Sebastian Szyller

CONTACT INFORMATION

Sebastian Adam Szyller
Helsinki, Finland

contact@sebszyller.com
sebszyller.com
GitHub , LinkedIn , Twitter

EDUCATION

PhD, Security and Privacy of Machine Learning **Mar. 2020 – Aug. 2023**
Aalto University, Espoo, Finland

- Thesis: *Ownership and Confidentiality in Machine Learning.*
- Supervisor: N. Asokan.
- Committee: Sébastien Gambs (opp.), Alina Oprea, Bimal Viswanath.
- Finnish AI Society Best Dissertaion Award & Aalto University Best Dissertation Award.

M.Sc., Machine Learning and Data Mining **Sept. 2017 – Feb. 2020**
Aalto University, Espoo, Finland

- Thesis: *Adversary Detection in Online Machine Learning Systems.*
- Supervisor: N. Asokan.
- Participant in the *Doctoral Track*. Graduation with honours.

***Erasmus+* Student Exchange, Computer Science** **Jan. 2016 – June 2016**
Turku University of Applied Sciences, Turku, Finland

B.Sc., Computer Science **Sept. 2013 – Mar. 2017**
Lodz University of Technology, Lodz, Poland

- Thesis: *Decryption of Feistel Network Based Algorithms Using Machine Learning.*
- Supervisor: Laurent Babout.

EMPLOYMENT

Research Scientist – Machine Learning Security **June 2023 – Now**
Intel Labs, Helsinki, Finland / Remote (Portland, USA)

- Research on privacy, robustness and provenance in (generative) machine learning.
- Security of RAG-enabled, and multimodal systems.
- Academic collaborations and outreach.

Doctoral Researcher **Mar. 2020 – June 2023**
Aalto University, Secure Systems Group, Espoo, Finland

- Research on model and data confidentiality in machine learning.

Visiting Researcher **Oct. 2022 – Nov. 2022**
University of Waterloo, CrySP, Waterloo, Canada

- Research on empirical metrics for membership privacy.

Research Assistant **Sept. 2017 – Feb. 2020**
Aalto University, Secure Systems Group, Espoo, Finland

- Research on model ownership and watermarking in machine learning.

Machine Learning and Security Research Intern **July 2019 – Sept. 2019**
Huawei Technologies, Helsinki, Finland

- Applied research on federated learning and differential privacy.

Big Data and Scala Software Engineer **June 2016 – July 2017**
GFT Poland, Lodz, Poland

- Development of a Scala-based microservice architecture to facilitate data ingestion.
- Implementation and optimization of various financial instruments on Apache Spark.

SAP/ABAP Developer Intern **July 2015 – Sept. 2015**
Accenture Lodz Delivery Center, Lodz, Poland

- Development of web services for processing WSDLs and handling SOAPUI.

Quality Assurance Tester **July 2014 – Oct. 2014**
Exact Systems, Lodz, Poland

- Quality assurance and reporting of mechanical defects for various electronical appliances.

Professional Interests

I work on topics at the intersection of machine learning and security/privacy such as: model ownership and provenance, adversarial robustness, and differential privacy.
I'm interested in different ways that we can protect ML models and data to enable secure and trustworthy analysis – both in terms of the technical details as well as legislation compliance.

Technical Skills

- Machine Learning, Federated Learning, Algorithm Design and Analysis, Security & Privacy of Machine Learning, Threat Modeling
- Python, common machine learning and analysis libraries (PyTorch, SciPy stack)
- Scala and big data tools (Apache Spark, NiFi, Hive/Impala, mongoDB)
- JavaScript (Node, Svelte), Rust, C/C++, SQL

Technical Reports

2. V. Duddu, **S. Szyller**, N. Asokan *SHAPr: an Efficient and Versatile Membership Privacy Risk Metric for Machine Learning* (arXiv:2112.02230)
1. **S. Szyller**, V. Duddu, T. Gröndahl, N. Asokan *Good Artists Copy, Great Artists Steal: Model Extraction Attacks Against Image Translation Models* (arXiv:2104.12623)

Published Research

9. V. Duddu, **S. Szyller**, N. Asokan *SoK: Unintended Interactions Among Machine Learning Defenses and Risks* IEEE S&P 2024 (Distinguished Paper Award) (arXiv:2312.04542)
8. R. Zhang, J. Liu, **S. Szyller**, K. Ren, N. Asokan *False Claims Against Model Ownership Resolution* USENIX 2024 (arXiv:2304.06607)
7. M. Phute, A. Helbling, M. Hull, S. Peng, **S. Szyller**, C. Cornelius, D. H. Chau *LLM Self Defense: by Self Examination, LLMs Know They Are Being Tricked* ICLR 2024, Tiny Papers (arXiv:2308.07308)
6. **S. Szyller**, R. Zhang, J. Liu, N. Asokan *On the Robustness of Dataset Inference.* TMLR 2023 (arXiv:2210.13631)
5. **S. Szyller** & N. Asokan *Conflicting Interactions Among Protection Mechanisms for Machine Learning Models.* AAAI 2023 (Spotlight) (arXiv:2207.01991)
4. **S. Szyller**, B. Atli, M. Marchal, N. Asokan *DAWN: Dynamic Adversarial Watermarking of Neural Networks.* ACM MM 2021 (arXiv:1906.00830)
3. B. Atli, **S. Szyller**, M. Juuti, M. Marchal, N. Asokan *Extraction of Complex DNN Models: Real Threat or Boogeyman?* AAAI 2020, EDSMLS (arXiv:1910.05429)
2. M. Marchal & **S. Szyller** *Detecting eCommerce Fraud Using Scalable Categorical Clustering.* ACSAC 2019 (arXiv:1910.04514)
1. M. Juuti, **S. Szyller**, M. Marchal, N. Asokan *PRADA: Protecting Against DNN Model Extraction Attacks.* IEEE EuroS&P 2019 (arXiv:1805.02628)

Presentations & Talks

7. *Whose Model Is It Anyway? Ownership and Intellectual Property in the Era of Machine Learning*, **Lodz University of Technology**, September 2024
6. *False Claims Against Model Ownership Resolution*, **USENIX Security**, August 2024
5. *Conflicting Interactions Among Protection Mechanisms for Machine Learning Models*, **AAAI**, February 2023
4. *Conflicting Interactions Among Protection Mechanisms for Machine Learning Models*, **Vector Institute**, November 2022
3. *Conflicting Interactions Among Protection Mechanisms for Machine Learning Models*, **University of Waterloo**, October 2022
2. *DAWN: Dynamic Adversarial Watermarking of Neural Networks*, **ACM Multimedia**, October 2021
1. *Extraction of Complex DNN Models: Real Threat or Boogeyman?*, **Cluster of Excellence for Cyber Security**, November 2020

Teaching Experience

Co-lecturer, Aalto University

· Research Seminar on Security & Privacy of Machine Learning **Spring 2022**

· Research Seminar on Security & Privacy of Machine Learning **Spring 2021**

Teaching Assistant, Aalto University

· Research Seminar on Security & Privacy of Machine Learning **Fall 2019**

Advised Students

MSc Thesis, Vasisht Duddu (University of Waterloo) **2022**

· *Towards Effective Measurement of Membership Privacy Risk for Machine Learning Models*

MSc Thesis, Eleonora Micozzi (Aalto University) **2021**

· *Guarantees of Differential Privacy in Over-parameterised Models*

Internship, Chloe Sham and Dmytro Shynkevych (University of Waterloo) **2021**

· Implementation of model evasion, differential privacy and watermarking schemes.

Service & Outreach

Programme Committees

· WWW'22, AAAI'23'24, IEEE S&P'24

Member of C2PA **Jan. 2024 – Now**
Technical Working Group, Intel Representative

Member of the Board, Aalto Debating Society **Jan. 2021 – Dec. 2021**
Aalto University, Espoo, Finland

Member of the Student Government **Oct. 2014 – Dec. 2016**
Lodz University of Technology, Lodz, Poland

· Vice-president since March 2015.

Member of the Scholarship Committee **Oct. 2014 – Jan. 2016**
Lodz University of Technology, Lodz, Poland

Languages

Polish, native speaker
English, native proficiency, C2 (CPE)
German and Spanish, communicative, A2
Finnish, actively learning, A2

INTERESTS Film & street photography, mechanical keyboards, industrial design, bouldering, fixed-gear cycling.