Yiyong **Liu**

INFORMATICS

Stuhlsatzenhaus 5, 66123 Saarbrücken, Germany

yiyong.liu@cispa.de | liu199604.github.io | Liu199604

“Live for die.”

Research Interest

- Trustworthy Machine Learning (Security, Privacy, and Robustness)
- Security and Privacy of Large Reasoning Models
- Safety and Alignment of AI Agents
- Adversarial Machine Learning

Education

CISPA & Saarland University *Saarbrücken, Germany*
PH.D. CANDIDATE IN COMPUTER SCIENCE *Oct. 2022 - Present*

CISPA & Saarland University *Saarbrücken, Germany*
PH.D. PREPARATORY PHASE IN COMPUTER SCIENCE *Apr. 2021 - Sep. 2022*

Zhejiang University *Hangzhou, China*
B.E. IN CONSTRUCTION ENGINEERING *Aug. 2015 - Jun. 2019*

Publication

[1] Membership inference attacks by exploiting loss trajectory

Yiyong Liu, Zhengyu Zhao, Michael Backes, Yang Zhang

In ACM SIGSAC Conference on Computer and Communications Security (CCS 2022).

[2] Auditing Membership Leakages of Multi-Exit Networks

Zheng Li, **Yiyong Liu**, Xinlei He, Ning Yu, Michael Backes, Yang Zhang

In ACM SIGSAC Conference on Computer and Communications Security (CCS 2022).

[3] SoK: Data Reconstruction Attacks Against Machine Learning Models: Definition, Metrics, and Benchmark

Rui Wen*, **Yiyong Liu***, Michael Backes, Yang Zhang (*Equal Contribution)

In USENIX Security Symposium (USENIX Security 2025).

Awards

2015	**Scholarship**, ZengXianzi Scholarship	*Zhejiang University*
2016	**Scholarship**, Third-Class Scholarship	*Zhejiang University*
2016	**Scholarship**, Excellent Student Scholarship	*Zhejiang University*
2017	**Scholarship**, Third-Class Scholarship	*Zhejiang University*

Related Skills

Computer Science	Data Structure, Algorithm Analysis, good at using pytorch, great understanding of classical models of machine learning
Programming language	c/c++, python