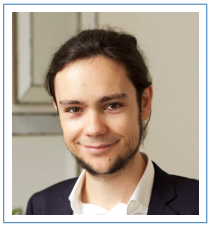
Dr. Martin Gubri

Research Lead

Tübingen, Germany
✉ martin[at]gubri[dot]eu
gubri.eu
21 Oct. 1991



Education

2019–2023 **PhD Degree**, *SnT, University of Luxembourg, Luxembourg*.
PhD thesis in adversarial machine learning titled "What Matters in Model Training to Transfer Adversarial Examples". Supervised by Prof. Dr. Yves Le Traon.

2014–2015 **Specialized Master, Data Science**, *Ensae ParisTech*, *with high honor*.
Courses on both theoretical & applied aspects of Machine Learning, Data Analysis & Econometrics with a emphasis on Computer Science & Big Data Processing and Analysis.

2012–2014 **Master's Degree, Statistics and Econometrics**, *Toulouse School of Economics*, *with highest honor (head of the class, $1^{st}/28$)*.
Courses on Statistics, Data Analysis, Econometrics, Applied Mathematics & Programming.

Since 2013 **MOOC**, *Coursera*.
Bayesian Methods for ML, Neural Networks and DL, Improving Deep Neural Networks, Structuring ML Projects, Computing for Data Analysis, Data Analysis, Computer Networks.

2011–2014 **Magister Degree, Economics and Statistics**, *Toulouse School of Economics & Paul Sabatier University*, *with highest honor ($3^{rd}/16$)*.

2011–2012 **Bachelor's Degree, Economics and Mathematics**, *Toulouse School of Economics*.

2009–2011 **Classe préparatoire aux grandes écoles**, *Lycée Montaigne*, Bordeaux.
Khâgne & Hypokhâgne B/L, Humanities, Social Science & Mathematics.

Work and volunteer experiences

From Sep. 2023 **Research Lead**, *NT Parameter Lab GmbH*, Tübingen, Germany.
Research interests at the intersectin of AI trustworthiness and large language models, including adversarial machine learning, privacy, alignment, and uncertainty. Supervised by Prof. Dr. Seong Joon Oh.

Sep. 2019 — Aug. 2023 **Doctoral Researcher**, *SnT, University of Luxembourg, Luxembourg*.
Doctoral researcher with a working contract. Research interests in adversarial machine learning, with a focus on the transferability of adversarial examples.

Oct. 2017 — Aug. 2019 **Freelance Data Scientist & Independent Researcher**, *France*.
Freelance working on Machine Learning, Data Analysis, Sampling design, Data Visualization, Big Data & Web Scraping. Independent Researcher on ML Security between contracts.

Since Jan. 2017 **Information Security contributions**.
26 XSS, RCEs on TeleMath server, etc. See full list of vulnerabilities on gubri.eu.

Oct. 2015 — Oct. 2016 **Ford–Mozilla Technology Exchange Fellow**, *ONG Derechos Digitales*, Chile.
Developer, Statistician and Technology Referent for various projects and reports about Computer Security, Privacy, Personal Data Protection, Cryptography, Net Neutrality & Censorship.

Jun. 2015 — Aug. 2015 **R package development**, *Google Summer of Code for the R foundation*.
Implementation of state-of-the-art predictors for Spatial Regression Models in *spdep*, the main R package for Spatial Statistics. Mentored by Roger Bivand & Giovanni Millo.

Jun. 2014 — Apr. 2015 **Junior Researcher in Applied Spatial Econometrics**, *BIwhere consulting*, Paris.
Development of new methods to improve predictions using spatial interaction models for a business use case (intern for 3 months and employee later).

Since 2012 **Member of the board of directors of Framasoft**.
Major French-speaking FLOSS promoting non-profit organization. Co-president in 2017.

Publications

1. Ulmer, D., Gubri, M., Lee, H., Yun, S., and Oh, S. J. Calibrating large language models using their generations only. In *ACL 2024* (2024) (details)
2. Gubri, M., Ulmer, D., Lee, H., Yun, S., and Oh, S. J. Trap: Targeted random adversarial prompt honeypot for black-box identification. In *ACL 2024 (findings)* (2024) (details)
3. Kim, S., Yun, S., Lee, H., Gubri, M., Yoon, S., and Oh, S. J. ProPILE: Probing privacy leakage in large language models. In *NeurIPS 2023 (spotlight)* (details)
4. Gubri, M. *What Matters in Model Training to Transfer Adversarial Examples*. PhD thesis, Unilu - University of Luxembourg, Luxembourg, June 2023 (details)
5. Gubri, M., Cordy, M., and Traon, Y. L. Going Further: Flatness at the Rescue of Early Stopping for Adversarial Example Transferability (pdf)
6. Gubri, M., Cordy, M., Papadakis, M., Traon, Y. L., and Sen, K. LGV: Boosting Adversarial Example Transferability from Large Geometric Vicinity. In *ECCV 2022* (2022) (pdf, video, poster, code)
7. Gubri, M., Cordy, M., Papadakis, M., and Traon, Y. L. Efficient and Transferable Adversarial Examples from Bayesian Neural Networks. In *UAI 2022* (2022) (pdf, poster, code)
8. Franci, A., Cordy, M., Gubri, M., Papadakis, M., and Traon, Y. L. Influence-Driven Data Poisoning in Graph-Based Semi-Supervised Classifiers. *CAIN 2022* (2022), 77–87 (pdf)
9. Ghamizi, S., Cordy, M., Gubri, M., Papadakis, M., Boystov, A., Le Traon, Y., and Goujon, A. Search-Based Adversarial Testing and Improvement of Constrained Credit Scoring Systems. In *FSE 2020* (2020), ESEC/FSE 2020 (doi, pdf, code)
10. Gubri, M. Adversarial Perturbation Intensity Achieving Chosen Intra-Technique Transferability Level for Logistic Regression. *arXiv:1801.01953 [cs, stat]* (2018) (pdf, code)

Computer skills

Significant Contributions	Torchattacks, Adversarial Robustness Toolbox (ART), spdep R package, Common Voice
Programming	Python, R (incl. tidyverse), Bash, Html/Css, JS
Deep Learning	PyTorch, TensorFlow (basis), Keras (basis)
ML/Stats	Scikit-learn, Statsmodels, R Stats Package, Spdep, Rpart, Caret, *i.a.*
Database	MySQL, PostgreSQL, SQLite, MongoDB, DynamoDB, Oracle
VCS	Git
OS	GNU/Linux, MacOS, Windows
Office	LaTeX, LibreOffice, MS Office
Miscellaneous	Matplotlib, Seaborn, Scrapy, Flask, Bottle, Pig, Hdfs

Linguistic skills

English	Fluent (TOEFL iBT Score in 2019: 101/120)
French	Native language

Spanish	Fluent
German	Basic

Interests

FLOSS, ML, Security, Statistics, Privacy, Climbing, Skiing