Journal of Analysis and Computation (JAC)
(An International Peer Reviewed Journal), www.ijaconline.com,
ISSN 0973-2861Volume XVII, Issue I, Jan-June 2023

ADVANCEMENTS IN REINFORCEMENT LEARNING-AN OVERVIEW

Sumedha Sharma[1], Tanisha Jain[2], Sanidhya Chaturvedi[3], Swapnil Saraswat[4], Ritvik Sharma[5]
JAIPUR ENGINEERING COLLEGE AND RESEARCH CENTRE (JECRC) FOUNDATION
Jaipur, Rajasthan, 302004, India

ABSTRACT

In the discipline of machine learning, reinforcement learning (RL) is a well-known study area that focuses on sequential decision-making in dynamic contexts. An extensive overview of reinforcement learning is provided in this publication, covering its key concepts, methodologies, and challenges. RL involves mapping situations to actions to maximize the associated rewards, having an agent discovering the behaviours that result in the greatest rewards through trial and error. Key challenges in RL, such as deriving optimal policies, credit assignment, dealing with complex environments, and temporal correlations, are explored. Additionally, the paper delves into the concept of transfer learning, where knowledge is transferred across related tasks to enhance RL performance. The use of transfer learning in single-agent and multi-agent systems is discussed, highlighting methods like instance transfer, representation transfer, and parameter transfer. This paper provides valuable insights into the foundations of RL and its application in solving real-world problems, offering a basis for further research and advancements in this exciting field.

Keywords- Reinforcement learning, sequential decision-making, challenges, transfer learning, real-world problems.

1 INTRODUCTION

One paradigm, reinforcement learning (RL), has distinguished itself as a leader in enabling machines to learn and make judgements in the dynamic field of artificial intelligence (AI). Via interaction with their surroundings, learning via trial and error, and receiving feedback in the form of rewards or penalties, agents can interact with their environment using this dynamic technique. Due to its capacity to handle complicated issues, reinforcement learning has grown in popularity across a variety of industries, from robotics to banking. However, RL algorithms' low interpretability has prevented their widespread use and comprehension. Here we

introduce explainable AI (XAI), a crucial research field with the goal of fostering transparency and comprehension in AI systems.

This research paper explores the intersection of XAI and RL, with a specific focus on the decision-making processes of RL agents. While RL has achieved remarkable success in

Games like GO and Atari, its lack interpretability hampers trust and acceptance. To address these challenges, this study delves into XAI approaches within RL, shedding light on the explanations and interpretability of RL agents' decision-making. Furthermore, it emphasizes the importance of transparency and understanding of AI systems, especially as AI finds applications in sensitive areas with social, ethical and security implications.

Reinforcement learning (RL), a technique used to develop autonomous agents that can learn the best behaviour by interacting with their environment, has gained interest in the field of artificial intelligence (AI). Traditional RL methods, however, were restricted to low-dimensional problems and had scaling issues. The ability to use RL in high-dimensional spatial domains like photos and natural language has been made possible by the development of deep learning enabled by deep neural networks.

DRL algorithms have achieved impressive successes, surpassing human-level performance in Atari games, defeating world champions in games like Go, and excelling in multiplayer video games. Beyond gaming, DRL has found applications in robotics, dialog systems, education, autonomous vehicles, smart grids, and recommender systems. However, practical implementation of DRL remains challenging, particularly in terms of sample efficiency and handling real-time systems.

Hierarchical reinforcement learning (HRL) offers a promising solution by breaking down complex problems into manageable tasks. HRL addresses various RL challenges, including reward definition, research, sampling efficiency, transfer learning, lifelong learning and interpretability. While DRL has shown potential, challenges such as specifying dense rewards, exploration efficiency, and sample inefficiency persist. Lifelong learning, transfer learning and interpretability are also areas that require further research and HRL offers possible solutions.

Reinforcement learning (RL) has become a powerful tool in various domains, optimizing agent behavior by maximizing rewards in an environment. However, the design of the reward function poses challenges. Reward hacking occurs when agents find ways to maximize rewards without performing the intended task, leading to undesired behaviours. Reward shaping involves finding the right balance between goal definition and guidance. Infinite rewards are challenging for traditional RL algorithms, requiring arbitrary, finite values. Aggregating multiple reward signals necessitates trade-offs, which may not be straightforward.

The reward shaping issue has gained attention, particularly in intrinsic versus extrinsic motivation. Extrinsic motivation focuses on defined goals, while intrinsic motivation serves as guidance. Learning intrinsic motivation automatically has been proposed, but relying solely on extrinsic goals or using scalar rewards can be challenging.

RL and deep neural networks (DNNs) are combined in deep reinforcement learning (DRL), which has produced outstanding results, surpassing human-level performance in games like Atari and chess. However, DNN decisions can lack transparency, leading to the emergence of Explainable Artificial Intelligence (XAI) techniques.

Integrating XAI with DRL is crucial for correctness and safety. Explanations shed light on model decisions and facilitate strategy analysis. Prompt delivery of explanations is vital, especially in time-sensitive applications.

This review explores the intersection of DRL and XAI, discussing advancements in deep learning's impact on RL. It delves into challenges in multi-task learning, resource management, scalability, and the importance of explain ability for trust and safety in critical operations. Integrating RL and DNNs with XAI holds promise for advancing AI technologies. Understanding their strengths, limitations, and future directions is crucial for building trustworthy and reliable AI systems.

1.1Setup For Reinforcement Learning:

An agent is positioned in an environment (E) and interacts with it at specific time intervals in a standard reinforcement learning setting. The agent finds itself in a certain state (St) at each point in time (t), at which point it acts in the environment. The environment then reacts by changing the current state (St) to a new state (St+1) at a new time step (t+1), and it gives the agent a reward that represents the worth of the action they took in the earlier state (St). The typical ecology of a reinforcement learning environment is shown in Figure 1 for any given time step (t).

Figure 1 ecosystem of reinforcement learning

1.2 Key Challenges of Reinforcement Learning:

- In order to choose the best course of action, an agent must use a brute-force approach based only on reward values.

- The RL agent must estimate the highest predicted discounted future reward for an action when it is taken in a particular state. The credit assignment dilemma is another name for this issue.

- Environments with 3D nature can lead to a large number of continuous state and action pairs.

- In a complex environment, an agent's activities are highly relied upon for observations, which can possess robust temporal connections.

2 Applications of Reinforcement Learning:

2.1 Self-Driving Cars:

Reinforcement learning can be used to optimize the trajectory of self-driving cars, considering factors such as speed limits, drivable zones, and avoiding collisions. The RL agent learns to generate optimal trajectories based on feedback and rewards and Parking and our algorithm can also help to develop automatic parking facilities, enabling self-driving cars to learn how to parkin different situations.

2.2 Gaming:

Reinforcement learning is widely used in gaming, allowing agents to learn optimal strategies and improve game play.

Games provide an ideal environment for reinforcement learning agents to explore and learn. Reinforcement algorithm has proved success in different games, shows as chess Go, Atari 2600 games, lot many more.

Deep Mind's Alpha Zero and Alpha Star are notable examples of reinforcement learning agents that have achieved superhuman performance in games.

2.3 Hyper parameters Selection for Neural Networks:

Determining the architecture and hyper parameters of neural networks can be treated as learning issues.

RNN-based models have been trained using reinforcement learning to generate hyper parameters for neural networks.

This approach has shown competitive results compared to other optimization methods.

2.4 Marketing:

Reinforcement learning can be employed in marketing for personalized recommendations and targeted advertising.

RL agents can learn customer preferences and optimize marketing strategies to maximize engagement and conversions.

2.5 Intelligent Transportation Management:

By dynamically modifying signal timing schedules, adaptive traffic signal control (ATSC) helps lessen traffic congestion.

The ATSC problem has been addressed using multi-agent reinforcement learning techniques that combine game theory and reinforcement learning.

The exploration-exploitation trade-off, the curse of dimensionality, stability, and no stationary are among the difficulties.

2.6 Natural Language Processing (NLP):

Several NLP tasks, including text creation, machine translation [12, 13], and conversation systems, have incorporated reinforcement learning.

In dialogue systems, reinforcement learning is applied to learn dialogue policies and optimize system responses.

2.7 Dialogue Systems:

Dialogue systems interact with natural language and can be task-oriented or chatbot-based. In the modular approach, reinforcement learning is frequently employed to teach discussion policies.

In the end of approach, reinforcement learning enhances dialogue system performance by optimizing system responses.

The Alexa Prize competition has seen the application of reinforcement learning in building conversational bots.

3 Some New Evolution In RL:

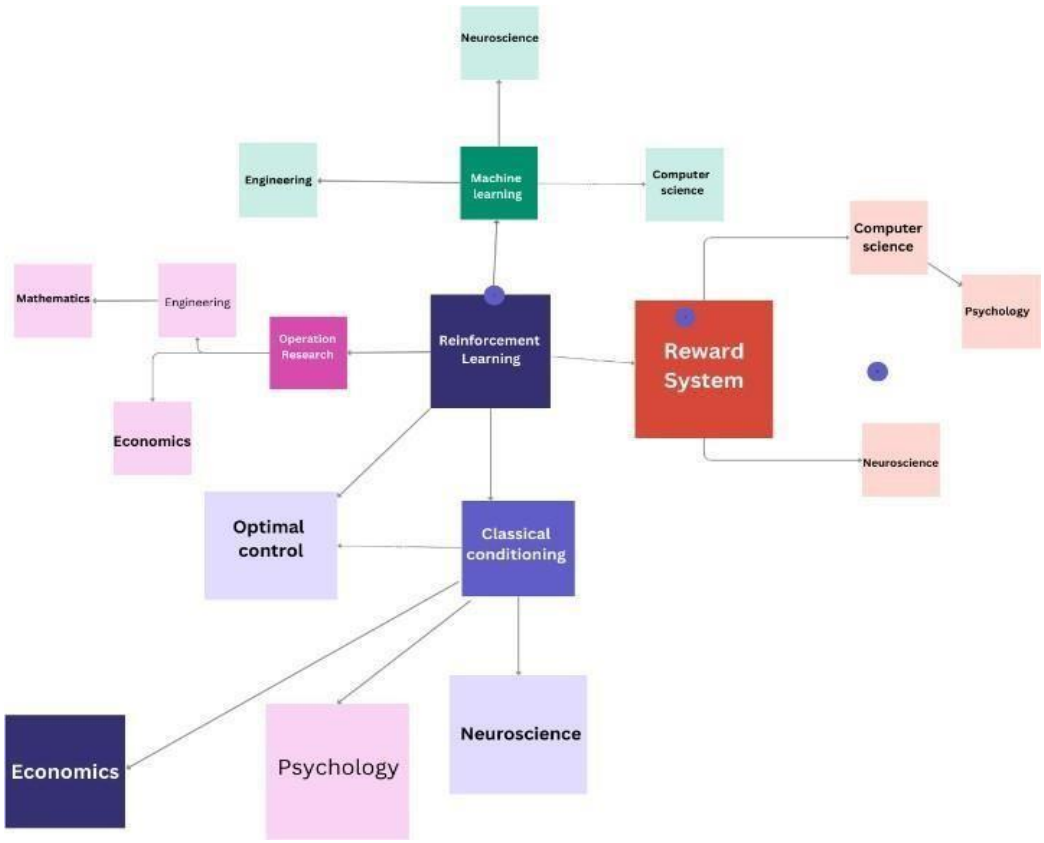

Figure 2 New evolutions in RL

3.1 Topological Q-Learning :

As and when the agent perceives the environment, it employs the topological ordering of the states. The two steps of this approach are task learning and exploration optimization. During the task-learning phase, the agent constructs the topological ordering of the states; during the exploration optimisation, the exploration is guided by an internal incentive function. In order to provide directed exploration, this algorithm uses an intrinsic or internal incentive system.

The capacity of Topological Q-Learning to generalize knowledge across states with comparable topological features is one of its main benefits. The agent can transfer learned experiences from one state to another by making use of the underlying topology, even if they were not explicitly encountered during training. Particularly in complicated situations where classical Q-Learning suffers with generalization, this trait can result in more effective learning.

3.2 Epoch-Incremental RL :

The interactions between an agent and its environment are broken down into episodes, or epochs, in reinforcement learning algorithms. A series of interactions between agents and environments make up each episode. Typically, the whole number of iterations is unknown beforehand. Each episode comes to a unique conclusion known as the terminal state.

Learning incrementally: A learning paradigm known as incremental learning, sometimes referred to as online learning or lifetime learning, involves continuously updating a model's knowledge as new data becomes available.

The interactions between agents and environments happen in incremental mode. The action-value function, which represents the agent's policy, is then modified as a result of these interactions. When the terminal condition is reached and the reinforcement signal is assigned, the incremental mode is terminated. After then, the epoch mode starts, during which all of the previously active states' distances from the terminal state are estimated using the environment model. Each state is also given an action an enables the best possible approach to the terminal state.

3.3 Multi-scale RL:

Utilizes specific mathematical operations to produce an abstract of the state-space graph. It produces various levels of abstraction. Consequently, action selection is carried out utilising multi-scale on the reduced abstraction map SoftMax selection [33].

The main concept behind Multi-Scale RL is grasping policies and value functions at each level, allowing the agent to reason and act at different levels of granularity. The lower-level policies focus on fine-grained control and short-term decisions, while the higher-level policies focus on coarse-grained control and long-term planning. This hierarchical structure facilitates more efficient learning and more robust decision-making.

3.4 Hierarchical RL(HRL):

Hierarchical Reinforcement Learning (HRL) is a framework that addresses the challenge of learning in complex tasks by leveraging hierarchical structures and decomposing the

problem into multiple levels of abstraction. It aims to improve learning efficiency, enable faster decision-making, and facilitate the handling of tasks with long time horizons.

A hierarchy of sub-goals is used to break down a bigger goal. Each subtask could still be broken down smaller tasks, with primitive actions often being the lowest level of tasks[34].

HRL gives us multiple benefits during training and exploration:

1. Since top prominent is to multiple environmental steps, episodes are relatively shorter, thus propagating rewards faster and improving learning.

2. Since exploration happens at a higher level, it is able to learn more meaningful policies and thus take more meaningful actions than those taken on an atomic level. Example, agent would learn better policies to reach the goal at the level of grasping objects than at the level of understanding joint movements of fingers.

3.5 RL in the realms of Game Premise:

Game theory is typically referred to as the mathematics of disputes and is used in a variety of disciplines, including economics, psychology, artificial intelligence (AI), sociology, etc. In game theory applied to real-world situations, the policy is the strategy that maps every conceivable state of an action with respect game's players.

- Multi-Agent RL (MARL) games come in the following varieties:
- Static Games: autonomous decision-making.
- Stage Games: the guidelines vary depending on the stage.
- Repeated Games: When a series of games are played.

The "Nash equilibrium (NE)"25, which is a compilation of all players'/agents' strategies, is a crucial component of all gaming theories. Every player or agent aims to reach equilibrium, or the best results for himself, but this is impossible without the help of other agents. Now, from the perspective of game theory, this is what we need to remember. In their study of algorithms for solving RLs for stochastic games, Bowling and Veloso found that the majority of them— including Shapely, Policheck and Avi-Itzhak, Van der Wal, Fictitious Play, and others— used variations of the RL's temporal difference learning component.

4 Comprehensive Analysis of Literature:

Anna Goldie and AzaliaMirhoseini [1] the paper discusses the importance of placement optimization, the use of reinforcement learning, and the lessons learned from research. The placement task is described as a reinforcement learning problem and a general description of deep reinforcement learning is provided. The challenges encountered during training, insights into the efficacy of reinforcement learning techniques, and potential enhancements or refinements to the approach are likely to be included in these lessons.

Christian Wirth, Riad Akrour, Gerhard Neumann and Johannes Fürnkranz [2] By directly learning from an expert's preferences rather than relying on a hand-designed numeric reward function, preference-based reinforcement learning (PbRL) algorithms offer an alternative to

traditional RL methods. PbRL's ability to solve the problem of reward shaping, learn from nonnumeric rewards, and reduce reliance on expert knowledge has made it popular. Their statement acknowledges the current shortcomings of PbRL algorithms, raises unanswered research questions for further investigation, and provides a brief overview of practical problems that have been successfully resolved using PbRL methods.

Deepanshu Mehta [3] The study in a variety of Reinforcement Learning-related fields are compiled in this study article. In the rapidly growing field of reinforcement learning, there are numerous learning algorithms that can be used, including Markov Decision Processes (MDPs), Temporal Difference (TD) Learning, Advantage Actor-Critic (A2C), Asynchronous Advantage Actor-Critic (A3C), Deep Deterministic Policy Gradient (DDPG), and Evolution Strategies (ES).Some of the most recent applications of this technology include neural scene rendering, brain-computer interface, stock predictions, trading, sports betting, proving challenging mathematical theorems, healthcare, astronomy, business, manufacturing, chat bots, self-driving cars, and superhuman video game performance.

Heyang Qin, Syed Zawad and Yanqi Zhou [4] The end made by them is that the proposed quick AI serving booking structure, in light of a District based Support Learning (RRL) approach, can successfully work on the presentation of AI as-a-Administration (MLaaS) frameworks. Reduced inference latency and a lower risk of violating Service-Level-Objectives (SLOs) result from the RRL approach's ability to identify the best parallelism configurations for various workloads.

Researchers claim that their method performs better than current cutting-edge techniques. In support of this claim, they provide both theoretical and experimental evidence. Near-optimal solutions can be found over 8 times faster using the RRL method, and inference latency and SLO violations can be reduced by up to 49.9% and 79.0%, respectively.

Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage and Anil Anthony Bharath [5]

It highlights the inherent advantages of deep neural networks, particularly in the field of visual reinforcement learning. They come to the conclusion that deep reinforcement learning research is currently very active, pointing to ongoing developments and advancements in this field.

Leslie Pack Kaelbling , Michael L. Littman and Andrew W. moore [6] Their paper gives an exhaustive study of the field of support gaining from a software engineering viewpoint. Its goal is to be usable by researchers with prior experience working with machine learning. The paper sums up both the verifiable underpinnings of the field and an expansive determination of flow research. Some of the key issues in reinforcement learning are covered in this paper, including the exploration-exploitation trade-off, the use of generalization and hierarchy, dealing with hidden state.

These are of many manifestations of necessary bias: Molding, nearby support signals, Impersonation, Issue disintegration, Reflexes. Reflexes can be used to improve the safety and efficiency of robot learning, according to recent research by Millan (1996).Learning techniques for approximating, decomposing, and incorporating bias into problems still requires a lot of work and a lot of interesting questions.

Lindsay Wells and Tomasz Bednarz [7] As a result of the growing demand for transparency and trust in AI systems, they concluded that Explainable AI (XAI) has been gaining traction. In sensitive areas where AI has societal, ethical, and safety implications, this need is especially critical.

This review specifically looks at the use of XAI in the field of reinforcement learning (RL), even if previous research in XAI has mostly focused on Machine Learning (ML) for categorization, decision, or action. 25 investigations, including 5 snowball-examined tests, were chosen for inspection out of 520 list items. Visualisation, query-based explanations, policy summary, human-in-the-loop cooperation, and verification are a few of the XAI for RL developments that have been noticed.

Marc G. Bellemare , Will Dabney and Remi Munos [8] They concluded in this paper that reinforcement learning relies heavily on the value distribution rather than just the expected value. The theoretical findings in the contexts of policy evaluation and control are presented in the paper.

The paper highlights how the value distribution affects learning in the approximate setting by combining theoretical and empirical evidence. The devisor reason that thinking about the full conveyance of profits, instead of simply the normal worth, is pivotal for accomplishing better execution and understanding the elements of support study.

Matthias Hutsebaut-Buysse and Kevin Mets and Steven Latré [9] they concluded that apromising strategy for resolving difficult sequential decision-making issues is HRL. While pure random exploration is typically unproductive and sample inefficient in these situations, HRL makes use of temporal and state abstractions to overcome these issues. Reusing behaviours, making RL systems easier to understand, and semi-automatically discovering and learning abstractions are all made possible by HRL. The Choices structure and the objective contingent methodology are presented as instances of HRL procedures that work with the learning of deliberations and their usage in perplexing, high-layered conditions. The main goal is to progress the development of HRL agents that can understand abstractions and use them efficiently when interacting with complex surroundings.

N. R. Ravishankar and M. V. Vijayakumar [10] the paper discusses how various algorithms are progressing within each dimension and introduces the idea of classifying RL as a three-dimensional problem. It begins with a review of fundamental RL classifications and discusses some well-known but older algorithms. The paper then looks at recent trends and gives a high-level overview of the RL landscape as a whole. .In order to provide readers with an accessible overview and abstractions, they emphasize that they have deliberately avoided complex mathematical equations and concepts.

They concentrate primarily on gaming theory because this area has a more challenging problem to answer and the way the problem is represented itself adds a new dimension. Examples include matrix games and stochastic games, both of which have a Nash equilibrium state as its solution.

Nelson Vithayathil Varghese and Qusay H. Mahmoud [11] The paper's goal is to review the current research issues surrounding multitasking in the deep reinforcement learning field and

present cutting-edge solutions to these issues. We compare and contrast the solutions offered by DISTRAL, IMPALA, and PopArt for scalability. The new heading, known as profound support learning, has fundamentally worked on the presentation of support learning based savvy specialists in sans model methodologies.

Nuo Xu [12] He emphasised the importance of artificial intelligence as a widely debated interdisciplinary area. AI has emerged as a crucial component of computer-based intelligence, encompassing several subfields. Two essential algorithms, the Partially Observable Markov Decision Process (POMDP) and the Markov Decision Process (MDP), are used to examine reinforcement learning. These equations provide a framework for comprehending and putting support learning practises into practise.

Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau ,DoinaPrecup and David Meger [13] They came to the conclusion that sustaining advancement in the field necessitates reproducing and accurately assessing the improvements provided by novel deep reinforcement learning (RL) methods. Due to this lack of reproducibility, it is challenging to establish whether purported advancements above the prior state-of-the-art are significant.

They likewise recommend rules to further develop reproducibility and call for conversations on the most proficient method to limit squandered exertion brought about by non- reproducible and handily confounded results.

Robert N. Boutea,B , JorenGijsbrechts C, Willem Van Jaarsveldd and Nathalie Vanvuchelena [14] They came to the conclusion that sequential decision-making tasks, such as early developments in inventory control, have shown promise for deep reinforcement learning (DRL).

Sequential decision-making tasks can benefit from DRL algorithms, but their practical usefulness is limited by the complexity of the algorithms' design and the computational demands placed on them. This report proposes future research areas to improve and broaden the range of DRL applications in inventory control.

Ronald Parr and Stuart Russell [15] The finish of the exploration support learning with ordered progressions of somewhat determined machines is that this approach takes intoconsideration the fuse of earlier information to oblige the inquiry space and works with information move across various issues. The research's method can be seen as a bridge between behavior-based or reactive approaches to control and reinforcement learning. It provides algorithms for learning and problem-solving with hierarchical machines that are demonstrated to be convergent.

Thomas Hickling ,Abdelhafid Zenati, Nabil Aouf and Phillippa Spencer [16] Their paper concludes that the lack of interpretability in Deep Reinforcement Learning (DRL) algorithms sparked the development of the Explainable Artificial Intelligence (XAI) field. Since its introduction in 2015, researchers and the general public lack understanding and trust due to its lack of interpretability.Their research aims to determine which XAI approaches are best suited to various applications. Researchers have the ability to improve the trustworthiness of DRL models and the interpretability of their models by comprehending which approaches are most effective in particular situations. This also aims to point out any methods that haven't been used enough, pointing out areas where more research and application can be done.

Yan Duan, ohn Schulman, Xi Chen, Peter L. Bartlett, Ilya Sutskeve and Pieter Abbeel[17] They discovered that the proposed technique, RL2, which turns the reinforcement learning algorithm into a recurrent neural network (RNN) and trains it using data**, is effective.**, shows promising results in bridging the gap between deep reinforcement learning and animals' capacity to quickly learn new tasks using prior knowledge**.** Markov Decision Process (MDP) by slowly learning the algorithm through a general-purpose RL algorithm and encoding it in the weights of the RNN. The condition of the "quick" RL computation on the current, already-hidden MDP is stored in the RNN's enactments.

5 CONCLUSION

Ongoing research explores the combination of RL with cognitive and neuroscience, investigating the relationship between the human brain's dopamine response, action choices, and RL algorithms. This emerging field aims to understand cognitive search and planning processes but is still in its early stages. Although RL has made significant progress in all dimensions, the state space dimensionality has rapidly evolved. Researchers are continuously exploring the balance between three aspects of RL: optimization, strategy-based approaches, and state space dimensionality, aiming to achieve a "Learning Equilibrium." Explainable AI (XAI) is an important area of research in RL, focusing on ethics, trust, transparency, and safety. Researchers are developing explanations and advanced visualization techniques to improve the interpretability and transparency of RL models. Despite the successes of RL, there are challenges that require attention, including scalability, sample efficiency, generalization, interpretability, and the integration of RL with other traditional AI approaches. Theoretical advancements are needed to enhance our understanding of neural networks and their properties within the context of RL. Overall, RL has significant potential for commercial success and can address complex real-world problems. Ongoing research and advancements in RL algorithms, integration with other AI techniques, and the development of XAI methods are key areas of focus to enhance the application and impact of RL.

6 REFERENCES

[1] Anna Goldie and AzaliaMirhoseini Placement Optimization with Deep Reinforcement Learning, 29 March 2020.

[2] Christian Wirth , Riad Akrour , Gerhard Neumann , Johannes Furnkranz ,A Survey of preference – Based Reinforcement Learning Methods .

[3] Deepanshu Mehta , State – of -the -Art Reinforcement Learning Algorithms ,12 December – 2019.

[4] Heyang Qin, Syed Zawad and Yanqi Zhou ,Swift Machine Learning Model Serving Scheduling: A Region Based Reinforcement Learning Approach , 9, November 17–22, 2019.

[5] Kai Arulkumaran , Marc peter Deisenroth , Miles Brundage , Anil Anthony Bharath ,A Brief survey of deep Reinforcement Learning , 28 sep 2017.

[6] Leslie Pack Kaelbing , Michael L.Littman ,Andrew W. Moore . Reinforcement Learning

[7] Lindsay Wells and Tomasz Bednarz , Explainable AI and Reinforcement Learning – A Systematic Review of Current Approaches and Trends , 20 May 2021.

[8] Marc G. Bellemare ,Will Dabney , Remi Munos , A Distributional Perspective on Reinforcement Learning.

[9] Matthias Hutsebaut-Buysse and Kevin Mets and Steven Latré ,Structured Exploration Through Instruction Enhancement for Object Navigation, November 15, 2022.

[10] N.R. Ravishankar and M.V. Vijayakumar . Reinforcement Learning Algorithms : and classifications . January 2017.

[11] Nelson VithayathilVarghese ,Qusay H. Mahmoud , A Survey of Multi- Task Deep Reinforcement Learning ,22 August 2020.

[12] Nuo Xu , Understanding the Reinforcement Learning ,2019.

[13] Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, DoinaPrecup, David Meger Deep Reinforcement Learning that Matters.

[14] Robert N. Boutea,b, JorenGijsbrechts , Willem van Jaarsveldd, Nathalie Vanvuchelena Deep reinforcement learning for inventory control: A roadmap.

[15] Ronald Parr and Stuart Russell, Swift Machine Learning Model Serving Scheduling:A Region Based Reinforcement Learning Approach, 19, November 17–22, 2019.

[16] Thomas Hickiling ,Abdelhafid Zenati , Nabil Aouf , Phillippa Spencer , Explainability in Deep Reinforcement learning , a review into Current Methods and Applications , 7 March 2023.

[17] Yan Duan, John Schulman, Xi Chen, Peter L. Bartlett, Ilya Sutskever, Pieter Abbeel Fast Reinforcement Learning via slow RL , 10 Nov 2016.