

Tommaso Mencattini

MSC STUDENT AT EPFL · RESEARCH INTERN AT ISTA, ETH ZURICH, AND GLADIA · PAPERS AT ICML, ACL, NEURIPS, ECAI

[✉ tommaso.mencattini@epfl.ch](mailto:tommaso.mencattini@epfl.ch) | [💻 tommasomncttn](https://github.com/tommasomncttn) | [🔗 tommasomencattini](https://www.linkedin.com/in/tommasomencattini/) | [✉ Tommaso Mencattini](https://scholar.google.com/citations?user=QWzgkxUAAAAJ&hl=en)

Education

École Polytechnique Fédérale de Lausanne (EPFL)

M.S. IN DATA SCIENCE (GPA: 5.5/6)

Lausanne, Switzerland

2024 – Present

Vrije Universiteit Amsterdam

B.S. IN ARTIFICIAL INTELLIGENCE (MAJOR) AND MATHEMATICS (MINOR) (GPA: 9.2/10)

Amsterdam, Netherlands

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

SUMMER SCHOOL IN AI SAFETY (3 MONTHS)

Remote

Summer 2024

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

Work Experience

Locatello Group, ISTA

RESEARCH STUDENT

Vienna

07/2025 – Ongoing

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

GLADIA, Università La Sapienza

RESEARCH STUDENT

Rome

04/2024 – Ongoing

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

LRE Lab, ETH Zurich

RESEARCH STUDENT

Zurich

06/2023 – 08/2023

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

KAI Lab, Vrije Universiteit

RESEARCH ASSISTANT

Amsterdam

01/2023 – 06/2023

- Mitigated hallucinations in LLMs using knowledge graphs for content planning (*Pl: 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 1 st 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 Luzzi, M. Macrì, **T. Mencattini** et al. “CICERO: A GPT2-based writing assistant in e-justice”. **ECAI** 2023.
- T. Kutsyuk*, **T. Mencattini*** et al. “Do Sparse Autoencoders (SAEs) transfer across base and finetuned language models?” **NeurIPS** 2024 (Workshop).