Pragya Srivastava

PhD Student in Computer Science | The University of Texas at Austin

@ pragya8srivastava@gmail.com GitHub Google Scholar +91-7617478783

Education

The University of Texas at Austin **2026 – Present**
PhD in Computer Science | Advisor: Prof. Amy Zhang

Indian Institute of Technology Delhi (IIT Delhi) **July 2019 – May 2023**
B.Tech in Engineering Physics; GPA: **8.7/10.0**

Research Interests

Reinforcement learning, with an emphasis on offline and model-based RL, long-horizon value learning and generalization; reliable learning and alignment of large language models.

Research Experience

The University of Texas at Austin **2026 – Present**
Graduate Research Assistant | Advisor: Prof. Amy Zhang
Project: Horizon-Scalable Q-Learning [**S.1**]

Google DeepMind **July 2024 – 2026**
Pre-doctoral Researcher | Advisors: Dr. Aravindan Raghuveer, Dr. Karthikeyan Shanmugam, Prof. Doina Precup
Projects: Causally Robust Reward Modeling [**C.5**], Multi-objective LLM Alignment, Sample-Efficient RL for Post-training Gemini Models, Self-Verification via Unit-Test Generation for SWE

The University of Texas at Austin **March 2024 – June 2024**
Research Intern | Advisors: Peihao Wang, Prof. Atlas Wang
Project: Theoretical Analysis of Bottlenecks in Structured State Space Models [**C.3**]

Microsoft Research, India **May 2022 – February 2024**
Research Intern | Advisors: Dr. Amit Sharma, Dr. Amit Deshpande, Tanuja Ganu
Project: Multi-hop reasoning over semi-structured data [**C.1**], In-context Learning [**C.2**], Active Learning for Reward Models [**C.4**]

ShareChat AI, India **February 2023 – May 2023**
Machine Learning Engineer Intern | Advisor: Srijan Saket
Project: Multimodal Feature Encoders for Enhanced *Early-Stage* Video Recommendations

Selected Publications

S=Preprint, C=Conf./Workshop, *=Equal Contribution

[S.1] **Horizon-Scalable Q-Learning**
Haoran Xu, Pragya Srivastava, Xinjie Liu, Siddhant Agarwal, Amy Zhang
arXiv preprint **[Preprint'26]**

[C.6] **Benchmarking and Improving Monitors for Out-Of-Distribution Alignment Failure in LLMs**
Dylan Feng*, Pragya Srivastava*, Anca Dragan, Cassidy Laidlaw
Evaluations & Datasets Track **[NeurIPS'26]**

[C.5] **Robust Reward Modeling via Causal Rubrics**
Pragya Srivastava*, Harman Singh*, Rahul Madhavan*, Gandharv Patil, Sravanti Addepalli, Arun Suggala, Rengarajan Aravamudhan, Soumya Sharma, Anirban Laha, Aravindan Raghuveer, Karthikeyan Shanmugam, Doina Precup
International Conference on Learning Representations **[ICLR'26]**

[C.4] **Outlier-Aware Preference Optimization for Large Language Models**
Pragya Srivastava , Sai Soumya Nalli, Amit Deshpande, Amit Sharma
Bidirectional Human-AI Alignment **[ICLR'25 Bi-Align Workshop]**

[C.3] **Understanding and Mitigating Bottlenecks of State Space Models through the Lens of Recency and Over-smoothing**
Peihao Wang, Ruisi Cai, Yuehao Wang, Jiajun Zhu, Pragya Srivastava, Zhangyang Wang, Pan Li
International Conference on Learning Representations
First-author workshop version at ICML 2024 NGSM Workshop **[ICLR'25]**

[C.2] **NICE: To Optimize In-Context Examples or Not?**
Pragya Srivastava, Satvik Golechha, Amit Deshpande, Amit Sharma
Annual Conference of the Association for Computational Linguistics **[ACL'24]**

[C.1] **Evaluating LLMs' Mathematical Reasoning in Financial Document Question Answering**
Pragya Srivastava, Manuj Malik, Vivek Gupta, Tanuja Ganu, Dan Roth
Annual Conference of the Association for Computational Linguistics **[ACL'24 Findings]**

Selected Projects

Horizon-Scalable Q-Learning (HQL) [S.1] 2026

Advisor: Prof. Amy Zhang

UT Austin

- Co-developed a horizon-scalable offline RL method that adaptively selects between one-step and multi-step TD targets to balance error accumulation and off-policy bias.
- The method uses critic-based, density-free off-policy correction, making it compatible with expressive behavior policies such as flow and diffusion models.
- Demonstrated improvements over prior horizon-reduction methods on long-horizon OGBench and OCBench tasks, and evaluated the effect of horizon length on controlled combination-lock tasks.

Self-Verification via Unit-Test Generation for SWE Problems 2026

Advisors: Dr. Rishi Saket, Shagun Sodhani

Google DeepMind

- Curated high-quality training datasets for unit-test generation (TestGen) on software engineering (SWE) problems.
- Designed reward functions for RL post-training to improve the effectiveness of generated unit tests.
- Integrated the curated datasets into the existing Gemini post-training pipeline, achieving **performance improvements on internal Gemini Thinking leaderboards**.

Robust Scalable Oversight for LLMs [C.6] June 2025 – September 2025

Advisors: Cassidy Laidlaw, Prof. Anca Dragan

UC Berkeley

- Co-developed MOOD, a benchmark for evaluating LLM safety monitors on previously unseen alignment failure modes.
- Studied hybrid monitoring pipelines combining guard models with out-of-distribution detectors, improving recall from **39% to 45%**.
- Demonstrated that augmenting guard models with OOD detection can deliver larger recall gains than scaling standalone guard models to 20 times as many parameters.

Robust Reward Modeling for Mitigating Reward Hacking February 2025 – July 2025

Advisors: Dr. Karthikeyan Shanmugam, Prof. Doina Precup

Google DeepMind

- Proposed a novel data-augmentation technique grounded in an explicit causal model to mitigate reward hacking, directly addressing failure modes in standard RLHF pipelines.
- Demonstrated **state-of-the-art performance** on RewardBench, outperforming existing baselines while achieving significant gains in downstream alignment tasks including Iterative On-Policy DPO and Best-of-N sampling.
- Built a modular framework that standardizes data processing and training loops across diverse model families.

Multi-Objective LLM Alignment December 2024 – February 2025

Advisors: Dr. Karthikeyan Shanmugam

Google DeepMind

- Proposed a theoretically grounded objective utilizing the Nash Social Welfare function to solve the problem of alignment under conflicting objectives, ensuring Pareto-optimal trade-offs.
- Led the end-to-end implementation of the proposed framework for Gemini Post-training, achieving superior convergence stability and **6%** metric improvements in complex domains like Math, Factuality, Safety and Tool Use.

Principled Curriculum Design for Post-training July 2025 – 2026

Advisors: Dr. Karthikeyan Shanmugam, Prof. Doina Precup

Google DeepMind

- Proposed a novel difficulty curriculum for sampling prompts with high learning signal during RL training to accelerate the training progress of a KL-constrained policy.
- Achieved $3\times$ **training efficiency (FLOPs)** on smaller models and **ranked #1 on internal downstream benchmarks, showing significant promise for scaling to larger model sizes as well**.

Natural Language Feedback for Efficient Policy Exploration July 2025 – 2026

Advisor: Dr. Aravindan Raghuveer

Google DeepMind

- Proposed a novel technique for incorporating natural language feedback to enable efficient exploration of the solution space.
- Tested various intervention methods to optimize the trade-off between guided training speed and final model autonomy at test time.

Simulating the Dynamics of Hamiltonian Systems Aug 2022 – Nov 2022

Advisor: Prof. Vishal K. Vaibhav

B.Tech Project, IIT Delhi

- Developed a novel algorithm to infer the energy of a Hamiltonian system and predict the trajectory of a physical system.
- Proposed a novel architecture **Recurrent Hamiltonian Neural Networks** for predicting the trajectory of physical systems with high fidelity by leveraging the **volume-preserving** property of symplectic transformations.

Few-Shot Transfer Learning for Remote-Sensing Images Dec 2021 – April 2022

Advisor: Prof. Stefano Ermon

Stanford University

- Worked on few-shot domain adapting a model trained on **fMoW** images to **fMoW-Sentinel** images.
- Experimented with feature transformation from target to source domain using **Pix2Pix** and **Style-Transfer** methods.

Domain- and Class-Specific Feature Decomposition Aug 2021 – March 2022

Advisor: Prof. Pengtao Xie

UC, San Diego

- Leveraged Differentiable Neural Architecture Search (**DARTS**) for learning to decompose the representation of an input image into its domain-specific and class-specific representations.
- Proposed a **multi-level optimization** problem for adapting a model to a target domain during inference time.

Cross-Lingual Few Shot Transfer for Natural Language Inference NLP Course Project

Advisor: Prof. Mausam

IIT Delhi

- Fine-tuned multi-lingual pre-trained models (mBERT, XLM-R) for the downstream task of inferring if one statement is an entailment, contradiction or neutral with respect to the other statement.

Named Entity Recognition in Chemical Procedures NLP Course Project

Advisor: Prof. Mausam

IIT Delhi

- Built neural networks for sequence labelling for the purpose of extracting named entities such as reagent, action, amount, temperature, etc.
- Benchmarked **BiLSTM-CRF with multi-head self-attention** and pre-trained language models (BERT) on the provided dataset.

Selected Accolades

One of 12 out of 20,000 applicants selected for the Google Pre-doctoral Program (2024–2026).

All India Rank in top 0.2% in JEE-Main among 1.5 million candidates.

All India Rank in top 2% in JEE-Advanced among 150,000 candidates.

Skills

Languages	C++, Python, Java, Bash, MATLAB
Libraries	PyTorch, TensorFlow, Keras, Hugging Face, NLTK, OpenCV, Rasa, spaCy, LangChain
Tools	Kubernetes, Docker, SQL (BigQuery), Weights & Biases, Optuna, Ray Tune, Git, HPC (Slurm), LaTeX

References

- Prof. Doina Precup ..... *Associate Professor, McGill University*
- Dr. Karthikeyan Shanmugam ..... *Senior Staff Research Scientist, Google DeepMind*
- Prof. Atlas Wang ..... *Associate Professor, The University of Texas at Austin*
- Dr. Amit Sharma ..... *Principal Researcher, Microsoft Research India*
- Dr. Amit Deshpande ..... *Principal Researcher, Microsoft Research India*
- Prof. Vishal K. Vaibhav ..... *Assistant Professor, IIT Delhi*
- Prof. Stefano Ermon ..... *Associate Professor, Stanford University*