Amin Rakhsha

arakhsha.github.io • aminr@cs.toronto.edu • Toronto, Canada

Summary

PhD candidate at the University of Toronto and researcher at the Vector Institute, with seven years of research experience in **reinforcement learning**. Led industry research on **inference-time scaling of LLM reasoning** and **long-horizon agent evaluation**, resulting in first-author papers at NeurIPS 2025 and Findings of ACL 2026.

Education

University of Toronto, Toronto, Canada
PhD in Computer Science — Sep. 2020–Oct. 2026 *(Expected)*
- Supervisor: Amir-massoud Farahmand

Sharif University of Technology, Tehran, Iran
BSc in Computer Engineering — Sep. 2020
- GPA: 19.52/20; ranked 2nd among 111 students

Research Experience

University of Toronto and Vector Institute, Toronto, Canada — Sep. 2020–Present
PhD researcher. Supervisor: Amir-massoud Farahmand
- Developed theoretically sound accelerated reinforcement learning algorithms, particularly useful for long-horizon tasks, by adapting techniques from numerical linear algebra, density estimation, and control theory.
- Designed planning algorithms that exploit imperfect models, matching model-free baselines with up to 3× fewer environment samples while remaining robust to model error.
- Proposed accelerated model-free algorithms with up to 75% lower computational cost than classical baselines.

Qualcomm AI Research, Amsterdam, Netherlands — Jun.–Sep. 2025
AI Research Scientist Intern
- Proposed and co-designed three highly configurable RL environments with programmatic task-success rewards and oracle assistance, enabling controlled, capability-specific evaluation of LLM agents.
- Produced targeted guidance for model post-training and agent design; one key finding was that long-context processing is a critical bottleneck for sub-8B models; published as first author at Findings of ACL 2026.

Autodesk, Toronto, Canada — Feb.–May 2025
AI Research Scientist Intern
- Proposed a bootstrap-based algorithm for reward-model-guided inference-time scaling of LLM reasoning, outperforming best-of-N in 25 of 30 model–task–reward-model settings by up to 11 percentage points.
- Led the project from conception through first-author publication at NeurIPS 2025.

Max Planck Institute for Software Systems (MPI-SWS), Saarbrücken, Germany — Jul. 2019–Jan. 2021
Research Intern (Jul.–Sep. 2019) and Collaborator (through Jan. 2021). Supervisor: Adish Singla
- Developed a theoretical framework for environment-poisoning attacks in reinforcement learning, deriving lower bounds on the cost of successful attacks and constructive attacks with provable success and cost guarantees.
- Work resulted in first-author papers in ICML 2020, JMLR 2021, and NeurIPS Workshop 2021.

Chinese University of Hong Kong (CUHK), Hong Kong — Jul. 2018–Jan. 2019
Research Intern (Jul.–Aug. 2018) and Collaborator (through Jan. 2019) with Anthony So; proved linear convergence of proximal-gradient methods for distributionally robust logistic regression with a fixed Wasserstein dual parameter.

Selected Honors and Awards

- **Silver Medal, International Mathematical Olympiad (IMO)** — Jul. 2016
- Borealis AI Global Fellowship — May 2022
- Three University of Toronto Research Awards — 2020, 2021, 2023
- Iran's National Elite Foundation Fellowship — 2015–2020

Publications

- Rakhsha, A., Hehn, T., Mazzaglia, P., Massoli, F. V., Behboodi, A., and Orekondy, T. LUMINA: Long-horizon understanding for multi-turn interactive agents. In Findings of the Association for Computational Linguistics **(Findings of ACL)**, 2026 [pdf]
- Rakhsha, A., Madan, K., Zhang, T., Farahmand, A.-M., and Khasahmadi, A. Majority of the Bests: Improving Best-of-N via Bootstrapping. In *Advances in Neural Information Processing Systems* ***(NeurIPS)***, 2025 [pdf]
- Lee, J.†, Rakhsha, A., Ryu, E., and Farahmand, A.-M. Deflated Dynamics Value Iteration. In Transactions on Machine Learning Research **(TMLR)**, 2025 [pdf]
- Bedaywi, M.*†, Rakhsha, A.*, and Farahmand, A.-M. PID Accelerated Temporal Difference Algorithms. In *Reinforcement Learning Conference* ***(RLC)***, 2024 [pdf]
- Rakhsha, A., Kemertas, M., Ghavamzadeh, M., and Farahmand, A.-M. Maximum Entropy Model Correction in Reinforcement Learning. In *International Conference on Learning Representations* ***(ICLR)***, 2024 [pdf]
- Rakhsha, A., Wang, A.†, Ghavamzadeh, M., and Farahmand, A.-M. Operator Splitting Value Iteration. In *Advances in Neural Information Processing Systems* ***(NeurIPS)***, 2022 [pdf]
- Rakhsha, A.*, Zhang, X.*, Zhu, X., and Singla, A. Reward poisoning in reinforcement learning: Attacks against unknown learners in unknown environments. In *NeurIPS Workshop on Learning and Decision-Making with Strategic Feedback* ***(StratML)***, 2021 [pdf]
- Rakhsha, A., Radanović, G., Devidze, R., Zhu, X., and Singla, A. Policy teaching in reinforcement learning via environment poisoning attacks. In *Journal of Machine Learning Research* ***(JMLR)***, 2021 [pdf]
- Rakhsha, A., Radanović, G., Devidze, R., Zhu, X., and Singla, A. Policy teaching via environment poisoning: training-time adversarial attacks against reinforcement learning. In *International Conference on Machine Learning (**ICML**)*, 2020 [pdf]

* Equal Contribution † Mentored Undergraduate/Master's Student Intern

Invited Talks

Theory of Reinforcement Learning Workshop, Edmonton, Canada May 2023

- Title: Operator Splitting Value Iteration [Video - 45 minutes] [Slides]
- Appeared in the RL theory seminars (advised by Csaba Szepesvári, Gergely Neu, and Ciara Pike-Burke).

Research Mentorship

Mentored three undergraduate and master's researchers on reinforcement-learning projects, leading to publications at NeurIPS 2022, TMLR 2024, and RLC 2024.

Teaching Experience

- Teaching assistant for Introduction to Artificial Intelligence, Design of Algorithms, Linear Algebra, Data Structures and Algorithms, and Probability and Statistics, 2017–2025.
- Designed assignments and course materials and instructed tutorials and discussion sections.

Technical Skills

Programming and ML: Python, PyTorch
LLM Training, Inference, and Evaluation: SkyRL, Hugging Face Transformers, vLLM, LM Evaluation Harness
Research Infrastructure: Slurm, Ray, Hydra, Weights & Biases