

SANGHYUN HONG

ASSISTANT PROFESSOR | COMPUTER SCIENCE | OREGON STATE UNIVERSITY

✉: sanghyun.hong@oregonstate.edu | 🏠: sanghyun-hong.com

RESEARCH SUMMARY

I am a computer scientist and educator dedicated to addressing threats to the trustworthiness and social responsibility of AI-enabled systems while fostering the development of the next-generation workforce capable of auditing and countering these risks. I received the Google Faculty Research Award 2023 and the Samsung Global Research Award 2023 and 2022. I am selected as a DARPA Riser (2022) and was invited as a speaker at USENIX Enigma (2021).

EDUCATION

University of Maryland, College Park, College Park, MD 09.2015 – 08.2021

Ph.D. in Computer Science

Dissertation: *Building Secure and Reliable Deep Learning from A System Security Perspective*

Academic advisor: [Prof. Tudor Dumitras](#)

M.S. in Computer Science

Scholarly paper: *Characterizing Program Behaviors in a Virtualized Infrastructure without Introspection*

Academic advisor: [Prof. Tudor Dumitras](#)

Seoul National University, Seoul, South Korea 03.2007 – 02.2015

B.S. in Electrical Engineering and Computer Science (*magna cum laude*)

Thesis: *A Power Saving Mechanism for the Smartphone Modem via Application-based Packet Piggybacking*

Academic adviser: [Prof. Seongsoo Hong](#)

EMPLOYMENT HISTORY

Oregon State University, Corvallis, OR 09.2021 – **Present**

Assistant Professor of Computer Science

Google Brain, Mountain View, CA 10.2020 – 01.2021

Research Intern in Privacy and Security Team

Host: [Dr. Nicholas Carlini](#) and [Dr. Alexey Kurakin](#)

Frame.io, New York, NY 11.2017 – 05.2018

Research Intern in Cloud Security Team

Host: [Dr. Abhinav Srivastava](#)

HONORS AND AWARDS

Honors

Google Faculty Research Award 2023

Samsung Global Research (GRO) Award 2023, 2022

DARPA Riser 2022

Future Faculty Fellow, University of Maryland, College Park 2020, 2021

Fellowships

Ann G. Wylie Dissertation Fellowship, University of Maryland, College Park 2020

KSEA-KUSCO Scholarships for Korean Graduate Students, KSEA 2017

Summer Research Fellowship, University of Maryland, College Park 2016

4-year Scholarship, Seoul National University Alumni Association (SNUAA) 2015

2-year Dean's Fellowship, University of Maryland, College Park 2015

SCHOLARLY AND PROFESSIONAL ACTIVITIES

PUBLICATIONS*

† indicates my advisees, ‡ denotes myself, and * indicates equal contributions.

Peer-Reviewed Conference Publications

- [C.1] [USENIX Sec’25] Seongho Keum, Dongwon Shin, †*Leo Marchyok, ‡Sanghyun Hong, and Soeul Son
Private Investigator: Extracting Personally Identifiable Information from Large Language Models Using Optimized Prompts.
[acceptance rate yet-unknown (cycle 2)]
- [C.2] [Interspeech’25] Seung-bin Kim, Hyun-seo Shin, Jungwoo Heo, Chan-yeong Lim, Kyo-Won Koo, Jisoo Son, ‡Sanghyun Hong, Souhwan Jung, Ha-Jin Yu
Enhancing Audio Deepfake Detection by Improving Representation Similarity of Bonafide Speech.
[acceptance rate 48.9%]
- [C.3] [IEEE S&P’25] Xuandong Zhao, Sam Gunn, Miranda Christ, Jaiden Fairoze, Andres Fabrega, Nicholas Carlini, Milad Nasr, ‡Sanghyun Hong, Florian Tramer, Sanjam Garg, Somesh Jha, Lei Li, Yu-Xiang Wang, and Dawn Song
SoK: Watermarking for AI-Generated Content.
[acceptance rate 15.1% (cycle 1)]
- [C.4] [ASIACCS’25] †*Eunjin Roh, *Sungwoo Jeon, Soeul Son, and ‡Sanghyun Hong
Evaluating Robustness of Reference-based Phishing Detectors.
[acceptance rate 10.3% (cycle 1)]
- [C.5] [NeurIPS’24] Yuxin Wen, †Leo Marchyok, ‡Sanghyun Hong, Jonas Geiping, Tom Goldstein, and Nicholas Carlini
Privacy Backdoors: Enhancing Membership Inference through Poisoning Pre-trained Models.
[acceptance rate 25.8%]
- [C.6] [ACSAC’24] Dongwon Shin, Suyoung Lee, ‡Sanghyun Hong, and Soeul Son
You Only Perturb Once: Breaking (Robust) Ad-Blockers with a Universal Adversarial Perturbation.
[acceptance rate 22.3%]
- [C.7] [CIKM’24] Fan Wu, Woojin Cho, †David Korotky, ‡Sanghyun Hong, Donsub Rim, Noseong Park and Kookjin Lee
Identifying Contemporaneous and Lagged Dependence Structures by Promoting Sparsity in Continuous-time Neural Networks.
[acceptance rate 23.2%]
- [C.8] [ICML’24] Victor Agostinelli III, ‡Sanghyun Hong, and Lihong Chen
Enabling Linear Transformers for Autoregressive and Simultaneous Tasks via Learned Proportions
[acceptance rate 27.5%]
- [C.9] [ICML’24][**Oral**] Woojin Cho, Minju Jo, Haksoo Lim, Kookjin Lee, Dongeun Lee, ‡Sanghyun Hong, and Noseong Park
Parameterized Physics-informed Neural Networks for Parameterized PDEs
[oral paper acceptance rate 1.5%]
- [C.10] [LREC-COLING’24] †Ojas Nimase and ‡Sanghyun Hong
When Do “More Contexts” Help with Sarcasm Recognition?
[acceptance rate 28.0%]

*My premier publication venues are [IEEE S&P, USENIX Security, ACM CCS] (security) and [ICML, ICLR, NeurIPs] (machine learning). These conferences are peer-reviewed and accept fewer than 10-30% of the total papers submitted.

- [C.11] [ICLR'24] Jinsung Jeon, Hyundong Jin, Jonghyun Choi, ‡Sanghyun Hong, Dongeun Lee, Kookjin Lee, and Noseong Park
Parallel-Structured All-Component Fourier Neural Operators for Recognizing Low-Quality Images
[acceptance rate 30.9%]
- [C.12] [AAAI'24] Woojin Cho, Seunghyeon Cho, Hyundong Jin, Jinsung Jeon, Kookjin Lee, ‡Sanghyun Hong, Dongeun Lee, Jonghyun Choi, and Noseong Park
Operator-learning-inspired Modeling of Neural Ordinary Differential Equations
[acceptance rate 23.8%]
- [C.13] [NeurIPS'23] †Zachery Coalson, Gabriel Ritter, Rakesh Bobba, and ‡Sanghyun Hong
BERT Lost Patience Won't Be Robust to Adversarial Slowdown
[acceptance rate 26.1%]
- [C.14] [ICML'23] Sicheng Zhu, Bang An, Furong Huang, and ‡Sanghyun Hong
Learning Unforeseen Robustness from Out-of-distribution Data Using Equivariant Domain Translator
[acceptance rate 28.0%]
- [C.15] [CHI'23] Sungbok Shin, ‡Sanghyun Hong, and Niklas Elmqvist
Perceptual Pat: A Virtual Human Visual System for Iterative Visualization Design
[acceptance rate 27.6%]
- [C.16] [SaTML'23] ‡Sanghyun Hong, Nicholas Carlini, and Alexey Kurakin
Publishing Efficient On-device Models Increases Adversarial Vulnerability
[acceptance rate 26.3%]
- [C.17] [NeurIPS'23][**Oral**] ‡Sanghyun Hong, Nicholas Carlini, and Alexey Kurakin
Handcrafted Backdoors in Deep Neural Networks
[oral paper acceptance rate 0.5%]
- [C.18] [CCS'23] Florian Tramèr, Reza Shokri, Ayrton San Joaquin, †Hoang Le, Matthew Jagielski, ‡Sanghyun Hong, and Nicholas Carlini
Truth Serum: Poisoning Machine Learning Models to Reveal Their Secrets
[acceptance rate 22.4% (track acceptance rate 8.9%); authors are in β - α order]
- [C.19] [VIS'22] Sungbok Shin, Sunghyo Chung, ‡Sanghyun Hong, and Niklas Elmqvist
A Scanner Deeply: Predicting Gaze Heatmaps on Visualizations Using Crowdsourced Eye Movement Data
[acceptance rate 24.7%]
- [C.20] [ISSTA'22] Geunwoo Kim, ‡Sanghyun Hong, Michael Franz, and Dokyung Song
XBA: Platform-Agnostic Binary Code Embedding Using Graph Convolutional Networks
[acceptance rate 27.2%]
- [C.21] [ICLR'22] Evani Radiya-Dixit, ‡Sanghyun Hong, Nicholas Carlini, and Florian Tramèr
Data Poisoning Won't Save You From Facial Recognition
[acceptance rate 32.9%]
- [C.22] [NeurIPS'21] ‡Sanghyun Hong, Michael-Andrei Panaitescu-Liess, Yiğitcan Kaya, and Tudor Dumitras
Qu-ANTI-zation: Exploiting Quantization Artifacts for Achieving Adversarial Outcomes
[acceptance rate 25.8%]
- [C.23] [ICLR'21][**Spotlight**] ‡Sanghyun Hong*, Yiğitcan Kaya*, Ionuț-Vlad Modoranu, and Tudor Dumitras
A Panda? No, It's a Sloth: Slowdown Attacks on Adaptive Multi-Exit Neural Network Inference
[spotlight paper acceptance rate 3.8%]

- [C.24] [ICICS'21] Bumjun Kwon, ‡Sanghyun Hong, Yuseok Jeon, and Doowon Kim
Certified Malware in South Korea: A Localized Study of Breaches of Trust in Code-Signing PKI Ecosystem
[acceptance rate 48.0%]
- [C.25] [ICLR'20] ‡Sanghyun Hong, Michael Davinroy, Yiğitcan Kaya, Dana Dachman-Soled and Tudor Dumitras
How to Own NAS in Your Spare Time
[acceptance rate 26.5%]
- [C.26] [USENIX Sec'19] ‡Sanghyun Hong, Pietro Frigo, Yiğitcan Kaya, Cristiano Giuffrida and Tudor Dumitras
Terminal Brain Damage: Exposing the Graceless Degradation in Deep Neural Networks Under Hardware Fault Attacks
[acceptance rate 15.5%]
- [C.27] [ICML'19] Yiğitcan Kaya, ‡Sanghyun Hong, and Tudor Dumitras
Shallow-Deep Networks: Understanding and Mitigating Network Overthinking
[acceptance rate 22.6%]
- [C.28] [BigData'18] ‡Sanghyun Hong, Noseong Park, Tanmoy Chakraborty, and Hyunjoong Kang
PAGE: Pattern-Query Answering via Knowledge Graph Embedding
[acceptance rate 19.7%]
- [C.29] [ICWS] Hyunjoong Kang, ‡Sanghyun Hong, Kookjin Lee, Noseong Park, and Soonhyun Kwon
On Integrating Knowledge Graph Embeddings into SPARQL Query Answering
[acceptance rate 20.0%]
- [C.30] [ACM HT'17] ‡Sanghyun Hong, Tanmoy Chakraborty, Sungjin Ahn, Ghaith Husari, and Noseong Park
SENA: Preserving Social Structure for Network Embedding
[acceptance rate 28.0%]

Peer-Reviewed Journal Articles

- [J.1] [JGR'24] Donsub Rim, Sanah Suri, ‡Sanghyun Hong, Kookjin Lee, and Randall J. LeVequestrong
A Stability Analysis of Neural Networks and Its Application to Tsunami Early Warning
Journal of Geophysical Research: Machine Learning and Computation, Volume 1, 2024.
- [J.2] [COSE'18] ‡Sanghyun Hong, Alina Nicolae, Abhinav Srivastava, and Tudor Dumitras
Peek-a-Boo: Inferring Program Behaviors in a Virtualized Infrastructure without Introspection
Computer & Security (COSE), 2018.

Workshop Papers

- [W.1] [NeurIPS RBFM'24] Eric Slyman, †Anirudh Kanneganti, ‡Sanghyun Hong, and Stefan Lee
You Never Know: Quantization Induces Inconsistent Biases in Vision-Language Foundation Models
- [W.2] [ICLR AI4DiffEqnsInSci'24] Woojin Cho, Minju Jo, Haksoo Lim, Kookjin Lee, Dongeun Lee, ‡Sanghyun Hong, and Noseong Park
Extension of Physics-informed Neural Networks to Solving Parameterized PDEs
- [W.3] [NeurIPS DistShift'23] Jaehoon Lee, Chan Kim, Gyumin Lee, Haksoo Lim, Jeongwhan Choi, Kookjin Lee, Dongeun Lee, ‡Sanghyun Hong, and Noseong Park
HyperNetwork Approximating Future Parameters for Time Series Forecasting under Temporal Drifts
- [W.4] [ICLR TrustworthyML'23] Sicheng Zhu, Bang An, Furong Huang, and ‡Sanghyun Hong
Learning Unforeseen Robustness from Out-of-distribution Data Using Equivariant Domain Translator
- [W.5] [WISP'22] Michael Curry, Byron Marshall, Forough Shadbad, and ‡Sanghyun Hong
Will SOC Telemetry Data Improve Predictive Models of User Riskiness? A Work in Progress
- [W.6] [ICML Con't time methods'22] Seunghyun Cho, ‡Sanghyun Hong, Kookjin Lee, and Noseong Park
AdamNODEs: When NODEs Meets Adaptive Moment Estimation

- [W.7] **[ISSTA Tool Demo'22]** Geunwoo Kim, ‡Sanghyun Hong, Michael Franz, and Dokyung Song
Improving Cross-Platform Binary Analysis Using Representation Learning via Graph Alignment
- [W.8] **[HotCloud'19]** ‡Sanghyun Hong, Abhinav Srivastava, William Shambrook, and Tudor Dumitraş
Go Serverless: Securing Cloud via Serverless Design Patterns
- [W.9] **[NeurIPS Crowdsourcing and ML'16]** Rock Stevens, Octavian Suci, Andrew Ruef,
‡Sanghyun Hong, Michael Hicks, and Tudor Dumitraş
Summoning Demons: The Pursuit of Exploitable Bugs in Machine Learning

Posters

- [P.1] **[NDSS'19 Poster]** ‡Sanghyun Hong, Tae-hoon Kim, Tudor Dumitraş, and Jonghyun Choi
Poster: On the Feasibility of Training Neural Networks with Visibly Watermarked Dataset

Preprints

- [M.1] **[arXiv'24]** ‡Sanghyun Hong, Nicholas Carlini, and Alexey Kurakin
Diffusion Denoising as a Certified Defense against Clean-label Poisoning Attacks
- [M.2] **[arXiv'21]** ‡Sanghyun Hong, Varun Chandrasekaran, Yiğitcan Kaya,
Tudor Dumitraş and Nicolas Papernot
On the Effectiveness of Mitigating Data Poisoning Attacks with Gradient Shaping
- [M.3] **[arXiv'20]** Yiğitcan Kaya, ‡Sanghyun Hong, and Tudor Dumitraş
On the Effectiveness of Regularization Against Membership Inference Attacks

INVITED TALKS

-
- | | |
|--|---------|
| [T.1] Pacific Northwest National Laboratory (PNNL) , Richland, WA
<i>A Sound Mind in A Vulnerable Body:
Practical Hardware Fault Attacks on DNNs</i> [Host: Dr. Bo Fang] | 03.2025 |
| [T.2] [Keynote] AI Security Summer Workshop , Gangwon-do, South Korea
<i>Importance in A Holistic Perspective in
Building Secure and Efficient AI Systems</i> [Host: Prof. Hyounghick Kim] | 07.2024 |
| [T.3] Korea University , Seoul, South Korea
<i>Great Haste Makes Great Waste:
Exploiting and Attacking Efficient Deep Learning</i> [Host: Prof. Sangkyun Lee] | 07.2024 |
| [T.4] KAIST , Daejeon, South Korea
<i>Great Haste Makes Great Waste:
Exploiting and Attacking Efficient Deep Learning</i> [Host: Prof. Yongdae Kim] | 07.2024 |
| [T.5] Soongsil University , Seoul, South Korea
<i>Great Haste Makes Great Waste:
Exploiting and Attacking Efficient Deep Learning</i> [Host: Prof. Souwhan Jung] | 06.2024 |
| [T.6] Hanyang University , Seoul, South Korea
<i>Great Haste Makes Great Waste:
Exploiting and Attacking Efficient Deep Learning</i> [Host: Prof. Seung-Hyun Seo] | 12.2023 |
| [T.7] OSU AI Seminar , Corvallis, OR
<i>Great Haste Makes Great Waste:
Exploiting and Attacking Efficient Deep Learning</i> [Host: Prof. Prasad Tadepalli] | 04.2023 |

†USENIX Enigma is a conference that focuses on communicating security ideas effectively through TED-style talks. The topics are selected competitively by a Program Committee, and the presentations undergo extensive peer-review and coaching following TED guidelines.

- [T.8] **TrustML Young Scientist Seminar**, Virtual Presentation 03.2023
*Great Haste Makes Great Waste:
Exploiting and Attacking Efficient Deep Learning* [Host: Dr. Jingfeng Zhang]
- [T.9] **KAIST**, Virtual Presentation 03.2023
A Systems Security Perspective for Building Trustworthy AI Systems [Host: Prof. Yongdae Kim]
- [T.10] **IEEE SaTML**, Raleigh, NC 02.2023
Publishing Efficient On-device Models Increases Adversarial Vulnerability
- [T.11] **[Oral] NeurIPS**, New Orleans, LA 12.2022
Handcrafted Backdoors in Deep Neural Networks
- [T.12] **IIT-Bombay**, Virtual Presentation 11.2022
A Systems Security Perspective for Building Trustworthy AI Systems [Host: Prof. Manoj Prabhakaran]
- [T.13] **DARPA Forward Conference**, Washington State University 09.2022
A Systems Security Perspective for Building Trustworthy AI Systems
- [T.14] **Samsung Advanced Institute of Research (SAIT)**, Virtual Presentation 07.2022
Building Secure and Reliable Deep Learning Systems from a Systems Security Perspective
- [T.15] **Tech Talk Tuesday**, Oregon State University 11.2021
Building Secure and Reliable Deep Learning Systems from a Systems Security Perspective
- [T.16] **Yonsei University**, Virtual Presentation 11.2021
*Qu-ANTI-zation: Exploiting Quantization Artifacts
for Achieving Adversarial Outcomes* [Host: Prof. Noseong Park]
- [T.17] **Oregon State University**, Eta Kappa Nu Induction Ceremony 11.2021
Lessons from My Journey Towards One-bit Flip [Host: Prof. Arun Natarajan]
- [T.18] **Cybersecurity Friday**, Oregon State University 10.2021
Why A Systems Security Perspective Matters in Adversarial Machine Learning?
- [T.19] **CMU**, Virtual Presentation 06.2021
*Building Secure and Reliable Deep Learning Systems
from a Systems Security Perspective* [Host: Prof. Lujo Bauer]
- [T.20] **[Spotlight] ICLR**, Virtual Conference Presentation 05.2021
A Panda? No, It's Sloth: Slowdown Attacks on Adaptive Multi-Exit Neural Network Inference
- [T.21] **Hardware.io**, Virtual Conference Presentation 02.2021
Practical Hardware Attacks on Deep Learning
- [T.22] **[TED Talk for Security] USENIX Enigma[†]**, Virtual Conference Presentation 02.2021
A Sound Mind in A Vulnerable Body: Practical Hardware Attacks on Deep Learning
- [T.23] **University of Tennessee**, Guest Lecturer 09.2020
Practical Hardware Attacks on Deep Learning [Host: Prof. Doowon Kim]
- [T.24] **ICLR**, Virtual Conference Presentation 05.2020
How to Own NAS in Your Spare Time
- [T.25] **USENIX Security Symposium**, Santa Clara, CA 08.2019
Terminal Brain Damage: Exposing the Graceless Degradation of DNNs under Hardware Fault Attacks
- [T.26] **Yonsei University**, South Korea 01.2019
*Can Machine Learning be Secure and Trustworthy
in the Presence of Micro-architectural Vulnerabilities?* [Host: Prof. Taekyung Kwon]
- [T.27] **USENIX Workshop on Hot Topics in Cloud Computing**, Boston, MA 07.2018
Go Serverless: Securing Cloud via Serverless Design Patterns

TEACHING AND MENTORING

TEACHING

Oregon State University, Corvallis, OR

Assistant Professor

CS 499/579: Cyber-security (Course website)	Spring 2025
CS 499/599: Trustworthy Machine Learning (Course website)	Winter 2025
CS 499/599: Trustworthy Machine Learning (Course website)	Fall 2023
[*Now this course is an elective in OSU's AI program, equivalent to AI 539]	
CS 370 : Introduction to Security (Course website)	Spring 2023
CS 499/599: Trustworthy Machine Learning (Course website)	Spring 2023
CS 344 : Operating Systems I (Course website)	Winter 2023
CS 344 : Operating Systems I	Spring 2022
CS 499/599: Machine Learning Security (Course website)	Winter 2022

STUDENTS ADVISING

PhD Students

Derek Lilienthal (AI)	2024 – Present
Jose Aguilar Escamilla (AI, co-advised with Huazheng Wang)	2023 – Present
Zachary Coalson (CS)	2025 – Present
Eunjin Roh (CS)	2025 – Present

Masters Students

Gabriel Ritter (CS, co-advised with Rakesh Bobba)	2021 – Present
---	----------------

Undergraduate Students

Aiden Gabriel (CS)	2025 – Present
Ajinkya Vijay Gokule (CS)	2023 – Present
Leo Marchyok (CS)	2022 – Present

Thesis Committee

Oluwatomi Hassan (MS, CS, OSU)	2024 – Present
Amelia Kawasaki (PhD, AI, OSU)	2023 – Present
John Conner Zontos (PhD, AI, OSU)	2023 – Present
Allen Yan (PhD, AI, OSU)	2023 – Present
Apurva Dilip Kokate (PhD, AI, OSU)	2022 – Present
Jarrod Jeffrey Isaac Hollis (PhD, CS, OSU)	2021 – Present
Avani Abhay Sathe (MS, CS, OSU)	2023 – Present
Opeyemi Ajibuwa (MS, CS, OSU)	2023 – Present
Sungbok Shin (PhD, CS, UMD) → PostDoc at Inria, France	2024
Philiph Lee (MS, CS, OSU)	2022
Ryan Kennedy (MS, CS, OSU)	2022

Alumni (and the 1st employment)

Eunjin Roh (MS, CS) → PhD student in my group at OSU	2023 – 2025
Tahmid Hasan Prato (MS, CS)	2023 – 2025

Ramya Jayaraman (MS, AI) → CoCreator, Inc.	2023 – 2024
Lucas Bell (MS, CS, co-advised with Dr. Yeongjin Jang)	2023 – 2024
Jonathan Keller (MS, CS, co-advised with Dr. Yeongjin Jang)	2023 – 2024
Hoang Le (MS, CS) → Promaxo, Inc.	2021 – 2023
Zachary Coalson (BS, CS) → PhD student in my group at OSU	2022 – 2025
Evan Mrazik (BS, CS) → NetSPI	2022 – 2024
Peter Mora-Stevens (BS, CS) → Uber, Inc.	2022
Ryan Little (BS, CS) → PhD student at UMD	2022

STUDENTS MENTORED

Research Interns at Maryland Cybersecurity Centers (MC2)

Michi Panaitescu	2020 – 2021
Ionuț-Vlad Modoranu	2019 – 2021
Khasmamad Shabanovi	2020
Stefan Matcovici	2019
Michael Davinroy	2018 – 2019
Alina-Elena Nicolae	2016

Undergraduate Research Interns from NSF-REU Program

Michael Davinroy	2018 - 2019
Stuart Nevans Locke	Summer 2018
Ian Rackow	Summer 2018
Kevin Kulda	Summer 2018

SERVICES

PROFESSIONAL ACTIVITIES

Program Chairs

- DSN-DSML**: Workshop on Dependable and Secure Machine Learning 2024
- RAID**: International Symposium on Research in Attacks, Intrusions and Defenses 2022

Area Chairs

- [§]**NeurIPS**: Conference on Neural Information Processing Systems 2025

Technical Program Committees

- [§]**IEEE S&P**: IEEE Symposium on Security and Privacy 2026
- [§]**USENIX**: USENIX Security Symposium 2023–**Present**
- [§]**CCS**: ACM Conference on Computer and Communications Security 2023–**Present**
- [§]**PETS**: Privacy Enhancing Technologies Symposium 2024–**Present**
- NDSS**: Network and Distributed System Security Symposium 2024
- RAID**: International Symposium on Research in Attacks, Intrusions and Defenses 2021 – 2023
- AISeC**: ACM Workshop on Artificial Intelligence and Security 2021 – 2024
- DLSP**: Deep Learning Security and Privacy Workshop 2021 – 2023, 2025
- WPES**: ACM Workshop on Privacy in the Electronic Society 2021
- ICLR TML**: Workshop on Towards Trustworthy ML: Rethinking Security and Privacy for ML 2020

Action Editor

- [§]**TMLR**: Transaction on Machine Learning Research 2023 - **Present**

Reviewer

- [§]**ICML**: International Conference on Machine Learning 2020 – **Present**
- [§]**ICLR**: International Conference on Learning Representation 2021 – **Present**
- [§]**NeurIPS**: Conference on Neural Information Processing Systems 2020 – **Present**
- ACL**: Annual Meeting of the Association for Computational Linguistics 2023
- TOPS**: ACM Transactions on Privacy and Security 2021
- IEEE Access**: IEEE Access 2021
- COSE**: Computer & Security 2019
- TCC**: IEEE Transaction on Cloud Computing 2018

[§]Indicates the main communities in which I actively contribute and serve.