Hongming Zhang

Email: hongmin2@ualberta.ca | Google Scholar | GitHub | Homepage

RESEARCH INTERESTS

My research centers on decision-making, with an emphasis on **deep reinforcement learning (DRL)** and planning across model-free, model-based, online, and offline settings. I develop sample-efficient algorithms by improving update rules, exploration, experience replay, and representation learning. My current work covers **RL-driven applications in game AI, LLMs, and robotics**.

EDUCATION

University of Alberta, Department of Computing Science 09/2020 ~ 06/2025
Ph.D. in Statistical Machine Learning, Supervisor: Martin Müller

Peking University, School of Mathematical Sciences 09/2016 ~ 06/2018
Master of Applied Statistics

Beijing Normal University, School of Mathematical Sciences 09/2012 ~ 06/2016
Bachelor of Science in Statistics

Aix-Marseille University, France Excellence Summer Schools (Dynamical Systems and Probability Theory) 07/2017 ~ 08/2017

WORK EXPERIENCE

Institute of Automation, Chinese Academy of Sciences (CASIA)

Lead research on LLM agents and LLM+RL 07/2025 ~ Present **Assistant Professor**

- Long-context & agent memory: Proposed a semantic-convolution long-context framework with logarithmic inference paths that beats baselines on RULER-HotpotQA across 7B-72B LLMs, and a coarse-to-fine memory system for agent planning that lifts AlfWorld/WebShop success rate by 10.4-31% while cutting online steps by 24%. [vLLM, verl]
- Tool use & self-improvement: Designed a multi-step tool-use method combining a retrieval strategy with a dynamic experience pool for self-improvement, improving ToolQA by 6.1/4.7% (easy/hard) and bringing an adapted Qwen2.5-72B to GPT-4o-level on τ-bench.

Built and scaled reinforcement-learning systems 06/2018 ~ 08/2020 **Algorithm Engineer**

- Training systems & infra: Built an evolutionary self-play framework that trains RL agents from scratch, compatible with model-free (PPO, SAC, AlphaStar) and model-based (AlphaZero, MuZero, Dreamer) algorithms; scaled to multi-node (**hundred-GPU**) training. [PyTorch/TF, Ray/RLlib]
- Game algorithms & theory: Developed and analyzed game-solving methods across tree search and game theory, extending AlphaZero to continuous action spaces with bounded-depth search for real-time budgets, and analyzing the convergence of DRL+evolutionary learning.

SELECTED PROJECTS

FPDou (Sole Developer) 06/2024 ~ 02/2025

- FPDou masters DouDizhu, a large-scale, imperfect-information card game. It is built by combining Fictitious Play with alternating on/off-policy learning, to provide a general and scalable solution for complex card-game decision-making.
- It achieved SOTA performance on DouDizhu compared with both open-source and closed-source

models. **Ranked 1st among 452 bots** on Botzone platform.

TensorLayer 2.0 (Core RL Team Member) **Stars 7.4k+, Fork 1.6k+** (code) 05/2019 ~ 02/2021

- TensorLayer is a TensorFlow-based deep learning and reinforcement learning library designed for researchers and engineers. It provides an extensive collection of customizable neural layers to build advanced AI models quickly. Joined the TensorLayer RL team and contributed to the DRL tutorial with TensorLayer 2.0, covering algorithms such as PG, DDPG, SAC, TRPO, PPO.
- Co-authored the textbook "*Deep Reinforcement Learning: Fundamentals, Research and Applications*", responsible for chapters on "*Taxonomy of Reinforcement Learning Algorithms*", "*Combine Deep Q-Networks with Actor-Critic*", and "*AlphaZero*". Published by Springer.

RLzoo (Core Team Member) **Stars 600+, Fork 100+** (code) 05/2019 ~ 02/2021

- RLzoo is a collection of the most practical reinforcement learning algorithms, frameworks and applications. It is implemented with TensorFlow 2.0 and TensorLayer 2.0, to provide a hands-on fast-developing approach for reinforcement learning practices and benchmarks.
- Published at ACM MM 2021; widely adopted by the RL research community for benchmarking.

AlphaZero Gomoku (Sole Developer) **Stars 221, Fork 44** (code) 09/2017 ~ 06/2018

- AlphaZero Gomoku is an open-source implementation of the AlphaZero algorithm for the board game Gomoku. It is built with an asynchronous self-play training pipeline and root-parallel Monte Carlo Tree Search (MCTS), together with a transfer-learning scheme that bootstraps larger-board models from smaller ones, to enable efficient AlphaZero-style learning under limited budget.
- The model ranked 20th in **Gomocup 2018** (one 1080Ti GPU, 400 simulations).

PUBLICATIONS

*equal contribution, †corresponding author

- **PhD Thesis**

  In-Sample Deep Reinforcement Learning. Hongming Zhang, 2025.

- **Book**

1. *Deep Reinforcement Learning: Fundamentals, Research and Applications*, Springer Nature, 2020.
   (Electronic Edition 300,000+ downloads (website); selected as **Annual High-Impact Publications in Computer Science by Chinese researchers**).
   Hao Dong, Zihan Ding, Shanghang Zhang, eds.
   Hang Yuan, **Hongming Zhang**, Jinqing Zhang, Yanhua Huang, Tianyang Yu, Huaqing Zhang, Ruitong Huang.
   First-authored Chapters: *Taxonomy of reinforcement learning algorithms*, *Combine Deep Q-Networks with Actor-Critic*, *AlphaZero.*

- **Published Papers**

1. *Great Minds Think Alike: Contextual Tacit Communication for Decentralized LLM-Agent Cooperation.*
   Yue Pei, **Hongming Zhang**, Jiarui Guan, Jusheng Zhang, Liang Lin, Haogang Zhu, Ziliang Chen.
   International Conference on Machine Learning (**ICML**), 2026.
2. *Principled Fast and Meta Knowledge Learners for Continual Reinforcement Learning.*
   Ke Sun*, **Hongming Zhang***, Jun Jin, Chao Gao, Xi Chen, Wulong Liu, Linglong Kong.

International Conference on Learning Representations (**ICLR**), 2026.

3. *STAR: Efficient Preference-based Reinforcement Learning via Dual Regularization.*
   Fengshuo Bai, Rui Zhao, **Hongming Zhang**, Sijia Cui, Ying Wen, Yaodong Yang, Bo Xu, Lei Han.
   Advances in Neural Information Processing Systems (**NeurIPS**), 2025.
4. *Coarse-to-Fine Grounded Memory for LLM Agent Planning.*
   Wei Yang, Jinwei Xiao, **Hongming Zhang**†**,** Qingyang Zhang, Yanna Wang, Bo Xu†.
   Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2025.
5. *Self-Guided Function Calling in Large Language Models via Stepwise Experience Recall.*
   Sijia Cui, Aiyao He, Shuai Xu, **Hongming Zhang,** Yanna Wang, Qingyang Zhang, Yajing Wang, Bo Xu.
   Conference on Empirical Methods in Natural Language Processing (**EMNLP Findings**), 2025.
6. *β-DQN: Improving Deep Q-Learning By Evolving the Behavior.*
   Hongming Zhang, Fengshuo Bai, Chenjun Xiao, Chao Gao, Bo Xu and Martin Müller.
   International Conference on Autonomous Agents and Multiagent Systems (**AAMAS Oral**), 2025.
7. *Latent Landmark Graph for Efficient Exploration-exploitation Balance in Hierarchical Reinforcement Learning.*
   Qingyang Zhang, **Hongming Zhang**, Dengpeng Xing, and Bo Xu.
   Machine Intelligence Research (MIR), 2025.
8. *Exploiting the Replay Memory Before Exploring the Environment: Enhancing Reinforcement Learning Through Empirical MDP Iteration.*
   Hongming Zhang, Chenjun Xiao, Chao Gao, Han Wang, Bo Xu, and Martin Müller.
   Advances in Neural Information Processing Systems (**NeurIPS**), 2024.
9. *A Distance-based Anomaly Detection Framework for Deep Reinforcement Learning.*
   Hongming Zhang*, Ke Sun*, Bo Xu, Linglong Kong, and Martin Müller.
   Transactions on Machine Learning Research (**TMLR**), 2024.
10. *Provable Representation with Efficient Planning for Partially Observable Reinforcement Learning.*
    Hongming Zhang*, Tongzheng Ren*, Chenjun Xiao, Dale Schuurmans, and Bo Dai.
    International Conference on Machine Learning (**ICML**), 2024.
11. *Monte Carlo Tree Search in the Presence of Transition Uncertainty.*
    Farnaz Kohankhaki, Kiarash Aghakasiri, **Hongming Zhang**, Ting-Han Wei, Chao Gao and Martin Müller.
    Association for the Advancement of Artificial Intelligence (**AAAI**), 2024.
12. *Replay Memory as An Empirical MDP: Combining Conservative Estimation with Experience Replay.*
    Hongming Zhang, Chenjun Xiao, Han Wang, Jun Jin, Bo Xu, and Martin Müller.
    International Conference on Learning Representations (**ICLR**), 2023.
13. *PiCor: Multi-Task Deep Reinforcement Learning with Policy Correction.*
    Fengshuo Bai, **Hongming Zhang**, Tianyang Tao, Zhiheng Wu, Yanna Wang, and Bo Xu.
    Association for the Advancement of Artificial Intelligence (**AAAI Oral**), 2023.
14. *Build generally reusable agent-environment interaction models.*
    Jun Jin, **Hongming Zhang**, and Jun Luo.
    Foundation Models for Decision Making, **NeurIPS Workshop**, 2022.
15. *Efficient Reinforcement Learning Development with RLzoo.*
    Zihan Ding, Tianyang Yu, **Hongming Zhang**, Yanhua Huang, Guo Li, Quancheng Guo, Luo Mai, and Hao Dong.
    ACM International Conference on Multimedia (**ACM MM**), 2021.
16. *RevCuT Tree Search Method in Complex Single-player Game with Continuous Search Space.*
    Hongming Zhang, Fangjuan Cheng, Bo Xu, Feng Chen, Jiachen Liu, and Wei Wu.

International Joint Conference on Neural Networks (**IJCNN Oral**), 2019.

- **Preprint/Under Review**

1. *TRACE: A Stability-Efficiency View of Adaptive Target-Network Updates in Deep Reinforcement Learning.*
   Hongming Zhang, et al.
2. *Rsync: Reward-Manifold Driven Adaptive Synchronization for Sample-Efficient Real-World RL.*
   Yishuai Cai, Fengshuo Bai, **Hongming Zhang**† et al.
3. *ConvMem: Convolutional Memory for Long-Context Reasoning.*
   Hongming Zhang, et al.
4. *FPDou: Mastering DouDizhu with Fictitious Play.*
   Hongming Zhang, et al.
5. *HyMem: Hierarchical Context Management for Long-Horizon Agents via Information Isolation.*
   XinQi Wang, **Hongming Zhang**†, et al.
6. *Failure-Aware Dual-Flow Control for Computer-Use Agents.*
   Jinwei Xiao, Sijia Cui, **Hongming Zhang**†, et al.
7. *Generalizable Opponent Exploitation in LLM Agents via Mixed Best-Responses Training.*
   Sijia Cui, XinQi Wang, **Hongming Zhang**†, et al.
8. *Double Check My Desired Return: Transformer with Value Validation for Offline RL.*
   Yue Pei, **Hongming Zhang**, et al.

HONORS & AWARDS

- UofA (PhD): Chinese Government Award for Outstanding Self-financed Students Abroad (Top 0.1%); Alberta Graduate Excellence Scholarship (Top 3%).
- CASIA: Fourth Prize, CETC Multi-Agent Reinforcement Learning Challenge.
- PKU (MSc): Academic Excellence Scholarship, second-class (Top 10%).
- BNU (BSc): Competition Scholarship, first-class (Top 1%); Academic Excellence Scholarship, second-class (Top 10%); Honorable Mention in Modeling, MCM/ICM; China Undergraduate Mathematical Contest in Modeling (2nd Prize in Beijing); Outstanding Student Leader.

ACADEMIC SERVICE

Conference Reviewer: ICLR, ICML, NeurIPS, ACL, EMNLP, AISTATS, UAI, ECAI, AAAI

TEACHING

TA CMPUT 455 Search, Knowledge and Simulation 2021 Fall, 2022 Winter, 2022 Fall, 2023 Fall

SKILLS

Programming: Python, R, MATLAB, Lua, SageMath, Scala
ML/DL: PyTorch, TensorFlow/TensorLayer, JAX, vLLM, verl, distributed training, multi-GPU
LLM stack: RAG, tool-use, agent, long-context, inference efficiency
RL: Model-free RL (PPO,SAC,AlphaStar), model-based RL (AlphaZero,MuZero,Dreamer), Offline RL
Others: LaTeX, Git, MySQL, Vibe Coding
Hobbies: Badminton, Football