The Artificial Intelligence and Machine Learning Review
www.scipublication.com

DOI: 10.69987/AIMLR.2020.10301

Reinforcement Learning in Autonomous Systems: A Review of Algorithms and Applications

Carlos Mendez[1], Sofia Ramirez[2], Juan Carlos Rojas[3]

Faculty of Engineering, Universidad Nacional de Cajamarca, Peru[1], School of Engineering, Universidad Nacional de Trujillo, Peru[2], Department of Information Systems, Universidad Nacional de San Agustin, Peru[3]

carlos.mendez@unc.edu.pe[1], sofia.ramirez@unt.edu.pe[2], juan.rojas@unsa.edu.pe[3]

Keywords

Reinforcement Learning, Autonomous Systems, Robotics, Self-Driving Cars, Unmanned Aerial Vehicles

Abstract

Reinforcement Learning (RL) has emerged as a powerful paradigm for enabling autonomous systems to learn and adapt to complex, dynamic environments. By leveraging the principles of trial and error, RL allows agents to optimize their behavior through interactions with their surroundings, receiving feedback in the form of rewards or penalties. This article provides a comprehensive review of RL algorithms and their applications in autonomous systems, including robotics, self-driving cars, and unmanned aerial vehicles (UAVs). These systems require the ability to make real-time decisions in unpredictable environments, making RL an ideal approach due to its adaptability and learning capabilities. The article begins by discussing the fundamental principles of RL, including key concepts such as Markov Decision Processes (MDPs), value functions, and policy optimization. It then explores state-of-the-art RL algorithms, such as Q-learning, Deep Q-Networks (DQN), Proximal Policy Optimization (PPO), and Actor-Critic methods, highlighting their strengths and limitations. For instance, while DQN has shown remarkable success in high-dimensional environments, it can struggle with continuous action spaces, which are better addressed by algorithms like PPO. To provide a structured overview, the article includes two tables summarizing key RL algorithms and their applications across various domains, offering readers a clear comparison of their features and use cases. Despite its potential, RL faces several challenges, including sample inefficiency, scalability, and safety concerns in real-world applications. The article concludes by discussing future research directions, such as improving sample efficiency through meta-learning, enhancing safety via robust RL techniques, and integrating RL with other machine learning paradigms like supervised and unsupervised learning. By addressing these challenges, RL can further advance the development of intelligent, autonomous systems capable of operating in increasingly complex and dynamic environments.

1. Introduction

Reinforcement Learning (RL) is a subfield of machine learning that focuses on training agents to make sequential decisions by interacting with an environment[1]. Unlike supervised learning, where the model is trained on labeled data, RL relies on trial and error, with the agent receiving feedback in the form of rewards or penalties. This paradigm is particularly well-suited for autonomous systems, which must operate in dynamic and uncertain environments. Autonomous systems, such

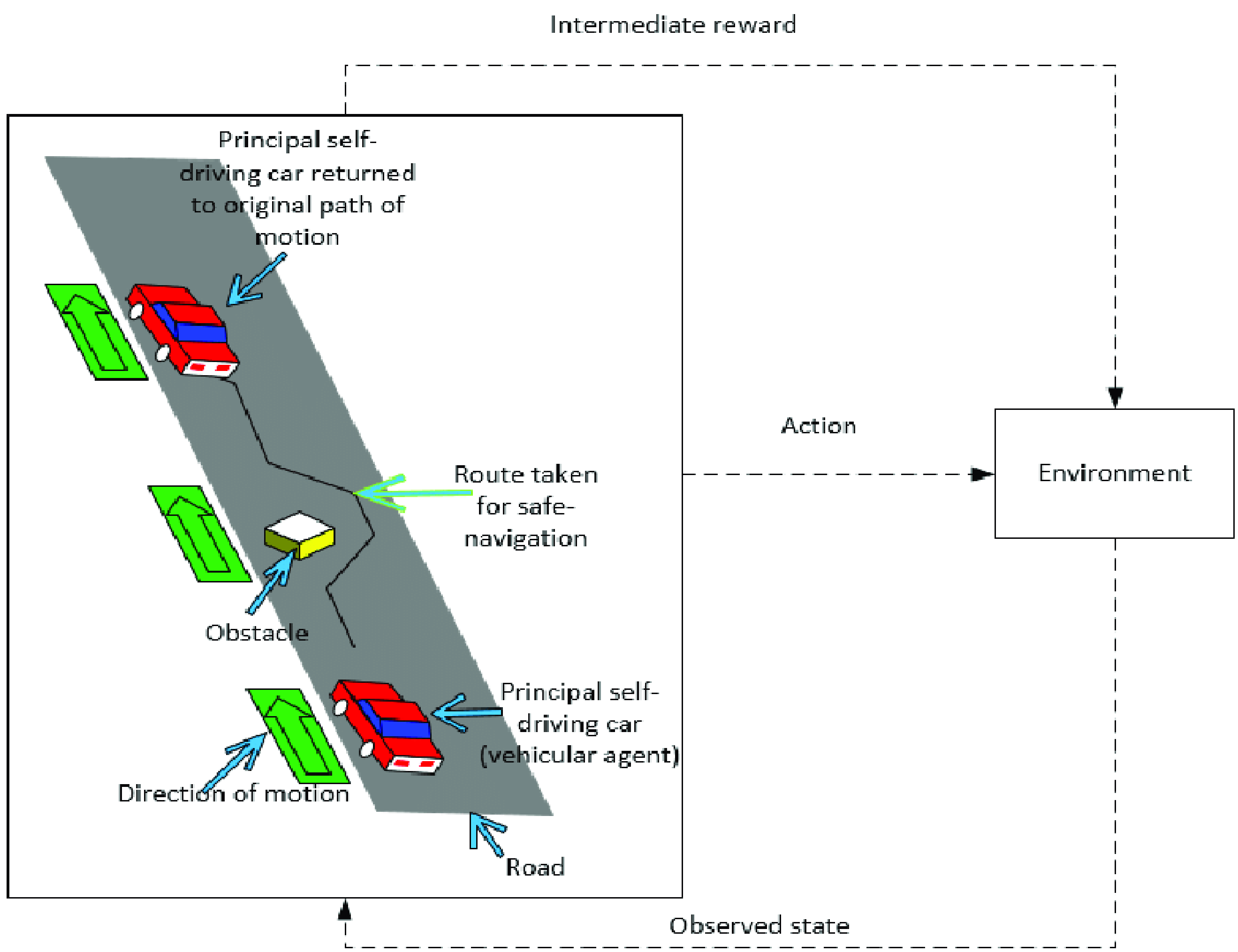
as self-driving cars, drones, and robotic manipulators, require the ability to learn from experience, adapt to new situations, and optimize their behavior over time. RL provides a framework for achieving these goals by enabling agents to learn optimal policies through exploration and exploitation[2].

The growing interest in RL for autonomous systems is driven by advancements in computational power, the availability of large-scale datasets, and the development of sophisticated algorithms. However, despite significant progress, several challenges remain, including sample inefficiency, scalability, and safety concerns. This article aims to provide a comprehensive review of RL algorithms and their applications in autonomous systems, highlighting both the successes and limitations of current approaches. By examining the state-of-the-art in RL, we hope to identify key research directions and inspire further innovation in the field[3].

2. Fundamentals of Reinforcement Learning

Reinforcement Learning is grounded in the framework of Markov Decision Processes (MDPs), which provide a mathematical model for decision-making in stochastic environments. An MDP is defined by a tuple (S,A,P,R,γ)(S,A,P,R,γ), where SS represents the set of states, AA denotes the set of actions, PP is the state transition probability function, RR is the reward function, and γγ is the discount factor. The goal of an RL agent is to learn a policy π:S→A$\pi:S\rightarrow A$ that maximizes the expected cumulative reward over time. The policy can be deterministic or stochastic, depending on the nature of the problem[4].

The learning process in RL involves two key components: exploration and exploitation. Exploration refers to the agent's ability to try new actions and discover their consequences, while exploitation involves leveraging known information to maximize rewards. Balancing these two aspects is critical for effective learning, as excessive exploration can lead to inefficiency, while excessive exploitation can result in suboptimal behavior. RL algorithms can be broadly categorized into model-based and model-free approaches. Model-based methods rely on a learned model of the environment to plan actions, whereas model-free methods directly learn a policy or value function without explicitly modeling the environment[5].



One of the most widely used RL algorithms is Q-learning, which is a model-free, off-policy method that learns the optimal action-value function Q(s,a)$Q(s,a)$. Q-learning updates the Q-values using the Bellman equation, which expresses the relationship between the value of a state-action pair and the values of subsequent states. Another popular algorithm is Deep Q-Networks (DQN), which extends Q-learning by using deep neural networks to approximate the Q-function. DQN has been successfully applied to a variety of tasks, including playing Atari games and controlling robotic systems. However, DQN suffers from limitations such as instability and sample inefficiency, which have spurred the development of improved algorithms like Double DQN, Dueling DQN, and Prioritized Experience Replay[6].

3. Reinforcement Learning Algorithms for Autonomous Systems

This section provides an in-depth review of RL algorithms that have been applied to autonomous systems. We categorize these algorithms into three groups: value-based methods, policy-based methods, and actor-critic methods. Each category is discussed in detail, with examples of their applications in autonomous systems.

3.1 Value-Based Methods

Value-based methods focus on learning a value function that estimates the expected cumulative reward for a given state or state-action pair. These methods are particularly effective for problems with discrete action spaces, where the optimal policy can be derived directly from the value function. Q-learning and its variants, such as Double Q-learning and Dueling DQN, are prominent examples of value-based methods. These algorithms have been widely used in autonomous systems, such as self-driving cars, where the action space includes discrete maneuvers like accelerating, braking, and steering[7].

One notable application of value-based methods is in the domain of robotic navigation. For instance, Q-learning has been used to train robots to navigate complex environments while avoiding obstacles. The robot learns a policy that maximizes the cumulative reward, which is defined based on factors such as reaching the goal, minimizing energy consumption, and avoiding collisions. However, value-based methods face challenges in high-dimensional state spaces, where the value function becomes difficult to approximate accurately. This limitation has motivated the development of deep reinforcement learning algorithms, which use neural networks to approximate the value function[8].

3.2 Policy-Based Methods

Policy-based methods directly optimize the policy without explicitly learning a value function. These methods are well-suited for problems with continuous action spaces, where the optimal policy cannot be easily derived from a value function. Policy Gradient (PG) algorithms, such as REINFORCE and Proximal Policy Optimization (PPO), are popular examples of policy-based methods. These algorithms use gradient ascent to optimize the policy parameters, with the goal of maximizing the expected cumulative reward[9].

Policy-based methods have been successfully applied to a variety of autonomous systems, including robotic manipulation and UAV control. For example, PPO has been used to train robotic arms to perform complex tasks such as grasping and object manipulation. The continuous action space of the robotic arm is well-suited for policy-based methods, which can directly optimize the joint angles and velocities. However, policy-based methods are often criticized for their high variance and sample inefficiency, which can make training slow and unstable. To address these issues, researchers have developed hybrid approaches that combine the strengths of value-based and policy-based methods[10].

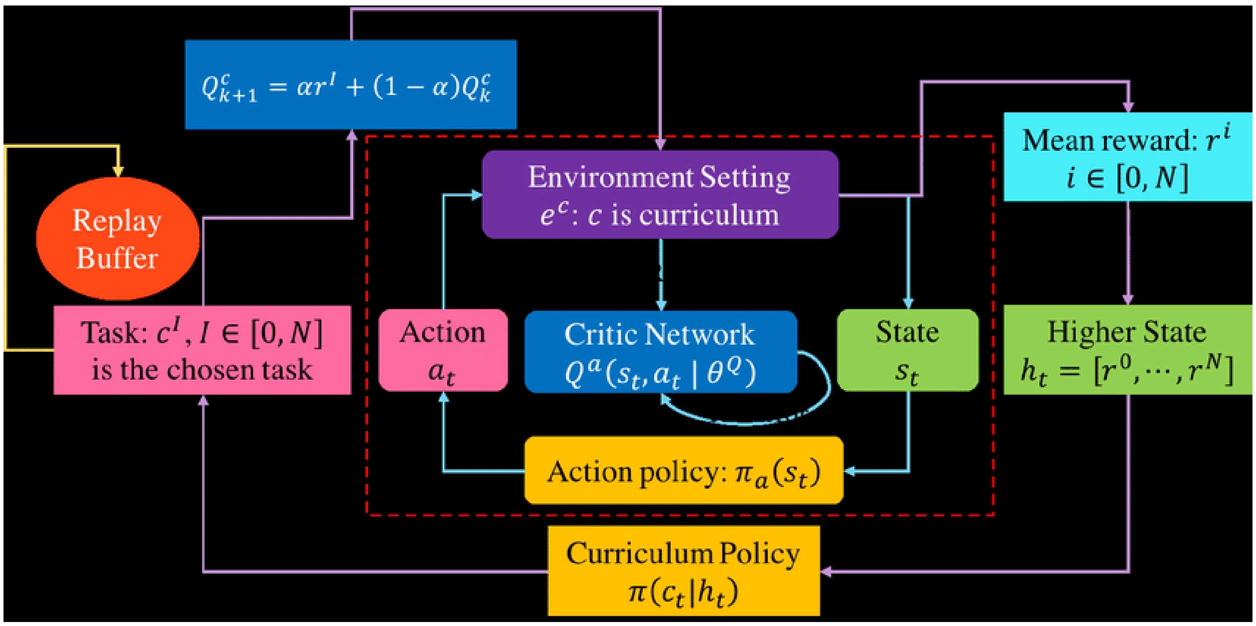
3.3 Actor-Critic Methods

Actor-critic methods combine the advantages of value-based and policy-based approaches by learning both a value function and a policy. The actor represents the policy, which is optimized using policy gradient methods, while the critic represents the value function, which is used to evaluate the actor's performance. This dual structure allows actor-critic methods to achieve more stable and efficient learning compared to pure policy-based methods. Examples of actor-critic algorithms include Advantage Actor-Critic (A2C), Asynchronous Advantage Actor-Critic (A3C), and Soft Actor-Critic (SAC).

Actor-critic methods have been widely used in autonomous systems, particularly in tasks that require continuous control and high-dimensional state spaces. For instance, SAC has been applied to self-driving cars, where the actor learns a policy for controlling the steering, acceleration, and braking, while the critic evaluates the quality of the actions. The ability of actor-critic methods to handle continuous action spaces and high-dimensional state spaces makes them a powerful tool for autonomous systems. However, these methods also face challenges, such as the need for careful tuning of hyperparameters and the risk of instability during training[11].

4. Applications of Reinforcement Learning in Autonomous Systems

This section explores the applications of RL in various autonomous systems, including robotics, self-driving cars, and UAVs. We discuss the specific challenges and opportunities associated with each application domain, as well as the RL algorithms that have been used to address these challenges[12].



4.1 Robotics

Robotics is one of the most prominent application domains for RL, as robots must operate in dynamic and uncertain environments. RL has been used to train robots for a wide range of tasks, including navigation, manipulation, and locomotion. For example, RL algorithms have been applied to teach robotic arms to perform complex manipulation tasks, such as picking and placing objects, assembling parts, and even playing table tennis. These tasks require the robot to learn precise control policies that can adapt to variations in the environment, such as changes in object position or orientation[13].

One of the key challenges in applying RL to robotics is the high dimensionality of the state and action spaces. Robots typically have multiple degrees of freedom, which makes it difficult to explore the entire state-action space efficiently. To address this challenge, researchers have developed techniques such as hierarchical RL, which decomposes the task into subtasks, and imitation learning, which leverages expert demonstrations to guide the learning process. Despite these advancements, RL in robotics remains a challenging area of research, with ongoing efforts to improve sample efficiency, scalability, and safety[14].

4.2 Self-Driving Cars

Self-driving cars represent another major application domain for RL, as they must navigate complex and dynamic environments while ensuring safety and efficiency. RL has been used to train self-driving cars for tasks such as lane keeping, obstacle avoidance, and traffic signal recognition. For example, RL algorithms have been applied to learn policies for controlling the steering, acceleration, and braking of a self-driving car based on sensor inputs such as camera images, lidar data, and radar measurements[15].

One of the key challenges in applying RL to self-driving cars is the need for safe and reliable decision-making. Unlike other domains, where mistakes may result in minor consequences, errors in self-driving cars can have severe safety implications. To address this challenge, researchers have developed techniques such as safe RL, which incorporates safety constraints into the learning process, and multi-agent RL, which enables self-driving cars to interact with other vehicles and pedestrians. Despite these advancements, RL in self-driving cars remains an active area of research, with ongoing efforts to improve robustness, interpretability, and scalability[16].

4.3 Unmanned Aerial Vehicles (UAVs)

UAVs, or drones, represent another important application domain for RL, as they must operate in dynamic and unstructured environments. RL has been used to train UAVs for tasks such as navigation, surveillance, and delivery. For example, RL algorithms have been applied to learn policies for controlling the flight path of a UAV based on sensor inputs such as GPS data, camera images, and inertial measurements. These tasks require the UAV to learn robust control policies that can adapt to variations in the environment, such as changes in wind speed or obstacles[17].

One of the key challenges in applying RL to UAVs is the need for real-time decision-making. UAVs typically operate in fast-changing environments, which require the RL algorithm to make decisions quickly and efficiently[18]. To address this challenge, researchers have developed techniques such as online RL, which enables the UAV to learn and adapt in real-time, and transfer learning, which leverages knowledge from previous tasks to accelerate learning. Despite these advancements, RL in UAVs remains a challenging area of research, with ongoing efforts to improve efficiency, robustness, and scalability[19].

5. Tables

Below are two tables summarizing key RL algorithms and their applications in autonomous systems.

Table 1: Summary of RL Algorithms

Algorithm	Type	Key Features	Applications in Autonomous Systems
Q-learning	Value-based	Model-free, off-policy	Robotic navigation, self-driving cars
DQN	Value-based	Deep neural networks	Atari games, robotic control
PPO	Policy-based	Stable, efficient	Robotic manipulation, UAV control
SAC	Actor-critic	Continuous control, high-dimensional	Self-driving cars, UAV navigation

6. Future Research Directions and Challenges

Reinforcement Learning (RL) has made remarkable strides in enabling autonomous systems to learn and adapt to complex environments. However, as the field continues to evolve, several critical challenges and research directions must be addressed to fully realize the potential of RL in real-world applications. Below, we outline six key areas that represent promising avenues for future research and innovation[20].

6. 1. Sample Efficiency

One of the most significant challenges in RL is its sample inefficiency. Many RL algorithms require a large number of interactions with the environment to learn effective policies, which can be impractical for real-world applications where data collection is expensive, time-consuming, or risky[21]. For example, training a self-driving car in real traffic or a robotic arm in a physical factory setting involves substantial costs and safety concerns. Future research should focus on developing more sample-efficient algorithms that can learn from limited data. Techniques such as meta-learning, transfer learning, and sim-to-real transfer are promising approaches. Meta-learning enables agents to leverage knowledge from previous tasks, while transfer learning allows them to apply learned policies to new but related tasks. Sim-to-real transfer, on the other hand, uses simulations to train agents before deploying them in the real world, significantly reducing the need for real-world data[22].

6. 2. Safety and Robustness

Ensuring the safety and robustness of RL-based autonomous systems is another critical challenge. Unlike traditional control systems, which are designed with explicit safety constraints, RL algorithms learn through trial and error, which can lead to unpredictable and potentially unsafe behavior. This is particularly concerning in safety-critical applications such as self-driving cars, UAVs, and medical robotics. Future research should focus on developing safe RL algorithms that incorporate safety constraints into the learning process. Techniques such as constrained RL, risk-sensitive RL, and robust RL are promising approaches. Additionally, formal verification methods can be used to certify the safety of learned policies before deployment[23].

Table 2: Applications of RL in Autonomous Systems

Application Domain	Key Challenges	RL Algorithms Used	Key Achievements
Robotics	High-dimensional state-action space	Q-learning, PPO, SAC	Object manipulation, locomotion
Self-driving cars	Safety, reliability	DQN, SAC, A3C	Lane keeping, obstacle avoidance
UAVs	Real-time decision-making	PPO, SAC, online RL	Navigation, surveillance, delivery

6. 3. Scalability

Scalability remains a major challenge in RL, particularly for applications with high-dimensional state and action spaces. Autonomous systems such as robotic manipulators, self-driving cars, and UAVs often operate in complex environments with a large number of variables, making it difficult for RL algorithms to explore and learn effectively. Future research should focus on developing scalable RL algorithms that can handle high-dimensional spaces efficiently. Advances in deep learning, such as the use of high-capacity neural networks and distributed training methods, have already made significant contributions in this area. However, further work is needed to improve the generalization and interpretability of these algorithms[24].

6. 4. Multi-Agent Reinforcement Learning

Many real-world applications of autonomous systems involve multiple agents interacting with each other in a shared environment. For example, self-driving cars must navigate alongside other vehicles, pedestrians, and cyclists, while UAVs must coordinate with other drones to perform tasks such as surveillance and delivery. Multi-agent RL (MARL) is a promising approach for addressing these challenges, but it introduces additional complexities such as non-stationarity, communication, and coordination. Future research should focus on developing MARL algorithms that can handle these complexities effectively. Techniques such as centralized training with decentralized execution, communication protocols, and cooperative learning are promising directions[25].

6. 5. Explainability and Interpretability

As RL-based autonomous systems become more prevalent, there is a growing need for explainability and interpretability. Many RL algorithms, particularly those based on deep learning, are often considered "black boxes" because their decision-making processes are not easily understandable by humans[26]. This lack of transparency can be problematic in safety-critical applications where human operators need to trust and understand the system's behavior. Future research should focus on developing explainable RL algorithms that provide insights into the decision-making process. Techniques such as attention mechanisms, interpretable models, and post-hoc explanations are promising approaches[27].

6. 6. Ethical and Societal Implications

The deployment of RL-based autonomous systems raises important ethical and societal questions. For example, the use of self-driving cars raises concerns about liability, accountability, and the impact on employment in the transportation sector. Similarly, the use of UAVs for surveillance and delivery raises concerns about privacy and security. Future research should focus on addressing these ethical and societal implications to ensure that RL-based systems are deployed responsibly and equitably. Interdisciplinary research involving experts from fields such as ethics, law, and social sciences will be essential in addressing these challenges[28].

7. Conclusion

Reinforcement Learning (RL) has emerged as a transformative paradigm in the field of autonomous systems, offering a robust framework for enabling machines to learn, adapt, and optimize their behavior in dynamic and uncertain environments. This article has provided a comprehensive review of RL algorithms and their applications across various domains, including robotics, self-driving cars, and unmanned aerial vehicles (UAVs). By examining the fundamental principles of RL, exploring state-of-the-art algorithms, and analyzing their strengths and limitations, we have highlighted the significant progress that has been made in this field. However, despite these advancements, several challenges remain, which must be addressed to fully realize the potential of RL in real-world applications[29].

One of the most pressing challenges in RL is sample inefficiency. Many RL algorithms require a large number of interactions with the environment to learn effective policies, which can be prohibitively expensive and time-consuming in real-world scenarios. This is particularly problematic for applications such as self-driving cars and UAVs, where data collection is often constrained by safety and logistical considerations. To address this issue, researchers are exploring techniques such as meta-learning, which enables agents to leverage knowledge from previous tasks, and sim-to-real transfer, which uses simulations to accelerate learning. These approaches have shown promise in reducing the sample complexity of RL algorithms, but further research is needed to make them more robust and scalable[30].

Another critical challenge is ensuring the safety and reliability of RL-based autonomous systems. Unlike traditional control systems, which are designed with explicit safety constraints, RL algorithms learn policies through trial and error, which can lead to unpredictable and potentially unsafe behavior. This is particularly concerning in safety-critical applications such as self-driving cars and UAVs, where errors can have severe consequences. To mitigate these risks, researchers are developing techniques such as safe RL, which incorporates safety constraints into the learning process, and robust RL, which ensures that the learned policies are resilient to uncertainties and disturbances. While these approaches have made significant strides, further work is needed to ensure that RL-based systems can operate safely and reliably in real-world environments.

Scalability is another major challenge in RL, particularly for applications with high-dimensional state and action spaces. Autonomous systems such as robotic manipulators and self-driving cars often operate in complex environments with a large number of variables, making it difficult for RL algorithms to explore and learn effectively. Advances in deep learning and parallel computing have helped to address this challenge by enabling the use of high-capacity neural networks and distributed training methods. However, scaling RL algorithms to real-world problems remains an open research question, with ongoing efforts to improve efficiency, generalization, and interpretability.

In addition to these technical challenges, there are also ethical and societal considerations that must be addressed as RL-based autonomous systems become more prevalent. For example, the deployment of self-driving cars raises questions about liability, accountability, and the impact on employment in the transportation sector. Similarly, the use of UAVs for surveillance and delivery raises concerns about privacy and security. As RL continues to advance, it is essential to engage in interdisciplinary research and dialogue to address these issues and ensure that the benefits of RL are realized in a responsible and equitable manner[31].

Despite these challenges, the potential of RL in autonomous systems is immense. RL has already demonstrated remarkable success in a wide range of applications, from robotic manipulation and navigation to autonomous driving and UAV control. As algorithms continue to improve and new techniques are developed, RL is expected to play an increasingly important role in shaping the future of autonomous systems. By addressing the challenges of sample inefficiency, safety, scalability, and ethics, researchers can unlock the full potential of RL and enable the development of intelligent, adaptive, and autonomous systems that can operate effectively in complex and dynamic environments.

In conclusion, Reinforcement Learning represents a powerful and versatile tool for enabling autonomous systems to learn and adapt to the challenges of the real world. While significant progress has been made, there is still much work to be done to overcome the technical and societal challenges that remain. By continuing to push the boundaries of RL research and innovation, we can pave the way for a future where autonomous systems are not only capable but also safe, reliable, and beneficial to society. The journey ahead is undoubtedly challenging, but the potential rewards are immense, making RL one of the most exciting and impactful areas of research in the field of artificial intelligence.

References

[1] W. Yuan, K. Hang, D. Kragic, M. Y. Wang, and J. A. Stork, “End-to-end nonprehensile rearrangement with deep reinforcement learning and simulation-to-reality transfer,” *Rob. Auton. Syst.*, vol. 119, pp. 119–134, Sep. 2019.

[2] C. Zhang *et al.*, “SA-IGA: a multiagent reinforcement learning method towards socially optimal outcomes,” *Auton. Agent. Multi. Agent. Syst.*, vol. 33, no. 4, pp. 403–429, Jul. 2019.

[3] B. Recht, “A tour of reinforcement learning: The view from continuous control,” *Annu. Rev. Control Robot. Auton. Syst.*, vol. 2, no. 1, pp. 253–279, May 2019.

[4] M. Bouton, K. D. Julian, A. Nakhaei, K. Fujimura, and M. J. Kochenderfer, “Decomposition methods with deep corrections for reinforcement learning,” *Auton. Agent. Multi. Agent. Syst.*, vol. 33, no. 3, pp. 330–352, May 2019.

[5] C. You, J. Lu, D. Filev, and P. Tsiotras, “Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning,” *Rob. Auton. Syst.*, vol. 114, pp. 1–18, Apr. 2019.

[6] Y. Tsurumine, Y. Cui, E. Uchibe, and T. Matsubara, “Deep reinforcement learning with smooth policy

update: Application to robotic cloth manipulation," *Rob. Auton. Syst.*, vol. 112, pp. 72–83, Feb. 2019.

[7] I. Yoon, A. Anwar, T. Rakshit, and A. Raychowdhury, "Transfer and online reinforcement learning in STT-MRAM based embedded systems for autonomous drones," *arXiv [cs.OH]*, 21-Apr-2019.

[8] R. Riveret, Y. Gao, G. Governatori, A. Rotolo, J. Pitt, and G. Sartor, "A probabilistic argumentation framework for reinforcement learning agents," *Auton. Agent. Multi. Agent. Syst.*, vol. 33, no. 1–2, pp. 216–274, Mar. 2019.

[9] S. A. Raza and M.-A. Williams, "Human feedback as action assignment in interactive reinforcement learning," *ACM Trans. Auton. Adapt. Syst.*, vol. 14, no. 4, pp. 1–24, Dec. 2019.

[10] P. Hernandez-Leal, B. Kartal, and M. E. Taylor, "A survey and critique of multiagent deep reinforcement learning," *Auton. Agent. Multi. Agent. Syst.*, vol. 33, no. 6, pp. 750–797, Nov. 2019.

[11] A. Murad, F. A. Kraemer, K. Bach, and G. Taylor, "Autonomous management of energy-harvesting IoT nodes using deep reinforcement learning," in *2019 IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, Umea, Sweden, 2019.

[12] Q. Liu and Q. Hui, "The formation control of mobile autonomous multi-agent systems using deep reinforcement learning," in *2019 IEEE International Systems Conference (SysCon)*, Orlando, FL, USA, 2019.

[13] I. Yoon, M. A. Anwar, R. V. Joshi, T. Rakshit, and A. Raychowdhury, "Hierarchical memory system with STT-MRAM and SRAM to support transfer and real-time reinforcement learning in autonomous drones," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 9, no. 3, pp. 485–497, Sep. 2019.

[14] I. Yoon, A. Anwar, T. Rakshit, and A. Raychowdhury, "Transfer and online reinforcement learning in STT-MRAM based embedded systems for autonomous drones," in *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Florence, Italy, 2019.

[15] Z. Huang, X. Xu, H. He, J. Tan, and Z. Sun, "Parameterized batch reinforcement learning for longitudinal control of autonomous land vehicles," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 49, no. 4, pp. 730–741, Apr. 2019.

[16] S. Jeong, G. Yoo, M. Yoo, I. Yeom, and H. Woo, "Resource-efficient sensor data management for autonomous systems using deep reinforcement learning," *Sensors (Basel)*, vol. 19, no. 20, p. 4410, Oct. 2019.

[17] L. Xin *et al.*, "Accelerated inverse reinforcement learning with randomly pre-sampled policies for autonomous driving reward design," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, 2019.

[18] M. Brittain and P. Wei, "Autonomous separation assurance in an high-density en route sector: A deep multi-agent reinforcement learning approach," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, 2019.

[19] A. Rodriguez-Ramos, C. Sampedro, H. Bavle, P. de la Puente, and P. Campoy, "A deep reinforcement learning strategy for UAV autonomous landing on a moving platform," *J. Intell. Robot. Syst.*, vol. 93, no. 1–2, pp. 351–366, Feb. 2019.

[20] A. Marinescu, I. Dusparic, and S. Clarke, "Prediction-based multi-agent reinforcement learning in inherently non-stationary environments," *ACM Trans. Auton. Adapt. Syst.*, vol. 12, no. 2, pp. 1–23, Jun. 2017.

[21] H. Wang *et al.*, "Integrating reinforcement learning with multi-agent techniques for adaptive service composition," *ACM Trans. Auton. Adapt. Syst.*, vol. 12, no. 2, pp. 1–42, Jun. 2017.

[22] J. Chen, B. Yuan, and M. Tomizuka, "Model-free deep reinforcement learning for urban autonomous driving," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, 2019.

[23] W. Wu and L. Gao, "Posture self-stabilizer of a biped robot based on training platform and reinforcement learning," *Rob. Auton. Syst.*, vol. 98, pp. 42–55, Dec. 2017.

[24] G. Ortali and I. Tollis, "A new framework for hierarchical drawings," *J. Graph Algorithms Appl.*, vol. 23, no. 3, pp. 553–578, Sep. 2019.

[25] T. Castermans, M. Van Garderen, W. Meulemans, M. Nöllenburg, and X. Yuan, "Short Plane Supports for Spatial Hypergraphs," *J. Graph Algorithms Appl.*, vol. 23, no. 3, pp. 463–498, Sep. 2019.

[26] D. Gonçalves, B. Lévêque, and A. Pinlou, "Homothetic triangle representations of planar graphs," *J. Graph Algorithms Appl.*, vol. 23, no. 4, pp. 745–753, Sep. 2019.

[27] K. Iwamura, T. Sugano, and N. Sugimura, "A study on autonomous distributed scheduling by using deep reinforcement learning," *Proc. Manuf. Syst. Div. Conf.*, vol. 2017, no. 0, p. 203, 2017.

[28] F. Frati, M. Hoffmann, and V. Kusters, "Simultaneous embeddings with few bends and crossings," *J. Graph Algorithms Appl.*, vol. 23, no. 4, pp. 683–713, Sep. 2019.

[29] E. Arseneva *et al.*, "Pole dancing: 3D morphs for tree drawings," *J. Graph Algorithms Appl.*, vol. 23, no. 3, pp. 579–602, Sep. 2019.

[30] T. Mchedlidze, A. Pak, and M. Klammler, "Aesthetic discrimination of graph layouts," *J. Graph Algorithms Appl.*, vol. 23, no. 3, pp. 525–552, Sep. 2019.

[31] R. Janssen, M. Jones, S. Kelk, G. Stamoulis, and T. Wu, "Treewidth of display graphs: bounds, brambles and applications," *J. Graph Algorithms Appl.*, vol. 23, no. 4, pp. 715–743, Sep. 2019.