

# Nils Lukas

Assistant Professor • MBZUAI • Abu Dhabi, UAE

[nils.lukas@mbzuai.ac.ae](mailto:nils.lukas@mbzuai.ac.ae) • [nilslukas.github.io](https://nilslukas.github.io)

Updated on July 19, 2024

<b>Research Interests</b>	Design safe and reliable Machine Learning systems in the presence of untrustworthy	
	<ol style="list-style-type: none"><li>1. <b>Providers:</b> Confidential computing via Homomorphic Encryption &amp; Secret Sharing.</li><li>2. <b>Data:</b> Mitigate data poisoning during training &amp; prompt injection during inference.</li><li>3. <b>Models:</b> Protect training data privacy through PII scrubbing &amp; differential privacy.</li><li>4. <b>Users:</b> Control misuse by detecting generated (mis)information with watermarking.</li></ol>	
<b>Education</b>	<b>University of Waterloo</b> , Canada	2019 - 02/2024
	Ph.D. in Computer Science <ul style="list-style-type: none"><li>▪ Advisor: <a href="#">Florian Kerschbaum</a></li><li>▪ Thesis: <a href="#">Analyzing Threats of Large-Scale Machine Learning Systems</a></li><li>▪ Awarded the <a href="#">Mathematics Doctoral Prize's Top Honour</a> 🏆</li></ul>	
	<b>RWTH-Aachen</b> , Germany	
	M.Sc. in Computer Science ( <i>w/Distinction</i> )	2016 - 2018
	B.Sc. in Computer Science	10/2012 - 2016
<b>Honors &amp; Awards</b>	<b>Mathematics Doctoral Prize</b> , <a href="#">University of Waterloo</a> [1 500 CAD]	2024
	<b>Best Poster Award</b> , <a href="#">Sponsored by David R. Cheriton</a> [300 CAD]	2023
	<b>Distinguished Contribution Award</b> , Microsoft MLADS conference	2023
	<b>David R. Cheriton Scholarship</b> , <a href="#">University of Waterloo</a> [20 000 CAD]	2022, 2023
	<b>Outstanding Reviewer</b> , ICML'22	2022
	<b>Best Poster Award</b> , <a href="#">Sponsored by Rogers</a> [1 000 CAD]	2019
	<b>KU Global Scholarship</b> , <a href="#">Korea University</a> [1.2 million KRW]	2016
	<b>MOGAM Scholarship</b> , <a href="#">RWTH-Aachen</a> [3 000 EUR]	2014
<b>Conference Publications</b>	<b>[USENIX'24]</b> <a href="#">PEPSI: Practically Efficient Private Set Intersection in the Unbalanced Setting</a> Rasoul Mahdavi, <b>Nils Lukas</b> , Faezeh Ebrahimiaghazani, Thomas Humphries, Bailey Kacsmar, John Premkumar, Xinda Li, Simon Oya, Ehsan Amjadian, Florian Kerschbaum. In the 33rd USENIX Security Symposium, 2024.	
	<b>[USENIX'24]</b> <a href="#">Fast and Private Inference of Deep Neural Networks by Co-designing Activation Functions</a> Abdulrahman Diaa, Lucas Fenaux, Thomas Humphries, Marian Dietz, Faezeh Ebrahimiaghazani, Bailey Kacsmar, Xinda Li, <b>Nils Lukas</b> , Rasoul Akhavan Mahdavi, Simon Oya, Ehsan Amjadian, Florian Kerschbaum. In the 33rd USENIX Security Symposium, 2024.	
	<b>[ICLR'24]</b> <a href="#">Leveraging Optimization for Adaptive Attacks on Image Watermarks</a> AR: 30.8% (2 250/7 262) <b>Nils Lukas</b> , Abdulrahman Diaa, Lucas Fenaux, Florian Kerschbaum. In the Twelfth International Conference on Learning Representations, 2024.	
	<b>[ICLR'24]</b> <a href="#">Universal Backdoor Attacks</a> AR: 30.8% (2 250/7 262)  <b>Media Coverage</b> Benjamin Schneider, <b>Nils Lukas</b> , Florian Kerschbaum. In the Twelfth International Conference on Learning Representations, 2024.	
	<b>[USENIX'23]</b> <a href="#">PTW: Pivotal Tuning Watermarking for Pre-Trained Image Generators</a> AR: 29.2% (422/1 444) <b>Nils Lukas</b> and Florian Kerschbaum. In the 32nd USENIX Security Symposium, 2023.	
	<b>[S&amp;P'23]</b> <a href="#">Analyzing Leakage of Personally Identifiable Information in Language Models</a> AR: 17.0% (195/1 147)  <b>Distinguished Contribution Award at Microsoft MLADS</b> <b>Nils Lukas</b> , Ahmed Salem, Robert Sim, Shruti Tople, Lukas Wutschitz, Santiago Zanella-Béguelin. In the 44th IEEE Symposium on Security and Privacy, 2023.	

	<b>[S&amp;P'22]</b> AR: 14.5% (147/1012)	<a href="#">SoK: How Robust is Image Classification Deep Neural Network Watermarking?</a> <b>Nils Lukas</b> , Edward Jiang, Xinda Li, Florian Kerschbaum. In the 43rd IEEE Symposium on Security and Privacy, 2022.
	<b>[ICLR'21]</b> AR: 28.7% (860/2997) 🏆 <b>Spotlight (Top 5%)</b>	<a href="#">Deep Neural Network Fingerprinting by Conferrable Adversarial Examples</a> <b>Nils Lukas</b> , Yuxuan Zhang, Florian Kerschbaum. The Ninth International Conference on Learning Representations, 2021.
	<b>[IH&amp;MMSEC'21]</b> AR: 40.3% (128/318)	<a href="#">On the Robustness of Backdoor-based Watermarking in Deep Neural Networks</a> Masoumeh Shafieinejad, <b>Nils Lukas</b> , Jiaqi Wang, Xinda Li, Florian Kerschbaum. Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security, 2021.
	<b>[ACSAC'20]</b> AR: 20.9% (104/497)	<a href="#">Practical Over-Threshold Multi-Party Private Set Intersection</a> Rasoul Mahdavi, Thomas Humphries, Bailey Kacsmar, Simeon Krastnikov, <b>Nils Lukas</b> , John Premkumar, Masoumeh Shafieinejad, Simon Oya, Florian Kerschbaum, Erik-Oliver Blass. Annual Computer Security Applications Conference (ACSAC), 2020.
	<b>[EuroS&amp;P'20]</b> AR: 20.9% (39/187)	<a href="#">Differentially Private Two-Party Set Operations</a> Bailey Kacsmar, Basit Khurram, <b>Nils Lukas</b> , Alexander Norton, Masoumeh Shafieinejad, Zhiwei Shang, Yaser Baseri, Maryam Sepehri, Simon Oya, Florian Kerschbaum. IEEE European Symposium on Security and Privacy (EuroS&P), 2020.
<b>Journal Publications</b>	<b>[AIP'18]</b>	<a href="#">SunFlower: A new Solar Tower Simulation Method for use in Field Layout Optimization</a> , Pascal Richter, Gregor Heimig, <b>Nils Lukas</b> , Martin Frank. AIP Conference Proceedings, Volume 2033, Issue 1, 2018.
<b>Working Papers</b>		<a href="#">Pick your Poison: Undetectability versus Robustness in Data Poisoning Attacks against Deep Image Classifiers</a> <b>Nils Lukas</b> and Florian Kerschbaum.
<b>Work Experience</b>	<b>Assistant Professor</b> , MBZUAI, Abu Dhabi, UAE <b>Research Intern</b> , Royal Bank of Canada, Borealis AI, Toronto ▪ Vertical Federated Learning, hosted by <a href="#">Kevin Wilson</a> <b>Research Intern</b> , Microsoft Research, Cambridge, UK ▪ Privacy for Language Models, hosted by <a href="#">Shruti Tople</a> & <a href="#">Lukas Wutschitz</a> <b>Research Assistant</b> , RWTH-Aachen, Aachen <b>Student Researcher</b> , DSA Daten- und Systemtechnik GmbH, Aachen <b>Software Engineer Intern</b> , A.R. Bayer DSP Systeme GmbH, Düsseldorf	from 08/2024 2024 2022 2014 - 2018 2016 2012
<b>Teaching</b>	<b>Teaching Assistant</b> , University of Waterloo ▪ CS458/658: Computer Security and Privacy ▪ CS246 - Object Oriented Programming <b>Co-Instructor</b> , RWTH-Aachen ▪ Course: Data-driven Medicine	2020, 2021 2021 2018
<b>Research Talks</b>	<b>Analyzing Leakage of Personal Information in Language Models</b> ▪ Microsoft M365, hosted by <a href="#">Robert Sim</a> ▪ Meta, hosted by Will Bullock ▪ MongoDB, hosted by <a href="#">Marilyn George</a> and <a href="#">Archita Agarwal</a> <b>How Reliable is Watermarking for Image Generators?</b> ▪ Google, hosted by <a href="#">Somesh Jha</a> ▪ University of California, Berkely, hosted by <a href="#">Dawn Song</a>	2024 2023 2023 2023 2023

Service	<b>Program Committee</b>	
	▪ IEEE Symposium on Security and Privacy (S&P)	2025
	▪ Recent Advances in Intrusion Detection (RAID)	2024
	<b>Artifact Evaluation Committee</b>	
	▪ The ACM Conference on Computer and Communications Security (CCS)	2023, 2024
	<b>Reviewer</b>	
	▪ International Conference on Learning Representations (ICLR)	2024
	▪ International World Wide Web Conference (TheWebConf)	2024
	▪ Recent Advances in Intrusion Detection (RAID)	2023
	▪ Neural Information Processing Systems (NeurIPS)	2022, 2023
	▪ International Conference on Machine Learning (ICML)	2022
	▪ The Conference on Information and Knowledge Management (CIKM)	2020
	<b>Other</b>	
	▪ <b>Sub-Reviewer</b> , Proceedings on Privacy Enhancing Technologies (PETS)	2021, 2022, 2023
	▪ <b>Session Chair</b> , IEEE Symposium on Security and Privacy (S&P)	2023
	▪ <b>Organizing Hackathon</b> , Workshop on Semantic Web Solutions for Large-Scale Biomedical Data Analytics (SeWeBMeDA)	2018
	<b>Student Board Member</b> , Cybersecurity and Privacy Institute	2022, 2023, 2024
	<b>School Advisory Committee on Appointments Liaison</b> , CrySP Lab	2022