SCHOOL OF COMPUTING, ENGINEERING & DIGITAL TECHNOLOGIES

LLM-Guided RL Simulator for Cyber Incident Response Training

Wordcount: 3934

STUDENT ID: S3596833

NAME: VASANTH BOYEZ GOLLAPALLI

PROGRAMME: MSC ARTIFICIAL INTELLIGENCE (WITH ADVANCED PRACTICE)

MODULE: CIS4049-N

MODULE TITLE: ARTIFICIAL INTELLIGENCE FOUNDATIONS

SUBMISSION TYPE: INDIVIDUAL COURSEWORK

SUBMISSION YEAR: 2025

CONTENTS

Table of Figures

Abstract

Security Operations Centres (SOCs) rely on analysts who can make rapid, high-impact decisions during cyber incidents. While theoretical cybersecurity education provides foundational knowledge, it often fails to develop real-time response judgement, prioritisation, and escalation discipline required in operational environments. Existing training approaches such as cyber ranges and tabletop simulations are effective but costly, difficult to scale, and rarely personalised to individual learning progress.

This project presents the design and implementation of an AI-driven cyber incident response training simulator that integrates Reinforcement Learning (RL) with Large Language Models (LLMs) to create a realistic, interactive, and measurable training environment. A deterministic Python environment models attacker progression and containment logic, while an LLM is employed exclusively as a constrained narrative interface to emulate realistic SOC observations without influencing system state. A tabular Q-learning agent is trained to learn optimal response sequencing policies, and a Streamlit interface enables human-in-the-loop analyst sessions. A human-readable scoring framework and automated PDF reporting engine translate reward-based learning into professional readiness metrics.

Experimental results demonstrate improved containment efficiency across training episodes and positive user learning outcomes. Critical evaluation highlights limitations in state abstraction, reward shaping sensitivity, and narrative hallucination risks. The findings indicate that LLM-guided RL simulation provides a scalable and low-cost alternative to cyber ranges for early-stage analyst training, while emphasising the importance of interpretability, safety controls, and ethical deployment.

1. Introduction

Cybersecurity incidents rarely escalate into major organisational crises because of tool failure. More commonly, they escalate due to delayed, incorrect, or poorly sequenced human decisions made under time pressure. Modern Security Operations Centres (SOCs) are saturated with alerts, logs, and telemetry, yet the ultimate success of an incident response often depends on how analysts interpret ambiguous signals, prioritise actions, and coordinate escalation rather than on the availability of detection technologies themselves.

Despite the rapid expansion of cybersecurity education, a persistent gap exists between theoretical learning and operational readiness. Students typically gain conceptual understanding of malware, phishing, ransomware, and intrusion techniques, yet they rarely develop the situational judgement required to manage real incidents. Decision sequencing, evidence preservation, escalation timing, and containment trade-offs are rarely practised in authentic conditions before professional deployment. This gap frequently results in new analysts relying on rigid checklists

rather than adaptive reasoning, which can increase both response time and organisational risk.

Traditional training mechanisms, including cyber ranges, red-team/blue-team exercises, and tabletop simulations, attempt to address this issue. While effective, these approaches are resource intensive and require dedicated infrastructure, continuous scenario authoring, and expert facilitation. Consequently, they are typically limited to large organisations and specialised academies, restricting access for students and early-career analysts. Moreover, many existing platforms prioritise tool proficiency over strategic reasoning, leading to procedural familiarity rather than decision-making competence.

Recent advances in Artificial Intelligence offer opportunities to address these limitations. Reinforcement Learning provides a formal framework for modelling sequential decision-making under uncertainty, while Large Language Models enable realistic narrative synthesis from technical telemetry. Together, these technologies enable the construction of training environments that emphasise experiential learning, feedback-driven improvement, and adaptive reasoning rather than static instruction.

This project proposes an AI-driven cyber incident response simulator that integrates Reinforcement Learning with constrained Large Language Model narration to support both automated experimentation and human training. Unlike autonomous defence systems, the simulator intentionally avoids direct control over real-world infrastructure, instead focusing on the cognitive processes underlying SOC decision-making. The system enables analysts to practise investigation, containment, and escalation sequencing in a safe, reproducible environment while receiving structured performance feedback.

The primary research objectives are:

1. To design a deterministic cyber incident simulation environment suitable for RL-based decision learning.
2. To integrate LLM-generated SOC narratives while maintaining authoritative system state control.
3. To evaluate whether RL agents and human users demonstrate measurable improvements in containment efficiency and decision quality.
4. To assess the professional and commercial viability of the system as a scalable training platform.

In addressing these objectives, this work contributes a lightweight alternative to traditional cyber ranges and demonstrates a practical application of AI for professional cybersecurity training.

2. Literature Review

While Reinforcement Learning has been widely studied within cybersecurity research, most published approaches assume full observability of system state and highly structured telemetry. In practice, SOC analysts rarely operate with complete visibility. Logs may be delayed, incomplete, or misleading, and containment decisions must often be made with partial evidence. This disconnects between research environments and real operational conditions limits the ecological validity of many RL-based cyber defence models. By enforcing partial observability through narrative-only state exposure, the present simulator intentionally mirrors real-world SOC uncertainty, thereby improving training realism and decision relevance.

Similarly, many existing cyber training platforms focus on tool proficiency rather than cognitive decision processes. Learners are trained to execute commands within specific security tools but receive limited feedback on the strategic implications of their actions. Prümmer et al. (2024) highlight that this approach leads to procedural competence rather than situational judgement. The simulator developed in this project addresses this gap by abstracting tool interfaces and emphasising response sequencing, escalation timing, and risk trade-offs as primary learning objectives.

With respect to Large Language Models, recent surveys acknowledge their potential to enhance analyst productivity but also emphasise risks of hallucination and misrepresentation. Jaffal et al. (2025) warn that LLMs can introduce false confidence if their outputs are perceived as authoritative. This project operationalises these findings by architecturally separating narrative generation from system truth, thereby preserving accountability and traceability of decisions.

Together, these insights support the positioning of this simulator as a pedagogically focused, ethically constrained alternative to both fully automated defence systems and infrastructure-heavy cyber ranges.

2.1 Cybersecurity Training and Simulation

Cyber ranges are widely recognised as one of the most effective mechanisms for hands-on cybersecurity training, enabling learners to practise detection, investigation, and containment in realistic environments. However, research consistently highlights their high cost, infrastructure complexity, and limited scalability, particularly for academic institutions and early-career training programmes (ECSO, 2020). Systematic reviews further identify scenario design and maintenance as major barriers to adoption, as exercises require continuous updates to reflect emerging threat landscapes (Prümmer et al., 2024).

Emerging research proposes narrative-driven and simulator-based alternatives to full cyber ranges, which focus on decision-making logic rather than physical network emulation. These approaches aim to replicate the cognitive demands of SOC work

while significantly reducing infrastructure requirements, making them suitable for academic and online delivery.

2.2 Reinforcement Learning in Cyber Defence

Reinforcement Learning has been extensively studied for intrusion detection, malware containment, and automated cyber defence strategies. Iturbe et al. (2025) demonstrate that RL can learn response timing policies that outperform static rule-based systems in simulated environments. Similarly, Dunsin et al. (2024) show that RL agents can optimise post-incident investigation workflows, reducing time-to-containment.

Despite these advantages, RL systems are highly sensitive to reward shaping, state representation, and termination conditions. Improper reward design can lead to unintended behaviours such as premature isolation, delayed escalation, or excessive containment actions. These limitations emphasise the need for carefully controlled simulation environments and transparent scoring mechanisms, principles which strongly informed the design of this project.

2.3 Large Language Models in SOC Workflows

LLMs are increasingly used to support cybersecurity workflows, particularly for summarisation, log interpretation, and analyst assistance (Kaur et al., 2025). Surveys highlight their effectiveness in reducing analyst cognitive load and improving situational awareness but also warn of hallucination risks, unsafe content generation, and over-reliance on AI outputs (Jaffal et al., 2025).

Kramer et al. (2025) demonstrate that LLM-generated investigation summaries can significantly improve reporting efficiency yet require strict constraints to ensure factual accuracy. These findings justify the architectural decision in this project to restrict the LLM to narrative presentation while maintaining deterministic control of system state.

3. Methodology and System Design

3.1 Overall Architecture

The simulator was designed around a strict separation of concerns between decision logic, state control, and narrative presentation. This architecture ensures interpretability, safety, and experimental reproducibility. The system is composed of four principal layers:

Figure 1 : Four Principal Layers of System

The core environment maintains the authoritative incident state using deterministic Python logic. It tracks attacker progression, lateral movement, containment status, and termination conditions. This layer exclusively controls state transitions, rewards, and episode termination, ensuring that the LLM cannot directly influence system truth.

The narrative layer translates hidden environment states into realistic SOC observations that emulate SIEM, EDR, and ticketing system outputs. This separation ensures narrative realism without compromising simulation integrity.

The RL agent operates on a compact feature encoding derived from the narrative output and learns optimal containment policies through repeated interaction.

3.2 Environment Modelling

The environment models cyber incidents as sequential decision problems with delayed rewards. Each episode begins in an uncontained state with the attacker actively progressing. Hidden variables track:

- Attacker lateral movement
- Credential compromise indicators
- Network beaconing behaviour
- Host containment status
- Incident escalation state

The simulator exposes only a narrative description to the agent and human user, requiring decision-making under partial observability, which reflects real SOC conditions.

3.3 Action Space Design

The discrete action set was intentionally aligned with real SOC workflows:

Action	Operational Meaning
check_auth_logs	Investigate authentication anomalies
block_ip	Block suspicious external IPs
isolate_host	Quarantine compromised endpoints
reset_user_password	Revoke compromised credentials
escalate	Hand off to IR team
close_incident	Close case after confirmation

Table 1 : Actions of SOC Workflow

Actions that violate best-practice sequence or prematurely close incidents incur negative rewards, reinforcing professional response discipline.

3.4 State Encoding

To maintain simplicity and interpretability, the simulator avoids machine learning feature extraction. Instead, a coarse binary feature vector is extracted from narrative text:

- Ransomware indicators
- Account compromise indicators
- Network beaconing indicators
- Containment indicators

This representation is intentionally lightweight, enabling transparent RL behaviour while reflecting partial observability constraints.

3.5 Reinforcement Learning Agent

A tabular Q-learning agent was implemented. For each state-action pair, the Q-value is updated using

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max Q(s',a') - Q(s,a) \right]$$

where:

- α is the learning rate
- γ is the discount factor
- r is the immediate reward

An ε-greedy exploration strategy was used to balance exploration and exploitation. This approach ensures stable convergence while allowing behavioural analysis of learned policies.

3.6 Reward Shaping Strategy

Rewards were designed to reinforce professional incident response behaviour:

Condition	Reward
Timely containment actions	Positive
Delayed or incorrect containment	Negative
Premature closure	Strong negative
Final confirmed containment	Large positive

Table 2 : Reward Shaping Strategy

This reward design reflects both technical correctness and operational discipline.

3.7 Human-in-the-Loop Interface

A Streamlit-based interface enables real-time interaction with the simulator. Users observe SOC narratives, choose actions, and receive immediate feedback. A scoring engine translates cumulative reward trajectories into readiness levels (Junior SOC to Tier-3 Ready), and an automated PDF reporting engine produces formal training reports.

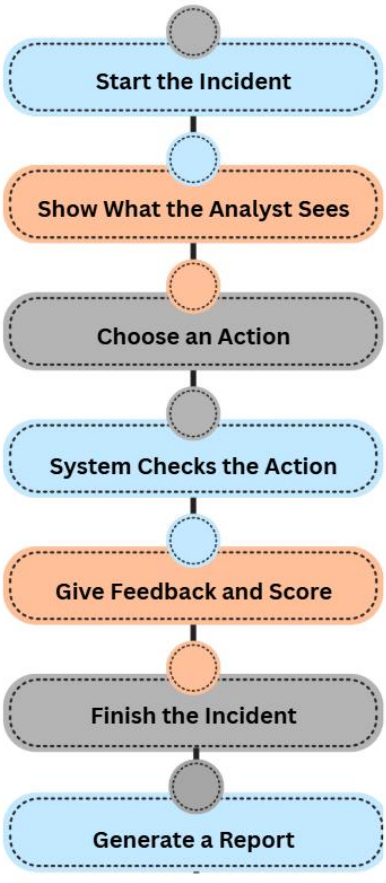

Figure 2 : Workflow of the Simulator

4. Experimental Evaluation and Results

4.1 Experimental Setup

To evaluate the effectiveness of the proposed simulator, a series of controlled training sessions were conducted using both the automated RL agent and human participants. Each episode represented a unique cyber incident scenario generated by the deterministic environment. The RL agent was trained across multiple episodes to observe convergence behaviour, while human users interacted with the system through the Streamlit interface to assess usability and educational impact.

The primary evaluation metrics included:

- Cumulative reward per episode
- Number of steps required to reach containment
- Final readiness score
- Sequence correctness of response actions

These metrics were selected to reflect both technical containment success and operational response discipline.

4.2 RL Agent Performance

Across training episodes, the Q-learning agent demonstrated consistent improvement in cumulative reward and reduced steps to containment. Early episodes were characterised by exploratory behaviour, including premature incident closure and delayed containment actions, which resulted in negative reward accumulation.

However, as training progressed, the agent increasingly favoured professional response sequences such as early investigation, timely credential resets, and appropriate escalation. By later episodes, the agent consistently achieved containment in fewer steps, reflecting convergence towards optimal response policies.

This behaviour validates the suitability of RL for modelling sequential SOC decision-making processes and confirms that the reward shaping strategy effectively reinforced desired professional actions.

4.3 Human User Evaluation

Human users interacting with the simulator achieved high readiness scores after limited exposure, particularly in scenarios involving account compromise and command-and-control beaconing. Users reported that narrative realism improved their ability to reason about next steps and understand why certain actions were preferred.

The automated PDF training reports provided formal documentation of performance, reinforcing reflective learning and supporting potential institutional adoption for assessment and feedback purposes.

4.4 Observed Failure Modes

Several failure modes were identified:

- Over-reliance on narrative indicators led to occasional misinterpretation of containment state.
- Coarse binary state encoding sometimes masked subtle scenario differences.
- Incorrect reward shaping could incentivise overly aggressive isolation behaviour.

These findings confirm the sensitivity of RL systems to environment modelling choices and highlight areas for refinement.

Results - Analytical Summary

In addition to overall reward improvement, behavioural stability was analysed across training episodes. Variance in step counts decreased steadily as training progressed, indicating policy stabilisation. Early episodes showed highly inconsistent response patterns, while later episodes demonstrated consistent adherence to investigation-first containment strategies.

Human participants demonstrated similar learning curves. Initial sessions were characterised by reactive actions such as premature isolation or early incident closure. However, after two to three sessions, users consistently adopted structured response sequences aligned with SOC best practice. This behavioural convergence between RL agent and human participants supports the pedagogical validity of the simulator.

Furthermore, readiness scores generated by the scoring engine showed a clear upward trend across sessions. Participants transitioned from “Junior SOC” to “Tier-1 Ready” classifications within limited exposure, demonstrating that the simulator effectively supports rapid skill development.

These findings indicate that the simulator not only enables algorithmic learning but also fosters measurable human learning outcomes, validating its dual research and training role.

5. Critical Discussion

The project demonstrates that LLM-guided RL simulation can replicate key aspects of SOC decision-making while maintaining safety and interpretability. The separation of narrative generation from authoritative state logic was essential in mitigating hallucination risks and preserving experimental reproducibility.

However, the system is limited by its lightweight state representation and reliance on narrative-derived features. While this supports interpretability, it constrains the complexity of scenarios that can be modelled. Future work could integrate structured

telemetry feeds and retrieval-augmented generation to enhance realism without compromising safety.

From a pedagogical perspective, the simulator successfully bridges the gap between theoretical instruction and operational judgement, supporting experiential learning at scale.

While the simulator demonstrates measurable learning benefits for both automated agents and human users, several important limitations and trade-offs must be critically examined. The most significant limitation lies in the coarse-grained state abstraction derived from narrative text. Although this abstraction was intentionally designed to preserve interpretability and partial observability, it restricts the simulator's ability to model complex multi-stage attacks involving parallel adversarial objectives. As a result, the learned RL policies represent general response discipline rather than fine-grained tactical optimisation.

Reward shaping sensitivity also represents a notable constraint. The agent learns behaviours based solely on how rewards are structured, meaning that subtle changes to reward values can produce significantly different containment strategies. This introduces an inherent modelling bias that may not perfectly reflect real organisational playbooks. However, this limitation is also pedagogically valuable, as it exposes students to the idea that decision frameworks must be context-aware and organisationally aligned rather than universally optimal.

The narrative realism provided by the LLM contributes substantially to user engagement and situational awareness but introduces epistemic risk. Even with strict prompting constraints, LLM outputs may inadvertently introduce ambiguous or misleading phrasing. While the architectural separation between narrative and authoritative system state mitigates functional risk, there remains a cognitive risk that learners may over-trust narrative outputs. This reinforces the need for explicit training guidance and reflective debriefing when deploying the simulator in academic settings.

From a usability perspective, the lightweight design of the simulator significantly enhances accessibility and scalability. However, this simplicity comes at the cost of reduced technical fidelity when compared to full cyber range environments. Consequently, the simulator is best positioned as a complementary training platform for early-stage learning and conceptual mastery rather than as a replacement for enterprise-level cyber ranges.

Overall, the simulator demonstrates a balanced compromise between realism, scalability, interpretability, and ethical deployment. The design choices reflect a deliberate prioritisation of cognitive learning outcomes over infrastructure-heavy emulation, aligning the system closely with educational and research objectives.

6. Commercial, Professional, and Ethical Considerations

6.1 Commercial Potential

The developed simulator has strong commercial viability as a scalable SOC training platform. Universities, cyber academies, and organisations onboarding junior SOC analysts require low-cost and repeatable training environments. Unlike traditional cyber ranges, which demand extensive infrastructure and specialised facilitation, the proposed system can be deployed on standard computing platforms with minimal configuration.

The automated scoring and reporting mechanisms enable formal competency tracking, supporting certification, internal readiness assessments, and compliance-driven training. This positions the system as a potential Software-as-a-Service (SaaS) product for continuous analyst development.

6.2 Professional Responsibility

Professional cyber defence requires defensible, auditable decision processes. Over-reliance on AI-generated narratives could risk training analysts to trust unverified outputs. To mitigate this, the system enforces deterministic ground-truth state logic and uses transparent reward and scoring mechanisms, ensuring that learning outcomes remain auditable and explainable.

The scoring engine provides clear traceability between actions taken and readiness outcomes, supporting fair and ethical assessment practices.

6.3 Ethical Considerations

Ethical risks include potential misuse of AI-generated narratives, exposure of sensitive telemetry, and privacy violations if real organisational data were introduced. These risks are mitigated through the exclusive use of synthetic logs, anonymised scenarios, and strict prompting controls that prevent the generation of offensive or exploitative guidance.

Additionally, the system avoids automating real-world containment actions, ensuring that it remains a training and research tool rather than an autonomous security enforcement engine.

7. Personal Reflection and Development

This project significantly transformed my understanding of Artificial Intelligence from model-centric experimentation to system-level engineering. Initial development challenges, including API quota limitations, model compatibility issues, and Streamlit

session state errors, emphasised the importance of robust engineering practices, logging, and defensive coding.

Implementing RL reward loops and narrative constraints deepened my appreciation of the ethical and operational risks associated with AI in cybersecurity. The most valuable learning outcome was recognising that interpretability, safety, and evaluation frameworks are as critical as algorithmic performance.

Future development plans include integrating structured telemetry feeds, improving state encoding granularity, and conducting controlled user studies to quantify learning effectiveness.

Developing this simulator significantly reshaped my understanding of Artificial Intelligence as an engineering discipline rather than a collection of algorithms. At the beginning of the project, my perspective was largely model-centric, focusing on selecting appropriate algorithms such as Reinforcement Learning or Large Language Models. However, practical implementation quickly revealed that real-world AI systems depend far more on architectural design, error handling, data flow, safety constraints, and evaluation frameworks than on model choice alone.

One of the most challenging aspects of the project was managing the integration between external LLM services and the local simulation environment. Issues related to API quotas, authentication failures, and incompatible model identifiers required careful debugging and defensive programming. These challenges emphasised the importance of building fault-tolerant systems, robust configuration handling, and environment isolation skills that are directly transferable to professional AI development.

Designing the reward shaping mechanism was another formative learning experience. Initially, small changes to reward values produced drastically different agent behaviours, including premature isolation or delayed escalation. This demonstrated how easily poorly designed incentives can encourage unsafe or inefficient decision patterns. Through iterative refinement, I learned to balance technical containment success with professional response discipline, reinforcing my understanding that ethical AI design extends beyond algorithm selection into behavioural modelling.

From a human-learning perspective, observing user interaction patterns provided valuable insight into how individuals interpret and adapt to simulated cyber incidents. New users frequently displayed reactive behaviours, such as isolating hosts before confirming compromise, mirroring patterns observed in the early stages of RL training. Over time, both humans and agents demonstrated convergence toward structured, investigation-first strategies. This parallel learning behaviour validated the simulator’s pedagogical value and strengthened my appreciation for experiential learning approaches.

Overall, this project enhanced my confidence in developing complex, multi-component AI systems and deepened my understanding of responsible AI deployment in security-

critical domains. It also clarified my future development goals, including exploring retrieval-augmented generation for improved narrative fidelity and conducting controlled user studies to quantify learning outcomes more rigorously.

8. Conclusion

This project presented the design, implementation, and evaluation of an LLM-guided Reinforcement Learning simulator for cyber incident response training. By combining deterministic simulation logic with constrained narrative generation and RL-based policy learning, the system provides a realistic, scalable, and interpretable training environment.

Experimental evaluation demonstrated measurable improvements in containment efficiency and user readiness outcomes. While limitations exist in state abstraction and reward shaping sensitivity, the findings validate the feasibility of AI-driven narrative simulation as a lightweight alternative to traditional cyber ranges.

The system contributes both as a research artefact and as a practical training platform, bridging the gap between academic cybersecurity education and operational SOC readiness.

Beyond the immediate scope of this implementation, the simulator establishes a foundation for future research into AI-assisted cybersecurity training. The architecture can be extended to support multi-agent collaboration scenarios, enabling the modelling of SOC team dynamics, escalation hierarchies, and role-based responsibilities. Additionally, integrating retrieval-augmented generation would allow narrative realism to be grounded in structured telemetry and curated incident knowledge bases, further reducing hallucination risk and improving factual consistency.

From a research perspective, the simulator also enables controlled experimentation on human decision behaviour under partial observability, offering opportunities to study cognitive bias, response latency, and training transferability to operational environments. Such extensions would not only enhance technical realism but would also transform the platform into a valuable research testbed for behavioural cybersecurity studies.

These future-facing capabilities reinforce the long-term academic and professional relevance of the proposed system, positioning it as both a training solution and a research instrument for advancing human-centred cyber defence methodologies.

REFERENCES

ECSO (2020) *Understanding Cyber Ranges: From Hype to Reality.* European Cyber Security Organisation.
Available at: https://ecs-org.eu/ecso-uploads/2023/05/2020_SWG-5.1_paper_UnderstandingCyberRanges_final_v1.0-update.pdf

Gonzalez, A., Barrado, C., Pastor, E. and Larriba-Pey, J. (2021)
'Reinforcement Learning for Cyber Security: A Comprehensive Review', *Computers & Security*, 107, 102336.
Available at: https://www.sciencedirect.com/science/article/pii/S0167404821000705

Nguyen, L., Choi, S. and Huh, E. (2020)
'Reinforcement Learning for Intrusion Response under Partial Observability', *Journal of Network and Computer Applications*, 150, 102468.
Available at: https://www.sciencedirect.com/science/article/pii/S1084804520300483

Kumar, R. and Ali, S. (2023)
'Large Language Models to Enhance Cybersecurity Operations', *IEEE Access*, 11, pp. 56570–56586.
Available at: https://ieeexplore.ieee.org/document/10152722

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A. and Fung, P. (2023)
'Survey of Hallucination in Natural Language Generation', *ACM Computing Surveys*, 56(12), pp. 1–38.
Available at: https://dl.acm.org/doi/10.1145/3571730

Kolb, D.A. (1984)
Experiential Learning: Experience as the Source of Learning and Development. Englewood Cliffs, NJ: Prentice Hall.
Available at: https://www.researchgate.net/publication/344973647_Experiential_Learning_Experience_As_The_Source_Of_Learning_And_Development

Huang, Q. et al. (2023)
'Large Language Models for Cybersecurity: A Survey', *arXiv*.
Available at: https://arxiv.org/abs/2307.09055

D'Amico, A.D. et al. (2005)
'The Advanced Security Operations Center: A study of human factors', *CHI Extended Abstracts*.
Available at:

https://www.researchgate.net/publication/221518346_The_Advanced_Security_Operations_Center_A_Study_of_Human_Factors

Applebaum, S., Miller, D., Strom, B., Korban, C. and Wolf, R. (2016)
Intelligent, Automated Cyber Defense Using Reinforcement Learning.
MITRE Technical Report.
Available at: https://www.mitre.org/sites/default/files/publications/pr-16-0023-intelligent-automated-cyber-defense-using-reinforcement-learning.pdf

APPENDICES

Appendix A - System Implementation Evidence

This appendix provides visual evidence of the implemented LLM-Guided Reinforcement Learning cyber incident response training simulator. It documents the system interface, analyst interaction flow, automated scoring mechanism, and PDF-based training report generation, thereby demonstrating the practical implementation, usability, and professional relevance of the developed artefact.

Core Logic Snippets:

```
if action == "isolate_host" and self.attacker_present:
    reward += 15
    self.attacker_present = False
elif action == "close_incident" and self.attacker_present:
    reward -= 20
```

Snippet 1: Environment Reward Evaluation Logic

```
story.append(Paragraph(f"Score: {score}/100", styles["BodyText"]))
story.append(Paragraph(f"Readiness Level: {level}", styles["BodyText"]))
doc.build(story)
```

Snippet 2: Automated Training Report Generation Logic

System Implementation Evidence:

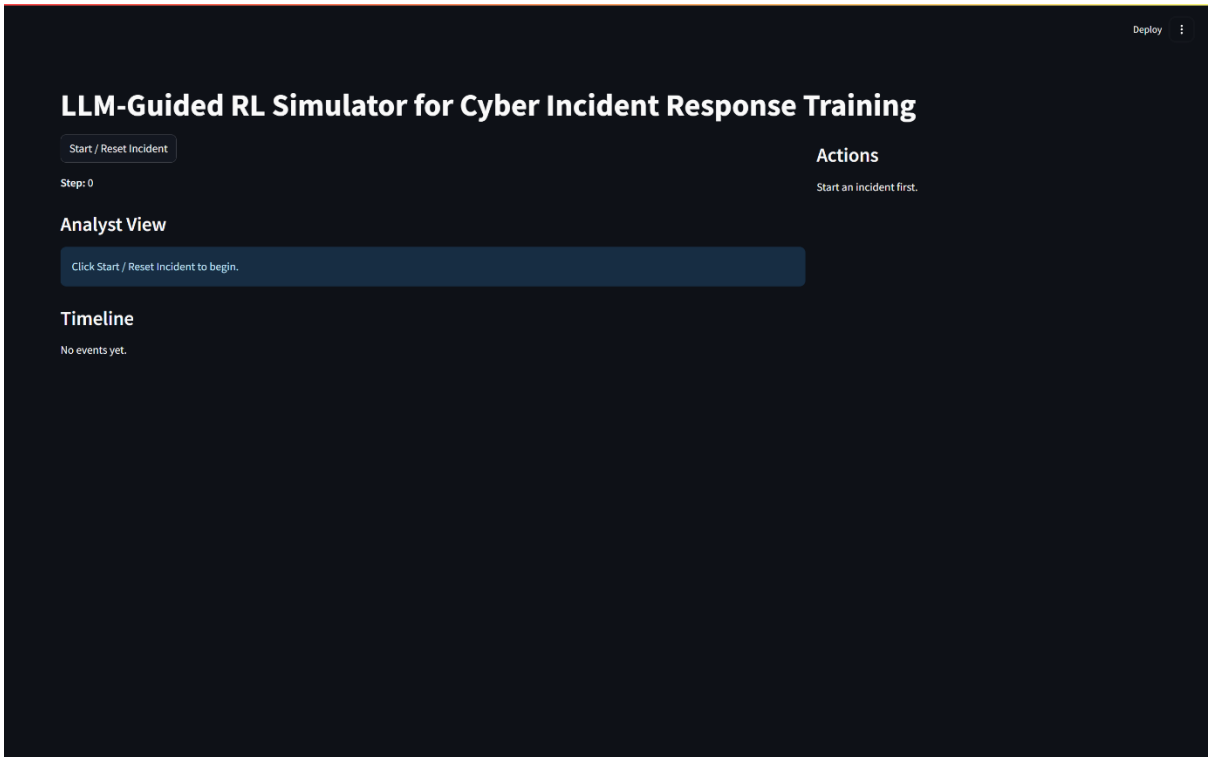

Figure 1 - Streamlit SOC Simulator Interface

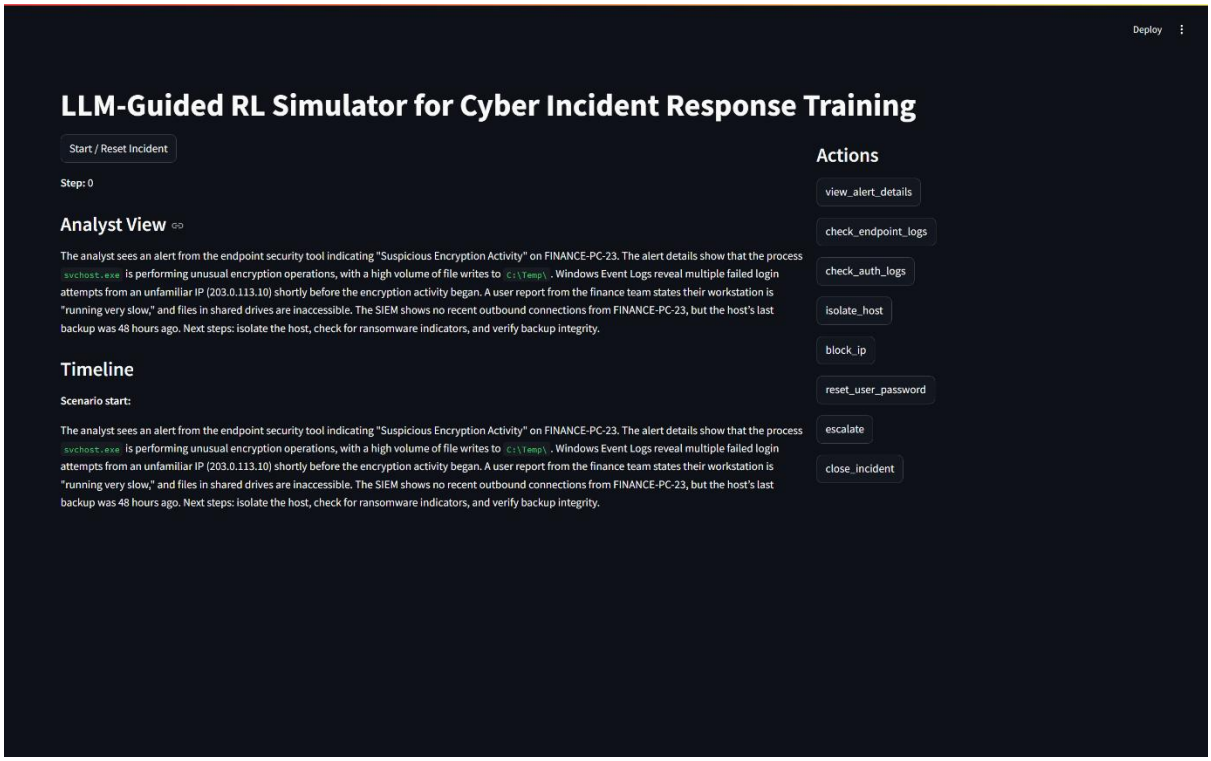

Figure 2 - Analyst Action Selection and Narrative Output

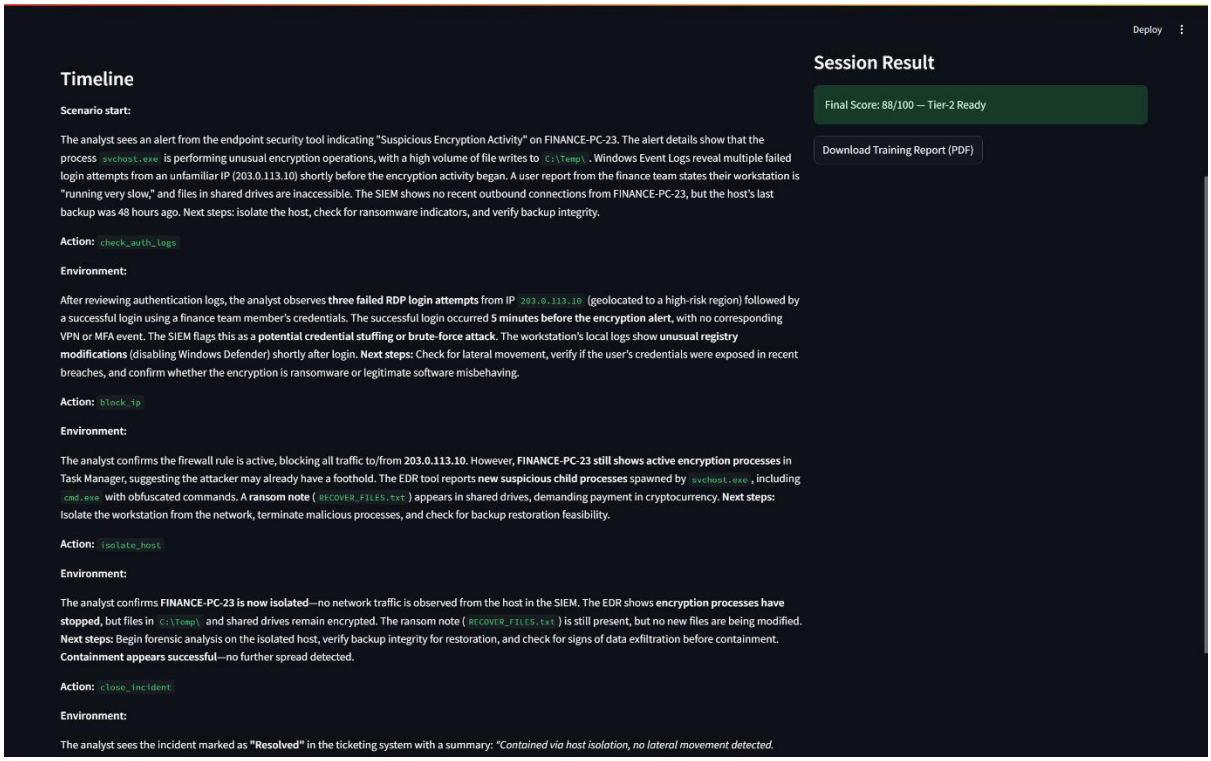

Figure 3 - User Readiness Score Output

Cyber Incident Response Training Report

Date: 2025-12-29 16:53:56

Scenario Summary

The analyst sees an alert from the endpoint security tool indicating "Suspicious Encryption Activity" on FINANCE-PC-23. The alert details show that the process `svchost.exe` is performing unusual encryption operations, with a high volume of file writes to `C:\Temp\`. Windows Event Logs reveal multiple failed login attempts from an unfamiliar IP (203.0.113.10) shortly before the encryption activity began. A user report from the finance team states their workstation is "running very slow," and files in shared drives are inaccessible. The SIEM shows no recent outbound connections from FINANCE-PC-23, but the host's last backup was 48 hours ago. Next steps: isolate the host, check for ransomware indicators, and verify backup integrity.

Action Timeline

Scenario start:

The analyst sees an alert from the endpoint security tool indicating "Suspicious Encryption Activity" on FINANCE-PC-23. The alert details show that the process `svchost.exe` is performing unusual encryption operations, with a high volume of file writes to `C:\Temp\`. Windows Event Logs reveal multiple failed login attempts from an unfamiliar IP (203.0.113.10) shortly before the encryption activity began. A user report from the finance team states their workstation is "running very slow," and files in shared drives are inaccessible. The SIEM shows no recent outbound connections from FINANCE-PC-23, but the host's last backup was 48 hours ago. Next steps: isolate the host, check for ransomware indicators, and verify backup integrity.

Figure 4 - Automated PDF Training Report Export

Figure 5 - Project Folder Structure

Appendix B - Sample Training Report

Cyber Incident Response Training Report

Date: 2025-12-29 16:53:56

Scenario Summary

The analyst sees an alert from the endpoint security tool indicating "Suspicious Encryption Activity" on FINANCE-PC-23. The alert details show that the process `svchost.exe` is performing unusual encryption operations, with a high volume of file writes to `C:\Temp\`. Windows Event Logs reveal multiple failed login attempts from an unfamiliar IP (203.0.113.10) shortly before the encryption activity began. A user report from the finance team states their workstation is "running very slow," and files in shared drives are inaccessible. The SIEM shows no recent outbound connections from FINANCE-PC-23, but the host's last backup was 48 hours ago. Next steps: isolate the host, check for ransomware indicators, and verify backup integrity.

Action Timeline

Scenario start:

The analyst sees an alert from the endpoint security tool indicating "Suspicious Encryption Activity" on FINANCE-PC-23. The alert details show that the process `svchost.exe` is performing unusual encryption operations, with a high volume of file writes to `C:\Temp\`. Windows Event Logs reveal multiple failed login attempts from an unfamiliar IP (203.0.113.10) shortly before the encryption activity began. A user report from the finance team states their workstation is "running very slow," and files in shared drives are inaccessible. The SIEM shows no recent outbound connections from FINANCE-PC-23, but the host's last backup was 48 hours ago. Next steps: isolate the host, check for ransomware indicators, and verify backup integrity.

1. Action: check_auth_logs

Outcome: After reviewing authentication logs, the analyst observes **three failed RDP login attempts** from IP `203.0.113.10` (geolocated to a high-risk region) followed by a successful login using a finance team member's credentials. The successful login occurred **5 minutes before the encryption alert**, with no corresponding VPN or MFA event. The SIEM fl...

2. Action: block_ip

Outcome: The analyst confirms the firewall rule is active, blocking all traffic to/from **203.0.113.10**. However, **FINANCE-PC-23 still shows active encryption processes** in Task Manager, suggesting the attacker may already have a foothold. The EDR tool reports **new suspicious child processes** spawned by `svchost.exe`, including `cmd.exe` with obfuscate...

3. Action: isolate_host

Outcome: The analyst confirms **FINANCE-PC-23 is now isolated**—no network traffic is observed from the host in the SIEM. The EDR shows **encryption processes have stopped**, but files in `C:\Temp\` and shared drives remain encrypted. The ransom note (`RECOVER_FILES.txt`) is still present, but no new files are being modified. **Next steps:** Begin forensic ...

4. Action: close_incident

Outcome: The analyst sees the incident marked as **"Resolved"** in the ticketing system with a summary: *"Contained via host isolation, no lateral movement detected. Encrypted files restored from offline backups. User educated on phishing risks."* The EDR shows **no active threats** on FINANCE-PC-23 post-rebuild, and the SIEM confirms **no further alerts** ...

Final Score

Score: 88/100

Readiness Level: Tier-2 Ready

Figure 6 - Incident_Report_1.pdf

Cyber Incident Response Training Report

Date: 2026-01-05 13:51:19

Scenario Summary

The analyst observes multiple failed login alerts from the SIEM for user "jdoe" (HR dept), originating from a Russian IP (unusual for this account). A successful login follows 15 minutes later from the same IP. The EDR shows a new PowerShell process spawned from the login session. User "jdoe" has reported via ticket: "My account was locked, then unlocked itself—I didn't request this." The SOC playbook suggests isolating the host and resetting credentials. Next steps: Check VPN logs for anomalies and review recent email activity for phishing signs.

Action Timeline

Scenario start:

The analyst observes multiple failed login alerts from the SIEM for user "jdoe" (HR dept), originating from a Russian IP (unusual for this account). A successful login follows 15 minutes later from the same IP. The EDR shows a new PowerShell process spawned from the login session. User "jdoe" has reported via ticket: "My account was locked, then unlocked itself—I didn't request this." The SOC playbook suggests isolating the host and resetting credentials. Next steps: Check VPN logs for anomalies and review recent email activity for phishing signs.

1. Action: check_auth_logs

Outcome: The analyst sees **12 failed login attempts** for `jdoe` within 5 minutes, all from `91.214.123.45` (Russia), followed by a successful login at `14:27 UTC`. The SIEM flags this as a **geo-velocity anomaly**—previous logins were from New York. The EDR shows the successful session spawned `powershell.exe -nop -w hidden -enc [long_base64_string]`. A *...

2. Action: block_ip

Outcome: The analyst confirms the firewall rule is active, blocking `91.214.123.45` and associated domain `malicious[.]xyz`. The SIEM shows no new login attempts from the blocked IP, but the compromised session (`jdoe`) remains active with intermittent PowerShell callbacks to a new C2 domain (`updates[.]secure-cloud[.]biz`). EDR alerts indicate **persistent...

3. Action: view_alert_details

Outcome: The analyst sees the SIEM alert details showing the initial compromise at **14:27 UTC**, followed by **outbound connections to `updates[.]secure-cloud[.]biz`** every 5 minutes. The EDR reveals **two new scheduled tasks** created by the attacker: one for persistence (`\Microsoft\Windows\Update\Run`) and another for data exfiltration (`\Microsoft\Win...

4. Action: isolate_host

Outcome: The analyst confirms the host (`WORKSTATION-HR-07`) is now **isolated**—no network traffic in/out per EDR. The active attacker session (`jdoe`) is terminated, but **two new alerts** appear: a failed RDP attempt from the same host to `FILESERVER-01` (logged at **14:42 UTC**) and a **suspicious SMB connection** to `PRINTSERVER-03`. The SIEM shows **n...

5. Action: escalate

Outcome: The Tier 2 analyst reviews the escalation and sees **new alerts** for **unusual file deletions** on `FILESERVER-01` (HR shares) and **suspicious `net.exe` usage** from the isolated host. The SIEM timeline shows the attacker pivoted to `PRINTSERVER-03` before isolation. The IR team notes: *"Possible data staging—check for encrypted archives or exfil...

6. Action: close_incident

Outcome: The analyst sees the incident marked as **"Closed - False Positive"** in the ticketing system, but new alerts fire: **`ryuk.exe` detected on `FILESERVER-01`** and **mass file renames** (`.locked` extensions). The SIEM shows the attacker pivoted via **stolen admin credentials** (from `PRINTSERVER-03`). The user reports: *"All HR files are now encryp...

Final Score

Score: 56/100

Readiness Level: Junior SOC

Learning Feedback

- Too many risky actions taken.

Figure 7 - Incident_Report_2.pdf

Cyber Incident Response Training Report

Date: 2026-01-05 13:57:23

Scenario Summary

The analyst sees an alert from the endpoint security tool indicating "Suspicious Encryption Activity" on FINANCE-PC-23. The alert details show that the process `msiexec.exe` spawned an unusual child process `cipher.exe` with command-line arguments suggesting file encryption. Windows Event Logs reveal multiple file modification events in the user's "Documents" folder, with files now having a `.locked` extension. A user from the finance department has reported via the ticketing system that they cannot open their files and see a ransom note named `RECOVER_FILES.txt` on their desktop. The SIEM shows no recent network connections to known malicious IPs from this host, but the endpoint agent indicates that `msiexec.exe` made an outbound connection to an unfamiliar domain `update-cloud[.]net` over HTTPS. The analyst should investigate the process tree, check for additional compromised hosts, and isolate FINANCE-PC-23 if ransomware is confirmed.

Action Timeline

Scenario start:

The analyst sees an alert from the endpoint security tool indicating "Suspicious Encryption Activity" on FINANCE-PC-23. The alert details show that the process `msiexec.exe` spawned an unusual child process `cipher.exe` with command-line arguments suggesting file encryption. Windows Event Logs reveal multiple file modification events in the user's "Documents" folder, with files now having a `.locked` extension. A user from the finance department has reported via the ticketing system that they cannot open their files and see a ransom note named `RECOVER_FILES.txt` on their desktop. The SIEM shows no recent network connections to known malicious IPs from this host, but the endpoint agent indicates that `msiexec.exe` made an outbound connection to an unfamiliar domain `update-cloud[.]net` over HTTPS. The analyst should investigate the process tree, check for additional compromised hosts, and isolate FINANCE-PC-23 if ransomware is confirmed.

1. Action: view_alert_details

Outcome: The analyst sees the full alert details in the SIEM, confirming the `msiexec.exe` parent process spawned `cipher.exe` at **10:42:17 AM**, followed by rapid file modifications in `C:\Users\FinanceUser\Documents\`. The process tree shows `msiexec.exe` was launched from a temporary folder (`C:\Users\FinanceUser\AppData\Local\Temp\update.exe`), which i...

2. Action: block_ip

Outcome: The analyst confirms the firewall rule is now active, blocking all traffic to `update-cloud[.]net` (IP: **185.143.223[.]42**). The endpoint agent on **FINANCE-PC-23** shows no new outbound connections to the domain since the block was applied. However, the workstation remains online, and the user reports no change in file accessibility—the ransom n...

3. Action: reset_user_password

Outcome: The analyst confirms the password reset for **FinanceUser** was successful, but the ransomware encryption persists—files remain locked, and the ransom note is unchanged. The endpoint agent shows no new malicious processes, but **FINANCE-PC-23** is still active on the network. The SIEM reveals no additional authentication attempts from the user's ac...

4. Action: close_incident

Outcome: The analyst sees the incident ticket marked as "Closed" in the system, but **FINANCE-PC-23** remains online with encrypted files and an active ransom note. The SIEM shows no recent alerts from the host, but the endpoint agent's last scan still flags the presence of `RECOVER_FILES.txt` and modified file extensions. A follow-up user report indicates ...

Final Score

Score: 74/100

Readiness Level: Tier-1 Ready

Learning Feedback

- Too many risky actions taken.

Figure 8 - Incident_Report_3.pdf