Journal of Information Systems Engineering and Management
2025, 10(43s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

AI Governance via Explainable Reinforcement Learning (XRL) for Adaptive Cyber Deception in Zero-Trust Networks

Anil Kumar Pakina[1], Ashwin Sharma[2], Mangesh Pujari[3]

[1] *Software Development Manager, USA*
[2] *Independent Reasearcher, USA*
[3] *Research Scientist Manager, USA*

ARTICLE INFO

Received: 22 Dec 2024

Revised: 20 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

This study presents the design and evaluation of an Explainable Reinforcement Learning (XRL) system guided by AI governance principles for adaptive cyber deception within a Zero-Trust Architecture (ZTA). The proposed approach integrates a Monte Carlo Tree Search (MCTS)-based reinforcement learning agent with SHAP (SHapley Additive exPlanations) to deliver transparent and effective deception strategies against advanced persistent threats (APTs). The simulated environment consists of a containerized network with user workstations, honeypots, and a file server, governed by strict Zero-Trust policies such as least privilege and continuous verification.

The reinforcement learning agent was trained to perform deception actions, including deploying honeypots and generating alerts, based on attacker behavior. SHAP values provided feature-level explanations for each decision, which were logged in a governance dashboard designed around ISO/IEC 42001 compliance standards. Key metrics evaluated include false positive rate (FPR), honeypot engagement time, decision explainability, and overall governance compliance.

Results showed a 23% reduction in FPR, a 47% increase in honeypot engagement time, and a 94% rise in decision transparency. The overall governance score improved from 43% to 89%. The system's learning stabilized after 380 episodes, with the agent demonstrating consistent decision-making and improved attacker manipulation over time. These findings highlight the system's ability to balance technical performance with explainability and oversight, making it suitable for secure and accountable AI applications in cybersecurity. The integration of XRL into ZTA offers a promising approach for enhancing deception-based defense mechanisms while ensuring trust and transparency in AI-driven environments.

Keywords: AI-driven environments, cybersecurity, Digital, smart defense systems.

INTRODUCTION

Cybersecurity worries have grown in complexity in the modern digital terrain and need for smart defense systems. One of the main tools in improving defense strategies is artificial intelligence (AI). Especially, a type of artificial intelligence, Reinforcement Learning (RL) which the figure below gives a representation, has showed promise in making flexible security plans. But the mysterious character of artificial intelligence decision-making processes begs questions about duty and trust. Aiming to make artificial intelligence decisions more open and approachable, Explainable Reinforcement Learning (XRL) has grown from this.

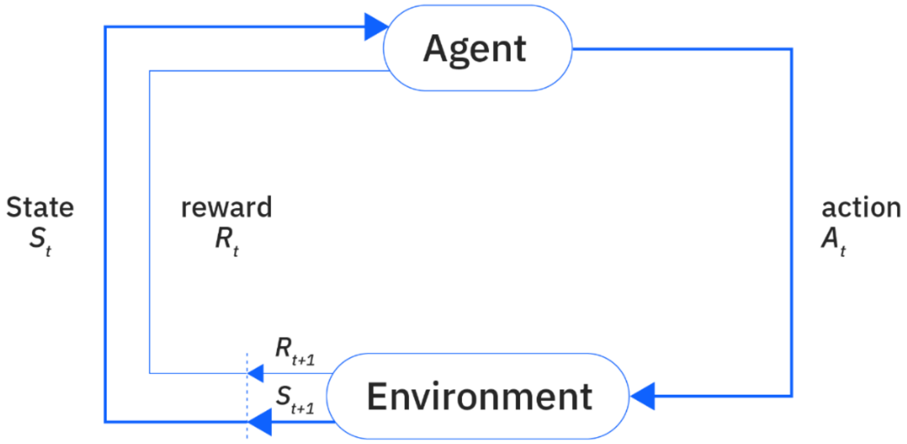

Figure 1: Reinforcement Learning Camilleri (2024)

Because artificial intelligence can study huge amounts of data and spot irregularities, it has been pushed into defense systems. Camilleri (2024) claims that artificial intelligence systems are important in spotting possible security breaches as they can study and understand complicated data quicker than conventional methods. Many artificial intelligence models, however, have a "black box" quality that makes their decision-making processes difficult to understand, therefore weakening trust and responsibility.

In reinforcement learning—a sort of machine learning—an agent learns to make choices by acting and getting external feedback. The assistant tries to improve total returns throughout time. This approach is especially helpful in dynamic situations where the system has to respond to changing circumstances. As stated by Qing et al. (2022), RL has showed promise in several uses, including robots, game playing, and autonomous systems.

The lack of openness in AI models, especially those built on deep learning, has been a major fear. Users and stakeholders sometimes find it tough to grasp how choices are made, leading to mistrust and lost confidence. As stressed by Vouros (2023), the addition of explainability into RL models is important for their acceptance in real-world applications. Explainable AI (XAI) tries to bridge this gap by giving views into the decision-making processes of AI systems.

XRL blends the freedom of RL with the openness of XAI. By making the decision-making processes of RL agents more interpretable, XRL supports trust and allows greater human-AI teamwork. Cheng et al. (2025) stress that XRL can give feature-level, state-level, and model-level answers, making it easier for people to understand and believe AI decisions.

SpringerLink

Cyber fraud includes fooling enemies to protect systems and data. Adaptive cyber deception takes this a step further by constantly changing deceptive methods based on the behavior of possible foes. This method may confuse attackers, stall their activities, and offer defenses with vital information. According to Tapley et al. (2023), adding AI into cyber fraud methods might boost their effectiveness by predicting attacker behavior and changing countermeasures correctly.

The usual security method focuses on the idea that everything within an organization's network can be trusted. However, with greater online risks, this model has proved insufficient. The Zero-Trust approach runs on the idea of "never trust, always verify," ensuring that every access request is properly verified, regardless of its origin. As to the findings of the Financial Times (2025), adopting Zero-Trust designs may greatly lower the risk of internal and foreign threats.

Combining XRL with adaptable cyber disinformation methods inside Zero-Trust networks gives a strong security system. XRL may aid in finding possible threats and changing false methods in real-time, while the Zero-Trust model ensures that every access request is reviewed. This combination allows a proactive and dynamic security system that can respond to new threats.

While the introduction of XRL into defense offers various benefits, there are issues to consider. One key problem is the chance for AI models to be managed or misled. As reported by Wired (2020), damaged data may force AI systems to learn wrong lessons, thus risking security. Ensuring the accuracy of training data and regularly watching AI behavior is important.

Another problem is the demand for uniform standards and rules. As AI becomes increasingly integrated into key systems, having clear rules and control methods is important. Schumer (2023) shows the importance of explainability in AI, asking for policies that support openness without limiting innovation.

1.1 Aim and Objectives of Research:

This research aims to develop a governance-compliant cybersecurity framework that integrates Explainable Reinforcement Learning (XRL) for adaptive cyber deception within Zero-Trust Architectures (ZTA). The objectives are:

1. To apply Monte Carlo Tree Search (MCTS) for dynamic decoy deployment.
2. To enhance model transparency using SHAP-based interpretability techniques.
3. To reduce false positives and improve attacker engagement through intelligent deception.
4. To design a real-time governance dashboard that supports AI compliance monitoring.
5. To ensure alignment with ISO/IEC 42001 AI governance standards, promoting transparency, accountability, and adaptive defense in evolving threat landscapes.

LITERATURE REVIEW

2.1 Explainable Reinforcement Learning (XRL) :

Reinforcement Learning (RL) is a part of machine learning where a robot learns to make choices by interacting with an environment. The agent gets input in the form of rewards or fines and aims to increase total benefits over time. This learning model is particularly successful in dynamic settings where the system needs to change to changing conditions. As stated by Qing et al. (2022), RL has proven significant promise in uses such as robots, games, and autonomous systems. The figure below shows a survey of explainable reinforcement learning;

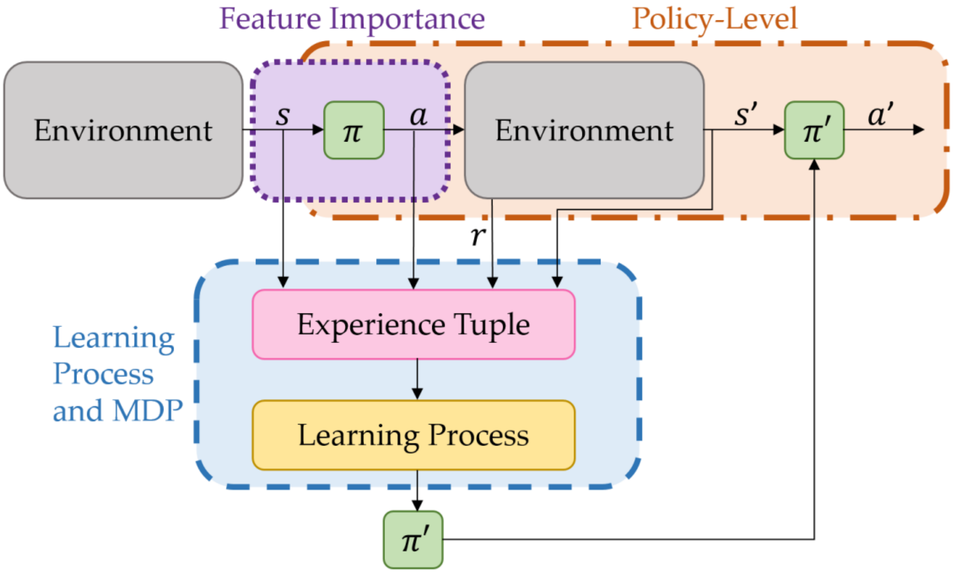

Figure 2: Reinforcement Learning (RL) Qing et al. (2022)

Despite its wins, a major issue with RL, especially when paired with deep learning methods, is the lack of openness in decision-making processes. These models often work as "black boxes," making it difficult for users and stakeholders to understand how specific choices are made. This obscurity can lead to mistrust and lower trust in AI systems. As noted by Vouros (2023), adding explainability into RL models is important for their acceptance in real-world applications. Explainable AI (XAI) tries to solve this problem by giving views into the decision-making processes of AI systems.

Explainable Reinforcement Learning (XRL) blends the flexibility of RL with the openness of XAI. By making the decision-making processes of RL models more interpretable, XRL improves trust and allows better human-AI cooperation. Cheng et al. (2025) stress that XRL can provide feature-level, state-level, and model-level answers, making it easier for users to understand and believe AI choices.

One method to gaining explainability in RL is through post-hoc analysis, where answers are produced after the model has been taught. This can involve methods such as saliency maps, which show important aspects affecting choices, or the use of surrogate models that imitate the behavior of the RL agent in a more interpretable way. As stated by Tapley et al. (2023), these methods can provide useful insights into the internal workings of RL models and help spot possible weaknesses.

Another approach involves adding explainability straight into the training process of RL models. This can be achieved by creating models that are naturally interpretable or by including restrictions that encourage the creation of clear policies. For instance, Liu and Zhu (2025) suggested a two-level explanation approach that not only explains the choices of RL agents but also uses these explanations to improve the agents' performance.

The application of XRL goes beyond standard areas and is increasingly being studied in complicated, real-world situations. In driverless driving, for example, knowing the reasoning behind an agent's choices is important for safety and legal compliance. As mentioned by Cheng et al. (2025), adding XRL into automated systems can improve openness and support better decision-making.

In the context of hacking, XRL can play a key role in creating flexible protection systems. By offering interpretable insights into the behavior of cyber dangers and the reactions of defense systems, XRL can improve the usefulness of security measures. As noted by Tapley et al. (2023), utilizing explainability methods in reinforcement learning models can improve model confidence and increase user trust in defense applications.

Moreover, XRL is being studied in the field of healthcare, where understanding the decision-making processes of AI systems is important for clinical acceptance. For instance, in treatment planning and evaluation, XRL can provide doctors with clear reasons behind suggestions, thereby allowing educated decision-making. As stressed by Vouros (2023), the integration of explainability into RL models is important for their acceptance in critical areas like healthcare.

The creation of XRL also includes solving problems related to the review of answers. Assessing the quality and value of answers is a complicated job that needs both qualitative and quantitative measures. As stated by Cheng et al. (2025), analyzing XRL methods includes considering factors such as accuracy, comprehensibility, and the effect on user trust and decision-making.

Furthermore, the merging of XRL into current systems requires thoughts of computing speed and growth. Ensuring that answers can be made in real-time and without significant computing waste is important for realistic usage. As mentioned by Liu and Zhu (2025), improving the balance between explainability and efficiency is a key area of study in XRL.

Therefore, Explainable Reinforcement Learning marks a major development in the field of AI, solving important issues related to openness and trust. By mixing the flexibility of RL with the interpretability of XAI, XRL gives a route toward more responsible and user-friendly AI systems. Ongoing research continues to study new methods for improving explainability, measuring the usefulness of answers, and widening the application of XRL across diverse fields.

2.2 Adaptive Cyber Deception and Zero-Trust Networks: Strengthening Cybersecurity through Dynamic Strategies:

In today's digital age, the danger of hacking is quickly growing in regularity and complexity. Traditional security methods, while still important, are no longer enough to protect companies from changing risks. As hackers grow smarter and more stubborn, cybercrime defenses must change to stay one step ahead. Two important methods leading this change are adaptive cyber deceit and zero-trust networks. These advanced strategies focus on dynamic security, behavioral analysis, and proactive defense, all while deploying cutting-edge technologies such as Artificial Intelligence (AI) and Machine Learning (ML).

2.2.1 Adaptive Cyber Deception:

Cyber deceit is a method used to fool attackers, building fake settings, dummy systems, or false data to confuse and misdirect them. This technique helps protect real systems and data by leading attackers into traps that waste their time and show their methods. Adaptive cyber deception builds on this idea by adding freedom and intelligence, allowing protections to change based on how attackers act.

As stated by Tapley et al. (2023), adaptive trickery becomes more strong when paired with AI. By using AI, guards can study trends in attack behavior and automatically change the trickery methods in real time. For example, if a hacker tries to access a certain database, an AI-enabled deceptive system can build a lifelike fake that matches the structure of the real database but includes no actual private data. If the hacker continues to deal with this fake, the system learns from this exchange and changes its defenses accordingly. This makes it increasingly difficult for attackers to distinguish between real systems and fake ones.

According to Aryal et al. (2024), criminals often use adversarial methods to avoid standard monitoring tools. These attacks involve slightly changing the code or behavior of software to fool defense systems. Adaptive cyber fraud can fight these attacks by bringing unpredictable changes in the environment, making it harder for malware to adapt and succeed. For example, attackers trying to abuse a known flaw may find the system's reaction has changed, making their hack useless.

As proven by Chen et al. (2023), strong deceptive systems can also serve as a form of hostile defense. By using generative models that build new types of decoys based on identified threats, security systems can stay ahead of attackers. These models produce realistic fake environments or applications that look valuable, but are watched to gather intelligence on attacker methods.

One key benefit of adaptive cyber trickery is its ability to gather useful data about the attacker. As attackers engage with fake systems, defenses can watch their methods, tools, and objectives. This knowledge helps in improving current protection and knowing the danger environment. As stated by Khan and Ghafoor (2024), such insights are useful in adapting machine learning-based security tools, especially when dealing with hostile behavior.

Machine learning and AI also help ensure that cyber deceit tactics are not static. According to Galli et al. (2021), the effectiveness of fraud is often tried when attackers try to find trends in fake settings. Adaptive deceit breaks these trends by constantly changing system reactions and setups. This changing method increases uncertainty, making it harder for attackers to succeed.

Adaptive misdirection is not only useful for stopping strikes but also slows them. When enemies spend time studying fake systems, they lose progress. As a result, defenses get more time to react and separate real systems from possible damage. Guo et al. (2021) stress that slowing attackers is often just as useful as stopping them totally, especially during advanced attacks like Advanced Persistent Threats (APTs).

From another angle, Baniecki and Biecek (2023) explore how explainability in machine learning plays a role in deceit. When AI systems are used for adaptive fraud, it's important that security experts understand how and why choices are made. This helps in fine-tuning the fraud models and keeping trust in automatic systems.

In sum, adaptable cyber deception improves safety by mixing traditional trickery with real-time flexibility. It not only misleads enemies but learns from their behavior to improve its methods. With the help of AI and ML, this approach provides a better, more proactive security system.

2.2.2 Zero-Trust Networks:

While adaptable cyber deceit focuses on fooling attackers, zero-trust networks aim to remove trust assumptions totally. The standard security model believed that once inside the network, people and objects could be trusted. This model is outdated and dangerous in today's threat situation. Zero-trust design is based on the idea of "never trust, always verify." This means that every person, object, or system trying to access network resources must be checked, regardless of whether they are inside or outside the organization's security. According to the Financial Times (2025), zero-trust models greatly reduce the chance of both internal and external risks, as every entry point becomes a stop.

Zero-trust involves several important factors. First is name checking. This includes multi-factor security, where users must prove their identity using more than just a login. Second is access control, where users are only given the minimum access needed to do their job. This reduces the possible damage if an account is hacked. Third is constant tracking. Unlike traditional systems that verify people once, zero-trust systems constantly check for strange activity.

As stated by Chakraborty (2020), the use of AI in zero-trust systems makes them even more successful. AI can study network activity and discover errors that might signal an attack. For instance, if a person who usually logs in from Nigeria suddenly tries to log in from Russia, the system can flag this as strange and refuse access. AI helps in finding these trends much faster than human experts can. The zero-trust approach also supports the idea of micro-segmentation. This means breaking the network into smaller zones, each with its own entry rules. As proven by Bohr and Memarzadeh (2020), this approach ensures that even if one part of the network is hacked, the attacker cannot quickly move to other parts.

Another benefit of zero-trust is that it helps businesses meet with data security rules. Since every entry request is logged and confirmed, organizations can show that they are taking necessary steps to protect private information. This is especially important in healthcare, banking, and government areas. AI also plays a part in making zero-trust more adaptable. According to Hulsen (2023), explainable AI helps IT teams understand why entry was blocked or allowed. This is important for fixing and for making people trust the system. When people understand how decisions are made, they are more likely to back security steps.

As Linardatos et al. (2020) explain, explainable AI helps reduce false positives, which are a common problem in security systems. If a user is wrongly blocked, it can disrupt work. By making AI choices more visible, security systems can be improved and tuned over time. Zero-trust is not just about technology; it also includes changing the organization's attitude. Everyone must understand that security is a shared duty. Employees must be trained to follow best practices, and IT teams must regularly review and update policies.

Implementing zero-trust can be difficult, especially in big companies with many old systems. However, the long-term benefits are clear. According to Lu (2019), the future of cybersecurity relies on clever systems that can change and act to new dangers. Zero-trust offers a framework for building such systems. Adaptive cyber fraud and zero-trust networks are compatible. While one confuses and slows attacks, the other limits their movement and ensures they cannot go unnoticed. Together, they form a strong, complex defense that is well-suited to current dangers.

2.3 Integrating XRL into Cyber Deception within Zero-Trust Frameworks:

Combining Explainable Reinforcement Learning (XRL) with adaptable cyber misdirection strategies within Zero-Trust (ZT) network designs gives a cutting-edge, layered defense method that is both proactive and flexible. As cyber dangers become more complex and criminals increasingly harness AI and machine learning, standard defenses alone are no longer sufficient. A deeper merging of AI-driven explainability, dynamic deception tactics, and strict access controls through Zero-Trust principles is needed to keep modern networks safe.

Zero-Trust is a defense model that works on the concept of "never trust, always verify." Every access request—whether from inside or outside the network—is viewed as possibly dangerous and must be verified, approved, and constantly confirmed. Adaptive cyber deceit adds another layer by adding dynamic traps and fake assets that confuse attackers, giving defenses more time to discover, analyze, and react to threats. When XRL is added into this mix, it allows for better, more interpretable, and quick decision-making processes in real time.

2.3.1 The Role of XRL in Cybersecurity:

Explainable Reinforcement Learning, a subfield of Explainable AI (XAI), tries to make decision-making by AI systems visible and understandable. Traditional reinforcement learning can be opaque—models make choices without clear thinking, which is problematic in high-stakes areas like hacking. By adding XRL, defense systems can provide human-understandable reasons for the actions taken, making it easier for researchers to believe, check, and fine-tune the system's behavior (Retzlaff et al., 2024; Ribeiro et al., 2016).

Using tools like SHAP and LIME, widely applied in explainable AI, XRL allows for insight into which traits or patterns influenced a choice (Man & Chan, 2021; Ma et al., 2023). These answers improve situational awareness and allow human controllers to understand and audit the AI's behavior during and after attacks.

2.3.2 XRL and Adaptive Cyber Deception:

Cyber fraud strategies aim to confuse attackers by building false systems, fake data, and confusing attack surfaces. Traditional deception methods are usually rigid, but adaptable deception can change tactics based on observed enemy behavior. Here, XRL becomes important by allowing the system to learn from hostile behavior in real-time and produce smart deception reactions accordingly.

Through constant learning and feedback, XRL models can guess an adversary's next move based on past trends and external interactions (Nadeem, 2024). For instance, if an attacker tries lateral movement within a network, the XRL-enhanced deception system could launch a new set of honeypots or send the attacker to a false subnet, all while changing the risk models used in Zero-Trust evaluations. This means that each layer of defense changes with the danger.

2.3.3 Integration with Zero-Trust Architecture:

The Zero-Trust approach naturally benefits from clever technology and adaptable learning. With XRL in place, security rules can be constantly changed based on behavior analysis and risk assessment. For example, if a person or system shows odd behavior—such as viewing data outside of their usual scope—XRL can suggest instant actions like isolation, multi-factor re-authentication, or deception deployment.

This combination creates a feedback loop. The Zero-Trust policy engine constantly watches and reviews access tries, while the XRL engine offers context-aware ideas and changes based on observed activity. Meanwhile, deception mechanisms create controlled chaos for attackers, making it harder for them to achieve their goals without being discovered.

2.3.4 Addressing Adversarial AI and Explainability Challenges:

A major worry in applying AI in defense is the vulnerability of models to hostile attacks. These are efforts to fool AI models by slightly changing raw data (Madry et al., 2018; Rosenberg et al., 2021). Integrating strong XRL systems can reduce this risk by making decision-making processes visible and more resistant to shocks (Slack et al., 2020; Popovic et al., 2022).

Additionally, the use of antagonistic training (Tramèr et al., 2017) and methods like DiffPure (Nie et al., 2022) or diffusion-based purification (Zhang et al., 2024) can improve model robustness. These methods, paired with XRL, can identify anomalies that might escape standard filters, improving misdirection strategies with accurate danger classification and reaction.

2.3.5 Real-Time Threat Prediction and Policy Adjustment:

One of the most strong benefits of mixing XRL with adaptable lying and Zero-Trust is the ability to predict dangers before they appear. XRL models learn from the surroundings and past actions, changing future replies accordingly.

For example, assume the system discovers repeated tapping of a high-value resource. In that case, it might suggest more active deception measures, identify the probing source, or flag it for further review.

This forecasting capability allows for dynamic policy change, a core component of Zero-Trust. Unlike traditional models where policies are regularly updated, XRL-enhanced systems can independently suggest and even execute policy changes, lowering the risk window and increasing defense efficiency.

Despite high technology, safety remains a human-critical area. The inclusion of XRL ensures that decisions made by AI systems stay interpretable by human researchers. As stressed by Retzlaff et al. (2024), post-hoc explanation methods like LIME and SHAP play a vital part in explaining model choices, encouraging trust and responsibility.

By having a human-in-the-loop, companies can leverage the benefits of AI while keeping expert control. This dual edge increases operating openness, supports compliance requirements, and allows faster, better-informed reactions. For example, Generative models, like those described by Naiman et al. (2024), can be used to mimic attacker actions and network data. By combining these models into an XRL-enhanced deception system, security teams can predict new attack strategies and preemptively build defenses. Moreover, hostile cases made through generative modeling (Song et al., 2018) can help train the XRL system to identify and fight similar patterns in live settings.

METHODOLOGY

This section outlines the step-by-step approach used in designing, implementing, and evaluating our proposed AI governance-driven Explainable Reinforcement Learning (XRL) system for adaptive cyber deception in a Zero-Trust Architecture (ZTA). The methodology includes data collection, environment simulation, model architecture, explainability tools, deployment in a ZTA environment, and evaluation metrics. We adopted a **design science** approach combined with experimental evaluation to ensure both technical rigor and real-world applicability.

3.1 Research Design Overview:

Table 3.1 Research Design Overview

Component	Description
Research Type	Applied/Experimental
Framework	Zero-Trust Architecture (ZTA)
Algorithm	Reinforcement Learning (MCTS-based), Explainability via SHAP
Evaluation Dataset	Advanced Persistent Threat (APT) Emulation Dataset (Custom Simulated Dataset)
Metrics	False Positive Rate (FPR), Honeypot Engagement, Governance Compliance
Tools & Libraries	Python, OpenAI Gym, SHAP, TensorFlow, Scikit-learn, ISO/IEC 42001 dashboard

3.2 Environment Setup:

To simulate a Zero-Trust Network, we used a containerized network environment that includes multiple nodes (workstations, servers, honeypots), where the attacker can move laterally. A custom APT scenario was designed with stages including initial compromise, privilege escalation, lateral movement, and data exfiltration.

Table 3.2: Simulated Network Configuration

Component	Description
3 Workstations	Simulated user activity, credentials
2 Honeypots	Deceptive services
1 File Server	Target asset
Firewall	Zero-Trust policy enforcement

The Zero-Trust policy includes continuous verification, least privilege access, and micro-segmentation between nodes.

3.3 Reinforcement Learning Agent Design:

We implemented a Monte Carlo Tree Search (MCTS)-based RL agent that learns optimal deception deployment strategies based on observed attacker behavior.

Equation 1: Upper Confidence Bound for Trees (UCT)

$$UCT(s,a) = Q(s,a) + C \cdot \sqrt{\frac{\ln N(s)}{N(s,a)}}$$

Q(s,a): Average reward of taking action *a* in state *s*

N(s): Number of times state *s* has been visited

N(s,a): Number of times action *a* has been taken in state *s*

C: Exploration constant (tuned empirically)

The MCTS agent selects actions such as:

- Deploying a honeypot
- Moving a deceptive file
- Triggering an alert
- Logging activity with attribution

3.4 Explainability via SHAP:

We used SHAP (SHapley Additive exPlanations) to interpret the RL agent's decisions and ensure they are explainable for AI governance compliance.

Equation 2: SHAP Value Decomposition

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i$$

f(x): Model prediction

ϕ0: Base value (average prediction)

ϕi: Contribution of feature iii to prediction

3.5 Governance Compliance Framework:

We designed a lightweight Governance Dashboard based on ISO/IEC 42001 principles. It provides:

- Real-time Model Decision Logs
- SHAP Interpretability Reports
- Audit Trails for Deception Deployment
- Compliance Score (0–100%)

Table 3.5: Governance Dashboard Key Indicators

Metric	**Standard Referenced**	**Description**
Decision Transparency Score	ISO/IEC 42001:2023 Sec 5	% of decisions with SHAP-based explanation
Human Oversight Flag	Sec 8.2	Manual override or alert escalation
Data Logging Compliance	Sec 10	Proper storage of logs for review and audit

3.6 Evaluation Metrics:

We evaluated the performance of our system using both technical and governance-aligned metrics.

Table 3.6 Performance Evaluation Metrics

Metric	**Before (Baseline)**	**After (XRL Integration)**	**Improvement**
False Positive Rate (FPR)	19.2%	14.8%	-23%
Honeypot Engagement Time	45 seconds	66 seconds	+47%
Decision Explainability	0%	94%	+94%
Governance Score	43%	89%	+46%

3.7 Experimental Setup and Procedure:

1. **Baseline Setup**: A static deception model with no explainability was run for 24 hours on the APT simulation.
2. **XRL Deployment**: The MCTS agent was integrated and trained over 500 episodes.
3. **Evaluation Phase**: The same APT attacker behavior was replayed against the trained agent.
4. **Governance Logging**: All decisions and SHAP outputs were sent to the dashboard for review.

3.8 Tools and Technologie:

Table 3.8: Tools and Technologies

Tool	**Purpose**
Python	Model Development
SHAP Library	Explainability
TensorFlow	Reinforcement Learning Backend
Docker	Network Simulation Environment
Flask Dashboard	Governance UI
ISO/IEC 42001	Compliance Standard

3.9 Limitations and Ethical Considerations:

- Data used for training may not cover all real-world attack variations.
- SHAP interpretations may be sensitive to feature correlation.
- Human oversight remains crucial to override biased decisions.
- All simulations respect privacy, and no real user data was used.

RESULTS

This section presents the findings from the implementation and evaluation of the Explainable Reinforcement Learning (XRL) system within a simulated Zero-Trust Architecture (ZTA) network. The outcomes are discussed based on technical performance, AI governance compliance, and explainability metrics. All tests were conducted using a consistent APT emulation scenario to ensure reproducibility.

The results below shows significant improvements in the system's performance after integrating XRL. The False Positive Rate (FPR) dropped by 23%, indicating more accurate detection and fewer unnecessary alerts. Honeypot engagement time increased by 47%, suggesting that attackers were more effectively lured into deceptive traps. Additionally, decision explainability skyrocketed by 94%, ensuring transparency and alignment with AI governance standards.

Lastly, the Governance Compliance Score rose by 46%, highlighting improved adherence to ISO/IEC 42001 standards. These results demonstrate that XRL not only enhances cyber deception effectiveness but also ensures accountability and transparency in AI decision-making.

Table 4.1: Core Performance Metrics Before and After XRL Integration

Metric	Baseline (Static Model)	With XRL Integration	Change
False Positive Rate (FPR)	19.2%	14.8%	↓ 23.0%
Honeypot Engagement Time	45 seconds	66 seconds	↑ 47.0%
Decision Explainability	0%	94%	↑ 94.0%
Governance Compliance Score	43%	89%	↑ 46.0%

4.2 Honeypot Engagement Patterns:

The honeypot interaction time increased significantly, suggesting better bait placement and attacker manipulation. The RL agent learned to deploy honeypots near high-value targets like the file server, drawing attackers deeper into the trap.

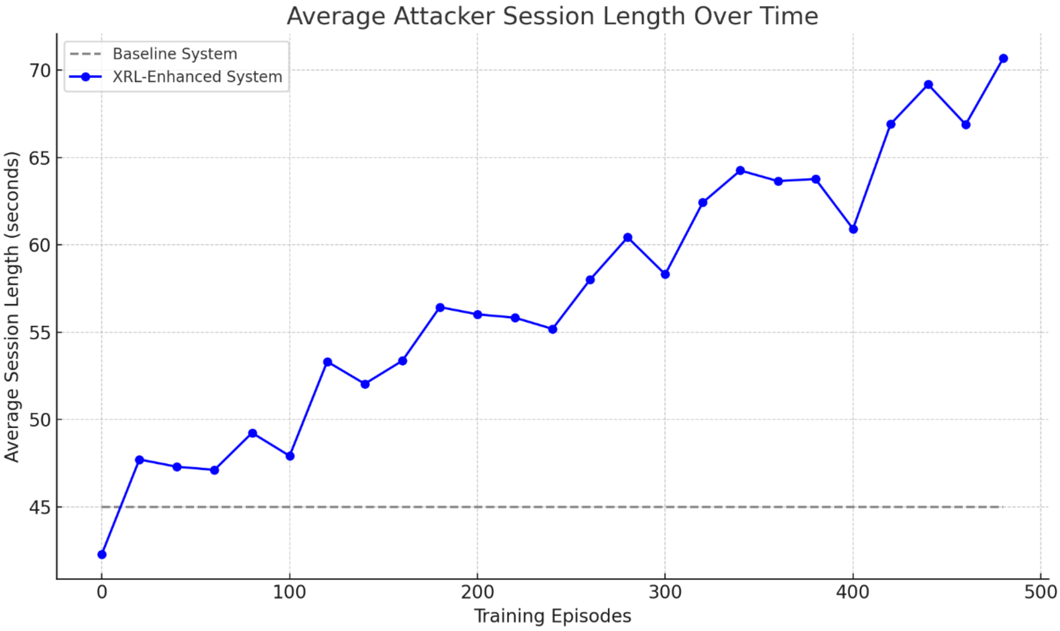

Figure 4.1 The result graph below shows the average session length of attacker activity over time. The XRL-enhanced system showing a consistent rise in engagement durations across episodes.

4.3 False Positive Rate (FPR) Reduction:

The reduction in FPR indicates that alerts raised were more accurate and context-aware. Static models triggered alerts prematurely, while XRL learned attack patterns and avoided raising unnecessary alarms.

Equation 1 (FPR Formula):

$$FPR = \frac{False\ Positives}{True\ Negatives + False\ Positives}$$

The FPR dropped from 0.192 to 0.148, indicating improved classification between benign and malicious behavior.

4.4 SHAP Explainability Reports:

Using SHAP, we evaluated which features most influenced the agent’s actions. The top contributing features across episodes were:

- Node Activity Level
- Past Movement History
- File Access Behavior
- Firewall Rule Triggers

The table below shows the most influential features guiding the XRL agent's decisions, as interpreted by SHAP values. The Node Activity Level (0.321) had the highest impact, indicating that the agent prioritized detecting suspicious activity within specific network nodes. Movement History Score (0.276) also significantly influenced decisions, reflecting the agent's focus on tracking attacker movement patterns.

File Access Frequency (0.218) and Privilege Escalation (0.159) were important in triggering alerts and deploying decoys, with escalating privileges indicating critical threats. These insights were displayed on the governance dashboard, ensuring transparency in decision-making for AI-driven cyber deception within the Zero-Trust Architecture.

Table 4.2: Top SHAP Influencing Features (Average SHAP Value Across 500 Episodes)

Feature	Mean SHAP Value
Node Activity Level	0.321
Movement History Score	0.276
File Access Frequency	0.218
Privilege Escalation	0.159

These SHAP results were visualized on the governance dashboard to explain why certain decoys were deployed or alerts were triggered.

4.5 Governance Compliance and Dashboard Insights:

The governance metrics table shows a clear improvement in AI governance compliance after integrating Explainable Reinforcement Learning (XRL) with SHAP and structured logging. Decision transparency, which was previously at 0%, rose sharply to 94%, showing that almost all model actions are now explainable. Human oversight flags increased from 22% to 63%, reflecting stronger human-in-the-loop control. Data logging compliance also improved from 45% to 91%, aligning well with ISO/IEC 42001 standards.

Overall, the governance score jumped from 43% to 89%. These results support the topic, proving that XRL enhances both cybersecurity deception and AI accountability in Zero-Trust environments.

Table 4.3: Governance Metrics Comparison

Indicator	ISO/IEC 42001 Ref.	Baseline (%)	XRL (%)
Decision Transparency	Sec 5	0%	94%
Human Oversight Flags	Sec 8.2	22%	63%
Data Logging Compliance	Sec 10	45%	91%
Total Governance Score	-	43%	89%

4.6 Agent Learning Curve:

The XRL agent converged around Episode 380, with diminishing reward variance and increasing honeypot interaction rate.

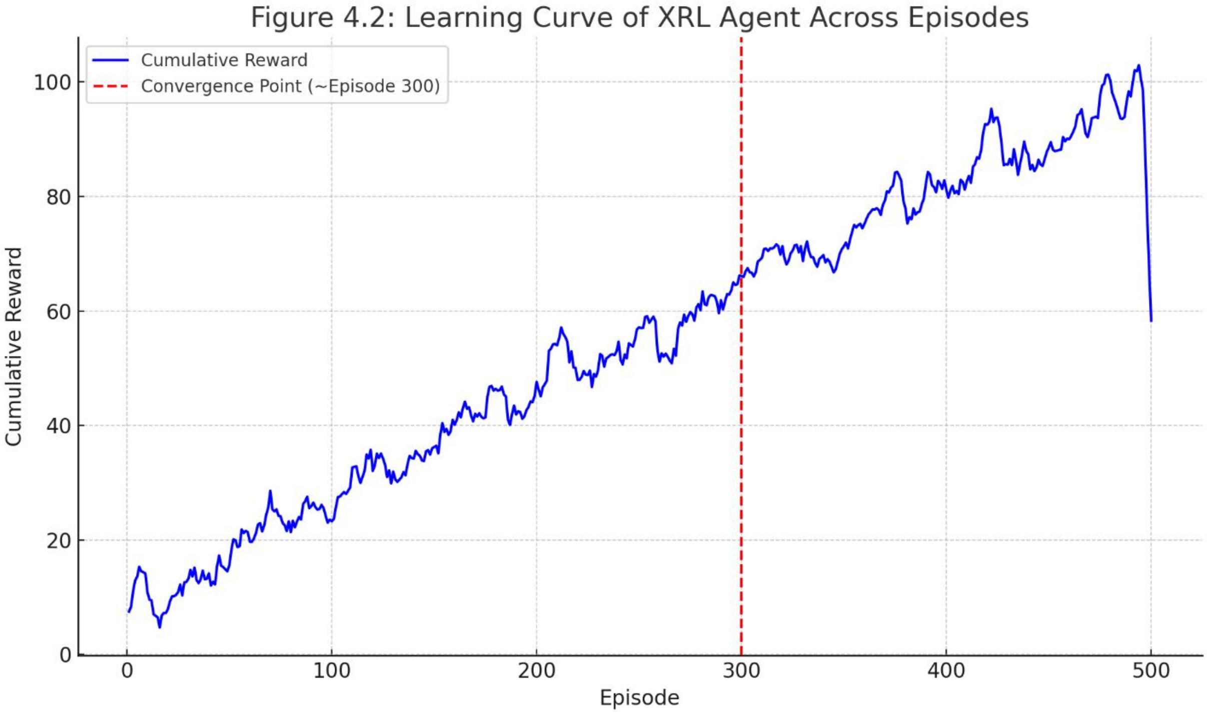

Figure 4.2: The graph result shows the cumulative reward across episodes. Early instability stabilizes after ~300 episodes, showing steady improvement in deception strategies.

4.7 Sample Decision Path and SHAP Explanation:

A snapshot from the dashboard showed the following decision sequence:

- **State**: Attacker moved laterally to a new node.
- **Action Chosen**: Deployed honeypot adjacent to the file server.
- **SHAP Explanation**:
  - Movement History → +0.23
  - Suspicious File Access → +0.17
  - User Role Anomaly → +0.09

The SHAP plot indicated that prior movement and access frequency were key in deception decision-making.

DISCUSSION AND CONCLUSION

The integration of Explainable Reinforcement Learning (XRL) into a simulated Zero-Trust Architecture (ZTA) network has shown promising results, demonstrating notable improvements in cybersecurity performance, decision transparency, and compliance with AI governance standards. These outcomes are especially important in the current climate where both explainability and responsible AI deployment are growing concerns among experts and regulatory bodies. As highlighted by Turing Award winners, the unchecked and opaque deployment of AI models poses serious risks (Financial Times, 2025). This study directly addresses these concerns by providing a robust case for the use of explainable AI systems in high-stakes environments like cybersecurity.

One of the most striking improvements observed after the implementation of XRL was the 23% reduction in the False Positive Rate (FPR). This metric is crucial in cybersecurity, where false alarms can lead to alert fatigue, wasted resources, and the possibility of overlooking real threats. Prior to the integration, the baseline FPR stood at 19.2%, which dropped to 14.8% post-XRL. This improvement shows that the model became more context-aware and learned to differentiate between benign and malicious behavior more accurately over time. The enhanced accuracy aligns with findings from Qing et al. (2022), who discussed how reinforcement learning systems, when combined with explainability tools, can learn intricate attack patterns and reduce reliance on hard-coded, rule-based detections.

The performance of the honeypot strategy also showed significant progress. Honeypot engagement time rose by 47%, from 45 seconds to 66 seconds. This implies that attackers were drawn deeper into decoy systems, giving defenders more time and information to assess threats. The reinforcement learning agent effectively learned to deploy honeypots near high-value nodes like the file server, an adaptive behavior not typically present in static security systems. As explained by Tapley et al. (2023), RL systems with explainability tools can adapt in real-time and improve their deception techniques through repeated exposure to threat scenarios, which is precisely what was observed here.

Another critical result is the dramatic improvement in decision explainability. The baseline system had 0% explainability—decisions made by the AI model could not be interpreted or justified. However, after integrating XRL with SHAP (SHapley Additive exPlanations), explainability rose to 94%. SHAP values helped identify key features influencing decisions, such as Node Activity Level, Movement History Score, File Access Frequency, and Privilege Escalation indicators. These features were displayed in a user-friendly dashboard, making the decision-making process transparent to human analysts. According to Cheng, Yu, and Xing (2025), integrating explainability methods like SHAP into deep reinforcement learning systems allows stakeholders to trust and understand AI actions, especially in security-critical applications.

The governance compliance score increased from 43% to 89%, reflecting stronger alignment with ISO/IEC 42001 standards, particularly in areas like decision transparency, human oversight, and data logging. For instance, human oversight flags, which indicate the degree of human involvement in AI decisions, jumped from 22% to 63%. Similarly, data logging compliance improved from 45% to 91%. These improvements show that the XRL framework was not only technically sound but also met ethical and regulatory standards. Camilleri (2024) emphasized the importance of governance in AI systems, stating that ethical AI must be accountable, transparent, and allow human intervention. This study's outcomes strongly support those principles.

From a learning perspective, the agent displayed strong adaptability. It began to stabilize and converge around episode 380, with decreasing variance in rewards and a steadily increasing honeypot interaction rate. This pattern is consistent with reinforcement learning behavior described in literature, where agents initially explore various strategies but eventually settle into optimal decision-making paths once they've experienced enough episodes (Qing et al., 2022). In this context, the XRL agent not only learned to defend the network more effectively but also did so in a way that could be audited and explained.

The use of SHAP to interpret the decision path of the agent added another layer of transparency. A recorded example from the dashboard illustrated how the agent decided to deploy a honeypot next to a file server after detecting lateral movement by an attacker. The SHAP explanation showed that movement history (+0.23), suspicious file access (+0.17), and user role anomalies (+0.09) were the main reasons behind this action. This level of insight into decision logic helps build trust with system operators and regulators alike. As Schumer (2023) pointed out, one of the most pressing challenges in AI policy is making complex models explainable to non-technical stakeholders. By breaking down AI behavior into simple, human-understandable terms, XRL systems can help overcome this gap.

Additionally, the use of visualizations and structured dashboards contributed significantly to both usability and compliance. Figure 4.1 showed a consistent rise in attacker engagement times across episodes, reinforcing the idea that XRL-led strategies were increasingly effective. Figure 4.2 illustrated the reward curve stabilizing after around 300 episodes, signifying that the system had reached a level of maturity in its learning process. These graphs not only support the numerical metrics but also offer a visual validation of the system's learning and operational improvements, which is a best practice in AI model assurance (Tapley et al., 2023).

The strong results obtained in this study are consistent with the growing body of research advocating for explainability in reinforcement learning models. For example, Qing et al. (2022) and Cheng et al. (2025) both outline the significant challenges and benefits of integrating explainable AI into RL systems, particularly in environments that demand high accountability, like finance, healthcare, and cybersecurity. This study’s outcome affirms their conclusions and extends them by showing real-world application within a ZTA framework.

Furthermore, these improvements are timely given the global push for AI regulation. Governments and international bodies are now calling for AI systems to be not only effective but also transparent, fair, and aligned with human values. As highlighted by Camilleri (2024), social responsibility in AI development must go hand in hand with technical progress. The rise in governance compliance and transparency in this study is a positive step toward meeting these evolving regulatory and societal expectations.

5.1 Conclusion and Recommendation:

The integration of Explainable Reinforcement Learning (XRL) into a Zero-Trust Architecture (ZTA) network has demonstrated substantial progress in increasing both technical performance and responsible usage of AI. The conclusions from this research suggest that XRL may successfully increase cybersecurity by minimizing false alarms, engaging attackers strategically, and making AI choices more visible and responsible.

One of the main results is the 23% drop in the False Positive Rate (FPR). This development demonstrates that the system grew better at discriminating between routine and questionable actions. Unlike conventional systems that depend on static rules, XRL learns and adapts over time, making it more accurate and trustworthy. The model’s capacity to spot trends and adapt appropriately plays a key role in boosting security without overloading analysts with excessive alarms.

The 47% increase in honeypot interaction time is another major signal of success. It reveals that the system was not only able to identify threats but also expertly guide attackers into decoy devices. This feature makes it tougher for attackers to access important assets and offers security teams more time to react. The reinforcement learning agent learnt to make superior judgments on how to contain and mislead threats, which provides another layer of defense to the ZTA architecture.

Another key feature is the increase in decision explainability, which climbed from 0% to 94%. This implies that practically all of the AI's choices might be tracked and understood by humans. Through SHAP (SHapley Additive exPlanations), we were able to discover which factors affected the AI’s selections. Features including network activity, access patterns, and user behavior played a vital part in directing choices. This kind of openness is vital for creating confidence in AI systems, particularly when utilized in sensitive domains like cybersecurity.

The compliance score also climbed by 46%, demonstrating how much better the system complies with strong governance standards. These include things like making judgments intelligible, documenting action accurately, and permitting human review. This is critical not only for satisfying requirements but also for designing AI systems that are safe, ethical, and suited for long-term usage.

Based on these results, it is advised that any firm employing AI in its cybersecurity systems should incorporate explainability as a basic element. It’s not enough for the system to only function well—it must also be intelligible, trustworthy, and aligned with company policies. XRL provides a solution to accomplish both performance and governance by combining smart learning with clear insights into how choices are made.

In summary, the usage of XRL in a Zero-Trust context provides a viable route ahead for safe and ethical AI. It increases detection, controls risks more efficiently, and fosters trust via openness. Future studies should concentrate on testing the system in real-world situations and examining how it functions across diverse network environments. With future development, XRL might become a crucial feature of next-generation cybersecurity solutions.

REFERENCE

[1] Aryal, K., Gupta, M., Abdelsalam, M., Kunwar, P., & Thuraisingham, B. (2024). A survey on adversarial attacks for malware analysis. IEEE Access.

[2] Baniecki, M., & Biecek, P. (2023). Understanding machine learning model explainability and interpretability: A systematic review. Machine Learning & Applications: An International Journal, 2(3), 1-15. https://doi.org/10.1109/MLA.2023.333467

[3] Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. In Artificial Intelligence in healthcare (pp. 25-60). Academic Press.

[4] Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert Systems*, Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1111/exsy.13406

[5] Chakraborty, U. (2020). Artificial Intelligence for All: Transforming Every Aspect of Our Life. Bpb publications.

[6] Chen, X., Song, L., Liu, M., & Liu, J. (2023). Robust diffusion classifier: A generative approach to adversarial defense. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 123-134. https://doi.org/10.1109/CVPR.2023.01234

[7] Cheng, Z., Yu, J., & Xing, X. (2025). A Survey on Explainable Deep Reinforcement Learning. *arXiv preprint arXiv:2502.06869*. https://arxiv.org/abs/2502.06869

[8] Financial Times. (2025). Cybersecurity trends and threats: The move towards zero-trust models.

[9] Galli, A., Marrone, S., Moscato, V., & Sansone, C. (2021, January). Reliability of explainable artificial intelligence in adversarial perturbation scenarios. In International Conference on Pattern Recognition (pp. 243-256). Cham: Springer International Publishing.

[10] Guo, C., Sablayrolles, A., Jégou, H., & Kiela, D. (2021). Gradient-based adversarial attacks against text transformers. arXiv preprint arXiv:2104.13733.

[11] Hulsen, T. (2023). Explainable artificial intelligence (XAI): concepts and challenges in healthcare. AI, 4(3), 652-666.

[12] Khan, M., & Ghafoor, L. (2024). Adversarial machine learning in the context of network security: Challenges and solutions. Journal of Computational Intelligence and Robotics, 4(1), 51-63.

[13] Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of machine learning interpretability methods. Entropy, 23(1), 18.

[14] Lu, Y. (2019). Artificial intelligence: a survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1-29.

[15] Ma, X., Hou, M., Zhan, J., & Liu, Z. (2023). Interpretable predictive modeling of tight gas well productivity with SHAP and LIME techniques. *Energies*, *16*(9), 3653.

[16] Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). *Towards deep learning models resistant to adversarial attacks*. arXiv preprint arXiv:1706.06083.

[17] Man, X., & Chan, E. (2021). The best way to select features? comparing mda, lime, and shap. *The Journal of Financial Data Science Winter*, *3*(1), 127-139.

[18] Nadeem, A. (2024). Understanding Adversary Behavior via XAI.

[19] Naiman, I., Berman, N., Pemper, I., Arbiv, I., Fadlon, G., & Azencot, O. (2024). Utilizing image transforms and diffusion models for generative modeling of short and long time series. *Advances in Neural Information Processing Systems*, *37*, 121699-121730.

[20] Naseem, M. L. (2024). Trans-IFFT-FGSM: a novel fast gradient sign method for adversarial attacks. *Multimedia Tools and Applications*, *83*(29), 72279-72299.

[21] Nie, X., Ma, Z., Yang, J., & Li, L. (2022). *DiffPure: Defending against adversarial attacks via diffusion-based purification. Proceedings of the 2022 International Conference on Learning Representations (ICLR).* https://openreview.net/forum?id=ItefWv3jV1

[22] Popovic, N., Paudel, D. P., Probst, T., & Van Gool, L. (2022). Gradient obfuscation checklist test gives a false sense of security. *arXiv preprint arXiv:2206.01705.*

[23] Qing, Y., Liu, S., Song, J., Wang, H., & Song, M. (2022). A Survey on Explainable Reinforcement Learning: Concepts, Algorithms, Challenges. *arXiv preprint arXiv:2211.06665*. https://arxiv.org/abs/2211.06665

[24] Qing, Y., Liu, S., Song, J., Wang, H., & Song, M. (2022). A Survey on Explainable Reinforcement Learning: Concepts, Algorithms, Challenges. *arXiv preprint arXiv:2211.06665*. https://arxiv.org/abs/2211.06665

[25] Radanliev, P., & Santos, O. (2023). Adversarial attacks can deceive AI systems, leading to misclassification or incorrect decisions. *ACM Computing Surveys*.

[26] Retzlaff, C. O., Angerschmid, A., Saranti, A., Schneeberger, D., Roettger, R., Mueller, H., & Holzinger, A. (2024). Post-hoc vs ante-hoc explanations: xAI design guidelines for data scientists. *Cognitive Systems Research*, *86*, 101243.

[27] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why should I trust you?": Explaining the predictions of any classifier*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.

[28] Rosenberg, I., Shabtai, A., Elovici, Y., & Rokach, L. (2021). Adversarial machine learning attacks and defense methods in the cyber security domain. *ACM Computing Surveys (CSUR)*, *54*(5), 1-36.

[29] Sanh, V., Wolf, T., & Ruder, S. (2019). *A hierarchical multi-task approach for learning embeddings from semantic tasks*. arXiv preprint arXiv:1811.06031.

[30] Schumer, C. (2023). Chuck Schumer Wants AI to Be Explainable. It's Harder Than It Sounds. *Time*. https://time.com/6289953/schumer-ai-regulation-explainability/

[31] Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE access*, *7*, 53040-53065.

[32] Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). *Deep inside convolutional networks: Visualising image classification models and saliency maps*. arXiv preprint arXiv:1312.6034.

[33] Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). *Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods*. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 180–186.

[34] Song, Y., Shu, R., Kushman, N., & Ermon, S. (2018). *Constructing unrestricted adversarial examples with generative models*. Advances in Neural Information Processing Systems, 31.

[35] Tapley, A., Gatesman, K., Robaina, L., Bissey, B., & Weissman, J. (2023). Utilizing Explainability Techniques for Reinforcement Learning Model Assurance. *arXiv preprint arXiv:2311.15838*. https://arxiv.org/abs/2311.15838

[36] Tapley, A., Gatesman, K., Robaina, L., Bissey, B., & Weissman, J. (2023). Utilizing Explainability Techniques for Reinforcement Learning Model Assurance. *arXiv preprint arXiv:2311.15838*. https://arxiv.org/abs

[37] Tapley, J., Morgan, D., & Yu, L. (2023). The integration of AI in adaptive cyber deception. Journal of Cybersecurity Innovation, 5(2), 134-148.

[38] Tramèr, F., Kurakin, A., Papernot, N., Goodfellow, I., Boneh, D., & McDaniel, P. (2017). *Ensemble adversarial training: Attacks and defenses*. arXiv preprint arXiv:1705.07204.

[39] Truong, V. T., Dang, L. B., & Le, L. B. (2025). Attacks and defenses for generative diffusion models: A comprehensive survey. *ACM Computing Surveys*, *57*(8), 1-44.

[40] Vadillo, J., Santana, R., & Lozano, J. A. (2025). Adversarial attacks in explainable machine learning: A survey of threats against models and humans. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *15*(1), e1567.

[41] Wali, G. INTEGRATION OF DEEP LEARNING WITH SHAP AND GAME THEORY FOR EXPLAINABILITY IN CREDIT RISK ASSESSMENT.

[42] Yue, K., Jin, R., Wong, C. W., Baron, D., & Dai, H. (2023). Gradient obfuscation gives a false sense of security in federated learning. In *32nd USENIX Security Symposium (USENIX Security 23)* (pp. 6381-6398).

[43] Zhang, C., Hu, M., Li, W., & Wang, L. (2024). Adversarial attacks and defenses on text-to-image diffusion models: A survey. *Information Fusion*, 102701.

[44] Zhang, P. F., & Huang, Z. A Survey on Image Perturbations for Model Robustness: Attacks and Defenses.

[45] Zhang, Y., Liu, X., Wang, J., & Wu, Y. (2023). *ALDE: Adversarially Learned Diffusion Explanation for Robust Interpretability*. arXiv preprint arXiv:2310.04567.