Ruizhe Shi

ruizshi03@gmail.com srzer

Education

Tsinghua University, Beijing, China **Sept. 2021 – Present**

*Undergraduate at Special Pilot Class in Computer Science (**Yao class**)*

- Major: *Computer Science and Technology* *GPA 3.93/4.00*
- Minor: *Chinese Language and Literature* *GPA 4.00/4.00*

University of Washington, Seattle, USA **Feb. 2024 – July 2024**

Research Visitor at Paul G. Allen School of Computer Science

Publications & Manuscripts

(∗ *indicates equal contribution.*)

[5] **(Manuscript)** Ruizhe Shi*, Runlong Zhou*, Simon S. Du. "The Crucial Role of Samplers in Online Direct Preference Optimization".[link]

[4] **(NeurIPS 2024)** Ruizhe Shi, Yifang Chen, Yushi Hu, Alisa Liu, Hannaneh Hajishirzi, Noah A. Smith, Simon S. Du. "Decoding-Time Language Model Alignment with Multiple Objectives".[link]

[3] **(ICML 2024)** Chenhao Lu, Ruizhe Shi*, Yuyao Liu*, Kaizhe Hu, Simon S. Du, Huazhe Xu. "Rethinking Transformers in Solving POMDPs".[link]

[2] **(ICLR 2024)** Ruizhe Shi*, Yuyao Liu*, Yanjie Ze, Simon S. Du, Huazhe Xu. "Unleashing the Power of Pre-trained Language Models for Offline Reinforcement Learning".[link]

[1] **(NeurIPS 2023)** Yanjie Ze, Yuyao Liu*, Ruizhe Shi*, Jiaxin Qin, Zhecheng Yuan, Jiashun Wang, Huazhe Xu. "H-InDex: Visual Reinforcement Learning with Hand-Informed Representations for Dexterous Manipulation".[link]

Selected Research Experiences

Optimization Theory of Online DPO **May 2024 – Sept. 2024**

Supervised by Prof. Simon S. Du *CSE, University of Washington*

- We study convergence rates of (online) DPO from optimization perspective, and show the impact of samplers through a theoretical separation and empirical experiments. Under review.
- We provide a rigorous analysis of DPO's convergence rates with different sampling strategies under the exact gradient setting, revealing a surprising separation: uniform sampling achieves linear convergence, while our proposed online sampler achieves quadratic convergence. We further adapt the sampler to practical settings by incorporating posterior distributions and logit mixing, demonstrating significant improvements over previous approaches.

Multi-Objective Language Model Alignment **Dec. 2023 – May 2024**

Supervised by Prof. Simon S. Du *CSE, University of Washington*

- We propose a training-free, simple yet effective decoding-time algorithm for multi-objective alignment of language models, with optimality guarantees. First-authored work accepted by **NeurIPS 2024**.
- We exploit a common form among a family of f-divergence regularized alignment approaches (such as PPO, DPO, and their variants) to identify a closed-form solution by Legendre transform, and derive an efficient decoding strategy. Theoretically, we show why existing approaches can be sub-optimal even in natural settings and obtain optimality guarantees for our method.

Representation Theory of Transformer in Decision-Making **Nov. 2023 – Jan. 2024**

Supervised by Prof. Huazhe Xu *IIIS, Tsinghua University*

- We challenge the common wisdom and prove theoretically and empirically that Transformers are not suitable for Partially Observable RL, while advocating Linear RNN as a promising alternative. Second-authored work accepted by **ICML 2024**.
- We establish that regular languages, which Transformers struggle to model, are reducible to POMDPs. This poses a significant challenge for Transformers in learning POMDP-specific inductive biases, due to their lack of inherent recurrence found in other models like RNNs, with empirical results highlighting the sub-optimal performance of the Transformer and considerable strength of LRU.

Training Language Model for Decision-Making **June 2023 – Sept. 2023**

Supervised by Prof. Huazhe Xu *IIIS, Tsinghua University*

- We leverage the power of pre-trained Language Models for low-level motion control in offline reinforcement learning. First-authored work accepted by **ICLR 2024**.
- We demonstrate the superiority of LaMo over DT-based and value-based offline RL algorithms. Specifically, we find that LaMo could successfully handle the challenging low-data regime while DT could not. This highlights the great potential of our cross-domain pre-training for sequential modeling.

Selected Awards & Honors

National Scholarship **Oct. 2024**
Ministry of Education of China *top scholarship; 0.2% domestically*

Yao Award (Silver Medal) **Sept. 2024**
IIIS, Tsinghua University *top scholarship; 3 students institute-wide*

Jiang Nanxiang Scholarship **Nov. 2023**
Tsinghua University *top scholarship; 1 student institute-wide*

China National Endeavor Scholarship **Oct. 2022**
Beijing Education Bureau *1 student institute-wide*

First Prize in National High School's Mathematics Competition of China **Oct. 2020**
Chinese Mathematical Society *top 20 province-wide*

Service

Conference Reviewer
NeurIPS 2024, ICLR 2025, AISTATS 2025

Yao Class Seminar Co-organizer **2024 Fall – Present**
Host weekly research seminars [link] *Tsinghua University*

Teaching Assistant **2024 Fall**
Natural Language Processing *Tsinghua University*

Voluntary Drop-in Tutoring **Oct. 2022 – July 2024**
Tutor freshmen in STEM Courses *Tsinghua University*

*I have **157** hours of officially recorded volunteering work.*

Selected Courses

Mathematics and Theory: Calculus (**A**$^+$), Linear Algebra (**A**), Abstract Algebra (**A**), Introduction to Complex Analysis (**A**), Probability and Statistics (**A**), Basic Topology (**93**), Introduction to Optimization (**A**), Theory of Computation (**A**), Physics of Information (**A**);
Programming and AI: Introduction to Programming in C/C++ (**A**$^+$), Intelligent Unmanned System (**A**$^+$), Type-safe Modern System Practice (**A**), Machine Learning (**A**), Artificial Intelligence: Principles and Techniques (**A**), Natural Language Processing (**A**).

Technical Skills

Programming Skills: Python, C/C++, LaTeX, Bash, Scala, Matlab.
Language Skills: Chinese Mandarin (native), English (CET-6, TOEFL 104 [R30/L26/S23/W25], GRE 327 [V157/Q170]).