Alireza Abdollahpoorrostam

M.Sc. Researcher, EPFL

Lausanne, Switzerland | alireza.abdollahpoorrostam@epfl.ch | Google Scholar | GitHub | LinkedIn

Research Summary

I am drawn to the mathematical foundations of deep learning: why neural networks fail under adversarial perturbations and distribution shift, and what principled algorithms follow. My research spans robustness, generalization, and post-training methods (model merging, model editing) for foundation models, with recent work on the optimal-transport geometry of adversarial training and distributionally robust optimization. I aim to derive rigorous guarantees and translate them into deployable algorithms, with work appearing at **NeurIPS**, **ICML**, and **ICLR**.

Education

École Polytechnique Fédérale de Lausanne (EPFL) *Sept. 2024 – Present*
M.Sc. in Communication Systems—Machine Learning track Lausanne, Switzerland

- **Chair of Risk Analytics and Optimization (RAO)**, advised by **Prof. Daniel Kuhn**—optimal-transport geometry of adversarial training and distributionally robust learning.
- Research Scholar at **Signal Processing Laboratory (LTS4)**, advised by **Prof. Pascal Frossard**—robust fine-tuning and weight-space merging for vision–language foundation models.

Amirkabir University of Technology (Tehran Polytechnic) *Sept. 2019 – Aug. 2024*
B.Sc. in Electrical Engineering & Computer Science Tehran, Iran

Publications

1. **A. Abdollahpoorrostam**, N. Dimitriadis, A. Hazimeh, P. Frossard. Model Soups Need Only One Ingredient. *International Conference on Machine Learning* (**ICML 2026**).
2. M. Salmani, **A. Abdollahpoorrostam**, S.-M. Moosavi-Dezfooli. A General Framework for Black-Box Attacks under Cost Asymmetry. *International Conference on Learning Representations* (**ICLR 2026**), .
3. **A. Abdollahpoorrostam**, M. Abroshan, S.-M. Moosavi-Dezfooli. SuperDeepFool: A New Fast and Accurate Minimal Adversarial Attack. *Advances in Neural Information Processing Systems* (**NeurIPS 2024**).

Under Review

5. **A. Abdollahpoorrostam**, E. Sharifian, B. Sen, M. Cuturi, D. Kuhn. **Brenier Meets Adversarial Training: Optimal Transport Geometry for Robust Learning.** *Submitted to Advances in Neural Information Processing Systems (NeurIPS)*, 2026.

Workshop Papers & Preprints

6. **A. Abdollahpoorrostam.** In Search of the Successful Interpolation: On the Role of Sharpness in CLIP Generalization. *NeurIPS 2024 AdvML-Frontiers Workshop.*
7. **A. Abdollahpoorrostam**, A. Sanyal, S.-M. Moosavi-Dezfooli. Unveiling CLIP Dynamics: Linear Mode Connectivity and Generalization. *ICML 2024 Foundation Models in the Wild Workshop.*

Research Experience

EPFL — Chair of Risk Analytics and Optimization (RAO) *Sept. 2025 – Present*
Graduate Researcher | Advisor: Prof. Daniel Kuhn Lausanne, Switzerland

- Bridging **optimal transport** and **adversarial training**: OT-geometric formulations of adversarial training via Brenier maps, toward principled, provably robust learning algorithms.
- **"Brenier Meets Adversarial Training: Optimal Transport Geometry for Robust Learning"** (with E. Sharifian, B. Sen, M. Cuturi, D. Kuhn)—*submitted to NeurIPS 2026.*

EPFL — Signal Processing Laboratory (LTS4) *Sept. 2024 – Aug. 2025*
Graduate Researcher | Advisor: Prof. Pascal Frossard Lausanne, Switzerland

- Published MonoSoup, a hyperparameter-free, data-free post-hoc method that recovers Model-Soup-level out-of-distribution accuracy from a *single* fine-tuned checkpoint by applying SVD to per-layer updates and re-weighting high-/low-energy directions; reduces ensembling cost from $\mathcal{O}(K)$ to $\mathcal{O}(1)$. **(ICML 2026)**

Trustworthy ML Collaboration with Dr. S.-M. Moosavi-Dezfooli *2022 – 2025*
External Research Collaborator Remote / Tehran / Lausanne

- Co-developed SuperDeepFool, a minimal-ℓ_2 adversarial attack with a geometric reformulation of the DeepFool iteration. **(NeurIPS 2024)**

- Asymmetric-cost black-box attack, designing Asymmetric Search and Asymmetric Gradient Estimation (AGREST) algorithms for black-box settings. **(ICLR 2026)**
- Concurrent collaboration with Dr. Amartya Sanyal (U. Copenhagen) on *linear mode connectivity* of robust fine-tuning. (**ICML 2024 FM-Wild Workshop**).

ACADEMIC SERVICE

- Reviewer: NeurIPS, ICML, ICLR

Technical Skills

Programming: Python, C/C++, Bash, MATLAB, LaTeX.
ML Frameworks: PyTorch, JAX.
Languages: Persian (native), English.