

# TINGHAO XIE

✉ [thx@princeton.edu](mailto:thx@princeton.edu) · 🔗 <https://tinghaoxie.com> · 🌐 vtu81

## 🎓 EDUCATION

**Princeton University**, Princeton, United States of America 08/2022 – Present

*Ph.D. student*, Electrical Computer Engineering (ECE)

- Advisor: **Prof. Prateek Mittal**

**Zhejiang University**, Zhejiang, China 09/2018 – 06/2022

*B.E.*, Computer Science and Technology (CS)

- **GPA**: 3.99/4.00 (92.07/100)
- **Rank**: 1st/186

**University of Oxford**, Oxford, United Kingdom 10/2021 – 12/2021

*Visiting Student*, Computer Science

- **GPA**: 4.00/4.00
- **Courses**: *Machine Learning, Computational Learning Theory*

## 💡 RESEARCH INTERESTS

Safe AI; Secure and Reliable AI Systems.

## 📖 PUBLICATIONS & MANUSCRIPTS

**Fine-tuning Aligned Language Models Compromises Safety, Even When Users Do Not Intend To!** 🔗, 📄, </>, 🖨️

Xiangyu Qi\*, Yi Zeng\*, **Tinghao Xie\***, Pin-Yu Chen, Ruoxi Jia, Prateek Mittal<sup>†</sup>, Peter Henderson<sup>†</sup>  
*Under Review*

*This work was exclusively reported by 🖨️ New York Times, and covered by many other social medias!*

**BaDExpert: Extracting Backdoor Functionality for Accurate Backdoor Input Detection** 📄, </>

**Tinghao Xie**, Xiangyu Qi, Ping He, Yiming Li, Jiachen T. Wang, Prateek Mittal  
*Under Review*

**Towards A Proactive ML Approach for Detecting Backdoor Poison Samples** 🔗, 📄, </>

Xiangyu Qi, **Tinghao Xie**, Jiachen T. Wang, Tong Wu, Saeed Mahloujifar, Prateek Mittal  
*USENIX Security Symposium 2023*

**Revisiting the Assumption of Latent Separability for Backdoor Defenses** 🔗, 📄, </>

Xiangyu Qi\*, **Tinghao Xie\***, Yiming Li, Saeed Mahloujifar, Prateek Mittal  
*ICLR 2023*

**Towards Practical Deployment-Stage Backdoor Attack on Deep Neural Networks** 🔗, 📄, </>

Xiangyu Qi\*, **Tinghao Xie\***, Ruizhe Pan, Jifeng Zhu, Yong Yang and Kai Bu  
*CVPR 2022 (Oral)*

## 👥 PREVIOUS RESEARCH EXPERIENCE

**Subnet Replacement Attack (SRA): A Graybox Backdoor Attack** 09/2021 – 11/2021

Advisors: Principal Researcher *Jifeng Zhu* (Tencent), Prof. **Kai Bu** (Zhejiang University)

Co-worker: Ph.d. Student **Xiangyu Qi** (Princeton University)

*Collaborator* with **Zhuque Lab, Tencent, China**

- Deployment-Stage Backdoor Attack on Deep Neural Networks (CVPR 22 Oral).

## Backdoor Restoration and Certification

05/2021 – 12/2021

Advisor: Prof. [Ting Wang](#) (Pennsylvania State University), Prof. [Shouling Ji](#) (Zhejiang University)

Remote Intern at **ALPS Lab**, Pennsylvania State University, United States

- Computing and tightening backdoor certification bounds ([%blog](#)).
- Faithfully restoring potential backdoor triggers ([%blog](#)).

## Enchecap: An Encrypted Heterogeneous Calculation Protocol

04/2020 – 05/2021

Advisor: Prof. [Jianhai Chen](#) (Zhejiang University)

Undergraduate Intern at **INCAS Lab**, Zhejiang University, China

- Designed `</>Enchecap`, a protocol securing heterogeneous computation for transmission and host memory.
- Implemented the protocol into a library, with 38% computational overhead and 19% overall overhead.

## ♡ HONORS AND AWARDS

---

Francis Robbins Upton Fellowship	08/2022
Outstanding Graduate Thesis	06/2022
Champion of Zhejiang University Bodybuilding Competition (70kg Level)	05/2022
Elite Liu Yongling Scholarship (1/802)	2020 – 2021
Tencent Scholarship (5/802)	2020 – 2021
The 2nd Class Prize in ASC 20-21 Student Supercomputer Challenge	01/2021
Narada Scholarship (1/372)	2019 – 2020
Champion of Zhejiang University DFM Hip-hop Crew Battle	2019

## 🏛️ SERVICES

---

Reviewers at Neurips 2023 Conference, Neurips 2023 Workshop BUGS

## ⚙️ SKILLS

---

- **Programming:** C/C++, Python, JavaScript, CUDA, Verilog, Shell, MATLAB, ActionScript, HTML.
- **Software:**  $\LaTeX$ , Vivado, Adobe {Photoshop, Premiere Pro, After Effects, Audition}.
- **Languages known:** English(fluent), Chinese(native), Cantonese(native).
- **Hobbies:** Climbing, Skiing, Dance(Hip-hop, House, etc.), Power Lifting, Swimming, Basketball, Billiards.