Frontier of Artificial Network

Department of Data Science
香港城市大學
City University of Hong Kong

# Towards Principled Reinforcement Learning: From Statistical Complexity to Representation Complexity


**Invited Speaker**
**Han Zhong**
***Peking University***


**Date:** Apr 29, 2026 (Wed)
**Time:** 15:30 (HKT)
**Zoom Meeting:** 801 137 0362

## Biography

Han Zhong is a Ph.D. student at Peking University. His research focuses on reinforcement learning and its connections to operations research, statistics, and optimization. He has published papers in leading journals and conferences, including Mathematics of Operations Research, Journal of the American Statistical Association, Journal of Machine Learning Research, ICML, NeurIPS, and ICLR.

## Abstract


Designing efficient RL algorithms requires addressing two key dimensions. The first is statistical complexity — how many samples do we need to learn a good policy? We propose a unified framework called the Generalized Eluder Coefficient that captures the sample efficiency of both model-based and model-free RL under general function approximation. This framework also extends naturally to preference-based learning for aligning large language models, leading to practical algorithms like Iterative DPO and Self-Exploring LM. The second, less explored dimension is representation complexity — what should we learn? We show that approximating the model, policy, and value functions in RL has fundamentally different difficulty levels, forming a strict hierarchy rooted in circuit complexity theory. In particular, value functions are the hardest to represent, which explains why discriminative critics in PPO-style methods struggle in long-horizon LLM reasoning tasks. Motivated by this finding, we propose Generative Actor-Critic, which replaces the scalar critic with a generative critic that reasons step-by-step before assigning credit. Experiments show it is more scalable, more robust, and achieves better performance than both value-free methods like GRPO and traditional PPO.