International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 12-No.2 (2026) pp. 546-557
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Unified AI Security Posture Management Framework for Multi-Cloud Large Language Model Deployments

Vaishali Mahavratayajula*

FNR Solutions, USA
* **Corresponding Author Email:** mahavratayajulavaishali@gmail.co - **ORCID:** 0000-0002-5247-7855

Article Info:

DOI: 10.22399/ijcesen.5279
Received : 25 March 2026
Revised : 25 May 2026
Accepted : 26 May 2026

Keywords

Large Language Models,
Multi-Cloud Security,
AI Security Posture Management,
Zero-Trust Architecture,
Cloud Security Compliance,
LLM Threat Detection

Abstract:

The rapid proliferation of large language models across multi-cloud environments has introduced new challenges for enterprise security posture management. Existing cloud security posture management tools operate individually, leading to blind spots and delayed threat response for distributed AI workloads. This survey presents emerging Unified AI Security Posture Management (UAI-SPM) frameworks for hosting LLM applications in multi-cloud environments. This article aggregates recent developments in industry, regulatory frameworks, and technological trends, while also addressing needs in security tooling, essential framework components, and novel architectural patterns. The article describes the tradeoffs between autonomy and security in AI systems, provides a literature review for applying zero trust to generative AI systems, and makes evidence-based recommendations for enterprises deploying LLMs at scale. Key findings indicate that attacks against AI services are pervasive and that the majority of risk exposure stems from identity and access misconfigurations. In this context, integrated posture management frameworks that enable continuous discovery, behavioral insights, and adaptive compliance will be foundational to security resiliency in an AI-driven world. Organizations will also have to deal with accelerating regulatory scrutiny around continuous monitoring, transparent governance, and verifiable security controls throughout the AI system lifecycle.

1. Introduction

1.1 Background and Operational Context

Rapid adoption of artificial intelligence, and in particular LLMs, has changed how enterprises operate in the finance, healthcare, and government sectors. Enterprises are deploying their LLM workloads across heterogeneous clouds offered by Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) to achieve optimal performance and resiliency for their LLM workloads. A distributed deployment model also requires a variety of security controls, compliance certifications, and operations among the cloud computing services it includes.

The security posture of an organization, which can be defined as the collective cybersecurity state of an organization (policies, configurations, monitoring, and incident response), is harder to enforce in multi-cloud setups. Including artificial intelligence and machine learning de-stresses manual risk assessments in Cloud Security Posture Management in favor of automated compliance checking and threat detection [1]. Pre-existing security approaches, such as siloed cloud security posture management (CSPM) and data security posture management (DSPM) tools from multiple vendors, each integrated with different cloud platforms, provide insufficient protection against dynamic threats that constantly evolve in the cloud environment. This can lead to blind spots, configuration drift, and slow threat detection.

Other vulnerabilities include misconfigured permissions and access policies, lack of visibility into data pipelines, inconsistent policy enforcement, and lack of coordinated threat response across the cloud. AI-enabled solutions for verifying compliance can help avoid inconsistent enforcement of compliance policies and security settings by continuously analyzing the infrastructure against externally mandated regulations and internal organizational policies and objectives [1]. ML algorithms can also be leveraged

to establish a baseline of normal behavior and detect deviations from this baseline, which may indicate a potential security incident or policy violation [2].

1.2 Problem Definition and Critical Challenges

Tools such as CSPM and general AI governance frameworks partially automate cloud security operations, but due to architectural limitations, they do not offer a holistic solution that combines AI monitoring, compliance testing, and remediation of LLM workloads. This fragmentation is compounded by heterogeneous multi-cloud infrastructures.Security vulnerabilities can therefore propagate across cloud boundaries, as a single misconfigured node or compromised credential could be leveraged to ease attacks across multiple connected systems and shared data repositories. Adaptive threat detection techniques must therefore evolve to take into account advanced attack vectors leveraging the heterogeneity of cloud services and inter-cloud communication protocols [2]. However, LLM deployments have unique attack surfaces such as direct model access, prompt injection, training data exfiltration, and leaked model weights, which existing tooling does not fully address. Additionally, AI services in production are retrained and redeployed iteratively, drastically shortening the time from development to deployment from weeks to hours, making classification rules and static rules insufficient.

Heterogeneity manifests itself in different ways in multi-cloud infrastructures on the cloud provider side of things. This includes, but is not limited to, differences in identity management, network security, encryption & logging. Adaptive threat detection on multi-cloud must overcome such heterogeneities to provide unified visibility and coordination with a distributed approach [2]. Further, AI-driven threat detection tools are typically black boxes that provide security alerts but do not attempt to interpret and explain the rationale or provide evidence for a behavior being flagged. This prevents security teams from trusting these alerts or auditing them for compliance. Together, these limitations prevent security operations from implementing advanced, cohesive mechanisms to observe, predict, and reduce propagation of security posture degradation across the LLM landscape.

1.3 Research Motivation and Strategic Significance

Enterprise customers have invested heavily in deploying multi-cloud LLMs for mission-critical use cases such as customer service automation, clinical decision support, and threat intelligence analysis. Compromise of these LLMs can lead to data breaches leaking sensitive data or model inversion attacks, regulatory fines, and loss of trust causing irreparable damage to the enterprise organization's reputation. AI-enabled automation can help organizations to deal with the increasing complexity of compliance requirements across jurisdictions and sectors [1].

Research has found that attacks on AI services have become pervasive in enterprise environments. Application programming interface vulnerabilities and identity-access misconfigurations are considered the top attack vectors. Attackers have also increasingly relied on adversarial machine learning, model poisoning, and inference-time attacks to alter the integrity and confidentiality of AI systems [2]. The speed and sophistication of these attacks demand that detection and response capabilities be automated beyond just human analyst response.

Recent regulations, such as the forthcoming European Union Artificial Intelligence Act, the National Institute of Standards and Technology AI Risk Management Framework, and the International Organization for Standardization standards, mandate that users monitor and transparently govern their AI systems and have effective security controls across the AI system lifecycle. This requires unified security architectures that deliver real-time visibility, predictive risk, and automated policy enforcement, especially in heterogeneous cloud environments. The increasingly perilous threat landscape and ever-tightening regulation create a compelling proposition for Unified AI Security Posture Management (USPM) frameworks that build analytics, continuous compliance, and automated defensive controls into the design, operationalization, and operation of LLMs. The full promise of USPM architectures is that they provide a means by which to realize zero trust and continuous assurance principles with LLM workloads, thereby ensuring these workloads are secure and compliant.

2. Research Gaps and Current Challenges

2.1 Visibility and Integration Deficiencies

Existing security tooling is architecturally limited to multi-cloud LLMs. Current observability and monitoring products also don't have broad interoperability, correlation, and visibility across multiple cloud environments to build a thorough security posture around data, activity,

configurations, and threats in an organization's infrastructure spanning multiple production environments and cloud providers. This fracturing, however, creates blind spots where advanced adversaries exploit seams between cloud boundaries, hiding behind authorized, cross-cloud integrations to undertake malicious actions and evade detection from conventional, single-platform analytics technologies.

The distributed environments lend themselves to AI-enabled compliance automation strategies that rely upon continuous data collection, validation, and reporting across heterogeneous systems [3]. Customary compliance models that revolve around periodic manual auditing and static rule enforcement are ill-suited to cope with the highly dynamic nature of cloud configurations and regulatory interpretations. Organizations operating in regulated industries may find it difficult to meet compliance obligations due to differing reporting formats, timelines for submission, and data retention in various jurisdictions.

Most AI security tools work by looking for anomalies and deviations in configurations, without including business context, model sensitivity classification, or operational criticality evaluation in their threat detection mechanism. For example, a machine learning anomaly might be an innocuous retraining of the machine learning model, or it could be a massive exfiltration of data, and existing tools cannot tell the difference. Security analytics business intelligence allows them to use risk-prioritized alerts to focus security teams' attention on the highest risks to high-value assets or business processes [3].

2.2 Adaptive Compliance and Remediation Limitations

Existing cloud security compliance frameworks are based on manual triggers to be updated to the latest regulatory requirements or changes to cloud service offerings. This creates time lapses during which organizations have to operate their cloud deployments under pre-existing security frameworks that do not reflect the latest cloud offerings, legal and regulatory requirements, or the threat vectors that target LLM deployments in particular. Automated remediation mechanisms can only respond to incidents based on established response playbooks. The inability to learn from past incidents or changing environmental conditions means the same attack can be repeated, and security teams will continually face the same threat without their defenses adapting.

Heterogeneity-based threats exploit the diversity of cloud provider platforms (e.g., different default configurations, non-standardized security controls, and different API authorization requirements) [4]. An example of such an attack is when an attacker uses one exploit on different platforms and the exploit behaves differently on different platforms (when it passes provider-specific intrusion detection systems). Configuration management frameworks that span heterogeneous cloud environments often provide translation layers that map common security policies to provider-specific implementation details while ensuring semantic consistency.

Security management becomes increasingly difficult as the number of cloud service providers, cloud services, and interconnections between the service environments grow [4]. Organizations cannot maintain their inventory of cloud resources, track the configuration drift across their distributed and multi-cloud environments, or guarantee the consistent implementation of security controls across their cloud environments. Automated validation of compliance must consider semantic equivalence, the preservation of policy intent between provider implementations of similar or equivalent security controls across platforms.

2.3 Predictive Capabilities and Explainability Deficits

Only a few predictive security posture simulation studies have been conducted on distributed AI systems. None of this work deals with attack path prediction, threat modeling, or proactive vulnerability discovery in multi-cloud LLM architectures. Most organizations are committed to reactive postures, lacking both knowledge of possible exploitation paths and the means to evaluate prior-to-exploitation postures. Security operations may use predictive analytics to inform threat hunting and preemptively employ security controls to interdict likely attack paths.

Explainability challenges in AI-based security tools arise when detection algorithms issue risk alerts without making clear how risks were determined or creating an auditable paper trail for security analysts to validate automated results, prioritize actions, or justify security investments. Many legal frameworks require transparency into automatic systems making decisions about security controls affecting data access, service availability, or compliance state [4]. Thus, organizations using AI security solutions must consider the trade-off between performance and interpretability, with the potential to sacrifice model performance in order to improve interpretability.

Industry research shows a disparity between security practice adoption and LLM deployment

speed. While most organizations have deployed LLMs, only a small fraction deploy security controls for all their enterprise AI workloads. This highlights common patterns of prioritizing rapid LLM deployment over thorough security, which often results in technical debt and risk across enterprise AI portfolios. The gap between AI adoption and security tooling maturity requires standardized sprints to both protect existing applications and implement patterns of secure-by-design for future deployments. Table 1 synthesizes the primary deficiencies in existing security tooling and frameworks for multi-cloud Large Language Model deployments, highlighting visibility constraints, compliance rigidities, and predictive capability limitations that necessitate unified posture management approaches.

3. Core Components of Unified AI Security Posture Management Framework

3.1 Continuous Discovery and Asset Inventory

Unified AI Security Posture Management (UAI-SPM) platforms constantly search for AI assets like trained AI models, inference runtimes, model development toolchains, data streaming pipelines, and service identities across multiple clouds. CSPM with integrated artificial intelligence and machine learning tooling improves asset discovery by identifying cloud assets using behavior-based pattern recognition instead of information from resource metadata or naming conventions [5]. These solutions leverage advanced agents that traverse cloud environments to build inventories of LLM instances, compute resources, data storage locations, and network connections.

The asset classification taxonomies of the USPM framework classify discovered assets according to models, sensitivities, required controls, and business criticalities. Machine learning can classify assets automatically based on usage, data flow, and access patterns, as well as recommend settings for the intensity of controls on assets [5]. Classification metadata allows for risk-based prioritization of security controls. Systems with high-value models used for processing sensitive data or for delivering essential business functions require increased protection and scrutiny. Inventory reconciliation processes are another important component to detecting unauthorized model deployment, shadow AI deployments, and abandoned compute resources that are potential attack surfaces within an organization's infrastructure.

One solution to some of the challenges of scale and maintenance arises from automation of the discovery process. In such multi-cloud deployments, where various services are continually being provisioned and deprovisioned, AI-driven discovery processes serve to maintain an up-to-date asset inventory and avoid the staleness problem that can be encountered with manual documentation [5].

3.2 Risk Assessment and Misconfiguration Detection

Dedicated AI misconfiguration detection tools extend CSPM capabilities to tackle misconfiguration risks in LLM systems. Typical misconfigurations identified by these tools include overly permissive IAM policies granting excessive model access, unprotected public inference endpoints, unverified configuration systems potentially exposing data to other systems, and the use of weak encryption to protect model weights or training datasets. Other machine learning-based anomaly detectors inspect the behavior of configuration elements for deviations from baseline behavior, flagging changes that pose security risks without necessarily violating policy rules [5].

Mature USPMs use analysis of AI telemetry data, behavioral profiling of models, and guardrail validation to identify weaknesses proactively and prevent risks from being exploited. By combining different information sources such as configuration compliance state, anomalous model behaviors, threat intelligence signals, and historical incident trends, composite risk scores can be computed for individual and aggregated AI portfolios. Explainability for AI-generated security decisions is a key requirement for enterprise adoption, as security teams need to understand how risks are being evaluated in order to validate and prioritize remediation efforts [6].Explainable AI methods such as feature importance analysis, decision tree visualization, and counterfactual explanations can enable security analysts to both query risk score predictions and understand the driving factors of security scores, enabling knowledge transfer and allowing human analysts to develop their intuition of risk patterns to complement an automated detection system [6]. Automated remediation workflows can quickly remediate a finding by fixing misconfigurations, revoking access rights, or quarantining workloads. Because of the need for explainability, these workflows must also build an auditable logic chain to show how the system arrived at the remediation decision.

3.3 Identity and Access Governance

Cloud service fragmentation, identity fragmentation, and LLM interaction fragmentation

are considered promising attack surfaces, as cloud security breaches are mainly tied to identity manifestations, not malware execution. Using identity attacks against network perimeter defenses can result in credential leakage, unauthorized model actions, and privilege escalation. Model Context Protocol, and similar systems for interacting with LLMs, add further identity management complexity, including service-to-service authentication, programming interface key management, and dynamic permission scoping.

With identity governance multi-cloud LLMs can centralize authentication, authorization and auditing across deployments, using enterprise identity providers to enforce policies consistently across architectures, enabling single sign-on and central credential lifecycle management [5]. Just-in-time access provisioning further reduces the exposure time of credentials. Continuous authentication monitoring detects anomalous patterns of access and usage, indicating credential theft or insider misuse. Behavioral analytics correlates identity activities with the patterns of model usage. These analytics could detect anomalies, such as admin credentials reaching customer-facing inference endpoints or dev accounts reaching production model weights.

Identity-based security decisions require explainability due to the potential for false positives to obstruct legitimate user activities or scheduled processes [6]. Security teams require the ability to quickly validate and re-enable legitimate identity activity blocked by an automated solution. Informed by why access was revoked or identity behaviors are flagged as risky, security teams can quickly undo their actions without risking bad actors circumventing the security system. Explainable access control systems provide an explanation of their authorization actions by disclosing the policies, risk factors, or behavioral anomalies present in their decisions.

3.4 Threat Detection and Behavioral Analytics

These frameworks may include anomaly detection algorithms, real-time telemetry analysis, or machine-learning-based behavioral analysis to detect anomalous behavior or unusual patterns in the use or inference of the models through prompt injection attempts, high or low query rates, abnormal patterns in model output, and unusual patterns in the access of model weights that might indicate data exfiltration or model theft. Machine learning models trained on historical security telemetry can detect multi-step or long-range attack campaigns, i.e., coordinated attacks that evade customary rule-based detection methods [5].

Threat intelligence information is used to augment the behavioral analytics with indicators of compromise, known attack signatures, and other behavior patterns of adversaries against the AI system. Multiple telemetry sources, such as network traffic, application programming interface (API) calls, resource consumption metrics, and application logs, can be used in combination. Advanced analytics to identify attack patterns that may be distributed over time or across multiple hosts and thus may not be visible to a single sensor. By providing interpretable results, threat detection helps security analysts distinguish security incidents from normal anomalies and reduce false positives, enabling them to concentrate their investigations on real threats. [6].

Explainable anomaly detection systems give security analysts explanations for why behaviors were identified as anomalous (i.e., which feature values were out of the expected range and/or how far out of range the value was) [6]. Transparency in threat detection encourages analysts to build their pattern recognition capabilities, and it creates a record of justifications for security actions taken. Organizations that are bound to regulatory compliance require security systems to be explainable, providing justification of the rationale for automatic decisions on data access, service availability, and incident response.

3.5 Compliance and Governance Alignment

USPM combines controls from the NIST AI Risk Management Framework, the European Union's AI Act, and the Open Web Application Security Project Top Ten vulnerabilities in LLMs. By incorporating these standards through automated compliance checking, USPM ensures compliance along with the corresponding documentation for auditing purposes for the configuration, access control, data processing, and operational procedures. Compliance dashboards provide a real-time overview of policy compliance status, exceptions, and remediation coverage across distributed LLM deployments.

Governance workflows allow teams to enforce approval for high-risk configurations and security reviews of any new model deployments. Business decisions made by the team are recorded for audit. Automated reporting generates compliance artifacts used for regulatory submissions and audits, such as configuration snapshots, access logs, change provenance, and incident response logs. Finally, the explainability of these compliance decisions plays a critical role in regulatory reporting. Organizations need to be able to show the existence and adequacy of controls along with justifying the automated

compliance decisions [6]. In the case of AI-based compliance systems, explainability is important for auditors who must confirm that automated checks for policy compliance are implemented accurately and that risk assessments are authentic. Table 2 delineates the foundational components comprising comprehensive Unified AI Security Posture Management frameworks, emphasizing automated discovery mechanisms, risk assessment capabilities, identity governance structures, behavioral analytics, and compliance alignment necessary for multi-cloud LLM security resilience.

4. Emerging Trends and Industry Developments

4.1 Zero-Trust Architecture for Multi-LLM Systems

Modern, zero-trust based multi-LLM architectures differ from perimeter-based security models that protect legacy cloud services by establishing trust in users and services inside the cloud network perimeter. In a zero-trust architecture, every interaction between users, services, and models is authenticated and authorized by default as opposed to implicitly trusted [7]. Verification mechanisms may be employed to establish the authenticity of these identities, the security posture of devices, the context in which requests are made, and the behavioral consistency of entities before granting access to LLM resources. The principle of least privilege dictates that an entity should only have access to what it needs to do its job.

Zero-trust principles for workloads with AI models include access controls at multiple levels of granularity, such as at the level of model, feature, and data element. Micro-segmentation segments individual models or model versions in order to avoid lateral movement and limit the impact if compromised [7]. Network segmentation strategies divide the cloud into independently secured enclaves, creating trust boundaries. Systems within the same cloud provider's infrastructure do not necessarily share a trust boundary and may require separate authentication and authorization controls. These permissions are automatically re-evaluated in changing risk contexts, such as when suspicious behavior is detected or threat intelligence indicates high risk.

Zero-trust architectures for multi-cloud environments have to deal with issues such as the lack of centralized identity management, the differences in security controls for network access between cloud service providers, and the complexities of communication between cloud service providers. Organizations use centralized policy decision points, which apply unified security policies, and distributed policy enforcement points in the cloud. Thus, regardless of the underlying platforms, the same security semantics apply, and so software-defined perimeters dynamically isolate networks, establish encrypted tunnels between authorized endpoints, and deny access to unauthorized sites and users irrespective of network location or device type.

4.2 Model Security and Adversarial Defenses

Model inversion attacks and related threats against ML systems pose a meaningful threat to companies that use LLMs in multi-cloud configurations. An attacker may be able to reconstruct the training data that was used to create a model from the parameters and inference behavior, which discloses sensitive information found in training datasets [8]. Membership inference attacks determine whether a certain record was part of a training set. This can create privacy problems when private or proprietary training data are used to build the model. Model extraction attacks may allow an opponent to recreate models and intellectual property and compete with expensive model training efforts.

Defenses against model inversion attacks include differential privacy, model watermarking, adversarial training, and inference query monitoring systems, among other techniques [8]. Differential privacy is a state-of-the-art learning algorithm that adds calibrated noise to a model during training and/or prediction, thereby concealing the training data without greatly affecting the model's accuracy. Model watermarking embeds identifying signatures into model parameters to prove ownership and detect counterfeits. Adversarial training is used to expose models to attacks at training time, improving their resilience against prompt injection attacks and output steering attacks at inference time.

Query monitoring detects suspicious patterns in inference requests, which may indicate model extraction or probing for open parameters [8]. Rate limiting and query filtering can defend against large model extraction attacks that require meaningful interaction with the target model. Output filtering mechanisms reduce the risk that models inadvertently regurgitate sensitive training data or otherwise produce content that violates content policies even in the face of adversarial attempts to circumvent model safety mechanisms. Model security defenses can be integrated into cross-cutting posture management workflows to standardize protections of multiple LLM deployments across heterogeneous cloud environments.

4.3 Regulatory Pressure and Compliance Evolution

Regulations are underway to establish security posture controls for AI/LLM systems, accelerating the convergence of management capabilities across the enterprise. The European Union Artificial Intelligence Act outlines a thorough set of requirements for AI systems offered on the European market. Many applications are classified as high-risk systems, which must undergo in-depth conformity assessment, monitoring, and documentation [9]. In addition to initial certification, these requirements may also include showing the continued effectiveness of security controls and incident reporting obligations and that updates are made in response to new vulnerabilities and changes in legal interpretation.

The EU system for high-risk AI systems introduces requirements relating to risk management systems, quality management systems, keeping technical documentation up to date, and maintaining documentation detailing how the system works, including how it makes decisions [10]. Such provisions will apply to high-risk LLMs that affect rights, safety, or access. Transparency and explainability, letting individuals affected by an AI system receive 'meaningful information' about how the system works, and enabling individuals to contest or appeal against an automated decision.

Prohibited AI technology includes systems that use subliminal manipulation, exploit weaknesses of vulnerable persons, use social scoring by public authorities, and use real-time remote biometric identification for the purpose of law enforcement in publicly accessible spaces, except where narrowly allowed [10]. Concerning corporate requirements, these compliance management systems are to identify cross-jurisdiction requirements that are relevant, verify that the deployed system is compliant, and be capable of storing audit logs. Given the global nature of cloud computing and AI services, organizations need to consider regulatory requirements in all jurisdictions where their systems are deployed and where their activities impact individuals.

Industry forecasts indicate that consolidated AI posture frameworks integrating CSPM and DevSecOps capabilities will become standard features of enterprise security stacks [9]. Global drivers include regulations, more advanced enterprises, and the acknowledgment that legacy security practices have failed to reduce AI risk. Early adopter organizations using unified frameworks report better visibility, faster response times, and stronger compliance than those using fragmented tool approaches. Table 3 examines contemporary security architectural developments addressing unique challenges in distributed Large Language Model environments, encompassing zero-trust implementations, adversarial defense mechanisms, and evolving regulatory compliance requirements shaping enterprise AI security strategies.

5. Debates, Forecasts, and Actionable Insights

5.1 AI Autonomy Versus Security Control Tensions

Many questions remain for the security community and the broader AI community about how tightly security controls can be implemented when AI workloads must be able to handle new inputs and changes to their environment without degrading their performance or operational flexibility. Other specific issues include latency penalties for real-time security checks, burdensome access control for collaborative development of models, and excessive compliance documentation efforts.

On the other hand, security advocates argue that posture management is required in order for AI services to be trusted with sensitive information and high-stakes decision-making in enterprises, where the impact of security incidents is broader and farther-reaching than in customary information technology. Attackers can steal AI models to give competitors an advantage, use training data to exploit privacy safeguards, or even modify the AI's outputs to force the business to fail. The frequent attacks on AI services in enterprise settings support the trend for security-first AI systems that prioritize defensive capabilities over rapid deployment [9].

Balanced approaches are integrated security models that combine strong security and low friction through the use of intelligent automation, risk-based controls, and developer-centered tooling. Organizations with mature AI security programs reported that it is less disruptive to integrate security into AI systems as they are developed rather than attaching security after deployment. Security-by-design methodologies with threat modeling, secure architecture patterns, and continuous validation throughout the development lifecycle enable rapid innovation without creating security technical debt.

5.2 Standardization and Framework Evolution

Technical communities are still assessing whether existing CSPM and extended detection and response frameworks provide adequate coverage for AI posture or if a new AI-SPM framework is needed. Early adopters have had success feeding AI

risk signals into existing security frameworks, as existing tools and analyst workflows could be employed to address AI posture. This allows enterprises to quickly deploy existing platforms that incorporate AI capabilities as the vendor matures its AI portfolio.

Alternatively, an external AI risk module could be architected separately from information technology security operations, with tailored architectures taking into account the unique threat models, operational patterns, and risk analysis of AI systems. Such dedicated architectures may optimize for AI workloads and provide better performance for risk analytics, including better fidelity and fewer false positives over customary security tools that lack customized analytics. The downside to this approach is increased siloing with gaps in visibility and lack of integrated incident response across mixed hybrid workloads.

As the AI security tooling ecosystem matures, industry experts have recognized that the old models may not be sufficient against advanced threats to machine learning systems [9]. Machine learning systems have unique attack surfaces such as training data poisoning, adversarial examples, model extraction, and inference manipulation, which existing network and endpoint security models cannot effectively defend against. There are some signs of consolidation of the AI security industry, with major security vendors acquiring startups and combining management into broader security platforms.

5.3 Actionable Implementation Recommendations

Multi-cloud LLM infrastructure enterprises should implement unified AI security posture management across the entire machine learning life cycle, from the development environment through the production runtime environment, across all public clouds. Integrating security policies and controls during the model development phase will help enterprises adopt a security-by-design approach and integrate security controls into the architecture, rather than bolting them on post-production. Continuous security validation throughout deployment pipelines helps to prevent insecure configurations from being deployed to production while maintaining fast release velocities through automated checks and fixes.

These identity governance principles are key enablers of unified architectures, where centralized identity controls prevent the emergence of fragmented risks from the Model Context Protocol and other domain-specific cloud-native service authentication protocols. Identity-centric security architectures that use least-privileged access controls, continuous authentication, and enable auditing of identity activities are recommended across LLM infrastructure [7]. Zero-trust architectures assume that every request for access to resources, both inside and outside an organization's periphery, is a potential threat and must be verified.

Security posture frameworks need to explicitly map to global regulatory and industry standards by mapping technical controls to their regulatory requirements and keeping auditable policy compliance documentation. Proactive compliance validation, through automated checks and manual audits, is needed as regulations evolve and the portfolio of organizations' AI systems grows, thus requiring continuous behavioral analytics of AI usage patterns to detect, in advance of a successful security breach, attempts to exploit, to adversarially manipulate, or emerging threats to the models. Organizations should also seek out explainable AI techniques, as automated decision-making must be transparent and auditable for compliance or when security teams accept the risks of automated systems [6].

5.4 Criticisms and Implementation Challenges

Unified AI security posture management frameworks are early-stage technologies, and most commercial options lack an independent assessment of their security or peer-reviewed research about their effectiveness. Organizations considering purchasing a vendor's offering should ask for empirical evidence of the framework's detection capabilities, false positive rates, and operational cost in addition to vendor materials. Proof-of-concept deployments in representative environments (where conditions and requirements would be similar to those in actual deployments) can assist in evaluating proposed solutions.

Behavior detection systems can create many false positive alerts that may need to be tuned according to an organization's environment and usage. Poorly tuned alert generation can cause alert fatigue for security staff and cause the users to overlook genuine threats. Successful systems require considerable tuning up front and continuing adjustment as the AI workloads, attack patterns, and threat landscape evolve over time [9]. In particular, it is necessary to monitor the false positive rates and the time taken to detect a true positive, as well as the analyst satisfaction with automated alerts.

Multi-cloud posture monitoring has costs associated with cloud application programming interface calls for constant scanning, storage of telemetry for an

extended period of time, computers for processing the analysis, and personnel costs associated with security operations that are not fully automated. Organizations also need to assess the total cost of ownership associated with unifying their tools and adopting a unified framework. Automating low-level, high-volume routine work with humans stepping in for complex investigations is key to resource-efficient and sustainable security. Table 4 addresses critical debates, standardization approaches, and practical challenges confronting organizations implementing unified security posture management for multi-cloud Large Language Model infrastructure, balancing security imperatives with operational requirements.

Table 1: *Critical Gaps in Current Multi-Cloud LLM Security Approaches [3, 4]*

Gap Category	Current Limitation	Impact on Security Posture
Visibility and Integration	Fragmented monitoring across individual cloud platforms without cross-platform correlation capabilities	Creates blind spots enabling adversaries to exploit cloud boundary seams and evade single-platform detection systems
Compliance Frameworks	Static rule sets requiring manual updates for regulatory changes and cloud service modifications	Temporal gaps where organizations operate under outdated security policies inadequate for emerging threats
Contextual Intelligence	Anomaly detection without business context, model sensitivity classification, or operational criticality	Inability to differentiate routine model operations from catastrophic data exfiltration attempts
Remediation Mechanisms	Predetermined response playbooks without adaptive learning from incident outcomes	Recurring security issues addressed symptomatically without preventing root causes or evolving defenses
Predictive Capabilities	Reactive incident response lacking attack path forecasting and proactive vulnerability identification	Organizations remain vulnerable to exploitation pathways discoverable through adversarial scenario simulation

Table 2*: Essential Components of Unified AI Security Posture Management Architecture [5, 6]*

Framework Component	Primary Function	Key Capability
Continuous Discovery and Asset Inventory	Automated identification and cataloging of AI assets across multi-cloud platforms	Behavioral pattern recognition for detecting models, runtimes, toolchains, and data pipelines with real-time reconciliation
Risk Assessment and Misconfiguration Detection	Proactive vulnerability identification specific to LLM deployment architectures	AI telemetry analysis integrated with behavioral profiling and guardrail validation for composite risk scoring
Identity and Access Governance	Centralized authentication, authorization, and auditing across heterogeneous cloud environments	Just-in-time access provisioning with continuous verification detecting anomalous credential usage patterns
Threat Detection and Behavioral Analytics	Real-time anomaly detection identifying suspicious model usage and inference patterns	Machine learning models recognizing coordinated attacks through cross-correlation of multiple telemetry streams
Compliance and Governance Alignment	Automated validation against regulatory frameworks and organizational policies	Real-time compliance dashboards with auditable documentation for regulatory submission and external audit

Table 3: *Emerging Security Paradigms for Multi-Cloud LLM Deployments [7, 8]*

Security Paradigm	Architectural Approach	Strategic Implementation
Zero-Trust Architecture	Continuous verification of every user, service, and model interaction without implicit trust assumptions	Micro-segmentation isolating models with centralized policy decision points enforcing least-privilege access
Model Inversion Defense	Multi-layered protection against training data reconstruction and intellectual property theft	Differential privacy mechanisms combined with model watermarking and adversarial training for inference robustness
Query Monitoring Systems	Behavioral analysis detecting model extraction attempts and systematic vulnerability probing	Rate limiting and filtering mechanisms preventing large-scale attacks requiring extensive target model interaction
European Union AI Act Compliance	Comprehensive regulatory framework classifying high-risk AI systems requiring conformity assessment	Risk management documentation with continuous monitoring and transparent governance throughout system lifecycle
Prohibited AI Practices	Regulatory restrictions preventing manipulative, exploitative, and invasive AI implementations	Compliance management validating deployed systems across jurisdictions with auditable verification processes

Table 4: Implementation Considerations for Unified AI Security Posture Management [9, 10]

Implementation Dimension	Primary Consideration	Strategic Recommendation
AI Autonomy vs Security Control	Tension between comprehensive security frameworks and AI workload performance requirements	Security-by-design methodologies embedding controls during development rather than post-deployment retrofitting
Framework Standardization	Debate between integrating AI signals into existing CSPM platforms versus dedicated AI-SPM architectures	Risk-based evaluation considering organizational maturity, existing infrastructure, and AI-specific threat landscape
Explainability Requirements	Transparency demands for automated security decisions affecting data access and compliance status	Implementation of feature importance analysis and counterfactual explanations enabling analyst validation and trust
Detection System Tuning	False positive generation requiring environment-specific calibration and continuous refinement	Metrics tracking alert accuracy and analyst satisfaction guiding iterative improvement as workloads evolve
Resource Allocation	Total cost of ownership including API costs, storage expenses, compute requirements, and personnel overhead	Strategic automation targeting high-volume routine tasks while preserving human expertise for complex investigations

6. Conclusions

As enterprises deploy large language models and AI workloads at scale across multi-cloud environments, unified AI security posture management has become a necessity. Static security models with limited visibility and slow incident response are no longer adequate for the dynamic threat landscape and tightening regulatory

requirements that characterize modern AI deployments. Adoption of full-stack AI-native platforms, continuous discovery, behavioral analytics, dynamic compliance, and automated remediation by security-focused organizations can address security resilience challenges. The near-universal targeting of AI services and the widespread exploitation of identity as a primary attack surface further underscore the urgency of addressing existing and emerging vulnerabilities in AI-first organizations. As regulatory demands for security controls, governance, and continuous monitoring across the AI system lifecycle grow, organizations that build unified posture management capabilities are achieving improved detection, faster response, better compliance posture, and reduced security operational costs through advanced automation, as well as addressing the compliance requirements associated with various regulations such as the GDPR and CCPA. Despite this, the maturity gap between security and AI deployment, the difficulty of integrating security tools, and the resources required to implement security practices continue to pose challenges that need to be weighed against the need to effectively protect systems from risk. Future work is likely to see tighter integration of security platforms into AI development toolchains, greater explainability of automated security decisions, and open interoperability standards to support heterogeneous vendor ecosystems. The advent of multi-cloud LLM deployments across the breadth of enterprise operations poses myriad technical and security challenges. Unified security posture management capabilities will separate organizations that are able to safely unlock the value of AI from those that experience catastrophic failures leading to reputational and business collapse. They are a qualitative change in how organizations think about and implement enterprise-scale AI systems and services that run across globally distributed, interconnected cloud environments.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
- **Use of AI Tools:** The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

References

1. Anil Chowdary Inaganti, et al., "Cloud Security Posture Management (CSPM) with AI : Automating Compliance and Threat Detection," Artificial Intelligence and Machine Learning Review, 2021. [Online]. Available: https://scipublication.com/index.php/AIMLR/article/view/129/123
2. Ramanpreet Kaur, et al., "Artificial intelligence for cybersecurity: Literature review and future research directions," Information Fusion, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253523001136
3. Ebimor Yinka Gbabo, et al., "Constructing AI-Enabled Compliance Automation Models for Real-Time Regulatory Reporting in Energy Systems," World Journal of Innovation and Modern Technology, 2025. [Online]. Available: https://iiardjournals.org/get/WJIMT/VOL.%209%20NO.%206%202025/Constructing%20AI-Enabled%20Compliance%2027-39.pdf
4. Chongzhou Fang, et al., "Assessing and Mitigating Heterogeneity-Driven Security Threats in the Cloud," ACM Digital Library, 2025. [Online]. Available: https://dl.acm.org/doi/full/10.1145/3746228
5. Gregory Coppola, et al., "Enhancing Cloud Security Posture for Ubiquitous Data Access with a Cybersecurity Framework Based Management Tool," 2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10316003
6. Gaith Rjoub, et al., "A Survey on Explainable Artificial Intelligence for Cybersecurity," IEEE Transactions on Network and Service Management, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10143992
7. Scott Rose, et al., "Zero-Trust Architectures," NIST Special Publication 800-207, 2020. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf
8. Wencheng Yang, et al., "Deep Learning Model Inversion Attacks and Defenses: A Comprehensive Survey," Artificial Intelligence Review, 2025.

[Online]. Available: https://link.springer.com/article/10.1007/s10462-025-11248-0

9. Yulong Wang, et al., "Adversarial Attacks and Defenses in Machine Learning-Empowered Communication Systems and Networks: A Contemporary Survey," IEEE Communications Surveys & Tutorials, 2023. [Online]. Available: https://dl.acm.org/doi/10.1109/COMST.2023.3319492
10. Tambiama Madiega, "Artificial Intelligence Act," European Parliamentary Research Service, 2024. [Online]. Available: https://www.dirittobancario.it/wp-content/uploads/2024/09/Documento-riassuntivo-dellAI-Act-Parlamento-europeo-settembre-2024.pdf