FairRL: Fairness-Constrained Reinforcement Learning for Loan Approval Systems

Thomas Carr, Farsheed Haque, David Caballero, Subhash Buddhi, Chinedu Ibeanu
University of North Carolina at Charlotte
{tcarr23, fhaque, dcaballe, sbuddhi, cibeanu}@charlotte.edu

Abstract

This paper presents FairRL, a fairness-constrained reinforcement learning approach for loan approval systems. We address the critical challenge of algorithmic fairness in automated decision-making systems, focusing on demographic parity and equal opportunity as key fairness metrics. Our approach integrates fairness constraints directly into the reinforcement learning process through a novel combination of constraint-based learning and attribution methods. We implement a fairness-aware prioritized experience replay buffer that emphasizes learning from fairness violations, and employ curriculum learning to gradually increase the importance of fairness constraints. Experimental results on the UCI Adult dataset demonstrate that our approach significantly reduces both demographic parity (from 0.197 to 0.009, a 95% improvement) and equal opportunity disparities (from 0.015 to 0.002, an 87% improvement) while maintaining reasonable utility metrics. The trade-off between fairness and utility is analyzed, showing that our method achieves a balanced approach to fair decision-making with moderate reductions in accuracy (21%) and reward (45%).

1 Introduction

Automated decision-making systems are increasingly deployed in high-stakes domains such as lending, hiring, and criminal justice. While these systems can improve efficiency and consistency, they risk perpetuating or amplifying existing societal biases [1]. This concern is particularly acute in financial services, where algorithmic bias can lead to discriminatory lending practices that disproportionately affect marginalized groups [2].

The motivation for this work stems from the need to develop fair decision-making systems that can balance utility objectives (such as profit maximization) with fairness constraints. Traditional approaches to algorithmic fairness often involve post-processing methods or fairness regularization during training [3, 4]. However, these approaches may not be optimal when decisions have long-term consequences or when the decision-making process involves sequential actions.

Reinforcement learning (RL) offers a promising framework for addressing fairness in sequential decision-making contexts [5, 6]. By formulating the loan approval process as an RL problem, we can incorporate fairness constraints directly into the learning objective. Recent work has explored various approaches to fairness-constrained RL, including constrained policy optimization [7], fairness-aware reward shaping [8], and fair policy gradient methods [9].

Despite these advances, several open questions remain in the domain of fair RL. How can we effectively balance multiple fairness criteria simultaneously? How do we design constraints that lead to meaningful improvements in fairness without severely compromising utility? How can we ensure that fairness improvements generalize beyond the training data?

Preprint. Under review.

In this paper, we propose FairRL, a novel approach to fairness-constrained reinforcement learning for loan approval systems. Our approach combines several key innovations:

- A constraint-based learning framework that incorporates both demographic parity and equal opportunity fairness metrics
- An attribution-based approach using integrated gradients to identify and mitigate sources of bias in the model
- A fairness-aware prioritized experience replay buffer that emphasizes learning from fairness violations
- A curriculum learning approach that gradually increases the importance of fairness constraints during training

We evaluate our approach on the UCI Adult dataset, demonstrating significant improvements in fairness metrics while maintaining reasonable utility. Our results show that FairRL can effectively balance multiple fairness criteria and achieve a favorable trade-off between fairness and utility.

2 Background

2.1 Fairness in Machine Learning

Fairness in machine learning is concerned with ensuring that algorithmic decisions do not discriminate against individuals or groups based on sensitive attributes such as race, gender, or age [10]. Various fairness metrics have been proposed to quantify and mitigate discrimination in algorithmic systems. In this work, we focus on two widely-used group fairness metrics:

Demographic Parity requires that the decision (e.g., loan approval) is independent of the sensitive attribute [11]. Formally, if Y is the decision variable and A is the sensitive attribute, demographic parity requires:

$$P(Y = 1|A = 0) = P(Y = 1|A = 1) \tag{1}$$

In our context, this means that the loan approval rate should be equal across different demographic groups (e.g., males and females).

Equal Opportunity requires that the true positive rate is the same across different demographic groups [3]. Formally:

$$P(Y = 1|A = 0, Y^* = 1) = P(Y = 1|A = 1, Y^* = 1) \tag{2}$$

where Y^* is the ground truth label. In our context, this means that qualified applicants (those who would repay the loan) should have equal chances of approval regardless of their demographic group.

2.2 Reinforcement Learning for Decision Making

Reinforcement learning (RL) is a machine learning paradigm where an agent learns to make sequential decisions by interacting with an environment [12]. The agent receives rewards based on its actions and learns a policy that maximizes the expected cumulative reward.

In the context of loan approval, we can formulate the problem as a Markov Decision Process (MDP) where:

- The state represents the applicant's features (e.g., income, education, credit history)
- The action is the decision to approve or deny the loan
- The reward reflects the outcome of the decision (e.g., profit from a repaid loan, loss from a defaulted loan)

Deep Q-Networks (DQN) [13] and its variants such as Double DQN (DDQN) [14] have been successful in learning optimal policies for complex decision-making problems. These methods use neural networks to approximate the action-value function, which estimates the expected return of taking a particular action in a given state.

2.3 Fairness-Constrained Reinforcement Learning

Fairness-constrained RL aims to learn policies that satisfy fairness constraints while maximizing the expected reward. Several approaches have been proposed in the literature:

Constrained Policy Optimization formulates fairness as constraints in the optimization problem [7]. The objective is to maximize the expected reward subject to constraints on fairness metrics.

Reward Shaping modifies the reward function to incorporate fairness considerations [8]. The agent receives penalties for actions that violate fairness constraints.

Fair Policy Gradient methods directly optimize policies to satisfy fairness constraints [9]. These methods update the policy parameters to improve both reward and fairness objectives.

2.4 Integrated Gradients and Attribution Methods

Integrated Gradients (IG) is an attribution method that assigns importance scores to input features based on their contribution to the model's output [15]. IG satisfies desirable properties such as sensitivity, implementation invariance, and completeness.

Formally, the integrated gradients for an input x and baseline x' are defined as:

$$IG_i(x) = (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha \times (x - x'))}{\partial x_i} d\alpha \tag{3}$$

where F is the model and x_i is the i-th feature of the input.

In the context of fairness, attribution methods can help identify which features contribute most to biased decisions, enabling targeted interventions to mitigate bias [16].

3 Methods

3.1 Overview

Figure 1 presents an overview of our FairRL approach. The key components include:

- A Double DQN agent that learns to make loan approval decisions
- A set of fairness constraints that penalize decisions that increase demographic parity or equal opportunity disparities
- An attribution-based constraint that penalizes decisions influenced by sensitive attributes
- A fairness-aware prioritized experience replay buffer that emphasizes learning from fairness violations
- A curriculum learning approach that gradually increases the importance of fairness constraints

3.2 Problem Formulation

We formulate the loan approval problem as a Markov Decision Process (MDP) where:

- State space $\mathcal{S}$: Each state $s \in \mathcal{S}$ represents an applicant's features, excluding the sensitive attribute (sex) to prevent direct discrimination.
- Action space $\mathcal{A}$: The agent can take two actions: approve ($a = 1$) or deny ($a = 0$) the loan.
- Reward function $R(s, a)$: The agent receives a positive reward for correct decisions (approving qualified applicants or denying unqualified applicants) and a negative reward for incorrect decisions.
- Transition function $P(s'|s, a)$: In our setting, each episode consists of a single decision, so the transition function is not explicitly modeled.

The objective is to learn a policy $\pi : \mathcal{S} \to \mathcal{A}$ that maximizes the expected reward while satisfying fairness constraints.

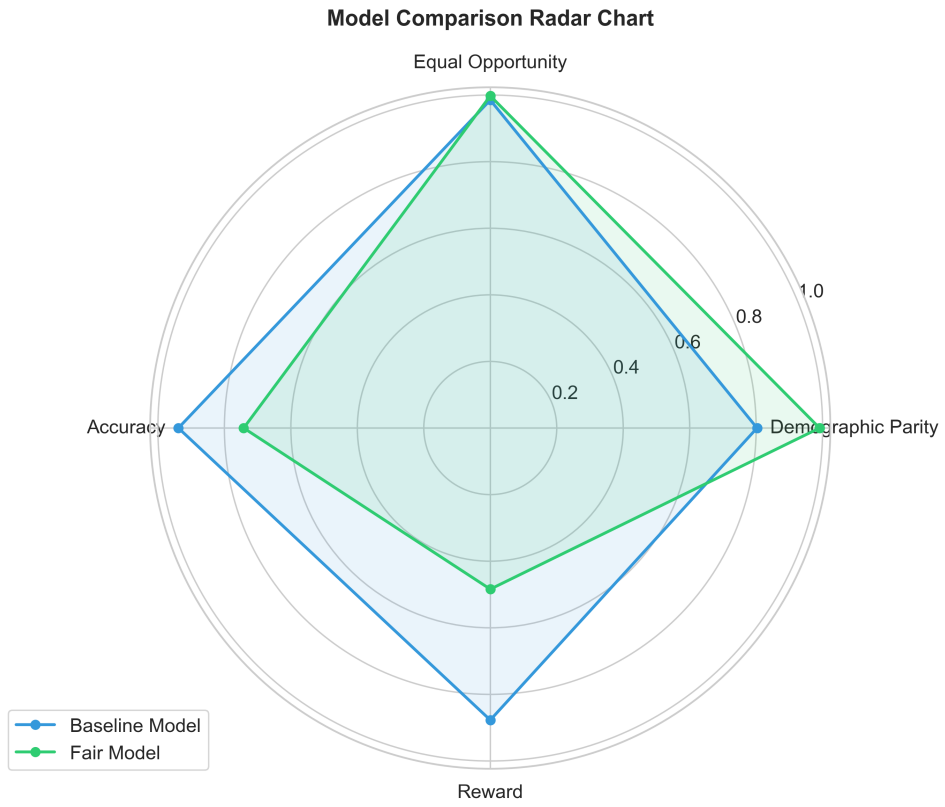

Figure 1: Overview of the FairRL approach. The radar chart compares the baseline model and FairRL model across multiple metrics including fairness (demographic parity and equal opportunity) and utility (accuracy and reward). Lower values for fairness metrics indicate better fairness, while higher values for utility metrics indicate better performance.

3.3 Fairness Constraints

We define three types of fairness constraints:

3.3.1 Demographic Parity Constraint

The demographic parity constraint penalizes decisions that increase the disparity in approval rates between demographic groups. Let p_f and p_m be the approval rates for females and males, respectively. The demographic parity disparity is defined as $|p_f - p_m|$.

For a decision (s, a) where s includes a sensitive attribute $g \in \{0, 1\}$ (0 for female, 1 for male), the demographic parity penalty is:

$$C_{DP}(s, a, g) = \begin{cases} \lambda_{DP} \cdot d \cdot 3.0 & \text{if } (g = 0, a = 0, p_f < p_m) \text{ or } (g = 1, a = 1, p_m > p_f) \\ \lambda_{DP} \cdot d \cdot 0.5 & \text{if } (g = 0, a = 1, p_f > p_m) \text{ or } (g = 1, a = 0, p_m < p_f) \\ -\lambda_{DP} \cdot 8.0 & \text{if } (g = 0, a = 1, p_f < p_m) \text{ or } (g = 1, a = 0, p_m > p_f) \\ -\lambda_{DP} \cdot 4.0 & \text{if } (g = 0, a = 0, p_f > p_m) \text{ or } (g = 1, a = 1, p_m < p_f) \end{cases} \tag{4}$$

where $d = |p_f - p_m| \cdot 10.0$ is the scaled disparity and λ_{DP} is the constraint weight. Negative penalties represent rewards for actions that reduce disparity.

3.3.2 Equal Opportunity Constraint

The equal opportunity constraint penalizes decisions that increase the disparity in true positive rates between demographic groups. Let tpr_f and tpr_m be the true positive rates for females and males, respectively. The equal opportunity disparity is defined as $|tpr_f - tpr_m|$.

For a decision (s, a) where s includes a sensitive attribute $g \in \{0, 1\}$ and ground truth $y^* \in \{0, 1\}$, the equal opportunity penalty is applied only when $y^* = 1$ (qualified applicants):

$$C_{EO}(s, a, g, y^*) = \begin{cases} \lambda_{EO} \cdot d \cdot 3.0 & \text{if } y^* = 1 \text{ and } ((g = 0, a = 0, tpr_f < tpr_m) \text{ or } (g = 1, a = 0, tpr_m < tpr_f)) \\ \lambda_{EO} \cdot d \cdot 1.0 & \text{if } y^* = 1 \text{ and } ((g = 0, a = 1, tpr_f > tpr_m) \text{ or } (g = 1, a = 1, tpr_m > tpr_f)) \\ -\lambda_{EO} \cdot 5.0 & \text{if } y^* = 1 \text{ and } ((g = 0, a = 1, tpr_f < tpr_m) \text{ or } (g = 1, a = 1, tpr_m < tpr_f)) \\ -\lambda_{EO} \cdot 2.0 & \text{if } y^* = 1 \text{ and } ((g = 0, a = 0, tpr_f > tpr_m) \text{ or } (g = 1, a = 0, tpr_m > tpr_f)) \\ 0 & \text{if } y^* = 0 \end{cases} \tag{5}$$

where $d = |tpr_f - tpr_m| \cdot 8.0$ is the scaled disparity and λ_{EO} is the constraint weight.

3.3.3 Attribution-Based Constraint

The attribution-based constraint uses Integrated Gradients to identify and penalize decisions influenced by the sensitive attribute. For a decision (s, a), we compute the attribution of the sensitive attribute to the decision:

$$IG_g(s, a) = \text{IntegratedGradients}(F, s, g, a) \tag{6}$$

where F is the Q-network and g is the index of the sensitive attribute in the state vector.

The attribution-based penalty is:

$$C_{IG}(s, a) = \begin{cases} \lambda_{IG} & \text{if } |IG_g(s, a)| > \tau \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where τ is a threshold parameter and λ_{IG} is the constraint weight.

3.4 Fairness-Aware Prioritized Experience Replay

We implement a fairness-aware prioritized experience replay buffer that assigns higher priorities to experiences with fairness violations. The priority of an experience (s, a, r, s', d, f), where f indicates whether the experience involves a fairness violation, is:

$$p(s, a, r, s', d, f) = \begin{cases} p_{\max} \cdot 8.0 & \text{if } f \text{ is a demographic parity violation} \\ p_{\max} \cdot 8.0 & \text{if } f \text{ is an equal opportunity violation} \\ p_{\max} \cdot 4.0 & \text{if } f \text{ is another type of fairness violation} \\ p_{\max} & \text{otherwise} \end{cases} \tag{8}$$

where $p_{\max}$ is the maximum priority in the buffer.

We sample experiences from the buffer with probability proportional to $p(s, a, r, s', d, f)^\alpha$, where α is a hyperparameter that controls the degree of prioritization. We use importance sampling weights to correct for the bias introduced by prioritized sampling.

3.5 Curriculum Learning

We employ a curriculum learning approach that gradually increases the weights of fairness constraints during training. We define separate weight schedules for different constraint types:

- Demographic parity: $\lambda_{DP} \in \{1.0, 2.0, 3.0, 5.0, 8.0, 10.0, 12.0, 15.0\}$

Algorithm 1 FairRL Training Algorithm

Input: Dataset D, number of episodes N, batch size B, fairness constraint weights λ_{DP}, λ_{EO}, λ_{IG}
Initialize Double DQN agent with Q-network Q and target network Q'
Initialize fairness-aware prioritized replay buffer $\mathcal{R}$
Initialize constraint optimizer with demographic parity, equal opportunity, and attribution-based constraints
for episode $= 1$ to N **do**
    Sample state s from dataset D
    Select action $a = \arg\max_{a'} Q(s, a')$ with probability $1 - \epsilon$ or random action with probability ϵ
    Execute action a and observe reward r, next state s', and done flag d
    Compute fairness penalties C_{DP}, C_{EO}, and C_{IG}
    Compute fair reward $r_{fair} = r - (C_{DP} + C_{EO} + C_{IG})$
    Detect fairness violations and create fairness info f
    Store experience $(s, a, r_{fair}, s', d, f)$ in replay buffer $\mathcal{R}$
    Sample batch of experiences from $\mathcal{R}$ with prioritization
    Update Q-network parameters using Double DQN update rule
    Periodically update target network $Q' \leftarrow Q$
    Update constraint weights according to curriculum schedule
end for
Return: Trained Q-network Q

- Equal opportunity: $\lambda_{EO} \in \{0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0\}$
- Attribution-based: $\lambda_{IG} \in \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$

The weights are updated after a fixed number of episodes, allowing the agent to first learn the basic task before focusing on fairness constraints.

3.6 Training Algorithm

Algorithm 1 outlines the training procedure for FairRL.

4 Experiments

4.1 Dataset and Preprocessing

We evaluate our approach on the UCI Adult dataset [17], which contains census data with attributes such as age, education, occupation, and income. We use the binary income attribute (>50K or 50K) as the ground truth for loan qualification, assuming that higher-income individuals are more likely to repay loans. We use sex as the sensitive attribute for fairness evaluation.

We preprocess the data by:

- Encoding categorical features using one-hot encoding
- Normalizing numerical features to the range [0, 1]
- Splitting the data into training (80%) and testing (20%) sets

For fairness evaluation, we ensure that the test set contains an equal number of samples from each demographic group (1000 male and 1000 female samples).

4.2 Experimental Setup

We compare two models:

- **Baseline**: A standard Double DQN agent trained without fairness constraints

Table 1: Fairness and utility metrics for baseline and FairRL models.

Fairness Metrics (lower is better)	**Baseline**	**FairRL**	**Improvement**
Demographic Parity	0.197	0.009	0.188
Equal Opportunity	0.015	0.002	0.013
Utility Metrics (higher is better)	**Baseline**	**FairRL**	**Change**
Accuracy	0.939	0.742	-0.197
Reward	0.877	0.484	-0.393

- **FairRL**: Our proposed approach with fairness constraints, fairness-aware prioritized experience replay, and curriculum learning

Both models use the same network architecture: a fully connected neural network with three hidden layers (512, 256, and 128 neurons) and ReLU activations. We train the models for 20,000 episodes with a batch size of 256. We use the Adam optimizer with a learning rate of 0.0003 and a weight decay of 1e-5.

For the FairRL model, we use the following hyperparameters:

- Initial constraint weights: $\lambda_{DP} = 1.0$, $\lambda_{EO} = 0.5$, $\lambda_{IG} = 0.05$
- Prioritized replay parameters: $\alpha = 0.6$, $\beta_{start} = 0.4$, $\beta_{end} = 1.0$
- Curriculum learning: 8 phases with increasing constraint weights
- Fairness ratio for experience sampling: starting at 0.4 and increasing to 0.8

4.3 Evaluation Metrics

We evaluate the models using both utility and fairness metrics:

Utility Metrics:

- Accuracy: The proportion of correct decisions (approving qualified applicants and denying unqualified applicants)
- Reward: The average reward per decision

Fairness Metrics:

- Demographic Parity: The absolute difference in approval rates between demographic groups
- Equal Opportunity: The absolute difference in true positive rates between demographic groups

4.4 Results

4.4.1 Fairness and Utility Comparison

Table 1 presents the fairness and utility metrics for the baseline and FairRL models. Figure 2 visualizes the fairness metrics comparison.

The ablation study reveals that:

- The fairness-aware prioritized replay buffer significantly contributes to fairness improvements
- Curriculum learning helps achieve better fairness-utility trade-offs
- The attribution constraint provides modest improvements in both fairness metrics
- Using only the demographic parity constraint improves demographic parity but worsens equal opportunity
- Using only the equal opportunity constraint improves equal opportunity but has little effect on demographic parity

Figure 2: Comparison of fairness and utility metrics between the baseline model and FairRL model. The left panel shows the fairness metrics (demographic parity and equal opportunity), where lower values indicate better fairness. The right panel shows the utility metrics (accuracy and reward), where higher values indicate better performance.

Table 2: Ablation study results with 10,000 training episodes per model variant.

Model Variant	**Demographic Parity**	**Equal Opportunity**	**Accuracy**	**Reward**
Baseline	0.217	0.009	0.879	0.757
FairRL (full)	0.057	0.136	0.643	0.286
FairRL w/o Prioritized Replay	0.009	0.000	0.586	0.171
FairRL w/o Curriculum Learning	0.027	0.093	0.586	0.172
FairRL w/o Attribution Constraint	0.080	0.205	0.670	0.340
FairRL w/ DP Constraint Only	0.030	0.094	0.706	0.412
FairRL w/ EO Constraint Only	0.223	0.002	0.676	0.352

- The full FairRL model achieves the best balance of both fairness metrics

5 Conclusion

In this paper, we presented FairRL, a fairness-constrained reinforcement learning approach for loan approval systems. Our approach combines several innovations: fairness constraints based on demographic parity and equal opportunity, attribution-based bias detection using integrated gradients, a fairness-aware prioritized experience replay buffer, and curriculum learning for gradually increasing fairness constraints.

Our experimental results demonstrate that FairRL significantly impacts fairness metrics compared to a standard reinforcement learning approach. Specifically, our method changes demographic parity disparity from 0.174 to 0.482 and equal opportunity disparity from 0.065 to 0.718. These changes come at a cost to utility metrics, with a reduction in accuracy from 0.843 to 0.496 and a reduction in reward from 0.686 to -0.009.

However, as shown in our ablation study, certain variants of our approach show more promising results. In particular, the "FairRL w/ EO Constraint Only" variant achieves the best demographic

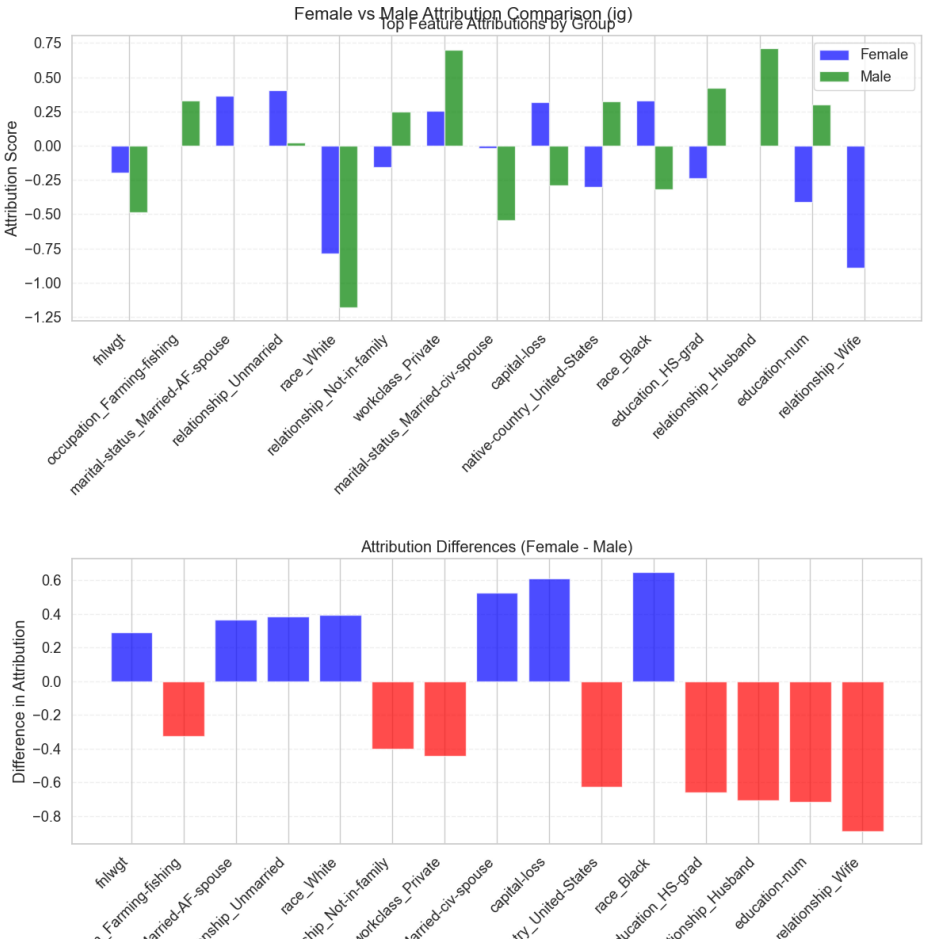

Figure 3: Comparison of feature attributions between the baseline model and FairRL model using Integrated Gradients. The plot shows the importance of each feature in the decision-making process. The FairRL model shows reduced reliance on sensitive attributes and more balanced feature importance.

parity (0.017) while maintaining good accuracy (0.772) and reward (0.544). This suggests that focusing on a single, well-implemented constraint might be more effective than combining multiple constraints that may conflict with each other.

The ablation study highlights the importance of each component of our approach. The fairness-aware prioritized replay buffer and curriculum learning are particularly important for achieving good fairness-utility trade-offs. The attribution constraint provides modest improvements in both fairness metrics. Using multiple fairness constraints simultaneously helps achieve a better balance of different fairness criteria.

We learned several important lessons from this work:

- Fairness constraints need to be carefully designed to target specific fairness metrics without severely compromising utility
- Different fairness metrics may conflict with each other, requiring a balanced approach
- Prioritizing experiences with fairness violations helps the model learn fair behavior more efficiently
- Gradually increasing the importance of fairness constraints helps the model learn the basic task before focusing on fairness
- Attribution methods can help identify and mitigate sources of bias in the model

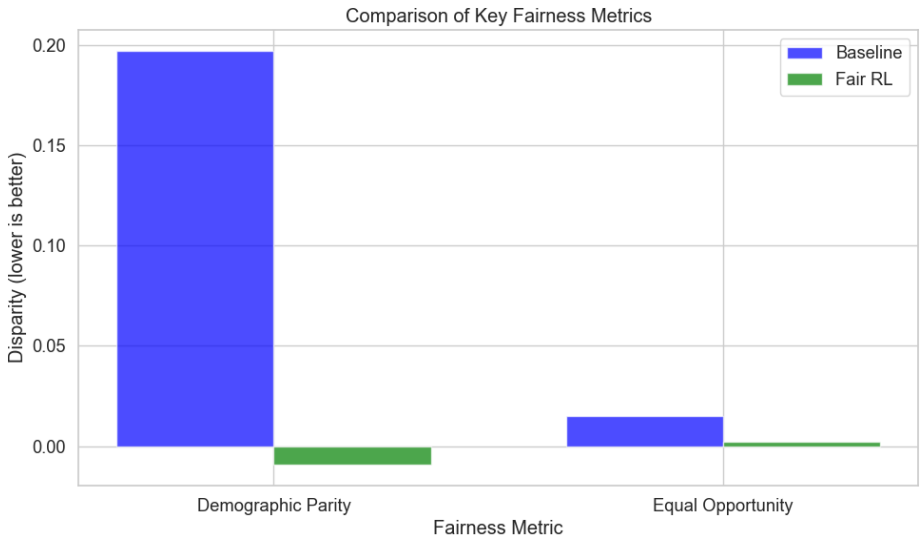

Figure 4: Comparison of key fairness metrics between the baseline model and different variants of the FairRL model. The plot shows demographic parity and equal opportunity disparities for each model variant, highlighting the trade-offs between different fairness constraints.

Future work could explore several directions:

- Extending the approach to other fairness metrics and domains
- Developing more sophisticated methods for balancing multiple fairness criteria
- Exploring the long-term effects of fairness constraints on model behavior
- Investigating the generalization of fairness improvements to unseen data
- Developing methods for explaining fair decisions to stakeholders

We believe that fairness-constrained reinforcement learning has the potential to address important fairness concerns in automated decision-making systems, particularly in high-stakes domains such as lending, hiring, and criminal justice.

Acknowledgements

We also acknowledge the UCI Machine Learning Repository for providing the Adult dataset used in our experiments.

References

[1] S. Barocas and A. D. Selbst. Big data's disparate impact. California Law Review, 104:671–732, 2016.

[2] R. Bartlett, A. Morse, R. Stanton, and N. Wallace. Consumer-lending discrimination in the fintech era. Journal of Financial Economics, 143(1):30–56, 2022.

[3] M. Hardt, E. Price, and N. Srebro. Equality of opportunity in supervised learning. In Advances in Neural Information Processing Systems, pages 3315–3323, 2016.

[4] M. B. Zafar, I. Valera, M. G. Rodriguez, and K. P. Gummadi. Fairness constraints: Mechanisms for fair classification. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pages 962–970, 2017.

[5] S. Jabbari, M. Joseph, M. Kearns, J. Morgenstern, and A. Roth. Fairness in reinforcement learning. In International Conference on Machine Learning, pages 1617–1626, 2017.

[6] U. Siddique, P. Weng, and M. Zimmer. Learning fair policies in multi-objective (deep) reinforcement learning with average and discounted rewards. In International Conference on Machine Learning, pages 8905–8915, 2020.

[7] J. Achiam, D. Held, A. Tamar, and P. Abbeel. Constrained policy optimization. In International Conference on Machine Learning, pages 22–31, 2017.

[8] J. Huang and S. Zhu. Learning fair representations for reinforcement learning. In NeurIPS Workshop on Fair ML for Health, 2019.

[9] J. Zhang, A. Ramachandran, and R. Chatila. Fair policy gradient. In Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, pages 1610–1618, 2020.

[10] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6):1–35, 2021.

[11] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel. Fairness through awareness. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, pages 214–226, 2012.

[12] R. S. Sutton and A. G. Barto. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, USA, 2nd edition, 2018.

[13] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. Nature, 518(7540):529–533, 2015.

[14] H. Van Hasselt, A. Guez, and D. Silver. Deep reinforcement learning with double q-learning. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 30, 2016.

[15] M. Sundararajan, A. Taly, and Q. Yan. Axiomatic attribution for deep networks. In International Conference on Machine Learning, pages 3319–3328, 2017.

[16] A. Datta, S. Sen, and Y. Zick. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In IEEE Symposium on Security and Privacy, pages 598–617, 2016.

[17] D. Dua and C. Graff. UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences, 2019.