

EDUCATION

- **University of Illinois Urbana-Champaign** Urbana, IL, USA
Ph.D. of Computer Science Aug. 2024 - May. 2029
 - **Research Field:** Self-evolving AGI Systems, Large-scale Generative Models, Data-efficient Policy Optimization, Trustworthy & Interpretable AI, RLHF, Reinforcement Learning Theory & Algorithms, MoE, MoA.
 - **Service:** Served as the reviewer at top-tier machine learning conferences, including ICML, ICLR, NeurIPS, AAAI and AISTATS.
 - **Relevant Course:** Machine Learning Algorithms for Large Language Models (LLMs).
- **Tsinghua University** Beijing, China
M.Eng. of Computer Technology; GPA 3.97/4.0, Top 1.3% Sept. 2021 - Jun. 2024
 - **Service:** Served as the reviewer at top-tier machine learning conferences, including ICML, ICLR, and NeurIPS.
 - **Relevant Course:** Stochastic Processes (A). Big Data Systems (A⁺). Digital Processing of Speech Signals (A). Data visualization (A⁺).
 - **Language Proficiency:** Attained a score of **101** on TOEFL, with a performance of 26 in Reading, 25 in Listening, 23 in Speaking, and 27 in Writing.
- **Nankai University** Tianjin, China
B.Eng. of Intelligent Science and Technology; GPA 93.28/100 (3.9/4.0), 1/83 Sept. 2017. - Jun. 2021
 - **Honors:** Ranked 1st in my major and obtained the National Scholarship twice (Top 1).
 - **Service:** Served as the minister at the Faculty Student Council.

RESEARCH INTEREST

- **Primary:** Sample-efficient reinforcement learning algorithms with a focus on self-evolving systems, human feedback integration (RLHF), and large-scale generative models. Particular emphasis on theoretical foundations for collapse-free continuous/lifelong learning and optimal policy convergence.
- **Secondary:** Bridging the gap between theoretical guarantees and practical scalability in AI systems, with emphasis on interpretable reinforcement learning algorithms that achieve both mathematical rigor and state-of-the-art performance in challenging real-world applications.

RESEARCH HIGHLIGHTS

- **Self-Evolving RLHF for Flow Matching Generative Models**
Sept. 2024 - Present
 - **Role:** **1)** Introduced a self-evolving RLHF framework (ORW-CFM-W2) that enables flow matching models to continuously optimize through online reward feedback without relying on human-collected datasets or likelihood calculations. **2)** Derived a tractable Wasserstein-2 distance bound for flow matching models, providing the first theoretical guarantee for collapse-free policy evolution. **3)** Established a unified perspective connecting flow matching fine-tuning with traditional KL-regularized RL, enabling controllable reward-diversity trade-offs.
 - **Achievements:** **1)** Achieved state-of-the-art alignment with orders of magnitude less data while maintaining generation diversity through theoretically-grounded regularization. **2)** Validated the framework's effectiveness by successfully fine-tuning large-scale models like Stable Diffusion 3 across challenging tasks including spatial understanding and compositional generation. **3)** Published a paper called "Online Reward-Weighted Fine-Tuning of Flow Matching with Wasserstein Regularization" at **ICLR 2025** as first author.
- **Behavior Control in Reinforcement Learning**
Sept. 2021 - Aug. 2024
 - **Role:** **1)** Introduced a unified framework called LBC to achieve behavior control in RL. **2)** Provided a unified perspective on diverse RL methods for behavior control and potential enhancements. **3)** Validated LBC's efficacy through rigorous theoretical support and extensive empirical experiments.
 - **Achievements:** **1)** Surpassed 24 human world records and attained the pinnacle of performance among reinforcement learning algorithms across most tasks. **2)** Published a paper titled "Learnable Behavior Control: Eclipsing Human World Records in Atari Games through Sample-Efficient Behavior Selection" at **ICLR 2023** with oral presentation as first author.

• Sample-Efficient Reinforcement Learning

Sept. 2020 - Aug. 2024

- **Role:** **1)** Introduced a sample-efficient Reinforcement Learning (RL) framework known as GDI, which achieved human-level performance by optimizing the data distribution of RL agents. **2)** Supported the efficacy of GDI with a robust foundation, including both theoretical proofs and an extensive array of experiments conducted in Atari. **3)** Provided a unified perspective on various RL algorithms with GDI.
- **Achievements:** **1)** Outperformed prior SOTA method Agent57 with **500x less data** and **twice** the average performance. **2)** Published a paper called "Generalized Data Distribution Iteration" at **ICML 2022** as first author.

PUBLICATIONS

1. Wang, Z., **Fan, J.**, et al. Variational Supervised Contrastive Learning. The Thirty-Ninth Annual Conference on Neural Information Processing Systems 2025 (**NeurIPS 2025**).
2. **Fan, J.**, et al. Adaptive Divergence Regularized Policy Optimization for Fine-tuning Generative Models. The Thirty-Ninth Annual Conference on Neural Information Processing Systems 2025 (**NeurIPS 2025**).
3. **Fan, J.**, et al. Online Reward-Weighted Fine-Tuning of Flow Matching with Wasserstein Regularization. International Conference on Learning Representations 2025 (**ICLR 2025**).
4. Ye Li, Chen Tang, Yuan Meng, **Jiajun Fan**, et al. PRANCE: Joint Token-Optimization and Structural Channel-Pruning for Adaptive ViT Inference, IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**).
5. **Fan, J.**, et al. Learnable Behavior Control: Breaking Atari Human World Records via Sample-Efficient Behavior Selection. International Conference on Learning Representations 2023 (**ICLR 2023**), **oral presentation, ranked 5/4176**.
6. **Fan, J.**, Xiao, C. Generalized Data Distribution Iteration. International Conference on Machine Learning 2022 (**ICML 2022**).
7. Hao Wang, Chen Zhichao, **Jiajun Fan**, et al. Optimal Transport for Treatment Effect Estimation. The Conference on Neural Information Processing Systems 2023 (**NeurIPS 2023**).
8. Wang E., Lian J., **Fan J.**, et al. Enhancing Sequential User Modeling with Large-kernel Convolution: A Lightweight Approach. The 30th SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, under review.
9. Wang E., Can Z., Yang Y., Pan L., **Fan J.**, et al. Unbiased Recommender Learning from Implicit Feedback: A Weak Supervision Perspective. The 30th SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, under review.
10. Wang E., **Fan J.**, et al. SPEC: Constructing Reliable Sequential User Model using Slide-window Spectrum. ACM TheWebConf 2024 Conference, under review.
11. E. Wang, H. Li, T. Liu, Y. Yang, **Fan J.**, X. Liu, and Z. Chen, "Unbiased recommender learning from implicit feedback: A progressive proximal transport approach," in ACM TheWebConf 2024 Conference, under review.
12. Wang, H., Chen, Z., **Fan, J.**, et al. Entire Space Counterfactual Learning: Tuning, Analytical Properties and Industrial Applications. The IEEE Transactions on Neural Networks and Learning Systems (**TNNLS**), under review.
13. **Fan, J.**, Xiao, C., & Huang, Y. GDI: Rethinking What Makes Reinforcement Learning Different From Supervised Learning. In the proceedings of **AAAI-22** Workshop on Reinforcement Learning in Games, 2021.
14. Xiao, C., Shi, H., **Fan, J.**, & Deng, S. An Entropy Regularization Free Mechanism for Policy-based Reinforcement Learning. arXiv preprint arXiv:2106.00707.
15. Xiao, C., Shi, H., **Fan, J.**, & Deng, S. CASA: A Bridge Between Gradient of Policy Improvement and Policy Evaluation. In the proceedings of Deep Reinforcement Learning Workshop **NeurIPS 2022**, 2022.
16. **Fan, J.** A Review for Deep Reinforcement Learning in Atari: Benchmarks, Challenges, and Solutions. In the proceedings of **AAAI-22** Workshop on Reinforcement Learning in Games, 2021.
17. **Fan, J.**, Ba, H., Guo, X., & Hao, J. Critic PI2: Master Continuous Planning via Policy Improvement with Path Integrals and Deep Actor-Critic Reinforcement Learning. arXiv preprint arXiv:2011.06752.

PATENTS

- Unified framework for model-free reinforcement learning algorithms Fan, J., Xiao, C. Unified framework for model-free reinforcement learning algorithms. CN112766497A[P].
- Hyperparameter tuning algorithm based on multi-arm gambling machine optimizer Fan, J. Hyperparameter tuning algorithm based on multi-arm gambling machine optimizer. CN112926629A[P].
- An unbiased estimation algorithm of behavior value function Fan, J., Xiao, C. An unbiased estimation algorithm of behavior value function. CN112926628A[P].
- Policy gradient algorithm based on double robust qualification trace Fan, J., Xiao, C. Policy gradient algorithm based on double robust qualification trace. CN112926735A[P].
- Asynchronous multi-arm gambling machine hyperparameter optimizer based on electoral college voting mechanism Fan, J. Asynchronous multi-arm gambling machine hyperparameter optimizer based on electoral college voting mechanism. CN112949850A[P].
- Real-time multi-hyperparameter controller Fan, J. Real-time multi-hyperparameter controller. CN113052252A[P].
- Hyperspace multi-coupling parameter optimizer based on multi-arm gambling machine combined with democratic voting Fan, J. Hyperspace multi-coupling parameter optimizer based on multi-arm gambling machine combined with democratic voting. CN113052253A[P].
- Fast and generalizable hyperspace coupling multi-parameter nonlinear optimizer Fan, J. Fast and generalizable hyperspace coupling multi-parameter nonlinear optimizer. CN113052248A[P].
- Reinforcement learning algorithm based on generalized combination strategy space Fan, J. Reinforcement learning algorithm based on generalized combination strategy space. CN113052312A[P].

AWARDS

- Outstanding Graduates (1%) Tianjin, China, 2021
- Excellent Graduation Thesis of Nankai University Tianjin, China, 2021
- Tang Lixin Scholarship (1%) Tianjin, China, 2021
- **National Scholarship**, Nankai University (1/83) Tianjin, China, 2020
- Nomination for Zhou Enlai Scholarship Tianjin, China, 2020
- **National Scholarship**, Nankai University (1/83) Tianjin, China, 2019
- 3rd Prize, Robocup@HOME Education World Final Sydney, Australia, 2019
- Bronze Medal, **ACM / ICPC** Asia Regional Contest Xuzhou, China, 2019
- National 2nd Prize, National College Students Mathematical Contest in Modeling (5%) Tianjin, China, 2018
- The First Prize Scholarship, Nankai University (2/83) Tianjin, China, 2018

RESEARCH EXPERIENCE

- **Collapse-free Self-evolution for Large Generative Model** Ph.D. Candidate
UIUC, Urbana, IL, USA Aug. 2024 - Present
 - **Role:** Introduced a theoretically-grounded flow matching framework (ORW-CFM-W2) that enables large generative models to **continuously evolve through human feedback** while preventing policy collapse.
 - **Achievements:** Developed a Wasserstein-regularized learning framework that achieved optimal policy convergence in challenging tasks like text-image alignment, while maintaining generation diversity through principled regularization. The method **significantly reduced computational and data requirements** compared to prior approaches.
- **Reinforcement Learning for Robotics** Research Assistant
Mila, Montreal, Canada Oct. 2023 - May. 2024
 - **Role:** Explored an efficient methodology for robot control in reinforcement learning and validated its effectiveness within Mujoco control task.
 - **Achievements:** Combined the priority of both model-free RL and design-based RL, and improved in several metrics. Introduced the behavior control into design-based RL, and improved the sample efficiency.

- Behavior Control in Reinforcement Learning** Research Assistant
Jun. 2021 - Oct. 2023
 Shenzhen, China
 - Role:** Pioneered a groundbreaking methodology for behavior control in reinforcement learning and validated its effectiveness by obliterating the most challenging Atari human world records.
 - Achievements:** Presented a cutting-edge paper titled "Learnable Behavior Control: Breaking Atari Human World Records via Sample-Efficient Behavior Selection" at **ICLR 2023** as an oral presentation, garnering a prestigious rank of **5 out of 4176** submissions.
- High Sample-efficiency Reinforcement Learning Paradigm** Research Assistant
Sep. 2020 - Jun. 2021
 Beijing, China
 - Role:** **1)** Innovated the foundational paradigm of RL by extending the Generalized Policy Iteration (GPI) model to a more all-encompassing version, the Generalized Data Distribution Iteration (GDI). **2)** Consolidated a multitude of vast-scale RL algorithms into the GDI architecture, which can be considered a specialized form of GDI. **3)** Substantiated the theoretical superiority of GDI over GPI through rigorous empirical analysis, providing irrefutable evidence of the framework's enhanced capabilities.
 - Achievements:** **1)** Attained State-of-the-Art (SOTA) in the Atari57 benchmark, outperforming the acclaimed Agent57 model with **500x less data** and even surpassing **22 human world records**. **2)** Authored a cutting-edge paper titled "Generalized Data Distribution Iteration," which was published at **ICML 2022**, a highly esteemed academic venue.
- Model-free Reinforcement Learning Algorithm** Research Assistant
Sep. 2020 - Jun. 2021
 Beijing, China
 - Role:** **1)** Spearheaded a comprehensive framework for model-free reinforcement learning titled CASA, which unifies and enhances existing approaches for more efficient and effective learning. **2)** Introduced a novel mechanism for policy-based reinforcement learning rooted in bandit theory, free from entropy regularization, which enhances the training process and leads to better performance.
 - Achievements:** Authored **two papers** entitled "CASA: A Bridge Between Gradient of Policy Improvement and Policy Evaluation" and "An Entropy Regularization Free Mechanism for Policy-based Reinforcement Learning", which have advanced the state-of-the-art in the field of reinforcement learning. Additionally, secured **nine patents**, highlighting the pioneering and innovative contributions to the field.
- Model-based Reinforcement Learning Algorithm** Research Assistant
Apr. 2020 - Sep. 2020
 Tianjin, China
 - Role:** Introduced a groundbreaking method for achieving unparalleled performance on the challenging MuJoCo environment, which represents a significant breakthrough in the field of reinforcement learning.
 - Achievements:** Presented a cutting-edge paper titled "Critic PI2: Master Continuous Planning via Policy Improvement with Path Integrals and Deep Actor-Critic Reinforcement Learning."
- Multifunctional Home Service Robot** Research Assistant
Sep. 2018 - Jul. 2019
 Tianjin, China
 - Role:** **1)** Devised a comprehensive solution for the entire project, along with a finite state machine diagram. **2)** Employed cutting-edge ROS-based algorithms for autonomous navigation and RRT path planning.
 - Achievements:** Presented the project in the 2019 ROBOCUP Sydney World Finals and won **3rd Prize**.

ACADEMIC SERVICE

- Reviewer of The 28th International Conference on Artificial Intelligence and Statistics AISTATS 2025
- Reviewer of Forty-Second International Conference on Machine Learning ICML 2025
- Reviewer of The Thirteenth International Conference on Learning Representations ICLR 2025
- Reviewer of The 39th Annual AAAI Conference on Artificial Intelligence AAAI 2025
- Reviewer of The Thirty-Eighth Annual Conference on Neural Information Processing Systems NeurIPS 2024
- Reviewer of The 30th SIGKDD Conference on Knowledge Discovery and Data Mining KDD 2024
- Reviewer of The Forty-first International Conference on Machine Learning ICML 2024
- Reviewer of The Twelfth International Conference on Learning Representations ICLR 2024
- Reviewer of The Thirty-seventh Annual Conference on Neural Information Processing Systems. NeurIPS 2023
- Reviewer of The Fortieth International Conference on Machine Learning. ICML 2023
- Reviewer of The Thirty-sixth Annual Conference on Neural Information Processing Systems. NeurIPS 2022