haque**[*], Antonio Mastropaolo* *Oct 2025*
  "A Systematic Literature Review of Parameter-Efficient Fine-Tuning for Large Code Models"

  - [*]Equal contribution

- **IEEE Access**
  Md. Zahidul Haque, *Sakib Zaman, Jillur Rahman Saurav, Summit Haque, Md. Saiful Islam, Ruhul Amin* *April 2023*
  "B-NER: A Novel Bangla Named Entity Recognition Dataset with Largest Entities and Its Baseline Evaluation"

- **ICCA '22: Proceedings of the 2nd International Conference on Computing Advancements.**
  Mobasshir Bhuiya Shagor, **Md. Zahidul Haque Alvi***, Khandaker Tabin Hasan* *March 2022*
  "CID: Cow Images Dataset for Regression and Classification"

PROJECTS

MayaNERC: A Pipeline to Biomedical NER from Patient Label Text
Advisor: Mohammad Ruhul Amin *April 2021 - February 2022*

- Applied entity recognition to a different motivation in biomedical patient-level text. Giving a generative solution to a patient who writes about his or her health condition and submits it to a doctor does not require all of the tokens in the sentence, and these words have been studied in over 272 diseases covering five classes. Validation was done on 111 types of diseases that can be verified by only passing a list of recognition where we finalized 17 types of entity.

- To establish a baseline for future comparisons, we present the dataset with numerous strong state-of-the-art benchmark analyses.

ON GOING PROJECTS

Currently monitoring two undergraduate-level quality theses in Natural Language Processing (NLP) and working as a Researcher at a non-profit startup called Bengali.AI Bengali.AI since December 2022.

End-To-End Entity Linking on Bangla Newspapers and Wikipedia Data   *July 2022 - Present*

- First Bangla Entity Linking (EL) dataset based on Bangla newspapers and Wikipedia data (Following AIDAConLL EL dataset)

- A potential Candidate list using Bangla Wikipedia data.

- Employed an End-To-End Entity Linking technique to examine the performance of our Bangla EL dataset and Candidate list.

BL-NERC: An Introduction to Bengali Literature Named Entity Recognition Corpus   *August 2022 - Present*

- First named entity recognition corpus based on Bengali literature from 16 authors (data is collected from BAAD16 Dataset).

- Manually annoted 18K data out of 30K on four entities (person, location, organization, & time).

Bengali Common Voice Speech Dataset for Automatic Speech Recognition   *February 2022 - Present*

- Launched Kaggle featured competition from Bangladesh on 475 hours of speech data. Raised 5K USD in funds for the competition. The details of the dataset can be found here: ARXIV.

- The Mozilla Common Voice platform collects 2000 hours of speech from more than 22K+ contributors.

PARTICIPATION'S

- **DL Sprint - BUET CSE Fest 2022**
  Bengali Automatic Speech Recognition Competition   *July 2022 - September 2022*

- **Robi Datathon - 2.0**
  Pre-Assessment on tebular data   *June 2022*

TEACHING EXPERIENCE

- **William & Mary**   Williamsburg, VA, USA
  Teaching Assistant - Department of Computer Science   *Fall 2023 - Present*

  - **Fall 2025**: CSCI 100 (Prompt Engineering) - Grading — Prof. Dr. Antonio Mastropaolo

  - **Spring 2025**: CSCI 312 (Principles of Programming Languages) - Grading — Prof. Dr. Martin White

  - **Fall 2024**: CSCI 141-03 (Introduction to Programming) - Grading and Lab Instruction (Labs 3 & 13) — Prof. Dr. Matthew Chapman

  - **Spring 2024**: CSCI 312 (Principles of Programming Languages) - Grading — Prof. Dr. Robert Michael Lewis

  - **Fall 2023**: CSCI 312 (Principles of Programming Languages) - Grading — Prof. Dr. Tim Davis

CERTIFICATIONS

- **Cisco Certified Network Associate**
  "Cisco Certifications"   *September 2020 - September 2023*

EXTRA CURRICULARS

- Hobbyist traveller, love to track.