Proceedings of CONF-SEML 2025 Symposium: Machine Learning Theory and Applications
DOI: 10.54254/2755-2721/158/2025.TJ23483

The Application and Challenges of Deep Reinforcement Learning in Complex Environments

Wenhan Wang

Electronic Information Engineering, China University of Petroleum (Beijing), Beijing, China
dndisjs138@outlook.com

Abstract: Deep reinforcement learning (DRL), as an important branch of machine learning, has shown strong potential for application in complex environmental decision-making problems in recent years. This article systematically reviews the current application status and development trends of DRL in fields such as gaming and virtual environments, robot control, resource management, and healthcare. Through a comprehensive analysis of existing literature, this paper has summarized the technical roadmap and solutions of DRL in addressing core challenges such as high-dimensional state spaces, sparse rewards, and partial observability, including hierarchical reinforcement learning frameworks, mixed reward designs, and memory based reinforcement learning methods. Meanwhile, this article delves into the opportunities and challenges faced by cutting-edge research directions such as multi-agent systems, security, and interpretability. Based on current research progress, possible paths for the future development of DRL have been proposed, including improving algorithm robustness, integrating interdisciplinary methods, and engineering considerations in practical deployment, and providing reference for the further development of deep reinforcement learning.

Keywords: Deep reinforcement learning, Complex environment, Multi agent system, Sparse rewards, interpretability

1. Introduction

In recent years, with the rapid development of artificial intelligence technology, reinforcement learning (RL) has shown great potential in many fields as a method of obtaining optimal decision strategies through interaction with the environment. Especially when deep learning techniques are introduced into reinforcement learning, Deep reinforcement learning (DRL) rapidly emerges, marking the birth of a new algorithmic paradigm. This combination not only fully utilizes the expressive power of deep neural networks in high-dimensional data, but also overcomes the limitations of traditional RL in handling large-scale state spaces, making decision-making problems in complex environments possible [1]. The rise of deep reinforcement learning began with the breakthrough of deep Q-networks (DQN), which achieved performance beyond traditional algorithms in multiple classical tasks by approximating Q-value functions using deep neural networks. Since then, various algorithms based on policy gradients, participant criticism, etc. have been proposed and continuously applied in games (AlphaGo, AlphaStar, and OpenAI Five), multi robot systems (collaborative path planning and dynamic task allocation) [2], energy management, and healthcare. Meanwhile, deep reinforcement learning not only demonstrates powerful decision-making abilities

DOI: 10.54254/2755-2721/158/2025.TJ23483

in virtual environments, but also gradually permeates into complex real-world scenarios, providing new ideas for solving practical problems. However, complex environments often come with various challenges, such as high-dimensional state spaces, sparse rewards, partial observability, and dynamic changes. This not only makes the learning process exceptionally difficult, but also places higher demands on the algorithm's generalization ability and stability. For example, when faced with sparse rewards, agents often struggle to obtain sufficient positive feedback, resulting in slow convergence speed; How to accurately capture and remember key information in partially observable or constantly changing environments has become another urgent problem to be solved.

This article aims to comprehensively review the current application status, key technologies, and future challenges of deep reinforcement learning in complex environments. Firstly, the basic concepts of reinforcement learning and the development history of deep reinforcement learning will be introduced, and its advantages and limitations in dealing with high-dimensional state spaces and sparse reward problems will be elaborated; Secondly, this article will also focus on discussing cutting-edge technologies such as multi-agent collaboration, meta reinforcement learning, security, and interpretability, attempting to provide theoretical foundations and practical guidance for how to build more efficient, robust, and intelligent decision-making systems in the future.

2. The fundamental theory of DRL

2.1. Mathematical framework for reinforcement learning

The theoretical basis of reinforcement learning is based on Markov Decision Process (MDP), which describes the interaction process between the agent and the environment through five elements: state space S, action space A, state transition probability P, reward function R, and discount factor γ. The core assumption of MDP is Markov property, which means that the future state depends only on the current state and action. The calculation of state value functions and action value functions based on the Bellman equation provides a mathematical foundation for reinforcement learning, while the limitations of traditional table methods in dealing with high-dimensional state spaces have promoted the combination of deep neural networks and reinforcement learning, forming a new paradigm of deep reinforcement learning [1, 3].

2.2. Core algorithms of deep reinforcement learning

The core algorithm system of deep reinforcement learning mainly includes three representative methods: DQN approximates the Q-function through neural networks and introduces experience replay and target network mechanisms, effectively solving the limitations of traditional Q-learning in high-dimensional space; The strategy gradient method directly optimizes strategy parameters, which is particularly suitable for dealing with continuous action space problems; The actor critic algorithm cleverly combines the advantages of value function evaluation and strategy optimization, and constructs a more efficient reinforcement learning framework through the collaborative work of the Actor and Critic components. The improved versions derived from these basic algorithms, such as Asynchronous Advantage Actor-Critic (A3C) and Proximal Policy Optimization (PPO), further enhance the training efficiency and stability of the algorithms [1, 2, 4].

2.3. Extension techniques for deep reinforcement learning

To improve the performance of deep reinforcement learning algorithms, researchers have developed multiple key technologies: experience replay and target network mechanisms effectively solve the problems of data correlation and training stability; The dual Q-learning method alleviates the problem of Q-value overestimation by decoupling the process of action selection and value evaluation;

DOI: 10.54254/2755-2721/158/2025.TJ23483

Hierarchical reinforcement learning reduces the difficulty of solving complex problems through task decomposition strategies; And meta reinforcement learning enhances the algorithm's adaptability to new tasks. The collaborative development of these technologies greatly expands the application boundaries of deep reinforcement learning [1, 3, 4].

3. The application of DRL in complex environments

From Figure 1, it can be seen that the application of dimension of state space is mainly for games, the application of real-time requirements is mainly for robots, and computational complexity and privacy sensitivity are mainly used in the field of healthcare.

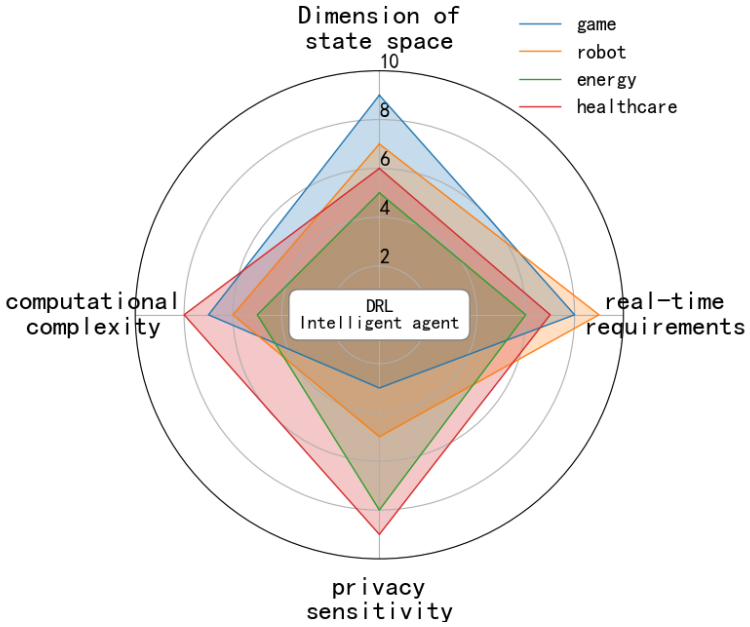

Figure 1: Overview of multi domain applications of DRL in complex environments (picture credit: original)

3.1. Games and virtual environments

DRL is driving a revolution in real-time interactive experiences in gaming and virtual environments. Intelligent systems represented by AlphaGo and AlphaStar have demonstrated the potential of DRL in complex strategic decision-making, while OpenAI Five has validated the possibility of multi-agent collaboration. These systems have achieved decision-making abilities beyond humans through the DRL algorithm, particularly excelling in complex strategy games. In real-time game streaming scenarios, DRL is used to dynamically optimize transmission bit rates to improve user experience quality (QoE). For example, Del Rio et al. proposed a multi site optimization framework based on an asynchronous dominance factor criticism algorithm, which reduces image blocks by 20% and data block loss by 15% in 5G virtualization environments by adjusting the bit rate in real-time [5]. In addition, DRL is also used in virtual environments to address the transmission challenges of high frame rate (HFR) video streams. By combining multiple access edge computing (MEC) and software defined network (SDN), DRL can dynamically adapt to network conditions to ensure low latency and high image quality. Current research is exploring the combination of DRL and computer vision technology to improve the accuracy of object recognition and interaction in virtual environments.

DOI: 10.54254/2755-2721/158/2025.TJ23483

3.2. Robot control and automation

In the field of robot control and automation, DRL is gradually becoming a key technology for solving complex decision-making problems. Researchers have made significant progress in applying DRL to various scenarios such as robot path planning, robot arm control, and autonomous driving. In terms of robot path planning, Liu et al. proposed a method based on multi-agent reinforcement learning, which can effectively handle obstacle avoidance problems in dynamic environments. This method significantly improves the success rate of long-distance navigation by decomposing the global path into local sub targets. The process structure is shown in Figure 2. The experimental results show that the navigation success rate of this method can reach over 99% in testing environments containing dynamic obstacles [6]. Robot arm control is another important application direction. The DRL algorithm, through end-to-end training, enables the robot arm to autonomously perform fine operations such as grasping and assembly, demonstrating stronger adaptability and robustness. Especially when dealing with unstructured environments, DRL methods can adjust control strategies in real-time to cope with sudden changes. The current research hotspots include how to improve the sampling efficiency of the DRL algorithm and enhance the system's generalization ability in real-world environments.

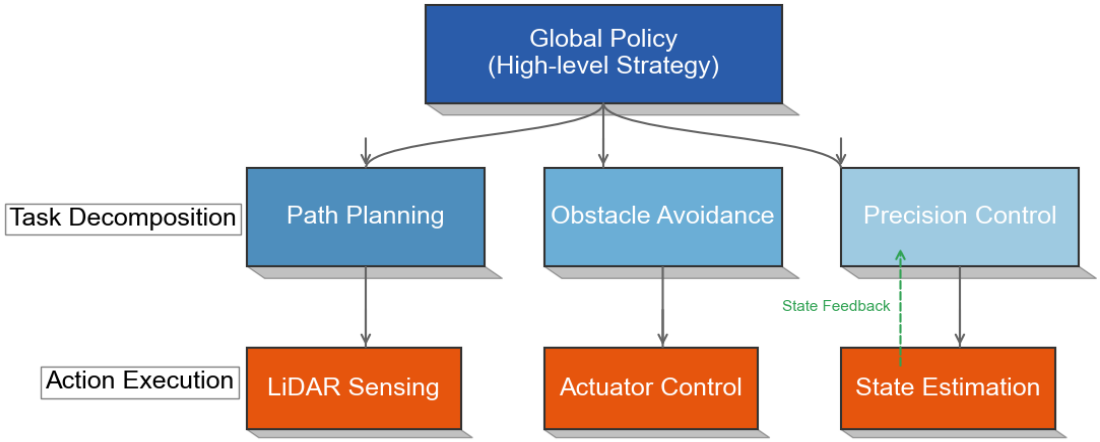

Figure 2: Hierarchical reinforcement learning framework diagram [6]

3.3. Resource management and optimization

In the field of energy and resource management, DRL provides innovative solutions for intelligent decision-making in dynamic pricing environments. Pokorn et al. proposed a household energy trading system based on the DQN algorithm, which autonomously optimizes the coordinated scheduling of photovoltaic power generation and battery energy storage by analyzing real-time electricity price (RTP) and time of use (TOU) modes. Research has shown that compared to traditional rule-based strategies, this approach can reduce energy costs by 47.66%. The core lies in designing a three-stage composite reward function, including battery state penalties, market price related rewards, and direct cost feedback [7]. In the scenario of grid level resource allocation, DRL achieves dynamic optimization by handling high-dimensional state spaces such as load demand, power generation output, and network topology. Typical applications include: microgrid power dispatch based on multi-agent reinforcement learning, balancing supply and demand relationships through distributed decision-making; Data center energy efficiency management, utilizing DRL to adjust computing load and cooling system power consumption in real-time. The current challenges mainly focus on improving sample efficiency and long-term strategy stability.

DOI: 10.54254/2755-2721/158/2025.TJ23483

3.4. Medical and health

DRL has demonstrated unique application value in the healthcare field. In terms of medical image analysis, DRL achieves efficient processing of high-resolution medical images through a serialized decision mechanism. Research has shown that compared to traditional supervised learning methods, DRL can reduce computational memory usage by 83% in lesion localization tasks. In personalized treatment, DRL systems based on strategic gradient algorithms such as PPO and DDPG can adjust treatment plans according to real-time physiological parameters of patients. For example, in the adjustment of diabetes insulin dose, the DRL model fine tuned the recommended dose through continuous action space output, and clinical trials showed that its control effect was better than that of the traditional PID controller (the time to reach the blood glucose standard increased by 22%) [8]. In addition, breakthroughs have been made in the application of DRL in surgical robot control. By combining imitation learning and reinforcement learning, robots can autonomously complete fine operations such as suturing and knotting, with an average operating error of less than 0.3mm [8].

4. The challenges and development directions of DRL in complex environments

4.1. High dimensional state space and action space

One of the core challenges of DRL is the problem of high-dimensional state space and action space. As the complexity of the environment increases, the number of combinations of states and actions grows exponentially, a phenomenon known as "state space explosion". For example, in autonomous driving tasks, intelligent agents need to process high-dimensional input data from cameras, LiDAR, and other sensors, while facing continuous control decisions, thus forming a very large state action space. To address this challenge, deep learning models such as Convolutional Neural Networks (CNNs) are widely used to extract low dimensional feature representations from high-dimensional raw inputs such as images and videos to reduce computational complexity. On the other hand, applying attention mechanisms in reinforcement learning can help agents focus on key state features. In terms of action space, continuous action problems are usually optimized using policy gradient based methods. Hierarchical RL is widely used in discrete action spaces to reduce the computational difficulty of each decision step by decomposing complex tasks into subtasks [9].

4.2. Sparse rewards and exploration problems

Effective exploration in sparse reward environments is a key challenge in deep reinforcement learning. Taking the TriFinger robot cube manipulation task as an example, the sparse reward mechanism for target trajectory tracking makes it difficult for agents to obtain effective feedback through random exploration. The team proposed a hybrid reward design scheme that combines sparse xy plane position error rewards with dense z-axis height error rewards, and accelerates initial training by explicitly optimizing the height error rewards. At the same time, by replacing the xy coordinates with only the original z-coordinate target, the post experience replay (HER) mechanism has been improved, avoiding punishment for enhancing behavior due to target modification. This method increases the success rate of tasks in simulated training from 70% to over 90%. After introducing knowledge transfer (KT) technology, reusing learned localization strategies to initialize network parameters or guide experience collection significantly reduces ineffective exploration. In the extended task of cube direction control, this method reduced the average position deviation from 0.134 meters to 0.02 meters and increased the direction deviation from 142 degrees to 76 degrees, verifying its effectiveness in sparse reward scenes. Figure 3 shows the success rate of sparse reward optimization [10].

DOI: 10.54254/2755-2721/158/2025.TJ23483

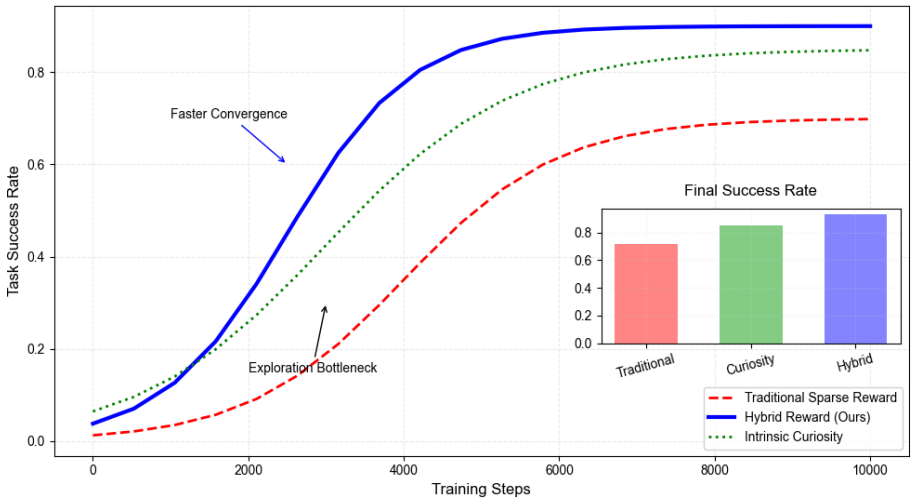

Figure 3: Comparison of sparse reward optimization techniques [10]

4.3. Partial observability and uncertainty

In complex environments, deep reinforcement learning also faces issues of observability and uncertainty. Partial observability stems from the fact that intelligent agents can only obtain local observations of the environmental state and cannot directly perceive the global state, which is particularly prominent in scenarios such as robot navigation and medical decision-making. For example, in human-machine collaboration systems, robots need to infer the intentions of human operators through limited sensor data, which increases the difficulty of decision-making due to the lack of information. Researchers propose using the POMDP framework to model partial observability, inferring hidden states through historical observation sequences, and combining empirical replay techniques to improve the accuracy of state estimation [11]. However, when the environment dynamics increase or observation noise increases, state estimation errors may accumulate, leading to strategy failure. Uncertainty includes two aspects: environmental randomness and model cognitive bias. The randomness of the environment, such as changes in traffic flow and fluctuations in patient physiological indicators, requires algorithms to have robustness; The uncertainty of the model arises from the deviation between the training data and the true distribution, which may lead to overfitting of the strategy. To address this issue, a Bayesian reinforcement learning framework is proposed, which treats policy parameters as random variables and quantifies uncertainty through variational inference [11].

4.4. Multi agent reinforcement learning

Deep reinforcement learning faces the challenge of data privacy protection in multi-agent systems, and federated learning provides new ideas for solving this problem through distributed training and parameter aggregation (as shown in Figure 4). Research has shown that combining federated learning with deep reinforcement learning algorithms such as DQN, DDPG, and PPO can improve model performance while protecting data privacy. For example, the accuracy of federated DQN in Atari games increased by 39.9%, while the performance of federated DDPG in continuous control tasks improved by 80%. This fusion method not only solves the privacy leakage risk in multi-agent collaboration, but also improves learning efficiency through distributed training, providing feasible technical solutions for fields with strict privacy requirements such as autonomous driving and intelligent healthcare [12]. Future research can further optimize communication protocols, improve processing methods for nonindependent and identically distributed data, explore stronger privacy

DOI: 10.54254/2755-2721/158/2025.TJ23483

protection mechanisms, and promote the practical application of federated reinforcement learning in complex multi-agent systems [12].

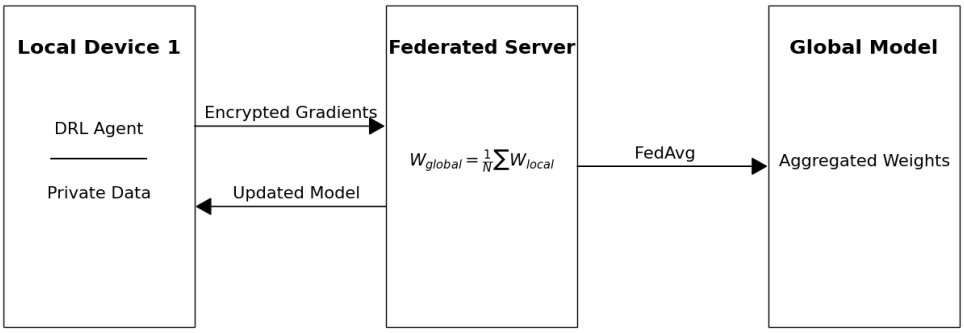

Figure 4: Privacy protection architecture for federated learning and DRL[12]

4.5. Safety and interpretability

In the application of deep reinforcement learning in financial markets, security and interpretability have become key challenges. On the one hand, although simulation environments such as PyMarketSim provide a secure testing platform for strategic training, avoiding direct risks in the real market, strategies still face potential security threats such as overfitting and market manipulation during deployment. Although PSRO and other methods can evaluate policy equilibrium, real-time risk monitoring mechanisms still need improvement. On the other hand, trading strategies driven by deep neural networks such as TRON agents have high expressive power, but their complex decision-making process leads to insufficient interpretability, making it difficult to gain the trust of regulatory agencies and investors. In the future, it is necessary to explore new technologies such as neural differential equations to improve the transparency of models, or to combine rule-based models with deep reinforcement learning to enhance the interpretability of strategies while maintaining performance [13].

4.6. Integration of reinforcement learning and neuroscience

The interdisciplinary research between deep reinforcement learning and neuroscience has injected new vitality into the development of this field. Neuroscience not only provides theoretical inspiration for deep reinforcement learning, but also verifies the effectiveness of the algorithm through experiments. Meanwhile, reinforcement learning simulates complex cognitive functions, such as model-based reinforcement learning, which simulates the ability of humans to construct cognitive maps. The combination of the two has promoted bidirectional progress: discoveries in neuroscience have facilitated algorithm innovation, such as distributed coding inspiring distributed reinforcement learning; The theoretical hypothesis was validated by reproducing neural phenomena through algorithms, such as using time difference models to predict animal learning behavior. In the future, it is necessary to further integrate neuroscience data to enhance the biological rationality of models, such as introducing motivational mechanisms in exploring the use of balance, or utilizing modular brain structures to enhance model adaptability, in order to promote the design of more brain like intelligent agents [14].

5. Conclusion

This article provides a comprehensive overview of the current applications and challenges of DRL in complex environments. DRL has demonstrated remarkable success across diverse fields such as

DOI: 10.54254/2755-2721/158/2025.TJ23483

gaming, robot control, resource management, and healthcare. By leveraging core algorithms like DQN, policy gradient methods, and actor-critic frameworks, DRL effectively addresses high-dimensional state spaces, sparse rewards, and partial observability. Extension techniques such as hierarchical reinforcement learning, meta-reinforcement learning, and hybrid reward designs further enhance the adaptability and robustness of DRL systems. These advancements have enabled breakthroughs in tasks ranging from autonomous driving to personalized medical treatments, showcasing the transformative potential of DRL in real-world scenarios. However, significant challenges remain. The issue of data efficiency persists, particularly in environments with dynamic changes or limited feedback. Multi-agent systems introduce complexities related to coordination, privacy, and scalability, necessitating innovative solutions like federated learning. Safety and interpretability are critical concerns, especially in high-stakes domains such as healthcare and finance, where transparent and reliable decision-making is paramount. Additionally, the integration of interdisciplinary approaches, such as neuroscience-inspired mechanisms, offers promising avenues for improving the biological plausibility and adaptability of DRL models. For instance, insights from neural coding and cognitive mapping could refine exploration strategies and enhance model generalization.

References

[1] *Wang, X. et al., (2024) Deep Reinforcement Learning: A Survey. IEEE Transactions on Neural Networks and Learning Systems, 35, 5064-5078.*

[2] *Li, Z., Shi, N., Zhao, L., and Zhang, M. (2024) Deep reinforcement learning path planning and task allocation for multi-robot collaboration. Alexandria Engineering Journal, 109, 408-423.*

[3] *Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Al Sallab, A. A., Yogamani, S., and Pérez, P. (2021) Deep reinforcement learning for autonomous driving: A survey. IEEE, 23, 4909-4926.*

[4] *Kang, Y., Park, H., Smit, B., and Kim, J. (2023) A multi-modal pre-training transformer for universal transfer learning in metal–organic frameworks. Nature Machine Intelligence, 5, 309-318.*

[5] *Rio, A., Serrano, J., Jimenez, D., Contreras, L. M., and Alvarez, F. (2024) Multisite gaming streaming optimization over virtualized 5G environment using Deep Reinforcement Learning techniques. Computer Networks, 244, 110334.*

[6] *Liu, Z., Chen, B., Zhou, H., Koushik, G., Hebert, M., and Zhao, D. (2020) Mapper: Multi-agent path planning with evolutionary reinforcement learning in mixed dynamic environments. IEEE, 11748-11754.*

[7] *Pokorn, M., Mohorčič, M., Čampa, A., and Hribar, J. (2023) Smart home energy cost minimisation using energy trading with deep reinforcement learning. ACM, 361-365.*

[8] *Thakur, P., and Talwandi, N. S. (2024) Deep Reinforcement Learning in Healthcare and Bio-Medical Applications. In 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT). IEEE, 5, 742-747.*

[9] *Yan, J., Luo, B., and Xu, X. (2024) Hierarchical reinforcement learning for handling sparse rewards in multi-goal navigation. Artificial Intelligence Review, 57, 156.*

[10] *Wang, Q., Sanchez, F. R., McCarthy, R., Bulens, D. C., McGuinness, K., O'Connor, N., and Redmond, S. J. (2023) Dexterous robotic manipulation using deep reinforcement learning and knowledge transfer for complex sparse reward-based tasks. Expert Systems, 40, 13205.*

[11] *Angelotti, G., Chanel, C. P., Pinto, A. H., Lounis, C., Chauffaut, C., and Drougard, N. (2024) Offline risk-sensitive rl with partial observability to enhance performance in human-robot teaming. arXiv.*

[12] *Kim, J., Kim, G., Hong, S., and Cho, S. (2024) Advancing Multi-Agent Systems Integrating Federated Learning with Deep Reinforcement Learning: A Survey. IEEE, 55-57.*

[13] *Mascioli, C., Gu, A., Wang, Y., Chakraborty, M., and Wellman, M. (2024, November). A Financial Market Simulation Environment for Trading Agents Using Deep Reinforcement Learning. ACM, 5, 117-125.*

[14] *Subramanian, A., Chitlangia, S., and Baths, V. (2022) Reinforcement learning and its connections with neuroscience and psychology. Neural Networks, 145, 271-287.*