Lucas Magnana

Research engineer in deep learning

I completed my PhD at INSA Lyon in February 2024. Since then, I have been working as a research engineer at Inria, focusing on the audit and privacy/security of Large Language Models (LLMs).

Contact

Phone
06 15 42 09 61

Email
lucas.magnana@gmail.com

Website
http://perso.citi-lab.fr/lmagnana

Github
https://github.com/LucasMagnana

Work Interests

Skills

- NLP / LLMs
- HuggingFace
- Reinforcement Learning
- Deep Learning
- PyTorch
- Machine Learning
- Python

Experience

Sep 2025 - Now
Inria (CITI Laboratory, PRIVATICS team) | Villeurbanne

Involvment in PANAME and NoLefa projects

- **Objective** : Implement libraries to audit machine learning models' security, robustness and fairness
- **Technologies used** : *Python*, *PyTorch*, *LLMs*, *Transformers*, *HuggingFace*

- Contribution to the French PANAME project
- *PANAME* (Privacy Auditing of AI Models): standardized library for auditing model privacy
- *Objective*: enable organizations to easily audit their models
- *My work*: co-development of the LLM module (model wrappers, privacy attacks, example scripts, unit and functional tests)

- Contribution to the Work Package 2 (WP2) of the NoLefa project
- *Objective*: development of an AI testing suite to support compliance with the AI Act
- *Scope*: robustness, fairness, and training data quality
- *My work*: initial use case → re-identification attack on a LLM trained on medical data

Jun 2024 - Now
Inria (CITI Laboratory, PRIVATICS team) | Villeurbanne

Research engineer

- **Objective :** Help healthcare actors evaluate and reduce their LLMs' weaknesses against data leaks and adversarial attacks
- **Technologies used** : *Python*, *PyTorch*, *LLMs*, *Transformers*, *HuggingFace*, *Docker*, *Opacus*, *Machine Unlearning*

- *Problem*: LLMs' data leakage and adversarial vulnerabilities (medical use cases)
- *Solution*: Privacy-Preserving Language Modeling (PPLM) finetuning strategy
  - PPLM prevents models from learning both direct and indirect identifiers
  - Masked language models (MLMs): masking of identifiers prohibited
  - Causal language models (CLMs): identifiers replaced with padding tokens
- *Results*: best utility-privacy trade-off compared to protected and unprotected baselines

- *Problem*: cost of re-training/re-finetuning a model with PPLM
- *Solution*: machine unlearning with PPLM as reconstruction phase
  - Finetuned model: 10-20% of the original finetuning duration
  - Pre-trained model: few epochs to forget specific information
- *Results 1*: utility-privacy trade-off comparable to fully protected finetuning
- *Results 2*: drastic reduction of regurgitation for pre-trained model

- *Research question*: Does LLMs' language impact privacy risks ?
- *Solution*: attack/evaluate LLMs pre-trained on different languages
  - Extraction attack
  - Counterfactual Memorization evaluation
  - Membership Inference Attack
- *Results*: Significant differences between languages for all attacks and evaluations

2021 - 2022 and 2025 - 2026
INSA Lyon

Teaching

- I've had the pleasure to teach for 3 semesters the basics of algorithmic to first year students at INSA Lyon.

Language

- French
- English

Life interests

- Reading books
- Community life
- Music instruments
- Human sciences
- Cooking
- Boxing

2022 - 2023
Inria (CITI Laboratory, Agora team) | Villeurbanne

PhD Project

- **Objective :** Make specific road segments more attractive to cyclists by modifying traffic lights in a safer and smarter way
- **Tools used :** *Python, PyTorch, SUMO, Deep Reinforcement Learning (3DQN, PPO)*

- *Problem*: Making road segments attractive to cyclists is costly in space (bike lanes)
- *Solution*: Intelligent traffic lights secured for cyclists
  - Simulations using SUMO (Simulation of Urban MObility) and vehicle counters data
  - Green phases added specifically for cyclists → explosion of the waiting time
  - Deep Reinforcement Learning to dynamically control the traffic light's phases
- *Results*: Deep Q-Network (3DQN) outperforms other tested traffic light control methods.

2020 - 2022
Inria (CITI Laboratory, Agora team) | Villeurbanne

PhD Project

- **Objective :** Understand cyclists' route choices without *a-priori* knowledge such as bike infrastructures
- **Tools used :** *Python, PyTorch, Sklearn, Jupyter Notebook, Recurrent Neural Networks (LSTM), Clustering algorithms (DBSCAN, K-means)*

- *Problem*: Route choice models for cyclists uses a-priori knowledge → sources of biases
- *Solution*: Implicit route choice models for cyclists using GPS tracks only
  - Quantitative and qualitative analysis of GPS tracks
  - Clustering on GPS tracks to find preferred road segments
  - LSTM to find relevant preferred road segments from any origin/destination
  - Weighting of a road graph using relevant preferred road segments
  - Shortest path on weighted road graph
- *Results*: Constructed paths are closer to real behaviors than commercial solutions

Education

Sept 2015 - Aug 2018
UIT and Bachelor in computer science
University of Lyon

Sept 2018 - Aug 2020
Master of computer science
University of Lyon

Oct 2020 - Feb 2024
PhD in Computer Science
INSA Lyon

Publications

- Implicit GPS-based bicycle route choice model using clustering methods and a LSTM network | PLOS ONE
- A DRL solution to help reduce the cost in waiting time of securing a traffic light for cyclists | Journal of Cycling and Micromobility Research
- Towards the Anonymization of Language Modeling. | arXiv preprint arXiv:2501.02407
- Leverage Unlearning to Sanitize LLMs | arXiv preprint arXiv:2510.21322
- The Model's Language Matters: A Comparative Privacy Analysis of LLMs | EACL 2026

Letters of Recommandation

Here is a list of people who can send a letter of recommendation on my behalf :

- Hervé Rivano (University Professor at INSA Lyon) | herve.rivano@insa-lyon.fr
- Antoine Boutet (Associate Professor at Inria) | antoine.boutet@inria.fr
- Nicolas Berkouk (AI Scientific Expert at CNIL) | nberkouk@cnil.fr