Zafaryab Haider

Ph.D. Candidate | ECE, UMaine, Orono,ME | zafaryab.haider@maine.edu | Zafaryab

Interests

Large Language Models (LLMs), Safe and Trustworthy Artificial Intelligence

Education

University of Maine (UMaine), Orono, ME. USA — August 2022 – Present
Ph.D in Electrical and Computer Engineering — (Pursuing)

Aligarh Muslim University (AMU), Aligarh, UP. INDIA — August 2011 – July 2013
M.Tech. in Computer Engineering — (GPA 9.33/10.0 (Hons.))

UP Technical University (now AKTU), Noida, UP. INDIA — August 2007 – July 2011
B.Tech. in Computer Science and Engineering — (Percentage 76.2% (Hons.))

Research Papers

Published

- Rahman, MH., **Haider, Z.**, Rizvee, M. M. , Shomaji, S. and Chakraborty, P., *Intelligent Layer Sharing (ILASH): A Predictive Neural Architecture Search Framework for Multi-Task Applications*, in *IEEE Access*, vol. 13, pp. 132767-132781, 2025, doi: 10.1109/ACCESS.2025.3592039.
- Rahman, M H., **Haider, Z.** and Chakraborty, P. *An Automated Multi-Parameter Neural Architecture Discovery Framework using ChatGPT in the Backend*, *Nature Scientific Reports, 2025.*
- **Haider, Z.**, Rahman, H., Devabhaktuni, V., Moeykens, S. and Chakraborty, P.. *CoBRa: Consensus Based Reward System for Mitigating Malicious Human-in-the-Loop Feedback in LLM*, *Nature Scientific Reports, 2025.*
- Aqeel, S., **Haider, Z.**, & Khan, W. (2023). *Towards digital diagnosis of malaria: How far have we reached?*, *Journal of Microbiological Methods*, *204*, 106630.
- Khan, D. A., Zamir, M. A., Umar, M. S., & **Haider, Z.** (2022, November). *Assistive Stick for Visually Impaired People.* In *2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)* (pp. 1-6). IEEE.

Submitted

- **Haider, Z.**, Rahaman, Md H., Moeykens, S., Devabhaktuni, V., & Chakraborty, P., *How a Bit Becomes a Story: Semantic Steering via Differentiable Fault Injection*, arXiv preprint: 2512.14715. CVPR 2026.
- Nafi, A. A. N., Rahaman, H., **Haider, Z.**, Mahfuz, T., Suya, F., Bhunia, S., & Chakraborty, P., *DAASH: A Meta-Attack Framework for Synthesizing Effective and Stealthy Adversarial Examples*, 2025. arXiv preprint: 2508.13309. CVPR 2026.

Abstract Accepted

- "North Atlantic Right Whale Calls Monitoring: Signal Processing Meets Deep Learning" accepted for lecture presentation in 184^{th} Meeting of the Acoustical Society of America in Chicago, Illinois, on May 9, 2023, in session 2aSP, Signal Processing Potpourri.

Patents

Filed

- *BIT-FLIP INDUCED SEMANTIC DRIFT IN MULTIMODAL TRANSFORMER MODELS*, U.S. Provisional Patent Application No.63/930,131, **Haider Z.**, Rahman MH, Moeykens S, Chakraborty P.
- *FRAMEWORK FOR MITIGATING MALICIOUS RLHF FEEDBACK IN LLM TRAINING USING CONSENSUS BASED REWARD*, U.S. Provisional Patent Application No. 63/768,361, **Haider Z.**, Rahman MH., Moeykens S., Chakraborty P.
- *METHODS AND SYSTEMS FOR THE AUTOMATED GENERATION OF NEURAL NETWORK ARCHITECTURES*, U.S. Provisional Patent Application No. 63/627,621, Chakraborty P., Rahman MH., **Haider Z.**

Scholarships/ Fellowships

Graduate Aptitude Test in Engineering (GATE) Scholarship
Maulana Azad National Fellowship (MANF)
Teaching and Research Assistantship by TEQIP-II
Graduate Assistantship @ UMaine

Awards

SentriAlign: Training-Signal Integrity for LLMs, *(CO-PI)*, Duration: January 14, 2026 – May 11th, 2026, Funding Agency: MIRTA 8.0 Commercialization Accelerator, Amount: $25,000

Invited Lecture / Research Talks / Poster Presentations/ Demos

BLADE: Bit-level FauLt Analysis via Differentiable Estimation, Poster Presentation at 3rd Nelms Annual IoT Conference, UF, Gainesville, FL. December 2025

Enhancing RLHF Robustness with Multi-Agent Consensus Rewarding Demo at 3rd Nelms Annual IoT Conference, UF, Gainesville, FL. December 2025

A framework for mitigating malicious RLHF feedback in LLM training using consensus based reward at UMaine, Orono, USA. April 2025

AI and its use cases at Chanakya National Law University, Bihar, India. April 2025

COBRA: Mitigating Malicious RLHF Feedback in LLM Training With Consensus-Based Reward at WHiSPerS: Wearable Health monitoring Sensors for Privacy and Security, UNH, USA. March 2025

AI Security at UMaine, Orono, USA. March, 2024

LLMs and Security at Chanakya National Law University, Bihar, India. 2024

Invited as Trainer to train the lecturers of Sayid Hamid Senior Secondary School on Google Tech.
Invited by UGC-HRDC to speak on MOOCs and their way forward.
Delivered educational lectures for SWAYAM course.[1, 2]

Employments

University of Maine August 2022 – Present
Orono, ME. USA.

- **Graduate Assistant**, *ARCSIM*: Artificial intelligence (AI) and Machine Learning consultant.
- **Research Assistant**, *ECE*: Cybersecurity Education, Security of AI.
- **Teaching Assistant**, *ECE*: Sequential Logic Lab (ECE275) and Microprocessor Lab (ECE473).

Aligarh Muslim University January 2015 – August 2022
***Assistant Professor**, Zakir Husain College of Engineering & Technology, Aligarh, UP. INDIA.*

Aligarh Muslim University August 2013 – February 2014
***Guest Faculty**, Zakir Husain College of Engineering & Technology, Aligarh, UP. INDIA.*

Aligarh Muslim University August 2011– 2013
Zakir Husain College of Engineering and Technology, Aligarh, UP. INDIA.

- **Teaching Assistant**, *CE*: Data Structure Lab (CO291), Minor Project Lab (CO395) and Software Lab-II (CO494).

Teaching Experiences

[COC4010]	Information Security	2020 - 2021
[COA3600]	Data Structures	2020 - 2022
[COE0431]	Internet Tools	2015 - 2019
[COA1910]	C Programming	2015 - 2022
[COA446N]	Selected Topics in Computer Engineering-I	2015 - 2019
[COA447N]	Selected Topics in Computer Engineering-II	2013
[COC0203]	Object Oriented Programming	2013

Projects

Game graphics for Custom Operating System 2023
Through this project, I gained foundational knowledge of custom operating systems and explored the complexities of managing system calls between kernel and user spaces. Building on this understanding, I developed a Pong game that runs in user space and interacts with a PS2 keyboard through kernel-level integration for gameplay.

Detecting and Classifying Northern Right Whale calls via Deep Learning 2022
Monitoring endangered marine mammals like North Atlantic Right Whales through acoustic signals is vital but challenged by underwater signal distortion and imbalanced call types. Traditional methods struggled with classification, but spectrogram preprocessing and a self-supervised deep learning model improved accuracy by addressing data imbalance.

BEIT v2: Masked Image Modeling with Vector-Quantized Visual Tokenizers for Breast Cancer Images 2022
Addressed the challenge of scarce annotated data with self-supervised pretraining methods, drawing from BERT's success in NLP to propose BEIT, and later BEITv2. Aimed to test BEITv2 on a Breast Cancer Image dataset, findings revealed the ImageNet pretrained model excelled in accuracy (0.968) and convergence, while the model pretrained on cancer images showed faster convergence than a randomly initialized model, achieving 0.948 accuracy after 30 epochs compared to 0.812.

Data Science project on Healthcare **(Collaborating with M Lab, AMU)**
Conducted a literature review to identify gaps in healthcare CDSS implementation, shifting focus from analyzing hematological parameters to detecting infected cells in blood smear reports automatically. (On-Halt)

Extension to Lex and YACC **Guide:** Mr. M. R. Warsi, 2012
In this project, I tried to provide a PASCAL-like interface to write Lex and YACC codes, thereby providing them flexibility and better readability.

Mentoring Experience

Human Activity Recognition Using Smartphone Sensors 2019-20
Primary physical activities are identified through sensors available in smartphones.The activities identified are walking, lying down, sitting, standing, ascending, and descending stairs.

Disease Risk Prediction 2018-19
To develop a prototype Intelligent Heart Disease Prediction System (IHDPS) using three data mining modeling techniques, namely, Decision Trees, Naïve Bayes, and Neural Networks.

2π Surveillance System 2017-18
To set-up a security system with minimum hardware that can detect motion in the premises and immediately click pictures in such event to be used for inspection later by the owner in case of suspicion, and can detect fire and immediately alarm the owner for immediate action.

Hajj Guide via Offline GPS 2016-17
The idea is to provide all relevant information regarding Hajj as reminders based on the location of the person.

SyncWorld 2012-13
SyncWorld is a file storage and synchronization Service. It allows users to create a special folder on their computers, which then synchronizes with the server and then can be accessed from anywhere regardless of which computer they are using.

Technical Skills

Programming and Scripting Languages: C, C++, Java, Python, JavaScript, SQL, HTML, CSS, Shell Scripting
Parsing & Compiler Tools: Lex, YACC
Operating Systems: Windows, Linux
Tools/IDEs: LaTeX, NetBeans, Eclipse, VSCode, Google Colab, Kaggle, Jupyter Notebook, Cursor
Database Tools: Oracle, MySQL, SQLite
Frameworks & Libraries: ReactJS, scikit-learn, PyTorch, Pandas, NumPy, Matplotlib
Embedded Systems & Hardware: Raspberry Pi, Verilog, VCU118 FPGA, Vivado, Vitis, Quartus

Professional Bodies

Member of IEEE (100018946)

References

Dr. Prabuddha Chakraborty, Assistant Professor, ECE, UMaine.
✉ prabuddha@maine.edu

Dr. Shane Moeykens, Director, UMaine ARCSIM and EPSCoR.
✉ shane.moeykens@maine.edu

Dr. Vijay Devabhaktuni, Professor and Founding Chair, ECE, ILSTU.
✉ vdevabh@ilstu.edu

Date: December 18, 2025
Place: Orono, ME, USA ZAFARYAB HAIDER