Vishnu Asutosh Dasu

582-203-9641 | vdasu@psu.edu | linkedin.com/in/vdasu | Personal Website | github.com/vdasu

Ph.D. Researcher in Trustworthy AI and Software Security with top-tier publications in ICCV, NDSS, ICSE, and AAAI. Expert in securing the end-to-end AI lifecycle, addressing challenges in model robustness, fairness alignment, and privacy.

Education

Pennsylvania State University — State College, PA
Doctor of Philosophy, Computer Science and Engineering — *Aug. 2024 – May 2028*

Pennsylvania State University — State College, PA
Master of Science, Computer Science and Engineering — *Aug. 2022 – May 2024*

Manipal Institute of Technology — Manipal, India
Bachelor of Technology, Computer Science and Engineering — *July. 2016 – July 2020*

Selected Publications [Google Scholar]

(* Equal Contribution)

- **Dasu** et. al, *Attention Pruning: Automated Fairness Repair of Language Models via Surrogate Simulated Annealing*, International Conference on Software Engineering **[ICSE 2026]**
- Zheng, **Dasu** et. al, *Improving Noise Efficiency in Privacy-preserving Dataset Distillation*, International Conference on Computer Vision **[ICCV 2025]**
- Abadi*, **Dasu***, Sarkar*, *Privacy-preserving Data Deduplication for Enhancing Federated Learning of Language Models*, Network and Distributed Systems Security Symposium **[NDSS 2025]**
- **Dasu** et. al, *Neufair: Neural network fairness repair with dropout*, International Symposium on Software Testing and Analysis **[ISSTA 2024]**
- Rashid, **Dasu**, et. al, *Chain-of-Thought Driven Adversarial Scenario Extrapolation for Robust Language Models*, Annual AAAI Conference on Artificial Intelligence, **[AAAI 2026]**
- **Dasu** et. al, *PROV-FL: Privacy-preserving Round Optimal Verifiable Federated Learning*, ACM Workshop on Artificial Intelligence and Security, **[AISEC, CCS 2022]**

Relevant Experience

Graduate Research Assistant — August 2024 – Present
Pennsylvania State University — *State College, PA*
- Research assistant in the Systems & Internet Infrastructure Security (SIIS) Lab
- Topics: Fairness of LLMs, Code Generation LLMs, Automated Code Translation, Trustworthy AI, Agentic AI Security
- 5 peer-reviewed publications (3 first-author) at top-venues, 3 under review (1 first-author), 2 in progress (1 first-author)

Security Researcher — Sept 2020 – June 2022
Tata Research Development and Design Centre (TRDDC) — *Remote*
- Designed a private and verifiable federated learning protocol that aggregates gradients in 1 round of communication.
- Developed an autoencoder-based approach for insider threat detection that improved detection rate by 50% over baseline.
- 1 peer-reviewed publication at top-venue (AISEC, ACM CCS 2022)

Technical Skills

Languages: Python, C/C++, Java, Rust, HTML/CSS, JavaScript, SQL
Frameworks: PyTorch, Tensorflow, Transformers, PEFT, Numpy, Pandas, Scikit-learn, OpenSSL, Angr, GMP, LaTeX, Git
AI/ML Expertise: LLMs, VLMs, Code LLMs, Synthetic Data Generation, Trustworthy AI (Fairness, Privacy, Robustness), Quantization, Agentic AI Security, Pre-training, Alignment, and Post-training of LLMs
Areas of Expertise: Security & Privacy, Applied Cryptography, Software Engineering

Ongoing Projects

- **Verifiable and Secure Code Translation from C to Rust**
- **Robustness of Multi-Modal Language Models against Adversarial Perturbations**
- **Secure and Private Agentic AI Systems**
- **Information-Theoretic Private Federated Learning**
- **Trustworthy and Robust Code Generation LLMs**

Awards

- **Distinguished Artifact Reviewer Award, ACM CCS 2025**
- **Internet Society Fellowship, 2025**
- **Student Travel Grant, ICCV 2025**
- **TCS Citation Award** 3× Recipient: Awarded for outstanding research contributions to Tata Research.