

MASTER'S RESEARCH INTERNSHIP

Privacy, Incentives and Learning Agents

Mathematical foundations of trustworthy AI

Host	Toulouse School of Economics	Location	Toulouse, France
Supervisor	Achraf Azize	Duration	4–5 months
Department	Mathematics and Statistics	Start	April 2027

Scientific context

AI systems increasingly act on behalf of users and interact with other decision-makers. A personal assistant, for example, may need to disclose information to obtain a useful service, even when the recipient's objectives differ from those of its user. Such interactions motivate a mathematical question at the interface of privacy and alignment: **how should information protection be understood when learning and strategic behaviour interact?**

Differential privacy provides rigorous guarantees on information disclosure [1]. Relating these guarantees to agents' interests and decisions also requires modelling incentives. Work on endogenous participation and data acquisition [2, 3], together with recent models of privacy among learning agents [4], provides starting points. Performative prediction offers a complementary perspective on learning when a deployed system changes the environment from which it receives data [5].

Research objective

The project will investigate privacy, incentives and statistical learning in tractable models of strategic interaction. The aim is to understand information protection and its consequences for learning and decision-making. This offers an entry point into the foundations of AI alignment through the study of information and incentives.

The emphasis is on **mathematical modelling and proofs**. A focused problem, adapted to the internship's duration, will be chosen with the student.

Approach and expected outcomes

The internship will combine probability, statistical learning and game theory. Depending on the direction, it may also draw on information theory, mechanism design or information design. The work will proceed in three stages:

1. **Foundations and literature.** Understand the relevant privacy guarantees and strategic models, reproduce a benchmark, and identify a precise question.
2. **Modelling and analysis.** Formulate a tractable setting and establish mathematical results about the interaction between information, incentives and learning.
3. **Synthesis and evaluation.** Examine the scope of the results through examples, counterexamples and, where useful, numerical experiments; prepare a research report and presentation.

The expected contribution is a well-justified model and a rigorous analysis of a focused case. Evaluation will consider the proofs, comparisons with relevant benchmarks and the role of the assumptions. Numerical illustrations will use synthetic data generated from the model.

Candidate profile and training

This offer is intended for a final-year Master's student (M2 or equivalent) in mathematics, statistics, machine learning or theoretical computer science. We are looking for:

- a strong background in probability and mathematical reasoning;
- an interest in theoretical research and in the foundations of trustworthy AI;
- curiosity about the interaction between learning, information and incentives.

Prior knowledge of game theory or differential privacy is welcome but **not required**; the relevant foundations can be acquired during the internship. Programming skills are useful for exploring examples. The project connects to Master's coursework in statistical learning, probability, optimization and, depending on the direction, online learning.

Host and scientific environment

Toulouse School of Economics (TSE), a constituent institution of Université Toulouse Capitole, brings together research in economics, mathematics and quantitative social sciences. Its Department of Mathematics and Statistics offers an environment spanning game theory, probability, statistics and machine learning, with close links to economic theory.

The internship will be supervised by **Achraf Azize**, Assistant Professor at TSE and holder of a *Chaire de Professeur Junior*. His research focuses on privacy, online learning, and AI alignment. The student will benefit from regular supervision and the department's research community and seminars.

Practical information and applications

The internship will take place at TSE in Toulouse for **4–5 months**, with a start around **April 2027**. Exact dates can be adapted to the student's academic calendar. A continuation into a PhD on related topics may be considered, subject to mutual interest and doctoral admission. Dedicated funding for the PhD is available through Achraf Azize's Chaire.

To apply, send a **CV**, **recent academic transcripts**, and a **short statement of research interests** to achraf.azize@tse-fr.eu. Include your availability and, optionally, a report illustrating your mathematical work. Informal enquiries are welcome.

Contact: Achraf Azize | achraf-azize.github.io

Selected references

- [1] C. Dwork and A. Roth. *The Algorithmic Foundations of Differential Privacy*. Foundations and Trends in Theoretical Computer Science, 2014.
- [2] A. Ghosh and K. Ligett. *Privacy and Coordination: Computing on Databases with Endogenous Participation*. ACM EC, 2013.
- [3] A. Fallah, A. Makhdoumi, A. Malekian and A. Ozdaglar. *Optimal and Differentially Private Data Acquisition: Central and Local Mechanisms*. Operations Research, 72(3), 2024.
- [4] N. Ananthakrishnan, T. Ding, M. Werner, S. P. Karimireddy and M. I. Jordan. *Privacy Can Arise Endogenously in an Economic System with Learning Agents*. FORC, 2024.
- [5] J. Perdomo, T. Zrnic, C. Mendler-Dünner and M. Hardt. *Performative Prediction*. ICML, 2020.