Haoyang Hong

Ph.D. Student, Oregon State University

Reinforcement Learning Theory · Bandits · Statistical Inference

honghao@oregonstate.edu theorymagic.github.io GitHub Google Scholar

My research focuses on the **theoretical foundations** of sequential decision-making, including **KL-regularization**, **reward hacking**, and **statistical inference** under adaptive data collection.

Education

Oregon State University Sep. 2024 – Sep. 2029 *(expected)*
Ph.D. in Artificial Intelligence Advisor: Huazheng Wang

University of Science and Technology of China Sep. 2020 – Jul. 2024
B.S. in Statistics, School of the Gifted Young

Publications

*Equal contribution.

Online KL-Regularized Reinforcement Learning with Function Approximation under Misspecification
Haoyang Hong*, Zichen Wang*, Quanquan Gu, Huazheng Wang
RLC 2026

Studies how **model misspecification** affects online reinforcement learning with KL regularization and function approximation.

When Can You Poison Rewards? A Tight Characterization of Reward Poisoning in Linear MDPs
Jose Efraim Aguilar Escamilla*, **Haoyang Hong***, Jiawei Li*, Haoyu Zhao, Xuezhou Zhang, Sanghyun Hong, Huazheng Wang
ICML 2026

Studies when **reward perturbations** can induce a target policy in linear MDPs.

Beyond Static Bias: Adaptive Multi-Fidelity Bandits with Improving Proxies
Muyun Lu*, **Haoyang Hong***, Huazheng Wang, Ying Lin
arXiv:2605.08558, 2026

Studies **evaluation allocation** across fidelity levels when low-cost proxy feedback improves with repeated use.

Design-Based Bandits under Network Interference: A Trade-off between Regret and Inference
Zichen Wang*, **Haoyang Hong***, Chuanhao Li, Haoxuan Li, Zhiheng Zhang, Huazheng Wang
NeurIPS 2025

Studies the **regret–inference trade-off** when one individual's treatment can affect others' outcomes.

Research Experience

Microsoft Research Asia (MSRA) Jul. – Oct. 2026
Research Intern Mentor: Wei Chen

- Investigated extensions of existing **reward hacking theory**, focusing on how optimizing a proxy reward affects the true reward and how additional feedback can guide proxy refinement.
- Studied **adaptive reward proxy refinement** using feedback from more capable LLMs, aiming to reduce supervision costs during training while identifying when further evaluation is needed.
- Conducted preliminary experiments with **coding agents and robotics tasks** to assess proxy refinement, discrepancies between proxy and true rewards, and evaluation cost.