

# Zaiyan Xu

---

<b>Contact Information</b>	Graduate Research Assistant Department of Electrical and Computer Engineering Texas A&M University	Email: <a href="mailto:zxu43@tamu.edu">zxu43@tamu.edu</a> Webpage: <a href="https://www.zaiyanxu.com">https://www.zaiyanxu.com</a> GitHub: <a href="https://github.com/zaiyan-x">https://github.com/zaiyan-x</a>
<b>Research Interests</b>	Reinforcement Learning, Distributionally Robust Optimization, Reinforcement Learning from Human Feedback, Large Language Model (LLM) Alignment	
<b>Education</b>	<b>Texas A&amp;M University</b> , College Station, TX Ph.D. in Computer Engineering Advisor: Dr. Dileep Kalathil	Aug. 2020 - Present Anticipated graduation Jun. 2026 GPA: 3.86
	<b>University of Illinois at Urbana-Champaign</b> , IL B.S. in Statistics & Computer Science and Actuarial Science Cum Laude, Highest Distinction in CS and Statistics, High Distinction in Actuarial Science	Aug. 2015 - Jul. 2020 GPA: 3.92
<b>Honors and Achievements</b>	<ul style="list-style-type: none"><li>▪ NeurIPS 2023 Top Reviewer</li><li>▪ Dept. of Electrical and Computer Engineering Graduate Merit Fellowship, TAMU, 2020</li><li>▪ Willis Towers Watson Actuarial Science Scholarship, Dept. of Mathematics, UIUC, 2018</li></ul>	
<b>Work Experience</b>	<b>Amazon AGI</b> , New York, NY Applied Scientist Intern Developed a compute-aware framework to enable fair, matched-budget comparisons of RL post-training methods for reasoning large language models by modeling learning curves in total roll-outs, standardizing training horizon and global batch size across model scales, and evaluating multiple methods on GSM8K and Eurys validation splits.	Aug. 2025 - Present
	<b>Mitsubishi Electric Research Laboratories</b> , Cambridge, MA Research Intern Developed a distributionally robust reinforcement learning method with explicit safety constraints. Proposed a Lyapunov-based characterization of cost functions that satisfy both safety and robustness requirements.	May 2023 - Aug. 2023
	<b>National Center for Supercomputing Applications</b> , Champaign, IL Undergraduate Researcher (NCSA SPIN Program) Worked on speech recognition and auto-captioning with a focus on engineering lectures. Developed several wrappers for CMU Sphinx engine and streamlined the model training process by automating audio slicing, caption partitioning.	Jun. 2019 - May 2020
<b>Publications</b>	<b>Distributionally Robust Large Language Model Finetuning</b> 1. Zaiyan Xu, Sushil Vemuri, Kishan Panaganti, Dileep Kalathil, Rahul Jain, Deepak Ramachandran. "Robust LLM alignment via distributionally robust direct preference optimization", in <i>NeurIPS 2025</i> . <b>Sample-efficient Robust Reinforcement Learning</b> 2. Kishan Panaganti, Zaiyan Xu, Dileep Kalathil, Mohammad Ghavamzadeh. "Bridging Distributionally Robust Learning and Offline RL: An Approach to Mitigate Distribution Shift and Partial Data Coverage", in <i>7th Annual Learning for Dynamics &amp; Control Conference (L4DC)</i> , 2025. 3. Zaiyan Xu*, Kishan Panaganti*, Dileep Kalathil. "Improved Sample Complexity Bounds For Distributionally Robust Reinforcement Learning", in <i>International Conference on Artificial Intelligence and Statistics (AISTATS)</i> , 2023.	

#### **Robust Reinforcement Learning with Neural Network Function Approximation**

4. Kishan Panaganti, Zaiyan Xu, Dileep Kalathil, Mohammad Ghavamzadeh. "Robust Reinforcement Learning Using Offline Data", in *NeurIPS 2022*.

#### **Sample-efficient Distributionally Robust Imitation Learning**

5. Kishan Panaganti\*, Zaiyan Xu\*, Dileep Kalathil. "Distributionally Robust Behavioral Cloning for Robust Imitation Learning", in *the 62nd IEEE Conference on Decision and Control (CDC)*, 2023.

#### **Reinforcement Learning for Hardware Security**

6. Chen Chen, Zaiyan Xu, Mohamadreza Rostami, David Liu, Dileep Kalathil, Ahmad-Reza Sadeghi, Jeyavijayan Rajendran. "ReFuzz: Reusing Tests for Processor Fuzzing with Contextual Bandits", in *NDSS 2026*.

(\* denotes equal contribution)

#### **Skills**

**Languages and Platforms:** Python, C, C++, R, SQL, PyTorch, Ray, Hugging Face TRL, OpenRLHF, verl, Gymnasium, MuJoCo, CVXPY

**Cloud Services:** Amazon AWS, Google Compute Engine (GCE)

#### **Summary of Selected Research**

##### **Distributionally Robust Large Language Model Finetuning** 2024-2025

Contributed to the development of two novel distributionally robust direct preference optimization algorithms, Wasserstein DPO (WDPO) and Kullback-Leibler DPO (KLDPO), to address preference distribution shift in LLM alignment. These methods leverage principled minimax approaches with scalable gradient descent-style learning, ensuring robust performance under shifting user preferences across diverse regions, demographics, and cultural trends. We provided formal sample complexity guarantees and demonstrated substantial alignment improvements in empirical experiments on LLaMA-3.2-1B/3B and LLaMA-3.1-8B models.

##### **Sample-efficient Robust Reinforcement Learning** 2023

Contributed to the development of Robust Phased Value Learning (RPVL), a distributionally robust RL algorithm designed for tabular episodic learning and capable of handling mismatch between training and testing environments. Our method achieves an  $\tilde{O}(|S||A|H^5)$  sample complexity, improving upon existing bounds by a factor of  $|S|$ , and supports multiple divergence-based uncertainty sets (including total variation, chi-square, KL, and Wasserstein).

##### **Robust Reinforcement Learning with Neural Network Function Approximation** 2022

Implemented a distributionally robust reinforcement learning algorithm to systematically handle uncertainties due to distributional shifts in the MDP transition model between training and testing environments. Designed a neural network architecture using PyTorch to explicitly encode the dual variables arising from the distributionally robust optimization formulation. This architecture was trained simultaneously with the robust Q-function, eliminating the need for external optimization solvers. Empirically validated the approach on MuJoCo benchmarks, demonstrating strong robustness and superior performance of the proposed Robust Fitted Q-Iteration (RFQI) algorithm.

#### **Professional Services**

Conference reviewer: AAAI (2025), ICLR (2024, 2025), NeurIPS (2023 **Top Reviewer**, 2024), ICML (2023, 2024, 2025), AISTATS (2023, 2024), American Control Conference (2023), IEEE Conference on Decision and Control (2023, 2024, 2025), L4DC (2023, 2024, 2025)

#### **References**

Dr. Dileep Kalathil  
Dept. of Electrical and Computer Engineering  
Texas A&M University, College Station, TX  
Email: dileep.kalathil@tamu.edu