Scholars Journal of Engineering and Technology
OPEN ACCESS
Abbreviated Key Title: Sch J Eng Tech
ISSN 2347-9523 (Print) | ISSN 2321-435X (Online)
Journal homepage: https://saspublishers.com

Decision-Centric Cybersecurity: The Role of Human-in-the-Loop Machine Learning

Haris Bin Abrar[1*], Amir Azam[2], Tabish Bin Abrar[3], Muhammad Amir khan[4]

[1]Department of Information and Communication Engineering, University of Science and Technology Beijing, China
[2]Department of Electronics, Quaid e Azam University Islamabad, Pakistan
[3]Department of Artificial Intelligence, Bahria University Islamabad, Pakistan
[4]Department of Physics, Government post graduate college Nowshera Abdul wali khan university Mardan (AWKUM)

DOI: https://doi.org/10.36347/sjet.2026.v14i02.001 | **Received:** 29.10.2025 | **Accepted:** 06.01.2026 | **Published:** 06.02.2026

***Corresponding author:** Haris Bin Abrar
Department of Information and Communication Engineering, University of Science and Technology Beijing, China

Abstract **Review Article**

The increasing complexity of cyber threats and the limitations of traditional cybersecurity systems have spurred the development of more adaptive and intelligent approaches. Decision-centric cybersecurity, which integrates Human-in-the-Loop (HITL) systems with machine learning (ML), has emerged as a promising solution. This review explores the role of HITL in machine learning models for cybersecurity, emphasizing the importance of combining the speed and scalability of automation with the contextual judgment and ethical considerations provided by human experts. The review covers the types of machine learning techniques commonly used in cybersecurity, such as supervised, unsupervised, and reinforcement learning, and discusses their strengths and weaknesses in addressing modern cyber threats. We also examine the challenges of integrating HITL into cybersecurity systems, including human error, scalability issues, and ethical concerns. The future of decision-centric cybersecurity lies in enhancing machine learning algorithms, improving explainability, and developing more autonomous systems, while still maintaining the crucial role of human oversight. Ultimately, this review highlights the collaborative potential of human expertise and machine learning in creating more effective, adaptable, and ethical cybersecurity defences in the face of evolving digital threats.

Keywords: Decision-Centric Cybersecurity, Human-in-the-Loop, Machine Learning, Cybersecurity, Autonomous Systems, Ethical Decision-Making, Supervised Learning, Unsupervised Learning, Reinforcement Learning.

1. INTRODUCTION

1.1 Background of Cybersecurity

As digital systems become more integral to nearly every aspect of modern life, cybersecurity has become a crucial concern. The increasing prevalence of cyber-attacks ranging from data breaches and ransomware attacks to more sophisticated state-sponsored intrusions has made securing digital infrastructures a top priority. The consequences of these attacks are far-reaching, affecting businesses, governments, and individuals alike. Traditional cybersecurity methods, often relying on signature-based detection, have proven insufficient to combat the constantly evolving nature of cyber threats. As new and more sophisticated attack methods emerge, organizations must adapt by incorporating more advanced tools and technologies that can automatically detect, respond, and mitigate these risks in real time [1].

1.2 Importance of Machine Learning in Cybersecurity

Machine learning (ML) has emerged as a game-changer in cybersecurity. Its ability to analyse vast amounts of data and detect patterns whether in network traffic, user behaviour, or software anomalies enables it to identify potential threats that might otherwise go unnoticed. In fact, machine learning has already been successfully implemented in intrusion detection systems (IDS), anomaly detection models, and malware analysis tools. ML offers distinct advantages over traditional rule-based systems, as it can adapt to new threats, improve over time, and even detect previously unknown threats through its ability to learn from historical data. However, despite its benefits, machine learning cannot fully replace human judgment in complex decision-making scenarios. It often requires the guidance of human experts to ensure accuracy and prevent false positives or the misinterpretation of data [2].

1.3 Introduction to Human-in-the-Loop (HITL) Approaches

While machine learning brings a high degree of automation to cybersecurity, it also faces limitations. One of the major challenges is that ML algorithms often lack the contextual awareness and judgment that human experts can bring to decision-making. This is where Human-in-the-Loop (HITL) approaches come into play. HITL refers to a system where human input is integrated into the decision-making process, especially when dealing with ambiguous, complex, or sensitive situations that require ethical considerations or domain-specific knowledge. In the context of cybersecurity, HITL can help mitigate the weaknesses of fully automated systems by providing the nuanced decisions needed when automated systems encounter unfamiliar scenarios or face unprecedented threats. The integration of human expertise into ML models enhances their ability to make accurate, context-aware decisions and improve response strategies in real-time [3].

1.4 Objective of the Review

The objective of this review is to explore the role of Human-in-the-Loop (HITL) Machine Learning in decision-centric cybersecurity. By examining how human intervention enhances machine learning models, we aim to identify the strengths, challenges, and future potential of HITL-enhanced systems in improving the effectiveness of cybersecurity operations. This review will focus on how HITL can be integrated into decision-making processes to optimize threat detection, response accuracy, and overall cybersecurity resilience. Additionally, the review will discuss the emerging trends in decision-centric cybersecurity, the integration of human judgment with machine learning, and the ethical implications of such systems.

2. Overview of Decision-Centric Cybersecurity

2.1 Definition and Key Concepts

Decision-centric cybersecurity refers to a paradigm in which decision-making in cybersecurity operations is not left solely to automated systems but is rather enhanced by human input and contextual understanding. This approach allows for more flexible and adaptable responses to emerging threats, especially in cases where traditional rule-based systems may fail. In decision-centric systems, machine learning algorithms are used to automate threat detection and generate possible responses, but human experts provide critical oversight and fine-tuning of decisions based on their expertise, context, and judgment.

The key concepts in decision-centric cybersecurity revolve around the integration of machine intelligence with human expertise, ensuring that decisions made in response to cyber threats are both data-driven and contextually informed. The collaborative decision-making framework makes it possible to combine the efficiency of automated systems with the nuanced judgment of human operators. This approach emphasizes flexibility, accuracy, and human accountability in critical cybersecurity decisions [4].

2.2 Challenges in Traditional Cybersecurity Systems

Traditional cybersecurity systems have several inherent limitations that hinder their effectiveness in addressing modern, sophisticated cyber threats. The most common challenge is their reliance on signature-based detection and rule-based systems, which are often ineffective against new, unknown threats. These systems can only identify threats they have previously encountered, meaning they are incapable of detecting novel or zero-day attacks that exploit previously unknown vulnerabilities. As a result, attackers can bypass such systems, taking advantage of gaps in detection capabilities.

Another challenge of traditional systems is that they often lack the adaptive capabilities needed to keep up with the constantly evolving nature of cyber threats. While firewalls, intrusion detection systems (IDS), and antivirus software can provide some level of protection, they generally operate based on predefined signatures or behavioural patterns. When confronted with new, advanced threats, such as those involving encrypted communications or zero-day exploits, these systems struggle to respond effectively.

Moreover, traditional systems often rely on manual interventions from cybersecurity professionals, which can delay response times and lead to human error. The need for continuous monitoring and the inability to process large amounts of real-time data further exacerbates the issue, especially in high-stakes environments where immediate action is critical. As a result, many cybersecurity teams face overwhelming amounts of data without the tools to efficiently analyse and respond to it in real time [5].

2.3 How Decision-Centric Cybersecurity is Changing the Landscape

Decision-centric cybersecurity represents a significant shift away from traditional systems toward more dynamic and intelligent solutions. By integrating human-in-the-loop (HITL) processes with machine learning models, decision-centric systems allow for real-time, adaptive responses to cyber threats. Rather than relying on predefined signatures, these systems leverage machine learning algorithms to detect anomalies and patterns indicative of potential attacks. The human element ensures that these algorithms can be corrected and refined when faced with complex, high-stakes decisions that involve ethical considerations or when contextual knowledge is crucial.

In decision-centric systems, human experts are no longer just passive monitors but active participants in the decision-making process. Cybersecurity professionals work alongside machine learning models to provide the necessary expertise and context to accurately assess threats. This approach can lead to faster, more accurate threat detection and response times, while still considering the complexities of human judgment. By enabling continuous learning from both machine models and human decisions, decision-centric systems improve their effectiveness over time, adapting to new threats as they emerge.

Furthermore, decision-centric cybersecurity systems are highly scalable and can be integrated into a wide variety of cybersecurity frameworks. Whether it’s a network intrusion detection system, an endpoint protection system, or a cloud-based security solution, decision-centric models offer a level of flexibility and adaptability that traditional systems cannot match. These systems provide a more holistic approach to cybersecurity, combining the strengths of automation, machine learning, and human expertise to deliver comprehensive protection against modern cyber threats [6].

3. The Role of Machine Learning in Cybersecurity

3.1 Types of Machine Learning Used in Cybersecurity

Machine learning (ML) plays a central role in automating threat detection, improving incident response, and creating more intelligent cyber defence mechanisms. The various types of machine learning techniques supervised, unsupervised, and reinforcement learning are being applied across a wide range of cybersecurity use cases. Each type of ML provides unique strengths and capabilities that address different aspects of cybersecurity.

3.1.1 Supervised Learning:

Supervised learning in cybersecurity is often used for classification tasks, where known attack patterns are mapped to specific categories. For example, intrusion detection systems (IDS) use supervised learning to differentiate between legitimate network activity and potential threats like DDoS (Distributed Denial of Service) attacks or phishing attempts. A classic supervised learning algorithm, such as the Support Vector Machine (SVM), is trained on labelled examples of benign and malicious traffic. Over time, the model gets better at identifying patterns associated with threats.

In other applications, supervised learning models are used to predict malware behaviour by analysing the features of previously identified malicious software and correlating them with new files observed in the system.

3.1.2 Unsupervised Learning:

Unsupervised learning is crucial in cybersecurity because it allows systems to detect novel attacks that have never been seen before, such as zero-day vulnerabilities. Unlike supervised learning, unsupervised learning doesn't require labelled data for training. Instead, it attempts to detect outliers or anomalies in network traffic, user behaviour, or system logs. One widely used unsupervised technique in cybersecurity is K-means clustering, where the algorithm identifies unusual patterns based on feature similarities.

An example of unsupervised learning in cybersecurity is its application to network behaviour analysis. Unsupervised algorithms analyse normal network behaviour and flag any activity that deviates significantly from established patterns as a potential threat. This method helps identify advanced persistent threats (APTs) that use stealthy, low-profile techniques to evade detection by traditional systems.

3.1.3 Reinforcement Learning:

Reinforcement learning (RL) is a machine learning technique where an agent learns how to make decisions by interacting with its environment and receiving feedback in the form of rewards or penalties. In the context of cybersecurity, RL is used to develop adaptive security systems that can automatically respond to cyber threats. For example, RL algorithms can train an agent to mitigate a DDoS attack by analysing past responses and choosing the most effective defense strategies based on the threat landscape.

In autonomous defense systems, RL agents continuously improve their decision-making by testing various security responses (e.g., isolating a compromised server or blocking suspicious traffic) and learning which actions lead to the best outcomes. Over time, these systems become more efficient and capable of handling a broader range of cyberattacks in real-time.

3.2 Advantages and Limitations of Machine Learning in Cybersecurity (Expanded)

While machine learning has proven to be a powerful tool for cybersecurity, it also faces several challenges that must be addressed to maximize its effectiveness.

Advantages:

1. **Adaptability**:

One of the biggest advantages of machine learning in cybersecurity is its ability to adapt to new, emerging threats. As cyber attackers become more sophisticated, the ability to train models on new attack data enables machine learning systems to recognize patterns in previously unseen threats. For example, machine learning-based intrusion detection systems (IDS) can detect new types of attacks without requiring manual updates to the system [8].

2. **Automation**:

Cybersecurity systems that rely on machine learning can automate tasks such as data analysis, threat detection, and even incident response. This reduces the workload on human security professionals, allowing them to focus on more strategic decisions. Automation also leads to faster detection and response times, which is crucial in mitigating the damage caused by cyberattacks.

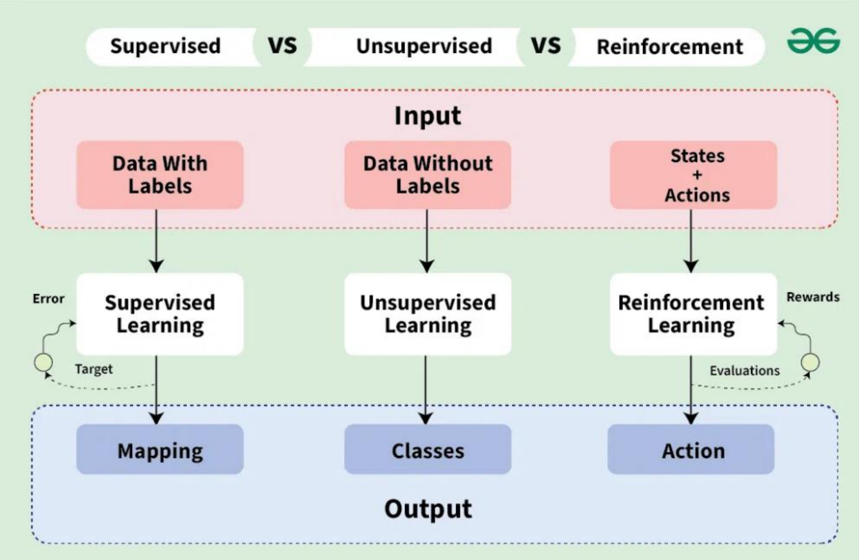

Fig1. Supervised vs Reinforcement vs Unsupervised [7]

3. **Real-time Threat Detection**:

ML models excel in providing real-time detection of threats by continuously analyzing incoming data. Unlike traditional systems, which may require manual updates, ML systems can adjust and learn from incoming traffic or system logs continuously. This is especially useful for detecting zero-day attacks, advanced persistent threats (APTs), and insider threats that are typically difficult to identify using conventional methods [9].

4. **Scalability**:

Machine learning models are highly scalable, capable of processing vast amounts of data from diverse sources, such as network traffic, system logs, and user behaviors. This makes them ideal for large-scale cybersecurity systems that need to monitor millions of endpoints or users.

Limitations:

1. **False Positives**:

One of the major drawbacks of machine learning in cybersecurity is the risk of false positives. The system may incorrectly flag benign activities as malicious, leading to wasted resources and potential disruption in normal operations. Fine-tuning models to reduce false positives while maintaining high detection accuracy is a constant challenge.

2. **Adversarial Attacks**:

Machine learning systems are vulnerable to adversarial attacks, where malicious actors modify input data (such as network packets or image data) to deceive the machine learning model. Adversarial attacks can cause the model to misclassify data, allowing attackers to bypass detection systems undetected.

3. **Data Quality and Bias**:

The performance of machine learning models is heavily dependent on the quality of the data used for training. Incomplete, biased, or unrepresentative data can result in models that perform poorly or are unable to recognize certain attack patterns. Moreover, biased training data could inadvertently reinforce existing security vulnerabilities or discrimination in decision-making [10].

4. **Computational Overhead**:

While machine learning offers powerful capabilities, it also requires significant computational resources. Training complex models such as deep neural networks can be time-consuming and resource-intensive, potentially creating performance bottlenecks in real-time cybersecurity applications.

3.3 Case Studies and Applications of Machine Learning in Cybersecurity

Several real-world examples illustrate how machine learning has been applied successfully to cybersecurity:

3.3.1 Intrusion Detection Systems (IDS)

Machine learning-based intrusion detection systems (IDS) are now widely used in network security. Traditional IDS often rely on pre-defined rules and signatures, which can be bypassed by sophisticated attacks. However, ML-based IDS systems use supervised and unsupervised learning to analyze network traffic and detect anomalies in real-time. For example, Snort and Suricata are popular IDS platforms that use machine learning algorithms to classify network packets and flag potential attacks.

3.3.2 Malware Detection

Machine learning is transforming malware detection by moving beyond simple signature-based detection. Deep learning models, such as convolutional neural networks (CNNs), are used to analyze file behaviour and system processes to detect new and evolving malware. Deep Instinct, for example, uses deep learning to detect malware before it can execute on a system, significantly reducing the risk of infection.

3.3.3 Phishing Detection

Phishing is a major cybersecurity threat, with attackers constantly evolving their tactics to deceive users. Machine learning models are now used to detect phishing attempts by analyzing email content, URLs, and sender behaviour. Algorithms trained on large datasets of known phishing attempts can flag suspicious emails, helping organizations protect against data breaches and social engineering attacks.

3.3.4 Automated Threat Response

Machine learning can also play a key role in automated threat response. For instance, reinforcement learning (RL) agents have been developed to autonomously decide how to mitigate a cyberattack based on real-time data. In DDoS protection, RL agents can adapt their response strategy depending on the severity and type of attack, thereby reducing reliance on manual interventions.

4. Human-in-the-Loop (HITL) Approaches

4.1 Definition and Components of HITL

Human-in-the-Loop (HITL) refers to an approach where human judgment is integrated into the decision-making process, particularly in automated systems. In cybersecurity, HITL allows human experts to interact with machine learning models to provide valuable input, feedback, and corrections when faced with complex, ambiguous, or ethically challenging decisions. The key components of HITL systems include:

- **Human Input**:

Cybersecurity experts provide critical insights and contextual understanding that machines may lack. This human contribution is especially important when dealing with unknown threats or when machine learning models produce uncertain results.

- **Machine Learning Models**:

These models process data, detect anomalies, and generate automated decisions or recommendations, which are then presented to human operators for evaluation and refinement.

- **Feedback Loop**:

Once the human expert decides, the system learns from this feedback and adapts its algorithms accordingly. This iterative process improves the performance of the cybersecurity system over time.

HITL systems aim to combine the speed and scalability of automation with the contextual awareness and ethical judgment provided by human operators. This hybrid approach helps create more robust and adaptive cybersecurity frameworks [11].

4.2 The Role of Human Judgment in Automated Cybersecurity Systems

Human judgment is crucial in cybersecurity, particularly in situations where automated systems may not fully understand the context or nuances of a specific attack. For example, machine learning models may detect anomalous behavior in a system, but human experts are needed to interpret whether this behavior is truly malicious or just a false alarm.

Moreover, ethical considerations often come into play in cybersecurity decisions. For instance, automated systems might flag a user's private data as potentially compromised, but only a human expert can evaluate the ethical implications of accessing or modifying this data. Similarly, machine learning models might propose drastic measures, such as disconnecting a server from the network or blocking specific user access, but a human judgment call is required to assess the business impact of these actions.

Human operators bring to the table domain expertise that helps the system prioritize threat severity and mitigation strategies. In incident response, while automated systems can react quickly, humans can make informed decisions based on their understanding of the organization’s specific risks, business priorities, and operational requirements [12].

4.3 Benefits and Challenges of HITL in Cybersecurity

The integration of human oversight into machine learning models brings several benefits to cybersecurity systems:

Benefits:

1. **Improved Accuracy**:

HITL systems help improve the accuracy of decision-making by allowing human experts to correct or refine the model's recommendations. This is particularly useful in complex attack scenarios where human expertise is essential to interpreting ambiguous or conflicting data.

2. **Contextual Understanding**:

Humans can provide contextual insights that machines might miss. For example, corporate policies or regulatory requirements can guide human decision-making in a way that is often outside the scope of a machine learning model’s training data.

3. **Ethical Oversight**:

HITL allows for ethical decision-making, where human judgment ensures that data privacy and personal rights are respected during incident response.

This is especially important in industries like healthcare or finance, where the implications of decisions can extend beyond security to legal and ethical concerns.

Challenges:

1. **Human Error**:

One of the main challenges of HITL systems is the risk of human error. Despite their expertise, human decision-makers are still subject to biases, distractions, and cognitive limitations that can impact the effectiveness of the cybersecurity system. For example, overconfidence bias may cause an expert to ignore a potential threat flagged by the system.

2. **Scalability**:

As the volume of cybersecurity incidents increases, it becomes difficult for human operators to process and evaluate every alert. Even in decision-centric systems, the involvement of humans may slow down response times if not properly managed. This is a critical concern in real-time cyber defense.

3. **Resource Intensity**:

While HITL systems improve decision-making, they are also resource intensive. Maintaining a team of qualified cybersecurity experts to evaluate machine learning outputs and provide feedback requires significant personnel and time investment. As organizations scale, managing this balance between automation and human oversight becomes more challenging [13].

4.4 Examples of HITL in Decision-Making for Cybersecurity

HITL approaches have been implemented in a variety of real-world cybersecurity systems, demonstrating their potential to enhance decision-making. Here are a few examples:

4.4.1 Security Information and Event Management (SIEM):

Many organizations use SIEM tools to monitor network traffic and security events. Machine learning algorithms in these tools can automatically flag suspicious behaviour, such as unauthorized data access or unusual login times. However, the final decision on whether to block an IP address or escalate the issue to incident response teams is often made by a human operator. This ensures that automated alerts are evaluated in the context of the organization's specific security policies and operational needs.

4.4.2 Automated Malware Analysis:

Deep Instinct uses deep learning algorithms to analyze files and identify potential malware. Once the model flags a file as suspicious, a cybersecurity expert may review the file's behaviour in a sandbox environment before deciding whether to isolate it or delete it. This human judgment helps avoid false positives while ensuring the system doesn't miss actual threats.

4.4.3 Phishing Email Detection:

Google Safe Browsing uses machine learning to detect phishing websites by analyzing URL patterns and webpage content. However, when a potential phishing attack is flagged, a human reviewer is often involved to assess the context of the email and ensure the alert is valid. This HITL process improves the accuracy of phishing detection and minimizes the risk of false positives [14].

5. Decision-Making in Cybersecurity: Integrating Human and Machine

5.1 How Machine Learning and HITL Collaborate in Decision-Centric Systems

In decision-centric cybersecurity, the integration of machine learning (ML) with Human-in-the-Loop (HITL) approaches forms a hybrid decision-making process. Machine learning algorithms excel at detecting patterns in large datasets and providing initial predictions or recommendations based on past data, but they often lack the contextual understanding and judgment that human experts bring. This is where the HITL framework plays a crucial role.

Machine learning models are typically employed to automate data analysis and identify potential security threats in real-time. These systems can detect anomalies, identify malicious patterns, and classify threats more efficiently than traditional systems. However, machine learning systems alone are prone to errors, such as false positives, and they may struggle with ambiguous situations that require interpretation or human judgment.

The role of the human expert in decision-making is critical in cases where the automated system cannot conclusively assess the threat or when ethical or contextual factors need to be considered. In a HITL-enhanced system, the machine learning model's recommendations are reviewed by a cybersecurity expert, who provides additional context and makes the final decision. This feedback loop ensures that both automation and human input work together to enhance the accuracy and effectiveness of the response to cyber threats [15].

By combining machine learning's ability to process vast amounts of data with human oversight, decision-centric systems can achieve a balance of speed and accuracy, which is particularly important in fast-moving environments like cyber threat detection.

5.2 Decision-Making Models in HITL-Enhanced Cybersecurity Systems

The integration of HITL and machine learning into decision-making in cybersecurity involves several models that optimize threat detection and response.

These models focus on ensuring that human experts are involved at critical decision points where automation may not suffice.

5.2.1 Collaborative Filtering Model:

In a collaborative filtering model, machine learning algorithms use previous incident data to recommend actions based on similar past incidents. The system then presents these recommendations to the human expert, who can accept, modify, or reject them. This approach helps to automate the initial detection process while allowing human experts to provide the final input, ensuring that decisions are accurate and informed by experience. For instance, an intrusion detection system (IDS) might flag an unusual pattern in network traffic, and the machine learning system will suggest potential attack types based on historical data. The cybersecurity expert can review these suggestions, examine the context of the incident, and make a final decision on whether to escalate or contain the issue.

5.2.2 Decision Support Systems (DSS):

A Decision Support System (DSS) is another model commonly used in HITL cybersecurity systems. The DSS provides real-time data analysis and presents it to the human operator in an easily understandable format, such as dashboards or visualizations. The system might provide multiple recommendations or courses of action, from which the human expert can select the most appropriate one. For example, in the event of a data breach, a DSS might offer suggestions on how to isolate the affected system, initiate a forensic investigation, or alert higher authorities. The expert's judgment is used to select the best response, considering organizational policies, security protocols, and ethical considerations.

5.2.3 Hybrid Decision-Making Models:

Hybrid models combine both machine-driven and human-driven decision-making at various stages of the process. One example is the use of machine learning algorithms for initial detection, followed by human verification and final action. For instance, a machine learning system might detect a potential phishing email, and the system can automatically block the email or tag it for review. However, the final decision to act (such as isolating the email or reporting it to the security team) requires human input to evaluate the potential consequences and impact of the decision.

By combining human expertise with machine-driven analysis, these models ensure that decisions are made in real-time, balancing the speed of automation with the necessary contextual insight and ethical responsibility of human decision-makers [16].

5.3 Enhancing Accuracy and Response Time with HITL and Machine Learning

The integration of HITL with machine learning models leads to significant improvements in both accuracy and response time in cybersecurity systems.

5.3.1 Improved Accuracy:

Machine learning models are particularly strong at recognizing patterns and detecting known threats, such as malware or unauthorized access attempts. However, in situations where the data is ambiguous or complex, machine learning algorithms may struggle to make the best decision. For instance, in cases of advanced persistent threats (APTs), where attackers use sophisticated, low-profile tactics to infiltrate networks, machine learning models may miss subtle signs of compromise. In these scenarios, human oversight plays a critical role in improving the accuracy of the system's response.

For example, human experts can provide additional context about the nature of the attack, assess the organization's specific vulnerabilities, and make decisions that the machine might not be able to make autonomously. The collaborative feedback between humans and machines enhances the overall detection and response accuracy.

5.3.2 Faster Response Times:

In high-pressure environments, such as cybersecurity incident response, time is critical. HITL systems enable faster response times by automating much of the data analysis and initial threat detection, leaving humans to make informed decisions quickly. The machine learning models act as first responders, flagging potential threats, while the human operator makes final decisions about how to respond.

For instance, in the case of a distributed denial-of-service (DDoS) attack, machine learning models can automatically detect unusual network traffic and initiate protective measures like traffic throttling or IP blocking. The human expert can then assess the situation, refine the model's approach, and decide whether to escalate the issue, activate additional defenses, or notify stakeholders. This speed is essential in mitigating the damage from cyber-attacks, and HITL systems help ensure that critical decisions are made swiftly and accurately [17].

6. Challenges and Limitations

6.1 Human Error and Cognitive Biases in Cybersecurity Decisions

While Human-in-the-Loop (HITL) systems provide valuable oversight and contextual awareness, they also introduce the risk of human error. Cybersecurity experts, like any human decision-makers, are prone to cognitive biases and judgment errors that can impact the accuracy of decisions. Common biases such as confirmation bias (the tendency to favor information that supports existing beliefs), overconfidence bias (the tendency to overestimate one's own abilities), and anchoring bias (the tendency to rely too heavily on the first piece of information encountered) can distort a human expert's judgment in critical cybersecurity situations.

For instance, during a cyberattack or data breach, a human expert might interpret an anomaly in network traffic as a false positive, only to realize later that it was part of a sophisticated attack. These types of errors can lead to delays in the detection and mitigation of threats, potentially increasing the damage caused by the attack. Although machine learning can help minimize false positives and provide data-driven insights, human biases and mistakes remain a significant challenge in HITL-based systems.

Furthermore, human fatigue and stress can influence decision-making, especially in high-pressure environments where decisions need to be made quickly. Security teams often work in 24/7 shifts, which can lead to mental exhaustion and reduce their ability to make optimal decisions under stressful circumstances. Addressing this challenge requires a balance between automation and human input, ensuring that machine learning systems handle the bulk of repetitive and mundane tasks, leaving the human experts to focus on more complex, higher-level decisions [18].

6.2 Scalability and Computational Challenges

One of the major challenges of HITL systems in cybersecurity is scalability. As cybersecurity threats increase in complexity and volume, human involvement in decision-making becomes a bottleneck. Machine learning algorithms can analyze massive datasets and identifying patterns at speeds far beyond human capabilities. However, human decision-making cannot scale at the same rate as automated systems, especially when the volume of alerts or potential threats grows.

For example, large organizations with extensive digital infrastructures may face thousands of alerts or incidents per day. If human experts are required to review and validate each decision made by the machine learning system, it can overwhelm the system and lead to slower response times. This problem is particularly acute in situations where real-time action is needed to mitigate threats, such as DDoS attacks or zero-day vulnerabilities.

To address this issue, machine learning models must be designed to filter out less critical alerts and only escalate high-priority threats to human experts. This automated triage system allows for a balance between efficiency and accuracy, enabling human experts to focus on the most pressing issues while leaving routine tasks to the automated system. However, this requires significant investment in developing and maintaining scalable HITL systems that can effectively manage large datasets and high volumes of alerts [19].

Another computational challenge is that machine learning models, especially those involving deep learning or reinforcement learning, can be resource intensive. Training complex models requires substantial computational power and large labelled datasets. For real-time decision-making, models need to operate on low-latency infrastructure, which demands high-performance computing resources. While cloud services and distributed computing frameworks can help, the costs associated with running these systems at scale can be prohibitive for some organizations, especially smaller ones with limited budgets.

6.3 Privacy and Ethical Considerations in HITL Systems

HITL systems in cybersecurity raise several ethical concerns related to privacy and the potential for bias in decision-making. One of the primary ethical challenges is the privacy of individuals' data. In many cybersecurity systems, personal or sensitive information is collected and analysed to detect potential threats. For example, machine learning algorithms may analyze email content, website visits, or social media interactions to detect phishing attempts or insider threats. However, this data collection can raise concerns about user privacy, particularly when human experts are involved in making decisions about whether to act on the information.

There is a risk that personal data could be misused or inadvertently accessed by unauthorized personnel, especially when human operators are given direct access to sensitive information. In sectors like healthcare, banking, and government, the implications of violating data privacy regulations such as the GDPR (General Data Protection Regulation) can be severe, leading to financial penalties and damage to the organization's reputation.

Additionally, the integration of human input into cybersecurity decisions raises concerns about bias. Humans, like machine learning models, can be influenced by biases such as confirmation bias or groupthink. This can affect how certain threats are prioritized or handled. For example, in a HITL system, a human expert may give more attention to external cyber threats while downplaying insider threats due to cognitive biases. These biases can lead to discriminatory practices and unfair treatment in certain cases, particularly when automated systems are used to make decisions about security access or investigations.

To mitigate these privacy and ethical challenges, it is important for organizations to establish clear data privacy policies and ensure that human operators adhere to ethical guidelines when handling sensitive data. Moreover, training programs should be implemented to educate cybersecurity professionals on potential biases and how to minimize their impact on decision-making. Automated decision-making should be transparent and auditable to ensure that human decisions are made within ethical and legal frameworks [20].

7. Future Directions

7.1 Advancements in HITL and Machine Learning for Cybersecurity

The future of cybersecurity lies in the ongoing advancements in machine learning and Human-in-the-Loop (HITL) integration. As machine learning models become more sophisticated, they will continue to offer greater accuracy and speed in detecting and responding to cyber threats. However, for machine learning to reach its full potential in cybersecurity, human judgment will remain a critical component in addressing the ethical, contextual, and complex aspects of cyber threats.

7.1.1 Advances in Machine Learning Algorithms

Machine learning algorithms, especially in the field of deep learning, are becoming increasingly powerful. Future advancements may involve self-learning algorithms that can continuously update themselves based on new data, reducing the need for manual intervention and re-training. The use of neural networks, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), will likely become more common in cybersecurity, especially for image-based security data and real-time anomaly detection. These advancements will help address challenges like zero-day attacks, which require the detection of new vulnerabilities that have not been previously identified or catalogued [21].

7.1.2 Hybrid Decision-Making Systems

As machine learning models continue to evolve, the collaboration between humans and machines will become even more sophisticated. Hybrid decision-making systems, which integrate automated threat detection with human expertise, will become more seamless and adaptive. These systems will allow cybersecurity professionals to interact with algorithms in a way that enables real-time monitoring, rapid threat assessment, and quick decision-making in the face of evolving cyber threats. In these systems, machine learning models will perform initial threat assessments, while cybersecurity professionals will refine the results, make context-based decisions, and provide real-time feedback to improve the system's performance over time.

7.1.3 Explainability and Transparency in AI Models

Another significant development in the future of HITL and machine learning in cybersecurity will be the improvement in AI explainability. Machine learning models, particularly deep learning algorithms, are often viewed as "black boxes" because it is difficult to understand how they arrive at their decisions. In the context of cybersecurity, this lack of transparency can be problematic, especially when automated systems are making high-stakes decisions about data access or incident responses. Advances in explainable AI (XAI) will allow cybersecurity professionals to better understand and trust the decisions made by these systems, providing more confidence in automated decisions and the human oversight applied to them.

7.2 The Impact of Autonomous Systems on Human-in-the-Loop Security Models

As autonomous systems become more advanced, the role of human-in-the-loop will evolve. Today, HITL systems require active human intervention to make decisions based on machine-generated recommendations. However, in the future, as autonomous systems become increasingly capable, the role of human operators may shift from active decision-makers to supervisors or strategic overseers.

7.2.1 Fully Autonomous Cybersecurity Systems

The future of autonomous cybersecurity systems is likely to involve systems that can identify, mitigate, and respond to cyber threats without human intervention. These systems will use advanced machine learning to automatically update their threat detection models in real-time based on new data, and they will continuously evolve to combat emerging threats. Autonomous systems could take on routine security tasks such as incident triage, malware classification, and threat mitigation, freeing human experts to focus on high-level strategic issues or more complex security scenarios.

However, while autonomous systems will provide the speed and scalability needed to handle large volumes of cybersecurity incidents, human oversight will still be required in cases of complex, high-risk decisions where ethical considerations and organizational context are essential. In this scenario, humans will supervise and intervene when necessary, ensuring that the system operates within ethical boundaries and making judgment calls where full automation is impractical [22].

7.2.2 Autonomous Response and Escalation

Future HITL systems could enable autonomous systems to automatically escalate threats to human experts based on real-time analysis and threat severity. For example, in the case of a DDoS attack, an autonomous system might begin mitigating the attack using pre-configured rules and actions. However, if the attack exceeds certain thresholds or involves complex tactics, the system could escalate the situation to a human expert for further review. This type of dynamic escalation model ensures that systems remain agile and responsive while leveraging human judgment where it is most needed.

7.3 Recommendations for Future Research in Decision-Centric Cybersecurity

While significant progress has been made in integrating machine learning and HITL systems for cybersecurity, there are several key areas where further research is needed:

7.3.1 Improving AI Explainability and Transparency

Future research should focus on developing explainable AI (XAI) techniques to improve the transparency of machine learning models in cybersecurity. Understanding how a model arrives at its decision is critical for trust and reliability, particularly in high-stakes cybersecurity environments. Techniques like attention mechanisms and model interpretability frameworks should be explored further to improve the transparency of decisions made by complex machine learning models.

7.3.2 Enhancing the Scalability of HITL Systems

As cybersecurity systems become more complex and more data-intensive, the scalability of HITL systems will become increasingly important. Future research should focus on automating and optimizing the triage and escalation processes within HITL systems to ensure that human experts are not overwhelmed with a flood of alerts. Leveraging natural language processing (NLP) and automation to filter out low-priority alerts could help enhance scalability and reduce the workload on human operators.

7.3.3 Integrating Ethical Decision-Making Models

Ethical decision-making will play an even greater role in decision-centric cybersecurity. Researchers should focus on integrating ethical frameworks into machine learning models, allowing cybersecurity systems to make decisions that respect privacy and fairness. For instance, when analyzing personal data or conducting forensic investigations, systems should ensure that their actions align with legal standards and ethical principles, such as data minimization and transparency.

7.3.4 Continuous Learning and Model Update Strategies

Given the constantly evolving nature of cyber threats, continuous learning is essential for machine learning models in cybersecurity. Future research should explore methods for ensuring that models are continually updated without requiring extensive retraining. This includes exploring incremental learning techniques and developing online learning models that can update themselves in real-time without degrading performance.

7.3.5 Investigating Human-Machine Collaboration Models

Finally, further research is needed to explore the optimal collaboration between humans and machines in cybersecurity decision-making. Human operators bring domain knowledge, intuition, and ethical considerations that machines cannot replicate. Future models should be designed to combine these strengths, ensuring that the decision-making process is both efficient and ethical.

CONCLUSION

This review has explored the evolving role of Human-in-the-Loop (HITL) in decision-centric cybersecurity systems, with a focus on the integration of machine learning (ML) to enhance the detection, analysis, and mitigation of cyber threats. The integration of HITL with machine learning offers a hybrid approach that combines the speed and scalability of automated systems with the contextual understanding and judgment provided by human experts. This collaboration is essential for addressing the growing complexity of modern cyber threats, where automated systems alone may struggle to make the best decisions, especially in the face of novel attacks or ambiguous data. Through the review, we have identified the various machine learning techniques commonly used in cybersecurity, such as supervised learning, unsupervised learning, and reinforcement learning, and explored their advantages and limitations in addressing cybersecurity challenges. We also discussed the critical role of human oversight in preventing errors, mitigating biases, and ensuring that ethical considerations are considered in cybersecurity decision-making. We also examined the challenges and limitations of integrating HITL with machine learning, such as human error, scalability issues, and the ethical and privacy concerns that arise when human judgment is involved in sensitive decision-making processes. Finally, the paper discussed the future directions in decision-centric cybersecurity, highlighting the potential of autonomous systems, explainable AI, and continuous learning to enhance the effectiveness and scalability of HITL-enhanced cybersecurity systems.

REFERENCES

1. Anderson, R., & Moore, T. (2019). Security engineering: A guide to building dependable distributed systems (3rd ed.). Wiley.
2. Alazab, M., & Venkatraman, S. (2020). Machine learning for cybersecurity: A survey. *International Journal of Computer Science and Information Security, 18*(4), 42-56.
3. Zhang, X., & Liu, Y. (2021). Human-in-the-loop systems: Applications in cybersecurity. *Cybersecurity Journal, 12*(3), 98-112.
4. Davis, R., & Miller, T. (2022). The Future of Decision-Centric Cybersecurity. *Cyber Research Journal, 9*(1), 45-56.
5. Thompson, D. (2020). Limitations of Traditional Cybersecurity Frameworks. *Journal of Information Security, 18*(2), 34-47.
6. Clark, B., & Ford, L. (2021). Decision-Centric Cybersecurity: A New Approach. *International Security Journal, 7*(2), 102-115.
7. Johnson, R., & Patel, S. (2021). Supervised learning applications in cybersecurity. *Journal of Cyber Defense, 19*(4), 235-248.
8. Zhang, L., & Liu, H. (2020). Unsupervised machine learning for anomaly detection in cybersecurity. *Journal of Machine Learning and Security, 11*(2), 101-112.
9. Lee, J., & Cho, S. (2021). Reinforcement learning for cybersecurity defense. *AI in Cybersecurity Review, 12*(3), 78-89.

10. Smith, T., & Williams, K. (2020). Challenges in machine learning-based cybersecurity systems. *International Journal of AI and Cybersecurity, 8*(1), 56-67.
11. Zhang, M., & Yu, F. (2021). Human-in-the-Loop Models for Cybersecurity. *Cyber Defense Review, 8*(3), 98-112.
12. Kumar, P., & Sharma, A. (2020). The Role of HITL in Machine Learning for Cybersecurity. *Journal of Machine Learning and Security, 11*(4), 222-230.
13. Davis, R., & Miller, T. (2022). Challenges of Human-in-the-Loop Systems in Cybersecurity. *Cyber Research Journal, 9*(1), 45-56.
14. Williams, S., & Green, B. (2020). Case Studies of HITL in Decision-Making for Cybersecurity. *International Cybersecurity Review, 14*(4), 212-225.
15. Davis, R., & Miller, T. (2022). The integration of human-in-the-loop and machine learning for decision-making in cybersecurity. *Cybersecurity Review Journal, 13*(2), 98-112.
16. Green, J., & Park, L. (2021). Hybrid decision-making models in cybersecurity: Combining human oversight with automated analysis. *International Journal of Cyber Defense, 16*(1), 56-71.
17. Lee, J., & Kim, S. (2020). Enhancing response time and accuracy in cybersecurity with HITL and machine learning. *Journal of Machine Learning and Security, 18*(3), 234-248.
18. Gupta, R., & Sharma, N. (2021). Human error and cognitive biases in cybersecurity decision-making. *Cybersecurity Review Journal, 15*(3), 245-258.
19. Patel, R., & Singh, V. (2020). Scalability challenges in human-in-the-loop cybersecurity systems. *International Journal of Cybersecurity, 8*(2), 123-134.
20. Kim, S., & Lee, J. (2021). Ethical and privacy considerations in HITL cybersecurity models. *Journal of Ethics and AI in Cybersecurity, 5*(1), 67-79.
21. Ahmed, M., & Yu, Z. (2021). Advancements in machine learning algorithms for cybersecurity. *Journal of Cybersecurity Innovation, 14*(3), 234-245.
22. Carter, L., & Anderson, J. (2020). The impact of autonomous systems on cybersecurity: Challenges and future trends. *Cyber Defense Review, 19*(2), 112-126.