Uwaisu Abubakar Umar, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Cloud Computing Security in the Age of AI: A Systematic Review Using Artificial Intelligence as Both Research Instrument and Object of Study

[1]Uwaisu Abubakar Umar, [2]Ibrahim Haruna Ibrahim,
[3]Aliyu Aminu Dahiru, [4]Emmanuel Martin Teman
Department of Computer Science, Modibbo Adama University, Yola, Nigeria[1,2,4]
Department of Computer Science, Usman Danfodio University, Sokoto, Nigeria[3]

Abstract- The rapid migration of organizational infrastructure to cloud computing has introduced unprecedented scalability and efficiency alongside increasingly complex security vulnerabilities. This review examines the evolving relationship between artificial intelligence (AI) and cloud security through an AI-Augmented Systematic Literature Review (ASLR), in which AI serves simultaneously as a research instrument for synthesizing literature and as the object of study within cloud security architectures. Drawing on studies published between 2020 and 2026 sourced from IEEE Xplore, ACM Digital Library, and SpringerLink, the review addresses three objectives: examining how cloud infrastructure supports the deployment of resource-intensive machine learning (ML) models, investigating AI's role as an adaptive defensive layer, and identifying new security vulnerabilities introduced by embedding AI into cloud networks. Findings show that cloud platforms enable resource-intensive ML through elastic compute allocation, architectural abstraction, and intelligent orchestration; that AI strengthens defense through predictive threat detection, continuous policy adaptation, and privacy-preserving federated learning; and that this same integration introduces distinct new vulnerability classes, including agent privilege escalation, lifecycle-specific exploitation, cascading cross-layer failures, serverless architectural weaknesses, and adversarial manipulation. The review concludes that AI functions as a double-edged capability within cloud ecosystems, simultaneously strengthening and expanding the attack surface, and argues for lifecycle-aware, zero-trust security models validated under real-world rather than solely simulated conditions.

Keywords— Artificial Intelligence, Cloud Computing, Security, Threat Detection, Vulnerability.

I. INTRODUCTION

Cloud computing has fundamentally altered the paradigm of IT infrastructure, shifting organizations from localized, capital-intensive data centers to distributed, on-demand services. The rapid increase in adoption of Cloud Computing has transformed how global enterprises and academic institutions manage, store, and process data. Cloud architectures have driven unprecedented technological efficiency by offering scalable and on-demand resources seamlessly. However, this decentralized paradigm has simultaneously expanded the attack surface for malicious actors, uplifting Cybersecurity to the top of cloud adoption challenges. As cloud infrastructures become more complex, traditional, perimeter-based security measures are proving inadequate against increasingly sophisticated, automated, and distributed cyber threats.

To bridge this defensive gap, the integration of Artificial Intelligence (AI) and Machine Learning (ML) has emerged as a critical necessity. AI-driven security frameworks capable of continuous monitoring, inconsistency recognition, and autonomous incident response are necessary in modern Threat Detection. So, understanding how these intelligent algorithms behave, scale, and secure cloud environments has become a primary objective within contemporary computer science research. In this context, AI serves

as the core object of study, representing the frontline of modern defensive computing.

Despite the widespread agreement on AI's effectiveness in cloud security, the volume of emerging research makes it difficult to synthesize a cohesive understanding of the field's current state. To address this, this paper presents a Systematic Literature Review (SLR) of recent advancements in AI-driven cloud security. The study adopts a dual-layered approach to artificial intelligence: AI is not merely the subject being investigated, but it is also deployed as the primary research instrument. By leveraging advanced Natural Language Processing (NLP) and Large Language Models (LLMs) to systematically retrieve, filter, and synthesize the vast corpus of academic literature, this methodology minimizes human selection bias and significantly accelerates the analytical process.

Objectives

This review maps the current landscape of AI-based issues and threat mitigation in cloud environments, identifying both breakthroughs and persistent vulnerabilities. This review is guided by the following research objectives:

- To examine how cloud infrastructure supports the deployment of resource-intensive machine learning models.
- To investigate the role of artificial intelligence as a defensive layer within cloud security architectures.
- To identify the new security vulnerabilities introduced by integrating artificial intelligence into cloud networks.

II. METHODOLOGY: AN AI-CENTRIC FRAMEWORK

We adopt a dual approach that places AI at the center of both how we review and what we review, with full disclosure in accordance with PRISMA 2020 and the emerging PRISMA-AI extension. This systematic literature review followed the PRISMA 2020 statement (Page et al., 2021) and the PRISMA-AI extension (Cacciamani et al., 2023) to identify, screen, and synthesize studies on security vulnerabilities and threat detection in cloud-based machine learning environments. A structured search across IEEE Xplore, ACM Digital Library, Scopus, and arXiv used keywords related to cloud computing, deep learning, security, and MLaaS. Records were screened against predefined inclusion and exclusion criteria. Data extraction captured study objectives, cloud deployment models, security threats, mitigation strategies, and evaluation metrics. PRISMA-AI was particularly relevant given the review's focus on AI-driven security frameworks and ML model lifecycle management. The final corpus was synthesized narratively and organized thematically to address the research questions.

AI as a Research Instrument

Traditional Systematic Literature Reviews (SLRs) are constrained by the manual screening and synthesis capacity of researchers. We address this with an AI-Augmented Systematic Literature Review (ASLR), detailed in the subsections below.

Search Strategy. We searched three databases on 31 May 2026: IEEE Xplore, ACM Digital Library, and SpringerLink. Date range: January 2020–May 2026. Boolean string: ("cloud computing" OR "multi-cloud") AND ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("security" OR "threat" OR "vulnerability").

Eligibility Criteria. Inclusion: (1) peer-reviewed journal/conference papers or arXiv preprints with full text in English; (2) primary studies, reviews, or meta-analyses; (3) focus on AI methods applied to IaaS, PaaS, or SaaS security. Exclusion: opinion pieces, editorials, non-English studies, or studies focused solely on on-premises networks without cloud relevance.

Data Extraction. Two reviewers extracted study design, cloud deployment model, AI technique (and its role as instrument vs. object), dataset, evaluation metrics, reported performance, and stated limitations using Covidence. AI-assisted extraction was used for initial field population, followed by 100% human verification.

Thematic Synthesis via Large Language Models. We used LLMs to extract core findings, compare

methodological frameworks across selected papers, and group the literature into distinct security domains, effectively eliminating human fatigue and accelerating the synthesis process. Each paper was parsed by an instruction-tuned LLM that extracted objective, AI technique, cloud layer (IaaS, PaaS, SaaS), dataset, evaluation metrics, reported performance, and stated limitations, and generated a one-sentence finding tagged by primary threat type.

Validation and Risk of Bias. All AI outputs were logged with timestamps, model version, and prompts to ensure auditability (PRISMA Item 7). Risk of bias was assessed using ROBIS for reviews and a modified PROBAST checklist for primary AI studies, with particular attention to data leakage, inadequate external validation, and selective reporting.

AI as the Object of Study

Artificial intelligence is the primary object of investigation in this review, with the literature focusing on how algorithmic systems operate within cloud environments. Unlike studies that treat AI solely as a methodological tool for conducting research, the present review positions AI systems themselves as the central subject of analysis. The surveyed literature examines the design, implementation, and behavior of machine learning models and other algorithmic components as they are deployed within cloud infrastructures, with particular attention to the security implications that arise from this interaction. Specifically, we investigate:

- How cloud infrastructure supports the deployment of resource-intensive machine learning models.
- How AI acts as a defensive layer within cloud security architectures.
- What new security vulnerabilities are presented by incorporating AI into cloud networks.

Findings: The Dual Role of AI in Cloud Security

The synthesized literature reveals a consistent, three-part pattern in how artificial intelligence and cloud computing intersect. We begin by exploring how cloud infrastructure has evolved to support the demands of resource-intensive machine learning, examining the mechanisms that make such deployment practical. From there, the analysis proceeds to consider how these same intelligent systems are being recursively applied to the cloud environment itself, engendering a paradigmatic shift in security architecture toward adaptive, autonomously responsive defensive frameworks. Finally, we address a less comfortable but equally important question: what new risks emerge when AI becomes deeply embedded in cloud environments? Together, these three threads paint a fuller picture of AI's evolving and often enigmatic role within the cloud ecosystem.

How Cloud Infrastructure Supports Resource-Intensive ML Models

Deploying resource-intensive ML models that require substantial computational power for training, large memory footprints for inference, or low-latency responses in production presents significant infrastructure challenges. Cloud infrastructure has emerged as a primary solution to these challenges, offering flexible compute, specialized hardware, and intelligent planning that collectively enable the deployment of models that would be impractical on local or static systems. This section examines five key studies that illuminate the specific mechanisms through which cloud environments address the demands of resource-intensive ML workloads.

Cost-effective compute scaling for inference workloads. Lemos et al. (2026), directly addressed the question of how cloud infrastructure can support ML inference without prohibitive costs by evaluating CPU-only virtual machines across AWS, GCP, and Azure. Their findings demonstrated that while GPU instances remain necessary for latency-critical applications, optimized CPU instances can deploy resource-intensive models at acceptable performance levels for a significantly reduced cost. This reveals a core support mechanism of cloud infrastructure: the ability to match instance types to workload requirements, allowing practitioners to trade performance against cost in granular, deliberate ways. The study establishes that cloud platforms support resource-intensive ML not simply by providing the most powerful hardware, but by

offering a spectrum of choices that accommodate diverse deployment constraints.

Distributed architectures for the full ML lifecycle. García et al. (2020), addressed the question from an architectural perspective through the DEEP-Hybrid-DataCloud proposal, which supports resource-intensive ML models by distributing computation across hybrid cloud environments. The architecture specifically targets the full lifecycle from data preprocessing through model training to production deployment by dynamically allocating GPU-accelerated resources for compute-heavy training phases while reserving cost-effective CPU resources for lighter tasks. This demonstrates that cloud infrastructure supports ML deployment not as a monolithic resource pool but as a composable environment where different stages of the ML pipeline draw upon appropriately scaled resources. The distributed architecture is particularly relevant for resource-intensive models whose training demands exceed the capacity of any single node.
Abstraction layers for multi-cloud deployment. Wei et al. (2022), confronted the operational challenge of deploying resource-intensive models across heterogeneous cloud environments. Their unified management framework introduces a provider-neutral policy layer that abstracts away the differences between cloud platforms, enabling models to be deployed across AWS, GCP, Azure, and others without reconfiguration. This abstraction mechanism is critical for resource-intensive workloads because it prevents vendor lock-in, allowing organizations to move models to the most cost-effective or better-performing provider as needs evolve. The framework also simplifies the management of complex, multi-stage ML pipelines by allowing users to define high-level deployment policies such as "maximize throughput under a $500 budget" that the system automatically translates into provider-specific configurations.

Intelligent, dynamic resource allocation. Xu et al. (2022), demonstrated that cloud infrastructure supports resource-intensive ML models most effectively when resource allocation is adaptive rather than static. Their deep reinforcement learning approach continuously monitors workload patterns and system states, learning optimal policies for when to scale cloud resources up or down, and when to supplement local edge capacity with rented cloud instances. For resource-intensive models, this dynamic allocation is transformative: it eliminates the need for over-provisioning to handle peak demand while ensuring that sufficient resources are available when computational demands spike. The cost savings achieved through this approach compared to static allocation directly address one of the primary barriers to deploying large models in production environments.

ML-powered cloud management as an enabling feedback loop. Khan et al., (2022) provided the broader context by reviewing how ML techniques are themselves used to manage the cloud infrastructure that hosts ML models. Their survey documents a paradigm in which cloud platforms support resource intensive models through data-driven management approaches predictive scaling, anomaly detection, and automated scheduling that respond to workload patterns in real time. This bidirectional relationship is central to understanding how cloud infrastructure supports demanding ML deployments: as models grow more resource-intensive, the infrastructure managing them becomes more intelligent, creating a feedback loop where cloud platforms continuously improve their ability to accommodate large-scale ML workloads. The review also highlights ongoing challenges, including the management overhead of ML-based systems themselves and the difficulty of generalizing management models across diverse workloads.

Across these five studies, three consistent mechanisms emerge through which cloud infrastructure enables the deployment of resource-intensive ML models. Firstly, elastic compute elasticity ensures that models can access the precise quantity and type of computational resources they require at any given moment, from CPU-only inference to multi-GPU distributed training. Secondly, abstraction layers whether architectural (DEEP-Hybrid-DataCloud), policy-based, or provider-neutral, shields ML practitioners from infrastructure complexity while enabling deployment across heterogeneous environments. Thirdly,

intelligent resource orchestration, exemplified by the deep reinforcement learning approach and the ML-driven management reviewed by Khan et al., optimizes the allocation of cloud resources dynamically, balancing performance requirements against cost constraints. Together, these mechanisms position cloud infrastructure not merely as a host for ML models but as an active, adaptive environment that responds to the unique and variable demands of resource-intensive machine learning.

III. AI ACTS AS A DEFENSIVE LAYER IN CLOUD SECURITY

Traditional cloud security models have largely relied on static, rule-based approaches with predefined signatures, fixed firewall policies, and manual threat detection that struggle to keep pace with progressively sophisticated and evolving cyber threats. As attack vectors grow more complex and cloud environments become more dynamic, researchers have turned to artificial intelligence as a responsive and adaptive defensive layer that can learn, predict, and react to threats in real time. This section blends three studies that scrutinize how AI enhances cloud security architectures, identifying both the mechanisms through which AI provides defensive capabilities and the limitations that remain to be addressed.

Federated learning for collaborative threat detection. Asha and Kanaga (2026) developed a framework that integrates Long Short-Term Memory (LSTM) networks with federated learning to enhance cloud firewall capabilities. Their approach addresses a fundamental limitation of traditional firewalls: the inability to detect novel or zero-day attacks without preconfigured signatures. By employing LSTM a recurrent neural network architecture well-suited for analyzing sequential data, the framework learns patterns of malicious network traffic over time, enabling it to identify threats that differ from normal behavior. The federated learning component further strengthens this defensive layer by allowing multiple cloud nodes to collaboratively train a shared detection model without exposing sensitive raw data, preserving privacy while improving detection accuracy across the distributed environment. The framework achieved a 94.7% detection rate, demonstrating that AI can serve as an intelligent, privacy-preserving defensive layer that improves collectively across the cloud network.

Adaptive, real-time policy adjustment. Kareem et al. (2025) proposed an adaptive security model that moves beyond static defense by continuously adjusting security policies in response to evolving threat conditions. Unlike traditional architectures where firewall rules and access controls remain fixed until manually updated, this AI-driven model monitors network behavior in real time, detects anomalies, and autonomously reconfigures security policies to mitigate emerging risks. The model's improved detection success over static baselines illustrates a key mechanism through which AI functions as a defensive layer: adaptability. Rather than relying on a fixed set of rules that attackers can learn to circumvent, AI-enabled security architectures maintain a dynamic defense posture that evolves alongside the threat landscape. This adaptability is particularly serious in cloud environments, where workloads shift, new instances spin up, and traffic patterns vary endlessly.

Deep learning for specialized, high-stakes environments. Dixit et al. (2026) focused on applying AI-driven threat detection to healthcare cloud environments a domain where security failures carry particularly severe consequences due to the sensitivity of patient data and regulatory requirements. Their framework uses deep learning models to detect malicious activity with high accuracy, effectively identifying intrusions that might evade traditional signature-based systems. The study achieved strong operational security outcomes, demonstrating that AI can function as a defensive layer not only in general-purpose cloud architectures but also in specialized, high-stakes contexts where precision and reliability are paramount. This work extends the scope of AI-driven security beyond generic threat detection to domain-specific applications, suggesting that AI defensive layers can be tailored to the unique threat profiles and compliance requirements of different cloud environments.

Across these three studies, AI acts as a defensive layer within cloud security architectures through three primary mechanisms. Firstly, predictive detection; using models like LSTM and deep neural networks to identify threats based on behavioral patterns rather than static signatures, enabling the detection of novel and zero-day attacks. Secondly, continuous adaptation allowing security policies to evolve in real time as threat conditions change, replacing the rigid rule collections of traditional architectures with dynamic, learning-based defenses. Thirdly, distributed intelligence leveraging approaches like federated learning to improve detection across the cloud network collaboratively while preserving data privacy. However, a significant shared limitation across all three studies is their reliance on simulated rather than real-world datasets for training and evaluation. While the reported detection rates and accuracy figures are promising, the absence of validation in production cloud environments raises questions about generalizability, performance under real-world traffic conditions, and susceptibility to adversarial attacks against the AI models themselves. This gap underscores the need for future research to test AI-driven defensive layers in operational cloud settings before these approaches can be widely adopted as reliable security mechanisms.

New Vulnerabilities Introduced by AI in Cloud Networks

The integration of artificial intelligence into cloud networks introduces capabilities that are transformative for both defenders and attackers. While AI strengthens security in many respects, its incorporation simultaneously expands the attack surface in ways that traditional cloud security models do not account for. AI agents, ML models, and automated decision-making systems become new targets, new vectors, and new points of failure. This section synthesizes five studies that collectively examine the novel vulnerabilities arising from embedding AI into cloud architectures, identifying distinct classes of risk and the emerging consensus on how to address them.

Agent privilege vulnerabilities. Goel (2026), identified a foundational security risk introduced by AI agents in cloud environments: the tendency to grant these agents excessive permissions overprivileged tools, broad access to cloud APIs, and stored credentials that, if compromised, provide attackers with powerful lateral movement capabilities. Unlike traditional software components with narrowly scoped permissions, AI agents are often given wide access to accomplish open-ended tasks, creating a privilege mismatch that attackers can exploit. Goel argues that AI agents must be treated as privileged principals in their own right, subject to the same if not stricter access controls, audit logging, and credential management as human administrators. The study highlights that credential leakage from AI agents represents a uniquely dangerous vulnerability because these agents operate autonomously and at machine speed, meaning a single compromised agent can cause disproportionate damage before human operators can respond.

Lifecycle exploitation. Czaja et al. (2025), examined how the very process of developing and deploying machine learning models in cloud environments creates new attack surfaces at each stage of the ML lifecycle. During training, models are vulnerable to data poisoning attacks, where adversaries inject malicious samples into training data to corrupt model behavior. During inference, models face adversarial attacks carefully crafted inputs designed to cause misclassification or unexpected outputs. Czaja et al. emphasize that these vulnerabilities are fundamentally different from traditional cloud security threats because they target the model's logic and learned behavior rather than infrastructure misconfigurations or network protocols. Each phase of the ML lifecycle requires specialized defenses: robust data validation and sanitization for training, adversarial training and input filtering for inference, and continuous monitoring for model drift that may indicate an ongoing attack.

Cascading decision-layer vulnerabilities. Radanliev et al. (2026), proposed an analytical framework demonstrating that vulnerabilities in agentic AI systems do not remain isolated, they propagate across different layers of decision-making. In a cloud network where AI agents operate at multiple levels (individual model decisions, organization layers,

policy enforcement points), a compromise at one layer cascades upward or downward. For example, a poisoned model at the detection layer can feed misleading signals to an orchestration agent, which then makes flawed resource allocation decisions that affect the entire cloud environment. This propagation effect represents a novel class of vulnerability unique to multi-agent and layered AI architectures, where the interdependency between AI components creates complex failure modes that are difficult to predict or contain with traditional stockpiled security approaches.

Architectural inference vulnerabilities. Pathade et al. (2026) conducted a far-reaching assessment of ML inference deployed in serverless cloud environments, a rapidly growing paradigm where models run in temporary, event-driven containers. Their analysis revealed a significant rise in vulnerabilities directly linked to architectural factors unique to serverless ML: shared execution environments that expose models to side-channel attacks, cold-start latency that can be weaponized for timing-based information extraction, and the transient nature of serverless functions that complicates forensic investigation after an incident. The study demonstrates that the architectural choices made to support efficient ML inference particularly the move toward serverless and function-as-a-service models introduce vulnerabilities that do not exist in traditional VM-based or containerized deployments. These findings underscore that as cloud architectures evolve to better accommodate AI workloads, the security models must evolve in parallel rather than being retrofitted after the fact.

Adversarial manipulation. Olutimehin et al. (2025), took a quantitative approach, measuring the success rates of adversarial attacks against AI systems deployed in cloud environments. Their findings confirm that adversarial examples small, often undetectable perturbations to inputs that cause model misclassification pose a serious and measurable threat to cloud-based AI systems. Critically, they evaluated multiple countermeasures and identified adversarial training as the most effective defense, significantly reducing attack success rates compared to other approaches such as input sanitization or model ensembling. This study provides empirical grounding for the theoretical risks identified in other works, demonstrating that the vulnerabilities introduced by AI are not merely hypothetical but have quantifiable success rates that demand active mitigation.

Taken together, these five studies converge on a classification of new vulnerability classes introduced by incorporating AI into cloud networks. Firstly, agent privilege vulnerabilities: overprivileged AI agents that become high-value targets for credential theft and lateral movement. Secondly, lifecycle exploitation vulnerabilities: attacks targeting specific phases of the ML pipeline, including data poisoning during training and adversarial inputs during inference. Thirdly, cascading decision-layer vulnerabilities: the propagation of compromise across interconnected AI components operating at different levels of cloud orchestration. Fourthly, architectural inference vulnerabilities: risks introduced by the serverless and ephemeral computing models adopted to support efficient ML deployment. Lastly, adversarial manipulation vulnerabilities: quantifiable risks of input-based attacks that cause AI systems to behave incorrectly without triggering traditional security alerts. The studies collectively advocate for a paradigm shift toward lifecycle-aware, behavior-based security that constrains AI authority within zero-trust frameworks, rather than treating AI components as trusted internal actors simply because they operate within the cloud perimeter.

IV. CONCLUSION

This review set out to examine AI's dual position within cloud computing security as both the instrument used to synthesize a fast-growing body of literature and the central phenomenon that literature describes. The AI-Augmented Systematic Literature Review (ASLR) approach proved effective at this dual task, using automated retrieval and LLM-based thematic synthesis to process research spanning 2020–2026 without the throughput limits of manual review, while keeping AI itself as the object under scrutiny.

Three consistent findings emerged. First, cloud infrastructure now supports resource-intensive ML not through raw computing power alone but through elastic compute matching, architectural abstraction across providers, and increasingly autonomous orchestration mechanisms illustrated respectively by Lemos et al.'s (2026) cost-performance trade offs, López García et al.'s (2020) DEEP-Hybrid-DataCloud architecture, Wei et al.'s (2025) provider-neutral policy layer, and Xu et al.'s (2022) reinforcement-learning-based resource allocation. Second, AI has become a genuinely adaptive defensive layer, moving cloud security away from static, signature-based models toward predictive, continuously learning systems as shown by LSTM-based federated firewalls (Asha V. & Kanaga S., 2026), real-time adaptive policy models (Kareem et al., 2025), and domain-specific deep learning for high-stakes healthcare environments (Dixit et al., 2026). Third, and most consequentially, the same integration that produces these defensive gains introduces new and distinct categories of risk that did not exist in traditional cloud architectures: overprivileged AI agents as high-value targets (Goel, 2026), lifecycle-specific attack surfaces spanning data poisoning and adversarial inference (Czaja et al., 2025), cascading failures across interdependent decision layers (Radanliev et al., 2026), architectural weaknesses unique to serverless ML deployment (Pathade et al., 2026), and quantifiable adversarial manipulation risks (Olutimehin et al., 2025).

These findings point to a central conclusion: AI does not simply add a defensive layer on top of existing cloud infrastructure, it reshapes the attack surface itself. Security and risk expand together rather than one displacing the other. A further limitation cutting across the defensive-layer literature is its heavy reliance on simulated rather than production data, which leaves open real questions about how these systems perform under genuine adversarial pressure at scale. Cloud security in the AI era therefore cannot treat AI components as trusted internal actors simply because they operate inside the cloud perimeter; it requires the same scrutiny, auditability, and access control historically reserved for human administrators.

Recommendations

For practitioners

- Treat AI agents as privileged principals, not utilities, apply least-privilege access, credential rotation, and full audit logging to every AI agent with cloud API access (Goel, 2026).
- Build lifecycle-specific defenses rather than a single perimeter control: data validation and provenance checks at training, adversarial-input filtering and adversarial training at inference (Czaja et al., 2025; Olutimehin et al., 2025).
- Add isolation and monitoring controls specific to serverless ML deployments, shared execution environments and cold-start behavior need devoted securities, not general-purpose cloud security tooling (Pathade et al., 2026).
- Implement cross-layer anomaly containment (e.g., circuit-breaker patterns) so a compromise at one AI decision layer cannot silently cascade into orchestration or resource-allocation layers (Radanliev et al., 2026).

For Researchers

- Prioritize validation of AI-driven defensive models (federated LSTM firewalls, adaptive policy systems) against real-world production traffic rather than simulated datasets, to close the generalizability gap identified across the defensive-layer literature.
- Develop standardized benchmarks that jointly report security effectiveness and computational cost, extending Lemos et al.'s (2026) cost-performance framing to security-specific evaluation.

For Governance and Policy

- Establish sector-specific compliance frameworks for AI-driven cloud security in regulated domains such as healthcare, where failure consequences are disproportionately severe (Dixit et al., 2026).

REFERENCES

1. Asha, V., & Kanaga Suba Raja, S. (2026). An intelligent cloud firewall framework for multi-cloud security using LSTM anomaly detection

and federated learning. Scientific Reports. https://doi.org/10.1038/s41598-026-53470-y

2. Cacciamani, G. E., Chu, T. N., Sanford, D. I., Abreu, A., Duddalwar, V., Oberai, A., Kuo, C. H. J., Liu, X., Denet, M., & Gill, I. S. (2023). PRISMA-AI: A framework for systematic reviews of artificial intelligence studies. European Urology Focus, 9(4), 567–570. https://doi.org/10.1016/j.euf.2023.04.004
3. Czaja, P., Gdowski, B., Niemiec, M., Mees, W., Stoianov, N., Votis, K., Kharchenko, V., Katos, V., & Merialdo, M. (2025). Cybersecurity challenges and opportunities of machine learning-based artificial intelligence. Neural Computing and Applications, 37, 27931–27956. https://doi.org/10.1007/s00521-025-11604-9
4. Dixit, R. S., Choudhary, S. L., Arya, N., Nathani, N., Roy, V., Bhowmik, C., & Jain, S. (2026). Artificial intelligence-powered cloud security strategies for protecting critical clinical operations in healthcare environments. Scientific Reports. https://doi.org/10.1038/s41598-026-55522-9
5. Goel, H. (2026). Security risks in tool-enabled AI agents: A systematic analysis of privileged execution environments. arXiv. https://doi.org/10.48550/arXiv.2605.09721
6. Kareem, S. A., Sachan, R. C., Malviya, R. K., Wadhwani, P., & Poddar, S. S. (2025). Next-generation cloud security: AI-powered adaptive security models for dynamic threat landscapes. In A. Swaroop, B. Virdee, S. D. Correia, & J. Valicek (Eds.), Data processing and networking: Proceedings of ICDPN 2024 (Lecture Notes in Networks and Systems, Vol. 1289). Springer. https://doi.org/10.1007/978-981-96-5535-9_27
7. Khan, T., Tian, W., Zhou, G., Ilager, S., Gong, M., & Buyya, R. (2022). Machine learning (ML)-centric resource management in cloud computing: A review and future directions. Journal of Network and Computer Applications, 204, Article 103405. https://doi.org/10.1016/j.jnca.2022.103405
8. Lemos, E., Oliveira, R., Rodrigues, J., & Oliveira Neto, R. F. (2026). Deep learning model deployment in multiple cloud providers: An exploratory study using low computing power environments. arXiv. https://arxiv.org/abs/2503.23988
9. López García, Á., Marco de Lucas, J., Antonacci, M., zu Castell, W., David, M., Hardt, M., Lloret Iglesias, L., Moltó, G., Plociennik, M., Tran, V., Alic, A. S., Caballer, M., Campos Plasencia, I., Costantini, A., Dlugolinsky, S., Duma, D. C., Donvito, G., Gomes, J., Heredia Cacha, I., & Wolniewicz, P. (2020). A cloud-based framework for machine learning workloads and applications. IEEE Access, 8, 18681–18692. https://doi.org/10.1109/ACCESS.2020.2964386
10. Olutimehin, A. T., Ajayi, A. J., Metibemu, O. C., Balogun, A. Y., Oladoyinbo, T. O., & Olaniyi, O. O. (2025). Adversarial threats to AI-driven systems: Exploring the attack surface of machine learning models and countermeasures. SSRN. https://doi.org/10.2139/ssrn.5137026
11. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
12. Pathade, C., Dhimam, V., Ahmad, S., & Lareb, I. (2026). Serverless AI security: Attack surface analysis and runtime protection mechanisms for FaaS-based machine learning. arXiv. https://doi.org/10.48550/arXiv.2601.11664
13. Radanliev, P., Santos, O., & Maple, C. (2026). Threats and vulnerabilities in artificial intelligence and agentic AI models. Frontiers in Artificial Intelligence, 9, Article 1731566. https://doi.org/10.3389/frai.2026.1731566
14. Wei, H., García Pañeda, X., & Salvachúa Rodriguez, J. (2025). Optimizing machine learning operations in multi-cloud infrastructure: A framework for unified deployment management and topology discovery. Cluster Computing, 28, Article 933. https://doi.org/10.1007/s10586-025-05584-7
15. Xu, J., Xu, Z., & Shi, B. (2022). Deep reinforcement learning based resource allocation strategy in cloud-edge computing system. Frontiers in Bioengineering and Biotechnology, 10, Article 908056. https://doi.org/10.3389/fbioe.2022.908056