(814)-699-0939
rzh5514@psu.edu
Ruiquan5514.github.io

Ruiquan Huang

Research Interests

My research interests lie in broad areas of machine learning, ranging from reinforcement learning and bandits to statistical learning theory and deep learning. The overarching theme of his research is developing provable yet practically *efficient and trustworthy* machine learning algorithms and advancing the research frontier. His current research topics can be largely grouped into five aspects of efficiency and trustworthiness in ML: (1) **statistical efficiency**, (2) **computational efficiency**, (3) **safety**, (4) **privacy**, and (5) **robustness**. These aspects are within the contexts of reinforcement learning, federated online learning, and theoretical foundations of Large Language Models (LLMs).

Education

Ph.D., Electrical Engineering, The Pennsylvania State University (PSU) **Aug 2020 — May 2025**
M.S., Applied Mathematics, Columbia University **Aug 2018 — Dec 2019**
B.S., Applied Mathematics, University of Science and Technology of China **Aug 2014 — Jun 2018**

Publications (Google Scholar Profile)

[1] Non-asymptotic Convergence of Training Transformers for Next-token Prediction
R. Huang, Y. Liang, J. Yang, Conference on Neural Information Processing Systems (NeurIPS), 2024.

[2] Federated Online Prediction from Experts with Differential Privacy: Separations and Regret Speed-ups
F. Gao, **R. Huang**, J. Yang, Conference on Neural Information Processing Systems (NeurIPS), 2024.

[3] Provably Efficient UCB-type Algorithms For Learning Predictive State Representations
R. Huang, Y. Liang, J. Yang, International Conference on Learning Representations (ICLR), 2024. https://arxiv.org/abs/2307.00405

[4] Provable Benefits of Multi-task RL under Non-Markovian Decision Making Processes
R. Huang, Y. Cheng, J. Yang, V. Tan, Y. Liang, International Conference on Learning Representations(ICLR), 2024. https://arxiv.org/abs/2310.13550

[5] Temporal-Distributed Backdoor Attack Against Video Based Action Recognition
X. Li, S. Wang, **R. Huang**, M. Gowda, G. Kesidis, 38th AAAI Conference on Artificial Intelligence (AAAI), 2024. https://arxiv.org/abs/2308.11070

[6] Information-Theoretic Methods for Trustworthy and Reliable Machine Learning
S. Feng, M. Yin, **R. Huang**, Y. Wang, J. Yang, Y. Liang, IEEE Journal on Selected Areas in Information Theory, 2024.

[7] Federated Linear Contextual Bandits with User-level Differential Privacy
R. Huang, H. Zhang, M. Hejazinia, L. Melis, M. Shen, J. Yang, International Conference on Machine Learning (ICML), 2023.

[8] Near-optimal Conservative Exploration in Reinforcement Learning under Episode-wise Constraints
D. Li*, **R. Huang***, C. Shen, J. Yang, International Conference on Machine Learning (ICML), 2023. (*Equal contribution)

[9] Non-stationary Reinforcement Learning under General Function Approximation
S. Feng, M.Yin, **R. Huang**, Y. Wang, J. Yang, Y. Liang, International Conference on Machine Learning (ICML), 2023.

[10] FLORAS: Differentially Private Wireless Federated Learning Using Orthogonal Sequences
X. Wei, T. Wang, **R. Huang**, C. Shen, J. Yang, and H. V. Poor, IEEE International Conference on Communications (ICC), May 2023.

[11] Safe Exploration Incurs Nearly No Additional Sample Complexity for Reward-free RL
R. Huang, J. Yang, Y. Liang, International Conference on Learning Representations (ICLR), 2023.

[12] Improved Sample Complexity for Reward-free Reinforcement Learning under Low-rank MDPs
Y. Cheng*, **R. Huang***, J. Yang, Y. Liang, International Conference on Learning Representations (ICLR), 2023. (*Equal contribution)

[13] Federated Linear Contextual Bandits
R. Huang, W. Wu, J. Yang, C. Shen, Conference on Neural Information Processing Systems (NeurIPS), December 2021.

[14] Cascading Bandits With Two-Level Feedback
D. Cheng*, **R. Huang***, C. Shen and J. Yang, IEEE International Symposium on Information Theory (ISIT), June 2022. (* Equal contribution)

Preprints

[1] Robust Offline Reinforcement Learning for Non-Markovian Decision Processes
R. Huang, Y. Liang, J. Yang, under review, submitted to IEEE Transactions on Information Theory.

[2] How Transformers Learn Regular Language Recognition: A Theoretical Study on Training Dynamics and Implicit Bias
R. Huang, Y. Liang, J. Yang, under review, submitted to International Conference on Machine Learning (ICML), 2025.

[3] Augmenting Online RL with Offline Data is All You Need: A Unified Hybrid RL Algorithm Design and Analysis
R. Huang, D. Li, C. Shi, C. Shen, J. Yang, under review, submitted to Conference on Uncertainty in Artificial Intelligence (UAI), 2025.

[4] Differentially Private Wireless Federated Learning Using Orthogonal Sequences
X. Wei, T. Wang, **R. Huang**, C. Shen, J. Yang and H. V. Poor, under review, submitted to IEEE Transactions on Information Theory.

Talks & Presentations

- Towards Efficient and Trustworthy AI: Theory on Models and Algorithms
  University of Kentucky, Feb 2025.
- Non-asymptotic Convergence of Training Transformers for Next-token Prediction
  NeurIPS, Dec 2024.
- Federated Online Prediction from Experts with Differential Privacy: Separations and Regret Speed-ups
  Theory and Practice of Differential Privacy Workshop (TPDP), Aug 2024
  NeurIPS, Dec 2024.
- Federated Linear Contextual Bandits with User-level Differential Privacy
  Theory and Practice of Differential Privacy Workshop (TPDP), Sep 2023
  ICML, Jul 2023.
- Federated Linear Contextual Bandits
  NeurIPS, Dec 2021.

- Safe Exploration Incurs Nearly No Additional Sample Complexity for Reward-Free RL
  Intel, Aug 2022.

Awards & Honors

2024 — 2025	Joab and Marly Thomas Graduate Fellowship, PSU. (The **sole** recipient in the College of Engineering)
2024	Melvin P. Bloom Memorial Graduate Fellowship in Electrical Engineering, PSU.
2023 — 2024	Future Faculty Immersive Teaching (FIT) Fellow, PSU.

Teaching

Instructor **2023 — 2024**
EDSGN 100: Cornerstone Engineering Design PSU

Teaching Assistant **2017 — 2018**
Calculus USTC

Services

Conference Reviewer	NeurIPS (2021-present) ICML (2022-present) ICLR (2022-present) AAAI (2024-present) AISTATS (2024-present)