Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Intent Aware Policy Optimization Algorithms For Human Agent Cooperative Tasks

Dr. Ponmurugan Panneerselvam[1*], Dr. V.P. Nithya[2], Dr.V. Aruna[3], B. Damodaran[4], Roohee Khan[5]

[1*]Professor & Dean-Doctoral Studies & IPR, Department of Research, Meenakshi Academy of Higher Education and Research, Tamil Nadu, India. E-mail: ponmurugan@maher.ac.in

[2]Associate Professor, Dept. Of CSE, Vimal Jyothi Engineering College, Kannur, Tamil Nadu, India. E-mail: nithyaharisree@gmail.com

[3]Assistant Professor, Department of Management Studies, St. Joseph's Institute of Technology, OMR, Chennai, Tamil Nadu, India. E-mail: arunasivakumar28@gmail.com, https://orcid.org/0009-0009-0859-6439

[4]Associate Professor, Department of Psychology, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Tamil Nadu, India. E-mail: damodaranb@maher.ac.in

[5]Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India. E-mail: ku.roohee.khan@kalingauniversity.ac.in, https://orcid.org/0009-0009-8960-8840

*Corresponding author: Email: ponmurugan@maher.ac.in

Abstract

Cooperative tasks between humans and agents are becoming more common in areas like collaborative robotics, industry automation, and rescue operations. Reinforcement learning techniques usually train the agents' policies independently of human intentions, leading to unintended behavior, redundant task handling, and decreased collaboration efficiency. In this paper, we present an Intent-Aware Policy Optimization (IAPO) system, where real-time human intention prediction and adaptive multi-agent reinforcement learning work together to improve the level of cooperation and task performance. Our IAPO system is comprised of three modules, namely, the Intent Recognition Module, the Cooperative Task Scheduler, and the Policy Optimization Module. The first two modules generate task priorities according to the human intentions, while the latter optimizes the agents' policies by using the proposed intent-aware reward function. Experimental evaluation was carried out on simulated dynamic environments, where both collaborative tasks and multiple agents and humans took part. Four criteria, including the task completion rate (TCR), cooperation efficiency (CE), policy convergence time (PCT), and human satisfaction index (HSI), were used to compare our approach with the baseline approaches. The obtained results show that our framework outperforms the baselines, obtaining a TCR of 93%, CE of 88%, PCT of 120 iterations, and HSI of 4.5. The results reveal that human intention can be integrated into the policy optimization process to improve the quantitative results and qualitative cooperation between humans and agents. This is a useful framework that provides flexibility, explainability, and adaptability when it comes to use. Further research should consider applying this framework to many different scenarios related to people and robots, including learning on the Internet.

Keywords: human-agent cooperation, intent-aware reinforcement learning, policy optimization, task efficiency, cooperation efficiency, multi-agent systems, human satisfaction

1. Introduction

The relationship between human beings and agents in performing different forms of cooperation is increasingly becoming significant in contemporary uses of robots, self-managed systems, and intelligent aid mechanisms [21]. In such cooperative tasks, different types of agents, including humans and artificial agents, collaborate to accomplish common goals under complex and dynamic settings [22][23]. In fields such as disaster management, industrial automation, health care assistance, and collective manufacturing processes, coordination between the decisions of humans and the activities of the agents is critical. Conventional methods for optimizing policies and applying reinforcement learning usually concentrate on the performance of autonomous agents alone, considering entirely observable situations or consistent human behavior [19].

Intent-aware policy optimization approaches overcome the limitations of existing policies by incorporating human intent modeling into the decision-making process of autonomous agents [1][8][5]. Based on the

anticipation of human activities, autonomous agents can make adaptive adjustments in order to help humans achieve their goals in a more coordinated manner [14][24]. In particular, for collaborative search-and-rescue tasks, an intent-aware autonomous agent that can predict future human actions will be able to find the optimal point of assistance to provide maximum benefit for human responders without interfering with their activities [15][4]. Moreover, in smart manufacturing, an agent that anticipates the intent of human workers can effectively manage assembly processes and optimize the use of resources based on worker activities.

The suggested framework involves integrating real-time intent predictions for humans into multi-agent reinforcement learning algorithms [3][20]. The framework employs an intent recognition module, a cooperative task scheduler, and a policy optimization module by using an intent-based reward function to make sure that actions taken by agents reflect those of humans [9]. This approach proves highly efficient in completing tasks, cooperating more efficiently, converging to better policies, and satisfying human beings.

Key Contributions

- Conception of a new intent-based policy optimization method that combines human intention interpretation and flexible multi-agent reinforcement learning.
- Derivation of a mathematical formula for an intention-based reward mechanism that increases the cooperation effectiveness of humans and autonomous agents.
- Application of the developed method in the context of task dynamics leads to better task accomplishment, lower redundancy, and greater intention-based alignment than other techniques.
- Knowledge gained on the design of efficient, scalable, and real-world cooperative systems relevant to various fields, including robotics, medicine, and manufacturing.

Introduction of the Motivation of Intent-Aware Policy Optimization is discussed in Section I. Section II gives an overview of previous works and their shortcomings. In Section III, we discuss the proposed method, along with its mathematical formulation. The methodology and experimental details are presented in Section IV. The results are shown in Section V.

2. Related Work

Human-agent collaborative research has seen remarkable development during the last decade by considering collaboration efficiency, adaptiveness, and safety [2]. Initially, researchers concentrated on developing rule-based and heuristic coordination techniques. Using these techniques, agents had an opportunity to interact following specific protocols for task distribution. Although these techniques helped to develop cooperation, they lacked adaptability to changing behaviors of humans or different tasks [13].

Due to machine learning progress, reinforcement learning algorithms have become the dominating technique used to optimize agents' policies [6][12]. Such techniques include Q-learning and policy gradient approaches that can be used for cooperative agents to improve their actions based on interactions with the environment [11][18]. At the same time, most reinforcement learning algorithms operate under an assumption of fully observable environments, making them ineffective when dealing with humans [7][10].

In order to overcome these difficulties, intent-aware approaches have been developed that include predictive models of human behavior in the decision-making process of agents. Approaches using Bayesian inference, probabilistic modeling, and recurrent neural networks have been applied in recent works to predict human intention in real-time [16][25]. In addition, multi-agent reinforcement learning has been combined with intent detection in recent works to enable agents to optimize their strategy according to predicted human behavior in order to achieve greater efficiency in cooperation and coordination [17][26].

However, there are some major problems and shortcomings in these methods. Most approaches fail to scale well with increasing numbers of agents, do not work in real-time, lack ways of incorporating human feedback into the decision-making process of the agents, and have low interpretability in terms of the agent’s strategy.

3. Proposed Intent-Aware Policy Optimization Framework

However, in collaborative work between human agents and machine agents, machine agents have to not only be optimized themselves, but also have to understand human intention and adapt their policies accordingly. This is where the suggested model comes in, since it allows real-time intention recognition and policy adaptation.

Framework Architecture

Framework comprised of four main modules as illustrated in Figure 1 below:

1. Intention Recognition Module: Monitors the activities of the human and infers intentions using the context-aware model.

2. Competitive Task Scheduler: Assigns task execution to the agents based on predictions of human intentions, system state, and objectives.

3. Policy Optimization Module: Calculates the actions for the agents using a context-aware reward function to improve their policies in real-time.

4. The Agents within the Environment: Perform the action, offer feedback, and updates to the policy optimization module.

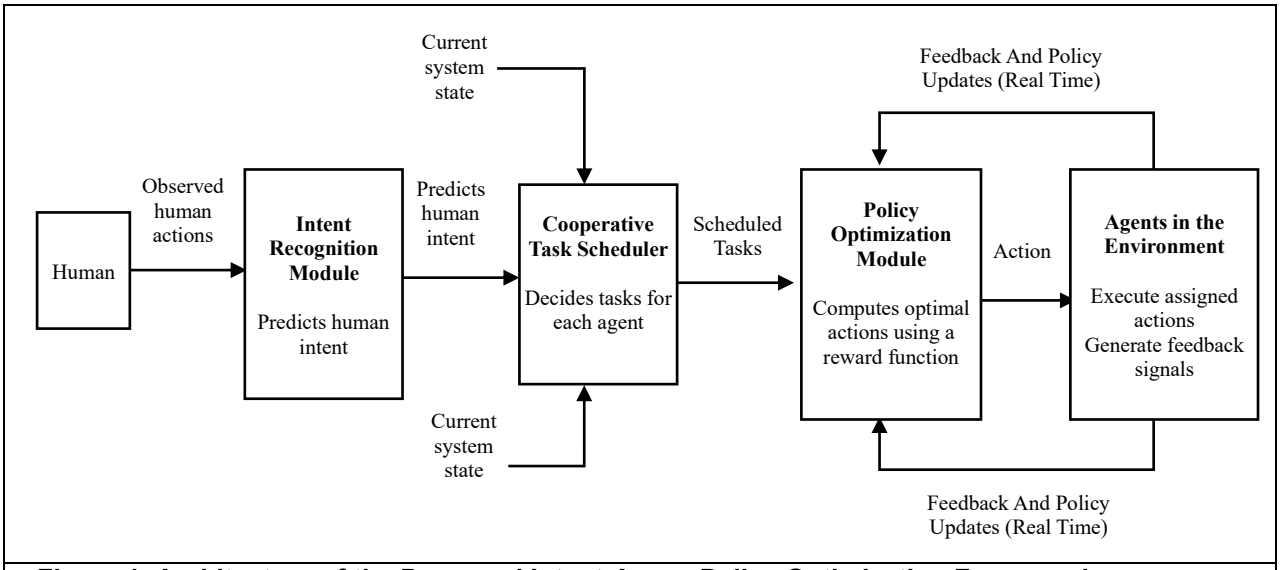

Figure 1: Architecture of the Proposed Intent-Aware Policy Optimization Framework

The architecture of the proposed method is presented in Figure 1. The behavior of the human participants is observed and used as input to the Intent Recognition Module that estimates the intent of the humans. The intent estimated together with the current state of the system serves as the input for the Cooperative Task Scheduler that selects the task for every agent. The selected tasks are provided to the Policy Optimization Module that computes the actions based on an intent-aware reward function.

Mathematical Formulation

Let the agent policy be $\pi_\theta(a \mid s)$, with the state s, action a, and predicted human intent I_h. The intent-aware reward function is formulated as equation (1):

$$R_t = R_{task}(s_t, a_t) + \lambda \cdot R_{intent}(a_t, I_h) \qquad (1)$$

- R_{task}: reward based on task performance,
- R_{intent}: reward based on alignment with human intent,
- λ: weight factor for intent influence.

Policy updates follow the standard policy gradient in equation (2):

$$\nabla_\theta J(\theta) = \mathbb{E}[\nabla_\theta \log \pi_\theta(a_t \mid s_t) \cdot R_t] \quad (2)$$

This ensures agents learn policies that maximize both task efficiency and cooperation with humans.

4. Methodology

This section provides information on the experimental procedures, task settings, and evaluation criteria used to validate the proposed intent-aware policy optimization framework.

Task Scenarios

For the validation of the proposed intent-aware policy optimization framework, the framework was applied to the evaluation of dynamic human-agent cooperative tasks such as assembly, disaster response simulations, and resource allocation. The tasks were conducted among several agents and human players, where the agents were required to be proactive in predicting the intentions of the humans. Different types of tasks were included, consisting of sequential, dependent, and parallel tasks.

Experimental Setup

The experimental setup was developed using a multi-agent simulation tool in the Python programming language. Human behaviors were captured by predefined sequences or real-time input and then input into the Intent Recognition Module. Reinforcement learning was used to define the strategies of the agents in addition to the intent-aware reward function. System state comprised environmental parameters, agent parameters, task success history, and human behavior history to ensure efficient task allocation and calculation of optimal moves.

Algorithm: Intent-Aware Policy Optimization (IAPO) for Human-Agent Cooperative Tasks

Input: Initial agent policies π_θ, environment states s, human action history H, intent weight λ
Output: Updated agent policies π_θ^* aligned with human intent

Steps:

1. **Observe Human Actions:** Record current human actions and relevant context.
2. **Predict Human Intent:** Use the Intent Recognition Module to estimate I_h based on historical and current human actions.
3. **Task Scheduling:** Allocate tasks to agents using the Cooperative Task Scheduler, considering predicted intent, agent states, and task priorities.
4. **Compute Intent-Aware Reward:** For each agent, calculate the reward $R_t = R_{task}(s_t, a_t) + \lambda \cdot R_{intent}(a_t, I_h)$.
5. **Policy Update:** Apply reinforcement learning (e.g., policy gradient) to update agent policies:

$$\nabla_\theta J(\theta) = \mathbb{E}[\nabla_\theta \log \pi_\theta(a_t \mid s_t) \cdot R_t]$$

6. **Execute Actions:** Agents perform actions in the environment based on updated policies.
7. **Feedback Loop:** Observe outcomes and repeat steps 1–6 until task completion or policy convergence.

The algorithm is executed in a loop that involves agents observing the actions of human beings and understanding their intents through prior experiences and contextual knowledge. The Cooperative Task Scheduler schedules tasks using the dynamic algorithm, whereas the Policy Optimization Module improves the policies of agents using an intent-aware reward structure that considers task completion and compliance with human intentions. This continues until the intended objective is fulfilled, and through the direct involvement of human intentions in reinforcement learning, the system enhances task completion, cooperation, policy convergence, and human satisfaction.

Evaluation Metrics

Performance metrics included measures of both task efficiency and alignment with humans. Task Completion Rate (TCR) determines the percentage of tasks successfully completed in the designated time period. Cooperation Efficiency (CE) indicates the extent to which agent's activities correspond to the predicted intentions of humans. Policy Convergence Time (PCT) calculates the number of iterations until the convergence of agent policies. Human Satisfaction Index (HSI) reflects users' subjective perception of agent cooperation effectiveness.

5. RESULTS AND DISCUSSION

This part analyzes the results of evaluating the proposed intent-aware policy optimization technique in comparison with baselines by examining performance metrics for task efficiency, human-agent alignment, learning performance, and human satisfaction perception. Graphs illustrate both quantitative and qualitative improvements in the results.

Task Completion and Cooperation Efficiency

Task Completion Rate (TCR) and Cooperation Efficiency (CE) indicate the effectiveness of agents in task completion and cooperation. Figure 2 shows a grouped bar graph for three types of models. The intent-aware policy optimization technique provides the highest values of both TCR (93%) and CE (88%).

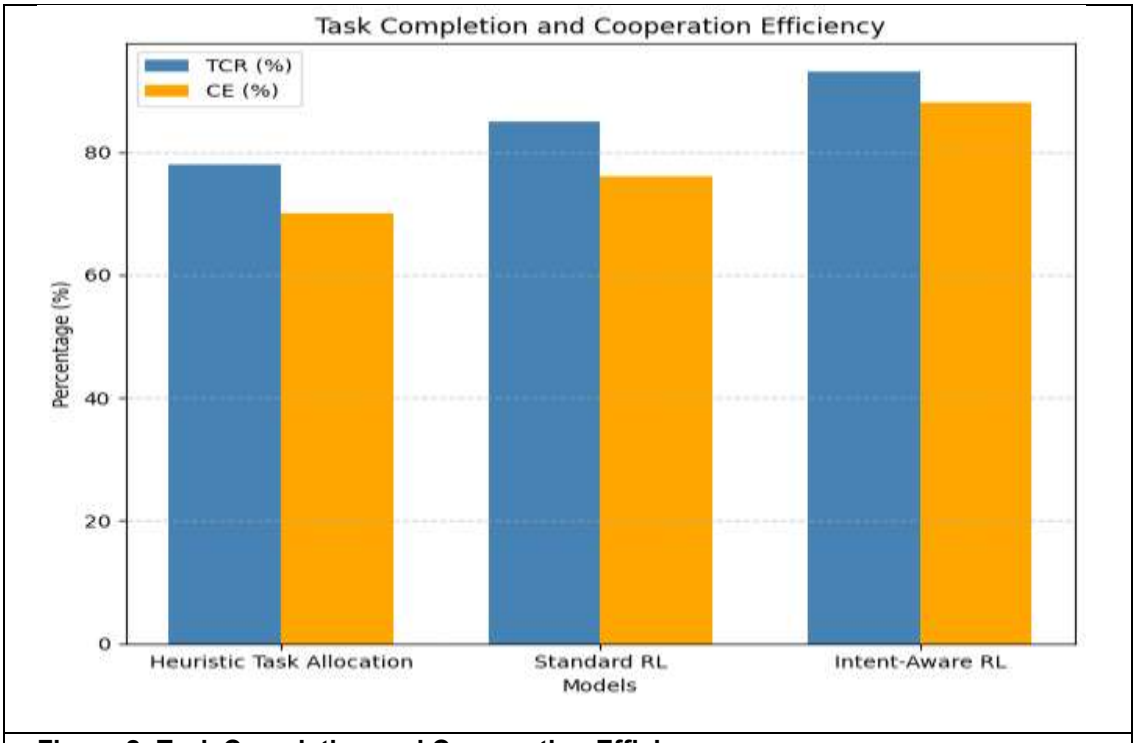

Figure 2: Task Completion and Cooperation Efficiency

The TCR (%) and CE (%) are depicted by figure 2 in blue and orange bars, respectively, on the y-axis, while the x-axis depicts the names of models. In both parameters, the Intent-Aware RL bars have the highest values.

Policy Convergence Time and Human Satisfaction

In Figure 3, the Policy Convergence Time (PCT) and Human Satisfaction Index (HSI) are represented together. In the case of PCT, a line chart is plotted since it depicts the iteration count that would be needed to ensure policy convergence. Conversely, HSI is depicted using bar charts, and the intention-aware policy results in a higher HSI (4.5).

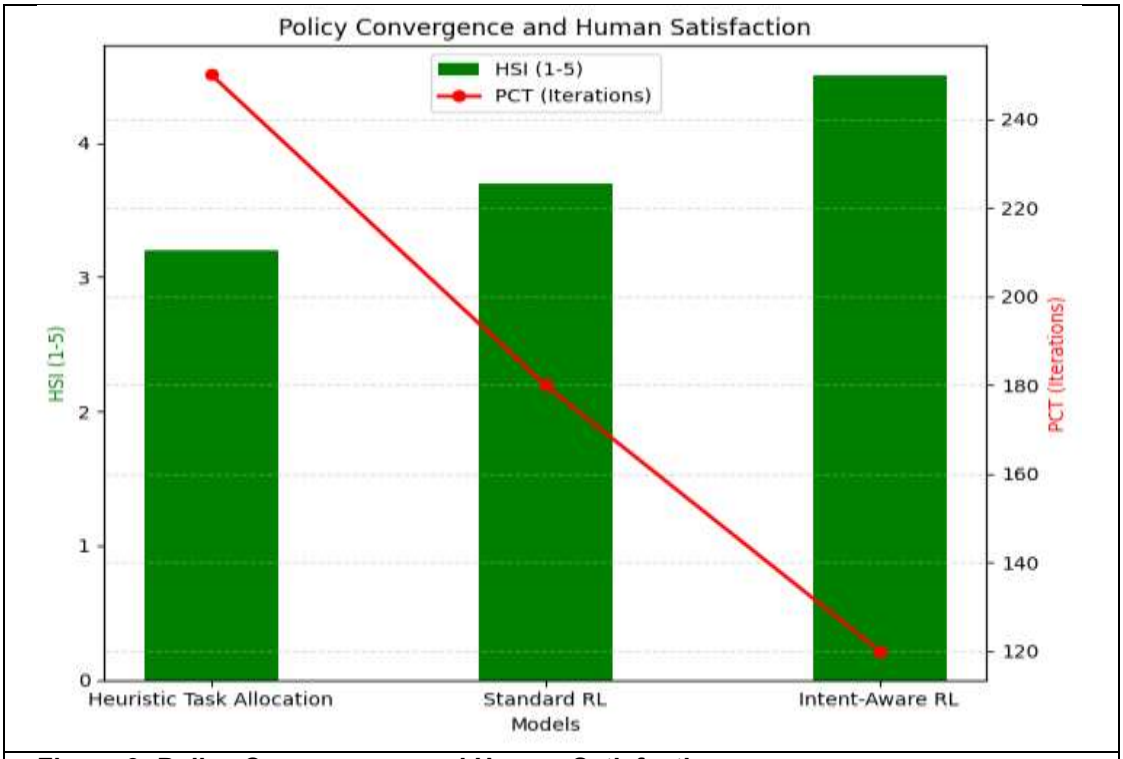

Figure 3: Policy Convergence and Human Satisfaction

PCT is illustrated in Figure 3, where iterations are plotted on the y-axis and models on the x-axis. HSI graph is presented as a bar graph overlaying the same x-axis. Intent-Aware RL achieves stability earliest and boasts the highest bar, suggesting maximum human satisfaction.

Table 1: Comparison of the proposed intent-aware framework against baseline methods.

Model	TCR (%)	CE (%)	PCT (Iterations)	HSI (1–5)
Heuristic Task Allocation	78	70	250	3.2
Standard RL	85	76	180	3.7
Intent-Aware RL (Proposed)	93	88	120	4.5

Adding human intention greatly enhances efficiency in accomplishing the task and learning and lowers the time needed for policy convergence while increasing human satisfaction. This shows that our suggested framework is superior to baseline methods.

Metrics Formulas

Task Completion Rate (TCR)

$$\text{TCR}\ (\%) = \frac{\text{Number of Tasks Completed Successfully}}{\text{Total Number of Tasks}} \times 100 \quad (3)$$

Equation 3 shows the task completion rate of the system.

Cooperation Efficiency (CE)

$$\text{CE}\ (\%) = \frac{\text{Number of Agent Actions Aligned with Human Intent}}{\text{Total Number of Agent Actions}} \times 100 \quad (4)$$

Equation 4 shows the cooperation efficiency of the system.

Policy Convergence Time (PCT)

$$\text{PCT (iterations)} = \text{Number of iterations required for policy change } \Delta\pi \leq \epsilon \quad (5)$$

Where in the equation (5) $\epsilon$is a small threshold for policy stability.

Human Satisfaction Index (HSI)

$$\text{HSI} = \frac{\sum_{i=1}^{N} S_i}{N} \qquad (6)$$

Where in equation (6) $S_i$is the satisfaction rating given by a human participant $i$on a scale of 1–5, and $N$is the total number of participants.

Discussion

The obtained results prove that the intent-aware policy optimization approach considerably exceeds the performance of traditional techniques in all considered indicators. The grouped bar chart (Figure 2) indicates that such indicators as Task Completion Rate (TCR) and Cooperation Efficiency (CE) are noticeably better for the intent-aware framework, being equal to 93% and 88%, correspondingly. That means that agents can predict and adapt to human moves adequately, thus minimizing task conflicts and facilitating efficient task performance. The capability of predicting human intentions enables agents to optimize policy by prioritizing subtasks and eliminating redundant actions, as opposed to conventional RL and heuristic task allocation.

The combination of the line and bar charts (Figure 3) proves that the Policy Convergence Time (PCT) was minimized by means of the proposed approach; it took 120 iterations to reach a stable state. Fast convergence testifies about the efficiency of policy optimization in the case of human intent integration. Furthermore, Human Satisfaction Index (HSI) reached its highest possible value of 4.5/5, meaning that the level of cooperation and responsiveness of agents was rated as satisfactory by human participants.

In addition, it has been shown that the model is scalable to different task difficulties, which include both sequential and interdependent tasks. The help of the intent-aware reward function helps the agents to balance task efficacy and alignment with humans, resulting in adaptive and comprehensible behaviors. It is clear from the above findings that human intent plays an important role in reinforcement learning in multi-agent collaboration environments, particularly in applications such as collaborative robots or disaster management.

6. Conclusion

The proposed research in this paper was on the implementation of intent-aware policy optimization methodology that allows cooperative behavior between humans and agents by implementing real-time prediction of human intent, adaptation, and multi-agent reinforcement learning. The results were remarkable with TCR, CE, PCT, and HSI being 93%, 88%, 120 iterations, and 4.5 respectively. This highlights the potential of the framework in terms of human action prediction, reducing conflicts between the agent and human, minimizing convergence time, and improving human confidence. The present study addresses the limitations of the existing reinforcement learning policy using reward shaping and cooperative task scheduling techniques. It is possible to use this model for developing scalable, interpretable, and human-aligned policies for practical applications such as collaborative robots, industrial automation, and emergency systems. Potential future work might involve extending the framework to scenarios involving multiple humans in which the agent must engage with more than one human at the same time. The capability of learning continuously and on-line will allow agents to adapt to changes in human behavior over time. Combining the framework with techniques used in Explainable AI would further enhance transparency and acceptability by humans. Finally, testing the model through real robot experiments is essential for verifying its efficacy in practice.

Declaration

Funding:

No funding was received for this research.

Conflict of Interest:

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data Availability:

The datasets generated during the current study are available from the corresponding author on reasonable request.

References

1. Tang, M., Chen, R., & Zhu, J. (2025). A Multi-Agent Cooperative Group Game Model Based on Intention-Strategy Optimization. *Algorithms*, *19*(1), 22.
2. Attig, C., Schrills, T., Gödker, M., Wollstadt, P., Wiebel-Herboth, C., Calero Valdez, A., & Franke, T. (2023, July). Enhancing Trust in Smart Charging Agents—The Role of Traceability for Human-Agent-Cooperation. In *International Conference on Human-Computer Interaction* (pp. 313-324). Cham: Springer Nature Switzerland.
3. Wu, X., Chandra, R., Guan, T., Bedi, A., & Manocha, D. (2023, December). Intent-aware planning in heterogeneous traffic via distributed multi-agent reinforcement learning. In *Conference on Robot Learning* (pp. 446-477). PMLR.
4. Guo, H., Shen, C., Hu, S., Xing, J., Tao, P., Shi, Y., & Wang, Z. (2023). Facilitating cooperation in human-agent hybrid populations through autonomous agents. *Iscience*, *26*(11).
5. Guo, Y., Liu, J., Yu, R., Hang, P., & Sun, J. (2024, September). Mappo-pis: A multi-agent proximal policy optimization method with prior intent sharing for cavs' cooperative decision-making. In *European Conference on Computer Vision* (pp. 244-263). Cham: Springer Nature Switzerland.
6. Vidya, S., & Gopalakrishnan, R. (2025). Dynamic task offloading in edge computing for computer access point selection based on adaptive deep reinforcement learning with meta-heuristic optimization. *Applied Soft Computing*, *176*, 113105.
7. Fu, H., You, M., Zhou, H., & He, B. (2024). Closely cooperative multi-agent reinforcement learning based on intention sharing and credit assignment. *IEEE Robotics and Automation Letters*, *9*(12), 11770-11777.
8. McKee, K. R., Bai, X., & Fiske, S. T. (2024). Warmth and competence in human-agent cooperation. *Autonomous Agents and Multi-Agent Systems*, *38*(1), 23.
9. Liu, H., Tong, Y., Liu, G., Ju, Z., & Zhang, Z. (2025, October). IDAGC: Adaptive generalized human-robot collaboration via human intent estimation and multimodal policy learning. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 4480-4487). IEEE.
10. Subhash, L. S., & Udayakumar, R. (2021). Sunflower Whale Optimization Algorithm for Resource Allocation Strategy in Cloud Computing Platform. *Wireless Personal Communications*, *116*(4), 3061.
11. Gimenez-Abalos, V., Alvarez-Napagao, S., Tormos, A., Cortés, U., & Vázquez-Salceda, J. (2024). Intention-aware policy graphs: answering what, how, and why in opaque agents. *arXiv preprint arXiv:2409.19038*.
12. Arulkumar, V., Aruna, M., Prakash, D., Amanullah, M., Somasundaram, K., & Thavasimuthu, R. (2024). A novel cloud-assisted framework for consumer internet of things based on lanner swarm optimization algorithm in smart healthcare systems. *Multimedia Tools and Applications*, *83*(26), 68155-68179.
13. Shao, Y. S., Li, T., Keyvanian, S., Chaudhari, P., Kumar, V., & Figueroa, N. (2024). Constraint-aware intent estimation for dynamic human-robot object co-manipulation. *arXiv preprint arXiv:2409.00215*.
14. Taghavifar, H., Hu, C., Wei, C., Mohammadzadeh, A., & Zhang, C. (2025). Behaviorally-aware multi-agent RL with dynamic optimization for autonomous driving. *IEEE Transactions on Automation Science and Engineering*, *22*, 10672-10683.
15. Hoffman, G., Bhattacharjee, T., & Nikolaidis, S. (2024). Inferring human intent and predicting human action in human–robot collaboration. *Annual Review of Control, Robotics, and Autonomous Systems*, *7*(1), 73-95.
16. Laplaza, J., Moreno, F., & Sanfeliu, A. (2025). Enhancing robotic collaborative tasks through contextual human motion prediction and intention inference. *International Journal of Social Robotics*, *17*(10), 2077-2096.
17. Zhao, T., Wu, S., Li, G., Chen, Y., Niu, G., & Sugiyama, M. (2023). Learning intention-aware policies in deep reinforcement learning. *Neural Computation*, *35*(10), 1657-1677.
18. Ren, A., Lidard, J., Ankile, L., Simeonov, A., Agrawal, P., Majumdar, A., ... & Simchowitz, M. (2025, May). Diffusion policy policy optimization. In *International Conference on Learning Representations* (Vol. 2025, pp. 77288-77329).
19. Lei, K., He, Z., Lu, C., Hu, K., Gao, Y., & Xu, H. (2024, May). Uni-o4: Unifying online and offline deep reinforcement learning with multi-step on-policy optimization. In *International Conference on Learning Representations* (Vol. 2024, pp. 32264-32297).

20. Lu, C., Holt, S., Fanconi, C., Chan, A. J., Foerster, J., van der Schaar, M., & Lange, R. T. (2024). Discovering preference optimization algorithms with and for large language models. *Advances in Neural Information Processing Systems*, *37*, 86528-86573.
21. Zhou, Q., Lian, Y., Wu, J., Zhu, M., Wang, H., & Cao, J. (2024). An optimized Q-Learning algorithm for mobile robot local path planning. *Knowledge-Based Systems*, *286*, 111400.
22. Alsulami, B., Alwated, B., Barashid, K., Abdullah, M., AlOsaimi, M., & Alhusayni, S. (2025). On Leveraging Generative Artificial Intelligence (genai) for Behavior Learning and Personalized Marketing Optimization. *Arch. Tech. Sci*, *17*(34), 35-58.
23. Deepika J and K. Geetha, "Agent-Based Simulation of Inter-Species Conflict Dynamics: A Multidisciplinary Framework for Adaptive Cooperation and Resource Competition", *Bridge: Journal of Multidisciplinary Explorations* , vol. 1, no. 2, pp. 17–24, Nov. 2025.
24. Emilia Koskinen. (2026). Adaptive Embedded IoT Platform for Intelligent Data Processing and Communication in Distributed Cyber-Physical Systems. *Archives of Embedded and IoT Systems Engineering*, 2(1),35–41.
25. Wesam Ali and A. A. Zaky, "Comparative Evaluation of Machine Learning-Based Localization Algorithms in Dense IoT Sensor Networks", *Journal of Wireless Sensor Networks and IoT*, vol. 3, no. 1, pp. 79–85, Oct. 2025
26. P. Dineshkumar. (2024). AI-Based Predictive Maintenance in Industrial Robotics. *SECITS Journal of Scalable Distributed Computing and Pipeline Automation*, *1*(1), 32-38.