Article

AI-Driven Threat Detection and Automated Incident Response for Securing Cloud Workloads

Anton Chagovec , Teodora Bakardjieva , Antonina Ivanova * , Fatima Sapundzhi * , Veselina Spasova , Andriana Ivanova

Posted Date: 20 May 2026

doi: 10.20944/preprints202605.1325.v1

Keywords: artificial intelligence; cloud security; threat detection; automated incident response; SIEM; XDR; Security Operations Centre; ransomware; machine learning; behavioral analytics

 doi:10.20944/preprints202605.1325.v1

Article

AI-Driven Threat Detection and Automated Incident Response for Securing Cloud Workloads

Anton Chagovec [1], Teodora Bakardjieva [1], Antonina Ivanova [1,*], Fatima Sapundzhi [1,2,*], Veselina Spasova [1,3] and Andriana Ivanova [1]

[1] Varna Free University "Chernorizets Hrabar", 84 Yanko Slavchev Str., Chaika Resort, 9007 Varna, Bulgaria
[2] South-West University "Neofit Rilski", 66 Ivan Mihaylov Str., 2700, Blagoevgrad, Bulgaria
[3] Department of Informational and Communication Technology, Faculty of Engineering, Nikola Vaptsarov Naval Academy, 73 Vasil Drumev, 9002, Varna, Bulgaria
* Correspondence: antonina.ivanova@vfu.bg (A.I.); sapundzhi@swu.bg (F.S.)

Abstract

The increasing adoption of cloud computing has expanded organizational attack surfaces and increased exposure to credential compromise, ransomware, and cloud misconfiguration. Conventional security information and event management (SIEM) systems, based primarily on static correlation rules and signature-based detection, often struggle to process heterogeneous cloud telemetry and prioritize high-severity incidents in real time. This study evaluates the operational impact of an artificial intelligence (AI)-augmented SIEM and Extended Detection and Response (XDR) architecture for cloud threat detection and automated incident response. A mixed-methods comparative case study was conducted across two enterprise-style security environments: an AI-augmented cloud-native SIEM/XDR architecture and a conventional baseline environment based on manual triage and signature-based controls. Three attack scenarios were analyzed: phishing-led account takeover, multi-stage ransomware, and shadow-IT data exfiltration. The AI-augmented environment reduced mean time to triage from 17.4 hours in the conventional baseline to 10.7 minutes and enabled ransomware containment in under five minutes through automated response playbooks. The results also showed improved prioritization of high-severity incidents, reduced analyst workload, and a high automated closure rate. However, limitations were observed in the calibration of behavioral models, vendor dependency, and detection gaps involving legitimate third-party services and password-protected content. The findings should be interpreted as operational evidence at an architectural level, not as an isolated evaluation of individual AI models.

Keywords: artificial intelligence; cloud security; threat detection; automated incident response; SIEM; XDR; Security Operations Centre; ransomware; machine learning; behavioral analytics

1. Introduction

Cloud computing is increasingly used as the primary infrastructure model in enterprise IT. Gartner forecasts global public cloud end-user spending to reach USD 723 billion in 2025, with Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) continuing to grow [1]. The expansion of cloud services increases the number of externally accessible systems, identities, APIs, and distributed workloads. SentinelOne reports that 80% of organizations have experienced increased attack activity in cloud environments, with data breaches, lateral movement, and cryptomining among the most frequently reported incident categories [2].

Conventional security mechanisms based on static rules and signature matching encounter limitations in large-scale cloud environments characterized by distributed infrastructure, dynamic workloads, and heterogeneous telemetry sources. Attack techniques based on credential compromise, zero-day exploitation, and multi-stage intrusion chains frequently bypass fixed

 doi:10.20944/preprints202605.1325.v1

detection rules. Large alert volumes also increase analyst workload and contribute to alert fatigue, where relevant incidents may be delayed during triage and investigation processes [3].

From an applied-science perspective, the main challenge is not only whether AI models can detect anomalies, but whether integrated AI-assisted SIEM/XDR workflows measurably improve operational security processes, including triage, containment, prioritization, and analyst workload reduction in cloud environments.

AI and ML techniques are increasingly incorporated into cloud security monitoring platforms to support behavioral analysis, anomaly detection, and automated incident correlation. Two technology categories are central to current enterprise cloud security operations: next-generation Security Information and Event Management (SIEM) systems and Extended Detection and Response (XDR) platforms. Existing studies and industry reports describe the operational capabilities of these platforms, although comparative empirical evaluations across multiple attack scenarios and conventional baseline environments remain limited [4].

1.1. Research Gap and Objectives

Current literature reveals several recurring limitations in the empirical evaluation of AI-driven cloud security monitoring systems. Many studies focus on algorithm-level detection accuracy, isolated attack scenarios, or vendor-specific platform descriptions, while fewer studies evaluate the operational effect of AI-assisted correlation and automated response on Security Operations Centre (SOC) workflows under production-style conditions. This study therefore focuses on operational metrics such as triage time, closure time, incident correlation ratio, false-positive activity, true-positive classification, and automated closure rate.

The study addresses the following research objectives:

- RO1: To examine the effect of AI-enhanced SIEM and XDR platforms on detection and response time across multiple cloud attack scenarios.
- RO2: To evaluate the reduction of false-positive incidents through behavioral analytics and AI-assisted event correlation in comparison with a conventional rule-based environment.
- RO3: To analyze the effect of AI-assisted incident correlation and automated response mechanisms on analyst workload and incident handling efficiency.
- RO4: To identify operational limitations and governance considerations associated with AI-supported incident response in enterprise cloud environments.

The main contributions of the study are follows:

- A comparative evaluation of AI-augmented SIEM/XDR and conventional security operations across three cloud attack scenarios.
- A production-oriented measurement model using triage time, closure time, incident correlation ratio, false-positive rate, true-positive rate, and automated closure rate.
- An analysis of AI-assisted correlation and automated response playbooks in phishing-led account takeover, ransomware, and shadow-IT exfiltration scenarios.
- Identification of operational limitations related to behavioral model calibration, legitimate third-party services, encrypted content, vendor dependence, and governance of automated response.

In this study, AI-assisted SIEM/XDR workflows are being evaluated at the architectural level, not as separate machine learning models. Unlike model-centric intrusion detection studies that evaluate isolated classifiers, this study evaluates AI-assisted SIEM/XDR workflows at an architectural level using production-style incident data, scenario-based analysis, and SOC-oriented operational metrics.

1.2. Cloud Security Threat Landscape

Cloud-related risks such as misconfiguration, weak identity and access management, insecure APIs, and insufficient monitoring are consistently discussed in cloud security literature and industry threat reports [5–8].

 doi:10.20944/preprints202605.1325.v1

Compromised credentials remain a common factor in cloud incidents, and the ENISA Threat Landscape report identifies ransomware and data theft among the most persistent threats affecting cloud-connected infrastructure [6,7]. Akamai reports that 84% of surveyed security professionals experienced API-related security incidents during 2024 [8]. At the same time, the global cybersecurity workforce shortage continues to affect operational response capacity. ISC² estimates a shortage of 4.76 million cybersecurity professionals worldwide [9].

1.3. SIEM Evolution and AI Integration

SIEM platforms were initially developed for log aggregation, event correlation, and compliance monitoring. Earlier systems relied primarily on predefined correlation rules and signature-based detection mechanisms. Cloud-native SIEM architectures increasingly incorporate behavioral analytics, anomaly detection, and ML-based correlation mechanisms to process large volumes of telemetry data from distributed infrastructure. Gartner defines modern SIEM platforms as systems that aggregate and analyze event data from cloud and on-premises infrastructure [10].

User and Entity Behavior Analytics (UEBA) extends SIEM functionality through behavioral profiling and anomaly detection based on historical activity patterns. AI-assisted correlation engines further aggregate related events from multiple telemetry sources into contextual incidents requiring analyst attention. CrowdStrike describes AI-powered SIEM environments as capable of reducing alert noise and improving prioritization of high-severity events [11].

1.4. XDR Platforms and AI Integration

Extended Detection and Response (XDR) platforms extend the scope of endpoint-focused monitoring by integrating telemetry from endpoints, networks, cloud workloads, identity systems, email infrastructure, and SaaS applications. Recent studies also examine telemetry normalization and orchestration mechanisms for scalable security monitoring across distributed multi-cloud infrastructures [12,13]. Palo Alto Networks defines XDR as a technology combining multiple telemetry sources for threat detection and response [14]. Gartner describes XDR platforms as integrated environments supporting automated investigation and cross-domain event correlation [15].

Behavioral analysis and ML-assisted anomaly detection represent central operational components in current XDR architectures. Reinforcement-learning approaches are also being investigated for adaptive policy optimization and automated response adjustment [16]. Large language models (LLMs) are increasingly incorporated into investigation workflows for contextual summarization, natural-language querying, and response orchestration support.

1.5. Automated Incident Response

Security Orchestration, Automation, and Response (SOAR) capabilities are increasingly integrated into SIEM and XDR platforms to support automated execution of response procedures following threat detection events [17]. Cloud security automation architectures combining AI-assisted detection, orchestration workflows, and automated response coordination are discussed in recent enterprise cloud security studies [18]. AI-assisted SOC environments have also been associated with improved incident prioritization and reduced operational workload during large-scale security monitoring [19]. Recent empirical evaluations of ATT&CK-aligned SIEM environments also report improvements in detection coverage, alert prioritization, and operational SOC workflows in enterprise monitoring scenarios [20]. Existing research also notes that comparative empirical evaluations involving multiple attack scenarios, heterogeneous infrastructures, and conventional baseline environments remain relatively limited [21].

1.6. Theoretical Framework

 doi:10.20944/preprints202605.1325.v1

This study is based on the Detect–Respond–Recover model of the NIST Cybersecurity Framework [22] and the Zero Trust Architecture principles defined in NIST SP 800-207 [23], which require continuous verification of entities independent of network location. Recent cloud security research also examines hybrid deep-learning approaches operating within Zero-Trust architectures for real-time threat detection in distributed cloud environments [24]. Dynamic capabilities theory [25] provides the conceptual basis for examining AI integration in SIEM/XDR environments through three operational dimensions: real-time behavioral anomaly detection across cloud telemetry, automated response orchestration through security playbooks, and adaptive ML model adjustment based on incident feedback. The evaluation metrics-MTTD, MTTR, false-positive rate, and incident-correlation precision-correspond to SOC performance indicators commonly used in MITRE ATT&CK-oriented detection engineering practice [4].

2. Materials and Methods

2.1. Research Design

This study employs a sequential mixed-methods design [26]. The quantitative phase measures operational detection and response metrics across three attack scenarios in two enterprise-style environments. The qualitative phase interprets selected production incident cases to explain how AI-assisted correlation and automated response affected handling in practice. Because the evaluation was conducted in production-style security environments, formal randomization and controlled replay of identical attacks were not feasible. Therefore, the reported metrics are interpreted as descriptive operational indicators rather than formal causal estimates.

The research design combines three layers of evidence: (i) quantitative incident metrics extracted from SIEM/ XDR and ticketing records; (ii) scenario-based analysis of three representative cloud attack chains, and (iii) qualitative interpretation of operational limitations and governance requirements.

2.2. Evaluation Environments

2.2.1. Environment A: AI-Augmented SIEM/XDR Architecture

Environment A represents the AI-enhanced architecture under primary evaluation. It comprises a cloud-native, data-lake-based SIEM with integrated UEBA and ML-driven correlation; an XDR layer providing unified telemetry ingestion from endpoints, cloud workloads, identity systems, email, and network sources; and a generative-AI investigation assistant supporting natural-language queries, incident summarization, and response playbook coordination. The environment includes enterprise – grade protection capabilities for endpoints, identities, cloud applications, and email threat monitoring. The evaluation focused on incident correlation and automated response behavior within the deployed production - style security environment. Enterprise-grade security licensing covers endpoint protection, identity risk scoring, cloud application security, and email threat protection.

Figure 1 illustrates the generalized automated incident response architecture deployed in Environment A, from multi-source telemetry ingestion through AI-assisted correlation to SOAR-based containment actions and analyst escalation. The architecture was applied across the three attack scenarios presented in Section 2.3.

 doi:10.20944/preprints202605.1325.v1

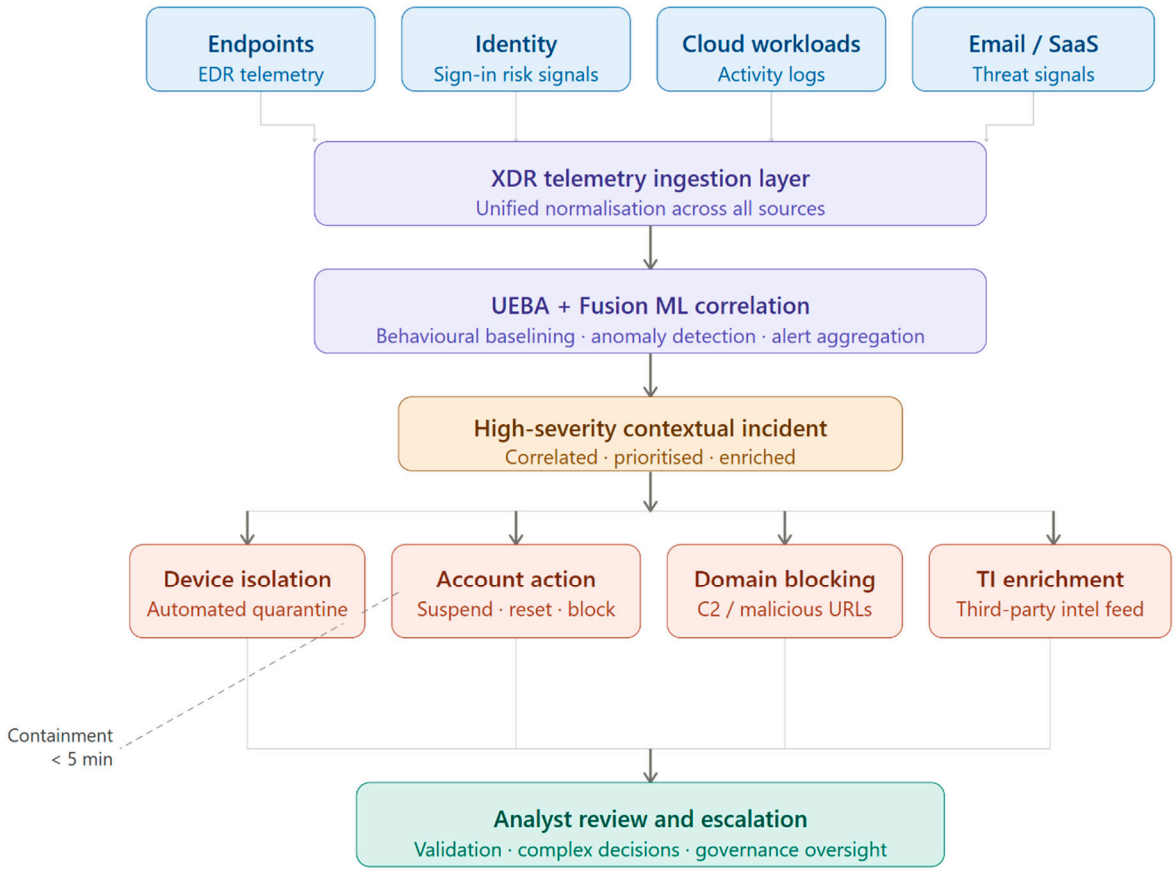

Figure 1. AI-Augmented SIEM/XDR incident detection and response architecture.

The comparison between Environment A and Environment B should therefore be interpreted as an architecture-level operational comparison rather than an isolated evaluation of AI algorithms. The two environments differ not only in the presence of AI-assisted analytics, but also in telemetry integration, automation maturity, identity protection, endpoint coverage, and response orchestration. The reported improvements indicate the operational effect of the integrated AI-enhanced SIEM/XDR architecture as a whole, rather than the casual contribution of a single AI module.

2.2.2. Environment B: Conventional Baseline

Environment B represents a conventional security monitoring baseline: legacy signature-based endpoint protection, on-premises Active Directory without cloud identity integration, perimeter firewall rules, limited cloud threat protection capabilities, and manual triage and response workflows. This configuration is representative of organizations that have not yet adopted AI-augmented SIEM/XDR and SOAR capabilities. Product-specific identifiers and selected operational details were generalized to preserve organizational confidentiality and reduce vendor-specific bias.

2.3. Attack Scenarios

The scenarios were extracted from records of production incidents and operational security exercises. Sensitive identifiers, tenant information, user data, infrastructure details, and product-specific operational information were anonymized or aggregated. Three attack chains were selected to represent prevalent and impactful threat categories targeting cloud workloads, consistent with ENISA threat landscape categories and cloud security literature [6,7].

Scenario 1 – Phishing-Led Account Takeover (S1): A phishing email was delivered to a target user via a legitimate third-party service, leading to credential theft. The attacker subsequently

 doi:10.20944/preprints202605.1325.v1

accessed cloud services from an anomalous geographic location, performed lateral movement within the cloud tenant, and exfiltrated data from a SaaS application.

Scenario 2 – Multi-Stage Ransomware (S2): Initial access was achieved via a compromised personal device belonging to a privileged administrator over VPN. The attack chain included credential theft tool deployment (Mimikatz-equivalent), defense evasion, command-and-control establishment (Cobalt Strike-equivalent), and ransomware execution targeting cloud-connected server workloads, spanning 220+ entities across two servers.

Scenario 3 – Shadow-IT Data Exfiltration (S3): A user transferred sensitive organizational data to an unsanctioned cloud application using a password-protected document distributed through a legitimate document-signing service. The use of encrypted content and trusted third-party infrastructure bypassed standard DLP inspection and delayed identification during the initial transfer stage (Table 1).

Table 1. Attack scenarios, telemetry sources, and evaluated AI/XDR functions.

Scenario	Definition	Initial Vector	Main stages of the attack	Telemetry Sources	AI/XDR Feature Assessment
S1:	Account takeover via phishing	Phishing via a legitimate third-party service	Credential theft; anomalous login; lateral movement; SaaS data access	Email, Identity, SaaS, Cloud Activity	Identity Risk Assessment; Multi-Source Correlation; Session Shutdown and Password Reset
S2:	Multi-stage ransomware	Compromised privileged device via VPN	Credential theft; defense evasion; C2; ransomware execution	Endpoint, VPN, Identity, Cloud-Connected Servers	Multi-Stage Fusion Correlation; Device Isolation; Account Deactivation; Enrichment
S3:	Shadow-IT data exfiltration	Unsanctioned cloud application and password-protected document	Sensitive data transfer; impossible travel; unusual download patterns	CASB/SaaS, Identity, Network, Cloud Application Logs	Behavioral Anomaly Detection; App Blocking; Analyst Escalation

2.4. Metrics and Analytical Approach

The following metrics were collected for both environments across all three scenarios. Based on the operational records that did not consistently separate detection, confirmation, and response timestamps, this study used mean time to triage (MTT) and mean time to closure (MTC) as operational metrics for SOC. They are treated as practical workflow metrics and not as direct substitutes for the standard metrics of MTTD and MTTR (Table 2):

- Mean Time to Triage (MTT): Elapsed time from incident registration to first analyst or automated review action.
- Mean Time to Closure (MTC): Elapsed time from initial triage to full incident resolution.
- Incident Correlation Ratio (ICR): Ratio of raw alerts to AI-correlated contextual incidents.
- False-Positive Rate (FPR): Proportion of reviewed incidents classified as False Positive upon investigation.
- True-Positive Rate (TPR): Proportion of reviewed incidents confirmed as genuine threats.

 doi:10.20944/preprints202605.1325.v1

- Automated Closure Rate (ACR): Proportion of incidents closed automatically without analyst intervention.

Table 2. Evaluation metrics and calculation method.

Metric	Definition	Calculation
MTT	Time from incident registration to first analyst/automated action	First triage timestamp-incident timestamp
MTC	Time from triage to closure	Closure timestamp-triage timestamp
ICR	Alert aggregation ratio	Total raw alerts/AI-correlated incidents
FPR	False-positive share	False positive/ reviewed incidents
TPR	Confirmed threat share	True positives/ reviewed incidents
ACR	Automated closure rate	Auto-closed incidents/ total incidents

Data were extracted from SIEM incident management tables (SecurityIncident) over a 30-day production measurement period. Fusion/Multi-stage AI-correlated incidents were identified by title pattern matching. Incident classifications (TruePositive, FalsePositive, BenignPositive, Undetermined) were recorded at closure. Classification outcomes were validated through operational incident review procedures conducted during the normal SOC workflow. Environment B metrics were derived from manual incident logs and ticketing records over an equivalent period. Statistical comparisons between environments are reported as percentage changes and effect ratios.

The comparison does not isolate the effect of AI as a single independent variable. Instead, it evaluates the operational impact of an integrated, AI-enhanced SIEM/XDR security architecture compared to a conventional monitoring baseline. Therefore, the results should be interpreted as operational differences at the architecture level, rather than as causal assessments of individual AI components.

The two evaluated environments differed in architecture, telemetry coverage, automation maturity, tool integration, and operational constraints, and no formal inferential testing was applied. Therefore, the comparison was based on descriptive operational indicators, percentage changes and effect ratios. As a result of this approach, operational performance can be interpreted from an architectural perspective under production style conditions, while more controlled studies would be needed to assess the causal contribution of each component.

To improve comparability, only incidents belonging to equivalent operational categories were included in the comparison: identity, endpoint, ransomware-related activity, cloud application abuse, and suspicious data transfer. Informational events and automatically generated low-risk notifications without analyst relevance were excluded from the Environment B false-positive estimate. For Environment A, classification labels were obtained from SIEM/XDR incident records. For Environment B, classifications were reconstructed from analyst ticket notes and closure categories. Because Environment B lacked automated metadata, its false-positive rate is reported as an estimated operational indicator.

2.5. Environment Configuration Summary

The evaluated environments differed substantially in architecture, security tooling, automation capabilities, licensing scope, and operational integration, representing two distinct maturity levels of enterprise cloud security monitoring that reflect deployment configurations commonly observed in contemporary organizational practice. Table 3 summarizes the main configuration parameters used in the comparative evaluation.

Where possible, product names, licensing labels, and operational identifiers were generalized to preserve confidentiality and to reduce vendor bias.

Table 3. Environment Configurations and Evaluation Parameters.

 doi:10.20944/preprints202605.1325.v1

Parameter	Environment A (AI-Augmented)	Environment B (Conventional Baseline)
SIEM Platform	Cloud-native data-lake SIEM with ML/UEBA and AI-assisted multi-stage correlation engine (like Fusion)	None integrated SIEM; manual log review only
XDR Layer	Integrated XDR layer covering endpoint, cloud, identity, e-mail and network telemetry (Vendor-agnostic XDR)	Siloed point tools: perimeter firewall and legacy endpoint protection (Cisco ASA/ Symantec SEP)
Endpoint Protection	Cloud-delivered endpoint protection with behavioral and static AI engines (Next-gen AI/ML)	Legacy signature-based endpoint protection (like Symantec SEP)
Identity Protection	Cloud identity protection with ML risk scoring and conditional access policies	On-premises identity management only; no cloud risk scoring
Email Security	AI-assisted e-mail threat protection with URL rewriting, attachment sandboxing, and post-delivery remediation.	Basic e-mail filtering only
Gen-AI Assistant	Generative AI investigation assistant supporting natural-language query, summarization, and playbook coordination (like GPT)	Not available
Response Automation	SOAR-integrated playbooks for auto-isolation, account locking, and domain/application blocking	Manual only; no automation
Licensing	Enterprise-grade integrated security licensing (like E5-equivalent)	Limited cloud productivity/ security licensing
Measurement Period	30 days (production)	30 days (production)

The product names and licensing labels were summarized to preserve confidentiality and reduce bias toward a specific vendor.

3. Results

3.1. Measurement Model Validation: Alert Correlation Performance

To validate the AI correlation model performance in Environment A, incident classification outcomes were analyzed across the full 30-day period. The Fusion/Multi-stage correlation engine generated 66 AI-correlated incidents from a base of 2,694 total logged alerts, corresponding to an incident correlation ratio (ICR) of approximately 41:1, indicating that each AI-correlated case aggregated an average of 41:1.

The classification profile of Fusion-correlated incidents indicates that the AI-assisted correlation model was effective in prioritizing investigation-relevant activity. Of the 66 Fusion incidents, 54 incidents, or 81.8%, were classified as BenignPositive or TruePositive and therefore required analyst attention. Specifically, 5 incidents, or 7.6% were confirmed as TruePositive active threats, while 49 incidents, or 74.2% were classified as BenignPositive suspicious but expected activity. The remaining 12 incidents, or 18.2% were closed without definitive classification. These results suggest that the

 doi:10.20944/preprints202605.1325.v1

correlation model functioned primarily as a prioritization and contextualization mechanism rather than as a binary threat classifier.

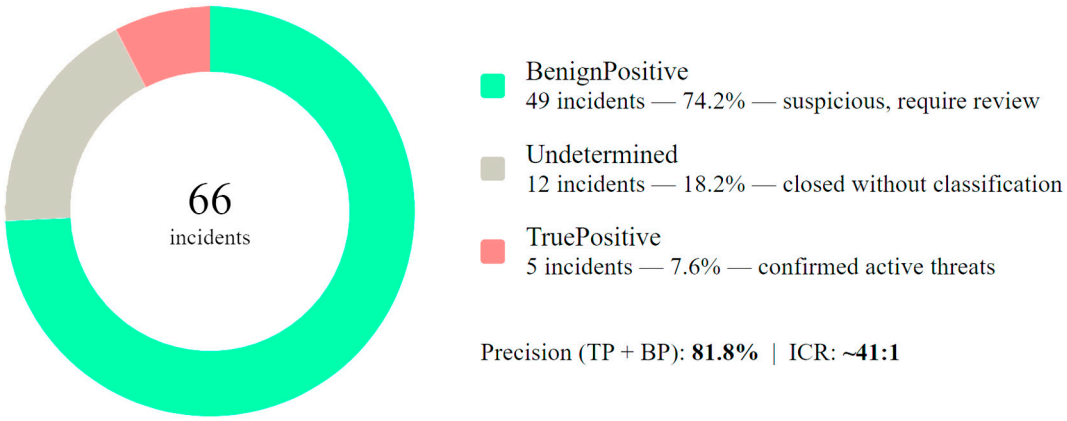

Figure 2. Classification outcomes of AI-correlated Fusion incidents in Environment A during the 30-day measurement period (n = 66).

In this study, incidents with benign positive characteristics (BenignPositive) are not treated as false alarms. They represent suspicious but expected activities that require validation by the analyst. Therefore, Fusion Engine is primarily evaluated as a prioritization and correlation mechanism, not just as a binary threat classifier.

The BenignPositive category accounts for the majority of Fusion-correlated incidents, indicating that the correlation engine predominantly surfaced suspicious activity requiring analyst review rather than confirmed active threats.

For the full incident population, the XDR layer automatically closed 89.42% of incidents (n = 2,341 of 2,628 non-Fusion incidents) without analyst intervention, corresponding to an automated closure rate (ACR) of 89.4%. The analyst reviewed population comprised 277 incidents, of which 258 (93.1%) were confirmed BenignPositive and 19 (6.9%) were confirmed TruePositive - a true-positive rate (TPR) of 0.73% of all incidents, representing the confirmed genuine threat caseload for human analysts over 30 days.

3.2. Structural Model Results: Severity Distribution

The incident classification results and severity distribution observed during the 30-day evaluation period in Environment A are summarized in Table 4, providing a disaggregated view of AI-correlated and non-AI incidents across all severity levels and outcome categories.

Table 4. Incident Classification and Severity Distribution (30-Day Period, Environment A, N = 2,694).

Metric	AI-Fusion Incidents (n=66)	Non-AI Incidents (n=2,628)	Total (N=2,694)
Auto-closed by XDR (Undetermined)	12 (18.2%)	2,341 (89.4%)	2,353
BenignPositive (analyst confirmed)	49 (74.2%)	258 (9.85%)	307
TruePositive (confirmed threat)	5 (7.6%)	19 (0.73%)	24
High severity	49 (74.2%)	200 (7.6%)	249
Medium severity	12 (18.2%)	266 (10.2%)	278
Low severity	1 (1.5%)	48 (1.8%)	49
Informational	4 (6.1%)	2,104 (80.4%)	2,108

 doi:10.20944/preprints202605.1325.v1

Incident Correlation Ratio (ICR)	~41:1 (alerts to incidents)	1:1 (no aggregation)	-

* Fusion incidents demonstrated a 74.2% High-severity rate versus 7.6% for non-AI incidents, indicating prioritization of higher-severity incidents by the AI correlation engine (ratio: 9.7×).

The severity distribution indicates that the Fusion ML engine preferentially surfaced higher-severity incidents from the broader alert population: High-severity cases increased from 7.6% in the non-AI incident population to 74.2% among AI-correlated cases, representing a 9.7-fold enrichment. This behavior suggests that the correlation layer supported prioritization of incidents requiring analyst attention and is consistent with RO3.

3.3. False-Positive Reduction (RO2)

Behavior-profiling mechanisms in the XDR layer were associated with lower volumes of false-positive activity incidents compared with Environment B, where alert classification relied primarily on static rules and manual review procedures. In Environment A, the proportion of analyst-reviewed incidents classified as FalsePositive was approximately 10.6% during the measurement period. For Environment B, the false-positive rate was derived from manually reviewed ticketing records using the same operational classification criteria applied in Environment A. Manual incident records from Environment B show significantly higher false positive activity, estimated at approximately 62% of the analyst-reviewed baseline signals. Since the baseline environment did not provide automated classification metadata, the value is reported as an estimate based on analyst-reviewed records.

In this study, BenignPositive category indicates were not treated as false positives because they represented suspicious activities requiring analyst validation. FalsePositive was reserved for incidents that were determined to have no security relevance after review.

The two environments differed in architecture, telemetry coverage, and operational workflows; therefore, the comparison is interpreted as a descriptive operational observation rather than a controlled statistical measurement. The observed difference between the evaluated environments indicates that behavioral analytics and AI-assisted correlation mechanisms can reduce repetitive analyst triage activity and improve alert prioritization efficiency in cloud security operations.

3.4. Detection and Response Timing (RO1, RO4)

Table 5 presents the comparative detection and response metrics for the evaluated environments, including mean triage time, incident closure time, false-positive rate, and ransomware containment performance.

Table 5. Detection and Response Time Comparison: Environment A vs. Environment B.

Metric	Environment A (AI-Augmented)	Environment B (Conventional)
Mean Time to Triage (MTT)	10.7 minutes	17.4 hours
Mean Time to Closure (MTC)	23.8 minutes	35.9 minutes*
MTT Improvement Factor	97.6× faster	- (baseline)
Ransomware containment time	< 5 minutes (automated playbook)	~90+ minutes (manual)
Ransomware response improvement	18× faster	- (baseline)
Automated closure rate (ACR)	89.4% of all incidents	0% (manual only)
False-positive rate (FPR)	~10.6%	~62% (estimated)
Analyst triage time reduction	approximately 99%	- (baseline)
SLA compliance (MTT < 20 min)	Yes (10.7 min)	No (17.4 hours)

 doi:10.20944/preprints202605.1325.v1

* MTC in Environment B is understated as it excludes incidents not yet resolved within the measurement window. MTT and ransomware containment time provide more reliable indicators than MTC. The MTT comparison (17.4 hours versus 10.7 minutes) represents an approximately 97.6-fold reduction in triage time.

Environment A achieved an MTT of 10.7 minutes, remaining below the contractual SLA threshold of 20 minutes. Environment B's MTT of 17.4 hours substantially exceeded the same threshold, reflecting the structural bottleneck of manual triage without automated pre-classification. These findings are consistent with RO1, indicating that AI-enhanced SIEM/XDR environments can reduce incident triage and response intervals from the hour scale to the minute scale in the evaluated context.

3.5. Qualitative Findings: Attack Scenario Outcomes (RO4)

3.5.1. Scenario S1: Phishing-Led Account Takeover

The ML Fusion engine correlated four distinct incident signals-malicious IP access, atypical geographic travel, unfamiliar sign-in properties, and password spray activity-into a single High-severity contextual incident (Incident #4163) within minutes of initial attack activity. Identity protection services detected the anomalous login pattern; automated playbooks suspended the compromised user session and triggered a password reset. The phishing vector was not intercepted at email delivery due to the use of a password-protected document delivered via a legitimate third-party service (a persistent limitation of current AI email filters). Post-compromise containment was achieved rapidly, and protection was extended to other potentially targeted users through automated classifier updates.

3.5.2. Scenario S2: Multi-Stage Ransomware

The ransomware scenario (Incident #167044) generated more than 220 entities across two servers and 15 correlated sub-incidents. The Fusion engine aggregated these signals into a single High-severity multi-stage incident within minutes of the first alert. Automated playbooks isolated affected devices, disabled compromised accounts, and initiated third-party threat intelligence enrichment. A senior analyst war room was convened given the high privilege level of the affected account (domain administrator). Total time from first alert to initial automated containment actions was under five minutes (18-fold improvement over the estimated manual response time of 90+ minutes), supporting RO4. The attacker's failure to account for the presence of cloud-delivered endpoint protection on targeted workloads was a decisive factor in rapid containment.

3.5.3. Scenario S3: Shadow-IT Data Exfiltration

Cloud application security monitoring detected impossible-travel anomalies and unusual download volume patterns consistent with data exfiltration. Automated policies blocked access to the unsanctioned application and generated an analyst alert. The scenario demonstrated the ability of behavioral monitoring to identify suspicious exfiltration patterns after the initial transfer. The password-protected document distributed through a legitimate third-party service was not intercepted at ingestion, which highlights a remaining limitation in content inspection and automated classification. This scenario illustrates the continued importance of layered controls and human oversight and indicates that AI-driven systems are not complete replacements for analyst judgment in novel social-engineering attacks.

4. Discussion

4.1. Interpretation of Findings

 doi:10.20944/preprints202605.1325.v1

The evaluation results indicate that AI-enabled SIEM/XDR architectures can improve detection prioritization and response speed in cloud security operations. Across the evaluated attack scenarios, the AI-augmented environment demonstrated shorter triage intervals and faster containment of high-severity incidents. The operational differences between the evaluated environments were most visible during multi-stage attack activity involving correlated telemetry from cloud workloads, identity systems, and endpoints.

Regarding RO1, the most visible operational effect was the reduction in triage time. With respect to RO2 and RO3, the AI-assisted correlation layer reduced alert fragmentation by aggregating related signals into contextual incidents and by escalating a greater proportion of high-severity cases for analyst review. Regarding RO4, the scenarios also revealed persistent limitations in the abuse of legitimate services, the verification of encrypted content, and the management of automated actions.

Environment A aggregated related alerts from multiple telemetry sources into contextual incidents with broader investigative context [17]. Incident activity associated with endpoints, cloud services, user identities, and network events was processed within unified investigation chains. The severity distribution of Fusion-correlated incidents also showed a substantially higher proportion of High-severity cases compared with the non-AI incident population.

In the ransomware scenario, automated response orchestration reduced the interval between initial detection and containment activity. Automated isolation procedures and identity-based response actions reduced the propagation window and limited the number of systems requiring manual remediation.

Differences were also observed in analyst workload associated with false-positive investigation and repetitive triage activity. Environment B relied primarily on static rules and manual review procedures. Investigation overhead increased during larger alert volumes, and prioritization required repeated analyst intervention. Environment A incorporated behavioral baselining and contextual event correlation before escalation to analysts, reducing the volume of alerts requiring direct review.

The qualitative findings also indicate the continued importance of human oversight in AI-assisted security operations. Password-protected documents distributed through legitimate third-party services were not intercepted during the initial delivery stage. These scenarios demonstrate the operational limitations of AI-supported monitoring in cloud environments and support the continued use of layered security controls, escalation procedures, and analyst supervision.

Behavioral analytics, SIEM/XDR integration, and SOAR automation improve SOC alarm prioritization and reduce alarm noise in modern environments. Unlike model-based intrusion detection studies that primarily look at classification accuracy, this study emphasizes operational SOC metrics and response behavior at the architectural level.

4.2. Practical Implications

The findings indicate that integrated SIEM/XDR environments with behavioral analytics and automated response capabilities can improve operational efficiency in enterprise cloud security monitoring. The evaluated architecture reduced the volume of incidents requiring manual review and shortened the response cycle for high-severity events.

Generative AI assistants were used during incident investigations to generate natural-language summaries of correlated alerts and provide contextual information related to affected entities and attack activity. In the evaluated environment, these functions reduced part of the manual effort associated with triage and multi-stage incident review. Analyst involvement remained necessary for incident validation and escalation decisions, particularly in scenarios involving ambiguous activity patterns or incomplete contextual data.

The comparison between the evaluated environments also illustrates the limitations of conventional rule-based monitoring approaches in large-scale cloud infrastructures. Manual triage and fragmented telemetry sources increased investigation time and reduced prioritization efficiency in the baseline environment.

 doi:10.20944/preprints202605.1325.v1

4.3. Limitations

Several limitations should be considered when interpreting the results. The evaluation was conducted within a single enterprise-style cloud environment and primarily within one vendor ecosystem. Different infrastructure configurations, telemetry coverage, organizational security practices, and operational maturity may influence the transferability of the reported results to other enterprise environments.

Behavioral profiling models also required an initial calibration period before stable classification performance was achieved. During this phase, anomaly detection precision was lower than in the later stages of operation. Certain attack vectors involving legitimate services and password-protected content were not detected during initial delivery stages, indicating limitations in behavioral and content-based inspection mechanisms.

The measurement period was limited to 30 days and focused on a selected set of attack scenarios. Additional long-term evaluations across heterogeneous multi-cloud environments may produce different correlation and response characteristics. The use of generative-AI assistants also introduces the possibility of incorrect contextual interpretation or excessive reliance on automated recommendations during incident analysis.

The absence of formal statistical testing represents a further methodological limitation. As the evaluation was conducted within live production environments, controlled randomization and hypothesis testing were not feasible. Findings should be interpreted as exploratory and case-specific rather than generalizable.

It is also possible that the same attack traces could not be reproduced under identical controlled conditions in both environments. This comparison reflects operational differences between two security architectures rather than an experimental controlled isolation of individual AI components. Differences in telemetry coverage, licensing level, analyst workflow maturity, and availability of automation may have contributed to the observed performance differences.

4.4. Ethical and Governance Considerations

Automated response mechanisms in production cloud environments require clearly defined governance procedures and escalation policies. Actions such as account suspension, device isolation, or automated blocking may affect operational continuity if triggered incorrectly. The evaluated ransomware scenario demonstrated the importance of combining automated response with analyst supervision and escalation procedures for high-impact incidents.

The increasing use of AI techniques in offensive operations introduces additional challenges for cloud security monitoring. AI-assisted social engineering, adversarial ML techniques, and automated evasion mechanisms may reduce the effectiveness of existing detection models. These developments require continuous model adaptation, monitoring procedures, and periodic validation of automated response mechanisms.

5. Conclusions

This study examined the application of AI-enabled SIEM and XDR technologies for cloud threat detection and automated incident response in comparison with a conventional security environment. The evaluation should be interpreted as a production-oriented comparative case study of two security architectures rather than as a controlled laboratory benchmark of individual AI models.

In the evaluated environment, the AI-augmented SIEM/XDR architecture was associated with reduced mean time to triage, lower observed false-positive activity, and more selective escalation of high-severity incidents through behavioral analysis and multi-source event correlation. Automated response playbooks reduced ransomware containment time and limited the number of incidents requiring direct analyst intervention. The evaluation also showed that generative-AI assistants can support investigation activities through contextual summarization and response coordination, while analyst supervision remains necessary for validation and escalation decisions.

 doi:10.20944/preprints202605.1325.v1

Several limitations were identified. The evaluation was conducted within a single enterprise-style environment and primarily within one vendor ecosystem. Certain social-engineering vectors involving legitimate services and password-protected content were not detected during initial delivery stages. Behavioral models also required an initial calibration period before stable operational performance was achieved. These observations indicate the continued need for governance procedures and human oversight in AI-supported security operations.

The findings can also inform the design of secure digital infrastructures for smart adaptive educational environments incorporating machine learning, IoT components, and educational robotics. Projects that combine learner profiling, cloud analytics, and robot-assisted feedback require not only pedagogical and technical mechanisms for personalization, but also robust cybersecurity monitoring and incident response procedures. From this perspective, the evaluated AI-assisted SIEM/XDR architecture provides a suitable security-oriented benchmark for future cloud-supported adaptive learning ecosystems in which connected robots and educational applications exchange sensitive operational and learner-related data.

Future work may include evaluation in heterogeneous multi-cloud environments, longer observation periods, and analysis of adversarial AI techniques and adaptive response mechanisms.

Author Contributions: All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the South-West University "Neofit Rilski", Blagoevgrad, Bulgaria.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and analyzed during the current study are not publicly available due to security and confidentiality restrictions associated with the evaluated enterprise cloud environment. Aggregated results and methodological information relevant to the study are provided in the manuscript.

Acknowledgments: This research was mainly carried out within the framework of Project KP-06-N100/7/10.12.2025 "Intelligent Adaptive Approaches to Supporting Primary Education for SEN through Machine Learning and Robotic Technologies".

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACR	Automated Closure Rate
EDR	Endpoint Detection and Response
FPR	False-Positive Rate
IaaS	Infrastructure as a Service
ICR	Incident Correlation Ratio
MTC	Mean Time to Closure
MTT	Mean Time to Triage
MTTD	Mean Time to Detect
MTTR	Mean Time to Respond
PaaS	Platform as a Service
SIEM	Security Information and Event Management
SOAR	Security Orchestration, Automation, and Response
SOC	Security Operations Centre
TPR	True-Positive Rate
UEBA	User and Entity Behavior Analytics
VPN	Virtual Private Network
XDR	Extended Detection and Response

 doi:10.20944/preprints202605.1325.v1

References

1. Gartner. Gartner Forecasts Worldwide Public Cloud End-User Spending to Total $723 Billion in 2025. Available online: https://www.gartner.com/en/newsroom/press-releases/2024-11-19-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-total-723-billion-dollars-in-2025 (accessed on 10 May 2026).
2. SentinelOne. Cloud Security Statistics 2024. Available online: https://www.sentinelone.com/cybersecurity-101/cloud-security/cloud-security-statistics/ (accessed on 10 May 2026).
3. Tariq, S.; Chhetri, M.B.; Nepal, S.; Paris, C. Alert Fatigue in Security Operations Centres: Research Challenges and Opportunities. *ACM Comput. Surv.* **2025**, *57*, 224. https://doi.org/10.1145/3723158.
4. Mgbemele, A.F. Advancing Cyber Threat Detection through SIEM-Based Automation and MITRE ATT&CK Aligned Analytics: A Systematic Review. *Asian J. Res. Comput. Sci.* **2026**, *19*, 233–254. https://doi.org/10.9734/ajrcos/2026/v19i1816.
5. Nasim, S., Pranav, P., Dutta, S. A Systematic Literature Review on Intrusion Detection Techniques in Cloud Computing. *Discov Computing* **2025,** 28, 107. https://doi.org/10.1007/s10791-025-09641-y
6. Sowmya, T.; Mary Anita, E.A. A Comprehensive Review of AI Based Intrusion Detection System. *Meas. Sens.* **2023**, *28*, 100827. https://doi.org/10.1016/j.measen.2023.100827.
7. ENISA. *ENISA Threat Landscape 2023*; European Union Agency for Cybersecurity: Heraklion, Greece, 2023. Available online: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023 (accessed on 10 May 2026).
8. Akamai. New Study Finds 84% of Security Professionals Experienced an API Security Incident. Available online: https://www.akamai.com/newsroom/press-release/new-study-finds-84-of-security-professionals-experienced-an-api-security-incident-in-the-past-year (accessed on 10 May 2026).
9. ISC². ISC² 2024 Cybersecurity Workforce Study. Available online: https://www.isc2.org/Insights/2024/10/ISC2-2024-Cybersecurity-Workforce-Study (accessed on 10 May 2026).
10. Ban, T.; Takahashi, T.; Ndichu, S.; Inoue, D. Breaking Alert Fatigue: AI-Assisted SIEM Framework for Effective Incident Response. Appl. Sci. 2023, 13, 6610. https://doi.org/10.3390/app13116610
11. González-Granadillo, G.; González-Zarzosa, S.; Diaz, R. Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures. Sensors 2021, 21, 4759. https://doi.org/10.3390/s21144759
12. CrowdStrike. AI SIEM: Next-Gen SIEM Powered by AI. Available online: https://www.crowdstrike.com/en-us/cybersecurity-101/next-gen-siem/aisiem/ (accessed on 10 May 2026).
13. Thandayutham, K. Federated Security Control Data Fabric: Scalable Telemetry Normalization and Orchestration in Multi-Cloud Environments. *Int. J. Cloud Comput. Serv.* **2025**, 11(4), 4414. https://doi.org/10.22399/ijcesen.4414
14. Palo Alto Networks. What is Extended Detection and Response (XDR)? Available online: https://www.paloaltonetworks.com/cyberpedia/what-is-extended-detection-response-XDR (accessed on 10 May 2026).
15. Gartner. Market Guide: Extended Detection and Response. Available online: https://www.gartner.com/reviews/market/extended-detection-and-response (accessed on 10 May 2026).
16. Meng, X. Advanced AI and ML Techniques in Cybersecurity: Supervised and Unsupervised Learning, Reinforcement Learning and Neural Networks in Threat Detection and Response. *Appl. Comput. Eng.* **2024**, *82*, 1–5. https://doi.org/10.54254/2755-2721/82/2024GLG0054.
17. Diogenes, Y.; DiCola, N.; Turpijn, T. *Microsoft Azure Sentinel: Planning and Implementing Microsoft's Cloud-Native SIEM Solution*; Microsoft Press, 2022.
18. Pitkar, H. Cloud Security Automation Through Symmetry: Threat Detection and Response. *Symmetry* **2025**, 17(6), 859. https://doi.org/10.3390/sym17060859.
19. Boughton, D.; Reid, I. The Role of Artificial Intelligence in SOC Operations: Adoption, Perception, and Workforce Impact. In *Proceedings of the 11th International Workshop on Socio-Technical Perspectives in Information Systems (STPIS 2025)*; CEUR Workshop Proceedings: 2025; Vol. 4134, p. 28.

 doi:10.20944/preprints202605.1325.v1

20. Winkler, A.M.; Sharma, P. Proactive Threat Detection in Enterprise Systems Using Wazuh: A MITRE ATT&CK Evaluation. *Comput. Secur.* **2025**, *159*, 104702. https://doi.org/10.1016/j.cose.2025.104702.
21. Farzaan, M.A.M.; Ghanem, M.C.; El-Hajjar, A.; Ratnayake, D.N. AI-Enabled System for Efficient and Effective Cyber Incident Detection and Response in Cloud Environments. *arXiv* **2024**. https://doi.org/10.48550/arXiv.2404.05602.
22. NIST. Framework for Improving Critical Infrastructure Cybersecurity (NIST CSF v1.1); National Institute of Standards and Technology: Gaithersburg, MD, USA, 2018. https://doi.org/10.6028/NIST.CSWP.04162018.
23. Rose, S., Borchert, O., Mitchell, S. and Connelly, S. Zero trust architecture. *NIST special publication*. **2020**, *800* (207), 1-52. https://doi.org/10.6028/NIST.SP.800-207
24. Lilhore, U.K.; Simaiya, S.; Alroobaea, R.; Baqasah, A.M.; Alsafyani, M.; Alhazmi, A.; Khan, M.M. SmartTrust: A Hybrid Deep Learning Framework for Real-Time Threat Detection in Cloud Environments Using Zero-Trust Architecture. *J. Cloud Comput.* **2025**, *14*, 35. https://doi.org/10.1186/s13677-025-00764-7.
25. Teece, D.J. Explicating Dynamic Capabilities: The Nature and Microfoundations of (Sustainable) Enterprise Performance. *Strateg. Manag.* J. **2007**, 28, 1319–1350.
26. Creswell, J.W.; Plano Clark, V.L. *Designing and Conducting Mixed Methods Research*, 3rd ed.; SAGE Publications: Thousand Oaks, CA, USA, 2017.