

Maitreyi Swaroop — Curriculum Vitae

Email: mswaroop@andrew.cmu.edu — Website: <https://maitreyiswaroop.github.io>

Research interests: LLM post-training; generalization and behavioral transfer under narrow finetuning and knowledge distillation; distributionally robust and causal machine learning.

EDUCATION

Carnegie Mellon University, Pittsburgh, USA August 2024 – Expected May 2029
Ph.D. Student, Machine Learning — Supervisor: [Prof. Bryan Wilder](#)
Department of Machine Learning, School of Computer Science

Indian Institute of Technology, Kharagpur, India July 2019 – May 2024
B.Sc. (Hons) + M.Sc. in Mathematics & Computing Department Rank: 2 of 63
Micro-specialization in Artificial Intelligence and Applications Overall GPA: 9.45/10.00

PUBLICATIONS

Peer-Reviewed Conference Publications

- **Distributionally Robust Feature Selection**
Maitreyi Swaroop, T. Krishnamurti, [B. Wilder](#).
NeurIPS 2025, Main Track. [[Paper](#)] [[Code](#)] [[Poster](#)]
Model-agnostic feature selection method for group-DRO, motivated by survey design for perinatal depression diagnosis.

Workshop Papers

- **interwhen: A Generalizable Framework for Verifiable Reasoning with Test-time Monitors**
V. K. Bhat, P. Chanda, A. Khandelwal, **Maitreyi Swaroop**, [V. N. Balasubramanian](#), et al.
ICLR 2026, VerifAI-2: The Second Workshop on AI Verification in the Wild. [[Paper](#)]
Framework for steering reasoning models via asynchronous test-time verifiers auto-generated from natural-language policies.
- **Learning Macro Variables with Auto-encoders**
Maitreyi Swaroop, E. Elmoznino, [D. Sridhar](#).
NeurIPS 2023 Workshop on Causal Representation Learning. [[Paper](#)] [[Code](#)]
Self-supervised method to learn macro variables and their relations, extending desiderata from Causal Feature Learning.

POSTERS / PREPRINTS

- **Representation Learning for Sample Efficient CATE Estimation** January 2026
Maitreyi Swaroop, S. Bhatt, S. Marie Rodriguez, T. Krishnamurti, [B. Wilder](#).
Poster, *Workshop on Bridging Predictions and Interventions, Simons Institute, Berkeley, CA*. [[Poster](#)][[Preprint](#)]
Representation learning for sample-efficient CATE estimation from multi-outcome administrative data.
- **What Must Generalist Agents Remember?**
K. Yamin, N. Deka, **Maitreyi Swaroop**, A. Ting, J. Schneider, [B. Wilder](#). [[Preprint](#)]
Uniformly near-optimal generalist agents encode domain-distinguishing information in memory. Such memory suffices to decode local transition dynamics.

RESEARCH EXPERIENCE

Research Internship — PIs: [Amit Sharma](#), [Vineeth Balasubramanian](#) June 2025 – August 2025
Microsoft Research, Bengaluru, India
• Developed monitor-guided reasoning methods for verifiable mathematical and symbolic reasoning in language models.

Research Internship — PI: [Prof. Dhanya Sridhar](#) May 2023 – May 2024
Mila, Quebec AI Institute, Montreal, Canada
• Developed the DeepCFL framework for learning macro variables in causal representation learning.

Master's Thesis — PI: [Prof. Partha Pratim Chakrabarti](#) May 2023 – May 2024
Indian Institute of Technology, Kharagpur, India
• Developed hybrid methods combining approximation algorithms, heuristics, and RL approaches for routing problems.
Awarded Best Masters' Thesis Project (2024).

Research Internship — PI: [Dr. Manuel Gomez Rodriguez](#) May 2022 – July 2022
Max Planck Institute for Software Systems, Kaiserslautern, Germany
• Threshold test to identify biased decision-makers using DPMMs and MCMC inference. [[Code](#)]

TEACHING

Teaching Assistant, 10-726: Machine Learning Theory (Course Instructor: Prof. Nina Balcan) Fall 2026
Carnegie Mellon University

AWARDS & HONORS

-
- **Best Project Work** – Awarded for the best Master's Thesis Project, Dept. of Mathematics & Computing, 2024.
 - **Indian Academy of Sciences Summer Research Fellowship (IASc-INSa-NASI)** – Awarded in Summer 2021.