# Ilias Tsingenopoulos

Postdoctoral Researcher at KU Leuven
Leuven, Belgium, 3000

itsiggen@gmail.com
ilias.tsingenopoulos@cs.kuleuven.be
LinkedIn: linkedin.com/in/itsiggen
+32 486 319 227


## About me

I am a postdoctoral researcher specializing in the intersection of Artificial Intelligence and Computer Security, with a proven track record in Adversarial Machine Learning and Reinforcement Learning. My research spans the theoretical and practical aspects of adversarial attacks and defenses across a broad range of AI systems and modalities: from circumventing web bot detection systems such as Google reCaptcha, to hardening commercial antivirus against adversarial malware, and more generally on adaptively optimizing attacks and defenses against each other under the competitive game they form.

I investigate the fundamentals of robust learning and its adversarial and counterfactual aspects, essential components towards achieving trustworthy and safe AI. As human decision-making increasingly transitions to automated, I am actively developing new techniques and methodologies for training models resilient to evasion and other failure modes. Currently, I focus on safeguarding LLM generation through the integration of alignment, guardrails, and other safety mechanisms to ensure safe and correct outputs in real-world deployments.

## Professional & Research Experience

**DistriNet, KU Leuven** — Leuven, Belgium
*Postdoctoral Researcher in AI Safety and Security* — *October 2024 – Present*
- Exploring principled approaches for safe LLM generation.

**DistriNet, KU Leuven** — Leuven, Belgium
*Doctoral Researcher in Adversarial Machine Learning (AML)* — *May 2019 – September 2024*
- Adversarial attacks and defenses on AI models.

**S2Lab, University College London** — London, UK
*Visiting Scholar* — *January, 2023 – April 2023*
- Rendering a commercial antivirus robust to evasive malware.

**DistriNet, KU Leuven** — Leuven, Belgium
*Research Assistant* — *May, 2018 – April 2019*
- Preventing abusive DNS registrations in the .eu domain.
- Optimizing black-box attack strategies.

**Information Technologies Institute, CERTH** — Thessaloniki, Greece
*Research Assistant* — *January, 2017 – April 2018*
- Formulating and writing research proposals.
- Research and implementation in several Horizon 2020 projects.

## Education

**Aristotle University of Thessaloniki** — Thessaloniki, Greece
*M.Eng in Electrical and Computer Engineering - 7.54/10* — *Sep. 2006 – Jul. 2016*
*Dissertation: Fuzzy Clustering Algorithms in Feature Subspaces*

**Technical University of Berlin** — Berlin, Germany
*Erasmus in Electrical Engineering and Computer Science Department* — *Feb. 2012 – Aug. 2012*

**High School** — Thessaloniki, Greece
*Arsakeio Thessalonikis – Final GPA: 19.3/20* — *Sep. 2002 – Jun 2005*

## Selected Publications

- **I. Tsingenopoulos**, J. Cortellazzi, B. Bosansky, S. Aonzo, D. Preuveneers, W. Joosen, F. Pierazzi, and L. Cavallaro. "How to Train your Antivirus: RL-based Hardening through the Problem-Space". In: 27th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2024).
- **I. Tsingenopoulos**, V. Rimmer, D. Preuveneers, F. Pierazzi, L. Cavallaro, and W. Joosen. "On Adaptive Decision-Based Attacks and Defenses". In: DLSP Workshop, IEEE S&P 2024.
- **I. Tsingenopoulos**, D. Preuveneers, L. Desmet, and W. Joosen. "Captcha me if you can: Imitation Games with Reinforcement Learning". In: 2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P 2022).
- **I. Tsingenopoulos**, A. M. Shafiei, L. Desmet, D. Preuveneers, and W. Joosen. "Adaptive Malware Control: Decision-Based Attacks in the Problem Space of Dynamic Analysis". In: 1st Workshop on Robust Malware Analysis (WoRMA 2022).
- C. J. Hernández-Castro, Z. Liu, A. Serban, **I. Tsingenopoulos**, and W. Joosen. "Adversarial Machine Learning". In: Security and Artificial Intelligence: A Crossdisciplinary Approach. Springer, 2022.

## Knowledge & Technical Skills

**Domain Expertise:**

- **Mathematical**: Linear Algebra, Non-convex/Derivative-free Optimization
- **Machine Learning**: Clustering, SVMs, Decision Trees/Random Forests, XGBoost
- **Deep Learning**: Convolutional/Graph/Recurrent Neural Networks, GANs
- **Reinforcement Learning**: Q-Learning, Policy Optimization, Multi-agent
- **Adversarial ML**: White/Black-box Attacks, Adversarial Training

**Development and Tools:**

- **Programming Languages**: Python, C, C++, Java, Matlab
- **Deep Learning**: PyTorch, Tensorflow, Keras
- **Reinforcement Learning**: Gym/Gymnasium, Stable Baselines, RLLib
- **Scientific Computing Libraries**: Numpy, Scipy, Pandas, Scikit-learn
- **Others**: Git, LaTex

## Summer Schools & Online Courses

- **LLM Agents MOOC Fall 2024**: (UC Berkeley)
- **Deep Reinforcement Learning**: CS 285 Fall 2021 (UC Berkeley)
- **M2L 2023**: Mediterranean Machine Learning Summer School
- **Security and Privacy in the Age of AI**: Organized and participated in editions 2022, 2023, 2024

## Languages

**Greek**: Native | **English**: Excellent - C2 | **German**: Very Good - B2/C1 | **Dutch**: Good - B1

## Other Interests

A curious and restless spirit, in my free time I enjoy science fiction, creative writing, and tabletop/computer games; I am also a competitive capoeirista and water polo player.