

# YIHAO SUN

(+86)13060070726 • [sunyh@lamda.nju.edu.cn](mailto:sunyh@lamda.nju.edu.cn) • [www.yihaosun.cn](http://www.yihaosun.cn)

## EDUCATION

### M.Sc. in Computer Science and Technology

Nanjing University, Jiangsu, China  
LAMDA Group, School of Artificial Intelligence

Sep 2021 - Expected Jun 2024

### B.Eng. in Software Engineering

Sichuan University, Sichuan, China  
School of Software Engineering

Sep 2017 - Jun 2021

## RESEARCH INTERESTS

My research interest is reinforcement learning (RL). Currently, I focus on offline reinforcement learning (Offline RL) and model-based reinforcement learning (MBRL). I am committed to improving the sample efficiency of online RL, deriving a performant policy from offline datasets, and learning a faithful environment model that enables reliable offline policy evaluation.

## PUBLICATIONS

### Model-Bellman Inconsistency for Model-based Offline Reinforcement learning. ([pdf](#)|[code](#))

ICML'23

Co-first author.

- Proposing a novel uncertainty quantification for uncertainty-driven model-based offline RL, which can better estimate the ideal uncertainty quantification, i.e., the Bellman error.
- Achieving SOTA on 20 of 27 evaluated benchmark datasets. Compared with MOPO, a representative model-based offline RL algorithm, the performance of our algorithm is improved by 42.4%.

### Model-based Reinforcement Learning with Multi-step Plan Value Estimation. ([pdf](#)|[code](#))

ECAI'23 (Oral)

Co-first author.

- Proposing multi-step plan value to mitigate the influence of compounding error.
- Showing a better utilization of the learned model and achieving a better sample efficiency than previous model-based RL approaches.

### Episodic Return Decomposition by Difference of Implicitly Assigned Sub-Trajectory Reward. ([pdf](#)|[code](#))

AAAI'24

Fourth author.

- Proposing a novel episodic return decomposition method called Diaster (Difference of implicitly assigned sub-trajectory reward).
- Experimental results show that our method outperforms previous methods in terms of both sample efficiency and performance, achieving SOTA on 6 of 7 tasks.

### Flow-to-Better: Offline preference-based reinforcement learning via preferred trajectory generation. ([pdf](#))

ICLR'24

Co-first author.

- Proposing a novel diffusion-based framework for offline PbRL, i.e., Flow-to-Better(FTB), which uses a trajectory diffuser to achieve trajectory-wise improvement, avoiding TD learning with inaccurate rewards.
- Outperforming previous offline PbRL methods by a large margin and achieving SOTA on 6 of 9 evaluated benchmark datasets.

## OPEN SOURCE PROJECTS

### OfflineRL-Kit: An elegant PyTorch offline reinforcement learning library. ([code](#))

180+ stars

95% code contribution.

- An elegant offline reinforcement learning library based on pure PyTorch.
- State-of-the-art offline RL algorithms, including model-free and model-based approaches.
- The first reproduction that matches or even exceeds the performance in the original papers of model-based offline RL.

**AWARDS AND HONORS**

---

**Xiaomi Outstanding Scholarship (10 recipients schoolwide)**

Sep 2023

**First Prize Scholarship in Nanjing University**

Nov 2021

**First Prize Scholarship in Sichuan University**

Nov 2018

**Excellent Student in Sichuan University**

Nov 2018