Antonio Mari

antonio.mari02@outlook.com — mari-antonio — antoniomari — Google Scholar Profile

Experience

Algoverse LLC - *Research Mentor* *Jul 2025 – Present*
- Mentoring 12 students on Mechanistic Interpretability research on Large Language Models and Diffusion Transformers.

Oracle Labs - *Research Assistant (6 months internship)* *Jul 2024 – Jan 2025*
- Developed a Machine Learning system for Cloud security defense, enhancing a Spark-based pipeline that processed billions of events daily and leveraging Transformer-based engineered features.
- Implemented and trained state-of-the-art anomaly detection models (PyOD library + custom implementations).
- Engineered custom data-loading mechanisms to handle large-scale training data, ensuring seamless model training.

University of Edinburgh - *Post Graduate Researcher (3 months internship)* *March 2023 – June 2023*
- Explored the connections between Probabilistic Circuits and tensor factorizations, conducting in-depth experiments.
- Designed and implemented complex PyTorch models from scratch, incorporating numerical stability techniques.
- Developed the foundation for a Python library for Probabilistic Circuits: *cirkit*.

Education

ETH, Eidgenössische Technische Hochschule Zürich - *MSc Thesis* *September 2025 – February 2026*
- Test-time Meta Reinforcement Learning: investigating test-time training to reinforcement learning, enabling policies to adapt and improve performance using information gathered during deployment.
- Designing a meta-learning framework to pre-train policies for rapid, computationally efficient adaptation, with the goal of reducing the gradient steps needed at test-time.

EPFL, École Polytechnique Fédérale de Lausanne - *MSc Data Science (**GPA: 5.83/6**)* *September 2023 – March 2026*
- Foundational machine learning, Large Language Models and Diffusion Models. **Relevant Courses**: Machine Learning, Applied Data Analysis, Markov Chains, Modern Natural Language Processing, Reinforcement Learning, Advanced Probability and Martingales, Visual Intelligence, Data visualization.
- **Part-time Teaching Assistant**: CS-552 Modern Natural Language Processing MSc course - prof. Antoine Bosselut.

University of Edinburgh - *Bachelor of Science Thesis* *March 2023 – July 2023*
- Unified overparameterized Probabilistic Circuits in an end-to-end pipeline, including existing and new SotA architectures
- Experimented on parameterizations based on tensor decomposition, showing how lower-rank layers are easier to train.

University of Bari Aldo Moro - *BSc Computer Science (**GPA: 29.97/30**)* *September 2020 – July 2023*
- Final Grade: 110/110 with Honors and Special Mention
- Samsung SmartThings, MSc Deep Learning, Programming, Software Engineering, Data Structures, Algorithms, Calculus, Databases, Numerical Analysis, Probability, Statistics, Computability, Information Retrieval, Knowledge Engineering.

Research Publications

Surkov V., Wendler C., **Mari A.** et al. arXiv:2410.22366
- "One-Step is Enough: Sparse Autoencoders for Text-to-Image Diffusion Models." *NeurIPS 2025*.

Loconte L*., **Mari A.***, Gala G.*, et al. – [Co-first author] arXiv:2409.07953
- "What is the Relationship between Tensor Factorizations and Circuits (and How Can We Exploit it)?" *Transactions on Machine Learning Research (TMLR)*, 2025. **Featured Certification.**

Mari A., Vessio G., Vergari A. Openreview:1btutFdIya
- "Unifying and understanding overparameterized circuit representations via low-rank tensor decompositions." In *6th Workshop on Tractable Probabilistic Modeling*, 2023.

Atsushi Y. et al., **Mari A.** – [Supervisor] Openreview:aslS4eRygE
- "Liminal Training: Characterizing and Mitigating Subliminal Learning in Large Language Models" *NeurIPS 2025 ResponsibleFM Workshop*.

Jamal A. et al., **Mari A.** – [Supervisor] Openreview:Fc7s3UrQr1
- "Diffusion Transformers use Sink Registers" *AAAI 2026 XAI4Science Workshop*.

Selected Projects

Diffusion Transformer Edit (Image composition) *February 2025 – June 2025*
- Designed a novel training-free method for real image composition based on SotA diffusion models (Flux) .
- Performed qualitative evaluation and a user study to prove superiority of our solution with respect to existing methods.
- **Best Project Award** at *"Visual Intelligence: Machines and Mind"* course (prof. Amir Zamir), consisting of compensation for virtual conference registration.

STEMerald (LLM): STEM-focused AI assistant *April 2024 – June 2024*
- Fine-tuned Gemma-2b using Direct Preference Optimization; reduced memory footprint to 2GB through quantization.
- Achieved high accuracy in multiple-choice QA across STEM-subject after alignment with university-level data.

Skills

Nat. Lang Italian (Native), English (Fluent, C1 cert.)
Prog. Lang Python, C, C++, Java, SQL, Prolog, Latex, HTML, CSS, JavaScript
Libraries numpy, pandas, scikit-learn, TensorFlow, PyTorch, matplotlib, seaborn, transformers, trl, pyspark, gensim, diffusers, react, (JAX, learning)