Shravya Kanchi

(540) 449-7120 | shravya@vt.edu | shrave.github.io | github.com/shrave | Google Scholar

Education

Virginia Polytechnic Institute and State University (Virginia Tech) — Blacksburg, VA
Ph.D. in Computer Science, GPA 3.9/4.0 — *Aug. 2021 – Expected May 2026*

IIIT Hyderabad — Hyderabad, India
MS by Research in Computer Science, GPA 9/10.0 — *Jul. 2018 – Aug. 2021*

IIIT Sricity — Sricity, India
B.Tech in Computer Science and Engineering with Honors, GPA 8.79/10.0 — *Jul. 2014 – May. 2018*

Selected Projects

Synthetic data to strengthen security defenses — Asia CCS 2026

- Engineered a novel VQVAE-based generative AI model for numerical tabular data generation, driving a 20% performance improvement across five varied ML-based security defenses.
- Outperformed five SOTA tabular data generators—CTABGAN, TVAE, TabDDPM, REaLTabFormer, and GReaT—across diverse model families (LLMs, GANs, VAEs) on security classification tasks.
- Conducted a comprehensive measurement study, identifying 3 major data challenges impacting security ML defenses by analyzing 35 top-tier security research articles.

Toxicity mitigation in conversational AI training pipelines — Under Submission

- Safety-aligned chatbots using synthetically crafted conversations via Direct Preference Optimization (DPO), achieving near-zero toxicity in conversational pipelines.
- Proposed context-aware and context-agnostic approaches prompting LLMs to identify toxic conversations which outperformed industry API services in a data poisoning setting.

LLM Security Test Automation (OpenAI Cybersecurity Grant Program) — Ongoing

- **Primary Researcher on OpenAI Cybersecurity Grant project** focused on building agentic automation for scalable vulnerability validation; designed multi-agent systems to generate compilable, CVE-triggering JUnit tests for open-source library benchmarking.
- Created benchmarking pipelines for exploit detection in vulnerable open-source libraries to enhance supply-chain security in 50 Java client codebases.

Publications

[**Asia CCS'26**] "Lessons Learned from Integrating Generative AI into ML-based Security Tasks" 1^{st} **author**. (Link)
[**ArXIV**] "A Defense Framework to Mitigate Toxicity While Fine-tuning Conversational AI" 1^{st} **author**. (Link)
[**IEEE S&P'24**] "An Analysis of Recent Advances in Deepfake Image Detection in an Evolving Threat Landscape." 3^{rd} author. (Link)
[**ACSAC'23**] "First Look at Toxicity Injection Attacks on Open-domain Chatbots." 3^{rd} author. (Link)
[**WWW Workshop'22**] "SEBI Regulation Biography" 3^{rd} author. (Link)
[**CODASPY'21**] "A multi perspective access control in a smart home." 1^{st} **author**. (Link)

Work Experience

Graduate Research Assistant — Aug. 2021 – Present
Virginia Tech

- Conducted research on secure and responsible generative AI, including synthetic tabular data generation to boost performance of security defenses, chatbot toxicity mitigation via DPO and LLM classifiers, and agent-based automation of jailbreak attacks outperforming SOTA baselines.

Graduate Research Assistant — Jan. 2021 – Jun 2021
IIIT Hyderabad & JP Morgan Chase

- Built the first named-entity labeled corpus for 7,500 SEBI sub-regulations with 7 proposed domain-specific entity types, and developed an overlapping NER tool achieving 87.47% precision.

Technical Skills

GenAI: Agentic Workflows, AI Safety Evaluation, LLMs, LoRA, model customization, DPO, Stable Diffusion, Transformers, Variational Autoencoder (VAE), Generative Adversarial Networks (GANs), prompt engineering, hyper-parameter search, BERT, Fine-tuning, PEFT.

Security: Malware detection, Vulnerability, Spam, Network IDS, concept drift, website privacy, anomaly detection.

Frameworks: Deep learing - Pytorch, Tensorflow, Keras, Flask, MongoDB, AutoML, CUDA, JUnit.

Developer Tools: Git, Linux, VS Code, Vim, Jupyter Notebook, Docker, Nmap, label-studio, SLURM

Libraries: NumPy, Huggingface, Transformers, RayTune, Matplotlib, gnuplot, Scikit-Learn, pandas, LightGBM, SpaCy

Languages: Python, C++, C, Shell, SQL