Sujai Hiremath

sujai1.github.io
sh2583@cornell.edu
Google Scholar
LinkedIn

TL;DR

- ORIE PhD at Cornell Tech (Exp. 27') working on causal RL and agentic security.
- Published 4 first-author papers (NeurIPS 24', UAI 25', NeurIPS 25', AISTATS 26') on efficient causal learning methods within 2 years of starting research.
- Interned at Amazon Research Tübingen, soon at Amazon Research San Francisco; (1) LLM-aided discovery, (2) causal RL/LLM post-training, (3) RCA for agentic security.

Education

Cornell Tech | New York, NY 2023 - 2027
PhD in Operations Research and Information Engineering | GPA: 3.9 *(expected)*
- Areas: Causal Inference, Reinforcement Learning, LLMs

California Institute of Technology | Pasadena, CA 2019 - 2023
BS in Applied and Computational Mathematics | GPA: 4.0
- Areas: Machine Learning, Mathematical Modelling, Deep Learning

Work Experience

Applied Scientist Intern | Amazon Research Tübingen, Germany 06.2025 - 11.2025
- Managers: Dr. Dominik Janzing, Dr. Shiva Kasiviswanathan, Dr. Elke Kirschbaum.
- Developed a method leveraging LLMs as unreliable experts to improve causal learning in finite samples. Validated theoretical results in Python experiments.
- Currently developing a causal reinforcement learning approach for sample-efficient training of LLMs in low-data/low-compute regimes. Validating theory with Python.

PhD Student Researcher | Cornell Tech 11.2023 - Present
- PIs: Dr. Kyra Gan, Dr. Promit Ghosal.
- Leveraged diffusion models, independence tests, and nonparametric regression for causal inference. Validated theory in experiments in Python (PyTorch, scikit-learn).
- Published 3 first-author papers at NeurIPS (2024, 2025) and UAI (2025) on improving finite-sample causal structure learning while relaxing assumptions.

Publications and Preprints

1. **Hiremath, S.***, et al. From Detection to Attribution: Identifying Malicious Documents in Prompt Injection Attacks on LLM Agents via Root Cause Analysis. *preprint*, 2026.
2. **Hiremath, S.***, et al. From Causal Structure to Efficient Representations: Deep Reinforcement Learning with Causal Rank Regularization. *preprint*, 2065.
3. **Hiremath, S.***, et al. From Guess2Graph: When and How Can Unreliable Experts Safely Boost Causal Discovery in Finite Samples? *AISTATS*, 2026.
4. Meier, D.* and **Hiremath, S.***, et al. When Additive Noise Meets Unobserved Mediators: Bivariate Denoising Diffusion for Causal Discovery. *NeurIPS*, 2025.
5. **Hiremath, S.***, et al. LoSAM: Local Search in Additive Noise Models with Mixed Mechanisms and General Noise for Global Causal Discovery. *UAI*, 2025.
6. **Hiremath, S.***, et al. Hybrid Top-Down Global Causal Discovery with Local Search for Linear and Nonlinear Additive Noise Models. *NeurIPS*, 2024.

Service & Awards

Service: Reviewer for NeurIPS 25', ICLR 25', AISTATS 25', CLeaR 26', UAI 26'.
Awards: NeurIPS Top Reviewer 25' | Cornell Fellowship 23'.