Integration of NLP and ML in cloud infrastructure security

Andrian Piskozub[1,*,†], Aziz Abibulaiev[1,†]

[1] *Lviv Polytechnic National University, Stepan Bandera Str.,12, Lviv, 79000, Ukraine*

Abstract

This paper explores the potential integration of Natural Language Processing (NLP) and Machine Learning (ML) technologies in securing cloud infrastructure. With the increasing complexity of multi-cloud and hybrid environments, stricter compliance requirements, and the rise of targeted attacks, there is a growing need for adaptive intelligent systems. Such systems should provide automated threat detection, behavioral analysis, real-time response, and dynamic access control aligned with Zero Trust principles. The paper also reviews and classifies existing solutions in this domain, including NLP modules for log, service message, and query analysis, ML modules for UEBA, threat classification, and risk assessment, as well as examples from AWS, Azure, and GCP cloud services. A critical analysis is provided on the limitations of current approaches (e.g., low explainability, overfitting issues, integration challenges with DevOps/IaC). The authors propose an original architecture of an intelligent security system with combined NLP/ML modules, IaC support, and modularity. The results confirm the effectiveness of this approach compared to classical systems. The article may be useful for researchers and practitioners implementing intelligent cybersecurity strategies in dynamic cloud environments.

Keywords

Machine Learning (ML), Natural Language Processing (NLP), cloud security, zero trust, cybersecurity

1. Introduction

Modern cloud computing opens new horizons for business scalability, infrastructure cost reduction, and accelerated deployment of digital services. At the same time, the growing volume of data, increased number of access points, decentralization of resources, and the prevalence of multi-cloud architectures complicate the task of ensuring cybersecurity. Traditional protection approaches, based on static rules and signatures, do not provide the required level of adaptability, scalability, or responsiveness in dynamic cloud environments [1, 2].

Particularly relevant challenges include slow incident response, inefficient processing of large volumes of logs and unstructured data, a high rate of false positives, and the inability to detect hidden threats. Identifying unauthorized services, secret leaks, access policy violations, and suspicious activity requires deeper contextual analysis than is possible with conventional tools such as WAFs or IAM systems.

In response to these challenges, Natural Language Processing (NLP) and Machine Learning (ML) technologies are emerging at the forefront, demonstrating strong potential for building intelligent, adaptive solutions. NLP enables the automated processing of log files, configuration documents, user queries, incident reports, and other unstructured information to detect signals of threats or anomalies. ML enables user and system behavior modeling, threat classification, attack prediction, and dynamic access control within the Zero Trust architecture.

The objective of this study is to identify effective approaches for automating cloud security using NLP and ML, analyze existing solutions, and develop an original architecture of an intelligent protection system tailored to the needs of modern multi-cloud infrastructure, where we achieved incident response time (<10 minutes), fast threat detection accuracy (up to 90%), depth of contextual analysis, and auditability in comparison with modern NLP/ML solutions.

CH&CMiGIN'25: Fourth International Conference on Cyber Hygiene & Conflict Management in Global Information Networks, June 20–22, 2025, Kyiv, Ukraine

[*] Corresponding author.

[†] These authors contributed equally.

andriian.z.piskozub@lpnu.ua (A. Piskozub); aziz.r.abibulaiev@lpnu.ua (A. Abibulaiev)

0000-0002-3582-2835 (A. Piskozub); 0000-0003-3037-8373 (A. Abibulaiev)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

The object of this research is the set of processes ensuring the security of cloud environments, including log analysis, access management, and threat detection. The subject of the research comprises methods and models for applying NLP and ML in cloud security systems, as well as mechanisms for their integration with monitoring, authorization, and incident response tools.

The main tasks of the study are to: analyze current problems and limitations of classical cloud security approaches; review scientific and applied solutions using NLP and ML in cybersecurity; identify the strengths and weaknesses of existing implementations; propose an original concept of an intelligent automated protection system for cloud infrastructure; and compare the effectiveness of standard approaches and the proposed architecture based on key metrics.

2. Challenges of traditional approaches to cloud infrastructure security

The traditional cybersecurity architecture, based on signature analysis, traffic filtering, and controlled access, has historically proven effective in environments with predictable topologies and centralized computing resources. Tools in this category include Security Information and Event Management (SIEM) systems, firewalls (WAF, NGFW), Identity and Access Management (IAM) solutions, and classic mechanisms like Access Control Lists (ACLs). However, in cloud infrastructure – characterized by dynamic scalability, automated resource deployment, and flexible, temporary access policies – these tools often prove ineffective or overly restrictive.

Firstly, most traditional tools rely on predefined rules or signatures. Detecting new, atypical threats that lack clearly defined characteristics (so-called zero-day attacks) is unlikely. While standard WAF solutions may effectively block SQL injections or XSS attacks, detecting obfuscated or social engineering intrusions is only partially possible – or not at all [3]. The same applies to SIEM systems: although they collect large volumes of data, the analytics they provide are often superficial and lack the necessary context for accurate incident response.

Secondly, the fragmented control across different layers of the cloud infrastructure leads to the emergence of “blind spots.” When events occur at the intersection of multiple services – for example, between an API gateway, a serverless component, and a cloud database – traditional SIEM/WAF tools lack full visibility and are unable to correlate events correctly. As a result, multi-stage, automated attacks often go unnoticed.

The third limitation is the high frequency of false positives, which is common in static security policies. For instance, upon detecting “unusual” user activity, the system might block operations even when the request is legitimate (such as during scheduled updates or backups). This hampers productivity for DevOps/SecOps teams, frustrates users, and often forces administrators to relax security rules – consequently increasing the risk of compromise.

The fourth major drawback is delayed response time. In many systems, incident alerts arrive only after an attack has already occurred. As noted in [4], the average response time for complex incidents in traditional security systems exceeds 6–12 hours – a critical delay in always-available cloud environments, especially for multi-stage attacks involving initial infiltration, lateral movement, and data exfiltration.

Another important shortcoming is the lack of semantic analysis, which makes traditional systems incapable of interpreting events at the content level. For example, a SIEM system may log a request to a specific API but cannot determine whether critical data was requested or it was just a routine availability check. This further complicates response efforts, as the severity of the event cannot be assessed without human intervention.

In conclusion, while traditional security tools remain important as a foundational defense layer, they cannot independently provide the flexibility, speed, or contextual awareness required in modern cloud infrastructure. This creates a strong need to complement them with intelligent systems based on NLP and ML – enabling higher levels of automation, detection of complex threats, and adaptation in real time.

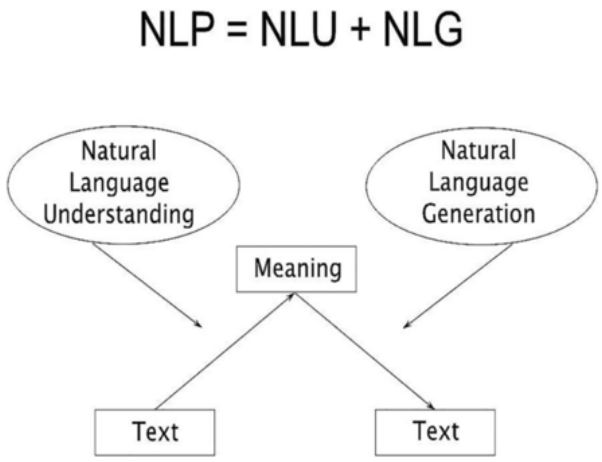


Figure 1: Understanding of NLP components.

3. Theoretical foundations of NLP and ML applications in cloud system security

The integration of NLP and ML technologies into cloud infrastructure cybersecurity opens new opportunities for delivering adaptive and intelligent protection of information resources. Unlike traditional tools based on fixed rules, NLP and ML enable the analysis of large volumes of heterogeneous data, the identification of patterns and anomalies, and the prediction of threats in dynamic environments.

The authors present the key theoretical principles underlying the use of NLP and ML in the context of cloud system protection. This includes functional mechanisms of NLP in security applications, ML-based incident detection approaches, and the types of threats these solutions can address.

3.1. Principles of NLP in cybersecurity

NLP in cybersecurity encompasses a range of algorithmic methods designed to analyze unstructured textual data for the detection of threats, anomalies, and security policy violations. The core of these approaches lies in the ability of NLP technologies to perform syntactic and semantic analysis of logs, alerts, security reports, service messages, access policies, and user-generated content [5, 6] (Figure 1).

Key tasks of NLP in the security domain include:

- Automatic classification of textual messages based on risk level;
- Detection of behavioral patterns and indicators of compromise in text logs;
- Semantic analysis of access requests to cloud resources;
- Identification of potentially dangerous instructions in documents or configurations.

A key advantage of NLP is the ability to process a large number of events in real time while accounting for context, which significantly improves threat detection accuracy and reduces false positives. In particular, models such as BERT or GPT can not only match keywords but also interpret their context within an information stream.

In many modern implementations, NLP is integrated into incident analytics automation (e.g., SOAR platforms) or log processing in SIEM systems, significantly enhancing their functionality [7]. NLP is also used to build chatbots and security virtual assistants capable of interpreting natural language queries from analysts and generating responses, thereby reducing the cognitive load on response teams [8].

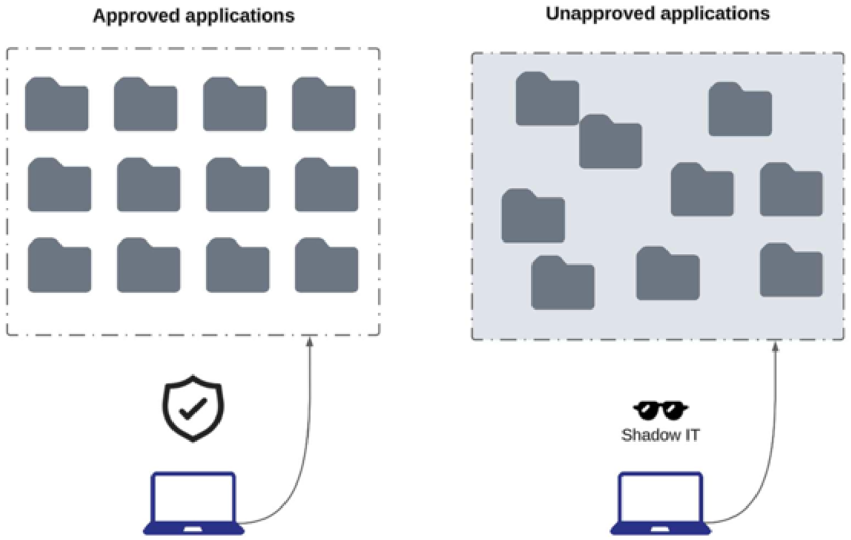


Figure 2: Applications from Shadow IT perspective.

Beyond traditional applications like log analysis, monitoring, and semantic query interpretation, NLP also offers promising capabilities in detecting signs of Shadow IT. This includes analyzing user behavior, incoming cloud service requests, and internal communications that may mention unauthorized or unregistered applications [9] (Figure 2).

Use of NLP in such scenarios enables:

- Identifying indications of unregistered services or external cloud environments;
- Detecting access policy violations disguised as normal activity;
- Automatically generating informative reports for security teams using natural language explanations;
- Improving the effectiveness of automated questionnaires and feedback forms in access and policy audits.

Another important area where NLP is actively integrated into cloud solutions is secret management. In combination with systems like HashiCorp Vault, NLP can be used to:

- Analyze textual configurations, CI/CD scripts, and IaC files (e.g., Terraform, Ansible) to detect exposed secrets, API keys, or hardcoded credentials [10];
- Contextually classify the content of environment variables, which are often sources of confidential data leaks;
- Automate the generation of access policies based on descriptive queries from security administrators;
- Integrate with chatbots that can explain secret management policies, assist in key rotation, and monitor cloud environment changes.

In current implementations, NLP is closely tied to the «Security as Code» approach [11], where all aspects of security (including access control, configuration validation, and activity monitoring) are expressed in code.

According to the authors, NLP is no longer merely a tool for reactive analysis, but an active component of the protection ecosystem–interacting with continuous deployment platforms, secret management policies, and Shadow IT restriction mechanisms. This paves the way for creating self-learning security systems that not only respond to events but also anticipate them, forming a truly intelligent approach to securing cloud systems.

3.2. Capabilities of ML in threat detection and prevention

Machine learning is a key component of modern threat detection systems in cloud infrastructure, as it enables efficient analysis of large data volumes, identification of hidden patterns, and prediction of potential attacks based on behavioral anomalies. ML algorithms provide a proactive security approach: instead of reacting to already-detected threats, the system learns to recognize threats before they occur.

The most common approaches in cloud security include classification, clustering, and reinforcement learning. Specifically:

- Classification models (e.g., Random Forest, XGBoost, SVM) are effectively used to determine the type of attack based on features extracted from network traffic or log data;
- Clustering algorithms, such as DBSCAN or k-means, help detect anomalous patterns in large volumes of user activity or telemetry data;
- Hybrid models can combine known attack signatures with the ability to learn from new incoming data, which is critical for defending against zero-day threats.

Flexibility, adaptability, and self-learning capabilities are key components of ML models form the foundation for building effective systems for data loss prevention (DLP), phishing protection, botnet activity detection, and risk assessment in multi-component cloud environments.

3.3. Types of attacks covered by NLP/ML solutions

he application of NLP and ML technologies in cloud infrastructure cybersecurity allows for effective detection and prevention of a wide range of attacks – including those that are subtle or undetectable by traditional signature-based systems.

The primary types of attacks that NLP/ML can help detect include [8]:

- Phishing attacks, identified by analyzing email and messaging content using NLP algorithms that detect social engineering patterns;
- Insider threats, where ML models detect deviations in user behavior that do not match typical profiles, enabling timely identification of internal risks;
- Zero-day attacks and unknown exploits, which ML systems can identify based on behavioral characteristics, even in the absence of known signatures;
- Unauthorized resource access attacks (e.g., privilege escalation, lateral or vertical movement)–where a combination of UEBA and NLP log analysis helps detect abnormal privilege usage patterns;
- DDoS and botnet attacks, where ML systems use traffic pattern clustering to detect anomalies and suspicious coordination of requests;
- Data exfiltration attempts, where NLP models identify attempts to transfer confidential information through unauthorized channels, especially via text (e.g., emails or chats).

Additionally, there is active development of automated security rule generation using LLM models, which are capable not only of analyzing events but also of generating appropriate access policies, SOAR playbooks, and actionable recommendations to counter specific threats [12, 13].

Because of their flexibility and contextual sensitivity, NLP and ML can cover both traditional and evolving threats–particularly in dynamic multi-cloud environments where conventional detection methods fall short in effectiveness.

4. Approaches to integrating NLP/ML in cloud infrastructure security

With the rise of multi-cloud and hybrid infrastructures, cybersecurity faces new challenges: complex cross-platform interactions, scalable threats, and the need to adapt protection policies to dynamic

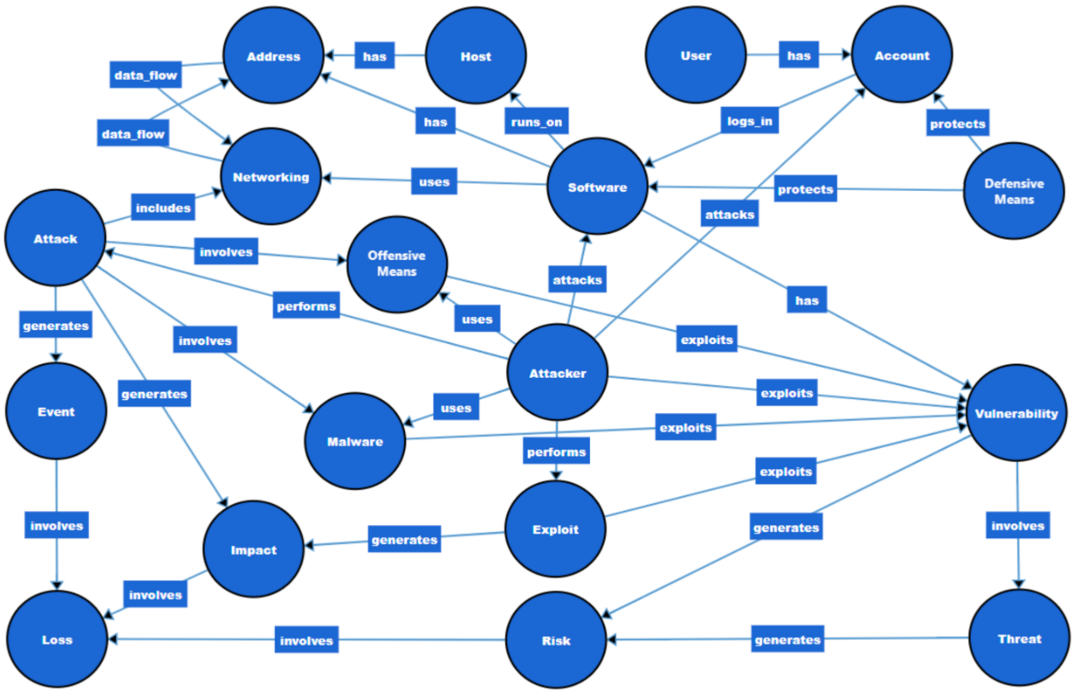


Figure 3: Ontology-based model of cybersecurity entities and their relationships.

environments in real time. In response, both the research community and industry are developing solutions based on ML and NLP technologies.

The application of NLP enables the automation of log analysis, security policy evaluation, and event interpretation. ML, in turn, offers deep user behavior analytics, adaptive attack response, incident classification, and risk assessment. Combining these technologies makes it possible to build intelligent systems capable of self-learning, threat prediction, and autonomous response. Automated threat detection is one of the most critical tasks in cloud security. Multi-cloud environments demand high-speed processing of large event volumes, making manual incident analysis infeasible. The use of NLP/ML enables early threat detection mechanisms, supporting a proactive approach to cybersecurity [14, 15].

Author [16] proposes an ontology-based knowledge representation model in the cybersecurity domain, implemented using NLP and supervised learning methods. A domain-specific ontology was developed, covering 18 core classes (e.g., Attacker, Exploit, Vulnerability, Software, Risk, etc.) and 33 types of relationships among them (exploits, performs, generates, involves, etc.) (Figure 3).

The architecture illustrates cause-effect relationships between attackers, vulnerabilities, events, and outcomes. This ontology became the basis for training named entity recognition model and relationship extraction, enabling automated semantic analysis of cybersecurity documents.

Recent research reinforces the significance of integrating NLP/ML tools into cloud infrastructure. Vakhula et al. [17] emphasized the value of the "security-as-code" approach for improving automation and dynamic policy updates in multi-cloud environments. Petrivskyi et al. [18] explored energy-efficient hybrid sensor network designs, which enhance security monitoring capabilities within cloud infrastructures. Milov et al. [19] introduced an agent-based modeling methodology to simulate antagonistic behavior in cyber systems, offering valuable insights for ML-based threat response models. In parallel, Shevchuk et al. [20] developed secure AAA service architectures, while Deineka et al. [21] proposed SOC 2-compliant classification mechanisms. Additionally, Martseniuk et al. [22] analyzed the role of centralized configuration repositories in ensuring secure and flexible infrastructure management for cloud-based services.

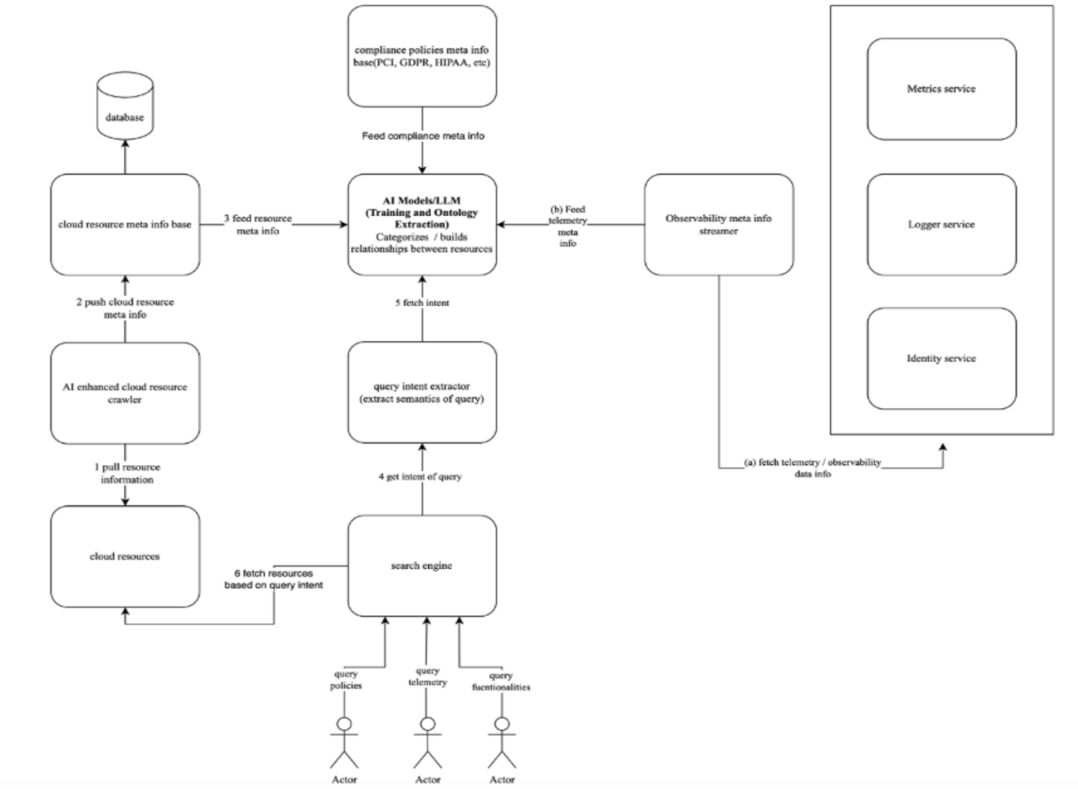

compliance policies meta info base(PCI, GDPR, HIPAA, etc)
database
Feed compliance meta info
cloud resource meta info base
3 feed resource meta info
AI Models/LLM (Training and Ontology Extraction) Categorizes / builds relationships between resources
(b) Feed telemetry meta info
Observability meta info streamer
Metrics service
Logger service
Identity service
2 push cloud resource meta info
5 fetch intent
AI enhanced cloud resource crawler
query intent extractor (extract semantics of query)
1 pull resource information
4 get intent of query
(a) fetch telemetry / observability data info
cloud resources
6 fetch resources based on query intent
search engine
query policies
query telemetry
query fucntionalities
Actor
Actor
Actor

Figure 4: Integration NLP/ML in cloud security.

4.1. Automated threat detection

In publication [23], the implementation of ML algorithms for analyzing behavioral patterns in ERP systems operating in the cloud is discussed. Classification algorithms based on historical data helped detect unusual user activity and trigger automated response procedures. At the same time, a context-aware model reduced the number of false positives by accounting for business operations.

A practical case [24] (Figure 4) demonstrates that ML-based threat detection combined with NLP enables adaptive response to continuously evolving threats. In particular – phishing attacks, data leaks, and botnet activity – ML-powered systems classify threats into multiple risk levels, facilitating automatic blocking or quarantining decisions.

Some approaches [25] focus on autonomous learning – where the model adapts to new threat types without full retraining – significantly reducing the detection delay.

It can be concluded that automated threat detection using NLP and ML significantly improves speed, accuracy, and overall security effectiveness in dynamic cloud environments. Key advantages include scalability, adaptability, fewer false positives, and reduced response time.

4.2. Data protection and DLP

Data protection in cloud environments, especially within multi-cloud architectures, requires more than just access control. It also demands mechanisms for detecting and preventing data leaks (Data Loss Prevention, DLP) [26]. According to the authors, traditional DLP systems based on signatures and manual rule configurations prove inefficient in conditions involving data mobility, replication, and encryption in the cloud. The integration of ML and NLP enables intelligent DLP solutions that are both adaptive and self-learning.

The architectural solution (Figure 5) explores AI-driven DLP strategies for multi-cloud environments in detail. Real-time automated scanning, data classification, contextual risk evaluation, and user activity

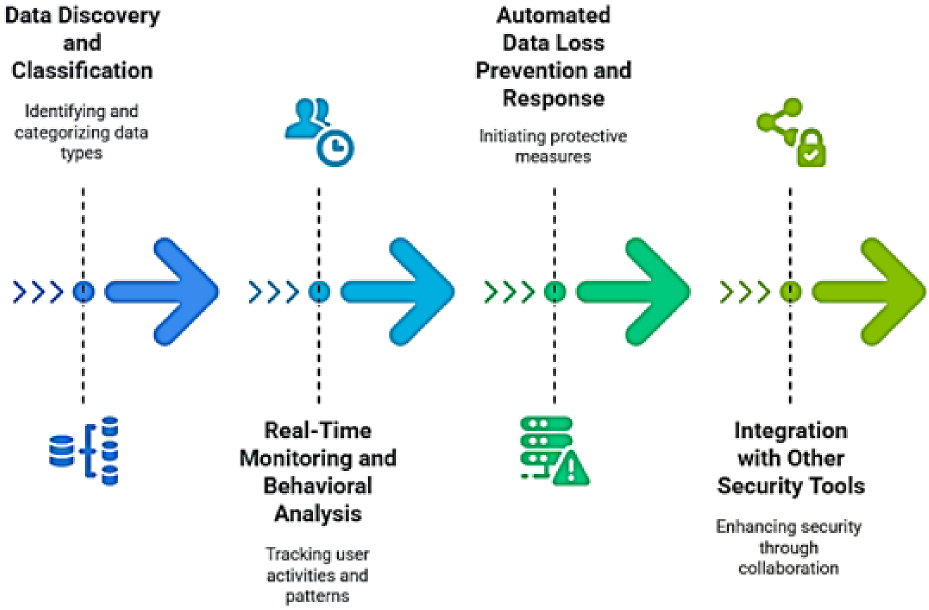

Data Discovery and Classification
Identifying and categorizing data types
Real-Time Monitoring and Behavioral Analysis
Tracking user activities and patterns
Automated Data Loss Prevention and Response
Initiating protective measures
Integration with Other Security Tools
Enhancing security through collaboration

Figure 5: Architecture decision of DLP strategy implementation, enhanced by NLP/ML [27].

monitoring are implemented. NLP algorithms identify sensitive data even in unstructured sources, such as messages, documents, or logs.

One of the key elements of such systems is phased processing: data is first identified and classified by sensitivity level, then ML models predict the likelihood of data leakage based on behavioral indicators. The NLP module detects PII (Personally Identifiable Information) in documents, while the ML module trains on prior incidents to flag risky behaviors. As a result, data leakage was reduced by over 40% compared to traditional DLP systems in a comparable environment.

Compliance with regulations such as GDPR, HIPAA, and ISO/IEC 27001 is also critical. The solution described in [28] integrates ML-based continuous monitoring mechanisms to dynamically update policies in accordance with regulatory changes.

In summary, the integration of NLP and ML into DLP modules ensures adaptive, context-aware data protection with minimal human intervention – especially important in highly dynamic cloud infrastructures.

4.3. Intelligent access control systems

Access control is a fundamental component of information security architecture, particularly in cloud environments where resources are dynamically scaled and users can access systems from anywhere in the world. Traditional access control models, such as RBAC (Role-Based Access Control) or ACLs (Access Control Lists), cannot promptly account for context, behavioral patterns, or access risk levels. As a result, intelligent access control systems integrating ML and NLP technologies are gaining popularity. NLP algorithms help detect hidden intentions in access requests and block unauthorized access to critical data (Figure 6).

Article [30] emphasizes the importance of dynamic, context-aware access control that adapts to current user conditions – their role, device, geolocation, data type, and request characteristics. These systems use ML models to construct behavioral profiles that are continuously updated based on user activity.

Because of the use of ML and NLP, intelligent access systems offer not only flexibility and contextual awareness, but also proactivity in preventing unauthorized access – an essential requirement for

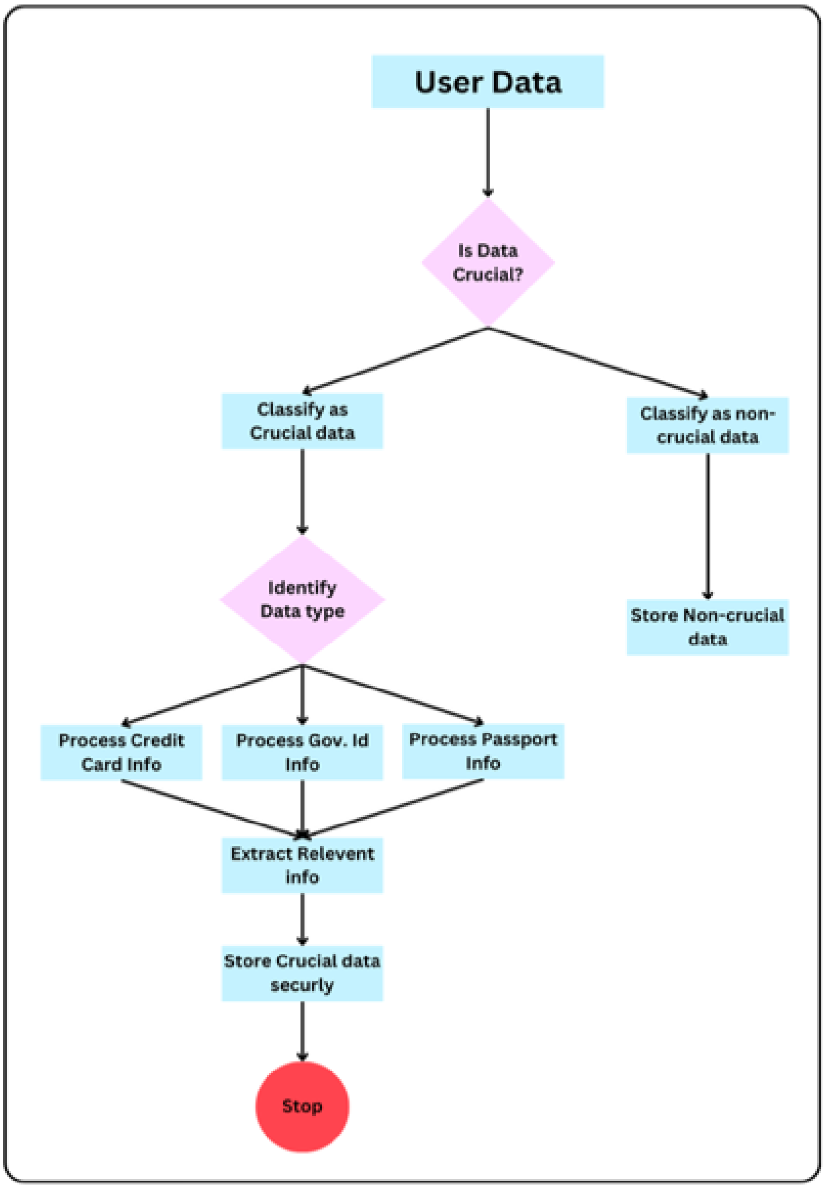


Figure 6: Preventing of storing crucial data by NLP [29].

protecting distributed cloud environments.

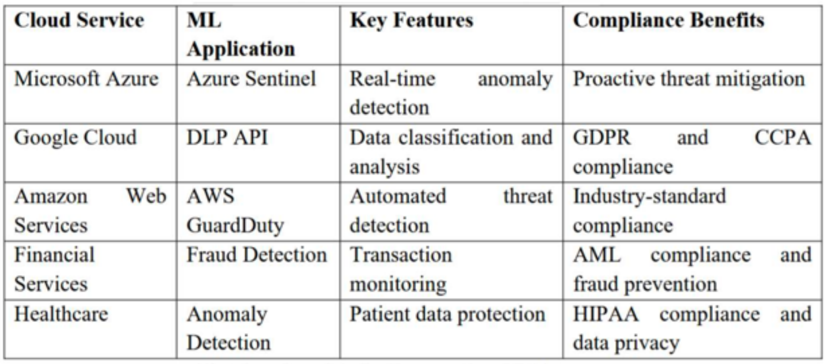
4.4. Application examples in leading companies and systems

The implementation of ML and NLP technologies in cloud infrastructure cybersecurity is no longer exceptional–it has become a strategic direction in IT security development among the world's leading companies. Real-world deployments of such approaches have shown significant success in threat detection, data protection, and automation of routine processes (Figure 7).

Amazon Web Services (AWS) has introduced several services such as Amazon Macie, which uses ML to detect PII in Amazon S3 storage, and Amazon GuardDuty, which identifies threats by analyzing event logs, network traffic, and user behavior [32].

In Google Cloud, the Chronicle platform enables behavior-based anomaly detection. It integrates large-scale data processing with ML models that analyze billions of events daily to identify patterns associated with targeted attacks.

In conclusion, leading cloud platforms and large enterprises demonstrate a wide range of use cases



Cloud Service	ML Application	Key Features	Compliance Benefits
Microsoft Azure	Azure Sentinel	Real-time anomaly detection	Proactive threat mitigation
Google Cloud	DLP API	Data classification and analysis	GDPR and CCPA compliance
Amazon Web Services	AWS GuardDuty	Automated threat detection	Industry-standard compliance
Financial Services	Fraud Detection	Transaction monitoring	AML compliance and fraud prevention
Healthcare	Anomaly Detection	Patient data protection	HIPAA compliance and data privacy

Figure 7: Comparison of ML-driven security and compliance use cases across cloud service providers [31].

for intelligent technologies that combine efficiency, scalability, and adaptability in response to modern cybersecurity challenges.

5. Analysis of problems and limitations of integrating NLP and ML in cloud infrastructure security

Despite the rapid advancement of intelligent solutions in the field of cloud security, the application of NLP and ML technologies is accompanied by a number of challenges. While many existing approaches show high performance in laboratory conditions or on synthetic datasets, in real-world environments they face issues related to scalability, adaptation to emerging threats, and compliance with ethical and legal standards. Another critical factor is the need to maintain model accuracy when training data is limited or in the case of zero-day attacks.

5.1. Model accuracy challenges

One of the main challenges in using ML and NLP for cloud infrastructure protection is ensuring consistently high model accuracy during real-world deployment. Despite impressive results on controlled datasets, models often experience performance degradation when faced with new or unpredictable data – particularly in the context of evolving cloud topologies, emerging attack types, or changing user behavior [24, 25].

Research [5] emphasizes that even well-trained models lose effectiveness when encountering non-standard log formats or textual fragments lacking keywords. Moreover, authors in [31] highlight the overfitting problem, where models perform well on known attack patterns but have low accuracy in detecting novel threats. This creates a misleading sense of effectiveness in testing environments and fails to guarantee real-world utility in production systems.

Another key challenge is class imbalance: security incident data is significantly underrepresented compared to normal activity. As a result, models may undervalue rare but critical incidents. Article [32] proposes using techniques such as oversampling, weighted training, and SMOTE to address this issue, but the authors note that these methods have limited effectiveness without quality manual tuning.

Despite substantial progress, model accuracy remains a limiting factor in deploying NLP/ML for cloud security. Moving forward, it is necessary to combine ML techniques with robust validation, adaptive learning, and interpretability mechanisms to ensure reliable and stable performance in dynamic environments.

5.2. Ethical and legal considerations

The integration of ML and NLP into cloud infrastructure cybersecurity systems introduces not only technical but also significant ethical and legal challenges. On the one hand, these technologies offer powerful tools for threat detection and protection; on the other hand, they pose risks of misuse, algorithmic bias, and violations of human rights.

From a legal perspective, particular attention must be paid to compliance with regulations such as GDPR, CCPA, and ISO/IEC 27001. Using NLP to analyze textual messages, correspondence, or logs requires strict control over privacy and the handling of personal data. According to GDPR, even partial analysis of personal information without user consent constitutes a violation – therefore, systems must implement built-in mechanisms to restrict access to PII (Personally Identifiable Information) [26].

Another critical concern is the deployment of automated decisions based on ML/NLP that affect user rights or freedoms – such as account blocking, resource isolation, or initiating defensive actions without human oversight. In such cases, regulatory standards require appeal mechanisms, decision explainability, and human intervention in the final decision-making process. It is evident that the development and implementation of intelligent cybersecurity systems must be guided not only by technical validity but also by ethical principles, transparency, data privacy, and legal compliance.

5.3. Alignment with zero trust / IAM / CIEM

The integration of intelligent systems based on NLP and ML into cloud security cannot be fully realized without alignment with modern concepts of trust and access management–specifically, Zero Trust Architecture (ZTA), Identity and Access Management (IAM), and the more dynamic Cloud Infrastructure Entitlement Management (CIEM).

According to [33], traditional IAM systems often struggle with the continuous changes in cloud environments–such as the creation of new services, temporary users, and external integrations. In such cases, CIEM provides a solution by using ML to continuously audit access rights, identify excessive privileges, and enforce least privilege policies based on behavioral data.

Alignment with Zero Trust becomes especially critical when deploying ML/NLP-based automated detection and response systems at scale. Without connection to identity verification and access control mechanisms, such solutions may become ineffective – or even dangerous –e.g., acting on unauthenticated or spoofed user requests [34].

Study [35] outlines a mechanism for integrating risk models with IAM systems. If a request's risk level – determined by ML – is high, the system initiates additional checks (e.g., MFA, or administrative approval). Clearly, ML does not operate in isolation, but in coordination with security policies. ML models analyze access history, detect anomalies, and recommend configuration changes.

Thus, to achieve high effectiveness and consistency in ML/NLP-based cybersecurity systems, it is essential to integrate them within the frameworks of Zero Trust, IAM, and CIEM – both in terms of policy enforcement and data exchange interfaces for decision-making.

6. Automation of threat response in the cloud with NLP and ML

Based on the analysis of the limitations of classical security systems, as well as the strengths and weaknesses of modern NLP/ML-based solutions discussed in the previous sections, we developed a concept for our own architecture of an intelligent cybersecurity system for cloud infrastructure. Unlike static solutions, the proposed model combines semantic analysis powered by NLP with behavioral modeling based on ML. This approach enables not only anomaly detection but also flexible automated response, scalable integration with DevOps processes, and alignment with the principles of Zero Trust and Security as Code.

It supports deep contextual analysis, scalability in multi-cloud environments, compatibility with IaC tools, and explainable decision-making. The proposed system architecture (Figure 8) implements a chain: event → analysis → decision → response. It is built around the integration of NLP modules (AWS Comprehend + Lambda with RoBERTa) and ML components (AWS SageMaker, Fraud Detector, Lambda) into a unified processing pipeline, which interacts with cloud infrastructure elements via corresponding response services (WAF, Security Groups, NACL, Route53, etc.).

Key components of the solution:

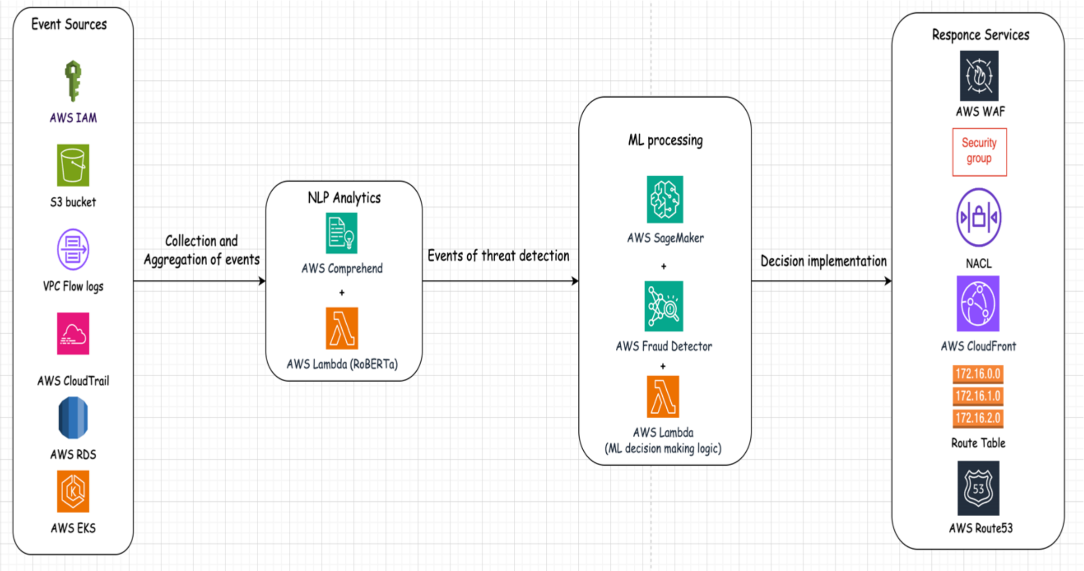


Figure 8: Proposed architecture implementation of automatic reaction on threats, based on NLP/ML.

- Event Sources: Include access log files (CloudTrail, VPC Flow Logs, WAF Logs), internal chat messages, technical tickets, CI/CD configurations, and other structured or unstructured data sources. These serve as input points for further analysis;
- NLP Analytics: Processes textual events using AWS Comprehend and a custom Lambda function powered by a pre-trained transformer model (RoBERTa). It performs request classification, detection of vulnerable patterns, and threat type categorization. The module generates a semantic threat vector representation, passed to the ML Core;
- ML Core: Analyzes the context of the event and behavioral data using a risk model. It leverages AWS SageMaker (for classification and prediction), Fraud Detector (to assess abuse probability), and custom anomaly detection modules implemented in Lambda. The output includes a risk score, threat classification, and recommended action;
- Security Reaction Layer: Automatically updates access policies, modifies WAF/Security Group/Route53 rules, and creates service events in EventBridge with updated security parameters).

7. Results analysis

A comparative evaluation was conducted to assess the effectiveness of the proposed intelligent cloud protection architecture against traditional solutions and existing NLP/ML approaches. The aim was to highlight the advantages provided by the integration of NLP and ML in threat detection, response, and access management.

The comparison was performed across several key criteria, including performance, accuracy, flexibility, explainability, and alignment with modern DevSecOps practices. Selected comparison criteria include: latency, threat detection accuracy, explainability, contextual awareness, and audit trail availability.

The results of the comparative analysis (Figure 9 and Figure 10) highlight significant advantages of the proposed architecture over classical security tools and demonstrate enhanced capabilities compared to other modern NLP/ML solutions [16, 18, 29, 34].

The proposed solution shows clear superiority in critical aspects of cloud security: performance, contextual understanding, scalability, DevOps compatibility, and decision explainability. Its implementation addresses the key limitations of traditional systems, which often rely on static rules, fragmented visibility, and manual operations.

Despite the achievements in integrating NLP/ML into cloud security, several areas remain for future research and enhancement. The next step in evolving this system is its transformation into a self-

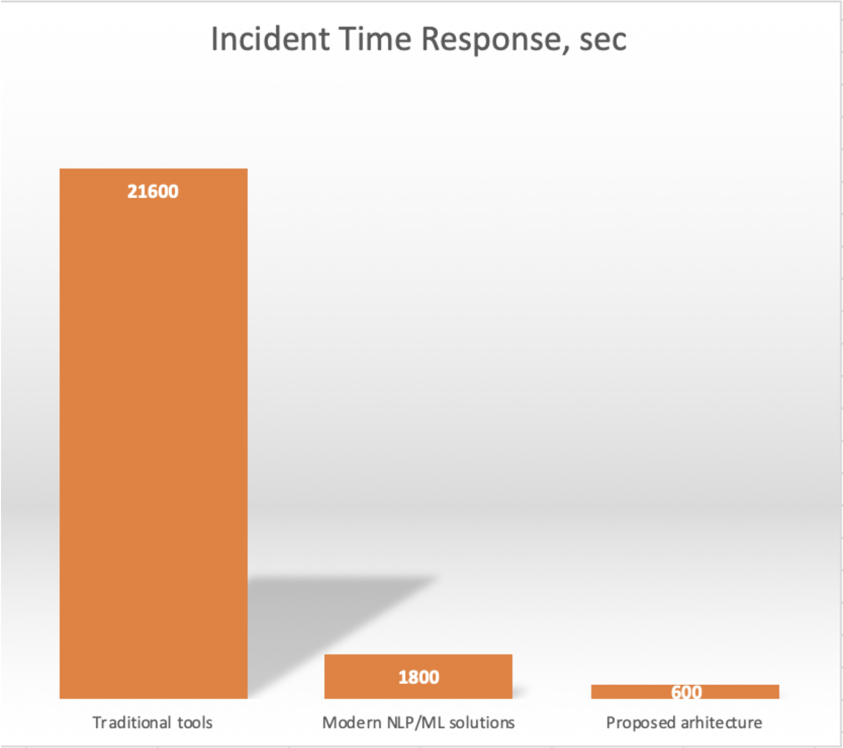


Figure 9: Comparison of incident time response ability.

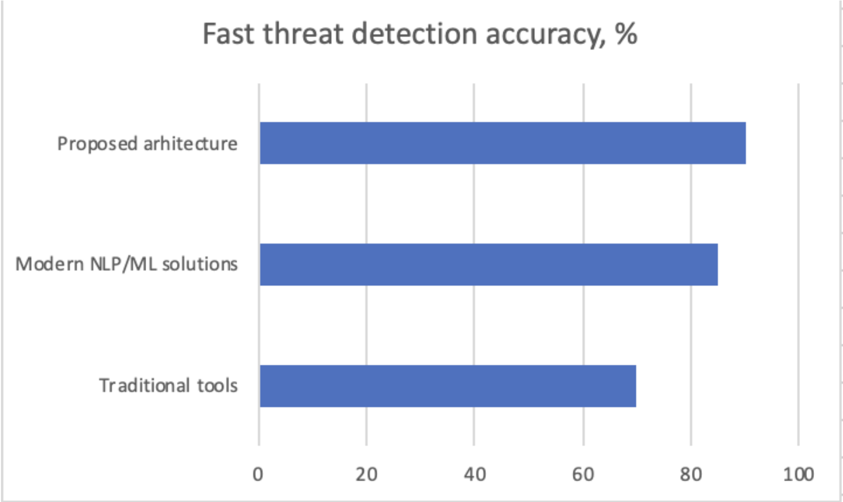


Figure 10: Comparison of the fast threat detection accuracy ability.

learning, transparent, and compliant platform capable of operating in real time across multi-cloud environments. This will empower organizations to effectively implement Zero Trust, ensure compliance and audit readiness, and reduce human error in security operations.

8. Conclusions

This paper presents a comprehensive study on the integration of Natural Language Processing (NLP) and Machine Learning (ML) technologies into the cybersecurity systems of cloud infrastructure. Based on the analysis of classical approaches, practical limitations, and modern implementations, we propose an original architecture for an intelligent security system aligned with the principles of Zero Trust, Security as Code, and DevSecOps:

1. Classical security tools, such as WAF, IAM, SIEM, and ACLs, lack the flexibility, scalability, and contextual awareness required in dynamic multi-cloud environments. Critical challenges include detecting Shadow IT, secret leaks, inter-service activity, and processing unstructured data.

2. Modern NLP and ML-based solutions (e.g., GuardDuty, Macie, Azure Sentinel, Chronicle) show clear progress but face limitations in explainability, often operate in isolation, lack full support for the DevOps lifecycle, and do not cover all data types or access scenarios.

3. The proposed architecture combines NLP analytics (AWS Comprehend + Lambda with RoBERTa) and ML modules (SageMaker, Fraud Detector, custom Lambda functions) to analyze textual messages, logs, access events, and user behavior. It includes support for XAI, CIEM, Security as Code, and automated policy updates via Infrastructure as Code (IaC).

4. Comparative evaluation demonstrated that the proposed solution outperforms traditional approaches across key metrics, including incident response time (<10 minutes), threat detection accuracy (up to 90%), depth of contextual analysis, and auditability.

The solution proposed by the authors represents a practical implementation of an intelligent, adaptive, and explainable cybersecurity architecture, suitable for deployment in today's cloud-native and highly dynamic environments. Its adoption can significantly reduce risk and response time, while delivering transparency, automation, and scalability in line with next-generation digital security requirements. Future development directions include integrating Explainable AI with natural language explanations, adopting compliance-as-code, enabling real-time dynamic access control, enhancing modularity, supporting LLM agents, and achieving full integration into multi-cloud environments.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] K. C. Sunkara, K. Narukulla, AI Enhanced Ontology Driven NLP for Intelligent Cloud Resource Query Processing Using Knowledge Graphs, 2023. doi:10.48550/arXiv.2502.18484, independent Research Report, IEEE Senior Members, Raleigh/San Jose, USA.

[2] M. M. Belal, D. M. Sundaram, Comprehensive review on intelligent security defences in cloud: Taxonomy, security issues, ML/DL techniques, challenges and future trends, Journal of King Saud University - Computer and Information Sciences (2022). doi:10.1016/j.jksuci.2022.08.035.

[3] J. Jaya, Application of deep learning in cloud security, in: Deep Learning Approaches to Cloud Security, Wiley, 2022. doi:10.1002/9781119760542.ch12.

[4] P. Nina, K. Ethan, AI-driven threat detection: Enhancing cloud security with cutting-edge technologies, International Journal of Trend in Scientific Research and Development 4 (2019) 1362–1374. URL: https://www.ijtsrd.com/papers/ijtsrd29520.pdf.

[5] J. S. Nimbhorkar, AI Enabled Cloud RAN Test Automation: Automatic Test Case Prediction Using Natural Language Processing and Machine Learning Techniques, M.sc. thesis, KTH Royal Institute of Technology, 2023. URL: urn:nbn:se:kth:diva-340090, ericsson AB, Stockholm.

[6] Z. Kilhoffer, M. Bashir, Cloud privacy beyond legal compliance: An NLP analysis of certifiable privacy and security standards, in: IEEE Cloud Summit, Washington, DC, USA, 2024, pp. 79–86. doi:10.1109/Cloud-Summit61220.2024.00020.

[7] K. C. Sunkara, K. Narukulla, AI enhanced ontology driven NLP for intelligent cloud resource query processing using knowledge graphs, 2025. doi:10.48550/arXiv.2502.18484.

[8] S. R. Mamidi, The role of AI and machine learning in enhancing cloud security, Journal of Artificial Intelligence General Science (JAIGS) (2024). doi:10.60087/jaigs.v3i1.161.

[9] Y. Martseniuk, A. Partyka, O. Harasymchuk, E. Nyemkova, M. Karpinski, Shadow IT risk analysis in public cloud infrastructure, in: CEUR Workshop Proceedings, volume 3800, 2024, pp. 22–31.

[10] Y. Martseniuk, A. Partyka, O. Harasymchuk, S. Shevchenko, Universal centralized secret data management for automated public cloud provisioning, in: CEUR Workshop Proceedings, volume 3826, 2024, pp. 72–81.

[11] V. Khoma, A. Abibulaiev, A. Piskozub, T. Kret, Comprehensive approach for developing an enterprise cloud infrastructure, in: CEUR Workshop Proceedings, volume 3654, 2024, pp. 201–215.

[12] D. M. Rakgoale, H. I. Kobo, Z. Z. Mapundu, T. N. Khosa, A review of AI/ML algorithms for security enhancement in cloud computing with emphasis on artificial neural networks, in: 4th International Multidisciplinary Information Technology and Engineering Conference (IMITEC), Vanderbijlpark, South Africa, 2024, pp. 329–336. doi:10.1109/IMITEC60221.2024.10851076.

[13] J. Al-Azzeh, M. A. Hadidi, R. Odarchenko, S. Gnatyuk, Z. Shevchuk, Z. Hu, Analysis of self-similar traffic models in computer networks. international review on modelling and simulations, International Journal of Computer Network and Information Security 10 (2017) 328–336. doi:10.15866/iremos.v10i5.12009.

[14] N. D. V. Talati, Scalable AI and data processing strategies for hybrid cloud environments, World Journal of Advanced Research and Reviews 10 (2021) 482–492. doi:10.30574/wjarr.2021.10.3.0289.

[15] Z. Hu, Y. Khokhlachova, V. Sydorenko, I. Opirskyy, Method for optimization of information security systems behavior under conditions of influences, International Journal of Intelligent Systems and Applications 9 (2017) 46–58. doi:10.5815/ijisa.2017.12.05.

[16] T.-M. Georgescu, Natural language processing model for automatic analysis of cybersecurity-related documents, Symmetry 12 (2020) 354. doi:10.3390/sym12030354.

[17] O. Vakhula, I. Opirskyy, O. Mykhaylova, Research on security challenges in cloud environments and solutions based on the "security-as-code" approach, in: CEUR Workshop Proceedings, volume 3550, 2023, pp. 55–69.

[18] V. Petrivskyi, O. Melnyk, S. Petrenko, O. Kot, Development of a modification of the method for constructing energy-efficient sensor networks using static and dynamic sensors, Eastern-European Journal of Enterprise Technologies 1 (2022) 15–23. doi:10.15587/1729-4061.2022.252988.

[19] O. Milov, S. Yevseiev, D. Bodnar, I. Opirskyy, Development of methodology for modeling the interaction of antagonistic agents in cybersecurity systems, Eastern-European Journal of Enterprise Technologies 2 (2019) 56–66. doi:10.15587/1729-4061.2019.164730.

[20] D. Shevchuk, O. Harasymchuk, A. Partyka, N. Korshun, Designing secured services for authentication, authorization, and accounting of users, in: CEUR Workshop Proceedings, volume 3550, 2023, pp. 217–225.

[21] O. Deineka, O. Harasymchuk, A. Partyka, A. Obshta, N. Korshun, Designing data classification and secure store policy according to SOC 2 type II, in: CEUR Workshop Proceedings, volume 3654, 2024, pp. 398–409.

[22] Y. Martseniuk, A. Partyka, O. Harasymchuk, V. Cherevyk, N. Dovzhenko, Research of the centralized configuration repository efficiency for secure cloud service infrastructure management, in: CEUR Workshop Proceedings, volume 3991, 2025, pp. 260–274.

[23] R. K. Jha, Strengthening smart grid cybersecurity: An in-depth investigation into the fusion of machine learning and natural language processing, Journal of Trends in Computer Science and Smart Technology 5 (2023) 284–301. doi:10.36548/jtcsst.2023.3.005.

[24] J. Wang, AI/ML-powered cybersecurity and cloud computing strategies for optimized business intelligence in ERP cloud, ResearchGate, 2023. doi:10.13140/RG.2.2.27926.66882.

[25] T. K. Vashishth, V. Sharma, B. Kumar, S. Chaudhary, R. Panwar, Enhancing cloud security: The role of artificial intelligence and machine learning, in: IGI Global, 2024. doi:10.4018/979-8-3693-1431-9.ch004.

[26] R. Muppalaneni, A. C. Inaganti, N. Ravichandran, Ai-enhanced data loss prevention (dlp) strategies for multi-cloud environments, Journal of Computing Innovations and Applications 2 (2024) 1–13. URL: https://ciajournal.com/index.php/jcia/article/view/9, accessed: 10 May 2025.

[27] P. Van Hau, Enhancing web application security: A deep learning and NLP-based approach for accurate attack detection, Journal of Science and Technology on Information Security (2023).

doi:10.54654/isj.v3i20.1008.

[28] H. Aldawsari, S. A. Kouchay, Integrating AI and machine learning algorithms in cloud security frameworks for enhanced proactive threat detection and mitigation, Journal of Engineering Technology and Management 74 (2024). URL: https://ciajournal.com/index.php/jcia/article/view/9, accessed: 11 May 2025.

[29] A. M. Buttar, F. Shahzad, U. Jamil, Conversational AI: Security features, applications, and future scope at cloud platform, in: Conversational Artificial Intelligence, 2024. doi:10.1002/9781394200801.ch3.

[30] M. R. Al Saidat, S. Y. Yerima, K. Shaalan, Advancements of SMS spam detection: A comprehensive survey of nlp and ml techniques, Procedia Computer Science (2024). doi:10.1016/j.procs.2024.10.198.

[31] J. N. Malaiyappan, S. Prakash, S. V. Bayani, M. Devan, Enhancing cloud compliance: A machine learning approach, AIJMR - Advanced International Journal of Multidisciplinary Research (2024). doi:10.62127/aijmr.2024.v02i02.1036.

[32] A. B. Haque, A. K. M. N. Islam, S. Hyrynsalmi, B. Naqvi, K. Smolander, GDPR compliant blockchains – a systematic literature review, IEEE Access 9 (2021) 50593–50606. doi:10.1109/ACCESS.2021.3069877.

[33] A. Chava, Application security and least privilege access in modern DevOps, The American Journal of Engineering and Technology (2024). doi:10.37547/tajet/Volume06Issue10-09.

[34] R. Komala, A. K. BR, M. Prasad, A. Shreyas, Smart governance among smart cities for legal consideration to international data migration in cloud using machine learning, nlp and blockchain smart contract, Preprints (2024). doi:10.20944/preprints202408.1028.v1.

[35] A. Manoharan, M. Sarker, Revolutionizing cybersecurity: Unleashing the power of artificial intelligence and machine learning for next-generation threat detection, International Research Journal of Modern Engineering and Technology and Science (IRJMETS) (2024). doi:10.56726/irjmets32644.