Huiying (Helen) Li

Contact Information

Address: 5730 S Ellis Ave, Chicago, IL, United States 60637
E-mail: huiyingli.biz@gmail.com
Homepage: `http://people.cs.uchicago.edu/~huiyingli/`

Education

The University of Chicago, Chicago, IL, United States *Sept. 2017 - Mar. 2023 (Expected)*
Ph.D. in Computer Science *GPA:* **4.0/4.0**

The University of Chicago, Chicago, IL, United States *Sept. 2017 - Mar. 2020*
M.S. in Computer Science

Fudan University, Shanghai, China *Sept. 2013 - Jun. 2017*
B.S. in Computer Science and Technology

Awards

- Siebel Scholarship (2021)
- CHI Honorable Mention Award (2020)
- Facebook Fellowship (2020)
- Two Sigma Fellowship Finalist (2020)
- National Scholarship, China (**Top 0.2%** in China) (2015)
- Outstanding Student, Fudan University (2014)

Industry Experience

PhD Software Engineer Intern, Meta *Jun. 2022 - Sept. 2022*
I implement my work "Blacklight: Scalable Defense for Neural Networks against Query-Based Black-Box Attacks." (USENIX Security'22) with the AI Security Team @ Meta.

Research Intern, Microsoft Research *Jun. 2020 - Sept. 2020*
I work with the Security + AI RIP group @Microsoft Research on measuring the utility of defenses against adversarial ML attacks.

Academic Experience

Research Assistant, SAND Lab, University of Chicago *Sept. 2017 - Present*
Supervised by Prof. Ben Y. Zhao and Prof. Heather Zheng

Teaching Assistant, University of Chicago
CMSC 23360 Advanced Networks *Spring 2021*
CMSC 23400 Mobile Computing *Winter 2018*
MPCS 52011 Introduction to Computer Systems *Autumn 2017*

Shadow Program Committee, IEEE Symposium on Security and Privacy, 2021

Research

SAND Lab, University of Chicago *Sept. 2017 - present*
Ph.D. Student Supervised by Prof. Ben Y. Zhao and Prof. Heather Zheng

- **ML Security and Robustness**
  - Attacks and defenses for DNN backdoor attacks.
    I have been working on discovering, understanding and mitigating attacks on Deep Neural Networks. We proposed the first detetcion and mitigation system, Neural Cleanse, to defend against DNN backdoor attacks. We also proposed latent backdoor, a more powerful and stealthy variant of backdoor attacks that functions under transfer learning and provide an effective defense against it. I'm also working on understanding backdoor attacks on time-varying models.
  - Defenses for adversarial attacks against DNNs.
    I have been working on detecting and mitigating adversarial attacks against Deep

Neural Networks. We designed the first global, scalable defense system against query-based black-box adversarial attacks against DNNs.

- **Human Privacy Protection**
  - Image ”Cloaking” for human facial privacy.
    We developed an algorithm and a software called Fawkes, which allow individuals to limit how unknown third parties can track them by building facial recognition models out of their publicly available photos.
  - Wearable jammer against commercial microphones.
    We designed and engineered a wearable bracelet which can disable all the microphones in the users' surroundings. Our device leverages the fact that when exposed to ultrasonic noise, commodity microphones will leak the noise into the audible range.

Publication

Huiying Li, Arjun Nitin Bhagoji, Ben Y Zhao, Haitao Zheng. “Can Backdoor Attacks Survive Time-Varying Models?” arXiv preprint arXiv:2206.04677 (2022).

Huiying Li, Shawn Shan, Emily Wenger, Jiayun Zhang, Haitao Zheng, Ben Y. Zhao. “Blacklight: Scalable Defense for Neural Networks against Query-Based Black-Box Attacks.” In Proceedings of *The 31th USENIX Security Symposium.* Boston, MA, Aug. 2022. **USENIX Security'22**

Huiying Li, Emily Wenger, Shawn Shan, Ben Y. Zhao, Haitao Zheng. “Piracy Resistant Watermarks for Deep Neural Networks.” arXiv preprint arXiv:1910.01226 (2020).

Shawn Shan, Emily Wenger, Jiayun Zhang, **Huiying Li**, Haitao Zheng, Ben Y. Zhao. “Fawkes: Protecting Privacy against Unauthorized Deep Learning Models.” In Proceedings of *The 29th USENIX Security Symposium.* Boston, MA, Aug. 2020. **USENIX Security'20**

Yuxin Chen*, **Huiying Li***, Shan-Yuan Teng*, Steven Nagels, Zhijing Li, Pedro Lopes, Ben Y. Zhao, and Haitao Zheng. “Wearable Microphone Jamming.” In Proceedings of *The CHI Conference on Human Factors in Computing Systems.* Honolulu, HI, Apr. 2020. **ACM CHI'20** (**Honorable Mention Award**)
** denotes equal contribution*

Yuanshun Yao, **Huiying Li**, Haitao Zheng, and Ben Y. Zhao. “Latent Backdoor Attacks on Deep Neural Networks.” In Proceedings of *The 26th ACM Conference on Computer and Communication Security.* London, UK, Nov. 2019. **ACM CCS'19**

Bolun Wang, Yuanshun Yao, Shawn Shan, **Huiying Li**, Bimal Viswanath, Haitao Zheng, and Ben Y. Zhao. “Neural Cleanse: Identifying and Mitigating Backdoor Attacks in Neural Networks.” In Proceedings of *The 40th IEEE Symposium on Security and Privacy.* San Francisco, CA, May 2019. **IEEE S&P'19**

Talks

Conference talk at USENIX Security 2022
Blacklight: Scalable Defense for Neural Networks against Query-Based Black-Box Attacks

Conference talk at ACM CCS 2019
Latent Backdoor Attacks on Deep Neural Networks

Invited talk at EE380 Stanford
Persistent and Unforgeable Watermarks for Deep Neural Networks