Sayak Saha Roy
PhD Candidate, University of Texas at Arlington
Website: `https://sayaksaharoy.com`
Email: sayak.saharoy@mavs.uta.edu

EDUCATION

- **2020-2025 (Expected)**, PhD in Computer Science, University of Texas at Arlington
  Dissertation: Innovative Strategies for Preventing and Counteracting Social Engineering Scams
- **2019-2020**, MS in Computer Science, University of Texas at Arlington
  Thesis: Studying the effectiveness and reliability of Anti-phishing tools towards malicious websites circulating through Twitter
- **2014-2018**, B.Tech in Computer Science and Engineering, graduated Summa Cum Laude, Maulana Abul Kalam Azad University of Technology

PEER REVIEWED PUBLICATIONS

1. Elham Pourabbas Vafa, Mohit Singhal, Poojitha Thota, **Sayak Saha Roy**."Learning from Censored Experiences: Social Media Discussions around Censorship Circumvention Technologies" - To appear in the 46th IEEE Symposium on Security and Privacy (IEEE S&P 2025), (14.3% acceptance rate).
2. **Sayak Saha Roy**, Poojitha Thota, Krishna Vamsi Naragam, Shirin Nilizadeh. "From Chatbots to Phishbots?: Phishing Scam Generation in Commercial Large Language Models" - The 45th IEEE Symposium on Security and Privacy (IEEE S&P 2024), (**Distinguished Paper Award**, 18% acceptance rate).
3. Ana Aleksandric, **Sayak Saha Roy**, Hanani Pankaj, Gabriela Mustata Wilson, Shirin Nilizadeh. "Understanding the Ripple Effects: Users' Behavioral and Emotional Response to Toxicity in Twitter Conversations" - The 18th International AAAI Conference on Web and Social Media (ICWSM 2024), (17% acceptance rate).
4. **Sayak Saha Roy**, Unique Karanjit, Shirin Nilizadeh. "Phishing in the Free Waters: A Study of Phishing Attacks Created using Free Website Building Services" - ACM Internet Measurement Conference (IMC 2023), (23% acceptance rate).
5. **Sayak Saha Roy**, Dipanjan Das, Priyanka Bose, Christopher Kruegel, Giovanni Vigna, Shirin Nilizadeh. "Unveiling the Risks of NFT Promotion Scams" - The 18th International AAAI Conference on Web and Social Media (ICWSM 2024), (17% acceptance rate).
6. **Sayak Saha Roy**, Unique Karanjit, Shirin Nilizadeh. "Evaluating the Effectiveness of Antiphishing Reports on Twitter" - Symposium on Electronic Crime Research (eCrime) 2021 (**Best Paper Award,** 28% acceptance rate).
7. **Sayak Saha Roy**, Unique Karanjit, Shirin Nilizadeh. "What Remains Uncaught: Characterizing Sparsely Detected Malicious URLs on Twitter" - NDSS Workshop on Measurements, Attacks and Defences for the Web (MADWeb) 2021 (35% acceptance rate).

INVITED TALKS

1. "Abusing Large Language Models for Phishing Scam Generation" for Comcast Cybersecurity Research, May 2024.
2. "Exploring the Risks of Large Language Models in Facilitating Social Engineering Attacks" for the US Communications Sector Coordinating Council, April 2024.

TEACHING

1. **CSE 6388: Advanced Topics in Data Driven Security and Privacy** - Every Spring semester from Spring 2020 onwards.

   Teaching Assistant, Mentor, Guest Lecturer.
2. **CSE 4380/5380: Information Security** - Spring 2021, Spring 2022, Spring 2024, Fall 2024. Guest Lecturer, Teaching Assistant.

MEDIA INTERACTIONS

1. "AI: the drive to combat LLM-generated phishing attacks", interviewed by Hugo Sedouramane of Orange Research, August 2024.
2. "UTA researchers work to prevent AI phishing scams", interviewed by Brian Lopez of Media Relations, UT Arlington, June 2024
3. "Doctoral student becomes Twitter Ambassador", interviewed by Jeremy Agor of Communications, UT Arlington, December 2022.
4. "University of Texas at Arlington Takes APWG eCrime Symposium Top Paper Award" covered by Yahoo Finance, December 2021

ACHIEVEMENTS

1. **STEM Fellowship by Comcast (2022-present):** Conducting research on characterizing the cybercriminal ecosystem on Telegram. Previously identified social media platforms such as Twitter, Facebook, Reddit as a reliable source for iudentifying active phishing threats and developing a real-time framework, Social ThreatFinder to identify and report such threats to stakeholders at scale.
2. **Distinguished Paper Award (2024):** Awarded at the 45th IEEE Symposium on Security and Privacy (S&P 2024) for "From Chatbots to Phishbots?: Phishing Scam Generation in Commercial Large Language Models".
3. **Student Travel Grant (2024):** Recieved for attending the 45th IEEE Symposium on Security and Privacy (S&P 2024).
4. **Most Impact Talk (2023):** Awarded at the 50th Anniversary of Computing at UTA for "A Study of Phishing Attacks Created using Free Website Building Services".
5. **Best Talk Award (2023):** Awarded at the Student Computing Research Festival for "Characterizing and Mitigating NFT Phishing Scams".

6. **Developer Student Ambassador, Twitter (2022):** Focused on creating and conducting workshops aimed at making data science research more approachable for individuals outside the STEM community.

7. **CSE Outstanding Lab Service (2022):** Awarded in recognition of outstanding performance in leading major projects, offering mentorship to emerging researchers, and increasing the research productivity of the Security and Privacy Research Lab.

8. **John S. Schuchman Outstanding Doctoral Student Award (2022):** Recognition for outstanding Academic and Research performance during the 2022 Academic year.

9. **Best Paper Award (2021):** For the paper "Evaluating the Effectiveness of Antiphishing Reports on Twitter" at the Symposium on Electronic Crime Research.

10. **Upsilon Pi Epsilon Award (2020):** Recognized by the International Honor Society for the Computing and Information Disciplines for Outstanding Academic Performance.

11. **Student Travel Grant (2021):** Received for attending the 30th USENIX Security Symposium.

SERVICE

1. **Program Committee Member/Conference Reviewer (2020 - Present):** Program Committee Member at IEEE European Symposium on Security and Privacy (Euro S&P 2025) and AAAI Conference on Web and Social Media (2024, 2025). Reviewer at USENIX Security Symposium (2020-2023), APWG Symposium on Electronic Crime Research (2020-2022), AAAI Conference on Web and Social Media (2022,2023) and The International Symposium on Research in Attacks, Intrusions, and Defenses (2020).

2. **Doctoral Student Mentor, UT Arlington (Fall 2020 - Present):** Mentored five undergraduate students (as part of the UT Arlington Undergraduate Research Program) and four graduate students, providing hands-on guidance on developing research methodologies and crafting well-structured academic papers. As of August 2024, the mentorship has resulted in four students successfully publishing their research in peer-reviewed conferences.

SELECTED PROJECTS

PhishLang

The first open-source anti-phishing browser extension that runs entirely on the client-side without the need for referring to online blocklists while using mininal system resources. The extension and source can be downloaded from `https://github.com/UTA-SPRLab/phishlang`. This tool is a result of our work on utilizing MobileBERT to aid in the early identification of evasive phishing threats ((`https://arxiv.org/pdf/2408.05667`) which has led to the takedown of **nearly 26k phishing websites** in a period of 3.5 months.

Social ThreatFinder

A real-time framework that identifies phishing attacks and other online scams from social media websites and instant messaging originating from popular social media platforms and instant messaging services, including Facebook, Telegram, Instagram, and Discord, and monitors their coverage

across several anti-phishing tools and blocklists to identify gaps in the detection ecosystem. Also regularly reports these attacks to the Antiphishing Working Group (APWG), social media services and the domain hosting provider. As of July 2024, Social ThreatFinder has helped in the takedown of over **68.2k unique malicious domains**. An open-source implementation of the project (based on Twitter) is available here: `https://github.com/UTA-SPLAB/SocialThreatFinder`

Is it Phish?

A RoBERTa-based classifier designed for identifying adversarial prompts that can used in generating phishing scams using commercial Large Language Models. This project has exposed several vulnerabilities in prompt engineering within models such as ChatGPT and Bard and has led to responsible disclosures to OpenAI, Google and Claude (Anthropic). The classifier is available on Huggingface: `https://huggingface.co/phishbot/ScamLLM` and has been downloaded over 1,200 times in just four months.