Guiv Farmanfarmaian

Zürich, Switzerland
+41 78 900 74 65 • guivff@gmail.com
guivff.github.io
Citizenship: US, Swiss, Canadian
Languages: English, French, Persian, Italian (C2), German (C2), Russian (A2), Spanish (A2)

PROFILE

Machine Learning researcher at ETH Zürich focused on **stable RL post-training** for LLM reasoning, with hands-on experience in RLVR and test-time compute. Master's thesis develops methods for preventing mode collapse, recovering learning signal on hard mathematical reasoning problems, and evaluating test-time inference strategies. The teacher-hint conditioning work is confidential, with a manuscript in preparation. Strong mathematical foundation (**BSc Mathematics, ETH Zürich**), with a research agenda centered on **s**.

EDUCATION

ETH Zürich — MSc Computer Science (Machine Intelligence) *Zürich, Switzerland*
Sep 2023 – Feb 2026 • Major: Machine Intelligence • Minor: Theoretical Computer Science • GPA: 5.7/6.0 (top 10%)

- **Master's Thesis:** *Improving LLM Reasoning on Challenging Problems* — supervised by Dr. Amir Joudaki and Prof. Dr. Thomas Hofmann; **manuscript in preparation.**
- **Semester Thesis:** *Improving Investment Strategies with GNNs* — supervised by Florian Grötschla, Joel Mathys, and Prof. Dr. Roger Wattenhofer (in collaboration with a leading investment firm; under confidentiality).
- **Selected coursework:** Deep Learning, Large Language Models, Machine Perception, Probabilistic Artificial Intelligence, Advanced Algorithms, Convex Optimization, Information Theory.

ETH Zürich — BSc Mathematics *Zürich, Switzerland*
Sep 2020 – Sep 2023 • Focus: Number Theory, Theoretical Computer Science

- **Selected coursework:** Probability Theory, Number Theory I & II, Algorithms / Probability / Computing (APC).

RESEARCH EXPERIENCE

Master's Thesis — RL Post-training & Test-Time Compute for Math Reasoning *Zürich, Switzerland*
ETH Zürich — Data Analytics Lab (Hofmann Group) • June 2025 – Feb 2026

- **Stable RLVR via teacher-hint conditioning.** Designed a GRPO variant that prevents mode collapse during RL post-training: **+54% policy entropy** and **~9× lower KL divergence** vs. the GRPO baseline, with **+3 pp on unhinted MATH-500 pass@8*. Confidential; manuscript in preparation.***
- **Recovering RL signal on hard problems.** Developed a minimal-prefix solution-conditioning method that restores learning signal on problems with zero pass@16, yielding **+5 pp on the unhinted target benchmar**k, addressing the well-known cold-start problem in RLVR.
- **Test-time compute methods.** Implemented and evaluated recursive reasoning, self-refinement, and best-of-N inference strategies on hard math problems; analyzed where test-time gains plateau and how they could be distilled into base weights.
- **Engineering.** Built a distributed RL training pipeline in PyTorch with vLLM-based rollouts, profiled and optimized GPU utilization on multi-node clusters.

Semester Thesis — Graph Learning for Financial Forecasting *Zürich, Switzerland*
ETH Zürich — Wattenhofer Group, in collaboration with a leading investment firm • Mar 2024 – Sep 2024

- Built and evaluated GNN and Graph Transformer architectures on a large directed supply-chain graph for predicting portfolio-relevant financial metrics.
- Ran controlled ablations comparing graph-based models to non-relational baselines, isolating the contribution of supply-chain topology to predictive accuracy and downstream portfolio performance.

RESEARCH INTERESTS

- **Stable RL for reasoning.** Why does RLVR collapse, and what's the right inductive bias to keep policies exploratory while still optimizing reward? Particularly interested in entropy regularization, off-policy correction, and reference-policy conditioning.
- **Sample-efficient reasoning.** Reaching strong reasoning performance with substantially less data. Fewer environment interactions also means greater control over what skills the model acquires and from where, making training more steerable and auditable.
- **Distilling test-time compute into weights.** Test-time recursion and best-of-N improve performance dramatically, but inference cost grows. How can these gains be amortized into the base policy through self-distillation or expert iteration?
- **Tiny Recursive Models for symbolic reasoning.** Extending TRM-style architectures beyond ARC toward algebraic and symbolic reasoning where verifiable rewards are tractable.
- **Continual learning.** Architectures and algorithms that mitigate catastrophic forgetting and enable models to keep improving from new data without re-training from scratch.

PROFESSIONAL EXPERIENCE

Teaching Assistant — Computational Intelligence Lab (CIL) *Zürich, Switzerland*
ETH Zürich • Feb 2025 – Jun 2025

- Designed and graded coursework, ran exercise sessions, and held office hours covering matrix factorization, deep generative models, and modern representation learning.

Funding Circle Ltd. — Data Scientist *London, United Kingdom*
Nov 2019 – Aug 2020

- Built regression and clustering models to analyze loan-portfolio risk; deployed Random Forest classifiers for early default detection during company hackathons.
- Automated reporting pipelines in Python and R; built data ingestion and processing on AWS.

Reply Sytel — IT Consulting Intern *London, United Kingdom*
Apr 2017 – May 2017

- Wrote full-stack software prototypes and authored white papers on next-generation networking and Proof-of-Stake consensus.

TECHNICAL SKILLS

ML Research: PyTorch, vLLM, Hugging Face Transformers, TRL, DeepSpeed, distributed training (FSDP / DDP), W&B, Slurm, GraphGPS, PyG.

Programming: Python, C++, TypeScript, R, Java.

Mathematics: Number theory, probability theory, convex optimization, theoretical CS.