

Curriculum Vitae – Thomas Souverain

thomas.souverain@ens.psl.eu ; +33 (0)7 77 72 26 34

1. Education

Postdoctoral Fellowship on Ethics and Governance – CEA Paris-Saclay (2025–2026)

In charge of **Ethics & Compliance** for the OpenLLM consortium (Saclay – Shanghai 1st rank in continental Europe, Sorbonne, LORIA), funded by France 2030. Development of a frugal, **open-access Large Language Model** (LLM) aligned with EU law for the **Ministry of Education**. Designed explainability and safety modules; coordinated **classroom-based field evaluations**.

PhD in Applied Philosophy – ENS Ulm & French Employment Agency (2020–2024)

Thesis based on **fieldwork**, in partnership with **AI departments** of a credit company & the French Employment Agency (France Travail) : “**How AI Classifies Us: Explaining AI in Financial Services and Public Employment.**” Supervision: Prof. Paul Égré ; Machine Learning (ML) supervision: Dr Nicolas Meric (DreamQuark), Dr Gaëlle Jouis (French Employment Agency). Mixed **philosophical, sociological and computer science methodology**.

Master’s in Philosophy – ENS Ulm / Paris Sciences & Lettres University (2018–2020)

Epistemology & Philosophy of Science. Thesis (High Distinction): “Explainability Challenges in Credit Scoring AI.”

Classe Préparatoire B/L – Lycée Henri-IV, Paris (2015–2018)

Intensive bachelor training in **mathematics, philosophy, economics, sociology**, history and languages.

2. Professional & Research Experience

European Commission & Hub France AI – AI Expert (2019–2025)

Consulted by the **EU AI Office** on **2024 AI Act transparency** implementation. Lead of the Ethical AI Working Group, developing operational tools and **European standards** for AI providers.

French Employment Agency – PhD Fieldwork (2023–2024)

Large-scale qualitative & quantitative study on trust and adoption of AI tools for detecting illegal job postings. Mixed qualitative and quantitative study; design and iteration of the tool with practitioners.

Stanford University – Visiting Researcher Student (Jan–Apr 2023)

Joint ENS–Stanford project on graph neural networks integrating structural causal notions. Work at the Department of Philosophy (epistemology) & Computer Science.

Oxford University – Visiting Researcher Student (Sep–Dec 2022)

Developed a method to measure perceptions of “**fair AI**” **with non-technical users**; collaboration with the Institute for Ethics in AI and the Maison Française d’Oxford. Associate Member of **New College MCR**.

DreamQuark – Data Science Team (2020–2023)

Programming competitions and **ethical AI experimentation**. Recognitions: **UNESCO Top 100 of World Global Outsanding Projects** (2022); MAS Veritas Challenge Finalist (2021). Tools: Python, Pandas, PyTorch, Git.

3. Research Activities

Field Investigations

Embedded **ethnography in Machine Learning teams** (DreamQuark, French Employment Agency). Study of trust, explanation and fairness in **real-world AI deployments**.

Technical Work and International Collaborations

Open-source packages: causal-transparency testing for GNNs; FairDream fairness library for credit scoring; extensive use of Python, NumPy, Pandas, Git. Stanford (causal GNNs) and Oxford (fairness perception studies).

4. Publications

Peer-Reviewed Articles

- Souverain, T. (2025). [Watermarking Large Language Models in Europe: Interpreting the AI Act in Light of Technology](#). arXiv preprint arXiv:2511.03641. En relecture par la revue *Ethics & Information Technology*.
- Souverain, T. (2025). [AI to renew public employment services? Explanation and trust of domain experts](#). *AI and Ethics*, Special Issue 'Exploring the Mutations of Society in the Era of Generative AI', 1-16.
- Ledro, C., Nosella, A., Vinelli, A., Dalla Pozza, I., & Souverain, T. (2025). [Artificial intelligence in customer relationship management](#): A systematic framework for a successful integration. *Journal of Business Research*.
- Souverain, T., Nguyen, J., Meric, N., & Égré, P. (2024). [Implementing Fairness in AI Classification: The Role of Explainability](#), *Topoi*.

Practitioner-Oriented Publications

- Souverain T. et. al. (2023, reed. 2025, French) "Integrating Ethics in AI Systems", Hub France IA.
- Souverain T., Meric N., "State of Ethical AI: Challenges posed by ethics to companies developing AI", 2021, 110 p. Analysis of stakes, international strategies, and applications to integrate Ethics into AI development.
- [Generative AI as a New Idol?](#) (French) *Communio*, 2025, 65-76. <https://doi.org/10.3917/commun.301.0065>.

Editorial Work

Founder & Editor-in-Chief of a new **interdisciplinary AI ethics journal** associating research, industrial, and governance actors (ISSN, DOI assignment, >130 submissions, 20-article inaugural issue). **Scientific Chair of the associated conference** : <https://paris-conference.com/journal/>

5. Responsibilities

Academic Roles

Head of the 'Trust & AI' research group (Paris XIII, from 2025), **jury** and pedagogical committee member (M2 Digital Law), organiser of intensive citizen consultations drafting **generative-AI strategy inside universities** – ENS, Sciences Po Bordeaux, and Paris XIII; supervision of interns; member of Paris-Saclay **Ethics Committee**.

Research Dissemination & Policy

Leader of Hub France AI's Ethical AI Working Group; expert member of AFNOR's **AI standardisation** committee; contributor to MOSCAI, EHESS IA programme, European AI Alliance, Association of AI Ethicists.

6. Teaching (Governance and Ethics of AI)

General Approach

Developed full syllabi for future **AI practitioners (governance leaders, computer scientists, engineers, lawyers)**. Courses combine **applied ethics, European regulatory frameworks, and hands-on ML exercises**.

Tutorials / Labs

Professional role-play exercises (data scientist, CNIL auditor, product lead). **Programming labs** on explainability limits in credit scoring. Group-graded projects and exams.

Course Portfolio (Selection)

- Sorbonne Paris XIII – MSc Data Science (2021–2026): Ethics of AI; 6h–40h formats
- Institut Supérieur d'Électronique de Paris – M1 Engineering (2023/24): Introduction to AI Ethics
- EM Lyon – Master in DS & AI Strategy (2023/24): Ethics & Responsible AI
- Institut Catholique de Paris – M2 Digital Law (2024/25): Ethics of Data & AI (20h)
- CentraleSupélec – M2 AI specialization (2025/26): Explainability: Techniques, Law, Ethics
- ENS Paris – L3/M1/M2 Philosophy (2025/26): “What is a Philosophy of AI Models” (24h, 18 theses)

7. Talks & Outreach

Academic Audiences

Around **30 talks in 7 countries**, among others in CEA Saclay, Sorbonne LIP6, IRISA, IRIT, Institut Jean Nicod, Maison Française d'Oxford, Amsterdam.

AI Practitioners & Policymakers

Sciences Po, MSN Paris XIII, **Paris World Summit on AI for Action**, OpenLLM Consortium, DreamQuark, French Employment Agency. Focused on **large-scale AI-ethics training**.

General Public & Interdisciplinary Audiences

Quandela (AI & quantum), Institut Jacques Monod (biology), public debate series, **scientific mediation** videos ([Fr](#)).

8. Additional Interests

Humanitarian Work

Current projects supporting people with **reduced mobility and homeless communities** (housing and feeding initiatives) ; previously scout leader for 10 young adults focused on the same range of projects (2018-2022).

Association MigrENS

Support for teaching French to **refugees** resuming university studies (2018–2020).

Acting

Appearances in films/series on France 2 (2010–2014).

Mountain Running

[Winner](#) of the 2022 Verbier Saint-Bernard Trail (46 km, +3600 m, 7h18); 140 km edition in 2025 (36h13).