Mohammad Yaghini

mohammad.yaghini@mail.utoronto.ca
myaghini
m-yaghini
m-yaghini.github.io

PhD Student in Machine Learning

Education

Sept.2020 – Present **University of Toronto & Vector Institute**, *Ph.D. in Machine Learning*, Canada, CleverHans Lab (under the supervision of Prof. Nicolas Papernot)
Committee Members: *Nisarg Shah*, *Aleksandar Nikolov*, *Nicolas Papernot*

Sept.2017 – Oct.2019 **École Polytechnique Fédérale de Lausanne (EPFL)**, *Master's in Data Science*, School of Computer and Communication Sciences, Switzerland
Thesis: A Human-in-the-loop Framework to Construct Context-dependent Mathematical Formulations of Fairness

2011–2016 **Isfahan University of Technology (IUT)**, *B.Sc. in Electrical Engineering – Communications*, Iran
Thesis: An Energy-Efficient Cooperative Mechanism for Device-to-Device Communications

Awards and Honors

Feb.2022 Received the **2022 Meta PhD Research Fellowship** in Security and Privacy

Sept.2021 Received the 2021 Schwartz Reisman Institute for Technology and Society **Graduate Fellowship**

2019-2020 Declined **Ph.D. scholarships** from MPI-SWS (Saarbrücken), UCL (London), and NUS (Singapore)

2016 Declined **direct Ph.D. scholarships** from University of Michigan (Ann Arbor), University of Pennsylvania, and Virginia Tech (Blacksburg)

2011 Ranked in the **top 0.3% (99.6 percentile)** among 252,000 participants in the Nationwide University Entrance Exam, also known as *Concours* (Math-Physics)

Publications

Conference Proceedings

* Joint 1st author

† Equal Contribution

Ali Shahin Shamsabadi*, **M. Yaghini***, Natalie Dullerud*, Sierra Wyllie, Ulrich Aïvodji, Aisha Alaagib, Sébastien Gambs, and Nicolas Papernot. Washing The Unwashable : On The (Im)possibility of Fairwashing Detection. In *NeurIPS 2022*.

Bogdan Kulynych, **M. Yaghini**, Giovanni Cherubin, Michael Veale, and Carmela Troncoso. Disparate Vulnerability to Membership Inference Attacks. In *Privacy Enhancing Technologies (PETs) 2022*.

M. Yaghini, Andreas Krause, and Hoda Heidari. A Human-in-the-loop Framework to Construct Context-aware Mathematical Notions of Outcome Fairness. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES) 2021*.

Hengrui Jia*, **M. Yaghini***, Christopher A. Choquette-Choo, Natalie Dullerud, Anvith Thudi, Varun Chandrasekaran, and Nicolas Papernot. Proof-of-Learning: Definitions and Practice. In *IEEE Symposium on Security and Privacy (S&P) 2021*.

Pratyush Maini, **M. Yaghini**, and Nicolas Papernot. Dataset Inference: Ownership Resolution in Machine Learning. In *ICLR 2021*.

Naman Goel, **M. Yaghini**, and Boi Faltings. Non-Discriminatory Machine Learning Through Convex Fairness Criteria. In *AAAI Conference on Artificial Intelligence (AAAI) 2018*.

Mehdi Naderi Soorki, **M. Yaghini**, Mohammad Hossein Manshaei, Walid Saad, and Hossein Saidi. Energy-aware optimization and mechanism design for cellular device-to-device local area networks. In *Conference on Information Science and Systems (CISS) 2016*.

Workshops

M. Yaghini, Patty Liu, Franziska Boenisch, and Nicolas Papernot. Regulation Games for Trustworthy Machine Learning. In *NeurIPS Workshop on Regulatable ML 2023*.

M. Yaghini, Patty Liu, Franziska Boenisch, and Nicolas Papernot. Learning to Walk Impartially on the Pareto Frontier of Fairness, Privacy, and Utility. In *NeurIPS Workshop on Regulatable ML 2023*.

Selected Pre-Prints

M. Yaghini, Patty Liu, Franziska Boenisch, and Nicolas Papernot. Regulation Games for Trustworthy Machine Learning. *CoRR*, abs/2402.03540, 2024.

Adam Dziedzic*, Stephan Rabanser*, **M. Yaghini***, Armin Ale, Murat A. Erdogdu, and Nicolas Papernot. p-DkNN: Out-of-Distribution Detection Through Statistical Testing of Deep Representations. *CoRR*, abs/2207.12545, 2022.

Varun Chandrasekaran†, Hengrui Jia†, Anvith Thudi†, Adelin Travers†, **M. Yaghini**†, and Nicolas Papernot. SoK: Machine Learning Governance. *CoRR*, abs/2109.10870, 2021.

Experience

Research Assistant

Sep.2020–Present **Nicolas Papernot**, *CleverHans Lab*, UoT & Vector Institute
- Game Theoretic Modeling of ML Governance
- Intellectual Property of ML Models
- Privacy
- Algorithmic Fairness

June.2023–Sept.2023 **Florian Tramèr**, *Secure and Private AI (SPY) Lab*, ETH Zurich
- Systematic Canary Design for Auditing Differential Privacy Guarantees

March.2020–Sep.2020 **Reza Shokri**, *Privacy and Trust Group*, NUS (remote)
- Human-in-the-loop Explainable ML

Mar.2019–August.2019 **Andreas Krause**, *Learning and Adaptive Systems (LAS)*, ETH Zurich
- Master thesis on context-dependent mathematical formulations of fairness

Oct.2017–Dec.2019 **Carmela Troncoso**, *Security and Privacy Engineering Laboratory (SPRING)*, EPFL
- Quantifying privacy vulnerability and its disparity for ML models, defenses, and the trade-offs

Feb.2018–Jun.2018 **Robert West**, *Data Science Lab (DLAB)*, EPFL
- Designing mechanisms for truthful judgment aggregation to detect misinformation

Feb.2017–Aug.2017 **Boi Faltings**, *Artificial Intelligence Laboratory (LIA)*, EPFL
- Building a convex fairness metric for classifiers

Sep.2014–Aug.2016 **MohammadHossein Manshaei**, *Game Theory & Mechanism Design Group*, IUT
- Designing a game-theoretic mechanism to incentivize device-to-device communication for 5G networks

Teaching Experience

Course Instructor

Fall 2022 **ECE421 Introduction to Machine Learning**

Selected Teaching Assistantships

Fall 2021 **ECE1784/CSC2559 Trustworthy Machine Learning**, *Nicolas Papernot*, Graduate seminar assistant

Jun-Dec 2021 **ECE421 Introduction to Machine Learning**, *Nicolas Papernot*, Course development & Head TA

Fall '15, '16 **(Graduate) Game Theory**, *MohammadHossein Manshaei*, Homework design and problem solving

Academic Service

IEEE SatML 2023, IEEE S&P 2023, *Program Committee*

ICML 2024, NeurIPS 2023, JMLR, NeurIPS Workshop on Privacy in ML 2021, *Reviewer*

NeurIPS 2021, USENIX Security 2021, IEEE S&P (2022), *External Reviewer*

Industry Experience

Jun.2022–Sept.2022 **Microsoft Research**, *Privacy Research Intern*, Cambridge, UK (Remote)
- Analysis and empirical estimation of differential privacy trade-off curves for machine learning

References

- **Nicolas Papernot**, Assistant Professor, University of Toronto *nicolas.papernot@utoronto.ca*
- **Carmela Troncoso**, Assistant Professor, EPFL *carmela.troncoso@epfl.ch*