An AI-Enabled Framework for Securing Cloud Computing Against Internal and External Cybersecurity Threats

Swapnil Dimble[#1], Rohanshu Dhage[*2], Dr. Rupali Kalekar[#3]
[#1,2] *Student, Master of Computer Application;*
[#3] *Assistant Professor and Research Guide, Master of Computer Application, Zeal Institute of Business Administration, Computer Application and Research, Narhe, Affiliated to Savitribai Phule Pune University, Pune, Maharashtra, India*
[1]swapnildimble09@gmail.com, [2]dhagerohanshu23@gmail.com, [3]rupali.kalekar@zealeducation.com

Abstract
In the current time, the use of cloud environments has increased significantly, along with the applications, identities, data, APIs, network services, and distributed infrastructure that they support. These connections generate a large volume of security events, which can be difficult to identify on their own. External attackers capture system weaknesses through methods like phishing, malware, ransomware, network misuse and credential leaks. Meanwhile, insiders and compromised accounts pose different challenges by abusing valid access in ways that appear normal to users. This study presents an explainable, risk-based framework that leverages AI support. It integrates telemetry from various cloud sources, combining event correlation, hybrid AI, rule-based detection, user and entity behaviour analytics (UEBA), contextual risk scoring, proportional incident response, human review, and continuous feedback. In this thesis of ten research papers covering malicious insider detection, cloud intrusion detection, anomaly detection, generative approaches, explainable AI, incident response and multi-threat detection. The literature synthesis shows that multi-source behaviour evidence, sequence-aware analysis and hybrid AI-plus-rule designs are repetitive strengths whereas privacy, false alarms, adversarial manipulation, computational cost and limited real-world validation remain important barriers. The proposed framework is conceptual and experimentally not validated; also reported performance values from the reviewed studies are present only as evidence and not as results of the proposed system. The main contribution is a combined evaluation-ready blueprint that turns uniform security signals into explainable risk levels and balanced response decisions while keeping human review for consequential actions.

Keywords—Cloud security, Artificial intelligence, Insider threat detection, User and Entity Behaviour Analytics, Risk-based response, Automated incident response

I. INTRODUCTION

A. Background of the Study

In organizations, cloud services run applications, keep data, and support users across widely distributed systems. Security evidence spreads just as widely, appearing in audit logs, identity and access systems, network activity, endpoint and application records, API traffic, and data-access logs. Research indicates that artificial intelligence can examine this large body of telemetry, spot behaviour that falls outside normal patterns, and help resolve security questions that isolated events cannot answer on their own.

The present state of the threat landscape is beyond traditional external attacks. Phishing, malware, ransomware, brute force action, denial of service action, unauthorized access and other network attacks may come from within the organization. At the same time, employees, contractors, administrators and compromised accounts may have valid permissions which they use to carry out what may at first look like legitimate actions. Insiders' focus is on behavioural trends, time-based sequences and multiple sources of evidence, because harmful intent is often included in what otherwise appears to be normal activity [1], [2], [9], [10].

Rules define and govern known and policy-based behaviour, in contrast to AI methods, which offer adaptive detection, anomaly analysis and prioritization [5], [6]. The issue at hand is to take these features and put them into a single integrated security framework which covers internal and external threats, which reports why an alert is what it is, and which ties the alert to an appropriate response.

B. Problem Statement

In recent days AI security tools have improved day by day, but cloud defences still face attacks from internal as well as external sources that change quickly and make telemetry inconsistent, making it hard to check suspicious activity against legitimate user behaviour. Conventional controls struggle with unfamiliar activity, slow-moving attacks and leaked credentials. AI detectors bring their own problems: they can produce too many false alarms, demand large amounts of data and computing power, and leave analysts unsure why a decision was made. Automated containment creates another risk. A response that is too aggressive may cut off legitimate users or

interrupt critical workloads. The problem, then, is not simply labelling an event as malicious. Detection must be tied to context, clear explanations, risk ranking, and carefully controlled response measures that account for threats from both inside and outside the cloud.

C. Research Gap

The ten reported studies show large-scale progress, but they also present a set of related issues. First, internal and external threats are put into separate research categories. Studies which focus on insider issues look at user behaviour and anomaly score, and those that deal with external threats look at network, malware or multi-threat detection [1], [8]–[10]. The evidence from identity, network, endpoint, application and data access streams is not usually put together in the same design, which multi-modal studies report to improve confidence [2], [9]. Rule-based and AI-controlled measures are discussed in separate terms, although they have a practical benefit when put together [5], [6]. Explainability and human oversight are very much presented as issues or future work, although they should be design requirements. Privacy, adversarial manipulation, data quality, drift, latency and computational cost are important for deployment but are not put into the architecture from the start [3], [5]–[7]. Many systems stop at detection or alert generation, which does not include how risk should play a role in response, investigation, evidence preservation and feedback. These gaps support the case for a unified, explainable and risk-based framework which includes detection, prioritization, response and learning.

D. Aim of the Research

This research aims at developing a source-grounded, AI-enabled framework as a solution for the detection, priority ranking and response to internal and external cyber security threats in cloud computing, which we do via multi-source telemetry, hybrid AI and rule-based analysis, explainable risk scoring, proportional response and controlled feedback.

E. Research Questions

RQ1: What internal and external threats must an AI-enabled cloud security framework detect and address?

RQ2: What is reported in current research regarding AI-based threat detection and user and entity behavior analytics?

RQ3: In what ways could telemetry, AI models, UEBA, and policy rules be used to present risk assessments for people, and how should the right response be determined?

RQ4: What technical, operational, and governance protections are required to conduct a fair and responsible assessment of the framework?

F. Scope of the Study

The report looks at AI in security as it pertains to cloud computing, which includes external threats like phishing, malware, ransomware, abnormal network activity, brute force and unauthorized access, in addition to internal or compromised account issues such as privilege misuse, abnormal logins, suspicious data transfer, and configuration changes. It covers audit and access events, network activity, endpoint and application records, behaviour analysis, AI-based anomaly detection, policy rules, risk scoring, incident response and feedback. Physical security, interview results, live deployment and current operational data are beyond the scope of the present study.

G. Significance of the Study

The purpose of this framework is to show that researchers and practitioners can see security controls as a whole single process instead of separate alert systems. In terms of practice, it puts together an event with its context, explains the alert's importance, and chooses an appropriate response, which includes review material that may be required at a later date. In terms of research, the study brings together what we have learned from past studies and also puts out the gaps in that research which may be tested in the future. For organizations we present a blueprint which includes privacy, human review, traceability and model maintenance in the security design as an integral part of the system as opposed to adding them on as an afterthought.

II. LITERATURE REVIEW

A. Introduction to Literature Review

The literature covers malicious insider threat identification and user behavior analysis, AI for cloud security, intrusion detection and prevention, deep learning and generative methods for anomaly detection, and large-scale incident response architecture. The papers we looked at vary greatly in what they cover and in the type of evidence they present. Some are review articles or present conceptual frameworks, while others report on model performance from specific data sets. We treat the reported metrics more as background information than as a common measure. Each study is looked at in terms of the issue it addresses, the approach it takes, the evidence it puts forth, and what it does not do, which informs the framework presented here.

B. Insider Threat Detection and User Behaviour Analytics

Prabaharan [9] looks at the issue of insider threats in AI-powered cloud infrastructures by looking at user behavior as well as technical and context-based evidence. The framework looks at telltale signs of abnormal login times, unexpected data movement, privilege escalation, and configuration changes, but does not put forth any single one of these as proof of a threat. Instead, each event is placed in the user's normal behavior pattern, their given permissions, and related actions before a security decision is made. The risk scores that arise from this then determine the adaptive access controls and the responsive actions that are taken. In this study the approach supports risk profiles which are put together

from many sources, although the use of context data requires very defined lines around what is authorized and what is private.

Jiang, Jia, and Zhang [7] report on what they did with deep learning to detect atypical user behaviour and put out early warnings. They look at each action in relation to what came before it instead of in isolation, using autoencoders, long short-term memory (LSTM) networks and attention mechanisms to study trends over time. Under the reported parameters they achieved 96.8% accuracy and also issued alerts an average of 142 minutes in advance. That is good performance, but the larger point is that they found out which signals a single event-based detector would miss by looking at the sequence of events. They also saw that for slow forms of insider drift the results were not as strong, which in turn points out the need for more detection methods.

Alkharabsheh and colleagues [3] present an explainable solution for insider threat detection. They used Isolation Forest, principal component analysis (PCA) for reconstruction, autoencoders, and LSTM autoencoders, which enabled them to look at individual warning signs as well as trends in behaviour over time. SHAP values were used to present which features contributed to each risk score, and that in turn could trigger an action such as reduced user access. In testing, the system reported an AUC of 0.976 and an F1 score of 0.87. The success lies in combining static data with dynamic behaviour patterns, but the authors also note issues with pseudo-labeling, which in turn calls for research in more real-world settings.

Alzaabi and Mehmood [8] present a wide review of the issue of malicious insider threat and put the field in terms of rule-based, anomaly-based, signature-based and heuristic approaches. Their review is important because it notes that insider detection does not occur in a single category. Transparent rules do well with issues of broken explicit policy, but behavior and anomaly measures are better at finding what is out of the ordinary. At the core of the continuous operation of these frameworks are issues related to class imbalance, privacy, and integration.

C. *AI-Enabled Cloud Security, IDPS and Incident Response*

Olaoye [2] reviews AI-based intrusion detection and prevention for cloud security, which include machine learning, deep learning, reinforcement learning, generative models, federated learning and explainable AI. AI supports the detection of known and unknown patterns, but model drift, adversarial attacks, privacy issues and computational overhead are issues for continuous deployment. These points support a framework which has model maintenance and human intervention as integral elements instead of add-ons.

Shaffi and colleagues [10] discuss how AI can protect cloud systems by predicting threats, recognizing unusual behaviour, and responding automatically. They also consider privacy, bias, and compliance. Because the material is a draft manuscript, its claims should be read as design evidence, not conclusive proof that a complete system has been deployed. Farzaan and colleagues [6] link AI-assisted cyber-incident detection to incident response and digital forensics in cloud environments, drawing attention to data preprocessing, model testing, follow-up investigation, and the handling of evidence. Detection alone is not enough. A working cloud-security system must give responders useful context, preserve relevant records, and document every action taken.

Ademilua and Areghan [4] cover collecting and preparing data, building AI models, making decisions, and putting results back into the system. Approaches which include supervised and unsupervised learning, deep learning, natural language processing, reinforcement learning, and federated learning are put forward along with practical issues of data quality, adversarial attacks, interpretability, latency, compliance, and scalability. High detection accuracy is not the be-all and end-all; a workable framework requires defined rules, clear decision points, response protocols, ongoing monitoring, and governance.

D. *Generative Models and Multi-Threat Detection*

Sai Veena and colleagues [1] report a cyber threat detection system which makes use of many AI models for the analysis of text, images, video and QR codes, addressing issues like phishing, social engineering, deepfakes and malicious QR codes. The work does not address cloud security, which means it did not determine whether these models' results include detection of cloud account misuse. The design still has value because many forms of evidence may be put into one decision process. A cloud framework brings together signals from identity checks, network activity, endpoints, applications, and data access instead of using a detector which is built for only one information stream.

Vadisetty and colleagues [5] look at generative models such as variational autoencoders, GANs, and transformer-based models for real-time anomaly detection and risk reduction in cloud settings. They learn what is normal, identify what breaks that norm, and also produce synthetic data, which in some settings is very useful. The research also raises issues of interpretability, resistance to adversarial attacks, and computational cost. The framework in this research presents these models as options for certain tasks and not as a solution for all. The use of these models should be based on the data at hand, and they should be tested against simpler baseline models.

E. *Explainability, Hybrid Design and Risk-Based Response*

Existing research reports that these methods work well in different ways. Clear rules do best when the relevant indicators, policy violations, or compliance limits are defined in advance. Machine learning does a

better job at recognizing patterns which are hard to put into words, which includes atypical behaviour and behaviour which has meaning in the context of a sequence. To put it simply, rules address what is familiar, statistical models bring to light what is out of the ordinary, and context decides which event to pay close attention to. A person should look over any case which may greatly interfere with normal operation or production work.

Report [9] is a useful example of the connection from risk score to response tier, and Paper [2] describes how multi-modal context plays a role in an access control model score $R = \alpha S_b + \beta S_t + \gamma S_c + \delta S_i$, where S_b stands for behavioural deviation, S_t for technical-event severity, S_c for contextual risk and S_i for the relevance of threat intelligence. The coefficients are observed as design variables which will require correction in coming studies, and their values are used in the decision logic for evaluation. Across the studies looked at, explainability is a practical requirement, not a cosmetic feature. SHAP and other related methods show the contribution of features to a detection, and attention mechanisms show which elements of a temporal sequence played a role in a model's output [9], [10]. In this section we see that explainability does not apply in theory but in practice. When the system has the power to control access, bring in containment, or play a role in an investigation, explainability becomes very important. A risk score that does not present a reason which is easy to challenge, audit or correct is a problem. In the proposed framework explanation is a feature of each decision and not a separate report.

F. Deployment, Privacy, Robustness and Governance

Governance is a factor in technical reliability. Thresholds must be documented, response actions should have clear authorization, and model updates must be versioned and tested before release. The framework presents governance as a cross-cutting requirement that goes from data collection to feedback. Security telemetry should have a protected use for a given protection purpose, access should be controlled, and each automated or analyst-approved action must include in its report the cause of the event, the model or policy version which brought it about, and the final decision [5], [6], [9].

Deployment issues are reported in the body of the literature. Security data may be of poor quality, which is to say it may be incomplete, inconsistent or hard to label, particularly in the case of insider threats. AI systems put out to the public may also be a target of evasion, poisoning and distribution shift, and deep learning elements increase latency and computing cost [3], [5]–[7]. Privacy is a very basic issue, since behaviour analytics may reveal sensitive information on users and organizations. The literature reviews privacy-protective methods like federated learning and differential privacy, which should be chosen with regard to the threat model and operational requirements and not added in a random fashion.

G. Comparative Synthesis of the Supplied Studies

The reviewed studies are grouped by the part of the security lifecycle they best support. Insider-focused papers report on behavioural baselines, multimodal fusion, temporal analysis and explainability [7], [8]. Cloud and incident response papers present broader telemetry, deployment patterns, automated responses and forensic evidence handling [3], [5], [6], [8]. Generative and multi-threat studies report on flexible anomaly models and heterogeneous evidence handling [4], [7]. No single study out of the given set reports an experimentally validated framework with this wide a combined scope for internal and external cloud threats in a single closed-loop design. That is the main gap on which we base our present work.

H. Comparative Analysis

Table I sets the ten supplied studies side by side, identifying each study's main contribution and the limitation that matters for this framework. It separates the evidence reported by each study from the way that evidence informs the overall synthesis.

TABLE I. SYNTHESIS OF THE TEN SUPPLIED STUDIES AND THEIR RELEVANCE TO THE PROPOSED FRAMEWORK

Study	Main Focus	Methods	Contribution / Limitation
[8] Alzaabi & Mehmood (2024)	Insider-threat detection survey	Rule, behaviour, anomaly, signature and heuristic methods	Taxonomy of detection strategies addressing class imbalance, privacy and integration problems
[9] Prabaharan	Multimodal insider threats in AI-enabled cloud environments	UEBA, technical events, contextual risk assessment and adaptive controls	Combines risk signals from multiple sources and privacy-aware validation
[10] Shaffi et al.	AI-driven protection for cloud systems	Predictive analytics, behaviour analysis and automated response	Integrates detection and response; the supplied manuscript discusses the combined approach
[1] Sai Veena et al. (2026)	Detection of multiple cyber-threats	NLP, CNN and URL- or decision-based models	Combines evidence from different sources; not specific to cloud environments

[2] Olaoye (2025)	AI-driven cloud IDPS review	Machine learning, deep learning, reinforcement learning, generative techniques and explainable AI	Highlights adversarial threats, privacy and computational cost as persistent problems
[4] Ademilua & Areghan (2022)	AI cloud-security frameworks	AI-based frameworks for cloud security	Layered architecture combines controls and supports a workflow from data collection through feedback
[5] Vadisetty et al. (2023)	Generative cloud anomaly detection	VAEs, GANs and transformers	May adapt to changing threats, but expense and limited interpretability remain concerns
[6] Farzaan et al.	Cloud incident detection and response	AI-assisted detection, data preprocessing, investigation and digital forensics	Links detection with incident response and evidence preservation
[3] Alkharabsheh et al. (2026)	Explainable insider-threat detection and response	Isolation Forest, PCA, LSTM-AE, SHAP and risk-based containment	Produces explainable risk rankings; requires validation across wider settings
[7] Jiang et al. (2025)	Unusual behaviour detection and early warnings	Autoencoders, LSTM and attention-based sequence analysis	Supports temporal evidence and early warnings; transferability needs validation

I. Summary of the Literature Review and Research Gap

The literature points to five conclusions. Context matters: an action that looks harmless on its own may appear malicious once compared with a user's normal behaviour and earlier activity. Evidence improves when data from several sources are brought together, standardized, and correlated rather than examined as separate streams. The findings favour hybrid systems that pair artificial intelligence with rules, since neither approach, used alone, reliably handles every threat. Clear risk scores help translate model results into decisions that security teams can act on. Deployment still faces serious obstacles, including privacy concerns, false alarms, model drift, adversarial manipulation, latency, and the cost of computation. The proposed framework turns these findings into one connected workflow: it gathers and correlates data, detects problems, explains them, responds according to risk, and incorporates feedback. Its contribution lies in bringing those functions together, not in claiming that any single model or algorithm is new.

III. METHODOLOGY

A. Research Design

This study uses a conceptual, design-science research framework, supported by a structured thematic review of ten selected papers. That approach fits the research question because its main result is a proposed cloud security framework, not a statistical estimate based on newly collected data. The study does not include a live deployment or present a newly trained model.

B. Research Approach

The study used a qualitative, analytical approach. Each paper was read closely to determine which threat it examined, how the authors detected or addressed it, what data or setting they used, and what findings and limitations they reported. The papers were then grouped around recurring themes: insider activity and anomalous behaviour; cloud threats, intrusion detection and prevention systems (IDPS); correlating events from multiple sources; explainability and risk scoring; automated investigation and response; and practical concerns such as privacy, robustness, scalability, and model maintenance. These themes together shaped the framework.

C. Study Population and Sample

The study draws on academic and technical writing about AI-driven cloud security, with ten research papers making up the sample. Each paper addresses at least one part of the research problem. Because these papers were supplied for the project rather than identified through a formal database search, the work is best described as a bounded narrative synthesis, not a systematic review of the field.

D. Data Collection

The data used are secondary. They include results, methods, tables, architectures, reported metrics and issues contained in the papers. For this manuscript we did not collect any new cloud data set, user study, interview or operational security telemetry.

E. Research Instruments / Tools

The study used manual extraction and coding, comparative tables and an architectural diagram. The literature shows that Python-oriented machine learning workflows, containerization, log analysis stacks like ELK, stream processing and SHAP are good options for future prototype development [3], [9], [10]. These options are presented as what may work for the

implementation of the framework, not as what was used for its validation.

F. Variables

In the design of future experiments the independent variable is to be the detection configuration, which may be a rule-only, AI-only or hybrid approach, in addition to the number and type of data input modalities. As for dependent variables, we will look at precision, recall, F1 score, false positive and false negative rates, detection and response times, computational cost, and explanation stability. We will control and report the dataset, time split, label quality and decision thresholds.

G. Research Procedure

1) Define the issue of cloud security, separating threats from within the organization and those from outside.

2) Go over the ten papers and group them by methods used, evidence, presentation and how they implement limitations.

3) Put the results into themes, then identify what research is still missing.

4) Report the study's goals, the research questions it sets out to answer and the parameters of the research.

5) From what is seen to be common across the literature, build a six-stage framework.

6) Develop a risk score, which in turn is put into a proper response category.

7) Put out what measures and governance standards are needed to validate the approach going forward.

8) Pull the results together, state what they may mean in practice, and report what is not addressed by this study, without making claims of performance that are not there.

H. Proposed Framework Design

Proposed is a six-stage, closed-loop system. The stages are independent of each other, which allows detection models, data sources, thresholds and response actions to be put in or out as we go along. We have tried to include what the literature reports to be strong points in this framework, but have also included elements which we identify as still requiring empirical support.

Stage 1 – Telemetry Collection. Authorized security events are collected from cloud audit logs, identity and access management, network traffic or flow logs, endpoints and hosts, applications and APIs, data access logs and present security alerts. We will document the owner, time stamp, retention policy and use of each of these sources in the future. We should note down cases of missing visibility, as incomplete data may in fact invalidate our results.

Stage 2 – Preparation and Event Correlation. Events are put into a timeline which includes the owner of the event and the event type; repeats and reporting mistakes are looked for and corrected, and security features are created from login frequency, failed logins, access time, data transfer volume, change in privileges, role context and event sequences. We have put in place correlation which identifies related actions across our data sets, which in turn allows us to see whether what appears to be an out-of-the-ordinary event is in fact part of a larger, possibly normal, sequence of action.

Stage 3 – Hybrid AI Detection and UEBA. At this stage different detectors are used. Supervised learning is used to label known attack patterns, unsupervised learning identifies what is out of the ordinary when few labels are available, and sequence models are used when looking at time-based behaviour. UEBA compares present user or entity action against that which is normal for that individual, and rule-based detection puts in place explicit security policies. Generative models may be introduced for anomaly modelling or synthetic data, which adds value at the expense of cost and complexity [4], [5], [7].

Stage 4 – Explainable and Contextual Risk Scoring. Individual models' results are put together with operational context to present a total risk score, as given by

$$R = \alpha S_b + \beta S_t + \gamma S_c + \delta S_i \quad (1)$$

Here S_b is used for behavioural deviation and S_t for technical-event severity. S_c includes policy violations and sensitive resources which fall under the category of context, and S_i is the most relevant supporting threat intelligence. The coefficients alpha, beta, gamma, and delta are at present unvalidated design choices. In the next prototypes these will have to be tuned against labelled or expert-reviewed data. Each risk decision should include a note which identifies which pieces of evidence had the greatest impact on the score.

Stage 5 – Risk-Tiered Response and Human Oversight. Risk determines the scale of response, while operational risk should set the level of automation. For low-risk incidents we may log and monitor; for medium risk we may see stronger authentication or alerting take place; high risk may see the system restrict what it deems suspicious activity or quarantine a session; and in the case of critical events we may see pre-authorized containment and evidence preservation. We should, wherever possible, include human review in actions that may impact legitimate access or production performance [3], [5], [8], [9].

Stage 6 – Feedback, Governance and Model Maintenance. Past studies, analyst estimates, false positive reports, missed attacks, model drift, reducing data quality, and resource issues should inform feature reviews, threshold adjustments, rule changes and very controlled retraining. Each model or policy change will have a version, pre-deployment testing, and an auditable

trace. Privacy constraints, role-based access, retention rules and review processes must be present in all six stages, not just at data collection.

I. Data Analysis and Evaluation Measures

The present study does not report newly developed detection metrics; they will be used in the future for evaluation of the framework. For fair assessment going forward we should see the analysis of rule-based, AI-based and hybrid models. Internal and external scenarios should be separated at the time of testing before they are combined. The datasets used and how they were split should be documented, and we also support the idea of assessing label quality. Both operational as well as statistical measures should be presented.

J. Ethical Considerations

Behaviour tracking brings up issues of privacy and trust. Security tracking should be seen as applying to what is put in place for protection, and its purpose must be defined. Access to raw logs and behaviour profiles should be kept to a minimum, and what we do with that data should be in line with the organization's policy. The system should keep a record of the event which triggered the issue, which model or policy was in play at the time and what the result of that was. Human review and a way to correct false alarms are very important when we are using automation which may in turn restrict access. In the present study we did not recruit human participants and we did not collect primary personal data.

IV. RESULTS

A. Description of the Study Material

This section reports findings from a bounded narrative synthesis of the ten research papers provided. The analysis centers on the themes that shape the proposed framework, including detection of external cyberthreats and insider threats, analysis of user behavior, explainability, risk assessment, automated response, and cloud security governance. Across the studies, different methods are used to identify threats and investigate suspicious activity; their results and limitations provide the basis for this paper's conceptual design. Drawing on those findings, the proposed framework brings together security-event collection, threat detection, contextual risk assessment, explainable decisions, risk-based response, and continuing governance in one workflow. Table II maps the AI techniques to their roles in the framework, Table III and Fig. 1 present the risk tiers and corresponding response actions, and Table IV lists the proposed evaluation measures.

TABLE II. AI TECHNIQUES AND THEIR ROLES IN THE PROPOSED FRAMEWORK

Technique	**Role in the Framework**	**Primary Evidence Base**
Supervised ML (Random Forest, SVM)	Recognises known attacks, malware and intrusion classifications from a large set of labelled examples.	[3], [5], [6]
Unsupervised anomaly detection (Isolation Forest, autoencoder)	Identifies unusual behaviour, including previously unseen activity and data with weak labels.	[3], [9]
Sequence models (Long Short-Term Memory, Attention Mechanisms)	Tracks behaviour over time to detect trends and issue early warnings.	[9], [10]
User and Entity Behaviour Analytics (UEBA) baselines	Compares the present activity of a user or entity with its historical behaviour.	[1], [2]
Generative models (Variational Autoencoders, Generative Adversarial Networks, Transformers)	Supports optional anomaly modelling and training-data generation.	[5], [7]
Rule-based controls	Implements explicit policies and compliance requirements, including known indicators.	[5], [6]

TABLE III. RISK TIERS AND CORRESPONDING RESPONSE ACTIONS

Tier	**Conceptual Range**	**Primary Response**	**Human Review**
T0 – Observe	$R < 0.40$	Logging, baseline update, periodic audit	Routine
T1 – Caution	$0.40 \leq R < 0.65$	Alert, step-up MFA, throttling or DLP warning	SOC review
T2 – Containment	$0.65 \leq R < 0.85$	Privilege downgrade, token revocation, session quarantine	Real-time validation
T3 – High Risk	$R \geq 0.85$	Pre-authorized containment, segmentation, forensic capture	Emergency escalation

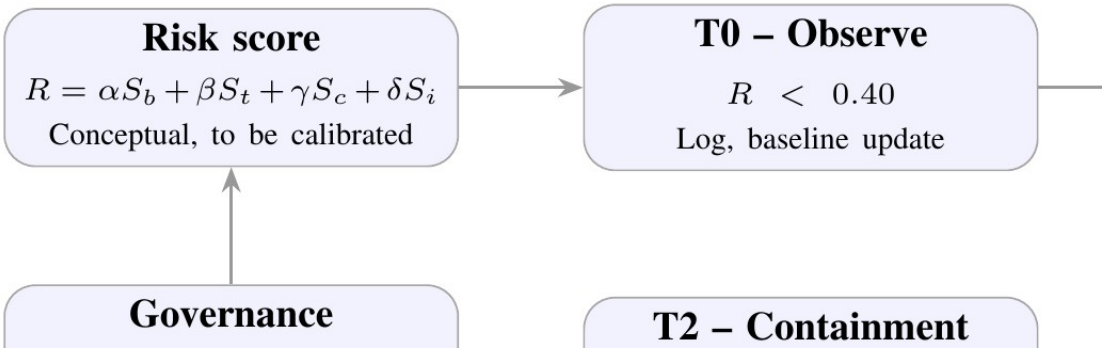

Fig1.Risk-based response logic with escalating response tiers and governance feedback.

TABLE IV. PROPOSED EVALUATION MEASURES FOR FUTURE EMPIRICAL VALIDATION

Measure	Purpose
Accuracy	Overall classification performance, which in itself is not enough for imbalanced security data.
Precision	Proportion of flagged events that are confirmed threats; relates to analyst alert burden.
Recall	Share of relevant threats detected; captures missed threats.
F1-score	Combined view of precision and recall.
False-positive / false-negative rate	Tracks analyst burden and missed attacks.
MTTD / MTTR	Measures detection and response speed.
Early-warning lead time	How far in advance a warning is given.
Resource cost and latency	Measures computation, memory and event-processing demands.
Explanation and score stability	Evaluates whether reasons and rankings are presentable and remain consistent across iterations.

B. Main Findings from the Literature Synthesis

1. Multi-Source Security Information. The reviewed literature considers different sources of security information, including user actions, technical events, network traffic, and cloud security logs. Combining information from multiple sources may provide a more complete picture for security assessment than examining isolated events. Based on this synthesis, the proposed framework includes event normalization and correlation before risk assessment. This is a design recommendation derived from the reviewed literature, not an experimentally proven improvement [2], [6], [8], [9].

2. Complementary Detection Approaches. The reviewed studies use several methods to detect threats, ranging from fixed rules and supervised classifiers to anomaly detection and sequence analysis. These methods address different detection needs and cannot simply substitute for one another. Taken together, they could form complementary components of the proposed framework. This possibility still requires practical testing and comparative evaluation [1], [5], [6], [10].

3. Temporal Analysis and Explainability. Security events should be examined over time, and detection decisions should be understandable to human analysts. Sequence-based approaches can reveal how events develop and in what order they occur. Explainability methods such as SHAP can indicate which inputs contributed to a model's output. These findings inform the proposed temporal-analysis and explainability components, although their performance within the complete framework remains to be evaluated empirically [9], [10].

4. Integration of Detection, Response, and Investigation. The reviewed literature highlights the importance of connecting threat detection with incident response and security investigation. An alert is more useful when it provides enough context for analysts to assess its priority and select an appropriate action. Relevant evidence must also be preserved for later examination. Accordingly, the proposed framework connects detection results with risk-based response decisions and evidence preservation. The present study does not experimentally evaluate this combined process [3], [5], [6], [8].

5. Governance, Privacy, and Operational Constraints. The reviewed studies identify several challenges in deploying AI-based security solutions, including privacy concerns, false positives, model drift, adversarial manipulation, processing delays, and

computational costs. Their treatment of these issues varies. These findings inform the proposed framework's governance measures, including controlled access to security data, model-performance monitoring, documented response protocols, and evaluation of model updates. These measures are design recommendations whose practical value remains to be established through future evaluation [3], [5]–[7], [9].

C. Objective-Wise Results

Table V summarizes selected performance evidence reported by the supplied literature, and Table VI aligns the research objectives with the outputs of the literature synthesis.

TABLE V. SELECTED PERFORMANCE EVIDENCE REPORTED BY THE SUPPLIED LITERATURE

Study	Reported Metric / Result	Interpretation for This Study
[1]	Multimodal insider detection achieved 0.96 precision and 0.97 recall, with a lower false-positive rate.	These findings support multi-source behavioural fusion, although each study was validated under its own specific conditions.
[3]	On the CIC-IDS 2017 dataset, Random Forest reached 99.82% accuracy. Performance was lower for some attack classes.	Performance depends on two factors: the dataset and the kind of threat being detected.
[9]	A hybrid, explainable insider-threat framework recorded an AUC of 0.976 and an F1-score of 0.87.	The results favour combining static and sequential analysis with explanations for detected anomalies.
[10]	In the related evaluation, the system reached 96.8% accuracy and issued warnings approximately 142 minutes in advance.	That result highlights the value of tracking behaviour over time and flagging threats before they fully emerge.
[5], [6]	Case studies and review articles report fewer false positives and faster response times.	These reports offer useful guidance for system design, but they do not independently validate results produced by the framework examined here.

TABLE VI. ALIGNMENT BETWEEN RESEARCH OBJECTIVES AND LITERATURE-SYNTHESIS OUTPUTS

Research Objective	Finding from the Synthesis	Output in This Study
Classify internal and external threats	The identified threats include misuse by insiders, compromised accounts, phishing, malware, ransomware, unusual network behaviour, and unauthorized access.	Together, these categories cover a broad range of possible threats.
Analyze AI detection techniques	The source set covers supervised and unsupervised methods, along with sequence analysis, UEBA, generative approaches, and rule-based systems.	These techniques are mapped across the relevant tasks.
Design an integrated framework	The literature points to a system that draws from multiple data sources, compares those signals, combines detection methods, explains its findings, triggers a response, and learns from feedback.	These elements form a proposed six-stage framework.
Develop risk-based response	The response should reflect both the assessed risk and the likely operational impact. Significant actions require human review. The framework therefore uses a conceptual score to assign cases to one of four response tiers.	The framework uses a conceptual score to assign cases to one of four response tiers.
Define evaluation and governance	Detection and response should be assessed alongside cost, explanation stability, privacy, and drift, not as separate concerns.	Future evaluation plan.

V. DISCUSSION

A. Interpretation of the Findings

The main point is that cloud threat management is in fact a process and not a single issue. A detector may note an anomaly, which is great, but that detection's value in the real world is in how related events are tied together, how the alert is put into play, what the reason is behind it and what the response is to the possible damage. From the insider perspective, the importance of user baseline performance and time-based behaviour is clear, and larger-scale cloud security studies point out that what is important is deployment, automation and incident response [2], [5], [6], [9], [10]. The presented model makes these points clear in one full circle.

B. Comparison with Previous Research

The framework presents a collection-to-feedback structure as per Ademilua and Areghan's work [4] and a detection-to-response slant, which is a theme in cloud IDPS and incident response research [5], [8]. It presents a behaviour-centered view of what is used in the insider threat literature [1], [2], [9], which we add to by including sequence-aware evidence from Jiang et al. [7], and we put forward generative approaches as an optional element as opposed to a required element [7]. We

differ from studies which look at one threat or model at a time; here we put all of it together in one workflow. This is to present an architectural synthesis and not a report of experimental superior performance.

C. Explanation of Unexpected or Inconsistent Findings

The study finds that results of what we may term "headline accuracy" are in fact a product of their context. Results play out differently in different data sets, with different threat models and label sets. In some insider threat studies we may be working with pseudo-labels or detection at an inferential level; in others which look at network attacks we may have clearly labelled simulated traffic. Some of the papers we looked at did not report in full the methods used or the publication details. These differences do not allow for a single performance metric to rank the studies. This paper therefore does not present high reported metrics as proof that the same results will translate to a different cloud environment; instead, what is reported is an indicator of what may be achieved.

D. Practical Implications

In the future we should see the framework implemented with a small and traceable set of telemetry sources and very clear threat models. Transparent rules should go after what we know for sure first, and then we can introduce behavioural detectors as we build up history. More complex models should follow only if they prove out via data that they add value. We should make risk scores which present the information that led to a decision, and we must also test automated responses for security improvement and operational impact. We should report on performance which includes false positive rates, response time and resource cost in addition to accuracy.

E. Theoretical and Design Implications

The research presents a design-level perspective in which detection, explanation, response and learning are elements of one security process. We put AI in a governed decision framework instead of presenting an AI model as the security system itself. We use contextual evidence to put forward alerts, explanations to support review, response stages to contain automation, and feedback for controlled improvement. This structure is presented as a test bed for the design on which future experimental work may be based.

VI. LIMITATIONS OF THE STUDY

This study reports a number of issues. We looked at 10 specific papers, which did not result from a systematic review of all relevant works. The studies which we looked at vary in what they look at, in the data they use, in what they define as a threat and in their evaluation methods, which makes it hard to compare results. The framework is still at the concept stage and we have not yet put it to the test in a real-world or large-scale setting. The material we looked at did not always provide enough of the same information to choose between different models, risk thresholds put out by the models, or response actions. Some of the published work we looked at did not report full publication information. Finally, in the area of behaviour monitoring and automatic containment, we raise some privacy and operational issues which machine learning does not solve.

VII. CONCLUSION

This study looked at how to put artificial intelligence into a structured framework for the protection of cloud systems against insider jobs, outside attacks, and issues related to compromised accounts. We looked at 10 provided research papers, which brought out that there is a need for multi-source telemetry, behaviour-based detection, a mix of AI and rule-based controls, explainable risk evaluation, a proportional response, human-in-the-loop review and continuous model care. We did not report that we have empirically proven the framework, but rather that we present a very clear, traceable design which can be put into practice and tested in a later prototype. For a strong cloud defense to work it should not rely on a single model or alert system. It should use a variety of signals, make its decisions easy to understand, put in place governance around automation, and improve from what is proven to work.

VIII. FUTURE SCOPE

1) Implement a proof of concept which uses the data sources and technical patterns which we identified from the literature.

2) Compare rule-based only, AI-only and hybrid detection in internal and external threat settings.

3) Study the value added by behavioural and context features via ablation studies.

4) Calibrate the risk model and response thresholds with expert-reviewed or properly labelled data.

5) Look at transformer-based sequence models, continuous learning and better performance with slow insider drift.

6) Look into privacy-preserving approaches like federated learning and differential privacy where it is operationally sound.

7) Test against adversarial evasion, poisoning and data quality issues.

8) Evaluate analyst workload, legitimate user disruption, explanation stability and evidence quality during pilot-scale deployment.

ACKNOWLEDGMENT

The authors wish to express their deep appreciation to Dr. Rupali Kalekar, their research guide, for the guidance given throughout the study. We would also like to extend our thanks to Dr. Dinesh Banswal for his great support, and we are also grateful to the Zeal Institute of Business Administration, Computer Application and Research (ZIBACAR) in Pune for providing the academic structure and resources which made this project a success.

FUNDING / CONFLICT OF INTEREST

This research received no external funding. The authors declare no conflict of interest.

REFERENCES

[1] K. S. Veena, V. Neelima, N. Nagaraju, P. Meghana, B. Sowmya, and G. Charyaka, "AI-based multi cyber threat detection," International Scientific Journal of Engineering and Management (ISJEM), vol. 5, no. 1, pp. 1–6, 2026.

[2] G. Olaoye, "AI-driven intrusion detection and prevention systems (IDPS) for cloud security," 2025.

[3] A. R. Alkharabsheh, G. Binsalma, M. Alhamni, R. Alshateri, S. Altaee, and M. Sweidan, "Artificial intelligence-based insider-threat detection: A hybrid explainable framework with automated response and privilege containment," Computers, vol. 15, no. 7, 2026.

[4] A. A. Ademilua and E. Areghan, "AI-driven cloud security frameworks: Techniques, challenges, and lessons from case studies," Communication in Physical Sciences, vol. 8, no. 4, pp. 684–696, 2022.

[5] R. Vadisetty, A. Polamarasetti, S. K. Rongali, S. K. Prajapati, and J. Bhanubhai, "AI-driven threat detection: Enhancing cloud security with generative models for real-time anomaly detection and risk mitigation," Journal of Computational Analysis and Applications, vol. 31, no. 3, pp. 532–543, 2023.

[6] M. A. M. Farzaan, M. C. Ghanem, A. El-Hajjar, and D. N. Ratnayake, "AI-enabled system for efficient and effective cyber incident detection and response in cloud environments," 2025.

[7] X. Jiang, R. Jia, and Z. Zhang, "Deep learning-based user behavior anomaly detection and threat early warning in cloud computing environments," Academia Nexus Journal, vol. 4, no. 3, pp. 1–24, 2025.

[8] F. R. Alzaabi and A. Mehmood, "A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods," IEEE Access, pp. 30907–30927, 2024.

[9] M. A. Prabaharan, "An advanced detection model for insider threats within AI-enhanced cloud computing infrastructures."

[10] S. M. Shaffi, S. Vengathathil, J. N. Sidhick, and R. Vijayan, "AI-driven security in cloud computing: Enhancing threat detection, automated response, and cyber resilience," Cyber Security and Applications.