Lucas Fenaux

GitHub | Scholar | Linkedin | Website | lucas.fenaux.apply@gmail.com

Summary

I am 3rd year PhD candidate at the University of Waterloo researching Adversarial Machine Learning and Machine Learning Security and Privacy topics under the supervision of Florian Kerschbaum. My research focuses on developing secure and private machine learning systems.

My research interests: Adversarial ML · ML Security · Federated Learning Security & Privacy · Watermarking · World Model Security.

Publications

Peer-Reviewed

Fenaux, Lucas, Christopher Srinivasa, and Florian Kerschbaum (2025). "On the Trade-Off Between Transparency and Security in Adversarial Machine Learning". In: *25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026)*. link.

Diaa, Abdulrahman*, **Lucas Fenaux***, Thomas Humphries*, et al. (2024). "Fast and private inference of deep neural networks by co-designing activation functions". In: *33rd USENIX Security Symposium (USENIX Security 24)*. **(* Co-first-authorship)**. link.

Fenaux, Lucas, Thomas Humphries, and Florian Kerschbaum (2023). "Gaggle: Genetic algorithms on the GPU using pytorch". In: *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*. link.

Lukas, Nils, Abdulrahman Diaa, **Lucas Fenaux**, and Florian Kerschbaum (2023). "Leveraging Optimization for Adaptive Attacks on Image Watermarks". In: *The Twelfth International Conference on Learning Representations*. link.

Preprints

Fenaux, Lucas, Zheng Wang, Jacob Yan, et al. (2026). "Hammer and Anvil: Toward a Theory of Backdoors in Federated Learning". In: *arXiv preprint arXiv:2509.08089*. Under submission at ACM CCS 2026. link.

Fenaux, Lucas*, Larris Xie*, Aditya Bang, et al. (2026). "Private Vertical Federated Inference for Time-Series". In: *arXiv preprint arXiv:2605.08343*. **(* Co-first-authorship)**. Under submission at NeurIPS 2026. link.

Education

2024 - present PhD (Computer Science) at **University of Waterloo** (CrySP Lab) (GPA: 98.0/100.0)

2021 - 2023 Master's Degree at **University of Waterloo** (CrySP Lab) (GPA: 93.5/100.0)
Thesis: "Analyzing Adversarial Examples: A Framework to Study Adversary Knowledge"

2017 - 2021 Honours Bachelor of Science at **University of Toronto** (GPA: 3.69/4.0)

Work Experience

Machine Learning Security Scientist May 2023 - Aug 2023

Royal Bank of Canada

- Developed LLM security guardrails and evaluation benchmarks for RBC's internal LLM gateway, improving detection of toxicity, prompt injection, and code or PII leakage.
- Designed and implemented a prototype Vertical Federated Learning inference pipeline.

Operations Developer May 2019 - Aug 2019

D1g1t Inc.

- Worked on data and infrastructure management, maintenance, and monitoring.
- Built financial reports for Wealth Management Firms using Windows SSRS (SQL) and API Usage Reports using AWS Quicksights.

Additional Research Experience

Reviewer

Conference on Neural Information Processing Systems (NeurIPS) 2026
International Conference on Learning Representations (ICLR) 2026

Research Assistant to Dr. Miriam Diamond May 2020 - May 2021

University of Toronto, SCDMS-SNOLAB

Developed supervised and unsupervised ML models to calibrate dark matter detectors (SuperCDMS); engineered feature representations in collaboration with physicists.

Research Assistant to Dr. John Percy Sep 2018 - Mar 2019

University of Toronto

Co-authored a paper on variable-star analysis using ASAS-SN data (JAAVSO).

Awards & Scholarships (Total $ Amount)

2024 - 2027	NSERC Postgraduate Scholarships – Doctoral (PGS D)	($120,000)
2024	Ontario Graduate Scholarship (OGS) [Declined]	($15,000)
2021 - 2023	David R. Cheriton Graduate Scholarship (University of Waterloo)	($20,000)
2020	Cross Disciplinary Internship Grant, Arthur B. McDonald Canadian Astro-particle Physics Research Institute	($10,000)
2017	President's Scholar of Excellence Scholarship (University of Toronto)	($10,000)

Projects & Miscellaneous

Poker AIs. poker
A pure Python-based implementation of various poker AI algorithms with a distributed training architecture. Currently contains PPO-based self-play models, with ongoing efforts to add CFR-based algorithms.

Gaggle: Genetic algorithms on the GPU using PyTorch. torch-gaggle
A Python package built on top of PyTorch to speed up genetic algorithm-based deep learning optimization, providing speedups of up to $\mathbf{9.7}\times$ over existing packages for deep learning applications.

RL for Network Dismantling Attacks and Defenses. network-dismantling
Research project on using Reinforcement Learning to improve network dismantling attacks. Also proposed a new kind of defense against network dismantling attacks using Reinforcement Learning. Built using PyTorch Geometric, NetworkX, and custom RL agents trained with DQN.

GSEF Board of Directors Member, Math Faculty Representative 2022 - present
University of Waterloo, Graduate School Endowment Fund (GSEF)
Serving as the Math Faculty representative on the GSEF Board, contributing to governance, funding allocation, and strategic planning for graduate student support.

Teaching Experience

Undergraduate Research Assistant Supervision
Larris Xie, Aditya Bang, Alex Zhang, Jacob Yan, George Wang, Nathan Chung.
Worked on research projects that are under-submission (preprints) .

Teaching Assistant

Winter 2023	CS459/659: Privacy, Cryptography, Network and Data Security
Summer 2022	CS458/658: Computer Security and Privacy

Skills

ML & Security	Adversarial Attacks & Defenses, Backdoors, Federated Learning, Generative AI, Watermarking, Reinforcement Learning, Diffusion Models, World Models.
ML Systems	GPU-accelerated training with PyTorch, distributed training with Ray, experience with large-scale experiments.
Frameworks	PyTorch, Torchvision, PyTorch Geometric, HuggingFace, Pandas, NumPy, Ray.
Programming	Python (expert), C, Java, Bash.
Language skills	Native French speaker, Fluent in English, B2/C1 proficiency in German.