

Francesco Massafra

Amsterdam, Netherlands | massafra32@gmail.com | +39 370 127 1535 | linkedin.com/in/francesco-massafra | github.com/FireWtap

Profile

AI researcher and M.Sc. Artificial Intelligence student focused on reinforcement learning, agentic planning, LLM evaluation, and reproducible deep-learning experiments. Experienced in Python/PyTorch research pipelines, model evaluation, failure analysis, and clean software engineering. Two-time post-graduation research grant recipient with 4+ years of software engineering experience.

Core Skills

AI/R&D: Reinforcement learning, agentic planning, latent world models, LLM evaluation, model probing, model evaluation, failure analysis, data mining, Python, PyTorch

Research Engineering: Reproducible experiments, pandas, NumPy, scikit-learn, Git, Docker, Linux, REST APIs, PostgreSQL

Additional ML/Data: XGBoost, SHAP, Scanpy, differential expression analysis, Java, React, Node.js

Research Experience

MEDICA Project, University of Pisa

Pisa, Italy

Research Assistant; Two-Time Post-Graduation Research Grant Recipient

2024 – 2025

- Won two post-graduation research grants and built reproducible ML pipelines for transcriptomics classification, feature attribution, statistical analysis, and cross-dataset integration; resulting work published as [arXiv:2603.05572](https://arxiv.org/abs/2603.05572).

Education

University of Amsterdam

Amsterdam, Netherlands

M.Sc. Artificial Intelligence

Sept 2025 – Present

- GPA: 8.57/10. Relevant coursework: Reinforcement Learning, Deep Learning, NLP, Computer Vision, Information Retrieval.

University of Pisa

Pisa, Italy

B.Sc. Computer Science

Sept 2021 – Nov 2024

- Final grade: 110/110; thesis on ML for Multiple Sclerosis biomarker discovery.

Selected AI Research Projects

Hi-LeWM: Hierarchical Planning in LeWorldModel [\[code\]](#)

2026

- Designed and evaluated a transformer-based hierarchical planner for goal-conditioned control, combining latent macro-actions, CEM/MPC planning, and receding-horizon execution over a frozen world model.
- Diagnosed off-manifold subgoal failures in long-horizon planning and improved robustness by constraining high-level search to data-supported macro-actions.

Open-Weight LLM Safety and Dataset Filtering [\[code\]](#)

2026

- Built reproducible LLM evaluation pipelines for prompt-based harmful-content classification across open-weight models, moderation datasets, and web-scale corpora.
- Implemented deterministic data mining, preprocessing, model-evaluation, and analysis workflows across Common Crawl, C4, FineWeb, and safety benchmarks.

Do Video Foundation Models Understand Intuitive Physics? [\[code\]](#)

2026

- Probed frozen V-JEPA, VideoMAE, and LTX-Video representations on IntPhys2 and MVP to evaluate whether pretrained video models encode intuitive-physics structure.
- Trained and compared linear, MLP, and temporal-attention readouts; analyzed layerwise emergence, temporal-control failures, and shortcut-driven benchmark behavior.

Software Engineering Experience

Wylit

Remote / Hybrid – Part-time

Full-Stack Developer

Sept 2021 – Aug 2025

- Built and owned full-stack production web platforms for 15+ client projects using React, Node.js, PostgreSQL, Docker, and Linux deployments, improving responsiveness and reducing load times by up to 40%.