



Multi-Agent Reinforcement Learning for Autonomous Decision-Making Systems: A Survey

Dr. Pradeep Laxkar

Associate Professor

Department of Computer Science and Engineering

ITM(SLS) University, Vadodara, Gujarat

pradeep.laxkar@gmail.com

Abstract—The autonomous systems are emerging in complex and dynamic contexts in which several intelligent agents have to coordinate, adapt and make decisions in the face of uncertainty. Single-agent reinforcement learning typically performs poorly in such environments due to issues with agent interactions, decentralized information, and non-stationary learning conditions. One promising framework that could enable multiple agents to learn cooperative, competitive, or mixed strategies, while continuously interacting with their environment and with each other, is Multi-Agent Reinforcement Learning (MARL). In this paper, a wide survey of MARL for autonomous decision-making systems is presented. It examines the core principles and models of MARL, such as cooperative and competitive learning paradigms, centralized training and decentralized execution (CTDE), and key algorithmic categories like value-based, policy-based, and actor-critic. In addition, the survey analyzes the use of MARL in autonomous decision-making architectures. It identifies its use in autonomous vehicles, unmanned aerial vehicle (UAV) swarms, multi-robot systems, smart grids, and industrial automation. Some of the main difficulties, such as scalability, coordination, credit assignment, safety, reliability, communication limitations, and partial observability, are also examined. Finally, the paper discusses current research trends and future directions to improve the scalability, robustness, explain ability, and real-world deployment of MARL-enabled autonomous systems. The survey provides an organized view of today's progress and challenges in this ever-changing field.

Keywords—Multi-Agent Reinforcement Learning (MARL), Autonomous Decision-Making Systems, Reinforcement Learning, Multi-Agent Systems, Autonomous Vehicles, UAV Swarms, Multi-Robot Systems.

I. INTRODUCTION

Autonomous decision-making systems are increasingly deployed in complex, dynamic environments such as autonomous vehicles, unmanned aerial vehicle (UAV) swarms, multi-robot systems, smart grids, and industrial automation. These systems require intelligent capabilities to perceive their surroundings, make decisions, and adapt to changing conditions with minimal human intervention. Traditional control approaches often struggle to address the uncertainty, scalability, and coordination requirements of such environments [1].

Reinforcement Learning (RL) has emerged as a powerful framework for learning optimal behaviours and is widely used for sequential decision-making tasks [2]. However, many real-world applications involve multiple interacting agents, making single-agent RL insufficient for handling complex collaborative and competitive scenarios [3]. Multi-Agent

Reinforcement Learning (MARL) extends RL to multi-agent environments, enabling agents to learn cooperative, competitive, or mixed strategies while considering the actions of other agents during learning. This capability is particularly valuable for autonomous decision-making systems operating in decentralized and partially observable environments [4].

Recent advances in Large Language Models (LLMs) have further expanded the capabilities of autonomous systems by enabling reasoning, communication, planning, and coordination among multiple intelligent agents. LLM-based multi-agent systems facilitate task decomposition collaborative problem-solving, and adaptive decision-making in complex environments. In parallel, Self-Healing AI has emerged as a promising paradigm that enables autonomous systems to detect faults, recover from failures, adapt to unexpected situations, and maintain operational reliability without human intervention. The integration of MARL, LLM-based agents, and self-healing mechanisms has opened new opportunities for developing robust, scalable, and resilient autonomous decision-making systems [5].

Despite significant advances in MARL algorithms and architectures, several challenges remain, including scalability, coordination, communication constraints, credit assignment, safety, reliability, and interpretability [6]. Furthermore, the growing diversity of MARL applications [7], together with the emergence of LLM-driven autonomous agents and self-healing systems, has led to fragmented research across multiple domains. Therefore, a comprehensive survey is needed to examine MARL concepts, algorithms, architectures, applications, and emerging research directions [8]. This paper presents a systematic review of MARL for autonomous decision-making systems and discusses future opportunities for building scalable, intelligent, resilient, and trustworthy autonomous systems.

A. Structure of the Paper

This paper is organized as follows: Section II presents the concepts and frameworks of Multi-Agent Reinforcement Learning (MARL). Section III discusses MARL algorithms and architectures. Section IV describes Autonomous Decision-Making Systems and their decision-making architectures. Section V reviews the existing literature related to MARL and autonomous systems. Section VI concludes the paper and outlines future research directions.

II. MULTI-AGENT REINFORCEMENT LEARNING: CONCEPTS AND FRAMEWORKS

Multi-Agent Reinforcement Learning (MARL) is a branch of Reinforcement Learning where multiple agents interact and

learn within a shared environment. Agents can cooperate, compete, or pursue mixed objectives while adapting their actions based on feedback from the environment. MARL is widely used for solving complex decision-making problems that require coordination, communication, and decentralized control in dynamic and uncertain environments [9].

A. Reinforcement Learning in Multi-Agent Environment:

Reinforcement Learning (RL) is a learning paradigm in which an agent interacts with an environment and learns to maximize cumulative rewards through trial and error [10]. In Multi-Agent Reinforcement Learning (MARL), multiple agents operate simultaneously within a shared environment, making the learning process more complex due to interactions among agents [11]. MARL supports cooperative, competitive, and mixed strategies, enabling agents to adapt their policies in dynamic and uncertain environments [12]. This approach is widely applied in domains such as robotics, autonomous vehicles, communication networks, and smart systems that require coordinated decision-making.

B. Concepts of Multi-Agent Reinforcement Learning

Multi-Agent Reinforcement Learning (MARL) is a branch of Reinforcement Learning (RL) in which multiple agents interact within a shared environment. Unlike single-agent RL, MARL enables agents to learn simultaneously while considering the actions and behaviors of other agents [13]. Depending on the application, agents may cooperate, compete, or pursue mixed objectives.

The fundamental concepts of MARL include:

- **Agent-Environment Interaction:** Each agent observes the environment, selects actions, receives rewards, and updates its policy to maximize cumulative rewards.
- **Cooperation:** Multiple agents work together to achieve a common objective and maximize a shared reward[14].
- **Competition:** Agents have conflicting goals and attempt to optimize their own rewards against others[15].
- **Mixed-Interest Scenarios:** Agents exhibit both cooperative and competitive behaviours simultaneously.
- **Coordination and Communication:** Agents exchange information or coordinate actions to improve collective performance.
- **Non-Stationarity:** The environment changes dynamically as multiple agents learn and adapt concurrently.
- **Decentralized Decision-Making:** Agents make independent decisions based on local observations without requiring centralized control.

C. Frameworks of Multi-Agent Reinforcement Learning

Several theoretical frameworks provide the foundation for modelling and solving MARL problems in complex environments:

- **Stochastic Games (SGs):** Extend Markov Decision Processes (MDPs) to multi-agent settings where state transitions and rewards depend on the actions of multiple agents.
- **Partially Observable Markov Decision Processes (POMDPs):** Model environments where agents have

limited or incomplete information about the system state.

- **Decentralized Partially Observable Markov Decision Processes (Dec-POMDPs):** Extend POMDPs to multiple agents operating under uncertainty while making decentralized decisions.
- **Centralized Training and Decentralized Execution (CTDE):** A widely adopted framework where agents are trained using global information but execute policies independently based on local observations.
- **Value Decomposition Networks (VDN):** A cooperative MARL framework that decomposes a global value function into individual agent value functions.
- **Multi-Agent Proximal Policy Optimization (MAPPO):** A policy-based framework that extends PPO to multi-agent environments for stable and efficient learning.
- **Multi-Agent Deep Deterministic Policy Gradient (MADDPG):** An actor-critic framework designed for mixed cooperative-competitive environments with continuous action spaces.

III. MARL ALGORITHMS AND ARCHITECTURES

Multi-Agent Reinforcement Learning (MARL) algorithms and architectures provide the foundation for enabling multiple agents to learn, coordinate, and make decisions in shared environments. MARL algorithms define how agents learn optimal policies, while architectures determine how information is shared and utilized during training and execution. Together, they address challenges such as coordination, scalability, non-stationarity, and partial observability in multi-agent systems. MARL algorithms can be broadly categorized into three groups:

A. Value-Based Algorithms

Value-based methods learn action-value functions that estimate the expected cumulative reward of actions. These approaches are effective for cooperative tasks and decentralized decision-making [16].

- Independent Q-Learning (IQL)
- Value Decomposition Networks (VDN)
- QMIX

B. Policy-Based Algorithms

Policy-based methods directly optimize agent policies without explicitly estimating value functions. They are particularly suitable for continuous and high-dimensional action spaces.

- Policy Gradient Methods
- Multi-Agent Proximal Policy Optimization (MAPPO)

C. Actor-Critic Algorithms

Actor-critic methods combine policy optimization (actor) with value estimation (critic), improving learning stability and coordination among agents.

- Multi-Agent Deep Deterministic Policy Gradient (MADDPG)
- MAPPO
- COMA (Counterfactual Multi-Agent Policy Gradients).

The architecture shown in Figure 1 illustrates the interaction among multiple agents operating within a shared environment. Each agent consists of two primary components: a Perception Module and a Policy Module[17]. The perception module processes observations obtained from the environment and extracts relevant state information. Based on these observations, the policy module selects actions that maximize the expected cumulative reward[18].

The actions generated by all agents are combined through a Joint Action Module, which produces a collective action affecting the environment. The environment then transitions to a new state and generates rewards based on the outcome of the joint action. Since the overall reward often depends on the contributions of multiple agents, a Credit Assignment Module is employed to distribute rewards fairly among agents. This module determines the contribution of each agent to the team performance and provides individual reward signals (r_1, r_2, \dots, r_n) for policy updates.

The information exchange between perception and policy modules enables coordination among agents, while the credit assignment mechanism improves cooperative learning by addressing the multi-agent credit assignment problem. Such architectures are widely used in cooperative MARL applications, including multi-robot systems, autonomous vehicles, smart grids, resource allocation, and distributed control systems

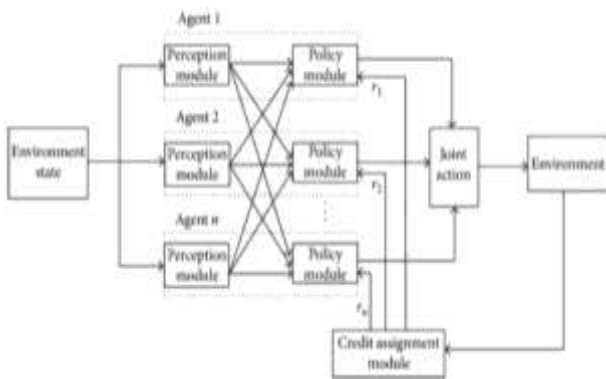


Fig. 1. MARL Architecture

Marl architectures define how agents are trained and how information is exchanged during learning[19].

D. Centralized Training and Decentralized Execution (CTDE)

The most widely adopted MARL architecture, where agents utilize global information during training but make decisions independently using local observations during execution.

E. Communication-Based Architectures

These architectures enable agents to exchange information and coordinate actions through learned communication protocols.

F. Graph Neural Network (GNN)-Based Architectures

GNNs model relationships among agents using graph structures, enabling efficient coordination in large-scale multi-agent systems.

G. Attention-Based Architectures

Attention mechanisms allow agents to focus on the most relevant information from neighboring agents or the environment, improving decision-making efficiency.

IV. AUTONOMOUS DECISION-MAKING SYSTEMS

Autonomous Decision-Making Systems (ADMS) are intelligent systems capable of perceiving their environment, processing information, and making decisions with minimal or no human intervention. These systems are designed to operate in dynamic and uncertain environments by continuously analyzing environmental conditions and selecting appropriate actions to achieve predefined objectives. Advances in Artificial Intelligence (AI), Machine Learning (ML), and Reinforcement Learning (RL) have significantly enhanced the autonomy, adaptability, and intelligence of such systems. Autonomous decision-making has become a fundamental capability in applications such as autonomous vehicles, unmanned aerial vehicles (UAVs), robotics, cybersecurity, smart transportation, and industrial automation, where real-time and reliable decision-making is essential[20].

A. Characteristics of Autonomous Systems:

Autonomous systems possess several key characteristics that enable intelligent and adaptive behaviours:

- **Autonomous Decision-Making:** Autonomous systems can independently observe the surrounding environment, analyze information and make decisions without constant human participation.
- **Multi-Agent Coordination:** Modern autonomous systems frequently involve multiple interacting agents that need to coordinate, compete or cooperate to efficiently reach some common objective.
- **Adaptability and Learning:** These systems have the ability to change their behavior in response to changes in the environment or experiences, which allows them to operate effectively in dynamic and uncertain environments.
- **Decentralized Operation:** Autonomous systems often operate in a decentralized way, with each agent making decisions based on immediate observations, yet playing a role in achieving overall system objectives.
- **Safety and Reliability:** Autonomous systems need to be safe and reliable, with certain applications like autonomous vehicles, UAV swarms, robotics and industrial automation involving safety-critical situations where failures can have serious consequences.

B. Decision-Making Process in Autonomous Systems

The decision-making process in autonomous systems typically consists of four interconnected stages:

- **Environment Perception:** Collection of information from sensors such as cameras, radar, LiDAR, and communication devices to understand the surrounding environment.
- **Situation Analysis and Decision Planning:** Processing environmental information to evaluate possible actions and select the most suitable strategy.
- **Action Execution:** Implementation of the selected action through control mechanisms or actuators.

- **Feedback and Learning:** Evaluation of action outcomes and continuous refinement of decision policies using learning algorithms.

C. Decision-Making Architectures:

Decision-making architectures define how information is processed and how decisions are generated within autonomous systems. These architectures influence system scalability, coordination efficiency, and adaptability.

1) Centralized Architecture

In centralized architectures, a central controller collects global information and makes decisions for all agents. This approach enables effective coordination but may suffer from scalability limitations and single points of failure.

2) Decentralized Architecture

Decentralized architectures allow agents to make decisions independently using local observations. These architectures improve scalability, robustness, and flexibility but may introduce coordination challenges among agents.

3) Hybrid Architecture

Hybrid architectures combine centralized and decentralized approaches to balance coordination and scalability. A widely adopted example is the Centralized Training and Decentralized Execution (CTDE) paradigm, in which agents use global information during training but make independent decisions during execution.

4) Learning-Based Architectures

Recent autonomous systems incorporate advanced learning mechanisms such as Actor–Critic networks, attention mechanisms, communication protocols, and Graph Neural Networks (GNNs). These architectures improve information sharing, coordination, and decision-making performance in complex multi-agent environments.

D. Challenges and Open Issues in Multi-Agent Reinforcement Learning (MARL)

MARL has been found to show great promise for autonomous decision-making systems. But there are some obstacles to using it in real-world applications. The challenges are the dynamics of multiple agents, dynamic environments, and efficient coordination and learning.

1) Scalability

One of the hard challenges in MARL is the scalability of the state-action space as the number of agents increases due to their exponential growth. More agents are added, more and more computation is required; convergence is slower [21]. The algorithms operating millions or even thousands of autonomous systems in a large-scale environment, like smart transportation systems, drone swarms, and robotic teams, must be efficient and be able to process thousands of interactions. New strategies include parameter sharing, hierarchical learning, and graph-based representations for increased scalability.

2) Coordination

Cooperative environments require effective coordination among the agents to achieve common objectives. Agents need to be able to learn behaviors that work together and be flexible in a changing environment. If there is poor coordination, conflicts, repeated efforts and suboptimal system performance may occur [22]. Robust coordination mechanism remains

difficult to develop, especially in low communication environments and when decentralized decision-making exists.

3) Credit Assignment

The problem of “credit assignment” is one of assigning credit to individual agents for the collective result. In cooperative MARL, agents are frequently rewarded together, and it is hard to determine which actions are responsible for success or failure [23]. The number of agents increases the difficulty of this challenge. Some strategies have been suggested to solve this problem, including value decomposition and counterfactual reasoning.

4) Safety and Reliability

In safety-critical applications like autonomous vehicles, healthcare and industrial automation, autonomous decision-making systems are frequently used[24] [25]. MARL agents are expected to make sound decisions and refrain from unsafe actions. Robustness to uncertainty, adversarial behaviors and unexpected environmental changes is still a big challenge to research. Safe MARL aims to build safety constraints into the learning process to enhance trustworthiness and reliability.

5) Communication and Partial Observability

The agents are only partially informed about the global state in many real-world situations. Limited observability can make it difficult to make decisions and coordinate effectively. Communication mechanisms can be used to share relevant information among agents[26], but excessive communication can increase computational and bandwidth demands. Designing efficient communication protocols to share information and resources remains an open problem in research.

V. LITERATURE REVIEW

The literature review explores MARL, autonomous decision-making systems, deep reinforcement learning, LLM-based multi-agent systems, and explainable AI, focusing on their applications, challenges, and future research directions.

J. Liang et al. (2025) systematically review and analyze their basic principles, applications in multi-agent systems, challenges faced, and corresponding solutions. Specifically, it discusses the challenges faced by MARL algorithms from four aspects: dimensionality, non-stationarity, partial observability, and scalability. Additionally, it surveys existing algorithm-training environments in the field of multi-agent systems and summarizes the applications of MARL algorithms across different domains [27].

J. Chai, Z. Zhao, Y. Zhu, and D. Zhao (2025) provide a comprehensive review of the latest advances in this field. By combining multi-task reinforcement learning with cooperative MARL, categorize and analyze the 3 major types of multi-task problems under multi-agent context, offering more fine-grained classifications and summarizing key insights for each. In addition, summarize commonly used benchmarks and discuss future directions of research in this area, which hold promise for further enhancing the multi-task cooperation capabilities of multi-agent systems and expanding their practical applications in the real world [28].

S. Govinda, B. Brik, and S. Harous (2025) This survey paper provides a comprehensive overview of the role of DRL in autonomous systems, focusing on recent advancements, applications, and challenges. Through a synthesis of existing literature and case studies, the paper elucidates key principles,

methodologies, and implications of integrating DRL into autonomous systems. The systematic examination of selected papers reveals recurring patterns, emerging trends, and identifies gaps and opportunities for further research [29].

S. Chen et al. (2024). This study is to comprehensively review the latest issues and research progress in decision-making and planning for autonomous vehicles in intersection environments. This paper reviews the research progress in the behavioural prediction of traffic participants in terms of machine learning-based behavioural prediction, probabilistic model behavioural prediction, and mixed-method behavioural prediction. Then, behavioural decision-making is divided into reactive decision-making, learning decision-making, and interactive decision-making, each of which is analyzed [30].

X. Li et al. (2024) present a comprehensive survey of these studies, offering a systematic review of LLM-based MAS. Adhering to the workflow of LLM-based multi-agent systems, synthesize a general structure encompassing five key components: profile, perception, self-action, mutual interaction, and evolution. This unified framework encapsulates much of the previous work in the field.

Furthermore, illuminate the extensive applications of LLM-based MAS in two principal areas: problem-solving and world simulation [31].

A. Willie (2024). This paper explores the critical need for explain ability and transparency in autonomous decision-making systems, focusing on the methods developed to make these systems more interpretable. Categorize the techniques into post-hoc explain ability methods, transparent model design, and hybrid approaches, analyzing their strengths, limitations, and applicability across different domains. Additionally, examine the ethical, legal, and social implications of explain ability and transparency, and propose future directions for enhancing these aspects in AI systems [32].

Table I summarizes recent studies on Multi-Agent Reinforcement Learning (MARL) and autonomous decision-making systems, highlighting key research areas, major findings, challenges, contributions, and limitations while identifying current trends and future research opportunities.

TABLE I. A SUMMARY OF THE STUDY ON MULTI-AGENT REINFORCEMENT LEARNING FOR AUTONOMOUS DECISION-MAKING SYSTEMS

Authors	Focus Area	Key Findings	Challenges	Key Contribution	Limitation
J. Liang et al. (2025)	Multi-Agent Reinforcement Learning (MARL) Survey	Reviewed MARL principles, applications, and training environments.	Dimensionality, non-stationarity, partial observability, scalability.	Comprehensive overview of MARL algorithms and application domains.	Limited discussion on real-world deployment and safety issues.
J. Chai, Z. Zhao, Y. Zhu, and D. Zhao (2025)	Cooperative Multi-Task MARL	Categorized multi-task learning problems in cooperative MARL and summarized benchmarks.	Task diversity, transferability and coordination among agents.	Fine-grained classification of multi-task MARL scenarios.	Focused mainly on cooperative environments, limited coverage of competitive environments.
S. Govinda, B. Brik, and S. Harous, (2025)	Deep Reinforcement Learning (DRL) for Autonomous Systems	Reviewed DRL applications, methodologies, and emerging trends in autonomous systems.	Sample efficiency, safety, interpretability and real-world deployment.	Identified research gaps and future directions for DRL-based autonomy.	Focuses on DRL broadly rather than MARL-specific challenges.
S. Chen et al. (2024)	Autonomous Vehicle Decision-Making	Analyzed behavioral prediction and decision-making methods at road intersections.	Dynamic traffic environments, uncertainty and interaction complexity.	Comprehensive review of planning and decision-making techniques for autonomous vehicles.	Restricted primarily to intersection-driving scenarios.
X. Li et al. (2024)	LLM-Based Multi-Agent Systems	Proposed a unified framework including perception, interaction, and evolution mechanisms.	Scalability, coordination, reasoning consistency and communication overhead.	Systematic survey of LLM-based multi-agent architectures and applications.	Limited empirical validation in real-world autonomous systems.
A. Willie (2024)	Explainable AI in Autonomous Decision-Making	Examined explainability methods, transparency mechanisms, and ethical implications.	Interpretability, trustworthiness, legal compliance, transparency.	Classification of explainable AI approaches for autonomous systems.	Limited focus on explainability techniques specifically designed for MARL systems.

VI. CONCLUSION AND FUTURE WORK

MARL has become an attractive approach to allow autonomous decision-making in complex and dynamic environments where multiple agents interact with each other. This survey covered the basics of MARL: learning paradigms, algorithms, architectures, and applications for autonomous systems such as autonomous vehicles, UAV swarms, multi-robot systems, and industrial automation. Value-based, policy-based, and actor-critic approaches were identified as strengths and the Centralized Training and Decentralized Execution (CTDE) framework as a way of tackling coordination and scalability issues. Although a number of developments have been made, there are still a lot of challenges that prevent the actual deployment in the real world, such as scalability to large-scale heterogeneous agent

populations, communication limitations, credit assignment problem, partial observability, safety considerations, and theoretical guarantees of convergence and reliability. The future research direction could include the development of scalable MARL algorithms, communication-efficient MARL algorithms, safe and explainable decision making, robustness in uncertain and adversarial contexts, and robust sim-to-real transfer. Moreover, MARL is poised for significant advances when coupled with other cutting-edge technologies, such as explainable artificial intelligence, federated learning, graph neural networks, and large language models. This is crucial for the development of autonomous systems with the ability to function effectively in real-world environments, and to build trustworthy, efficient and deployable autonomous systems using MARL.

REFERENCES

- [1] N. Tsakalakis, S. Stalla-Bourdillon, D. Huynh, and L. Moreau, "A typology of explanations to support Explainability-by-Design," *ACM J. Responsib. Comput.*, vol. 2, no. 1, Feb. 2025, doi: 10.1145/3708504.
- [2] S. Satyanarayana, S. Kilaru, and K. Venkatrao, "An Integrated Deep Learning and Reinforcement Learning Framework for Profit Maximization in Perishable Food Supply Chains," 2025, pp. 156–171. doi: 10.2991/978-94-6463-978-0_15.
- [3] D. K. Dake, J. D. Gadze, G. S. Klogo, and H. Nunoo-Mensah, "Multi-Agent Reinforcement Learning Framework in SDN-IoT for Transient Load Detection and Prevention," *Technologies*, vol. 9, no. 3, 2021, doi: 10.3390/technologies9030044.
- [4] I. Narayanasamy and V. Rajamanickam, "A Cascaded Multi-Agent Reinforcement Learning-Based Resource Allocation for Cellular-V2X Vehicular Platooning Networks," *Sensors*, vol. 24, no. 17, 2024, doi: 10.3390/s24175658.
- [5] T. P. Patel and P. Agarwal, "DeepServe: Hierarchical Model Placement and Dynamic Batching for Cost-Efficient Multi-Tenant LLM Inference at Scale," in *SoutheastCon 2026*, Huntsville, AL, USA: IEEE, 2026, pp. 1–6, April. doi: 10.1109/SoutheastCon63549.2026.11476566.
- [6] S. Chen, X. Chen, Q. Bao, H. Zhang, and C. U. I. Wong, "Adaptive Multi-Agent Reinforcement Learning with Graph Neural Networks for Dynamic Optimization in Sports Buildings," *Buildings*, vol. 15, no. 14, 2025, doi: 10.3390/buildings15142554.
- [7] B. Mohan, V. R. Surasani, and R. Kumar, "Autonomous Data Stewardship: Multi-Agent AI for Real-Time Master Data Management in Financial Services," in *2026 IEEE 16th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, Jan. 2026, pp. 0689–0698. doi: 10.1109/CCWC67433.2026.11393832.
- [8] A. M. Mequanenit, E. A. Nibret, P. Herrero-Martín, M. S. García-González, and R. Martínez-Béjar, "A Multi-Agent Deep Reinforcement Learning System for Governmental Interoperability," *Appl. Sci.*, vol. 15, no. 6, 2025, doi: 10.3390/app15063146.
- [9] C. R. Landolt, W. Christoph, R. Meier, and J. Jang-Jaccard, "Multi-Agent Reinforcement Learning in Cybersecurity: From Fundamentals to Applications," 2025. doi: 10.48550/arXiv.2505.19837.
- [10] D. Huh and P. Mohapatra, "Multi-agent Reinforcement Learning: A Comprehensive Survey," 2024. doi: 10.48550/arXiv.2312.10256.
- [11] S. Wang, Q. Yue, Z. Xu, P. Qiao, Z. Lyu, and F. Gao, "A Collaborative Multi-Agent Reinforcement Learning Approach for Non-Stationary Environments with Unknown Change Points," *Mathematics*, vol. 13, no. 11, 2025, doi: 10.3390/math13111738.
- [12] L. Jia and Y. Pei, "Recent Advances in Multi-Agent Reinforcement Learning for Intelligent Automation and Control of Water Environment Systems," *Machines*, vol. 13, no. 6, 2025, doi: 10.3390/machines13060503.
- [13] T. P. Patel, A. K. Elengovan, V. Ranganathan, M. Parikh, and D. Kole, "Self-Healing AI Systems Using Multi-Agent Learning," in *2026 International Seminar on Intelligent Business and Edge-Computing Research (ISIBER)*, 2026, pp. 7–12. doi: 10.1109/ISIBER68248.2026.11470173.
- [14] P. Sunehag, G. Lever, N. Sonnerat, and M. Jaderberg, "Value-Decomposition Networks For Cooperative Multi-Agent Learning," 2017. doi: 10.48550/arXiv.1706.05296.
- [15] J.-H. Wang, H. He, J. Cha, I. Jeong, and C.-J. Ahn, "Multi-Agent Reinforcement Learning for Efficient Resource Allocation in Internet of Vehicles," *Electronics*, vol. 14, no. 1, 2025, doi: 10.3390/electronics14010192.
- [16] K. Zhang, "Multi-Agent Reinforcement Learning : A Selective Overview of Theories and Algorithms," 2021. doi: 10.48550/arXiv.1911.10635.
- [17] L. G. Nov and C. Yu, "The Surprising Effectiveness of PPO in Cooperative," 2022. doi: 10.48550/arXiv.2103.01955.
- [18] R. Lowe, J. Harb, A. Tamar, P. Abbeel, and I. Mordatch, "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," 2020. doi: 10.48550/arXiv.1706.02275.
- [19] Q. Liu, J. Ma, and W. Xie, "Multiagent Reinforcement Learning with Regret Matching for Robot Soccer," *Math. Probl. Eng.*, vol. 2013, pp. 1–8, 2013, doi: 10.1155/2013/926267.
- [20] H. Qu and J. Zou, "Complex-Scene-Oriented Autonomous Decision-Making Method for UAVs," *Electronics*, vol. 15, no. 8, p. 1757, Apr. 2026, doi: 10.3390/electronics15081757.
- [21] S. Gronauer and K. Diepold, "Multi-agent deep reinforcement learning: a survey," *Artif. Intell. Rev.*, vol. 55, no. 2, pp. 895–943, Feb. 2022, doi: 10.1007/s10462-021-09996-w.
- [22] P. Hernandez-Leal, B. Kartal, and M. E. Taylor, "A survey and critique of multi-agent deep reinforcement learning," *Auton. Agent. Multi. Agent. Syst.*, vol. 33, no. 6, pp. 750–797, 2019, doi: 10.1007/s10458-019-09421-1.
- [23] T. Rashid, M. Samvelyan, C. Schroeder, G. Farquhar, J. Foerster, and S. Whiteson, "QMIX: Monotonic Value Function Factorization for Deep Multi-Agent Reinforcement Learning," in *Proceedings of the 35th International Conference on Machine Learning*, 2018, pp. 4295–4304.
- [24] S. Mahmud, D. P. Mishra, G. . G. Ramani, M. I. Patel, M. S. Soumik, and R. Manivannan, "Design of intelligent healthcare IT infrastructure using graph theory, network analysis, and artificial intelligence," *Int. J. Appl. Math.*, vol. 38, no. 12s, pp. 2267–2280, December, 2025.
- [25] V. Sanikal, "Software-in-the-Loop Validation of Electric Vehicle Battery Systems Using AI-Augmented Digital Twins," *Int. J. Emerg. Res. Eng. Technol.*, vol. 7, no. 1, pp. 265–270, March, 2026, doi: 10.63282/3050-922X.IJERET-V7I1P133.
- [26] S. K. Davuluri, V. Challagulla, V. Mudapaka, and U. Konka, "Telcoformrix: An AI-Augmented Framework for Declarative and Scalable Provisioning of Real-Time Communication Infrastructure (Work in Progress)," in *2025 IEEE International Conference and Expo on Real Time Communications at IIT (RTC)*, 2025, pp. 1–4. doi: 10.1109/RTC66985.2025.11211725.
- [27] J. Liang *et al.*, "A Review of Multi-Agent Reinforcement Learning Algorithms," *Electronics*, vol. 14, no. 4, 2025, doi: 10.3390/electronics14040820.
- [28] J. Chai, Z. Zhao, Y. Zhu, and D. Zhao, "A Survey of Cooperative Multi-Agent Reinforcement Learning for Multi-Task Scenarios," *Artif. Intell. Sci. Eng.*, vol. 1, no. 2, pp. 98–121, 2025, doi: 10.23919/AISE.2025.000008.
- [29] S. Govinda, B. Brik, and S. Harous, "A Survey on Deep Reinforcement Learning Applications in Autonomous Systems: Applications, Open Challenges, and Future Directions," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 7, pp. 11088–11113, 2025, doi: 10.1109/TITS.2025.3560379.
- [30] S. Chen, X. Hu, J. Zhao, R. Wang, and M. Qiao, "A Review of Decision-Making and Planning for Autonomous Vehicles in Intersection Environments," *World Electr. Veh. J.*, vol. 15, no. 3, 2024, doi: 10.3390/wevj15030099.
- [31] X. Li, S. Wang, S. Zeng, Y. Wu, and Y. Yang, "A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges," *Vicinagearth*, vol. 1, no. 1, p. 9, Oct. 2024, doi: 10.1007/s44336-024-00009-2.
- [32] A. Willie, "Explainability and Transparency in Autonomous Decision-Making Systems," 2024.