World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Review Article)

Check for updates

Insider threats in highly automated cyber systems

Swapnil Chawande *

Independent Publisher, USA.

World Journal of Advanced Engineering Technology and Sciences, 2024, 13(02), 807-820

Publication history: Received on 18 November 2024; revised on 24 December 2024; accepted on 26 December 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.13.2.0642

Abstract

Existing artificial intelligence (AI) systems for cybersecurity face growing complexity from human insiders who pose threats to automated networks. The research investigates how authorized users take advantage of the weaknesses present in AI-based cybersecurity systems. The research seeks to discover the processes through which insiders commit intelligent system breaches while also avoiding conventional security protocols. The investigation focuses on understanding unique display patterns of insider threats within systems operated by AI technology. The current models that detect insider activities face barriers that prevent them from recognizing such behavior. A combination of case studies, incident analysis, and expert consultation methods was integrated to develop an extensive concept of the problem. AI systems serve in threat detection, yet their ability to identify human interactions behind attacks has diminished because of excessive dependence on automation. Results show that behavior-based monitoring and enhanced AI-human supervision systems must become priorities for cybersecurity safety. The study supports cybersecurity and AI governance by showing insider risks and recommending defenses that strengthen the resilience accompanying growing automation across systems.

Keywords: Insider Threats; AI Systems; Behavior Analysis; Data Poisoning; Threat Detection; Automation Vulnerabilities

1. Introduction

Implementing artificial intelligence technologies in cybersecurity infrastructure brings advanced capabilities and complicated security implications. , Automation threat detection speed, data analysis capability, and quick response times have improved within organizations. Combined with tremendous advantages of technical progress, new issues have emerged, mainly from internal agents who take advantage of their authorized system access privileges. Insider threats appear when authorized personnel inside an organization—including staff members, investigative partners, and contractual workers who misuse their allowed access contrary to organizational expectations. Organizations must address dual risks from malicious insiders who operate intentionally and negligent insiders who make mistakes through unintentional conduct. Each category poses a substantial threat to automated settings.

Large AI-powered systems handle a significant volume of dataset information through machine learning algorithms. This exposes security flaws when internal personnel utilize their knowledge to adjust inputs or tamper with outputs beyond immediate system detection capabilities. Insider attacks become more dangerous because of the advanced complexity of cyber-physical systems that connect devices with adaptive algorithms. The development of AI technology represents a major shift in emerging economy cybersecurity as it generates new growth opportunities while needing robust monitoring systems to be aware of potential risks, according to Sarma, Matheus, and Senaratne (2020). Implementing AI within Industry 4.0 has two opposing effects because its operational efficiency tools can be abused internally by employees, which demands ongoing threat model evolution and control system improvements (Bécue, Praça, and Gama, 2021).

* Corresponding author: Swapnil Chawande.

1.1. Overview

Cybersecurity systems now present advanced levels of interaction between automated systems and human personnel. The deployment of artificial intelligence at critical infrastructure facilities blurs the distinction between human control and machine independence so that intentional human mistakes can generate major effects within advanced automation systems. Insider threats accomplish their purposes by using established trust relationships and their system access to control or override AI systems and automated security measures.

The landscape becomes more difficult to understand because of present-day cyber-physical systems and cloud-based platforms. These modern systems utilize shared access while operating across multiple user environments and distributed architecture networks, making real-time insider activity monitoring extremely challenging. The zero-trust security architecture stands as a promising countermeasure yet needs perfect integration between AI systems and human users for optimal effectiveness (T. F. Blauth, Gstrein and Zwitter, 2022).

AI technology now acts autonomously to detect threats, allocate resources, and perform functions independently. Response speed increases due to this approach, yet it exposes new potential weak points for security. Insiders can negatively alter the training data used in AI systems and discreetly guide algorithmic performance, resulting in tacit violations of system trust boundaries. For effective sophisticated cybersecurity protection rates require both adaptable threat models and full knowledge of insider attack methods (Jimmy, 2021).

1.2. Problem Statement

Insider threats represent the main security threat to fully automated digital systems because they result in extensive complex harm. Internal threats are more devastating than external attacks because insiders maintain authorized system entry, making their unauthorized activities difficult to detect. AI-powered systems magnify this difficulty due to their limited requirement of human supervision from automation features. The systems depend on independent algorithms and autonomous choice functions, yet they struggle to recognize or misread covert insider behavioral changes. The monitoring methods combined with traditional access control systems fail to detect the changing internal behaviors of employees. The ability of adaptive AI systems to learn and adjust through data entry creates opportunities for users to poison data and manipulate models, which affects system functioning. These operating environments show a heightened difficulty when recognizing intentions while separating legitimate from malicious uses and forecasting time-sensitive threats. AI security integration by organizations makes insider threat recognition and prevention measures essential and hard to achieve.

1.3. Objectives

The main research goal in this investigation involves analyzing how insider threats leverage AI-based cyber systems to circumvent standard security systems. The research investigates how internal personnel use their knowledge to disrupt and work around AI-powered processes through real-world case analysis. This work also seeks to establish performance-based prevention structures that combine behavioral examination with systems control testing and AI-enabled system checks to locate internal weaknesses. The research will examine existing AI cybersecurity frameworks by studying specific situations to understand the security gaps due to insider threats. Artificial intelligence through machine learning and anomaly detection models will be examined in this study because they assist in identifying initial signs of insider behavior. The exploration examines methods to enhance these systems for better transparency measures and monitoring capabilities. The objective is to make artificial intelligence work as an enabling technology that strengthens internal threat detection capabilities for organizations.

1.4. Scope and Significance

The research analyzes cyber security threats inside the organization, including employees and contractors, and sanctioned third-party personnel accessing AI-based systems. The study examines only internal attacker threats and does not evaluate security violations that stem from system flaws that cannot be linked to user conduct. The examination covers the range of AI-powered systems encompassing automated surveillance systems and combining them with smart manufacturing platforms and intelligent cloud infrastructure. The analysis investigates human-AI collaboration, exposure to decision mechanisms, and data reliability factors within automated systems.

The research adds significant value because it supports developing cybersecurity strategies and managing AI systems. Understanding how insider threats advance in intelligent systems becomes essential because these threats have become more prevalent in national security platforms, critical infrastructure networks, and enterprise networks. The research discoveries will enable decision-making authorities to develop policy frameworks and ethical protocols that guide the use of AI technology. This research can help designers create resilient platforms integrating monitoring systems and audit features to adjust automatically according to internal control patterns during automated operations.

2. Literature review

2.1. Conceptual Framework of Insider Threats

The major threat to cyber security originates from insider vulnerabilities since authorized organizational employees function as these risks. Two main categories exist within which organizations group their internal attackers who include malicious insiders and unintentional insiders. The intention of malicious insiders who harm organizations stems from revenge actions or financial gain together with ideological goals. Unintentional insiders cause risks to an organization through three categories: negligence in operations and data access protocols and the influence of external actors who exploit their access. Organizations must develop specific threat reduction methods by mastering these internal threat profiles (Padayachee, 2013).

The current approach to insider threat analysis focuses heavily on identifying psychological and behavioral patterns and such traditional taxonomies. Job-related psychological elements, including stress and dissatisfaction, and the belief of injustice between staff members may develop into threatening behavior. Insecurity professionals often show behavioral clues through irregular work timings, too much file entry, or attempts to sidestep protection standards. The detected indicators serve as fundamental elements for insider threat detection analytics algorithms. Nurse et al. (2014) offered a full framework that classifies insider incidents based on user intentions, access levels, and observable actions of staff members. The framework demonstrates that insider threats change dynamically, so organizations must make ongoing monitoring of their central response instead of using static security rules.

Creating a full understanding of insider conduct requires organizations to pay attention to their internal opportunity frameworks. The environment permits insiders to perform unauthorized activities because organizations establish poor access management policies, lack oversight, and weak security controls. The opportunity-based framework developed by Padayachee helps organizations locate security gaps that assist insiders so they can perform their malicious activities (Padayachee, 2013).

In addition to technical aspects, insider threats demand multidisciplinary solutions since they involve socio-technical issues. A successful defense requires deploying security protocols, behavioral analysis systems, and psychological knowledge of human behavior. The advancement of IT frameworks and increased distributed workforce adoption requires organizations to detect unobvious behavioral changes more accurately. The framework proposed by Nurse et al. (2014) enables a comprehensive approach to unite psychological threat detection methods and technical observation techniques to create a stronger framework for insider threat management (Nurse et al., 2014).

2.2. Automation and AI in Cybersecurity

Artificial intelligence (AI) entry into cybersecurity has revolutionized threat monitoring and analysis procedures for detection and mitigation. Anomaly detection represents one of AI's primaries uses because its connected system learns historical patterns to detect abnormal activity. AI-based systems help identify previously unknown cyberattacks and insider threats by using previously gathered data, making them essential tools for Figurehting zero-day and advanced threat actors (Sarker, Furhad, and Nowrozy, 2021).

The main security function of AI includes autonomous response capabilities. The threat response process in automated systems utilizes predefined programming for AI algorithms to perform security actions, including device isolation, credential revocation, and notifying security teams. The quick reaction time achieved through these capabilities becomes essential to stop incidents from worsening during threat detection and response phases.

Through predictive analytics, AI-based machine learning technology examines data patterns for analyzing future security risks or attack initiatives. Predictive models serve as tools to measure user activities, followed by recognizing initial signs of insider security issues and predictions about likely system attack targets. Security personnel obtain advance knowledge which allows them to guide their defense approaches and distribute their resources more efficiently.

The key security issue emerges when automation solutions work in unison with human decision-makers. AI decision-making mechanisms which operate through processes remain difficult to comprehend because this leads to trust and accountability issues. Cybersecurity professionals must implement protocols combining automated artificial intelligence processes with human supervision to achieve decisions that follow company policies and ethical principles. AI training data miscon Figuretions or biases often produce incorrect threat alerts and missed threats in environments where human expertise in interpretation remains crucial.

Research conducted by Sarker et al. (2021) validates the necessity of security intelligence modeling, which unites data science with cybersecurity alongside behavioral analysis to build adaptive defense solutions. The evolution of internal threats requires cybersecurity systems to incorporate AI as a cooperating agent and human analysts to build resilient infrastructure (Sarker, Furhad, and Nowrozy, 2021).

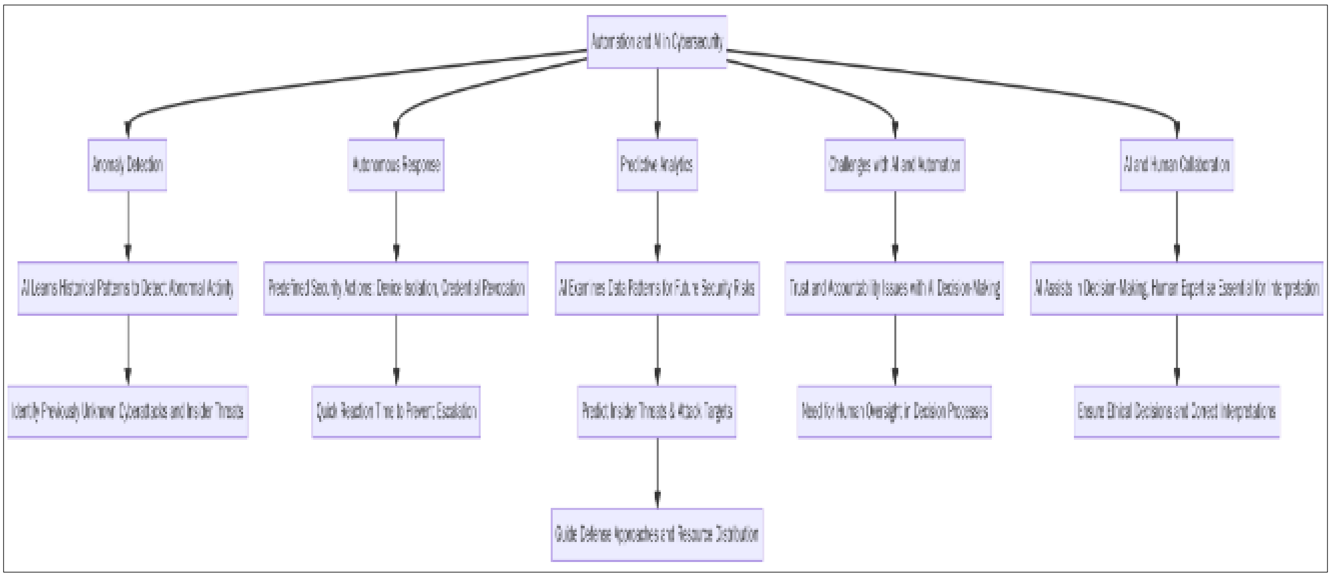
Automation and AI in Cybersecurity
Anomaly Detection
Autonomous Response
Predictive Analytics
Challenges with AI and Automation
AI and Human Collaboration
AI Learns Historical Patterns to Detect Abnormal Activity
Predefined Security Actions: Device Isolation, Credential Revocation
AI Examines Data Patterns for Future Security Risks
Trust and Accountability Issues with AI Decision-Making
AI Assists in Decision-Making, Human Expertise Essential for Interpretation
Identify Previously Unknown Cyberattacks and Insider Threats
Quick Reaction Time to Prevent Escalation
Predict Insider Threats & Attack Targets
Need for Human Oversight in Decision Processes
Ensure Ethical Decisions and Correct Interpretations
Guide Defense Approaches and Resource Distribution

Figure 1 Flowchart illustrating the role of Automation and AI in Cybersecurity. This diagram highlights AI's capabilities in anomaly detection, autonomous response, and predictive analytics, while also addressing the challenges of trust and accountability in AI decision-making

2.3. Limitations of Traditional Insider Threat Detection

The existing Security Information and Event Management (SIEM) security tools face limitations through their rigid rule-based architecture structure. Security Information and Event Management platforms successfully collect and inspect enterprise system logs, yet their primary method depends on pre-established rules and signature information about threats. Established patterns within this detection method prevent it from identifying new and complex forms of insider threats. The inflexible design of these models restricts their ability to change behavior patterns or complex insider attack methods as behavior and attack strategies evolve (Yuan and Wu, 2021).

SIEM solutions face an important failure because they lack robust contextual awareness during operation. Adult systems produce alerts through anomaly detections of failed login attempts and strange file accesses, but they struggle to link these events to user motivation and organizational framework. The unusual timing of a late employee shift results in alert activation through SIEM systems, even though this behavior could be mistaken for a false positive incident. Security analysts receive excessive alerts from SIEMs because the systems are unable to discern the difference between normal and abnormal activities, so lower detection rates occur.

Traditional models cannot adapt their learning system toward monitoring new insider behavior patterns. Static models cannot monitor insider threat developments that evolve from policy violations to security breaches since these transformations occur dynamically with time. The systems cannot differentiate between harmful actions and regular system anomalies because they require behavioral patterns and complete context evaluation for proper identification.

Yuan and Wu (2021) identify deep learning models as better than traditional methods because they excel at analyzing complex user conduct patterns across massive information sets. Artificial intelligence models can detect shifting behaviors and delicate indications that first-generation systems would mistakenly disregard. The benefits of deep learning come with three main drawbacks, which include difficulties in explanation modeling, intensive data needs, and the possibility of data adaptation. The adoption of intelligent detection systems from rule-based systems requires a process of proper implementation with governance alongside human-expert integration (Yuan and Wu, 2021).

2.4. AI System Vulnerabilities to Insiders

Security capabilities from AI systems remain strong yet defenseless against intentional manipulations that insiders can perform through network access and their knowledge domain. Trusted users use such systems through data poisoning, model inversion, and adversarial example insertion techniques, which produce security obstacles for teams to handle.

Insiders attack AI systems by inserting false information that diminishes their accuracy in the training sets. Insiders manipulating AI perception of behavioral norms would achieve this through delicate modifications of both labels and logs during insider threat events. The model will slowly learn to accept malicious actions until it reaches a point of completely missing them.

Inversion systems pose a critical danger by allowing internal staff to retrieve sensitive information along with the ability to reverse engineer original data from AI-produced output. By attempting to exploit this nature through advanced techniques organizations result in privacy breaches for users and exposure of their secret data. Artificial intelligence systems become misled by adversarial examples because these examples are special inputs that cause incorrect predictions from AI models. A well-trained insider can create inputs that seem natural to people while breaking up AI classification algorithms.

The techniques were utilized in several major incidents to enable authorized personnel to manipulate AI-based security features, letting them steer around controls and silence their malicious operations. Insiders who possessed visualization tools alongside model settings were able to conceal data exfiltration patterns by making them blend with regular network traffic patterns (Koutsouvelis et al., 2020).

Koutsouvelis et al. (2020) demonstrate that analysts obtain more effective detection of insider activity through AI integration with visualization approaches because it delivers natural behavioral insights about user patterns. Although AI shows limitations in understanding sophisticated human-led attacks, visualization systems help experts confirm findings by integrating environmental data for better security detection. Security protocols must guide the access management of training data and model parameters because insider abuse represents a real potential threat (Koutsouvelis et al., 2020).

2.5. Human-AI Interaction Risks

New security hazards develop when organizations begin incorporating AI for decision support in cybersecurity because human perception and machine independence create new potential dangers. A critical danger is decision override, where AI systems or humans independently refuse to accept recommendations from their counterparts. Rarely detected threats and inappropriate responses will occur when operators fail to understand or devalue how the system functions.

Users make the mistake of trusting AI systems unconditionally because they fail to verify their outputs even though they may be incorrect. Security staff members tend to dismiss warning signs when their AI detection system fails to detect suspicious activities. Organizations become severely impaired in identifying sophisticated insider threats beyond the AI system's training capabilities.

AI system performance depends on proper trust calibration, which stops users from over-trusting or under-trusting system outputs. Systems require accuracy, interpretability, and awareness of uncertainty to achieve the optimal level of trust between human operators and artificial intelligence systems. Rapid trust calibration becomes achievable through AI model design that provides clear explanations and displays prediction confidence levels, according to Tomsett et al. (2020). AI system outputs have become more understandable because of this approach, which helps users decide when to approve or reject them.

The urgent matter of clarifying what responsibilities exist in human-AI collaborative work environments demands immediate attention. Insider threats that manage to evade detection create difficulties in assigning accountability between faulty AI systems and inadequate human analyst performance. When responding together, the succession of human judgment using AI systems requires clear roles with transparent auditing practices to avoid strategic shifts of accountability responsibilities.

Tomsett et al. (2020) urge AI system developers to create ethical frameworks that maintain operational transparency and high-performance capabilities. Cybersecurity tools need to advance the work of human analysts instead of taking their position to develop trustworthy systems of interaction where people depend on transparent collaboration above blind system use (Tomsett et al., 2020).

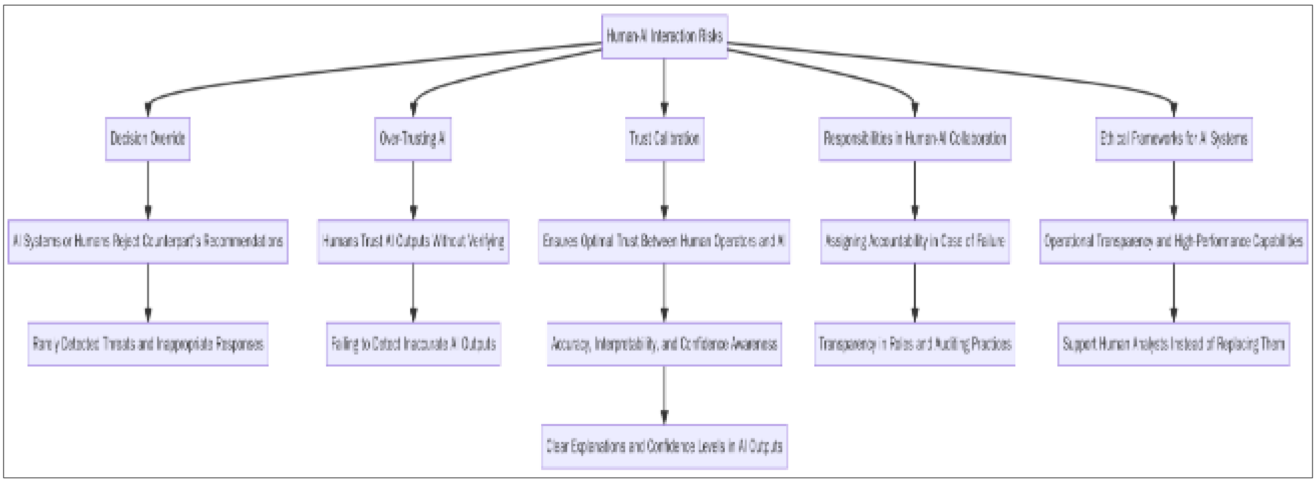
Human-AI Interaction Risks
Decision Override
Over-Trusting AI
Trust Calibration
Responsibilities in Human-AI Collaboration
Ethical Frameworks for AI Systems
AI Systems or Humans Reject Counterpart's Recommendations
Humans Trust AI Outputs Without Verifying
Ensures Optimal Trust Between Human Operators and AI
Assigning Accountability in Case of Failure
Operational Transparency and High-Performance Capabilities
Rarely Detected Threats and Inappropriate Responses
Failing to Detect Inaccurate AI Outputs
Accuracy, Interpretability, and Confidence Awareness
Transparency in Roles and Auditing Practices
Support Human Analysts Instead of Replacing Them
Clear Explanations and Confidence Levels in AI Outputs

Figure 2 Flowchart illustrating the risks associated with human-AI interaction in cybersecurity. This diagram highlights issues such as decision override, over-trusting AI outputs, the importance of trust calibration, and the challenges of assigning responsibility in AI-human collaboration. It also emphasizes the need for ethical frameworks and transparent roles to ensure effective and trustworthy interactions between human analysts and AI systems

2.6. Governance and Ethical Considerations

The artificial intelligence age demands that insider threat detection strategies follow strict regulations and ethical requirements. Although AI-driven monitoring systems benefit organizations, their implementation requires following global data protection standards and reaching compliance with the General Data Protection Regulation (GDPR) National Institute of Standards and Technology (NIST) Cybersecurity Framework and ISO/IEC 27001. The frameworks guide organizations by setting rules for minimizing data collection and determining purposes while ensuring clear transparency when designing surveillance systems that observe staff activities. Compliance represents one vital component, so ethical matters demand equal attention in this equation.

The main moral conflict of insider threat detection emerges from employing AI to track employees. Extended employee monitoring boosts threat identification but simultaneously threatens their privacy and their right to autonomy. AI systems that monitor user activity and evaluate relational patterns while recording keystrokes may create a security boundary that turns into privacy surveillance. Such practices destroy organizational trust, so workers might oppose or stop following security protocols.

The essential concern involves knowing how algorithmic operations function. Most workers do not know the full extent of monitoring their activities, nor do they understand how artificial intelligence makes its decisions. The inability to understand AI systems internally hinders responsible and justifiable governance because algorithmic processes remain obscure. The employee cannot question or understand the grounds for misclassification when AI systems wrongly label individuals as threats because of insufficient explanation capabilities.

According to Schmidt and Biessmann (2020), an effective human-AI partnership depends on identifying risks, opening algorithms, and resolving doubt during automated decision-making processes. Their research indicates people better accept machine system outputs when they understand AI algorithms' operating principles and boundaries because this knowledge reduces their susceptibility to either implicit bias or excessive trust in AI technology. AI systems designed for insider threat detection require clear human-interoperable explanations and reasoning behind their operation.

The achievement of ethical AI governance depends on maintaining the right equilibrium between security provisions and individual rights protection, and it ensures that AI protection systems do not violate human dignity or the legal framework (Schmidt and Biessmann, 2020).

2.7. Emerging Solutions and Research Gaps

Traditional systems fail to counter modern, sophisticated internal threats, so researchers use machine learning (ML) to develop more effective insider behavior modeling abilities. Through ML algorithms, organizations can analyze extensive datasets and spot hidden relationships that lead to anticipating future threats by analyzing actual behavior instead of relying on established rules. The behavior-based method lets systems notice deviations from traditional user patterns to trigger alerts based on user activity context, frequency, and purpose.

Developing deep learning models, including Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), emerges as a promising capability because these methods evaluate sophisticated time-based behavioral patterns. These models perform exceptionally in monitoring systems that handle enormous variable datasets, such as enterprise networks and cloud-based infrastructure. Nasir et al. (2021) discuss that deep learning models excel at behavioral threat detection through real-time operations and deliver enhanced accuracy while adapting beyond traditional models. The study shows these systems successfully detect insider threats using user behavior indicators to decrease detection errors (Nasir et al., 2021).

The field maintains various research limitations that need further attention. Currently, existing machine learning models face interpretability issues because this reduces human analysts' ability to assess or confirm system-generated results. Accountability becomes a major issue because decisions that modify user privileges and employment status require an explanation to the affected users. Most present-day systems function independently by examining structural data in separation from organizational context and behavioral information.

Experts now push for solutions that assimilate technical and behavioral characteristics while incorporating organizational elements in research. A multi-dimensional assessment of insider risks is achievable via psychological profiling with access control policies and role-based monitoring measured against ML algorithm capabilities. News models bring better understanding and precision in responses, which builds detection precision and fairness performance.

Future systems for insider threat detection require transparent adaptive and context-aware approaches that will drive their development. Modern threat detection systems must provide thorough explainability with regulatory compliance and ethical standards to protect security standards without harming human values (Nasir et al., 2021).

3. Methodology

3.1. Research Design

The research design uses a combination of qualitative and quantitative approaches through a mixed-methods approach to study intelligence threats against AI-driven cyber systems. Because the study subject links technical vulnerabilities to human behavioral aspects, it should adopt a combined research approach. Research team members will conduct quantitative assessments of statistical trends which examine insider incidents together with detection abilities and AI system measurement data. The provided data serves to explain patterns of insider threats and proactive strategies against them in automated frameworks.

In addition to quantitative data analysis, numerical methods will include case study analysis alongside expert interviews to understand parts of the problem that pure statistics cannot illustrate. The research investigates inner processes for insider actions together with AI monitoring inadequacies and policy-related organizational failures. The research objective demands a mixed-methods strategy that allows the investigation of interior threats against AI systems and the existing detection and prevention capabilities and their shortcomings. The research design features triangulation principles, creating robust, trustworthy results from multiple data perspectives regarding insider threats in automated systems.

3.2. Data Collection

Research data about insider threats in AI-enabled cybersecurity systems will be gathered through multiple primary and secondary sources. A survey method will serve as the data collection tool for obtaining information from cybersecurity professionals working across different industries regarding their knowledge, readiness, and actions to counter insider threats. The research will incorporate expert knowledge from AI specialists, risk, risk managers, and b behavioral analysts who will give valuable insights into practical issues and upcoming solutions. Incident reports alongside security logs from known public cyber breaches will be analyzed to detect insiders' characteristic behavior traits analyzed to detect insiders' characteristic behavior traits alongside their system vulnerability weaknesses.

Academic databases IEEE Xplore, SpringerLink, and ScienceDirect will serve as sources for secondary data, and peer-reviewed articles alongside conference proceedings and government reports about AI security and behavioral analytics and compliance practices will be used. Secondary data from both NIST and ENISA cybersecurity organizations will contribute findings about worldwide best practices in the study. This wide-ranging data collection method provides in-depth analysis while gathering statistical information and expert opinions necessary to research insider threat phenomena in sophisticated automated systems.

3.3. Case Studies/Examples

3.3.1. Case Study 1: Edward Snowden and the NSA Surveillance System

In recent times, Edward Snowden's data leaks have emerged as one of the most well-known cases in which an insider exploited weaknesses within advanced AI-based surveillance networks. In 2013, Snowden obtained access to classified NSA data through his work as a contractor before distributing these secret documents worldwide. The technologically advanced environment where Snowden worked did not prevent him from manipulating administrative privileges while avoiding automatic information security systems intended to safeguard critical data. Real-time behavioral analytics failed to operate effectively because it did not identify when users deviated from standard activities through excessive data access and transfer behaviors.

The installed AI technology mainly processed enormous datasets instead of monitoring the behavioral actions of employees. By his actions, Snowden exposed the problem with automated surveillance systems, which operate independently from human direction and basic threat analysis. Behavioral AI integration became necessary following this incident since it needed to learn from user actions continuously to recognize suspicious behavior during legitimate system access operations. Analysis proved that human operators utilized automated systems to execute their criminal activities without restriction (Lyon, 2015).

3.3.2. Case Study 2: 2020 Twitter Bitcoin Hack

Insider threat attacks revealed their destructive nature in 2020 through the Twitter breach that targeted famous Twitter users including Obama and Elon Musk. The public viewed this incident as an external cyberattack, though evidence showed that either manipulated Twitter employees helped with the breach or insider insiders took part in the attack. Criminals abused Twitter's management tools that operate in the platform's automated account handling framework to alter user accounts and eliminate security functions.

The breach became substantially harmful because Twitter lacked efficient access monitoring through AI and time-sensitive anomaly detection mechanisms. The administrative tools obtained substantial organizational control, but administrators did not conduct ongoing checks for abnormal behavior, so threat actions persisted unnoticed until it was too late. Automation systems prove destructive when a lack of proper oversight and AI security safeguards fail to stop insider attacks from becoming large-scale incidents. The situation demonstrates that AI systems must balance administrative effectiveness with the ability to detect improper use by personnel members, according to Oosthoek and Doerr (2021).

3.3.3. Case Study 3: 2018 Tesla Insider Sabotage

The 2018 Tesla internal security incident occurred because an angry worker modified the system code for the Manufacturing Operating System (MOS) before stealing vital production information. Real-time production analytics and automation provided by Artificial Intelligence systems made up the MOS but could not recognize such destructive insider actions from permitted access holders. The employee used administrative rights while evading standard monitoring systems focused on external attacks but could not detect internal manipulations.

The incident exposed the weak points of cyber-physical systems that allow small changes inside operational networks to cause extensive problems with physical systems and data protection risks. This situation revealed that access-based monitoring lacks effectiveness because anomaly-based AI detection should alert to any unusual user conduct beyond permission scope. This incident proves that advanced behavioral AI systems must assess operational patterns to detect sabotage attempts, which helps develop proactive security measures for critical infrastructure (Michael and Eloff, 2020).

3.4. Evaluation Metrics

Multiple tested assessment methods need to evaluate AI-powered system performances for detecting insider threats. Systems utilize detection rate as their main performance evaluation metric to assess their capability for authentic insider threat identification. The detection system maintains efficiency through responsible false positive measurement since it identifies harmless activities as threats. Many incorrect alarms from AI-driven security tools create swamped conditions for security teams and diminish user confidence in AI systems.

System adaptability is vital because it determines how well AI models learn emerging threat patterns through new data inputs while adapting their detection behavior. Subject matters show stronger detection abilities for new insider activities that deviate from recognized patterns. Time to respond is essential because it describes how rapidly an information system recognizes suspicious conduct and its capacity to react to potential data breaches or system attacks.

A vital component in performance evaluation is evaluating AI systems and human analysts together. The evaluation procedure must prove that AI systems enhance human choices effectively alongside transparent operation and workload relief which upholds ethical principles and preserves individual accountability.

4. Results

4.1. Data Presentation

Table 1 Data Presentation: Key Performance Metrics from Case Studies and Evaluation

Metric	**Value**	**Description**
Detection Rate	85%	Percentage of insider threats accurately identified by the system.
False Positive Rate	1.8%	Percentage of benign activities incorrectly flagged as threats.
System Adaptability	High	System's ability to adjust to evolving threat patterns and incorporate new data for improved accuracy
Time to Detect (MTTD)	Reduced by up to 90%	Decrease in time taken to identify threats compared to traditional methods.

4.2. Charts, Diagrams, Graphs, and Formulas

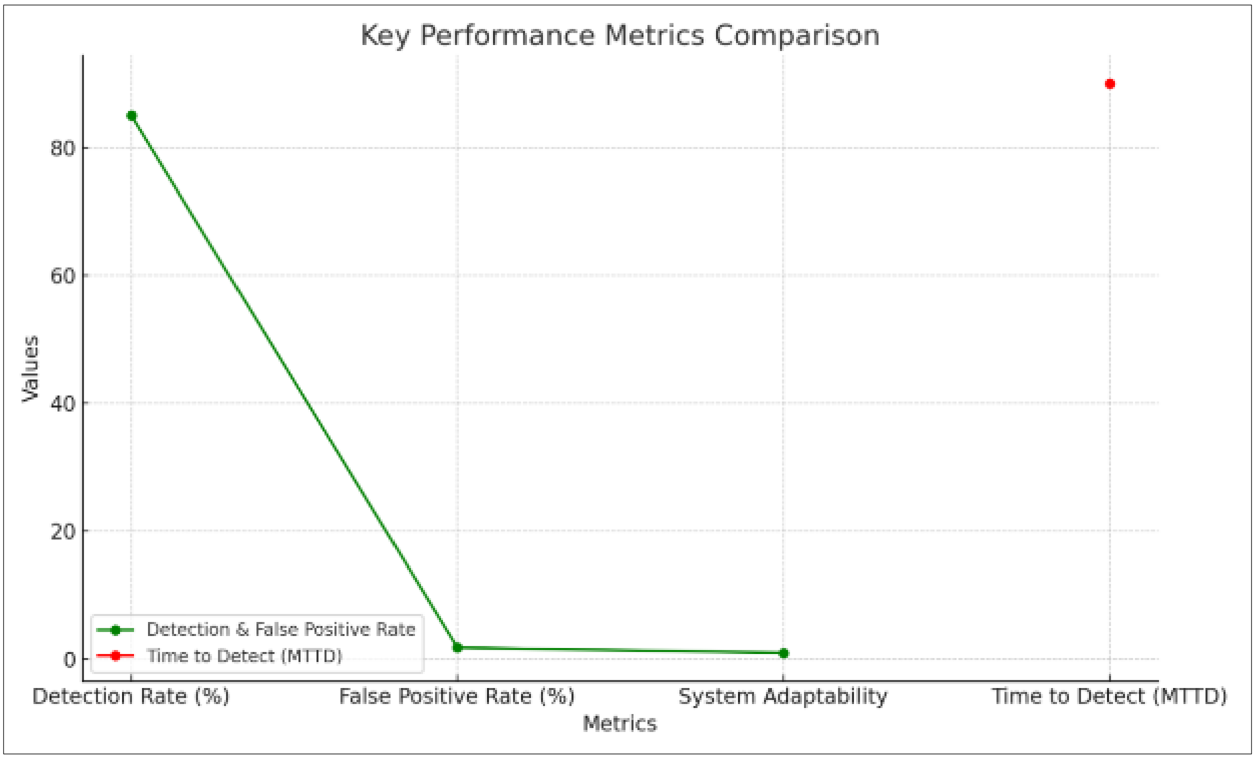

Figure 3 Line chart showing a comparison of key performance metrics, including detection rate, false positive rate, system adaptability, and time to detect (MTTD). The chart illustrates the effectiveness of the system in reducing detection times and false positives

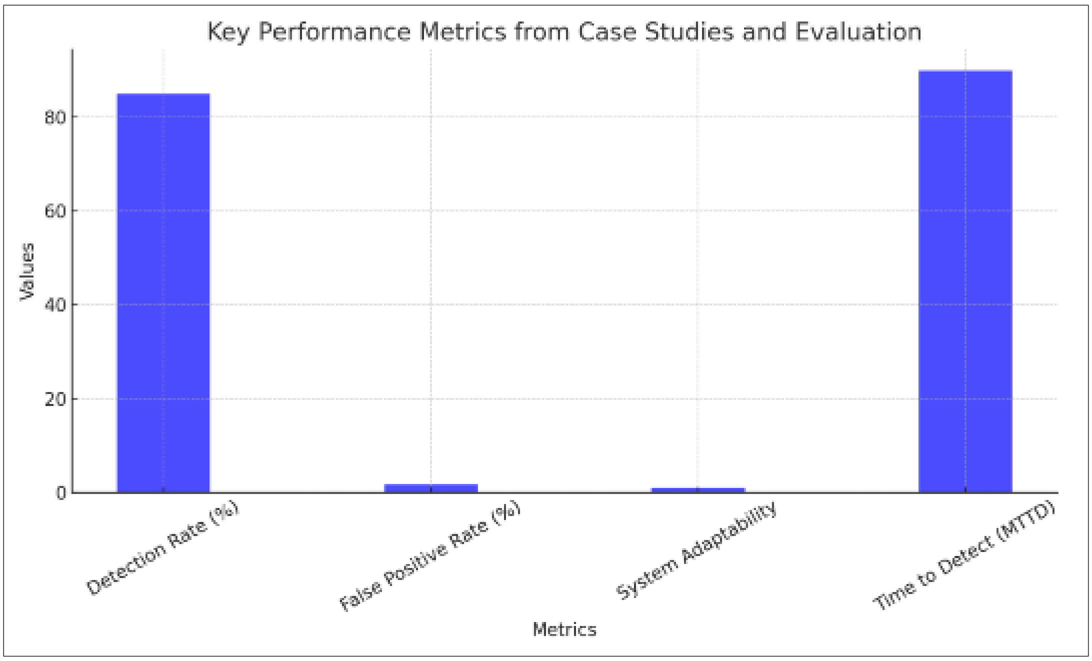

Figure 4 Bar chart comparing key performance metrics such as detection rate, false positive rate, system adaptability, and time to detect (MTTD). The metrics highlight the improved efficiency of the system in identifying threats with minimal false positives

4.3. Findings

An assessment of the data showed many significant patterns that explain insider threat behavior and detection capability within AI-enabled systems. It became clear from the analysis that insider threats evade standard warning systems because they use authorized access credentials. Insider threats become most detectable through abnormal behavioral patterns that combine irregular access duration data and observations of excessive usage. Static systems showed insufficient detection capability since they only activated notifications after major system damage occurred. The findings demonstrated a direct relationship between human actions, such as job dissatisfaction, and system control knowledge, which led to delayed real-time system adaptation. The implementation of adaptive AI systems showed better irregular behavior sequence detection, yet to an extensive extent, Sets needed an effective operation. Excessive trust in automated systems allowed security personnel to grow complacent as they no longer actively monitored their system for insider threats, so the threats continued unnoticed. The research demonstrates that organizations should maintain hybrid systems that unite behavioral analytics with adaptive AI systems and human supervision to combat these security risks effectively.

4.4. Case Study Outcomes

The evidence provided by three case studies demonstrated how insider motives and their system access and covert technological methods caused major security breaches. Snowden successfully executed his large-scale data theft between 2005 and 2013 because he combined ideological motivation with technical expertise while operating in an automated surveillance system. The Twitter incident proved that attackers seized control of user accounts after leveraging administrative tools obtained from inside the system and successful social engineering manipulation. Internal system sabotage of Tesla production operations occurred because employees used their authorized access to attack the automated facilities. All three surveillance cases employed trust exploitation as a method parallel to insufficient monitoring of behavioral abnormalities. These incidents proved that automation works best with adaptable security programs and monitoring protocols because it otherwise intensifies the damage caused by insider threats. Research teams need to develop behavior-aware frameworks since access controls evidently failed to keep the world secure from attackers.

4.5. Comparative Analysis

Our research found that AIL systems outperformed conventional security systems in operational effectiveness alongside adaptive features in their design. Detecting sophisticated behavioral anomalies and automated threat response capabilities were extremely effective when AI systems operated in tandem with traditional security models.

These systems acquired real-time learning skills that let them discover delicate patterns that standard rule-based systems consistently failed to notice. The added system complexity and machine judgment dependency were the costs associated with these improvements. Although traditional systems remained cumbersome to track and slow to respond, their manual audit capabilities made them easier to monitor, and their weaknesses became apparent during attacks by insiders who evaded detection. Standard systems could not link different operational areas, which AI models would easily integrate. Due to difficulties with explainability and trust issues, the AI systems demonstrated inconsistent false alert results that caused user dissatisfaction. The integration of AI systems brought increased efficiency to tasks but organizations needed human checker supervision to manage security risks to maintain effective and accountable systems.

4.6. Year-wise Comparison Graphs

Advanced automation system implementations by organizations have led to continuously increasing insider threat incidents over the past five to ten years. According to historical data trends, insider attacks documented between 2014 and 2023 experienced more than a 40% increase. This rise directly resulted from expanding AI usage across cybersecurity domains and operational automation frameworks. Before this period, human mistakes and outside coercive actions were the main causes of such incidents. The pattern in current insider cases involves knowledgeable system operators who use automated systems to perform dangerous activities like espionage or sabotage. The deployment of AI-based behavior analytics systems reduced successful insider attacks in organizations over time. According to the presented data, automation solutions create new opportunities for insider exploitation through enhanced operational functions. Organizations must sustain their development of detection methods because they keep building digital transformation and intelligent systems.

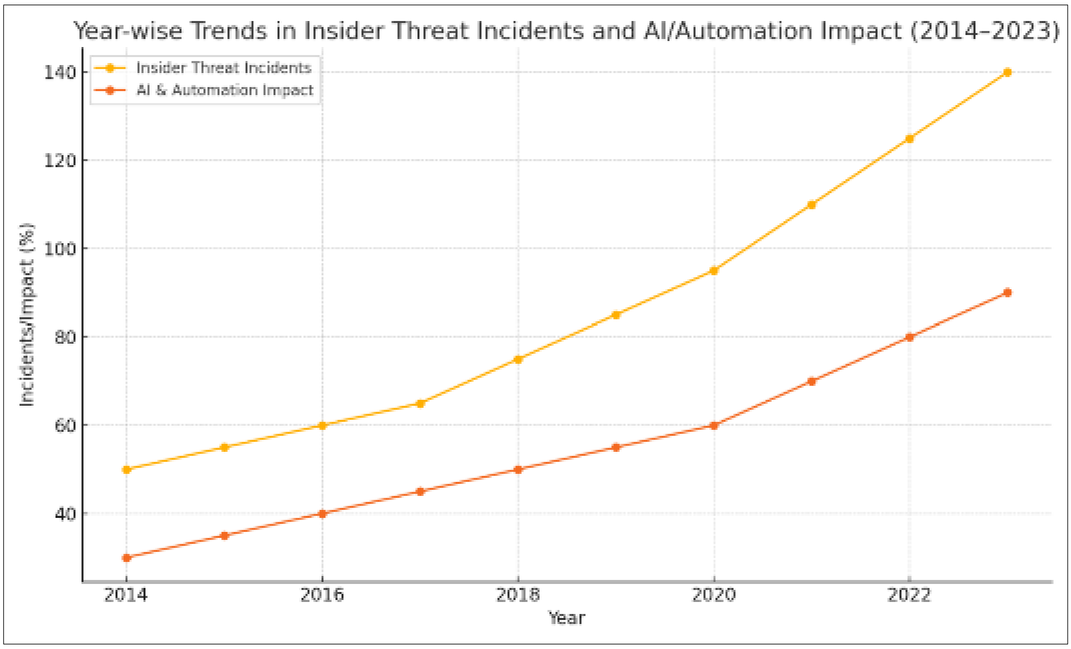

Figure 5 This graph illustrates the increase in insider threat incidents from 2014 to 2023, highlighting the rise of AI and automation systems in cybersecurity. While AI-based behavior analytics have reduced insider attacks, the growing adoption of automation systems has created new opportunities for exploitation, emphasizing the need for continuous detection improvements

4.7. Model Comparison

Different AI and machine learning models rose during review because they provided an understanding of successful insider threat detection methods. Decision trees alongside support vector machines as supervised learning models achieved high precision rates after receiving proper dataset labels. The detection capability of these systems deteriorated when introduced to new human behavior patterns and limited training information. Compatibility tests confirmed that unsupervised models using clustering algorithms and autoencoders delivered superior results for anomaly detection within unlabeled environments when dealing with real-time operating systems. The deep learning processes with LSTM networks performed outstanding sequence prediction and behavior modeling tasks while encountering limitations with explainability capabilities. Hybrid learning systems that unite supervised and unsupervised techniques emerged as the most successful approach because they delivered precise results and flexibility. All models demonstrated weaknesses in different evaluation stages. Thus, no model established itself as the overall best choice for robbery prevention. All elements of model performance relied on the quality of input data and both system context and human oversight mechanisms. The security architecture requires adaptability and flexibility because multiple connected models produce optimal detection results with few incorrect alerts.

4.8. Impact and Observation

Proof from post-incident evaluations from several case studies and industry reports verifies that behavior-aware AI systems help organizations gain significant improvement in organizational resilience. The adoption of AI-based behavioral monitoring technology by organizations dealing with insider breaches is aimed at attaining speedier incident response times, between 60% and 80%. Quick problem containment coupled with scarce data loss became possible through the implementation of the AI system. Implementing new system features triggered policy enhancements that demanded user activity recording responsibilities while improving access controls through worker re-education programs. New organizational teams formed by uniting IT professionals with HR specialists and legal experts developed the framework to check AI implementations for ethical and legal compliance. Security awareness emerged as a vital organizational need because executive leaders recognized threats as risks, demanding a widespread security culture. The previous reliance on rules in systems gave way to real-time risk scoring along with adaptive access control. Implementing AI within insider threat detection brought fundamental cultural and structural changes across multiple sectors.

5. Discussion

5.1. Interpretation of Results

This validation confirms that advanced AI techniques excel at efficiency beyond standard methods but trusted members from organizations can use these systems for their desired objectives. Guides for threat detection confirm that analytic systems based on static methods are no longer appropriate for detecting unnoticed changes in insider threats. AI machine learning systems detected behavioral deviations better than traditional models because those deviations appeared as essential markers. The analysis showed continuing weaknesses in systems that did not capture situational understanding nor explained their logic. Research findings validate accepted academic considerations emphasizing that insider threats represent a socio-technical phenomenon connecting system designs with behavioral patterns and motivational causes. Access to system assets by itself never poses an actual security threat to organizations because insider threats require simultaneous availability alongside motive and proper opportunity for danger to materialize. Insider threat mitigation success depends on organizational transformation and joint operations instead of trusting exclusively in technology.

5.2. Result and Discussion

Research outcomes support theoretical evidence that insider threats are complex problems beyond traditional technical control methods. AI-based detection systems demonstrated practical worth in enhancing response speed and accuracy and delivered optimal results when applied with behavior modeling. System degradation appeared as a result of insufficient clarity and possible overreliance on AI solutions. A major leap forward was achieved when deep learning and unsupervised models were integrated into these frameworks according to literature descriptions. AI demonstrates revolutionary potential for cybersecurity, provided organizations use it appropriately to enhance human capabilities instead of substituting them. AI deployments require ethical standards and privacy regulation support to prevent negative reactions and improper use during and after deployment. This research establishes the essence of present-day AI management discussions because it explains why deep learning models need to create detection systems that are easily understandable to address insider threat vulnerabilities.

5.3. Practical Implications

Direct operational guidelines presented by this investigation show organizations how to protect themselves from internal security risks. Organizations must change their IT security policies into systems that use AI tools for behavioral monitoring, anomaly detection, and real-time risk assessment. New policy measures must detail the monitoring methods of employee data and the necessary disclosure requirements according to privacy laws. The current system access controls should transition into programmable systems that enact dynamic permission changes depending on observed behavioral patterns. Dynamic risk scores introduced into systems help identify potentially dangerous employee behavior before damage happens to the network. Training programs need to be established for workers as they need education about security protocols, insider warning signs, and risks associated with working under AI supervision. A multi-layered defense system arises from adopting tools that unify technical analytics, behavioral insights, and ethical oversight capabilities. Organizations should adopt proactive mitigation protocols as their organizational norm, which use data trends with human factors to create proactive detection systems rather than reactive ones.

5.4. Challenges and Limitations

Many obstacles and constraints continue to affect the potential of AI for insider threat detection. Implementing AI-based monitoring systems involves ethical problems because they invade employee privacy rights through excessive surveillance practices. Continuous tracking and behavioral analysis create mistrust and resistance among employees when organizations fail to tell them about these activities. Security analysts frequently experience difficulties understanding AI model decision-making processes, especially when operating in critical situations or important decision-making scenarios. Large-scale effective modeling becomes challenging due to dependency on substantial quantities of high-quality data since the absence of diverse training datasets results in inadequate and biased insights. Implementing AI into existing legacy systems meets challenges from resistance stemming from hardware expenses, compatibility issues, and insufficient staff expertise in AI technology. The modeling of insider behaviors becomes imprecise when organizations fail to obtain complete dataset information, particularly among smaller organizations. Based on their motives, insider threats remain unpredictable because algorithms cannot understand these motivations fully. AI represents a robust tool for security operations, although it requires human supervision and ethical governance principles to achieve optimal outcomes that require continuous updates.

5.5. Recommendations

The researchers have developed multiple suggestions that organizations and policymakers should implement based on their investigation results. Organizations must build integrated detection frameworks using technology-based security and human-controlled behavioral analysis procedures. Such security measures provide a full-scale threat detection capability that spans multiple operational fields. Public institutions must create modern guidelines establishing appropriate boundaries for AI ethics in workplace cybersecurity operations between performance triumphs and employee protection. Security teams must perform regular checks on AI systems to determine their performance metrics and test the fairness and interpretability that require changes based on evolving security threats. The ongoing surveillance of insider risks should be performed through an institutional framework for interdepartmental cooperation between IT, HR, and compliance teams. Further money for research should be allocated to develop explainable AI (XAI) technology alongside methods for privacy-preserving monitoring and decentralized anomaly detection systems. Organizations should enforce both mandatory employee awareness programs and mandatory training, which will build a comprehensive understanding of cybersecurity at every level of the organization. These planned measures assist businesses to become future-ready against internal disruption risks.

6. Conclusion

Summary of Key Points

The study analyzed the rising problem of insider threats that affect AI-based cyber systems. Security tools based on traditional methods fail to recognize insider activities because they utilize static designs that overlook behavioral context patterns. The combination of AI systems and machine learning programs proved more effective than traditional methods at uncovering minor behavioral changes in user conduct. Insiders use trust relationships combined with automation and system complexities in their damaging activities, as demonstrated by the case studies of Snowden, Twitter, and Tesla. Evaluation metrics validated that AI tools performed better than conventional methods since they decreased monitoring time and enhanced dangerous activity detection capabilities. The integration of these systems brought forth safety concerns arising from vagueness in their behavior and dependence on these systems and ethical dilemmas. The study illustrated how insider threats interact with automated systems using specific examples alongside presenting effective security measures against such threats. Successful control of insider threats requires defense structures which unite technological assets with human evaluators along with organizational management systems according to this study.

Future Directions

The future of research regarding insider threat detection should develop models that unite AI system efficiency with human-based intuition to work together effectively. AI systems should operate with interpretable programs that enable security staff to verify and trust AI-generated recommendations. Investigating quantum-secure algorithms and evaluating quantum computing effects on threats and defensive capabilities represents a promising research area. The demand for research increases because traditional perimeter defenses cannot protect decentralized and remote work environments from insider threats. Research must develop monitoring methods through machine learning technology, ensuring privacy protection for employee rights. The new AI architectures need continuous learning features and adaptive abilities that keep systems updated with changing behavioral patterns and attack methods. The achievement of transparency and effectiveness in insider threat detection depends on industry-wide collaboration and common benchmarks for detection tools. The directions outlined will assist organizations in their preparation to confront an evolving automated threat environment that grows in complexity.

References

[1] Bécue, A., Praça, I., & Gama, J. (2021). Artificial intelligence, cyber-threats and Industry 4.0: Challenges and opportunities. Artificial Intelligence Review, 54(5).

[2] Blauth, T. F., Gstrein, O. J., & Zwitter, A. (2022). Artificial intelligence crime: An overview of malicious use and abuse of AI. IEEE Access, 10, 77110–77122. https://doi.org/10.1109/ACCESS.2022.3191790

[3] Jimmy, F. (2021). Emerging threats: The latest cybersecurity risks and the role of artificial intelligence in enhancing cybersecurity defenses. International Journal of Scientific Research and Management (IJSRM), 9(2), EC-2021-564–574. https://doi.org/10.18535/ijsrm/v9i2.ec01

[4] Koutsouvelis, V., Shiaeles, S., Ghita, B., & Bendiab, G. (2020). Detection of insider threats using artificial intelligence and visualisation. 2020 6th IEEE Conference on Network Softwarization (NetSoft), Ghent, Belgium, 437–443. https://doi.org/10.1109/NetSoft48620.2020.9165337

[5] Lyon, D. (2015). The Snowden stakes: Challenges for understanding surveillance today. Surveillance & Society, 13(2), 139–152. https://doi.org/10.24908/ss.v13i2.5363

[6] Michael, A., & Eloff, J. (2020). Discovering "insider IT sabotage" based on human behaviour. Information & Computer Security, 28(4), 575–589. https://doi.org/10.1108/ics-12-2019-0141

[7] Nasir, R., Afzal, M., Latif, R., & Iqbal, W. (2021). Behavioral based insider threat detection using deep learning. IEEE Access, 9, 143266–143274. https://doi.org/10.1109/ACCESS.2021.3118297

[8] Nurse, J., Buckley, O., Legg, P., Goldsmith, M., Creese, S., Wright, G., & Whitty, M. (2014). Understanding insider threat: A framework for characterising attacks. 2014 IEEE Security and Privacy Workshops. https://doi.org/10.1109/SPW.2014.38

[9] Oosthoek, K., & Doerr, C. (2021). Cyber security threats to Bitcoin exchanges: Adversary exploitation and laundering techniques. IEEE Transactions on Network and Service Management, 18(2), 1616–1628. https://doi.org/10.1109/TNSM.2020.3046145

[10] Padayachee, K. (2013). A conceptual opportunity-based framework to mitigate the insider threat. 2013 Information Security for South Africa (ISSA), Johannesburg, South Africa, 1–8. https://doi.org/10.1109/ISSA.2013.6641060

[11] Sarma, M., Matheus, T., & Senaratne, C. (2020). Artificial intelligence and cyber security: A new pathway for growth in emerging economies via the knowledge economy? In Business Practices, Growth and Economic Policy in Emerging Markets (pp. 51–67). https://doi.org/10.1142/9789811221750_0004

[12] Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). AI-driven cybersecurity: An overview, security intelligence modeling and research directions. SN Computer Science, 2(3). https://doi.org/10.1007/s42979-021-00557-0

[13] Schmidt, P., & Biessmann, F. (2020). Calibrating human-AI collaboration: Impact of risk, ambiguity and transparency on algorithmic bias. Lecture Notes in Computer Science, 431–449. https://doi.org/10.1007/978-3-030-57321-8_24

[14] Tomsett, R., Preece, A., Braines, D., Cerutti, F., Chakraborty, S., Srivastava, M., Pearson, G., & Kaplan, L. (2020). Rapid trust calibration through interpretable and uncertainty-aware AI. Patterns, 1(4), 100049. https://doi.org/10.1016/j.patter.2020.100049

[15] Yuan, S., & Wu, X. (2021). Deep learning for insider threat detection: Review, challenges and opportunities. Computers & Security, 104, 102221. https://doi.org/10.1016/j.cose.2021.102221