Novelty Assessment Report

Paper: Fair Policy Aggregation from Standard Policy Optimization
PDF URL: https://openreview.net/pdf?id=XNNDODynCl
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2026-01-01

Abstract

Currently the most powerful AI systems are aligned with human values via reinforcement learning from human feedback. Yet, reinforcement learning from human feedback models human preferences as noisy samples from a single linear ordering of shared human values and is unable to incorporate democratic AI alignment. In particular, the standard approach fails to represent and reflect diverse and conflicting perspectives of pluralistic human values. Recent research introduced the theoretically principled notion of quantile fairness for training a reinforcement learning policy in the presence of multiple, competing sets of values from different agents. Quite recent work provided an algorithm for achieving quantile fairness in the tabular setting with explicit access to the full set of states, actions and transition probabilities in the MDP. These current methods require solving linear programs with the size of the constraint set given by the number of states and actions, making it unclear how to translate this into practical training algorithms that can only take actions and observe individual transitions from the current state. In this paper, we design and prove the correctness of a new algorithm for quantile fairness that makes efficient use of standard policy optimization as a black-box without any direct dependence on the number of states or actions. We further empirically validate our theoretical results and demonstrate that our algorithm achieves competitive fairness guarantees to the prior work, while being orders of magnitude more efficient with respect to computation and the required number of samples. Our algorithm opens a new avenue for provable fairness guarantees in any setting where standard policy optimization is possible.

Core Task Landscape

This paper addresses: **Aggregating Multiple Reward Functions in Reinforcement Learning for Fair Policy Selection**

A total of **50 papers** were analyzed and organized into a taxonomy with **18 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Multi-Objective Reinforcement Learning Frameworks for Fairness**
- **Reward Function Design and Aggregation Techniques**
- **Constrained Reinforcement Learning for Fairness**
- **Multi-Agent Reinforcement Learning with Fairness Considerations**
- **Domain-Specific Applications of Fair Multi-Objective RL**
- **Theoretical Foundations and Algorithmic Innovations**

Complete Taxonomy Tree

- Aggregating Multiple Reward Functions in Reinforcement Learning for Fair Policy Selection Survey Taxonomy
- Multi-Objective Reinforcement Learning Frameworks for Fairness
  - Welfare-Based Multi-Objective Optimization (4 papers)
  - [7] Welfare and fairness in multi-objective reinforcement learning (Fan, 2022) View paper
  - [11] Scalable multi-objective reinforcement learning with fairness guarantees using lorenz dominance (Michailidis Dimitris, 2024) View paper
  - [12] Fair deep reinforcement learning with generalized gini welfare functions (Guanbao Yu, 2023) View paper
  - [20] FairDICE: Fairness-Driven Offline Multi-Objective Reinforcement Learning (Kim Woosung, 2025) View paper
  - Pareto-Optimal Policy Discovery with Fairness Criteria (4 papers)
  - [1] A multi-objective framework for fair reinforcement learning (A Cimpeana, 2023) View paper
  - [4] Group Fairness in Reinforcement Learning via Multi-Objective Rewards (J Blandin, 2024) View paper
  - [13] A Fairness-Oriented Multi-Objective Reinforcement Learning approach for Autonomous Intersection Management (Fabris, 2025) View paper
  - [23] Learning fair policies in multi-objective (deep) reinforcement learning with average and discounted rewards (Siddique, 2020) View paper
  - Multi-Objective Evolutionary and Meta-Learning Approaches (2 papers)
  - [5] Fairness-aware multiobjective evolutionary learning (Zhang Qing-quan, 2024) View paper
  - [22] Provable Multi-Party Reinforcement Learning with Diverse Human Feedback (Zhong Huiying, 2024) View paper
- Reward Function Design and Aggregation Techniques
  - Fairness-Aware Reward Shaping and Composition (5 papers)
  - [6] Large-scale reinforcement learning for diffusion models (Yinan Zhang, 2024) View paper
  - [9] CH-MARL: Constrained Hierarchical Multiagent Reinforcement Learning for Sustainable Maritime Logistics (Alqithami, 2025) View paper
  - [16] Optimizing Language Models with Fair and Stable Reward Composition in Reinforcement Learning (Jiahui Li, 2024) View paper
  - [25] Dynamic Fairness Calibration for Real-Time Decision Systems using Multi-Objective Reinforcement Learning (Harshraj Bhoite, 2025) View paper

- [42] Reward redistribution mechanisms in multi-agent reinforcement learning (A Ibrahim, 2020) View paper
- Multi-Reward Aggregation for Language Model Alignment (2 papers)
- [14] Countering reward over-optimization in llm with demonstration-guided reinforcement learning (Mathieu Rita, 2024) View paper
- [41] Multi-Reward GRPO Fine-Tuning for De-biasing Large Language Models: A Study Based on Chinese-Context Discrimination Data (Deng Yixuan, 2025) View paper
- Adaptive and Dynamic Reward Mechanisms (2 papers)
- [33] A Multi-Component Reward Function with Policy Gradient for Automated Feature Selection with Dynamic Regularization and Bias Mitigation (Paudel, 2025) View paper
- [43] Multimodal reward shaping for efficient exploration in reinforcement learning (Yuan, 2021) View paper

- Constrained Reinforcement Learning for Fairness (1 papers)
  - [19] Constrained Reinforcement Learning for Fair and Environmentally Efficient Traffic Signal Controllers (Ammar Haydari, 2024) View paper
- Multi-Agent Reinforcement Learning with Fairness Considerations
  - Cooperative Multi-Agent Systems with Equity Objectives (4 papers)
  - [17] Cooperation and Fairness in Multi-Agent Reinforcement Learning (Aloor, 2024) View paper
  - [29] Kindness in Multi-Agent Reinforcement Learning (Alamiyan-Harandi, 2023) View paper
  - [46] A Fairness-Aware Cooperation Strategy for Multi-Agent Systems Driven by Deep Reinforcement Learning (Zhixiang Liu, 2022) View paper
  - [48] Fairness in Cooperative Multiagent Multiobjective Reinforcement Learning using the Expected Scalarized Return (F Chouaki, 2025) View paper
  - Decentralized and Federated Multi-Agent Learning (3 papers)
  - [32] Scalable Multi-Agent Reinforcement Learning-Based Distributed Channel Access (Zhenyu Chen, 2023) View paper
  - [37] Federated Multi-Agent Reinforcement Learning for Privacy-Preserving and Energy-Aware Resource Management in 6G Edge Networks (Min Qi, 2025) View paper
  - [39] A Systematic Evaluation of Preference Aggregation in Federated RLHF for Pluralistic Alignment of LLMs (Mahmoud Srewa, 2025) View paper
  - Market-Based and Peer-to-Peer Multi-Agent Systems (3 papers)
  - [28] Achieving diverse objectives with ai-driven prices in deep reinforcement learning multi-agent markets (Panayiotis Danassis, 2021) View paper
  - [40] FairMarket-RL: LLM-Guided Fairness Shaping for Multi-Agent Reinforcement Learning in Peer-to-Peer Markets (Shrenik Jadhav, 2025) View paper
  - [44] Scalable Fairness Shaping with LLM-Guided Multi-Agent Reinforcement Learning for Peer-to-Peer Electricity Markets (Birva Sevak, 2025) View paper
- Domain-Specific Applications of Fair Multi-Objective RL
  - Recommender Systems with Fairness and Diversity (5 papers)
  - [2] Toward pareto efficient fairness-utility trade-off in recommendation through reinforcement learning (Ge, 2022) View paper
  - [3] Multi-objective reinforcement learning for recommender systems: a comprehensive survey of methods, challenges, and future directions (Fatima Ezzahra Zaizi, 2025) View paper
  - [10] DMOR: A dynamic multi-objective reinforcement framework with adaptive graph fusion for recommendation (Alamdari, 2025) View paper
  - [31] Multi-objective Reinforcementlearning Approach to Fairness in Ride Matching (Hsu, 2024) View paper
  - [50] Revisiting Fairness-aware Interactive Recommendation: Item Lifecycle as a Control Knob (Yun Lu, 2025) View paper
  - Traffic Signal Control and Transportation Systems (2 papers)
  - [26] MMD-TSC: An adaptive multi-objective traffic signal control for energy saving with traffic efficiency (Yuqi Zhang, 2024) View paper
  - [30] Multi-Objective Reinforcement Learning for Large-Scale Mixed Traffic Control (Iftekharul Islam, 2025) View paper
  - Wireless Network Resource Management (6 papers)
  - [8] Adaptive Proximal Policy Optimization for Efficient and SLA-Compliant Dynamic O-RAN Slice Resource Allocation (Nokwanda Shezi, 2025) View paper
  - [18] Multi-user reinforcement learning based multi-reward for spectrum access in cognitive vehicular networks (Lingling Chen, 2023) View paper
  - [24] Joint Optimization of Multi-UAV Deployment and User Association via Deep Reinforcement Learning for Long-Term Communication Coverage (Xu Cheng, 2024) View paper
  - [35] A Scalable Reinforcement Learning Framework for Ultra-Reliable Low-Latency Spectrum Management in Healthcare Internet of Things (Adeel Iqbal, 2025) View paper
  - [38] Dynamic Service Function Chain (SFC) Deployment for Autonomous Intelligent Systems (Xiao Kong, 2025) View paper
  - [45] Intelligent Priority-Aware Spectrum Access in 5G Vehicular IoT: A Reinforcement Learning Approach (Adeel Iqbal, 2025) View paper
  - UAV Deployment and Aerial Network Optimization (1 papers)
  - [34] UAV-Mounted Base Station Coverage and Trajectory Optimization Using LSTM-A2C with Attention (YM Worku, 2025) View paper
  - Agricultural and Environmental Decision Support (1 papers)
  - [21] Comparative Analysis of Multi-Agent Reinforcement Learning Policies for Crop Planning Decision Support (Mahajan Anubha, 2024) View paper
  - Healthcare and Service Scheduling Systems (2 papers)
  - [27] Bias in reinforcement learning: A review in healthcare applications (Smith, 2023) View paper
  - [47] Optimized Routing and Scheduling in Home Health Care Using Deep Reinforcement Learning (Omollo, 2026) View paper
  - Specialized Technical Applications (1 papers)
  - [49] NL2Lean: Translating Natural Language into Lean 4 through Multi-Aspect Reinforcement Learning (Yue Fang, 2025) View paper
- Theoretical Foundations and Algorithmic Innovations ★ (3 papers)
  - [0] Fair Policy Aggregation from Standard Policy Optimization (Anon et al., 2026) View paper
  - [15] From fair solutions to compromise solutions in multi-objective deep reinforcement learning (Junqi Qian, 2025) View paper
  - [36] Fairness in Preference-based Reinforcement Learning (Siddique, 2023) View paper

Narrative

Core task: Aggregating multiple reward functions in reinforcement learning for fair policy selection. The field addresses how agents can balance competing objectives—such as efficiency, equity, and stakeholder preferences—when a single scalar reward is insufficient. The taxonomy reveals six main branches that collectively span the landscape. Multi-Objective Reinforcement Learning Frameworks for Fairness and Theoretical Foundations and Algorithmic Innovations provide the conceptual backbone, developing Pareto-based methods (Pareto Fairness Utility[2], Multi-Objective Fair RL[1]) and welfare-theoretic approaches (Welfare Fairness MORL[7], Gini Welfare Functions[12]) that formalize trade-offs among objectives. Reward Function Design and Aggregation Techniques explores practical scalarization and weighting schemes, while Constrained Reinforcement Learning for Fairness enforces hard fairness constraints alongside primary objectives. Multi-Agent Reinforcement Learning with Fairness Considerations examines coordination and equity in decentralized settings (Kindness MARL[29], Fairness-Aware Cooperation DRL[46]), and Domain-Specific Applications of Fair Multi-Objective RL demonstrates these ideas in areas like traffic control, healthcare resource allocation, and recommendation systems (Multi-Objective RecSys Survey[3]).

Recent work highlights tensions between computational scalability and fairness guarantees, with some studies pursuing large-scale deployments (Large-Scale Diffusion RL[6]) and others refining theoretical notions of equity under uncertainty (Scalable Lorenz Dominance[11], Group Fairness Multi-Objective[4]). Fair Policy Aggregation[0] sits within the Theoretical Foundations branch, contributing algorithmic mechanisms for combining multiple reward signals into policies that respect fairness criteria. It shares conceptual ground with Fair to Compromise[15] and Preference-Based Fairness[36], both of which also grapple with how to aggregate stakeholder utilities without imposing a single dominant objective. Where Fair to Compromise[15] emphasizes negotiation-style trade-offs and Preference-Based Fairness[36] incorporates explicit user preferences, Fair Policy Aggregation[0] focuses on principled aggregation rules that ensure no group is systematically disadvantaged. This positioning reflects broader debates about whether fairness should emerge from constraint satisfaction, welfare optimization, or transparent aggregation of diverse reward functions.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. From fair solutions to compromise solutions in multi-objective deep reinforcement learning

Authors: Junqi Qian, Umer Siddique, Guanbao Yu, Paul Weng | **Year/Venue**: 2025 | **URL**: View paper

Abstract

â¦ (RL) where the expected vector returns are aggregated with a â¦ policies), we prove a general performance bound that justifies learning a policy using discounted rewards, even if a policy â¦

Relationship Analysis

Both papers belong to the Theoretical Foundations and Algorithmic Innovations category, focusing on theoretical guarantees for fair policy aggregation in multi-objective reinforcement learning. They overlap in addressing fairness when aggregating multiple reward functions, with both providing theoretical analysis of policy selection mechanisms. However, the original paper specifically develops algorithms for quantile fairness using standard policy optimization as a black-box with provable efficiency guarantees, while the candidate paper focuses on deriving performance bounds for compromise solutions and justifying the use of discounted rewards in fair multi-objective settings.

---

2. Fairness in Preference-based Reinforcement Learning

Authors: Siddique, Umer, Umer Siddique, Sinha, Abhinav, et al. (8 authors total) | **Year/Venue**: 2023 | **URL**: View paper

Abstract

In this paper, we address the issue of fairness in preference-based reinforcement learning (PbRL) in the presence of multiple objectives. The main objective is to design control policies that can optimize multiple objectives while treating each objective fairly. Toward this objective, we design a new fairness-induced preference-based reinforcement learning or FPbRL. The main idea of FPbRL is to learn vector reward functions associated with multiple objectives via new welfare-based preferences ra...

Relationship Analysis

Both papers belong to the Theoretical Foundations and Algorithmic Innovations category, focusing on fair policy aggregation in multi-objective reinforcement learning settings. They overlap in addressing fairness through welfare functions and policy optimization, but differ fundamentally in their approach: the original paper develops a black-box algorithm using multiplicative weights for quantile fairness with provable guarantees, while the candidate paper integrates fairness into preference-based RL by learning vector reward functions from welfare-based preferences rather than reward-based preferences. The original paper emphasizes theoretical efficiency and quantile-based fairness definitions, whereas the candidate focuses on practical preference learning with generalized Gini welfare functions.

Contributions Analysis

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Efficient algorithm for quantile-fair policy aggregation using policy optimization as black-box

Description: The authors develop an algorithm that achieves quantile-fair policy aggregation by making O(n) calls to a policy optimization subroutine, avoiding explicit dependence on the MDP's state or action space size. This contrasts with prior methods requiring full access to transition probabilities.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

Contribution 2: Optimal occupancy distribution for defining quantile fairness

Description: The authors propose using a distribution over policies induced by individually optimal policies (the optimal occupancy distribution) rather than the uniform distribution over all policies. This choice enables tractable quantile estimation and avoids exponential sample complexity issues.

This contribution was assessed against **2 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. UOEP: User-Oriented Exploration Policy for Enhancing Long-Term User Experiences in Recommender Systems

URL: View paper

Brief Assessment

UOEP Long-Term RecSys[53] focuses on user-oriented exploration in recommender systems using distributional critics for different user activity levels, not on policy aggregation or optimal occupancy distributions over policies in multi-agent MDPs.

2. Reinforcing Long-Term Performance in Recommender Systems with User-Oriented Exploration Policy

URL: View paper

Brief Assessment

User-Oriented Exploration Policy[52] focuses on user-level exploration in recommender systems using distributional critics for different user activity groups, not on policy aggregation or optimal occupancy distributions over policies in multi-agent MDPs.

Contribution 3: Multiplicative weights update method for computing quantile-fair policies

Description: The authors design an algorithm based on the multiplicative weights update method that computes quantile-fair policies efficiently, requiring only O(log n) policy evaluations instead of solving large linear programs over states and actions.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. AGNOSTICâEffective Quantitative Finance With Online Learning

URL: View paper

Brief Assessment

AGNOSTIC Quantitative Finance[51] applies multiplicative weights to quantitative finance problems with lookback windows and quantile thresholds for asset selection, not to computing quantile-fair policies in multi-agent reinforcement learning settings.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Fair Policy Aggregation from Standard Policy Optimization View paper
- [1] A multi-objective framework for fair reinforcement learning View paper
- [2] Toward pareto efficient fairness-utility trade-off in recommendation through reinforcement learning View paper
- [3] Multi-objective reinforcement learning for recommender systems: a comprehensive survey of methods, challenges, and future directions View paper
- [4] Group Fairness in Reinforcement Learning via Multi-Objective Rewards View paper
- [5] Fairness-aware multiobjective evolutionary learning View paper
- [6] Large-scale reinforcement learning for diffusion models View paper
- [7] Welfare and fairness in multi-objective reinforcement learning View paper
- [8] Adaptive Proximal Policy Optimization for Efficient and SLA-Compliant Dynamic O-RAN Slice Resource Allocation View paper
- [9] CH-MARL: Constrained Hierarchical Multiagent Reinforcement Learning for Sustainable Maritime Logistics View paper
- [10] DMOR: A dynamic multi-objective reinforcement framework with adaptive graph fusion for recommendation View paper
- [11] Scalable multi-objective reinforcement learning with fairness guarantees using lorenz dominance View paper
- [12] Fair deep reinforcement learning with generalized gini welfare functions View paper
- [13] A Fairness-Oriented Multi-Objective Reinforcement Learning approach for Autonomous Intersection Management View paper
- [14] Countering reward over-optimization in llm with demonstration-guided reinforcement learning View paper
- [15] From fair solutions to compromise solutions in multi-objective deep reinforcement learning View paper
- [16] Optimizing Language Models with Fair and Stable Reward Composition in Reinforcement Learning View paper
- [17] Cooperation and Fairness in Multi-Agent Reinforcement Learning View paper
- [18] Multi-user reinforcement learning based multi-reward for spectrum access in cognitive vehicular networks View paper
- [19] Constrained Reinforcement Learning for Fair and Environmentally Efficient Traffic Signal Controllers View paper
- [20] FairDICE: Fairness-Driven Offline Multi-Objective Reinforcement Learning View paper
- [21] Comparative Analysis of Multi-Agent Reinforcement Learning Policies for Crop Planning Decision Support View paper
- [22] Provable Multi-Party Reinforcement Learning with Diverse Human Feedback View paper
- [23] Learning fair policies in multi-objective (deep) reinforcement learning with average and discounted rewards View paper
- [24] Joint Optimization of Multi-UAV Deployment and User Association via Deep Reinforcement Learning for Long-Term Communication Coverage View paper
- [25] Dynamic Fairness Calibration for Real-Time Decision Systems using Multi-Objective Reinforcement Learning View paper
- [26] MMD-TSC: An adaptive multi-objective traffic signal control for energy saving with traffic efficiency View paper
- [27] Bias in reinforcement learning: A review in healthcare applications View paper
- [28] Achieving diverse objectives with ai-driven prices in deep reinforcement learning multi-agent markets View paper
- [29] Kindness in Multi-Agent Reinforcement Learning View paper
- [30] Multi-Objective Reinforcement Learning for Large-Scale Mixed Traffic Control View paper
- [31] Multi-objective Reinforcementlearning Approach to Fairness in Ride Matching View paper
- [32] Scalable Multi-Agent Reinforcement Learning-Based Distributed Channel Access View paper
- [33] A Multi-Component Reward Function with Policy Gradient for Automated Feature Selection with Dynamic Regularization and Bias Mitigation View paper
- [34] UAV-Mounted Base Station Coverage and Trajectory Optimization Using LSTM-A2C with Attention View paper
- [35] A Scalable Reinforcement Learning Framework for Ultra-Reliable Low-Latency Spectrum Management in Healthcare Internet of Things View paper
- [36] Fairness in Preference-based Reinforcement Learning View paper
- [37] Federated Multi-Agent Reinforcement Learning for Privacy-Preserving and Energy-Aware Resource Management in 6G Edge Networks View paper
- [38] Dynamic Service Function Chain (SFC) Deployment for Autonomous Intelligent Systems View paper
- [39] A Systematic Evaluation of Preference Aggregation in Federated RLHF for Pluralistic Alignment of LLMs View paper
- [40] FairMarket-RL: LLM-Guided Fairness Shaping for Multi-Agent Reinforcement Learning in Peer-to-Peer Markets View paper
- [41] Multi-Reward GRPO Fine-Tuning for De-biasing Large Language Models: A Study Based on Chinese-Context Discrimination Data View paper
- [42] Reward redistribution mechanisms in multi-agent reinforcement learning View paper

- [43] Multimodal reward shaping for efficient exploration in reinforcement learning View paper
- [44] Scalable Fairness Shaping with LLM-Guided Multi-Agent Reinforcement Learning for Peer-to-Peer Electricity Markets View paper
- [45] Intelligent Priority-Aware Spectrum Access in 5G Vehicular IoT: A Reinforcement Learning Approach View paper
- [46] A Fairness-Aware Cooperation Strategy for Multi-Agent Systems Driven by Deep Reinforcement Learning View paper
- [47] Optimized Routing and Scheduling in Home Health Care Using Deep Reinforcement Learning View paper
- [48] Fairness in Cooperative Multiagent Multiobjective Reinforcement Learning using the Expected Scalarized Return View paper
- [49] NL2Lean: Translating Natural Language into Lean 4 through Multi-Aspect Reinforcement Learning View paper
- [50] Revisiting Fairness-aware Interactive Recommendation: Item Lifecycle as a Control Knob View paper
- [51] AGNOSTICâ□□Effective Quantitative Finance With Online Learning View paper
- [52] Reinforcing Long-Term Performance in Recommender Systems with User-Oriented Exploration Policy View paper
- [53] UOEP: User-Oriented Exploration Policy for Enhancing Long-Term User Experiences in Recommender Systems View paper