Sifat Muhammad Abdullah

+1 (540)-449-2710 | sifat.abdullah577@gmail.com | https://sifatmd.github.io | Google Scholar

EDUCATION

Virginia Tech, Ph.D. in Computer Science, advisor: Dr. Bimal Viswanath — Jan 2021 - expected Dec 2025

BUET, B.S. in Computer Science and Engineering (GPA: 3.91/4.0) — 2015 - 2019

RESEARCH INTERESTS

Security of Multimodal LLMs & Text-to-Image (T2I) generation models, Multi-LLM reasoning, toxicity mitigation, deepfake detection.

SELECTED PUBLICATIONS

[**IEEE S&P'24**] **1st author**. "*An Analysis of Recent Advances in Deepfake Image Detection in an Evolving Threat Landscape*".

[**ACSAC'23**] **2nd author**. "*A First Look at Toxicity Injection Attacks on Open-domain Chatbots*".

[**IEEE S&P'23**] **2nd author**. "*Deepfake Text Detection: Limitations and Opportunities*". Dataset requested by **143 research groups**.

SELECTED PROJECTS

Adversarial Robustness of Multimodal LLMs | Ongoing work

- Defending MLLMs against diverse adversarial attacks using FLUX, GPT-4o, along with Kimi-VL-A3B-Thinking model with test-time reasoning.

Multi-LLM Reasoning | Submitted work during ML Internship

- Customizing LLaMA 3 & Qwen 3 for cooling data centers (DC), cutting carbon use by 24.3% over RL-models.
- Multi-LLM reasoning with QwQ-32B, reducing DC energy usage by 43.7% over single-LLM controllers.

Deepfake Image Detection | Published in **IEEE S&P'24**

- Studied 8 state-of-the-art deepfake image detectors using Diffusion and GAN-based text-to-image generators.
- Developed adversarial attacks using LoRA and Vision Foundation models without adding adversarial noise.
- Achieved more than 70% recall score degradation against most of the deepfake image detectors.

Toxicity Injection Attacks | Published in **ACSAC'23**

- Studied toxicity injection attacks on chatbots after deployment in a Dialog-based Learning setup.
- Proposed fully automated injection attacks using public LLMs eliciting up-to 60% response toxicity rate.

Deepfake Text Detection | Published in **IEEE S&P'23**

- Evaluated SOTA deepfake text detectors, e.g., BERT and GPT-2 based defenses on real-world datasets.
- Our adversarial attack achieves up-to 91.3% evasion rate while maintaining linguistic quality of text.

EXPERIENCE

HPE Labs – ML Research Associate Intern — May 2025 - Aug 2025

Virginia Tech SecML Lab – Graduate Research Assistant — Jan 2022 - Apr 2025 | Sep - Dec 2025

Virginia Tech – Graduate Teaching Assistant — Jan 2021 - Dec 2021

BUET DataLab – Graduate Research Assistant — Jan 2020 - Dec 2020

REVE Systems – Software Engineer — May 2019 - Dec 2019

ACHIEVEMENTS

- Pratt Fellowship, CS@VirginiaTech, 2025
- **Invited Talk**: VT Skillshop Series: Leveraging Creative Technologies (Oct 2023)
- *The Rise of the Chatbots* - Communications of the ACM: 2023

TECHNICAL SKILLS

- **GenAI Technologies**: MLLMs/VLMs, LLMs, T2I models, LoRA, Foundation Model Fine-tuning
- **Languages & Frameworks**: Python, C/C++, Bash, Java, PyTorch, TensorFlow, Keras, Django
- **Developer Tools**: Git, Vim, Jupyter Notebook, VS Code, Markdown, LaTeX, Linux, Docker