Mohammad Yaghini

✉ *mohammad.yaghini@mail.utoronto.ca*
in *myaghini*
m-yaghini
⌂ *m-yaghini.github.io*

PhD Student in ML, SRI Graduate Fellow

Education

Sept.2020 – Present **Ph.D. in Machine Learning**, *University of Toronto and Vector Institute*, Canada, CleverHans Lab (under the supervision of Prof. Nicolas Papernot)

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

Sept.2016 – Aug.2017 **Master's in Communication Systems**, *School of Computer and Communication Sciences*, École Polytechnique Fédérale de Lausanne, Switzerland – switched to Data Science in the 2nd year.

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

Publications

* Joint 1st author

† Equal Contribution

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

Bogdan Kulynych, **M. Yaghini**, Giovanni Cherubin, and Carmela Troncoso. Disparate Vulnerability: on the Unfairness of Privacy Attacks Against Machine Learning. In *22nd Privacy Enhancing Technologies Symposium (2022)*.

M. Yaghini, Andreas Krause, and Hoda Heidari. A Human-in-the-loop Framework to Construct Context-aware Mathematical Notions of Outcome Fairness. In *AIES '21: AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event, USA, May 19-21, 2021*, pages 1023–1033. ACM, 2021.

Hengrui Jia*, **M. Yaghini***, Christopher A. Choquette-Choo, Natalie Dullerud, Anvith Thudi, Varun Chandrasekaran, and Nicolas Papernot. Proof-of-Learning: Definitions and Practice. *42nd IEEE Symposium on Security and Privacy*, May 2021.

Pratyush Maini, **M. Yaghini**, and Nicolas Papernot. Dataset Inference: Ownership Resolution in Machine Learning. In *Proceedings of the 2021 International Conference on Learning Representations (ICLR 2021)*, May 2021.

Naman Goel, **M. Yaghini**, and Boi Faltings. Non-Discriminatory Machine Learning Through Convex Fairness Criteria. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18)*, pages 3029–3036, 2018.

Experience

Research Assistant

Sep.2020– Present **CleverHans Lab**, UoT/Vector Institute
- Trustworthy Machine Learning/Model Governance
- Intellectual Property of ML Models
- ML Security for Audio Domain

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

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

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

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

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

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

Academic Service

Nov.2021 **Journal of Machine Learning Research (JMLR)**, *Reviewer*

Aug.2021 **NeurIPS 2021 Workshop Privacy in Machine Learning**, *Reviewer*

Jul.2021 **NeurIPS 2021**, *External Reviewer*

Feb.2021 **USENIX Security 2021**, *External Reviewer*

Jan.2021 **IEEE Security and Privacy 2022**, *External Reviewer*

Industry Experience

Sept.2018– Feb.2019 **Expedia**, *Junior Data Scientist*, Geneva
- Building statistical models for advanced time-series forecasting using Spark

Voluntary Work

July.2017– June.2018 **EPFL Iranian Student Association (IRSA)**, *Public Relations*, Lausanne
- Moderating bi-weekly intellectual discussions on society, culture, technology, psychology, etc.

2013–2016 **IEEE IUT Student Branch**, *Active Member and Vice Chair in 2014*, Isfahan

Awards and Honors

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

2019-2020 Received **Ph.D.** offers from UoT/Vector Institute (Toronto, CA), EPFL (Lausanne, CH), MPI-SWS (Saarbrücken, DE), UCL (London, UK), and NUS (Singapore, SG)

2016 Received **Direct-Ph.D.** offers from University of Michigan (Ann Arbor, US), University of Pennsylvania, and Virginia Tech (Blacksburg, US)

2016 Received **Master's** offers from EPFL (Lausanne, CH), ETHZ (Zurich, CH), University of British Columbia (Vancouver, Canada)

2011–2016 Received **Gifted Student Award** (Sept. 2011) and **Merit-based admission** to MSc program in Communication Systems (Dec. 2014), Isfahan University of Technology

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)

Related Course Work

- Machine Learning
- Deep Learning
- Game Theory & Evolutionary Games
- Algorithms for Private Data Analysis
- Convex Optimization
- Statistical Learning Theory
- Information Theory & Signal Processing
- Algorithms for Collective Decision Making
- Data Visualization
- Statistics for data science

Computer Skills

Machine Lear.	Scikit, Pandas, Spark MLib, XGBoost	Languages	**Python**, Scala, Julia, MATLAB, Java, Javascript/Typescript, C
Deep Lear.	PyTorch, Keras	Big Data	Spark, Hive SQL, Kafka/SparkStreaming
Data Vis.	Plotly, D3.js, Matplotlib	Optimization	CVX, CVXOPT
Web Dev.	JS/TS, HTML, CSS, React	NLP	NLTK, Gensim

Languages and Test Scores

Persian Native proficiency

French Full proficiency (DELF B2: 76.5/100)

Turkish Speaking proficiency

English Full proficiency

References

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