

# Sifat Muhammad Abdullah

+1 (540)-449-2710 | [sifat.abdullah577@gmail.com](mailto: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 and Adversarial Robustness of Multimodal LLMs, LLMs & Deepfake defenses. Foundation model detection of T2I deepfakes. Defending Multimodal LLMs using test-time Reasoning. Toxicity mitigation in Large Language Models. Reasoning in Multi-LLM systems. Efficient LLM Fine-tuning.

## SELECTED PUBLICATIONS

[**NeurIPS MLForSys W'25**] **Co-author**. “*Sustainable Control of Geo-Distributed Datacenters by Distilling Numerical Experts into Adaptive LLM Agents*”.

[**IEEE S&P'24**] **1st author**. “*An Analysis of Recent Advances in Deepfake Image Detection in an Evolving Threat Landscape*”. Resources requested by **40 research groups**.

[**ACSAC'23**] **Co-author**. “*A First Look at Toxicity Injection Attacks on Open-domain Chatbots*”.

[**IEEE S&P'23**] **Co-author**. “*Deepfake Text Detection: Limitations and Opportunities*”. Resources requested by **158 research groups**.

## SELECTED PROJECTS

**Adversarial Robustness of Multimodal LLMs** | Ongoing work

- Defending MLLMs against diverse adversarial attacks using FLUX & GPT-4o image translation, along with Kimi-VL-A3B-Thinking model with test-time reasoning, gaining >98% CLIPScore in image captioning in one of the case studies.

**Protection Scheme Evaluation** | Under submission

- Studied robustness of 8 state-of-the-art defenses, including watermarking & text-to-image model style mimicry.
- Achieved up-to 100% attack success while preserving image utility, using GenAI-based image translation.

**Multi-LLM Reasoning** | Under submission

- Utilized multi-turn debate with multi-LLM reasoning by deploying QwQ-32B, reducing data center energy usage by 43.7% over single-LLM systems.

**Distilling Experts into Adaptive LLMs** | Published in **NeurIPS MLForSys W'25**

- Customizing LLaMA 3 & Qwen 3 for cooling data centers (DC) using efficient fine-tuning.
- Achieved 24.3% gain in energy consumption over rule-based controllers, along with explainability.

**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

<b>HPE Labs</b> – ML Research Associate Intern	May 2025 - Aug 2025
<b>Virginia Tech SecML Lab</b> – Graduate Research Assistant	Jan 2022 - Apr 2025   Aug - Dec 2025
<b>Virginia Tech</b> – Graduate Teaching Assistant	Jan 2021 - Dec 2021
<b>BUET DataLab</b> – Graduate Research Assistant	Jan 2020 - Dec 2020
<b>REVE Systems</b> – Software Engineer	May 2019 - Dec 2019

## ACHIEVEMENTS

---

- Pratt Fellowship, CS@VT 2025
- CCI SWVA Cyber Innovation Scholarship 2024 - 2025
- Invited Talk: VT Skillshop Series: Leveraging Creative Technologies 10/2023
- *The Dark Side of AI* - VPM News Focal Point 10/2023
- CCI Student Spotlight 2023
- *The Rise of the Chatbots* - Communications of the ACM 7/2023
- *The strengths and limitations of approaches to detect deepfake text* - TechXplore 11/2022
- BUET Dean's List Award 2015 - 2019

## TECHNICAL PROGRAM COMMITTEES

---

- Deepfake, Deception, and Disinformation Security Workshop (3D-Sec), 2025
- IEEE Transactions on Information Forensics and Security (IEEE TIFS), 2025
- 4th Workshop on the Security Implications of Deepfakes and Cheapfakes (WDC), 2025
- Pervasive and Mobile Computing (PMC) Journal, 2025

## 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

## REFERENCES

---

- **Bimal Viswanath**, Associate Professor, Department of Computer Science, Virginia Tech.
- **Peng Gao**, Assistant Professor, Department of Computer Science, Virginia Tech.
- **Murtuza Jadliwala**, Associate Professor, Department of Computer Science, UT San Antonio.