Tommaso **Mencattini**

MSc student at **EPFL** · Research Intern at **ISTA**, **ETH Zurich**, and **Gladia** · Papers at **ICML**, **ACL**, **NeurIPS**, **ECAI**

✉ tommaso.mencattini@epfl.ch | ⌨ tommasomncttn | 💼 tommasomencattini | 🜁 Tommaso Mencattini

Education

École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

M.S. in Data Science (**GPA: 5.5/6**)
2024 – Present

- Courses focused on the mathematics of data science, from measure-theoretic probability to reinforcement learning theory.

Vrije Universiteit Amsterdam
Amsterdam, Netherlands

B.S. in Artificial Intelligence (Major) and Mathematics (Minor) (**GPA: 9.2/10**)
2021 – 2024

- Ranked among the top 7 AI students in the Netherlands for achieving one of the **highest GPAs in a cohort of nearly 300 students** (KHMW Award).
- **Honours Programme:** 30 additional credits of advanced coursework (e.g., Partial Differential Equations) and research projects.

BlueDot Impact
Remote

Summer School in AI Safety (3 months)
Summer 2024

- Studied sparse autoencoder diffing and transfer. **Accepted at UniReps NeurIPS 2024 (Blog Post)**.

Work Experience

Locatello Group, ISTA
Vienna

Research Student
07/2025 – Ongoing

- Research on causality and multimodal foundation models. Co-first author submission to **ICLR 2026**: framework (theory + library) for unsupervised causal effect discovery in RCTs *(PI: Francesco Locatello)*.

GLADIA, Università La Sapienza
Rome

Research Student
04/2024 – Ongoing

- Scaled evolutionary model merging of LLMs to consumer GPUs; developed ability-estimation theory via Item Response Theory. Now **50×** faster, uses **2%** of the data, with **85%** performance retained *(PI: Emanuele Rodolà)*. Two first-author papers **accepted at ICML 2025, ACL 2025**.
- Co-first author submission to **ICLR 2026**: prove injectivity of LLM hidden representations and present the first exact inversion algorithm (**4.7 M views** on 🐦 tweeprint).

LRE Lab, ETH Zurich
Zurich

Research Student
06/2023 – 08/2023

- Applied causal inference to explain and assess LLM performance on math word problems *(PI: Mrinmaya Sachan)*.

KAI Lab, Vrije Universiteit
Amsterdam

Research Assistant
01/2023 – 06/2023

- Mitigated hallucinations in LLMs using knowledge graphs for content planning *(PI: Ilaria Tiddi)*.
- Designed methods to extract knowledge graphs from text (Text-to-Graph) and convert them into content planners (Graph-to-Text).

Honors & Awards

2022	**KHMW Young Talent Incentive Award,** National award presented to the student with the highest GPA in their degree program at a Dutch research university.	*Amsterdam*
2021	**Intelligent Systems Competition,** Ranked 1st out of 270 VU students in the final competition. Developed an intelligent system to play the trick-taking card game Schnapsen against other students' bots.	*Amsterdam*

Publications

- G. Nikolaou*, **T. Mencattini*** et al. "Language Models are Injective and Hence Invertible". Under Review.
- **T. Mencattini***, R. Cadei*, et al. "Exploratory Causal Inference in SAEnce". Under Review.
- **T. Mencattini***, A. R. Minut*, et al. "MERGE³: Efficient Evolutionary Merging on Consumer-grade GPUs". **ICML** 2025.
- S. Facchiano, G. Strano, D. Crisostomi, I. Tallini, **T. Mencattini** et al. "Activation Patching for Interpretable Steering in Music Generation". Under Review.
- A. R. Minut*, **T. Mencattini*** et al. "Mergenetic: A Simple Evolutionary Model Merging Library". **ACL** 2025 (System Demonstrations).
- M. Calamo, F. De Luzi, M. Macrì, **T. Mencattini** et al. "CICERO: A GPT2-based writing assistant in e-justice". **ECAI** 2023.
- T. Kutsyk*, **T. Mencattini*** et al. "Do Sparse Autoencoders (SAEs) transfer across base and finetuned language models?" **NeurIPS** 2024 (Workshop).