Shagoto Rahman

shagotor@uci.edu
(949) 233-4120

📄 Resume | 🎓 Google Scholar | 🌐 Website | 🔒 GitHub | 🔗 LinkedIn

Department of Computer Science,   Irvine, CA.
University of California, Irvine.
Irvine, California.

EDUCATION

2023 – Present	University of California, Irvine	Irvine, California.
	PhD	CGPA: **3.96/4.00**
	Department of Computer Science	

2023 – 2025	University of California, Irvine	Irvine, California.
	MS	CGPA: **3.96/4.00**
	Department of Computer Science	

2017 - 2022	Khulna University	Khulna, Bangladesh.
	B.Sc.	
	Computer Science and Engineering Discipline	CGPA: **3.72/4.00**

RESEARCH INTERESTS

- LLM Guardrail
- LLM Fine-tuning
- ML based Privacy and Security
- Generative AI

- Retrieval-Augmented Generation
- Hallucination Mitigation
- Responsible and Efficient AI
- Context / Bias Analysis in NLP

PUBLICATIONS

Shagoto Rahman and Ian Harris. 2025. *"Summary the Savior: Harmful Keyword and Query-based Summarization for LLM Jailbreak Defense"*. In Proceedings of the 5th Workshop on Trustworthy NLP (TrustNLP 2025), pages 266–275, Albuquerque, New Mexico. Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.trustnlp-main.17

D. Lee, S. Xie, **Shagoto Rahman** et al. (2023*),"Prompter says: A linguistic approach to understanding and detecting jailbreak attacks against large-language models"*, Proceedings of the 1st ACM Workshop on Large AI Systems and Models with Privacy and Safety Analysis, pp. 77–87. doi:10.1145/3689217.3690618.

Shagoto Rahman, M. Raihan, S. K. Mithila, "Smartphone-based Skin Cancer Detection using Image Processing and Convolutional Neural Network", *12th INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT), Kharagpur, India, 2021*. Available: 10.1109/icccnt51525.2021.9579966

Shagoto Rahman, S. Mithila, A. Akther and K. Alam, "An Empirical Study of Machine Learning-based Bangla News Classification Methods", 12th *INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT), Kharagpur, India, 2021*. Available: 10.1109/icccnt51525.2021.9579655

Shagoto Rahman, K.H. Talukder, S. K. Mithila, "An Empirical Study to Detect Cyberbullying with TF-IDF and Machine Learning Algorithms", *International Conference on Electronics, Communications, and Information Technology (ICECIT), Khulna, Bangladesh, 2021.* Available: 10.1109/ICECIT54077.2021.9641251

Shagoto Rahman et al., "Efficient Machine Learning Approaches to Detect Fake News of Covid-19", *Machine Intelligence and Data Science Applications*, pp. 513-525, 2022. Available: https://doi.org/10.1007/978-981-19-2347-0_40

M. Raihan, A. K. Bairagi, **Shagoto Rahman,** "A Machine Learning Based Study to Predict Depression with Monitoring Actigraph Watch Data", *12th INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT), Kharagpur, India, 2021.* Available: 10.1109/ICCCNT51525.2021.957961

K. Mumenin et al., "COVID-19 Fake News Detection on Social Media", *International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, 2021. Available: 10.1109/ic4me253898.2021.9768523

Manuscripts Under Review

Shagoto Rahman, Cornelia Pechmann, Ian G Harris (2025). *"Enhancing Intent Detection in Nicotine Replacement Therapy Tweets Using LLM: A Fine-Tuning Framework with Class Downsampling and Misclassification Refinement."* Submitted to Journal of Medical Internet Research (JMIR).

RESEARCH EXPERIENCES

07/2025 - 09/2025   **Graduate Research Student (Summer 2025)**
Paul Merage School of Business, University of California, Irvine
Research Advisor: Dr. Cornelia Pechmann, Dr. Ian G. Harris
Current Contributions:
- Awarded **"Chao Family Comprehensive Cancer Center (CFCCC) Pilot Awards"**, by the Anti-Cancer Challenge – 2024 and authored "Enhancing Intent Detection in Nicotine Replacement Therapy Tweets Using LLM: A Fine-Tuning Framework with Class Downsampling and Misclassification Refinement" with 28% overall improvement.
- Finetuned Llama 3 8B model with LoRA and QLoRA.
- Experimented with data augmentation and novel model losses, LLM confidence and uncertainty, and out-of-distribution.
- Experimented with rules based explicit knowledge injection in LLM, with LLM interpretability.
- Developed optimized Retrieval-Augmented Generation (RAG) methods with parameter-efficient fine-tuning, enhancing accuracy and relevance of an LLM-based chatbot for nicotine replacement therapy.

- Designed and deployed an optimized LLM-based chatbot for a smoking cessation GroupMe community, achieving high helpfulness ratings and integrating log analysis to drive user engagement.

03/2024 - Present   **Graduate Research Student**
Department of Computer Science, University of California, Irvine.
Research Advisor: Dr. Ian G. Harris
Current Contributions:
- Authored "Summary the Savior: Harmful Keyword and Query-based Summarization for LLM Jailbreak Defense", achieving 100% defense success against state-of-the-art jailbreak attacks on leading LLMs.
- Authored "Prompter Says: A Natural Language Processing Approach to Jailbreak Detection in Large Language Models" and conducted semantic and linguistic analysis to detect jailbreak attempts.
- Conducted research on embedding-based Retrieval-Augmented Generation (RAG) systems for response retrieval; implemented cross-encoder reranking and leveraged entailment techniques to mitigate hallucinations in LLM outputs.
- Performed parameter-efficient fine-tuning analyses using methods such as LoRa, QLoRa, and gradient accumulation in commodity GPUs to optimize computational efficiency.
- Designed and implemented dynamic channel slimming and early exit mechanisms leveraging contrastive learning to enhance model performance and efficiency.

09/2023 - 03/2024   **Graduate Research Student**
Department of Computer Science, University of California, Irvine.
Research Advisor: Dr. Chen Li.
Contributions:
- Analyzed Java User Defined Function (UDF) in Texera,
- Analyzed Runtime Compilation of Java Code inside Texera,
- Developed Boxplot operator in Texera,
- Developed CSV file scan over various OS system in Texera,
- Developed Parts-of-Speech Java operator in Texera,
- Analyzed KNIME and RapidMiner workflow models.

05/2021 - 11/2021   **North Western University, Khulna.**
Contributions:
- Authored "Efficient Machine Learning Approaches to Detect Fake News of Covid-19."
- Explored data mining, cleaning and preprocessing.
- Explored TF-IDF and bag-of-words features.
- Explored ML and DNN algorithms.
- Explored data balancing with SMOTE, ADASYN.

03/2021 - 08/2021   **DGTED Lab, Khulna University.**
Research Advisor: Dr. Kazi Masudul Alam.

Contributions:
- Authored "An Empirical Study of Machine Learning-based Bangla News Classification Methods."
- Explored data scrapping from local newspaper.
- Explored data mining, cleaning and preprocessing.
- Explored TF-IDF and bag-of-words features.
- Explored ML and DNN algorithms

03/2021 - 08/2021   **Khulna University.**
Research Advisor: Dr. Kamrul Hasan Talukder.
Contributions:
- Authored "An Empirical Study to Detect Cyberbullying with TF-IDF and Machine Learning Algorithms."
- Explored data mining, cleaning and preprocessing.
- Explored TF-IDF and bag-of-words features.
- Explored ML and DNN algorithms.
- Explored data balancing with SMOTE, Adasyn.

03/2020 - 12/2021   **Undergraduate Research Student**
Computer Science and Engineering, Khulna University.
Research Advisor: Dr. Kamrul Hasan Talukder.
Contributions:
- Authored "Smartphone-based Skin-cancer detection using Image Processing and Support Vector Machine." (Undergraduate Thesis).
- Explored image processing with Opencv.
- Explored geometric and color features in skin lesion.
- Explored OpenCV in smartphones (Java).

SEMINARS, CONFERENCES, AND WORKSHOPS

Attended multiple National and International Seminars, Conferences, and Workshops as:

- 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics, Venue: Albuquerque, New Mexico April 29–May 4, 2025

- TrustNLP: Fifth Workshop on Trustworthy Natural Language Processing Collocated with the 2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2025), Venue: Albuquerque, New Mexico May 3, 2025

- 2025 UC Irvine Anti-Cancer Challenge Kickoff Event, Venue: UC Irvine, June 5, 2025

- Seminar on Testing Quantumness in the NISQ Era, Venue: UC Irvine, September 29, 2023.

- Seminar on Privacy in the Age of AI and the Internet of Things, Venue: UC Irvine, October 13, 2023.

- Seminar on Towards Full-Stack Adaptivity in Permissioned Blockchain Systems, Venue: UC Irvine,

October 20, 2023.

- Seminar on Cascade: A Platform for Fast Edge Intelligence, Venue: UC Irvine, October 27, 2023.
- Seminar on Rethinking Security and Privacy for Emerging Computing Platforms, Venue: UC Irvine, November 3, 2023.
- Seminar on Intelligent 3D Multimedia Displays using Flying Light Specks, Venue: UC Irvine, January 12, 2024.
- Seminar on Computer Science Foundations, Positive Social and Business Impact, and Research Opportunities, Venue: UC Irvine, January 19, 2024.
- Seminar on Towards Environmentally Responsible AI, Venue: UC Irvine, January 26, 2024.
- Seminar on Hydro: A Compiler Stack for Distributed Programs, Venue: UC Irvine, February 9, 2024.
- Seminar on BGP-iSec: Improved Security of Internet Routing Against Post-ROV Attacks, Venue: UC Irvine, March 1, 2024.
- Seminar on The EFX problem, Venue: UC Irvine, March 8, 2024.
- Seminar on Towards quantum-resistant security, Venue: UC Irvine, March 15, 2024.
- Seminar on Fast and Simple Sorting Using Partial Information, Venue: UC Irvine, April 5, 2024.
- Seminar on MAD Games - Multi-Agent Dynamic Games: What can you learn from Autonomous Racing?, Venue: UC Irvine, April 12, 2024.
- Seminar on Constraint-Based Control Design for Long Duration Autonomy, Venue: UC Irvine, April 19, 2024.

- Seminar on Towards Computational Acoustic Cameras: Neural Deconvolution and Rendering for Synthetic Aperture Sonar, Venue: UC Irvine, April 26, 2024.
- Seminar on Distribution-matching multi-agent deployment for coverage, Venue: UC Irvine, May 3, 2024.
- Seminar on From Hierarchical Clustering to Phylogenetic CSPs, Venue: UC Irvine, May 10, 2024.

- Seminar on Specifying Goals to Deep Neural Networks with Answer Set Programming, Venue: UC Irvine, May 17, 2024.
- Seminar on Incentives and Expressiveness in multi-winner approval-based elections, Venue: UC Irvine, May 24, 2024.
- Seminar on The Journey of Building Global-Scale Sustainable Blockchain Fabric, Venue: UC Irvine, May 31, 2024.
- Seminar on Deep Learning for Biological Sequence Design towards Personalized Immunotherapy,

Venue: UC Irvine, June 07, 2024.

- International Conference on MACHINE INTELLIGENCE & DATA SCIENCE APPLICATIONS (MIDAS 2021), Venue: Comilla University, Bangladesh, December 26-27, 2021.
- First International Cyber Security Summit 2021. Venue: Khulna University, Bangladesh. October 22, 2021.
- An International Conference on "*Big Data, IoT and Machine Learning (BIM 2021)*", Venue: Cox's Bazar, Bangladesh. September 23-25, 2021.

- An International Conference on "*2021 International Conference on Electronics, Communications and Information Technology (ICECIT).*", Venue: Khulna University, Bangladesh. September 14-16, 2021.
- An International Conference on "*12th INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT).*", Venue: IIT Kharagpur, India (Remote). July 6-8, 2021.
- Online Workshop on Smart City and Smart Village*,* Venue: Khulna University, Bangladesh. February 27, 2021.
- Training on Growing and Understanding your Business with Facebook, Venue: Khulna University, Bangladesh. February 17, 2018.
- Ruby on Rails Workshop, Venue: Khulna University, Bangladesh. August 12, 2017.

HONORS AND AWARDS

Summer-25	Awarded "Chao Family Comprehensive Cancer Center (CFCCC) Pilot Awards", supported by the Anti-Cancer Challenge – 2024.
Spring-24	Computer Science Department Research Fellowship.
Summer-24	Computer Science Department Research Fellowship.
2021	Best Paper Award (International Conference on MACHINE INTELLIGENCE & DATA SCIENCE APPLICATIONS (MIDAS 2021) December 26-27, 2021.)
2021	Bangladesh Hi-Tech Park (ICT Ministry) Startup Incubation Allotment for Skin Cancer Detector.
2021	Bangladesh Hi-Tech Park (ICT Ministry) Startup Incubation Allotment for ChatBot for blind.
2017 – 2021	University Merit Scholarship: Class Position and Highest Academic Credentials (Undergraduate Program).
2017 – 2021	Board Scholarship on Higher Secondary Certificate Examination.
2014 – 2016	Board Scholarship on Secondary School Certificate Examination.
2012 – 2014	Board Scholarship on Junior School Certificate Examination.

2012   Prothom Alo Internet Genius Champion.
2009   Divisional Math Olympiad 1st Runner up.
2008   Divisional Math Olympiad 1st Runner up.

ADDITIONAL TRAININGS

Around 30 courses on Coursera are done. Some of them are:

- Machine Learning for All, Offered by the **University of London**, Instructor: Prof Marco Gillies (Professor, Computing Department, Goldsmiths, University of London).

- Python Data Analysis, Offered by **Rice University,** Instructors: Scott Rixner (Professor, Computer Science, Rice University) & Joe Warren (Professor, Computer Science, Rice University).

- Basic Data Processing and Visualization, Offered by the **University of California San Diego**, Instructors: Julian McAuley (Assistant Professor, Computer Science, University of California San Diego) & Ilkay Altintas (Chief Data Science Officer, University of California San Diego).

- Machine Learning Foundations: A Case Study Approach, Offered by the **University of Washington**, Instructors: Emily Fox (Amazon Professor of Machine Learning, University of Washington) & Carlos Guestrin (Professor, Computer Science, Stanford University).

- Crash Course on Python, Offered by **Google**, Instructor: Google Career Certificates.

- Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning, Offered by **DeepLearning.AI**, Instructor: Laurence Moroney (Lead AI Advocate, Google).

- Neural Networks and Deep Learning, Offered by **DeepLearning.AI**, Instructor: Andrew Ng (Instructor Founder, DeepLearning.AI & Co-founder, Coursera).

- Applied Plotting, Charting & Data Representation in Python, Offered by the **University of Michigan,** Instructor: Christopher Brooks (Assistant Professor, School of Information, University of Michigan).

- Machine Learning with Python, Offered by **IBM**, Instructors: SAEED AGHABOZORGI (Ph.D., Sr. Data Scientist, IBM) & Joseph Santarcangelo (Ph.D., Data Scientist at IBM).

- Natural Language Processing in TensorFlow, Offered by **DeepLearning.AI**, Instructors: Laurence Moroney (Lead AI Advocate, Google).

- Introduction to Data Science in Python, Offered by the **University of Michigan,** Instructor: Christopher Brooks (Assistant Professor, School of Information, University of Michigan).

- HTML, CSS, and Javascript for Web Developers, Offered by **Johns Hopkins University**, Instructor:

Yaakov Chaikin (Adjunct Professor, Graduate Computer Science, Johns Hopkins University).

- Speak English Professionally: In Person, Online & On the Phone, Offered by **Georgia Institute of Technology**, Instructor: Amalia B. Stephens (Lecturer, Georgia Tech Language Institute).

SKILLS AND TECHNIQUES

Programming:
- C
- C++
- C#
- Assembly
- Java
- PHP/MySQL
- Python
- MATLAB
- JavaScript
- Android Studio
- Visual Studio
- Unreal Engine
- Unity Engine
- React

Machine Learning:
- Supervised Machine Learning.
- Unsupervised Machine Learning.

Deep Learning:
- CNN
- Pytorch
- Cuda
- RNN, LSTM, BiLSTM

Dynamic Deep Learning:
- Dynamic Slimming (Channel)
- Early Exit (Width

Data Science:
- Data Cleaning
- Data Manipulation
- Data Mapping
- Pattern Recognition

Datatype Quantization:
- LoRa
- QLoRA

Large Language Model Optimization Techniques:
- Paged Optimization
- Gradient Accumulation

Natural Language Processing:
- TF-IDF
- Textual Recognition
- Word2Vec

Computer Vision:
- Object Detection
- TensorFlow
- OpenCV

Web Programming:
- PHP
- SQL Database
- React

Simulation:
- Unity 3D
- Unreal Engine
- Gazebo

PROJECTS HIGHLIGHTS

- Summary the Savior: Harmful Keyword and Query-based Summarization for LLM Jailbreak Defense [GitHub] [Video] [Paper]
- Enhancing Intent Detection in Nicotine Replacement Therapy Tweets Using LLM: A Fine-Tuning Framework with Class Downsampling and Misclassification Refinement. [GitHub]
- 3D-Face-Object-Surface-Reconstruction-Using-Calibrated-and-Uncalibrated-Photometric-Stereo [GitHub] [Video]
- Detecting Malicious Prompts for Large Language Models (LLMs) [GitHub] [Paper]
- Smartphone-based Skin Cancer Detection using Image Processing and Support Vector Machine. [GitHub] [Paper]
- Smartphone-based Skin Cancer Detection using Image Processing and Convolutional Neural Network [Paper]
- An Empirical Study to Detect Cyberbullying with TF-IDF and Machine Learning Algorithms. [Paper]
- Efficient Machine Learning Approaches to Detect Fake News of Covid-19. [Paper]
- An Empirical Study of Machine Learning-based Bangla News Classification Methods [Paper]
- A Machine Learning Based Study to Predict Depression with Monitoring Actigraph Watch Data [Paper]
- Bangla News Headline Sentiment Analysis [GitHub] [Report]
- Bangla news Similarity and Searching tool for Disabled people [GitHub] [Report]
- Bangla news summarizer for Disabled people [GitHub] [Report]
- Bangla intelligent ChatBot [GitHub] [Report]
- Military Simulator 3D using Unity Engine [Github]
- 3rd Person Zombie Shooter 3D using Unreal Engine [Video]
- Bangla next word predictor [GitHub] [Report]
- Khulna University Hall and Housing Management [GitHub]
- Railway management Web Program [GitHub]
- Railway management Visual Studio version [GitHub]
- Parser and Lexical Analyzer for compiling [Parser-GitHub] [Lexical-GitHub]
- Numerical Methods analyzer [GitHub]

TEACHING EXPERIENCES

Spring 2025	Graduate Teaching Assistant (ICS 33 - Intermediate Programming at UC Irvine.)
Winter 2025	Graduate Teaching Assistant (ICS 33 - Intermediate Programming at UC Irvine.)
Fall 2024	Graduate Teaching Assistant (ICS 33 - Intermediate Programming at UC Irvine.)
Winter 2024	Graduate Teaching Assistant (ICS 33 - Intermediate Programming at UC Irvine.)
Fall 2023	Graduate Teaching Assistant (ICS 32A- Python Programming and Libraries (Accelerated)

2018 – 2021   Tutoring: Higher Secondary Mathematics, Biology and Physics.

2017 – 2019   Prothom Alo Math Olympiad Camp.

INTERNSHIP EXPERIENCES

01/2021 - 06/2021   Research Intern (Online), Next Solution Lab, Dhaka, Bangladesh.

07/2020 - 12/2020   Software Engineering Mentorship Program Intern (Online), Durbin Labs LTD, Dhaka, Bangladesh.

REFERENCES

Dr. Ian G. Harris
Professor
Department of Computer Science,
University of California Irvine,
Donald Bren Hall, Room 3088
Irvine, CA 92697, USA
Email: harris@ics.uci.edu
Phone: +1 949 824 8842

Dr. Cornelia Pechmann
Professor
The Paul Merage School of Business SB1-4317
University of California Irvine,
Irvine, CA 92697-3125
Email: cpechman@uci.edu
Phone: 310-892-0619

Dr. Kamrul Hasan Talukder
Dean & Professor
Science, Engineering and Technology School
Computer Science and Engineering Discipline
Khulna University, Khulna-9208, Bangladesh.
Email: khtalukder@cse.ku.ac.bd
Mobile: +880-1925-687763