Fahim Tajwar

Website: https://tajwarfahim.github.io/ Email: ftajwar@cs.cmu.edu

EDUCATION

Carnegie Mellon University — **Pittsburgh, PA**
Doctor of Philosophy (PhD), Machine Learning — 2023 – Current
Advisor: Ruslan Salakhutdinov & Jeff Schneider

Stanford University — **Stanford, CA**
Master of Science (MS), Computer Science (AI/ML) — 2023
Bachelor of Science (BS) with Distinction, Mathematics — 2022

SELECTED PUBLICATIONS (* Equal Contribution)

Training a Generally Curious Agent — 2025
Fahim Tajwar*, Yiding Jiang*, Abitha Thankaraj, Sumaita Sadia Rahman, J Zico Kolter, Jeff Schneider, Ruslan Salakhutdinov
International Conference on Machine Learning (ICML), 2025 **(Spotlight)**

Preference Fine-Tuning of LLMs Should Leverage Suboptimal, On-Policy Data — 2024
Fahim Tajwar*, Anikait Singh*, Archit Sharma, Rafael Rafailov, Jeff Schneider, Tengyang Xie, Stefano Ermon, Chelsea Finn, Aviral Kumar
International Conference on Machine Learning (ICML), 2024

Conservative Prediction via Data-Driven Confidence Minimization — 2024
Caroline Choi*, Fahim Tajwar*, Yoonho Lee*, Huaxiu Yao, Ananya Kumar, Chelsea Finn
Transactions on Machine Learning Research (TMLR), 2024

Surgical Fine-Tuning Improves Adaptation to Distribution Shifts — 2023
Yoonho Lee*, Annie S. Chen*, Fahim Tajwar, Ananya Kumar, Huaxiu Yao, Percy Liang, Chelsea Finn
International Conference on Learning Representations (ICLR), 2023

When to Ask for Help: Proactive Interventions in Autonomous Reinforcement Learning — 2022
Annie Xie*, Fahim Tajwar*, Archit Sharma*, Chelsea Finn
Conference on Neural Information Processing Systems (NeurIPS), 2022

Do Deep Networks Transfer Invariances Across Classes? — 2022
Allan Zhou*, Fahim Tajwar*, Alexander Robey, Tom Knowles, George J. Pappas, Hamed Hassani, Chelsea Finn
International Conference on Learning Representations (ICLR), 2022

Scalable Deep Learning to Identify Brick Kilns and Aid Regulatory Capacity — 2021
Jihyeon Lee*, Nina R. Brooks*, Fahim Tajwar, Marshall Burke, Stefano Ermon, David B. Lobell, Debashish Biswas, Stephen P. Luby
Proceedings of the National Academy of Sciences (PNAS), 2021

INDUSTRY EXPERIENCE

Research Scientist Intern, Meta Platforms — May 2025 – August 2025
Software Engineer Intern, Meta Platforms — June 2022 – September 2022
Software Engineer Intern, Cadence Design Systems — June 2020 – September 2020

AWARDS

Top Reviewer, Conference on Neural Information Processing Systems (NeurIPS) — 2023
University Distinction, top 15% of the graduating class, Stanford University — 2022
Tau Beta Pi Engineering Honor Society — 2020
Bronze Medal, 48th International Physics Olympiad (**IPhO**), Indonesia — 2017
Bronze Medal, 47th International Physics Olympiad (**IPhO**), Switzerland Liechtenstein — 2016

TALKS & PRESENTATION

- Neural Information Processing Systems (NeurIPS) — November 2022
- International Conference on Learning Representations (ICLR) — April 2022
- ICML Workshop on Uncertainty & Robustness in Deep Learning (UDL) — July 2021
- Stanford Earth Summer Undergraduate Research (SESUR) — August 2019
- Stanford EE Research Experience for Undergraduates (REU) — August 2018

SERVICE

- Reviewer, Conference on Neural Information Processing Systems (NeurIPS) (**Top Reviewer, 2023**) — 2023, 2024, 2025
- Reviewer, International Conference on Learning Representations (ICLR) — 2024, 2025
- Reviewer, International Conference on Machine Learning (ICML) — 2024, 2025
- Reviewer, The IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR) — 2024
- Reviewer, Transactions on Machine Learning Research (TMLR) — 2024-2025
- Reviewer, International Joint Conference on Artificial Intelligence (IJCAI) — 2024