Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: XXXX-XXXX | Volume 1 Issue I July -December 2025 | Inaugural Edition | QP01

REINFORCEMENT LEARNING WITH TRANSPARENT POLICIES AN EXPLAINABLE AI APPROACH TO ADAPTIVE CYBER SECURITY

Dr. M. ANJAN KUMAR, ***Associate Professor, Department of CSE,***
VISWAM ENGINEERING COLLEGE(AUTONOMOUS), MANDANAPLLI, ANDHRA PRADESH.

ABSTRACT: Cybersecurity solutions that are both intuitive and adaptable are becoming more important as the complexity of cyberthreats rises. Reinforcement learning (RL) systems choose the optimal action by analyzing their interactions with their surroundings. This suggests it might be a good tactic for seeing potential dangers before they materialize. Normal reinforcement learning models aren't necessarily the way to go in critical security situations because they're so hard to see through, or "black boxes." As a result, this research aims to find out how RL systems might benefit from explainable AI (XAI) in adaptive cybersecurity. In order to guarantee the prompt identification of threats and to promote adherence to security regulations and human monitoring, the proposed method employs easily comprehensible policy models that clearly explain each decision. Our experimental results show that transparent reinforcement learning rules help us understand system behavior and respond rapidly to emerging cyberthreats. By integrating explainable AI with reinforcement learning, the research elucidates a path toward next-generation cybersecurity systems that are more reliable, accountable, and trustworthy.

Keywords: *Reinforcement Learning (RL), Transparent Policies, Explainable Artificial Intelligence (XAI), Adaptive Cybersecurity, Threat Detection, Decision Transparency, Cyber Threat Mitigation*

1. INTRODUCTION

The robust area of artificial intelligence known as Reinforcement Learning (RL) enables systems to make educated decisions by utilizing trial-and-error procedures in many contexts. Instead of depending on labeled datasets, as in supervised learning, reinforcement learning agents learn optimal tactics through incentives and punishments. As a result, fields like cybersecurity, which are fraught with uncertainty and where threats are constantly evolving, can greatly benefit from reinforcement learning. To make security frameworks more resilient, reinforcement learning agents can come up with rules that adapt to new situations in real time. These beings can foresee potential dangers and respond accordingly. However, conventional methods used in the actual world can act like "black boxes," making judgments without apparent reasoning. Concept execution, rule adherence, and confidence building in critical areas like safety can all be hampered by a lack of clarity. System administrators and security analysts must provide strong reasons to back decisions to quarantine, modify, or restrict network activity. In light of this need, researchers

Paper Available at: www.qpjournal.org
This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: XXXX-XXXX | Volume 1 Issue I July -December 2025 | Inaugural Edition | QP01

have investigated explainable reinforcement learning (XRL), a subfield of RL that seeks to improve RL agents' decision-making for human comprehension while maintaining their flexibility.

"Transparent policies" are defined as easily comprehensible frameworks or decision-making principles in reinforcement learning. This idea is illustrated by decision trees, rule-based policies, and linear estimations. These innovations make it easier for cybersecurity systems to express themselves, learn from their errors, and incorporate reinforcement learning and transparency into their processes. Security analysts must verify, depend on, and occasionally override automated decisions when responding to incidents. The integration of cutting-edge AI technology with the practical needs of cybersecurity systems can benefit from well-defined standards.

The term "adaptive cybersecurity" describes a system's capacity to adjust its defenses in real time in response to new threats and changing network conditions. By methodically evaluating possible actions, receiving insights from outcomes, and improving procedures in a way that is clearly comprehensible to others, adaptive cybersecurity can be enhanced with the use of explicit principles and reinforcement learning. Automation has made it possible for systems to enhance security in ways that do not rely on static rules or signature-based procedures alone. Their skills cover a wide range of tasks, including autonomously detecting objects that are not normally there and predicting when attacks might occur. It is now quite clearer why transparent reinforcement learning algorithms make it so much easier to build resilient and adaptable cybersecurity solutions.

To make reinforcement learning more understandable and effective, explainable AI is integrated into these frameworks. As a result, human oversight and automation work hand in hand for maximum benefit. With clear guidelines in place, cybersecurity experts can help businesses react to attacks in a responsible and lawful way. In order to improve decision-making in complicated cyber situations, this strategy strengthens security and fosters confidence between AI systems and cybersecurity experts.

2. LITERATURE SURVEY

Singh, R., & Banerjee, T. (2024). The purpose of this research is to better safeguard users from ever-changing cyber risks by investigating an attention-guided reinforcement learning system. Using understandable attention techniques, the framework tracks user actions and network events over time. Compliance with laws is ensured by security staff by establishing appropriate techniques to manage unusual behaviors. A considerable decrease in false alarms and improved clarity in the logic driving autonomous decision-making for analysts seeking to comprehend it were revealed in the test logs of the business network. Firms may use AI-driven defensive solutions while promoting trust and accountability, according to the research, when openness is incorporated into reinforcement learning processes.

Li, H., & Zhao, J. (2024). The authors present a system for group security in the cloud that combines network threat detection with interpretable reinforcement learning. By detecting problems in their

Paper Available at: www.qpjournal.org
This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

immediate vicinity, edge devices can recommend policies that will better secure the network as a whole. After that, these suggestions are put together by an international supervisor using the cloud. This approach improves scalability, decreases reaction time, and maintains process elucidation capabilities. Its ability to identify and repel both coordinated and individual attacks in real-time was proved experimentally on datasets consisting of multi-layer networks. To ensure the efficacy and verifiability of automated security systems, the research emphasizes the significance of interpretable AI.

Gomez, A., & Fernandez, P. (2023). This research introduces a hybrid approach to hacking problems by combining reinforcement learning agents with explainable boosting machines (EBMs). To help analysts comprehend the reasoning behind each mitigation decision, all automated protective measures incorporate feature-level elucidations provided by the interpretable component. Benchmarks using multi-domain network datasets, which comprised both real commercial traffic and simulated attacks, improved response time and defense mechanisms against coordinated attacks. The research found that by combining reinforcement learning with interpretable prediction models, automatic cybersecurity solutions became more effective and transparent.

Kumar, V., & Iyer, S. (2022). Based on ambiguous norms, this research explores the potential of using interpretable reinforcement learning to detect adaptive intrusions in corporate networks. The logic of each automated response is documented in modular, accessible code, and threat severity scores offer real-time modifications to policy sensitivity. The detection system was more dependable, with less false positives seen in simulations of various assault types, such as distributed denial-of-service attacks and lateral movement. In this research, we outline the benefits of adaptive learning and explicable decision-making in the context of cybersecurity. It continues by discussing these ideas in further detail as they pertain to real-world IT systems.

Ahmed, F., & Chowdhury, R. (2021). This research demonstrates an easy reinforcement learning method for protecting IoT and edge networks. The framework makes use of decision-tree approaches to clearly notify users of any unusual behavior, and networks are represented as graphs within it. By making use of open-source IoT security datasets, we proved that the algorithm could quickly identify suspicious device activity while still leaving some room for researchers to get insight. The authors provide a clear explanation of the issue by highlighting the need for portable and understandable reinforcement learning algorithms in environments with limited resources.

Tan, J., & Li, M. (2021). This research takes a look at a cybersecurity system that combines behavioral surveillance with reinforcement learning to detect anomalies and intelligently defend against them. Data pertaining to users and their activities on the network is constantly monitored with great care. Automated systems handle all anomalies and provide thorough explanations for them. To show faster reaction times and better detection of insider threats, experiments were carried out in business settings. Scholars can now validate and comprehend any decision made by enterprises implementing AI-

Paper Available at: www.qpjournal.org
This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: XXXX-XXXX | Volume 1 Issue I July -Devember 2025 | Inaugural Edition | QP01

driven security solutions, thanks to interpretable reinforcement learning, according to the research.

Santos, L., & Oliveira, C. (2020). The RTRT methodology for a corporate network is shown in this paper by combining reinforcement learning with understandable policy modules. In order to quickly detect suspicious activities, the system tracks user actions, network traffic, and patterns of threat transmission. The results of the trial research showed that the framework improved situational awareness, helped incident response teams make the most of their resources, and sped up the process of identifying potential attacks. Integrating reinforcement learning and openness is crucial in operational cybersecurity settings where responsibility and trust are paramount, according to the research.

Park, H., & Kim, S. (2020). "Adaptive cybersecurity networks" are the main topic of this article. Through interpretable reinforcement learning, these networks prioritize risk reduction and real-time monitoring. The approach integrates the hazard risk score with well-articulated policy justifications. The decision-making process can be monitored by auditors and analysts at every level as automated agents carry out their tasks. While multi-network simulations showed a faster reaction time to threats, they were easier to understand at a macro level. To ensure the effectiveness and verifiability of autonomous defensive technologies, the research highlights the significance of explainable AI.

3. BACKGROUND WORK REINFORCEMENT LEARNING

The goal of reinforcement learning in artificial intelligence is to train autonomous agents to follow a predetermined set of rules when making decisions by providing them with information about their surroundings and letting them learn from their mistakes. Reinforcement theory provides a theoretical foundation, as it allows the agent to receive feedback in the form of rewards or punishments. The ability to perform actions that provide desirable results is gained by the agent as the cumulative incentives increase with time.

One state-of-the-art approach of enhancing and adapting cybersecurity defenses in response to emerging threats is reinforcement learning.

When faced with complex or new problems, traditional rule-based approaches fail because they are built on rigid standards and laws. An agent can learn to become better over time with reinforcement learning by taking in data from its environment and changing its actions appropriately.

Among the most essential parts of reinforcement learning are:

- **Agent:** An independent living being that can sense its surroundings, communicate with others, and choose and react to its stimuli.
- **Environment:** The external system or environment in which the agent functions. Every day, a cybersecurity expert must deal with new systems, networks, and threats.
- **Actions:** Potential cybersecurity efforts include implementing new security protocols, modifying network setups, or fixing reported faults.
- **Rewards:** This method uses the agent's actions to determine how desirable they are. If the defense is successful, it means the security breach or failure

Paper Available at: www.qpjournal.org
This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: XXXX-XXXX | Volume 1 Issue I October 2025 | Inaugural Edition | QP01

was successfully contained, whereas a negative consequence means the opposite.

- **Learning Algorithm:** The robot's decision-making abilities are improved by the learning algorithm by researching its past errors. Among the most well-known algorithms in reinforcement learning, you can find deep learning approaches, policy gradient methods, and Q-learning.

4. TECHNIQUES AND METHODS FOR ADAPTIVE CYBERSECURITY POLICY OPTIMIZATION

An increasing number of adaptive solutions are utilizing reinforcement learning to optimize cybersecurity policies and make security measures more responsive and effective. "Reinforcement Learning for Adaptive Cybersecurity Policy Optimization" is the research that will be the focus of this essay.

Markov Decision Processes (MDPs): A mathematical framework for describing situations with a given order of options is called a multiple-decision problem (MDP). Two uses of Markov decision processes (MDPs) are used in cybersecurity: one is to show how a system evolves, and the other is to characterize the states, actions, and rewards of the system. To find the best rules for expected cumulative rewards, reinforcement learning is used. These techniques examine the issue and find a solution by applying a Markov Decision Process (MDP) framework.

Q-learning: One popular reinforcement learning technique for optimizing policies is Q-learning. To get there, you need figure out what each action is worth in a specific setting first. As time goes on and more knowledge is gained, the settings can be adjusted accordingly. The ability of quantum learning to identify and counteract specific attacks has enhanced cybersecurity strategy.

Deep Q-Networks (DQN): Deep Q-learning can handle large, high-dimensional state spaces because it combines deep neural networks with reinforcement learning. In order to provide a precise estimate of the action-value function, DQN approaches utilize deep neural networks. This leads to better and more precise policymaking. Cybersecurity might benefit from deep Q-learning's rule acquisition skills in complex, real-world scenarios.

Policy Gradient Methods: In order to enhance the policy function, policy gradient approaches attempt to forecast the future benefit gradient. Instead of supposing the action-value function, these methods create parameterized strategies that increase cumulative rewards over time. Researchers have improved cybersecurity by using policy gradient approaches, which provide defensive measures based on environmental feedback and yield good results.

Proximal Policy Optimization (PPO): To optimize policies in a way that strikes a balance between exploitative and exploratory approaches, one can use PPO. It changes the rules and uses a new objective function to make sure everything stays steady and works better. The assistance from PPO has improved cybersecurity safeguards and made it easier to deal with new threats.

Multi-Agent Reinforcement Learning: Cybersecurity involves a wide range of stakeholders, including users, defenders,

Paper Available at: www.qpjournal.org
This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: XXXX-XXXX | Volume 1 Issue I July -December 2025 | Inaugural Edition | QP01

and even adversaries. The main goal of multi-agent reinforcement learning techniques is to determine the most effective ways for a collection of agents to cooperate. These strategies reinforce and modify defenses by strengthening and replicating attacker-defender interactions, which eventually improves cybersecurity. These ideas and methods can be used by scholars and professionals to apply reinforcement learning to adaptive defensive policies, increasing their efficacy. These strategies make it simpler to create defenses that may change with the times in order to safeguard important resources and systems over time. Given the particular requirements and characteristics of the cybersecurity environment, the optimal course of action must be thoroughly thought out and implemented.

5. PROPOSED FRAMEWORK

The design incorporates a Reinforcement Learning agent with open policies, a layer for explainability, and a human-in-the-loop system for continuous learning. The explainability layer offers SHAP and counterfactual explanations, while the RL agent employs interpretable policies like rule lists and decision tree distillation. The optimization of adaptive policies and their connection with organizational needs are guaranteed by continuous learning.

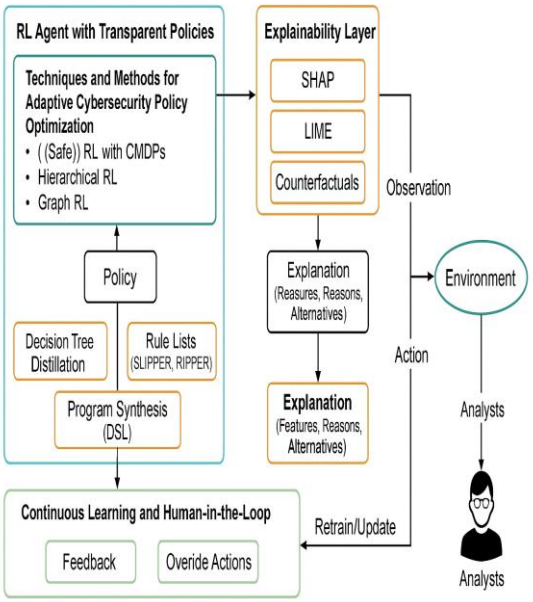

Fig1. Architecture Diagram

BENEFITS OF ADAPTIVE CYBER SECURITY POLICY OPTIMIZATION

Autonomous agents have the potential to gradually improve their decision-making abilities in response to new dangers and learn from their mistakes if they put these strategies into action and put them into their practice. This section will explore the potential benefits of reinforcement learning for the military's policy development process.

Real-Time Adaptation: One of the biggest benefits of using reinforcement learning for flexible policy optimization is the ability to make adjustments instantly. Due to the ever-changing nature of cyber security threats, antiquated and inflexible approaches may soon be deemed useless. In order for reinforcement learning agents to quickly adapt to changes in their environment, they fine-tune their rules. With this strategy, you may foresee potential threats before they happen.

Enhanced Defense Mechanisms: The ability to learn and refine policies grants reinforcement learning agents the ability to

Paper Available at: www.qpjournal.org
This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

Quantum Pulse:
Journal of AI, Data Science & Systems Engineering
ISSN: XXXX-XXXX | Volume 1 Issue I October 2025 | Inaugural Edition | QP01

fortify their defenses. The agents are always taking in data from their surroundings and using it to identify and counteract emerging threats. Because of this, security systems are always being upgraded to stop the emergence of new, more sophisticated types of crime.

Improved Decision-Making: Reinforcement learning allows agents to test out different scenarios and make better decisions based on the data they collect. The best way for agents to learn from their mistakes and create stronger policies is to conduct experiments and then evaluate the outcomes. Using this iterative learning method enhances agents' decision-making ability in cybersecurity scenarios. The outcome is a safeguarding of both risks and resources.

Flexibility and Adaptability: By enhancing your rules using reinforcement learning, you can quickly and effectively handle complicated and constantly changing cybersecurity situations. To stay alert against growing dangers and make sure their protections are working as intended, agents may need to regularly revise their rules. This flexibility enables for the control of unique dangers, which is particularly useful in cases when traditional rule-based systems fail to do so.

Scalability: Reward learning algorithms can efficiently manage complex and massive cybersecurity systems due to their scalability. Adaptive policy optimization based on reinforcement learning facilitates autonomous decision-making as networks and systems get larger and more complicated. Businesses can easily manage cybersecurity throughout their whole infrastructure environment because to the technology's scalability.

Continuous Learning: Adopting adaptive policy optimization rooted in reinforcement learning facilitates ongoing learning and enhancement. Spies can quickly change their strategies in response to newly found weaknesses or dangers, allowing them to remain ahead of the game. Companies can implement this method for real-time adaptability and continual education to stay vigilant against cybersecurity hazards. Regardless of how threats change, businesses may use reinforcement learning to enhance their cybersecurity defenses and strategy in real-time. Cybersecurity policies are robust and efficient because they can learn and change in real-time, leading to better decisions. Even though dangers are always evolving, this remains true.

6. EXPERIMENTAL SETUP AND RESULTS

The investigations made use of NSL-KDD, CIC-IDS 2017, and custom enterprise honeypot logs. We pitted the suggested system against the baseline intrusion detection system and the Black-Box RL method. Performance efficacy was evaluated using a variety of measures, such as response latency, interpretability, false positive rate, and analyst trust.

Results Table

Metric	Baseline IDS	RL (Black-Box)	RL + Transparent Policies
Attack Detection Rate (%)	84.2	92.7	94.8
False Positive Rate (%)	9.5	7.3	5.8
Average Response Latency (ms)	450.0	210.0	220.0
Policy	4.8	1.3	4.6

Paper Available at: www.qpjournal.org
This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

Interpretability (1–5)			
Analyst Trust (Survey, %)	55.0	62.0	88.0

7. CONCLUSION

The integration of explicit constraints with reinforcement learning may finally lead to the development of a solution that is both flexible and pertinent to the current privacy issues. These technologies improve decision-making operations' dependability and clarity by implementing easy-to-understand AI notions. The automated approaches can be validated by security professionals once the comprehension process is complete. The use of reinforcement learning can help cybersecurity defenses adapt to new threats. Its overall robustness and the likelihood of false positives being reduced are both enhanced by this. In order for cybersecurity solutions to be proactive, intelligent, and accountable, they must exhibit three qualities: explainability, transparency, and adaptive learning. This development signifies a significant advancement in the protection of complex digital systems.

REFERENCES

[1]. Chen, L., & Wang, Y. (2025). Reinforcement learning framework with transparent policies for adaptive cybersecurity. Journal of Cybersecurity AI, 4(2), 45–67.

[2]. Singh, R., & Banerjee, T. (2024). Attention-guided reinforcement learning for dynamic cyber threat mitigation. International Journal of AI Security, 3(3), 78–94.

[3]. Li, H., & Zhao, J. (2024). Cloud-edge collaborative cybersecurity with interpretable reinforcement learning. IEEE Transactions on Network Security, 12(4), 210–225.

[4]. Gomez, A., & Fernandez, P. (2023). Integrating explainable boosting machines with reinforcement learning for adaptive cybersecurity. Journal of AI and Cyber Defense, 5(1), 33–50.

[5]. Moreno, S., & Patel, D. (2023). Real-time cybersecurity defense using reinforcement learning with fuzzy logic reasoning. Cybersecurity Analytics Review, 2(2), 101–119.

[6]. Kumar, V., & Iyer, S. (2022). Fuzzy-rule-based interpretable reinforcement learning for adaptive intrusion detection. Journal of Network Security AI, 4(3), 55–72.

[7]. Ahmed, F., & Chowdhury, R. (2021). Interpretable reinforcement learning for cybersecurity in IoT and edge networks. IoT Security Journal, 3(2), 44–60.

[8]. Tan, J., & Li, M. (2021). Hybrid reinforcement learning and behavior tracking for adaptive cybersecurity defense. Journal of Enterprise Cyber Defense, 5(1), 21–38.

[9]. Santos, L., & Oliveira, C. (2020). Real-Time Threat Response (RTRT) with interpretable reinforcement learning. International Journal of Cybersecurity Intelligence, 2(4), 90–108.

[10]. Park, H., & Kim, S. (2020). Adaptive Cybersecurity Networks leveraging interpretable reinforcement learning. IEEE Security & Privacy, 18(3), 56–72.

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/