

Tingwei Zhang

🏠 ztingwei.com | ✉ tingwei@cs.cornell.edu | 🌐 [Tingwei-Zhang](#)

RESEARCH INTERESTS

Tingwei focuses on exploring security and privacy challenges in machine learning technologies, particularly in real-world scenarios and under adversarial conditions, to develop secure, ethical, and privacy-preserving AI systems.

EDUCATION

Cornell Tech 2023–Present

Ph.D. in Computer Science

– Advised by [Vitaly Shmatikov](#)

University of Virginia (UVA) 2020–2023

B.A. in Computer Science with Minor in Statistics

– Graduated with *Highest Distinction* in [Distinguished Majors Program](#) in computer science.

– Worked with Prof. [David Evans](#) and Prof. [Yuan Tian](#) on security of machine learning projects at [SRG](#) at UVA.

PUBLICATIONS

Conference Papers

Tingwei Zhang, C. Zhang, J. X. Morris, E. Bagdasarian, V. Shmatikov, "Self-interpreting Adversarial Images," in *Proceedings of the 34th USENIX Security Symposium (USENIX Security)*, Seattle, WA, USA, 2025. arxiv.org/abs/2407.08970 (Artifacts available, Artifacts functional, Results reproduced)

Tingwei Zhang*, R. Jha*, E. Bagdasarian, V. Shmatikov, "Adversarial illusions in multi-modal embeddings," in *Proceedings of the 33rd USENIX Security Symposium (USENIX Security)*, Philadelphia, PA, USA, 2024. arxiv.org/abs/2308.11804 ([Distinguished Paper Award](#), Artifacts available, Artifacts functional, Results reproduced)

S. Fnu, A. Suri, [Tingwei Zhang](#), J. Hong, Y. Tian, and D. Evan, "SoK: Pitfalls in evaluating black-box attacks," in *Proceedings of the 2nd IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, Toronto, Canada, 2024. arxiv.org/abs/2310.17534

Preprints

Tingwei Zhang, S. Fnu, R. Jha, C. Zhang, V. Shmatikov, "Adversarial hubness in multi-modal retrieval," in *Preprint*, 2025. arxiv.org/pdf/2412.14113

C. Zhang, [Tingwei Zhang](#), V. Shmatikov, "Adversarial Decoding: Generating Readable Documents for Adversarial Objectives," in *Preprint*, 2024. arxiv.org/pdf/2410.02163

See [Google Scholar profile](#) for a full list.

INTERNSHIP

Security Engineering Intern, Amazon AGI Security, Seattle, WA May–Aug. 2025

– Mentored by [Akash Charan Cholleti](#) and [Nur Gucu](#) on red-teaming audio language models.

– Developed an automated framework to detect security vulnerabilities in audio-based AI systems, integrating with existing AGI security testing tools.

TEACHING

Cornell Tech CS5434: Trustworthy AI, TA Fall 2025

Cornell Tech CS5450: Networked and Distributed Systems, Head TA Spring 2024

Cornell University CS2110: Object-Oriented Programming and Data Structures, TA Fall 2023

UVA CS4774: Machine Learning, TA Fall 2022

UVA CS4102: Algorithms, TA Spring 2022

HONORS & AWARDS

Distinguished Paper Award at USENIX Security	Aug. 2024
USENIX Security Student Grant, 2024	Aug. 2024
Dean's List of Distinguished Students, College of Arts & Sciences, UVA	2021 & 2022

TALKS & PRESENTATIONS

Self-interpreting Adversarial Images	Aug. 2025
<i>Conference Talk, USENIX Security Symposium 2025</i>	
Adversarial Illusions in Multi-Modal Embeddings	Apr. 2025
<i>Invited Talk, RSAC 2025</i>	
Attacking and Defending Multi-Modal Representations	Dec. 2024
<i>Invited Talk, University of Virginia CS, Research Seminar</i>	
Adversarial Illusions in Multi-modal Embeddings	Aug. 2024
<i>Conference Talk, USENIX Security Symposium 2024</i>	

SERVICES

Cornell University CS Student Reviewer: Ph.D. Admission Committee 2025
