

# Shoumik Saha

 [shoumiksaha.github.io](https://github.com/shoumiksaha)

 [smksaha@umd.edu](mailto:smksaha@umd.edu)

 [in/shoumik-saha](https://in.linkedin.com/in/shoumik-saha)

 [/ShoumikSaha](https://twitter.com/ShoumikSaha)

## RESEARCH INTEREST

---

- **Machine Learning** - Security & Reliability of GenAI, LLM Alignment & Hallucination
- **Computer Security** - Adversarial Attacks & Defenses, Malware

## EXPERIENCE

---

<b>Applied Scientist Intern</b> <i>Amazon AWS</i> Managers - <a href="#">Varun Kumar</a> , <a href="#">Zijian Wang</a>	<b>May 2025 – Aug 2025</b> <i>Full Time</i>
<b>Graduate Research Assistant</b> <i>Department of Computer Science</i> <a href="#">University of Maryland</a> Supervisor - <a href="#">Dr. Soheil Feizi</a>	<b>Aug 2023 – Present</b> <i>Full Time</i>
<b>Applied Scientist Intern</b> <i>Amazon AWS</i> Manager - <a href="#">Zijian Wang</a>	<b>Jun 2024 – Aug 2024</b> <i>Full Time</i>
<b>Graduate Research Assistant</b> <i>Maryland Cybersecurity Center</i> Supervisor - <a href="#">Dr. Tudor Dumitras</a>	<b>Aug 2022 – Jul 2023</b> <i>Full Time</i>
<b>Lecturer</b> <i>Department of Computer Science and Engineering</i> United International University	<b>Jul 2021 – Jul 2022</b> <i>Full Time</i>
<b>Research Assistant</b> <i>Data Science and Engineering Research Lab</i> Bangladesh University of Engineering and Technology (BUET)	<b>Mar 2021 – Jul 2022</b> <i>Part Time</i>

## EDUCATION

---

<b>University of Maryland - College Park (UMD)</b> <i>Ph.D. in Computer Science (Ongoing)</i> Advisor – <a href="#">Dr. Soheil Feizi</a>	<b>Aug 2022 – Present</b> <i>CGPA – 3.814/4.0</i>
<b>University of Maryland - College Park (UMD)</b> <i>M.S. in Computer Science</i> Advisor – <a href="#">Dr. Soheil Feizi</a>	<b>Aug 2022 – Aug 2024</b>
<b>Bangladesh University of Engineering and Technology (BUET)</b> <i>B.Sc. in Computer Science and Engineering</i> Thesis supervisor – <a href="#">Dr. Atif Hasan Rahman</a>	<b>Mar 2016 – Feb 2021</b> <i>Advanced GPA – 3.79/4.0</i>

## PUBLICATIONS

---

- **Breaking the Code: Security Assessment of AI Code Agents Through Systematic Jailbreaking Attacks**  
[[Shoumik Saha](#), J Chen\*, S Mayers\*, SK Gouda, Z Wang, V Kumar]  
[Under Review](#)([arXiv link](#))
- **Almost AI, Almost Human: The Challenge of Detecting AI-Polished Writing**  
[[Shoumik Saha](#), S Feizi]  
[ACL 2025](#) (*Association for Computational Linguistics*) ([Paper Link](#))  
Media Coverage: [The New York Times](#), [Plagiarism Today](#), [The Science Archive](#), [The Cheat Sheet](#)

- **Adversarial Paraphrasing: A Universal Attack for Humanizing AI-Generated Text**  
[Y Cheng, VS Sadasivan, **Shoumik Saha\***, M Saberi\*, S Feizi]  
*NeurIPS 2025* (Conference on Neural Information Processing Systems) ([Paper Link](#))
- **ProcVQA: Benchmarking the Effects of Structural Properties in Mined Process Visualizations on Vision-Language Model Performance**  
[KT Zinat\*, SM Abrar\*, **Shoumik Saha**, S Duppala, SN Sakhamuri, Z Liu]  
*EMNLP 2025* (Empirical Methods in Natural Language Processing) ([Paper Link](#))
- **LLM-Check: Investigating Detection of Hallucinations in Large Language Models**  
[G Sriramanan, S Bharti, VS Sadasivan, **Shoumik Saha**, P Kattakinda, S Feizi]  
*NeurIPS 2024* (Conference on Neural Information Processing Systems) ([Paper Link](#))
- **Fast Adversarial Attacks on Language Models In One GPU Minute**  
[VS Sadasivan, **Shoumik Saha\***, G Sriramanan\*, P Kattakinda, A Chegini, S Feizi]  
*ICML 2024* (International Conference on Machine Learning) ([Paper Link](#))  
Media Coverage: [The Register](#)
- **DRSM: De-Randomized Smoothing on Malware Classifier Providing Certified Robustness**  
[**Shoumik Saha**, W Wang, Y Kaya, S Feizi, T Dumitras]  
*ICLR 2024* (International Conference on Learning Representations) ([Paper Link](#))
- **MAlign: Explainable Static Raw-byte Based Malware Family Classification using Sequence Alignment**  
[**Shoumik Saha**, S Afroz, A Rahman]  
*Computers & Security Journal* ([Paper Link](#))
- **ML-Based Behavioral Malware Detection Is Far From a Solved Problem**  
[Y Kaya, Y Chen, Marcus Botacin, **Shoumik Saha**, F Pierazzi, L Cavallaro, D Wagner, T Dumitras]  
*SaTML 2025* (Secure & Trustworthy Machine Learning) ([Paper Link](#))
- **Contrastive Self-Supervised Learning Based Approach for Patient Similarity: A Case Study on Atrial Fibrillation Detection from PPG Signal**  
[**Shoumik Saha\***, S Shanto\*, A Rahman, M Masud, ME Ali]  
([arXiv Link](#))

## SCHOLARSHIPS & ACHIEVEMENTS

---

- Awarded 'Dean's Fellowship' from the University of Maryland
- Awarded 'Innovation Fund' for research from the ICT division of Bangladesh government
- Achieved 'Merit Stipends' from BUET in five out of seven terms
- Awarded 'Dean's Award' in Junior year for extra-ordinary result
- Achieved 'Talent-Pool Scholarship' in High School

## TEACHING

---

### Teaching Assistant

University of Maryland

Fall 2023

Data 200: Knowledge in Society: Science, Data and Ethics

### Lecturer

United International University

Fall 2021 - Summer 2022

Discrete Mathematics, Data Structure and Algorithm, Operating Systems

## SERVICES

---

- **Reviewer**, top-tier ML conferences such as ICML 2024, ICLR 2025, NEURIPS 2025, etc.
- **External Reviewer**, top-tier Computer Security conference, including USENIX 2023
- **Lab Instructor**, United International University
- **General Secretary**, BUET Photographic Society