

Valentin Lafargue

PhD student at [IMT](#), [INRIA](#), [ANITI 2](#), [IRIT](#)

valentin.lafargue@math.univ-toulouse.fr [Google Scholar](#)

PROFILE

PhD student at IMT, INRIA, ANITI 2 and IRIT focused on AI evaluation and auditing, with research on fairness, robustness and bias auditing in machine learning and generative AI. Passionate about interdisciplinary work, exploring how AI intersects with fields like Law, Sociology, Cultural Analytics, Psychology, and Education.

PUBLICATIONS

- [1] **Probing Cultural Signals in Large Language Models through Author Profiling**
Valentin Lafargue, Ariel Guerra-Adames, Emmanuelle Claeys, Elouan Vuichard, Jean-Michel Loubes.
Empirical Methods in Natural Language Processing (EMNLP) 2026. Conference paper.
- [2] **Exposing the Illusion of Fairness: Auditing Vulnerabilities to Distributional Manipulation Attacks**
Valentin Lafargue, Adriana Laurindo Monteiro, Emmanuelle Claeys, Laurent Risser, Jean-Michel Loubes.
ECML PKDD Applied Data Science Track 2026. Conference paper.
- [3] **Fairness is in the details: Face Dataset Auditing**
Valentin Lafargue, Emmanuelle Claeys, Jean-Michel Loubes.
ECML PKDD Applied Data Science Track 2025. Conference paper.
- [4] **Orthogonium: A Unified, Efficient Library of Orthogonal and 1-Lipschitz Building Blocks**
Thibaut Boissin, Franck Mamalet, **Valentin Lafargue**, Mathieu Serrurier.
Championing Open-source DEvelopment in ML Workshop @ ICML25. Workshop paper.

EDUCATION

2024 – present	PhD in Mathematics Disloyalty and Biases in Generative AI <i>IMT, INRIA, ANITI 2, IRIT</i> – Supervised by J.-M. Loubes and E. Claeys – Research on fairness, robustness and bias auditing in ML and generative AI – Teaching duties: Lecture, project/TER supervision, TDs & TP
2023 – 2024	Master 2 (MSc) , Data Science and Data Engineering <i>Université Paul Sabatier, Toulouse</i> Result: 16.91/20 (rank 1/37) Deep Learning, NLP & SVACS, advanced statistics, parallel databases and cloud computing, business management, NoSQL, ethical communication, law for data.
2022 – 2023	Master 1 , Data Science and Data Engineering <i>Université Paul Sabatier, Toulouse</i> Result: 16.26/20 (rank 1/40)

Statistics, optimization, time series, machine learning, data warehousing, data mining, advanced algorithmics, advanced Python.

2019 – 2022

Bachelor, Mathematics, Higher Education and Research

Université Paul Sabatier, Toulouse

Results: Year 1: 16.52/20 (rank 4/164); Year 2: 16.27/20 (rank 5/157); Year 3: 14.64/20 (rank 3/31)

EXPERIENCE

04/2024 – 09/2024

Research Internship: Robust Neural Network Architectures

IRT Saint Exupéry, Toulouse, France

- Implemented and benchmarked 1-Lipschitz and orthogonal dense/convolutional layers in a shared GitHub environment.
- Studied robustness to noise and adversarial attacks, and the benefits of 1-Lipschitz networks across MLP, CNN, GAN, AE and VAE architectures
- Researched latent-space smoothness and interpolation; built a toy time-series dataset for quantitative evaluation and proved theoretical advantages of Lipschitzness, confirmed experimentally
- Explored invertible neural networks and the pullback operator via gradients

04/2023 – 07/2023

Research Internship: Explainability & Fairness in Machine Learning

ANITI, Toulouse, France

- Built prediction models on tabular data (logistic regression, CART, GBM) and image classification with custom CNNs and fine-tuning
- Conducted explainability and fairness analysis of the resulting models; simulated econometric data to study bias
- Developed an interactive Shiny-Live Python interface to make fairness criteria accessible to non-experts
- Explored new approaches independently, an early introduction to research

2021 – 2022

Other Experience (pre-university)

Summer roles outside academia

Agricultural Technician, approved by the French Ministry of Agriculture, SPSMQ, Montauban (06/2022–08/2022) – field inspection and producer liaison. Commercial Employee, Intermarché Super, Saint-Paul-sur-Save (05/2021–08/2021) – stock management and customer service.

TEACHING

2026–2027	M1 SID UT	Math methods for supervised Machine Learning TDs/TPs
	M2 SID UT	Deep Learning Lectures/TDs/TPs
	M2 Mapi3 & IMA UT	Math for Machine Learning TDs/TPs
2025–2026	1st year INSA	Maths 0 TDs (Logic, Set theory, Applications, Sequence)
	M1 SID UT	Math methods for supervised Machine Learning TDs/TPs
	M2 Mapi3 & IMA UT	Math for Machine Learning TDs/TPs
2024–2025	M1–M2 SID UT3	Inter-promotional project supervision: Open data university, energy performance diagnosis group
	M1 SID UT3	TER supervisions, two students working on LLM evaluation and ML applications in education
	M2 SID UT3	Learning process for Big Data, presentation jury