

# Yue Gao

☎ +1 (608) 733-8789 ✉ gy@cs.wisc.edu 🏠 pages.cs.wisc.edu/~gy 🌐 ygao234 🔄 Lodour 📍 Madison, WI

## RESEARCH INTERESTS

---

**Trustworthy Machine Learning** (adversarial robustness, black-box evasion attacks and defenses)

**System Security** (machine learning systems, web-based applications and services)

## EDUCATION

---

**University of Wisconsin–Madison**

Madison, WI

*Ph.D. Candidate in Computer Science*

*Sep 2018 – May 2024 (expected)*

- Advisor: Prof. Kassem Fawaz
- Thesis: *Characterizing the Limitations of Defenses in Adversarial Machine Learning*

**Shanghai University**

Shanghai, China

*B.S. in Computer Science and Technology*

*Sep 2014 – Jul 2018*

- Major GPA: 3.99/4.00 (ranked 1/292)
- Advisor: Prof. Xiaodong Yue
- Thesis: *A Deep Neural Network based Image Compression Method*

## WORK EXPERIENCE

---

**Research Assistant @ University of Wisconsin–Madison**

Madison, WI

*Advised by Prof. Kassem Fawaz*

*Nov 2018 – present*

- Explore the weaknesses of evasion attacks and defenses for ML-based systems.
- Improve the security analysis of ML-based and web-based systems in black-box settings.

**Research Intern @ Microsoft Research**

Redmond, WA

*Mentored by Dr. Jay Stokes and Dr. Emre Kiciman*

*Jun 2021 – Sep 2021*

- Explore data-centric solutions for backdoor attacks on language models with domain knowledge.
- Design auditing frameworks for the continual update of backdoor-free language models.

**Research and Development Intern @ TuCodec**

Shanghai, China

*Mentored by Dr. Chunlei Cai*

*Jan 2018 – Jul 2018*

- Optimize learning-based image compression algorithms.
- Develop DNN-based applications on mainstream operating systems and deploy them to cloud services.
- Winner of the CVPR 2018 Challenge on Learned Image Compression.

## SELECTED PROJECTS

---

**Shareable and Explainable Attribution for Black-box Attacks on ML systems**

*Jan 2023 – Aug 2023*

- Characterize the attack's progression for forensic purposes and human-explainable intelligence sharing.
- Fingerprint and attribute zero-day attacks on their first and second occurrence, respectively.
- Discover specific minor implementation bugs in popular ML attack toolkits.

**The Role of Randomization in Adversarial Robustness**

*Feb 2022 – May 2022*

- Characterize the limitations of using randomization to defend ML models.
- Theoretically explain the source of robustness for randomized defenses against evasion attacks.

**Trustworthy Machine Learning in Real-World Systems**

*Sep 2020 – Jan 2021*

- Explore the security of ML systems under threats from multiple components.
- Demonstrate amplified threats from the interplay between multiple vulnerabilities.

**Defending against Evasion Attacks in Multimodal Scenarios (Collaborative)**

*Since 2019 (semiannual)*

- Improve adversarial robustness by enforcing physical constraints.
- Design defenses for multimodal tasks (e.g., remote sensing satellites and autonomous driving).
- Develop a usable code base for team members with varying tracks and technical backgrounds.

**Conference**

- [1] On the Limitations of Stochastic Pre-processing Defenses  
**Yue Gao**, Ilia Shumailov, Kassem Fawaz, and Nicolas Papernot  
*Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS)*, 2022
- [2] Rethinking Image-Scaling Attacks: The Interplay Between Vulnerabilities in Machine Learning Systems  
**Yue Gao**, Ilia Shumailov, and Kassem Fawaz  
*Proceedings of the 39th International Conference on Machine Learning (ICML)*, 2022  
*Oral Presentation (Top 2%)*
- [3] Experimental Security Analysis of the App Model in Business Collaboration Platforms  
Yunang Chen\*, **Yue Gao**\*, Nick Ceccio, Rahul Chatterjee, Kassem Fawaz, and Earlence Fernandes  
*31st USENIX Security Symposium (USENIX Security)*, 2022  
*Bug Bounty (\$1500)*
- [4] I Know Your Triggers: Defending Against Textual Backdoor Attacks With Benign Backdoor Augmentation  
**Yue Gao**, Jack W. Stokes, Manoj Prasad, Andrew Marshall, Kassem Fawaz, and Emre Kiciman  
*IEEE Military Communications Conference (MILCOM)*, 2022

**Workshop**

- [1] Variational Autoencoder for Low Bit-rate Image Compression  
Lei Zhou\*, Chunlei Cai\*, **Yue Gao**, Sanbao Su, and Junmin Wu  
*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018  
*Winner of the first Challenge on Learned Image Compression*

**Preprints**

- [1] SEA: Shareable and Explainable Attribution for Query-based Black-box Attacks  
**Yue Gao**, Ilia Shumailov, and Kassem Fawaz  
*arXiv*, 2023
- [2] Analyzing Accuracy Loss in Randomized Smoothing Defenses  
**Yue Gao**\*, Harrison Rosenberg\*, Kassem Fawaz, Somesh Jha, and Justin Hsu  
*arXiv*, 2020

---

**SELECTED HONORS & AWARDS**

<b>Slack Bug Bounty:</b> Medium Severity, \$1500	2022
<b>Top 10% Reviewers Award:</b> NeurIPS	2022
<b>CVPR Competition Winner:</b> Challenge on Learned Image Compression	2018
<b>National Scholarship:</b> China	2017
<b>Top 100 Elite Collegiate Award:</b> China Computer Federation	2017
<b>Scholarship for Exceptional Leadership:</b> Shanghai University	2017
<b>City Scholarship:</b> Shanghai	2016
<b>Outstanding Student Award:</b> Shanghai University	2016
<b>Outstanding Volunteer Award:</b> ACM ICPC Asia Regional Contest	2016
<b>Scholarship for Exceptional Innovation:</b> Shanghai University	2016
<b>Scholarship for Exceptional Academic Achievements:</b> Shanghai University	2015 – 2018
<b>Bronze Prize for Programming Contest:</b> ACM ICPC Asia East-Continent Final Contest	2015
<b>Bronze Prize for Programming Contest:</b> ACM ICPC Asia Shanghai Regional Contest	2015

---

**PROFESSIONAL ACTIVITIES**

<b>Reviewer:</b> NeurIPS and ICML	2022 – 2023
<b>External Reviewer:</b> USENIX Security Symposium	2021 – 2022
<b>External Reviewer:</b> IEEE Symposium on Security and Privacy	2021 – 2022
<b>External Reviewer:</b> ACM Conference on Computer and Communications Security	2019
<b>Team Leader:</b> Collegiate ICPC Team at Shanghai University	2016 – 2017

## TALKS

---

1. **The Vulnerabilities of Preprocessing in Adversarial Machine Learning** Apr 2023  
*TrustML Young Scientist Seminar, RIKEN AIP*
2. **On the Limitations of Stochastic Pre-processing Defenses** Oct 2022  
*University of Southern California (virtual)*
3. **The Interplay Between Vulnerabilities in Machine Learning Systems** Sep 2022  
*University of Michigan*
4. **Experimental Security Analysis of the App Model in Business Collaboration Platforms** Aug 2022  
*USENIX Security 2022*
5. **The Interplay Between Vulnerabilities in Machine Learning Systems** Jun 2022  
*ICML 2022*

## TEACHING AND MENTORING

---

<b>Teaching Assistant:</b> CS 368 (C++ for Java Programmers), University of Wisconsin–Madison	Fall 2018
<b>Guest Lecturer:</b> Advanced Algorithms & Data Structures, Shanghai University	2015 – 2017
<b>Problem Designer:</b> Undergraduate Programming Contests, Shanghai University	2015 – 2017
<b>Student Mentor:</b> Undergraduate Computer Science Coursework, Shanghai University	2015 – 2017

## TECHNICAL SKILLS

---

<b>Python</b>	Research (2018 – present), System Optimization (2018), Backend Development (2016 – 2017).
<b>PyTorch</b>	Research (2019 – present), Distributed Training (2020 – 2022).
<b>Docker</b>	Research (2018 – present), Computing Cluster (2017 – 2018).
<b>C / C++</b>	Linux Kernel (2019), ML System (2018), Programming Contest (2014 – 2018).
<b>TensorFlow</b>	Service Deployment (2018).
<b>Java EE</b>	Backend Development (2016).

## ARTICLES AND MEDIA COVERAGE

---

<b>CleverHans.</b> Can stochastic pre-processing defenses protect your models?	2022
<b>USENIX login.</b> Experimental Security Analysis of the App Model in Business Collaboration Platforms	2022
<b>Wired.</b> Slack's and Teams' Lax App Security Raises Alarms	2022