Fairness-Driven Explainable Learning in Multi-Agent Reinforcement Learning

Tariq Mahmood[1,†], Reza Shahbazian[1,*,†] and Irina Trubitsyna[1,†]

[1]*Department of Informatics, Modeling, Electronics and System Engineering (DIMES), University of Calabria, Italy*

Abstract

Fairness and explainability are important issues that must be addressed in multi-agent reinforcement learning (MARL) systems. In this study, we propose a novel approach that directly incorporates fairness constraints and layer-wise relevance propagation (LRP) into multi-agent training. Through the proposed method, explainability and fairness can be addressed simultaneously, improving the interpretability of agent's decisions and guaranteeing that agents are assigned tasks equitably. We evaluate the performance of the proposed method based on a resource allocation problem. The results show average fairness and explainability ratings of 0.921 and 0.931, respectively. Preliminary results show that this strategy greatly enhances system fairness and explainability while maintaining a competitive average system reward. Furthermore, by encouraging efficient resource use, the proposed method advances the principles of green artificial intelligence.

Keywords

Explainable AI, Reinforcement Learning, Multi-Agent Systems, Fariness

1. Introduction

The rapid advancement of artificial intelligence (AI) and, in particular, of machine learning (ML) has transformed numerous sectors. Given the tremendous impact these technologies have on decision-making processes, the concepts of fairness and explainability in AI have become a necessity. AI systems have been shown to have biases [1, 2], that may be in many shapes and forms. To avoid their negative impact, AI systems should aim toward fairness, *i.e. the absence of any prejudice or favoritism toward an individual or group based on their inherent or acquired characteristics* [3]. This is especially significant in multi-agent systems, as one agent's decisions have direct consequences for others in the same environment. Several studies have focused on fairness in machine learning algorithms (see [3, 4]).

On the other hand, the explainability of AI systems is another important research issue [5]. Explainable AI (XAI) systems seek to provide humans with clear and transparent explanations for their actions, which is essential for trusting and interacting with AI. Furthermore, by encouraging efficient resource use, these methods advance the principles of green artificial intelligence.

Multi-agent reinforcement learning (MARL) is a powerful way to deal with complex and dynamic environments where multiple agents interact and learn simultaneously. MARL can be used to generate through continuous learning and adaptation. The underlying interactions and dependencies among the agents, make it difficult to achieve both fairness and explainability simultaneously. The literature extensively addresses both fairness [6] and explainability [7]; however, there is low emphasis on satisfying both features at the same time. Different frameworks could be utilized to address fairness in multi-agent systems, including *Individual Fairness* that ensures similar agents are treated similarly and *Group Fairness* that seeks comparable results for various demographic groupings [8].

AAPEI '24: 1st International Workshop on Adjustable Autonomy and Physical Embodied Intelligence, October 20, 2024, Santiago de Compostela, Spain.

*Corresponding author.

†The names are in alphabetical order.

mahmood.tariq@dimes.unical.it (T. Mahmood); reza.shahbazian@unical.it (R. Shahbazian); i.trubitsyna@dimes.unical.it (I. Trubitsyna)

https://dottorato.dimes.unical.it/students/mahmood-tariq (T. Mahmood); https://sites.google.com/view/rezashahbazian/ (R. Shahbazian); https://sites.google.com/dimes.unical.it/itrubitsyna/home (I. Trubitsyna)

0009-0005-6870-7509 (T. Mahmood); 0000-0002-2313-6002 (R. Shahbazian); 0000-0002-9031-0672 (I. Trubitsyna)

Green AI highlights the need for AI systems that are not only powerful but also sustainable in terms of the environment. As an example, lets consider the following multi-agent reinforcement learning framework:

Example 1. *Think of a MARL system where N agents cooperate to accomplish M activities in a shared environment. The capability level c_i of each agent $i \in \{1, \ldots, N\}$ indicates its task-performance capacity (e.g., hardware availability). The difficulty level d_j determines the characteristics of each task $j \in \{1, \ldots, M\}$. The task complexity and agent capacity match to determine the agent's reward for completing a task. Each agent aims to choose tasks that maximize its predicted cumulative reward, $J_i(\pi_i)$.*

Agents are encouraged to choose tasks that are in line with their capabilities, which promotes effective resource utilization. This is achieved by using the match between the agent's capability and the task's difficulty as a metric for a given reward. The conventional objective is to maximize the *total cumulative* rewards $\sum_{i=1}^{N} J_i(\pi_i)$. Unlike the traditional methods, we propose to add fairness constraints to ensure *equitable (individual) reward distribution* among agents and explainability constraints that guarantee *transparent decision-making processes*.

This fair allocation of rewards among agents, independent of their innate capabilities, contributes to preventing disadvantages for agents with lower capabilities. This is significant because many stakeholders, geographical areas, or systems with differing resources and technological capabilities may be represented by these agents. Let's consider "World Community Grid" in our Example 1 as a sizable distributed computer network for scientific research. The agents stand in for individual computers or small clusters that have been donated by various global players. These systems' computing power and energy efficiency are reflected in their capability levels. If there was no fairness restriction, the system might constantly assign the hardest jobs to the agents with the highest capabilities, which would result in high energy usage and participation barriers. It should be emphasized that the major idea behind this study is *to integrate fairness and explainability into MARL*. Depending on the application, the fairness constraint may be defined differently.

On the other hand, understanding why certain decisions are made can lead to insights for improving the system, potentially identifying more efficient allocation strategies. By combining fairness and explainability, one can help to design more sustainable, efficient, and trustworthy AI systems that adhere closely to the ideals of Green AI.

In this study, we address individual fairness in the context of algorithmic fairness, meaning similar agents should receive similar treatment [3]. We use the concept of explainable guided learning (EGL) [9] and apply it to multi-agent reinforcement learning. We integrate fairness and explainability constraints directly into the learning process. We utilize layer-wise relevance propagation (LRP) to evaluate the explainability of agents' decisions and incorporate a fairness-driven reward adjustment mechanism to maintain equity among agents.

The structure of this paper is as follows: In Section 2, we briefly address the most recent related works and the preliminary. Section 3 outlines the problem statement, objective function, and proposed method. Section 4 presents the experimental setup and results, demonstrating the efficacy of the proposed algorithm. Finally, Section 5 concludes the paper, discussing the limits and future research direction.

2. Literature Review

Cimpean et al. [10] propose a universal framework for the establishment of fairness in RL agents, targeting fairness. Their methodology formulates the fairness as a sequential problem through Markov decision process (MDP), incorporating historical data on states, actions, rewards, and feedback. Pozanco and Borrajo [11] extend fairness considerations to cooperative multi-agent planning, known as MAP, with a focus on fair distribution of goals among the agents. They come up with two approaches: first, a fairness-driven optimization method for preprocessing the goals that have to be assigned to the agents, and second, a planning-based compilation to solve the assignment of goals and planning together. Fairness in RL through reward shaping has been explored by Jabbari et al. [12]. They use

a mechanism to adjust rewards to achieve fair outcomes. Weng et al. [13, 14] discuss fairness in RL by developing policies that account for multiple fairness constraints simultaneously. Chen et al. [15] bring fairness to actor-critic RL by putting forward modifications to the learning algorithm that ensure fair outcomes. Rodriguez-Soto et al. [16] study ethical behavior in multi-agent systems through multi-objective reinforcement learning.

Explainability in AI is critical for understanding decision-making processes. Traditional approaches often focus on post-hoc explanations, where the decisions of pre-trained models are interpreted. However, such methods come short in dynamic interactions where interpretable and transparent decisions must be made in a continuous fashion. Current research efforts, like LRP [17], try to decompose neural network decisions back to input features, hence making them more transparent. Most existing approaches, however, do not make explainability an integral part of the learning process and might therefore not be applicable in real-time decision scenarios. In a comprehensive survey, Milani et al. [7] categorize explainable reinforcement learning (XRL) techniques at a high level into three categories: Feature Importance, Markov decision process (MDP), and Policy-Level. They stress the importance of explainability in RL and discuss the recent techniques, including SHAP and LIME, to provide feature-level explanations and discuss various challenges.

In short, the existing methods mainly focus on fairness and explainability as independent issues. Therefore, we propose a single framework that could handle both in multi-agent systems simultaneously. Continuing, we provide some preliminary definitions:

- *Agent:* A multi-agent system entity that learns and makes decisions according to its environment and follows policy. An agent i can be represented by a tuple $\langle S_i, A_i, \pi_i, R_i \rangle$ where S_i is the set of states, A_i is the set of actions the agent can take, $\pi_i : S_i \to A_i$ is the policy that maps states to actions, and $R_i : S_i \times A_i \to \mathbb{R}$ is the reward function that provides feedback to the agent based on its actions in a given state.
- *Environment:* The environment or context within which agents act and interact. It is typically modeled by a Markov decision process (MDP) defined by a tuple $\langle S, A, P, R \rangle$ where S is the set of all possible states of the environment, A is the set of all possible actions that agents can take, $P : S \times A \times S \to [0, 1]$ is the state transition probability function, where $P(s'|s, a)$ denotes the probability of transitioning to state s' from state s after taking action a, and $R : S \times A \to \mathbb{R}$ is the reward function that assigns a reward based on the current state and action taken.
- *Fairness:* A measure ensuring that no agent is unjustly disadvantaged or advantaged in terms of the rewards obtained. Fairness can be measured through many metrics, such as *equality of opportunity* and *demographic parity* (e.g., race, gender). It should be noted that different frameworks exist for fairness in multi-agent systems, while we focus on individual fairness.
- *Explainability:* The explanation capacity of the model for the agent's decisions. Explainability is achievable through techniques that describe the process of the decision itself, like counterfactuals or feature attribution methods.

3. System Model

We consider a set of N agents interacting within a common environment. Each agent i, where $i \in [1..N]$, acts according to a policy π_i that specifies its actions as a function of observed states of the environment. The goal is to maximize the collective reward function while maximizing the agents' overall performance under fairness and explainability constraints. The problem is defined as follows:

$$\begin{aligned} &\max_{\pi_1, \pi_2, \ldots, \pi_N} \sum_{i=1}^{N} J_i(\pi_i), \\ &\text{subject to:} \quad F(J_1, J_2, \ldots, J_N) \geq \delta, \quad E(\pi_i) \geq \epsilon, \quad \forall i, \end{aligned} \tag{1}$$

where $J_i(\pi_i)$ is the expected cumulative reward (objective function) of agent i, F measures the fairness across the agents, δ is the threshold of fairness, E measures the explainability of the decisions, and ϵ is

the minimal level of explainability. The fairness constraint requires the equal treatment of all agents in terms of the rewards they receive for their actions. The explainability for the actions performed by agents is obtained by using the *Layer-wise Relevance Propagation (LRP)* method [17]. Intuitively, this method decomposes the output decision back to the input layer, providing the contribution of every input feature to the decision. Given a neural network function g and an input x, LRP seeks to assign a relevance score Re_j to each input feature x_j such that $g(x) = \sum_j Re_j$, where Re_j is the relevancy of feature x_j with respect to the network output.

3.1. Proposed Algorithm

The proposed method is presented in Algorithm 1. We first initialize the network parameters θ_i for each agent i, the learning rate α[1], and set threshold values for fairness and explainability (line 1). We use a softmax exploration strategy [2] to select actions as presented in Equation (2):

$$\pi_i(a_i|s_i;\theta_i) = \frac{\exp(Q_i(s_i,a_i;\theta_i))}{\sum_{a' \in A_i} \exp(Q_i(s_i,a';\theta_i))} \tag{2}$$

where $Q_i(s_i, a_i; \theta_i)$ is the estimated reward for action a_i in state s_i. We update the policy parameters according to the rule given in Equation (3):

$$\theta_i \leftarrow \theta_i + \alpha(t) \cdot r_i \tag{3}$$

where $\alpha(t)$ is the learning rate and r_i is the immediate reward. It should be noted that calculating $J_i(\pi_i)$ requires considering the entire sequence of states, actions, and rewards, which can be computationally expensive and impractical for frequent updates. Therefore, in Algorithm 1, we use r_i as the immediate reward for each agent. To evaluate fairness, we consider Jain's fairness metric [18] as defined in the Equation (4):

$$F(J_1, J_2, \ldots, J_N) = \frac{(\sum_{i=1}^{N} J_i(\pi_i))^2}{N \sum_{i=1}^{N} J_i(\pi_i)^2} \tag{4}$$

A score close to 1 indicates high fairness, meaning all agents are receiving rewards that are close to the average. In Algorithm 1 (lines 14–16), we check to make sure that the fairness is above the given threshold. In cases where the fairness is lower than the threshold, we perform the "Adjust" mechanism to ensure equitable reward distribution among agents. As a sample for "Adjust" mechanism, lets consider the code that we implemented in our experiment as given in Appendix.

To evaluate the explainability, we use the layer-wise relevance propagation (LRP) scores $Re_{i,j}$. The relevance metric is defined in Equation (5):

$$Re_{i,j} = \frac{|Q_i(s_i,a_j;\theta_i)|}{\sum_k |Q_i(s_i,a_k;\theta_i)|} \tag{5}$$

where $Q_i(s_i, a_j; \theta_i)$ is the Q-value for agent i taking action a_j in its current state s_i, and the summation in the denominator is over all possible actions k for agent i in state s_i. We assign relevance scores based on the absolute of Q-values for each action. Actions with higher absolute Q-values are considered more relevant to the agent's decision-making process. We consider the *Entropy* of this relevance (Equation (6)), where $H(Re_i)$ is the entropy of the relevance scores for agent i. Entropy is a quantitative indicator of relevance scores that makes comparing explainability amongst agents simple.

$$H(Re_i) = -\sum_j Re_{i,j} \log_2(Re_{i,j}) \tag{6}$$

[1] In our experiment, we use a constant value for the learning rate.

[2] We use a softmax strategy, however other strategies such as ϵ-greedy could also be considered.

The explainablity metric $E(\pi_i)$ is defined in Equation (7), where d is the number of features or inputs considered by the model. A higher value (close to 1) indicates that behaviors are more explainable.

$$E(\pi_i) = 1 - \frac{H(Re_i)}{\log_2(d)} \tag{7}$$

In Algorithm 1, we compare the explainablity with the predefined threshold (lines 19-21). In case that the $E(\pi_i)$ is less than the given threshold, we perform "Explainability boost", and prioritize the activities of the most relevant agents. A simple code for this part as implemented in our experiment is given in the appendix.

Algorithm 1 Proposed Algorithm for Multi-Agent Systems

1: **Initialize:** θ_i, α, ϵ, δ
2: **while** not converged **do**
3: **for** each agent $i = 1$ to N **do**
4: **for** each task $j = 1$ to M **do**
5: Calculate $Q_i(s_j, a; \theta_i)$ for all $a \in A_i$
6: Calculate $\pi_i(a|s_j; \theta_i)$ using Equation (2)
7: Choose action $a_{ij} \sim \pi_i(a|s_j; \theta_i)$
8: Execute a_{ij}, observe reward r_{ij}
9: **end for**
10: Update θ_i using Equation (3)
11: **end for**
12: Compute $J_i(\pi_i)$ for each agent
13: Compute fairness $F(J_1, J_2, \ldots, J_N)$ using Equation (4)
14: **if** $F(J_1, J_2, \ldots, J_N) < \delta$ **then**
15: Adjust θ_i for best and worst-performing agents
16: **end if**
17: Compute relevance scores Re_i using Equation (5)
18: Compute explainability $E(\pi_i)$ using Equation (7)
19: **if** $E(\pi_i) < \epsilon$ for any i **then**
20: Perform Explainability Boost θ_i for relevant actions
21: **end if**
22: **if** periodic reset condition met **then**
23: Partially reset θ_i to encourage exploration
24: **end if**
25: **if** pruning condition met **then**
26: Prune less relevant connections in θ_i
27: **end if**
28: **end while**
29: **return** $\theta_1, \theta_2, \ldots, \theta_N$

In lines 22-24 of Algorithm 1, we perform "partially reset". We perform this operation in a defined number of iterations to prevents the agents from converging to suboptimal policies by encouraging them to explore a wider range of actions. A code is given in the appendix. We also perform "pruning" (lines 25-27 of Algorithm 1) as a mechanism to enhance the efficiency of the learning process. By utilizing this mechanism, we remove the less relevant connections in the policy parameters θ_i. A sample code performed in our experiment is given in the appendix.

4. Experiment Results

We evaluate the performance of the proposed algorithm and compare it with a typical RL approach for our presented Example 1. We consider $N = 10$ agents and 100 tasks, assuming each task has a

difficulty level (we randomly assign a value in the range $[1-10]$), and each agent has a capability level (we assign randomly to each agent a value in the range $[1-5]$). This example is a simplified resource allocation in MARL. In the proposed method, we perform fairness and explainability adjustments, while the base method algorithm follows a standard reinforcement learning approach without the given explainablity and fairness constraints. The experiment software is available [3]. The parameters used in this experiment are summarized in Table 1. The results are illustrated in Figure 1 and Table 2.

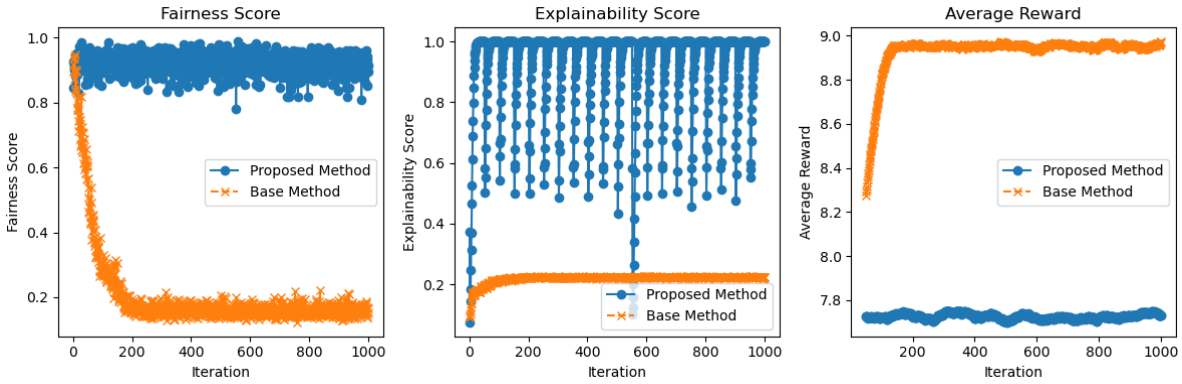

Figure 1: Comparison of the proposed and base method in terms of fairness, explainability and average reward.

Table 1
Parameters used in the performed experiment

Parameter	Value	Description
Number of Agents	10	Number of Agents
Number of Tasks	100	Number of Tasks
Task Difficulty Range	1 - 10	Assigned Randomly
Agent Capability Range	1 - 5	Assigned Randomly
Iterations	1000	Number of Iterations
Learning Rate	0.1	(α) (fixed learning rate)
Discount Factor	0.9	(γ) (reward calculation)
Fairness Threshold	0.8	The threshold for fairness
Explainability Threshold	0.6	The threshold for explainability

Table 2
Comparison of proposed and base method in terms of average system fairness, explainability and reward.

Metric	Proposed Method	Base Method
Average System Fairness	**0.921**	0.206
Average System Explainability	**0.931**	0.219
Average System Reward	7.726	**8.913**
Agent 1 Average Reward	6.342	3.346
Agent 2 Average Reward	5.464	2.720
Agent 3 Average Reward	9.193	2.444
Agent 4 Average Reward	8.407	5.792
Agent 5 Average Reward	9.230	64.673
Agent 6 Average Reward	5.256	1.748
Agent 7 Average Reward	7.936	2.487
Agent 8 Average Reward	9.924	2.357
Agent 9 Average Reward	7.778	1.790
Agent 10 Average Reward	7.728	1.777

[3]https://github.com/ShahbazianR/Fair-XAI-MARL.git

As can be seen in Table 2, the proposed method achieves significantly higher fairness compared to the base method. This means that the rewards are distributed more equitably among agents. The proposed method also excels in explainability. This suggests that the decisions made by agents in the proposed method are more transparent compared to those in typical reinforcement learning. These achievements comes with a price and the total average reward of the base method is higher. However, this slightly decrees in the total reward, enables the system to satisfy fairness, and provides better explainablity.

5. Conclusion

This study proposed a new approach for integrating fairness and explainability in multi-agent reinforcement learning systems. Our approach increased agent transparency by using LRP and systematic fairness requirements. The experimental results showed that, while keeping a competitive average system reward, our proposed method greatly increases both explainability and system fairness, with average scores of 0.921 and 0.931, respectively. The proposed method is a straightforward illustration of how explainability and fairness might be combined in multi-agent systems. However, the proposed algorithm needs to be improved with more dynamic tasks. The efficacy and scalability of proposed method needs to be verified by experimenting on large and varied datasets. A thorough investigation is also needed to evaluate the effects of hyperparameters on the algorithm's performance. As the future work, we expect that adding more dynamic and complex metrics for fairness and explainability could improve the system's equity and transparency. We would investigate the proposed concept in the practical fields, including healthcare, banking, and self-governing systems.

Acknowledgement

We acknowledge the support of the PNRR project FAIR - Future AI Research (PE00000013), Spoke 9 - Green-aware AI, under the NRRP MUR program funded by the NextGenerationEU. One of authors was also funded by the Next Generation EU -Italian NRRP, Mission 4, Component 2, Investment 1.5, call for the creation and strengthening of 'Innovation Ecosystems', building 'Territorial R&D Leaders' (Directorial Decree n. 2021/3277) - project Tech4You - Technologies for climate change adaptation and quality of life improvement, n. ECS0000009. This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them.

Declaration on Generative AI

During the preparation of this work, the authors used Quillbot and Grammarly in order to: Grammar and spelling check, and re-phrasing sentences. After using these tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] C. M. R. Haider, C. Clifton, Y. Zhou, Unfair ai: It isn't just biased data, in: 2022 IEEE International Conference on Data Mining (ICDM), IEEE, 2022, pp. 957–962.

[2] J. Kleinberg, S. Mullainathan, M. Raghavan, Inherent trade-offs in the fair determination of risk scores, arXiv preprint arXiv:1609.05807 (2016).

[3] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, A survey on bias and fairness in machine learning, ACM computing surveys (CSUR) 54 (2021) 1–35.

[4] S. Caton, C. Haas, Fairness in machine learning: A survey, ACM Computing Surveys 56 (2024) 1–38.

[5] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, et al., Explainable ai (xai): Core ideas, techniques, and solutions, ACM Computing Surveys 55 (2023) 1–33.
[6] A. Reuel, D. Ma, Fairness in reinforcement learning: A survey, arXiv preprint arXiv:2405.06909 (2024).
[7] S. Milani, N. Topin, M. Veloso, F. Fang, Explainable reinforcement learning: A survey and comparative review, ACM Computing Surveys 56 (2024) 1–36.
[8] O. Parraga, M. D. More, C. M. Oliveira, N. S. Gavenski, L. S. Kupssinskü, A. Medronha, L. V. Moura, G. S. Simões, R. C. Barros, Fairness in deep learning: A survey on vision and language research, ACM Computing Surveys (2023).
[9] Y. Gao, S. Gu, J. Jiang, S. R. Hong, D. Yu, L. Zhao, Going beyond xai: A systematic survey for explanation-guided learning, ACM Computing Surveys 56 (2024) 1–39.
[10] A. Cimpean, P. Libin, Y. Coppens, C. Jonker, A. Nowé, Towards fairness in reinforcement learning, in: Proceedings of the Adaptive and Learning Agents Workshop (ALA 2023), 2023, pp. 1–5.
[11] A. Pozanco, D. Borrajo, Fairness in multi-agent planning, arXiv preprint arXiv:2212.00506 (2022).
[12] S. Jabbari, M. Joseph, M. Kearns, J. Morgenstern, A. Roth, Fairness in reinforcement learning, in: International conference on machine learning, PMLR, 2017, pp. 1617–1626.
[13] P. Weng, Fairness in reinforcement learning, arXiv preprint arXiv:1907.10323 (2019).
[14] U. Siddique, P. Weng, M. Zimmer, Learning fair policies in multi-objective (deep) reinforcement learning with average and discounted rewards, in: International Conference on Machine Learning, PMLR, 2020, pp. 8905–8915.
[15] J. Chen, Y. Wang, T. Lan, Bringing fairness to actor-critic reinforcement learning for network utility optimization, in: IEEE INFOCOM 2021-IEEE Conference on Computer Communications, IEEE, 2021, pp. 1–10.
[16] M. Rodriguez-Soto, M. Lopez-Sanchez, J. A. Rodriguez-Aguilar, Multi-objective reinforcement learning for designing ethical multi-agent environments, Neural Computing and Applications (2023) 1–26.
[17] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, K.-R. Müller, Layer-wise relevance propagation: an overview, Explainable AI: interpreting, explaining and visualizing deep learning (2019) 193–209.
[18] N. Rezaeinia, J. C. Góez, M. Guajardo, On efficiency and the jain's fairness index in integer assignment problems, Computational Management Science 20 (2023) 42.

Appendix

```
# Adjust fairness
if fairness_score < fairness_threshold:
    max_agent = np.argmax(agent_rewards)
    min_agent = np.argmin(agent_rewards)
   adjustment = (agent_rewards[max_agent] - agent_rewards[min_agent]) * alpha
    policy_params[:, max_agent] -= adjustment
    policy_params[:, min_agent] += adjustment

# Explainability boost
for task_id in range(num_tasks):
    top_agent = np.argmax(relevance_scores[task_id])
    policy_params[task_id] *= 0.5
    policy_params[task_id][top_agent] += 0.5

# Periodic Reset
if iteration % 100 == 0:
   policy_params = 0.7 * policy_params + 0.3 * np.random.rand(num_tasks, num_agents)
```

```
# pruning
if iteration % 50 == 0:
    threshold = np.percentile(policy_params, 50)
    policy_params[policy_params < threshold] = 0
```