Edoardo **Debenedetti**

PhD Student in CS @ ETH Zürich

(+41) 76 699 43 27 | edebenedetti@inf.ethz.ch | https://edoardo.science | dedeswim | 0000-0003-3343-9477 | Edoardo Debenedetti

Education

ETH Zürich - Federal Institute of Technology Zürich — *Zürich, Switzerland*

PhD in Computer Science — *Aug. 2022 - 2026 (exp.)*

- Focus: **Real-world adversarial machine learning**, advised by **Prof. Florian Tramèr**.
- Fully funded by the **CYD Doctoral Fellowship**, awarded by the Armasuisse Cyber-Defense Campus.

EPFL - Federal Institute of Technology Lausanne — *Lausanne, Switzerland*

MSc in Computer Science — *Sep. 2019 - Apr. 2022*

- **GPA 5.63/6**, focus on **Machine Learning** ∩ **Security** ∩ **Privacy**.
- Master's Thesis about the **adversarial robustness of Vision Transformers** supervised by **Princeton University**'s **Prof. Mittal**.

Politecnico di Torino — *Turin, Italy*

BSc in Computer Engineering — *Sep. 2016 - Jul. 2019*

- **GPA 28.4/30**, graduation mark 110/110, **top 9%**.
- **Exchange year at 同济大学** (Tongji University), in Shanghai (China), supported by a **full scholarship** granted to the top 31% applicants.

Experience

Bloomberg LP — *London, United Kingdom*

Software Engineering Intern — *Jul. 2021 - Sep. 2021*

- Worked in the **Multi Asset Risk System** team, on the re-design and implementation of the configuration of a distributed logging library.
- Move the configuration of a **distributed logging library** from an internal technology to a **centralized SQL DB**, using a **cache** and a **C++ service**.
- The configuration is checked **~1M times per minute**, and the usage of the cache gave a **~23x speed improvement** w.r.t. querying the DB.

armasuisse Cyber-Defence Campus — *Lausanne, Switzerland*

Research Intern — *Aug. 2020 - Jan. 2021*

- Worked on **Machine Unlearning** and **Membership Inference Attacks** against Generative Models, supervised by **Prof. Mathias Humbert**.
- Adapt the **MIA** technique proposed by the *GAN-Leaks* work (by Chen et al.), to work after the removal some datapoints from the training set.
- The technique achieved **promising results** when attacking DCGAN trained on the CelebA dataset

Conference papers

- **Debenedetti, E.**, Sehwag, V., Mittal, P., *"A Light Recipe to Train Robust Vision Transformers"*, First IEEE Conference on Secure and Trustworthy Machine Learning, February 2023.
- Croce*, F., Andriushchenko*, M., Sehwag*, V., **Debenedetti*, E.**, Flammarion, N., Chiang, M., Mittal, P., Hein, M., *"RobustBench: a standardized adversarial robustness benchmark"*, Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track, 2021. *(* equal contribution)*. A preliminary version appeared at the ICLR 2021 Workshop on Security and Safety in ML Systems.

Manuscripts and workshop papers

- **Debenedetti, E.**, Severi, G., Carlini, N., Choquette-Choo, C. A., Jagielski, M., Nasr, M., Wallace, E., Tramèr, F., *"Privacy Side Channels in Machine Learning Systems"*, arXiv ePrint 2309.05610.
- **Debenedetti, E.**, Carlini, N., Tramèr, F., *"Evading Black-box Classifiers Without Breaking Eggs"*, 2nd ICML Workshop on New Frontiers in Adversarial Machine Learning, 2023. **Oral presentation**.

Honors and Awards

2023 **CYD Doctoral Fellowship,** full PhD funding for 4 years, worth **USD 516'000** (CHF 461'000), from Armasuisse CYD Campus and EPFL.

2021 **Google TPU Research Cloud Program,** extensive **hardware support for 8 months** to work on the Master's Thesis.

2021 **Best Paper Honorable Mention Prize,** ICLR Workshop on Security and Safety in ML Systems. **Top 2 out of 50** accepted papers.

Service

Reviewer

- **NeurIPS Datasets and Benchmarks Track**: 2022, 2023
- **CCS AISec workshop**: 2023

Open Source Maintainer

- **RobustBench**: adversarial robustness benchmarking library and model zoo.
  - More than 150 models spanning 3 datasets and 3 threat models.
  - **517 stars**, with 379 unique cloners in 2 weeks (measured in September 2023).
  - Refactored the code to improve the extensibility of the library.

  Repository at `https://github.com/RobustBench/robustbench`.

Conference service

- **Competition organizer at SaTML 2024**: organizing the *Large Language Models Capture-the-Flag*.
- **Volunteer at NeurIPS 2021**: helped with monitoring the website and technical issues.