Edoardo **Debenedetti**

PhD Student in CS @ ETH Zürich

(+39) 340 512 6541 | edoardo.m.debenedetti@gmail.com | edoardo.science | dedeswim | edoardo-debenedetti | 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**.

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 at 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

Reply — *Turin, Italy*

Software Engineering Intern — *Nov. 2018 - Feb. 2019*

- Developed a **chatbot** that answers questions about GDPR law, using **TypeScript**, **Redis**, **MongoDB**, **IBM Watson Assistant**, and **Docker**.
- Worked on **RPA**, using **Python**. One of the bots **decreased a task duration by 88%,** without requiring human intervention in it.

Publication

- 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. **131 citations**. *(* equal contribution)*. A preliminary version appeared at the ICLR 2021 Workshop on Security and Safety in ML Systems.

Honors and Awards

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.

Projects

RobustBench: a standardized adversarial robustness benchmark — *EPFL, Graded 6/6*

- **Benchmark and leaderboard** to analyze the robustness of image classifiers, and Model Zoo containing 60+ **PyTorch** models trained robustly.
- **Analysis** on the progress in robustness: worked on the Lipschitzness of robust models, and on a new ensemble black-box transfer attack, capable of **improving the success rate by up to 70%** w.r.t. transfer attacks carried with one model only.

Counteracting DeepFakes — *EPFL, Graded 5.75/6*

- Term **Research Project** run at **Prof. Troncoso**'s SPRING Lab at EPFL, worked on **poisoning** attacks via **Influence Functions** against DeepFakes.
- Tested the attack on MNIST-trained **autoencoder** which showed **major distortions** in its output after the poisoned training.

Leadership and Services

NeurIPS — *Virtual*

Voulunteer — *Dec. 2021*

- Volunteer at NeurIPS 2021 to help with monitoring the website and technical issues.

LeadTheFuture

Mentee — *Sept. 2019 - Current*

- Selected to be part of the leading mentorship organization for STEM students in Italy - Acceptance rate < 20%.
- Held **mentoring sessions** about MSc admissions at EPFL.

JEToP (PoliTo's Junior Enterprise) — *Turin, Italy*

International Manager — *Oct. 2018 - Jun. 2019*

- **Executive board** member responsible for external relations of the Junior Enterprise. Signed **8 new partnerships.**