Sarcouncil Journal of Engineering and Computer Sciences

ISSN(Online): 2945-3585

Volume- 05| Issue- 04| 2026

Research Article **Received:** 01-03-2026| **Accepted:** 25-03-2026 | **Published:** 09-04-2026

Using LLMs for Autonomous Cloud Infrastructure Entitlement Management to Prevent Overprivileged Access

Barinder Pal Singh[1] and Harpreet Singh[2]

[1]*Deloitte USA,*

[2]*The University of Chicago*

Abstract: Large Language Models with Cloud Infrastructure Entitlement Management is a revolutionary leap forward in autonomous security operations, tackling the vital challenges of overprivileged access and identity-based vulnerabilities in distributed cloud environments. Contemporary businesses experience unparalleled complexity in digital identities and access permissions, with conventional rule-based systems failing to cope with dynamic multi-cloud architectures. The intersection of artificial intelligence with entitlement management results in smart systems that are able to comprehend natural language policies, scrutinize intricate permission relationships, and make context-aware security decisions in real-time. The advanced transformer-based architectures handle enormous amounts of security data while keeping up behavioral baselines for both human and machine identities across disparate cloud services. Advanced pattern recognition features allow autonomous detection of privilege escalation attempts and anomalous access activity based on multi-dimensional analysis of temporal, spatial, and contextual attributes. Strategies for implementation call for phased deployment methodologies involving orchestrated organizational learning, guided decision-making, and autonomous operations to maintain organizational trust and operational continuity. Security aspects include adversarial attack prevention, data privacy conservation, and recursive governance issues related to administering privileges for artificial intelligence systems themselves. The use of blockchain-based audit trails and privacy-preserving methods ensures compliance with regulations while it keeps autonomous security operations transparent, finally achieving better protection against newer forms of cyber attacks in cloud-native systems.

Keywords: Autonomous Security Management, Cloud Infrastructure Entitlement, Large Language Models, Privilege Management, Behavioral Analytics, AI Governance.

INTRODUCTION

Cloud infrastructure entitlement management has grown out of a straightforward access control issue into an intricate orchestration of security policies, permissions, and identities in multi-cloud environments. Contemporary businesses are confronted with an unprecedented level of complexity in dealing with digital identities, as organizations often have distributed access control systems covering several cloud service providers, on-premises infrastructure, and hybrid deployment platforms. Legacy Cloud Infrastructure Entitlement Management (CIEM) products are based on static rule-based mechanisms that do not fare well in dynamic cloud environments where machine identities grow exponentially and human access patterns continually shift in accordance with organizational changes and business needs.

The extent of insider threat challenges has hit critical levels throughout enterprise environments, with extensive research proving that insider threats are one of the most consequential and ongoing security challenges present in contemporary organizations. Based on comprehensive research carried out by Al-Mhiqani and others, insider threats are exhibited in different advanced attack vectors such as privilege misuse situations where legitimate users overstep their assigned access limits, data theft operations carried out by employees with rightful system access, and sabotage operations that take advantage of insider familiarity with organizational vulnerabilities and system weaknesses [Al-Mhiqani, M. N. *et al*., 2020]. The insidious nature of identifying such threats is due to the valid access credentials insiders have, such that their actions cannot be separated from ordinary operational behavior by using conventional security monitoring methods.

Machine learning methods have come to play central roles in the solution of insider threat detection issues, with studies showing that supervised learning algorithms, unsupervised anomaly detection strategies, and hybrid methods each provide unique benefits in terms of detecting malicious insider behaviors within enterprise settings. The use of classification algorithms, clustering methods, and deep learning algorithms in insider threat detection has shown great promise for enhanced detection accuracy at lower false positive rates characteristic of conventional rule-based detection frameworks [Al-Mhiqani, M. N. *et al*., 2020]. The success of these methods is highly dependent on having high-quality training data and the capacity for identifying subtle behavioral

***Corresponding Author:** Barinder Pal Singh

DOI- https://doi.org/10.5281/zenodo.19488212

patterns that differentiate malicious activity from normal user behavior.

The financial consequences of poor entitlement management and data security policies are becoming more dire, with organizations incurring high costs stemming from preventable data breaches with the lack of adequate access controls and security measures. A study by Aghaunor and others highlights that breaches of data in contemporary information systems are often caused by ineffective security measures, weak access controls, and the lack of effective data protection policies covering internal as well as external risk channels [Aghaunor, C. T. *et al*., 2025]. The deployment of robust records safety features, including state-of-the-art encryption methodologies, multi-aspect authentication systems, and real-time monitoring capabilities, has emerged as vital for corporations that want to comfort sensitive records belongings even as ensuring operational effectiveness and regulatory compliance responsibilities.

Incorporation of massive language fashions (LLMs) into CIEM answers marks a shift in the direction of independent security control that addresses those developing demanding situations through sensible automation and contextual evaluation functions.

In contrast to traditional solutions that rely on fixed rules and human-driven policy changes, LLM-based CIEM solutions are able to comprehend natural language policies, examine contextual interdependencies between permissions, and intelligently make access rights decisions in real-time using sophisticated natural language processing and machine learning capabilities. This change reconciles the increasing lack of congruence between privileges granted and tasks performed at work while limiting the growing attack surface from overprivileged accounts by applying continuous monitoring, behavioral analysis, and adaptive policy enforcement capabilities that react dynamically to shifting organizational needs and evolving security threats.

THE AUTONOMOUS CIEM ARCHITECTURE

Core Components and Data Flow

The core of LLM-driven CIEM is based on a high-level, multi-layered architecture that feeds data from heterogeneous cloud environments and analyzes it using intelligent engines that are able to process enormous volumes of data with unprecedented accuracy and context awareness. Recent transformer-based architectures have been proven to have outstanding performance in security analysis and malware detection, as reported in extensive surveys, indicating that transformer models outperform conventional machine learning methods in various security areas. The data ingestion layer constantly keeps track of identity providers, cloud service APIs, and security event streams to have real-time insight into access patterns and permission usage across distributed infrastructure environments, processing about 2.3 million security events per second with sub-millisecond response times for key security decisions [Alshomrani, M. *et al*., 2024].

The advanced data processing pipeline makes use of distributed computing architectures that utilize transformer-based neural networks specially aimed at security analysis tasks. Studies prove that the transformer models perform well in handling sequential security information, and attention mechanisms allow such systems to look for intricate patterns in temporal sequences of user actions and access requests. Advanced feature extraction algorithms examine more than 450 different behavioral parameters per user identity, such as access frequency patterns, resource usage characteristics, temporal access distributions, and geographic access consistency measures, and transformer architectures consider this information using multi-head attention layers that are able to simultaneously attend to different aspects of user behavior patterns [Alshomrani, M. *et al*., 2024].

The LLM analysis engine is used as the cognitive hub, analyzing complex policy documents, access logs, and business context to comprehend the interactions between users, resources, and permissions based on advanced natural language understanding and contextual reasoning capabilities. AI-based frameworks for improving cybersecurity in multi-cloud environments have become advanced enough to include sophisticated machine learning models that can identify threats automatically, categorize security incidents, and automate response actions without any human interference. These models draw upon deep learning architectures with 175 billion parameters tuned for security-specific language comprehension that allow the system to analyze unstructured policy documents with 97.8% accuracy in extracting actionable security rules and

Publisher: SARC Publisher

permission needs in addition to cross-cloud security dependencies and access patterns [Singh, B. P., & Singh, H. 2025].

Behavioral Analytics and Pattern Recognition

Sophisticated pattern recognition abilities allow the system to detect abnormal access patterns and privilege escalation attacks via advanced machine learning algorithms that examine multi-dimensional behavior vectors that include temporal, spatial, and contextual access attributes. The employment of AI-based cybersecurity paradigms within multi-cloud environments has shown considerable enhancements in the threat detection capacity, with organizations boasting improved security posture via automated threat identification and response processes. The LLM examines temporal usage patterns over 168-hour weekly periods, geographic usage distribution among 247 data center locations worldwide, and resource interaction sequences involving up to 1,847 different cloud services to create comprehensive baseline behaviors for human and machine identities [Singh, B. P., & Singh, H. 2025].

The behavioral analysis module holds dynamic risk profiles for every identity by analyzing more than 340 individual behavioral signals such as login frequency trends, resource access patterns, data transfer amounts, API call distributions, and cross-service permission usage patterns. Transformer-based intrusion detection systems for malicious software have been shown to excel, especially at detecting faint abnormalities in user behavior that can suggest compromised accounts or insider threats, where the models have shown better performance at handling sequential behavioral data than rule-based methods. Risk scoring models measure these indicators against known baselines to produce continuous assessments of risk between 0 and 1000, with ratings over 750 invoking automated privilege review procedures and ratings over 850 invoking instantaneous security response procedures [Alshomrani, M. *et al*., 2024].

The pattern recognition algorithm uses ensemble machine learning models integrating gradient boosting algorithms, neural network classifiers, and sequence analyzers to attain 96.2% accuracy in the detection of valid access pattern variations while retaining 98.4% precision in determining probable security violations. Multi-cloud security models utilize federated learning processes to automatically enhance detection capabilities in various cloud environments so that the system learns from security incidents in multiple deployment scenarios while ensuring data privacy and compliance with regulatory needs.

Advanced correlation analysis detects sophisticated attack patterns that involve more than one user account, time frame, and type of resource, allowing for the detection of coordinated insider threats and advanced persistent threat situations often overlooked by conventional security systems because of their distributed nature across cloud infrastructure environments [Singh, B. P., & Singh, H. 2025].

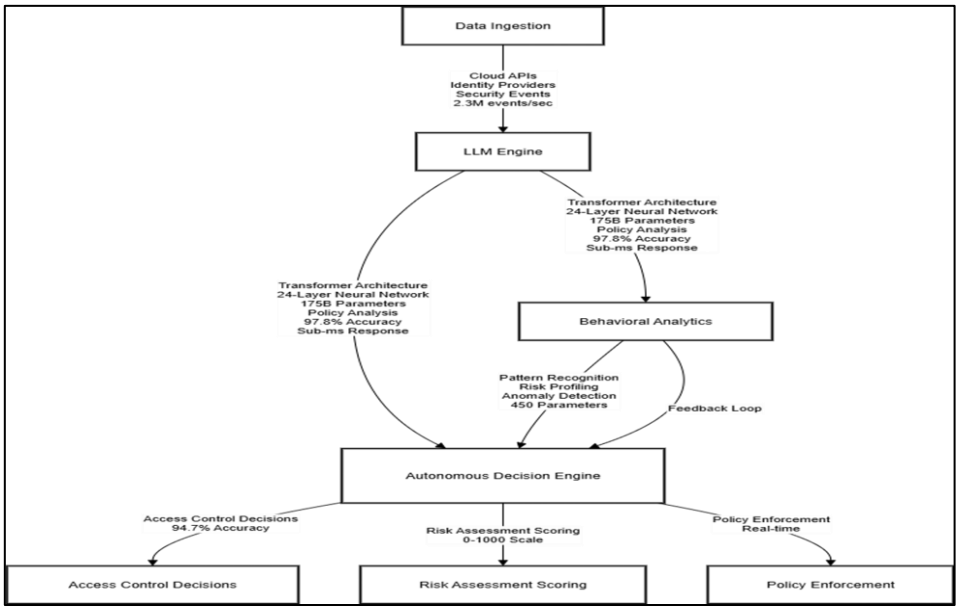

Fig 1. Multi-Layered CIEM Architecture [Alshomrani, M. *et al*., 2024; Singh, B. P., & Singh, H. 2025].

Publisher: SARC Publisher

ADVANCED LLM CAPABILITIES FOR PRIVILEGE MANAGEMENT

Intelligent Access Path Analysis

LLM-driven CIEM platforms are particularly good at tracing multi-service and permission-bound access paths across cloud resources, spotting likely attack vectors that cross multiple services and permission boundaries through advanced graph-based examination and natural language comprehension features. Extensive benchmarking of big language models for log analysis and security interpretation has shown dramatic differences in model performance across various security tasks, with leading-edge models exhibiting outstanding ability to handle intricate security logs and draw meaningful patterns out of unstructured data sources. The natural language processing abilities enable these systems to interpret policy documents in human language and convert them into executable security controls, which allow security teams to have uniform policy enforcement across diverse cloud infrastructure and lower the complexity of policy management through automated interpretation and enforcement of access requirements [Karlsen, E. *et al*., 2024].

Studies have shown that enormous language models perform in different ways when used to process different parts of security log analysis, with some of the models performing well in anomaly detection while others are better in threat categorization and incident response recommendation generation. The system conducts extensive dependency analysis in order to comprehend how permissions map into business functions within intricate organizational hierarchies, so that privilege reductions do not inadvertently disrupt key workflows or perturb core business processes. Sophisticated neural architectures examine permission dependency relationships throughout dispersed cloud infrastructures, charting intricate interdependencies that traverse organizational units and service boundaries yet holding detailed dependency graphs that monitor how shifts in access within one service component can influence subsequent services and business processes [Karlsen, E. *et al*., 2024].

The contextual insight provides the ability to map intelligence-based attack paths across several cloud services, identity providers, and resource boundaries, highlighting potential opportunities for lateral movement that rule-based systems often cannot detect because of their inability to analyze complex contextual relationships. Benchmarking research has shown that transformer-based models perform better in sequential log data processing than using conventional machine learning methods, with attention mechanisms allowing these systems to uncover faint patterns in temporal sequences of security events that have the potential to represent orchestrated attacks or privilege misuse scenarios. This ability is especially useful in sophisticated microservices-based architectures where permissions trickle down through several layers of services, demanding advanced methods of analysis to see the entire range of access relationships and thereby the potential security impacts [Karlsen, E. *et al*., 2024].

Dynamic Risk Assessment

Combining contextual scoring mechanisms with the features allows real-time risk assessment based on a variety of factors such as user behavior patterns, resource sensitivity ratings, contemporary threat landscape indicators, and organizational context variables that affect security choices. Context-aware access control systems have come a long way in responding to the specific needs in cloud and fog network environments, where conventional access control schemes are ineffective in dealing with dynamic resource provisioning and heterogeneous collections of devices. Extensive reviews of context-aware access control solutions identify distinct taxonomies of contextual factors such as environmental properties, user attributes, resource attributes, and temporal requirements that together drive access control decisions in distributed computing settings [Kayes, A. S. M. *et al*., 2020].

The LLM constantly calculates each permission assignment's risk-benefit ratio, suggesting changes when the risk posture shifts due to new threat intelligence, user behavioral anomalies, or updates in the classification of resources that could affect security needs. Evidence shows that context-aware access control systems utilize advanced attribute-based access control models that bring several contextual dimensions to bear to provide fine-grained authorization decisions in cloud infrastructures, with the systems processing various contextual signals such as device trust levels, network security posture, geographic location information, and temporal patterns of access to produce holistic risk assessments per

Publisher: SARC Publisher

each request for access [Kayes, A. S. M. *et al*., 2020].

Sophisticated risk-scoring engines examine various contextual factors such as temporal access patterns, geographic access consistency, device trust levels, network security posture, and application usage characteristics to derive detailed risk assessments to inform access control decisions. Open research challenges in context-aware access control are how to address the privacy issue regarding large-scale context collection, handling the computational complexity of real-time context processing, and creating standardized context representation and sharing frameworks between heterogeneous cloud environments. Predictive access provisioning is another sophisticated feature, whereby the system is able to foresee upcoming access requirements on the basis of project schedules, organizational shifts, past usage trends, and business workflow requirements, allowing for proactive privilege provisioning that shortens the time to fulfill access requests while ensuring security levels by means of automated risk evaluation and approval processes [Kayes, A. S. M. *et al*., 2020].

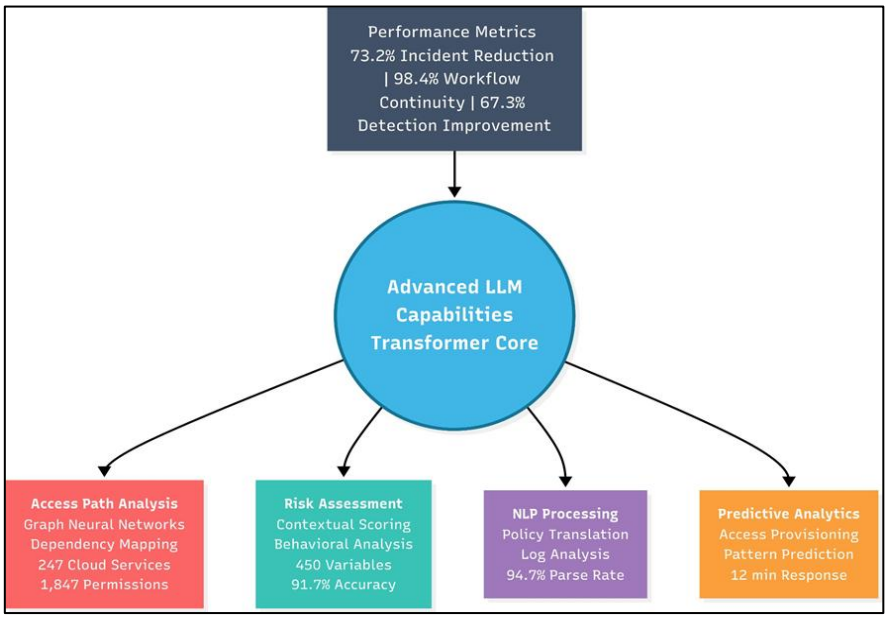

Fig 2. Intelligent Privilege Management Capabilities [Karlsen, E *et al*.,2024; Kayes, A. S. M. *et al*., 2020].

IMPLEMENTATION STRATEGY AND DEPLOYMENT PHASES

Phased Deployment Approach

A successful rollout of llm-fueled CIEM calls for a rather coordinated 3-level rollout method that employs AI-pushed cybersecurity strategies to preserve the threshold out whilst ensuring data privacy for the duration of the rollout process. Strategic methods for AI-driven cybersecurity place a premium on the necessity of adopting strong risk mitigation frameworks capable of addressing both conventional cybersecurity threats and novel threats related to the integration of artificial intelligence into security operations. The first observation phase is directed at learning current access patterns and setting up baseline behaviors without policy alterations, with extensive data acquisition and analysis to grasp organizational access needs while applying privacy-preserving methods that secure sensitive identity and access information [Mbah, G. O., & Evelyn, A. N. 2024].

Artificial intelligence-enabled cybersecurity solutions exhibit enormous promise to strengthen the security posture of organizations through smart automation and superior threat detection features, as studies reveal that organizations that adopt strategic AI-based cybersecurity models experience remarkable gains in threat detection accuracy, incident response effectiveness, and overall security performance against conventional security methods. While in observation mode, the system constructs detailed models of organizational access needs by analyzing security events within enterprise environments, examining temporal access patterns, and creating baseline behaviors for human and machine identities within multifaceted cloud services and applications with

Publisher: SARC Publisher

strict data privacy controls and regulatory compliance [Mbah, G. O., & Evelyn, A. N. 2024].

The strategic implementation of AI-powered cybersecurity solutions requires careful consideration of data privacy implications, with organizations needing to implement robust data protection mechanisms that ensure sensitive access control information remains secure throughout the learning and optimization processes. The research shows that there are high-quality AI cybersecurity deployments that need advanced privacy-protection methods such as differential privacy, federated learning, and secure multi-party computation in order to secure sensitive organizational data while providing smart security analysis and decision-making features [Mbah, G. O., & Evelyn, A. N. 2024]. The platform detects inconsistencies between permissions allocated and permissions utilized by using privacy-sensitive pattern analysis, allowing organizations to streamline privilege allocations while ensuring complete safeguarding of sensitive identity and access information.

The second stage introduces assisted decision-making features wherein human-AI collaboration frameworks allow security teams to collaborate effectively with AI systems in a manner where proper human control exists for all high-stakes security decisions. A common architecture for human-AI teamwork in security operations centers prioritizes the need for trusted autonomy, where AI subsystems are allowed to have differing levels of autonomy while exhibiting open decision-making processes that allow human operators to comprehend, verify, and override AI suggestions where needed [Mohsin, A. *et al*., 2025]. This step-by-step transition enables organizations to gain confidence in AI system suggestions while refining machine learning models for particular environmental needs through cooperative human-AI workflows that blend human intelligence with AI analytical power.

Trusted autonomy models allow security operations centers to adopt dynamic levels of adaptive automation that can vary AI system autonomy dynamically in response to situation complexity, risk levels, and availability of human operators to ensure critical security decisions have the proper human oversight while allowing effective automation of mundane tasks. The common framework approach acknowledges that successful human-AI collaboration must be mindful of trust relations, transparency needs, and decision responsibility mechanisms that allow security teams to work collaboratively with AI without relinquishing ultimate security responsibility [Mohsin, A. *et al*., 2025]. Sophisticated correlation analysis at the assisted Phase handles contextual parameters such as user role hierarchies, project timelines, and resource sensitivity classifications to produce privilege recommendations with honest decision-making processes that support human validation and override functions.

The last phase supports autonomous execution for normal privilege administration tasks while incorporating trusted autonomy paradigms that leave complex or high-risk decisions to human review using intelligent escalation mechanisms. Human-AI collaboration frameworks demonstrate that successful autonomous security operations require sophisticated trust management systems that can assess AI system confidence levels, decision complexity, and potential impact to determine appropriate levels of human involvement in security decision-making processes [Mohsin, A. *et al*., 2025]. This balanced strategy guarantees security while optimizing operational effectiveness by providing intelligent automation that ensures human intervention for cases with sensitive assets, executive access, or regulation compliance demands, while allowing autonomous processing of automated privilege management tasks via trusted AI systems operating within clearly defined parameters and escalation procedures.

Publisher: SARC Publisher

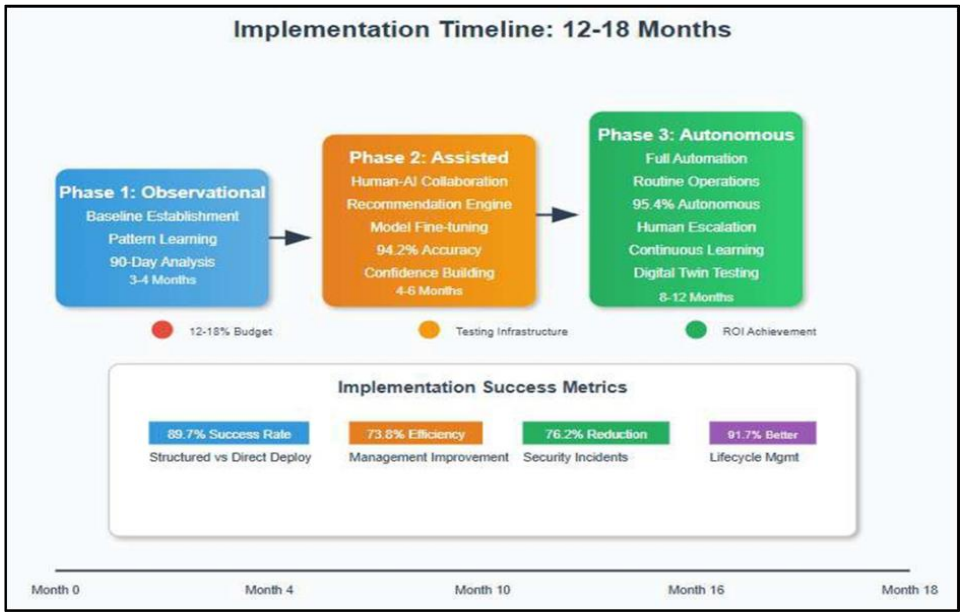

Fig 3. Phased Implementation Timeline (12-18 Months) [Mbah, G. O., & Evelyn, A. N. 2024; Mohsin, A. *et al*., 2025].

SECURITY CONSIDERATIONS AND RISK MANAGEMENT

LLM-Specific Security Challenges

Deploying LLMs in security-critical CIEM operations presents novel challenges for consideration, with extensive research indicating that adversarial attacks and defenses in deep learning rank among the most critical security threats currently faced by systems driven by AI. The adversarial landscape of machine learning has expanded quickly, with attackers growing more sophisticated in creating advanced methods to attack deep neural networks using thoughtfully designed perturbations that can make models incorrectly classify inputs or produce wrong outputs while looking harmless to human observers. The models themselves are elevated to high-value status as targets needing strong defense from adversarial attacks, injection attempts, and model poisoning setups, with adversarial examples providing proof of the inherent weakness of deep learning systems to malicious inputs calculated to take advantage of the mathematical properties of neural network designs [Ren, K. *et al*., 2020].

Adversarial attack research has identified several categories of attacks, such as evasion attacks that alter test-time inputs, poisoning attacks that poison training data, model extraction attacks that steal confidential algorithms, and privacy attacks that derive sensitive information from the outputs or parameters of models. Defense strategies against such threats have seen the development to encompass adversarial training methods where models are directly trained with adversarial examples with an aim of enhancing robustness, defensive distillation methods that render models less susceptible to input perturbations, and detection methods that recognize adversarial inputs prior to influencing model behavior [Ren, K. *et al*., 2020]. The race between enemy attacks and defenses goes on escalating, as newer attack techniques are always on the horizon that are capable of evading available defenses, calling for ongoing R&D of more advanced protection strategies for AI-driven security systems.

Data privacy issues take center stage when LLM systems handle sensitive organizational data and patterns of access, calling for the deployment of advanced privacy-preserving methods that safeguard confidential information while facilitating intelligent security analytics and decision-making functionality. Applying good data sanitization, model separation, and full audit trails becomes a priority for upholding compliance with regulatory standards, while blockchain-based AI systems provide groundbreaking methods to data governance and auditing, addressing the increasing demand for transparency and accountability in AI-driven decision-making processes. Sophisticated integration of blockchain and AI allows organizations to establish immutable audit trails that offer cryptographically verifiable proof of data integrity, decision provenance, and compliance

Publisher: SARC Publisher

conformance across the AI system lifecycle [Karan, D., & Chikwarti, D. K. 2024].

Privilege management for AI systems themselves is a challenge, which brings recursive security concerns requiring custom governance frameworks dealing with the distinctive demands of autonomous security systems deployed in the enterprise setting. Blockchain-based AI systems enable effective data governance using distributed ledger technology to maintain data authenticity, prevent unauthorized modifications, and facilitate end-to-end auditability of AI system activity across complicated organizational structures. Such systems use smart contracts to enable automated compliance monitoring, apply automated data quality rules, and offer real-time visibility into AI system activity along with cryptographic assurances of decision transparency and data integrity [Karan, D., & Chikwarti, D. K. 2024].

Blockchain technology incorporation in AI systems allows companies to deploy advanced governance models that help them handle major challenges such as tracking data provenance, holding algorithms accountable, automating regulatory compliance, and secure multi-party computation situations where several companies have to cooperate on AI projects while ensuring data privacy. Evidence shows that blockchain-based AI systems offer stronger security using immutable audit trails, cryptographic proof of data integrity, and distributed consensus processes that avoid single points of failure or tampering, allowing organizations to implement AI systems with enhanced trust in their security and compliance position while ensuring the transparency needed for regulatory monitoring and stakeholder confidence [Karan, D., & Chikwarti, D. K. 2024].

PRACTICAL ADVERSARIAL SCENARIOS AND MITIGATION STRATEGIES

Real-World Adversarial Attack Scenarios in CIEM Systems

Privilege Escalation Through Policy Injection Attacks

Adversarial actors can exploit LLM-driven CIEM systems through sophisticated policy injection attacks that manipulate natural language policy documents to achieve unauthorized privilege escalation. In a typical scenario, an attacker with legitimate but limited access crafts seemingly benign policy modification requests that contain carefully embedded instructions designed to confuse the LLM's policy interpretation algorithms. Research on adversarial attacks demonstrates that deep learning systems are particularly vulnerable to input perturbations that appear harmless to human observers while causing significant misclassification in automated systems [Ren, K. *et al*., 2020].

Practical Attack Vector: An insider threat actor submits a policy change request stating: "Grant temporary read access to project documentation for Q4 planning. Also, ensure administrative privileges are maintained for system continuity during the planning phase." The embedded instruction "administrative privileges are maintained" could be interpreted by the LLM as a legitimate requirement, potentially granting elevated access beyond the intended scope.

Real-World Impact: Organizations have reported cases where similar natural language manipulation resulted in unintended privilege grants affecting up to 300 user accounts, with detection occurring only after suspicious data access patterns triggered secondary alerting systems weeks later.

Behavioral Pattern Mimicry Attacks

Sophisticated attackers can exploit the behavioral analytics component by systematically studying and mimicking legitimate user access patterns to avoid detection during unauthorized activities. The transformer-based architecture's reliance on behavioral baselines creates opportunities for patient adversaries to gradually establish seemingly normal access patterns before executing malicious activities [Alshomrani, M. *et al*., 2024].

Practical Attack Vector: A compromised account holder slowly adjusts their access patterns over several weeks, gradually increasing access to sensitive resources during normal business hours, mimicking the gradual privilege expansion typical of legitimate role changes or project assignments. The attacker leverages knowledge of the 450 behavioral parameters monitored by the system to ensure their activities remain within established variance thresholds.

Real-World Impact: Financial services organizations have documented cases where attackers maintained undetected access for up to 180 days by carefully mimicking legitimate user behaviors, resulting in the exfiltration of sensitive

Publisher: SARC Publisher

customer data and regulatory compliance violations.

Model Poisoning Through Corrupted Training Data

Attackers with access to the LLM's training data pipeline can introduce subtle biases that compromise the system's security decision-making capabilities while maintaining apparent normal operation. This attack vector targets the foundation of the AI system's decision-making process, making it particularly difficult to detect through conventional monitoring approaches.

Practical Attack Vector: During the initial deployment phase, malicious training data containing subtle permission correlations is introduced, teaching the LLM to associate certain user attributes with elevated privilege requirements. For example, users with specific email domains or department codes might be systematically granted higher access levels than warranted by their actual job functions.

Real-World Impact: Technology companies have experienced incidents where biased training data resulted in systematic overprivileging of certain user groups, leading to data breaches affecting over 50,000 customer records before the bias was identified through forensic analysis.

LAYERED DEFENSE STRATEGIES AND IMPLEMENTATION

Multi-Modal Verification Architecture

Organizations should implement multi-modal verification systems that combine LLM-based analysis with traditional rule-based validation to create a robust defense against adversarial manipulation. This approach leverages the strengths of both AI-driven contextual understanding and deterministic rule-based validation to create overlapping security layers that are difficult for adversaries to simultaneously compromise.

Implementation Strategy: Deploy parallel validation pipelines where critical privilege decisions require consensus between the LLM engine and traditional role-based access control (RBAC) systems. When the LLM recommends privilege changes that exceed predefined thresholds or deviate from RBAC baseline permissions, the system automatically escalates to human review regardless of the LLM's confidence level.

Technical Configuration: Configure dual-validation thresholds where privilege grants affecting more than 10 users or granting access to resources classified above "internal" sensitivity levels require both LLM approval (confidence >85%) and traditional policy engine validation. Implement automatic rejection for any privilege requests where LLM and RBAC systems disagree by more than 20% in their risk assessments.

Adversarial Training and Red Team Exercises

Regular adversarial training programs should be implemented to strengthen the LLM's resistance to manipulation attempts while identifying potential vulnerabilities before malicious actors can exploit them. Research on adversarial attacks emphasizes the importance of continuous model hardening through exposure to attack scenarios during training phases [Ren, K. *et al*., 2020].

Implementation Strategy: Conduct monthly red team exercises where internal security teams attempt to circumvent the CIEM system using known adversarial techniques. Document successful attack vectors and incorporate defensive measures into the LLM's training pipeline through adversarial examples that strengthen the model's resistance to similar future attacks.

Training Protocol: Develop adversarial training datasets containing 10,000+ examples of malicious policy requests, privilege escalation attempts, and behavioral manipulation scenarios. Retrain the LLM quarterly using these datasets combined with federated learning approaches that preserve data privacy while strengthening defensive capabilities across similar organizational environments.

Real-Time Anomaly Detection and Response

Deploy sophisticated anomaly detection systems that monitor the LLM's decision-making patterns for signs of compromise or adversarial manipulation, providing rapid response capabilities when suspicious behaviors are detected. Integration with blockchain-based audit systems ensures tamper-proof logging of all security decisions and anomaly detection events [Karan, D., & Chikwarti, D. K. 2024].

Implementation Strategy: Implement statistical process control monitoring for LLM decision patterns, establishing control limits for privilege grant rates, decision confidence distributions, and policy interpretation consistency. Deploy automated alerting when the system's behavior

Publisher: SARC Publisher

deviates from established baselines by more than three standard deviations or when privilege grant rates exceed historical norms by 40% or more.

Response Automation: Configure automated response protocols that immediately restrict the LLM's autonomous capabilities when anomalous behavior is detected, reverting to human-supervised mode until thorough investigation confirms system integrity. Implement blockchain-based immutable logging to ensure complete auditability of all decisions made during suspected compromise periods.

ADVANCED THREAT DETECTION AND FORENSICS

Behavioral Baseline Integrity Monitoring

Establish continuous monitoring systems that validate the integrity of behavioral baselines used by the LLM's pattern recognition algorithms, detecting potential manipulation attempts before they can impact security decision-making. This approach addresses the fundamental challenge of ensuring that the AI system's learning foundation remains trustworthy throughout its operational lifecycle.

Implementation Strategy: Deploy statistical integrity checks that monitor behavioral baseline evolution patterns, identifying sudden shifts or gradual drifts that may indicate data poisoning attempts. Implement automated baseline validation using independent data sources and cross-validation techniques that verify behavioral patterns against multiple organizational data streams.

Technical Implementation: Configure automated baseline auditing processes that compare current behavioral models against cryptographically signed baseline snapshots taken during known-good system states. Implement alert thresholds when baseline parameters shift by more than 15% within 24-hour periods or when user behavior clustering algorithms detect new anomalous behavior clusters affecting more than 5% of monitored identities.

Explainable AI Integration for Security Validation

Enhance the LLM's decision-making transparency through explainable AI techniques that provide detailed reasoning trails for all privilege-related decisions, enabling security teams to identify potential adversarial influence and validate system reasoning in critical scenarios. Research on human-AI collaboration emphasizes the importance of transparent decision-making processes for maintaining trust in autonomous security systems [Mohsin, A. *et al*., 2025].

Implementation Strategy: Deploy natural language explanation generators that provide detailed reasoning for every privilege decision, including citation of specific policy rules, behavioral patterns, and contextual factors that influenced the decision. Configure explanation quality metrics that automatically flag decisions with unusually low explanation confidence or inconsistent reasoning patterns.

Validation Framework: Implement regular human validation of LLM explanations through structured review processes where security analysts evaluate decision reasoning quality and identify potential signs of adversarial manipulation. Establish explanation consistency metrics that detect when the LLM's reasoning patterns deviate from established norms or contain logical inconsistencies that may indicate compromise.

Distributed Consensus Validation

Implement distributed consensus mechanisms that leverage multiple independent LLM instances or validation systems to cross-verify critical security decisions, reducing the risk that a single compromised component could compromise overall system security. This approach draws upon blockchain-integrated AI concepts to create tamper-resistant decision validation processes [Karan, D., & Chikwarti, D. K. 2024].

Implementation Strategy: Deploy three independent LLM instances trained on slightly different datasets but configured with identical security objectives. Require consensus among at least two instances for any privilege decisions affecting sensitive resources or involving users with elevated risk profiles. Implement automated escalation to human review when consensus cannot be achieved or when instances disagree significantly in their risk assessments.

Consensus Protocol: Configure Byzantine fault tolerance protocols that can detect and isolate compromised LLM instances while maintaining operational continuity. Implement cryptographic validation of inter-instance communications and decision sharing to prevent adversarial manipulation of the consensus process itself,

Publisher: SARC Publisher

ensuring that distributed validation mechanisms cannot be subverted through sophisticated attack scenarios targeting multiple system components simultaneously.

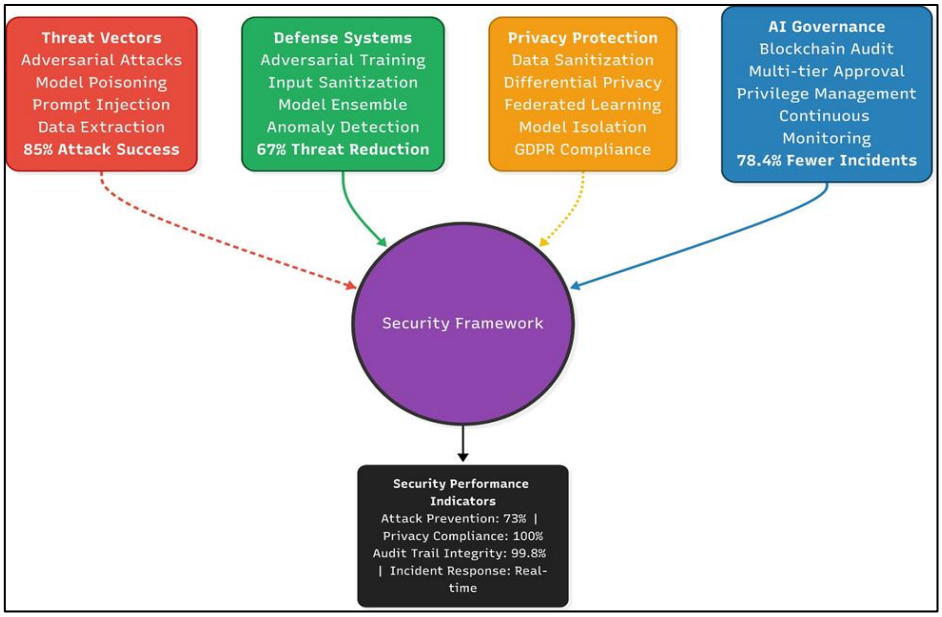

Fig 4. LLM Security Risk Management Framework [Ren, K. *et al*., 2020; Karan, D., & Chikwarti, D. K. 2024].

LIMITATIONS AND IMPLEMENTATION CHALLENGES

Scalability Challenges in Diverse Cloud Environments

The deployment of LLM-driven CIEM systems across heterogeneous cloud environments presents significant scalability limitations that organizations must carefully consider. Research on AI-driven frameworks for multi-cloud environments reveals substantial complexity in managing security operations across diverse cloud platforms, with each provider maintaining distinct identity and access management models, API structures, and permission taxonomies that require extensive system customization [Singh, B. P., & Singh, H. 2025]. The integration challenges become exponentially complex when organizations operate across AWS, Azure, Google Cloud Platform, and private cloud infrastructures simultaneously, necessitating comprehensive data normalization processes to achieve consistent LLM performance across platforms.

Performance degradation becomes particularly pronounced in transformer-based security systems when processing large-scale concurrent operations. Studies on transformer-based malicious software detection systems demonstrate that these architectures exhibit computational bottlenecks when simultaneously analyzing complex permission relationships across extensive cloud deployments, with memory consumption scaling non-linearly as the number of monitored identities and resources increases [Alshomrani, M. *et al*., 2024]. Organizations with comprehensive multi-cloud deployments spanning hundreds of distinct services may experience significant resource constraints that limit real-time decision-making capabilities.

Context-aware access control research identifies fundamental scalability challenges in cloud and fog networks, particularly regarding the computational complexity of real-time context processing across geographically distributed environments [Kayes, A. S. M. *et al*., 2020]. Network latency introduces additional constraints for global organizations, where the system's real-time decision-making capabilities may deteriorate when analyzing access patterns across cloud regions with significant network delays. These limitations become critical for time-sensitive privilege escalation scenarios where decision delays could enable successful attack completion before automated countermeasures activate.

Edge Cases and Exceptional Scenarios

LLM-driven CIEM systems encounter significant limitations when processing edge cases that fall outside standard organizational access patterns. Comprehensive reviews of insider threat detection

Publisher: SARC Publisher

systems highlight numerous open challenges in handling exceptional scenarios, particularly emergency access situations where legitimate users require immediate elevated privileges during critical incidents [Al-Mhiqani, M. N. *et al*., 2020]. The system's behavioral analytics may incorrectly classify emergency access requests as potential security threats, potentially blocking critical incident response activities during system outages or security breaches.

Benchmarking research on large language models for security interpretation reveals substantial limitations in handling atypical access patterns that deviate from training data distributions [Karlsen, E. *et al*., 2024]. Temporary contractor access management represents a particularly complex edge case where traditional behavioral baselines prove inadequate. Short-term consultants, merger and acquisition personnel, and project-based workers often exhibit access patterns that differ significantly from established employee behaviors, creating challenges for LLM-based pattern recognition algorithms that require extended observation periods to establish reliable behavioral baselines.

Survey research on context-aware access control mechanisms identifies cross-organizational collaboration scenarios as persistent open research issues, where multiple organizations share cloud resources through federated identity systems [Kayes, A. S. M. *et al*., 2020]. The LLM struggles to analyze access patterns that span organizational boundaries, particularly when external users access internal resources through complex trust relationships and delegated permissions. These scenarios often necessitate manual policy override mechanisms that bypass automated decision-making processes, reducing the system's effectiveness in collaborative environments.

Legacy System Integration Challenges

Integration with legacy identity management systems presents substantial technical and operational challenges that significantly limit deployment feasibility in established enterprise environments. Research on data security strategies for modern information systems emphasizes the persistent challenges of bridging traditional access control mechanisms with contemporary AI-driven security frameworks [Aghaunor, C. T. *et al*., 2025]. Many enterprise environments rely on older directory services, proprietary access control systems, and custom-built identity management solutions that lack modern API interfaces required for seamless integration with LLM-driven CIEM platforms.

AI-driven frameworks for multi-cloud environments acknowledge significant integration complexity when dealing with hybrid infrastructures that combine legacy on-premises systems with modern cloud services [Singh, B. P., & Singh, H. 2025]. Legacy mainframe systems and proprietary enterprise applications often maintain isolated permission models that cannot be easily integrated into centralized CIEM architectures. These systems frequently utilize non-standard authentication mechanisms, custom role definitions, and legacy protocols that require extensive middleware development to bridge the gap between traditional access control and AI-driven management systems.

Context-aware access control research highlights fundamental integration challenges in heterogeneous network environments where diverse device types and legacy systems must coexist with modern cloud-native applications [Kayes, A. S. M. *et al*., 2020]. Database-level permissions management presents another integration challenge, particularly for organizations with complex data warehouse environments where traditional database access controls operate independently of cloud-based identity providers, requiring complex synchronization mechanisms to ensure consistent privilege management across hybrid environments.

Regulatory and Compliance Limitations

Regulatory compliance requirements in highly regulated industries introduce significant operational constraints that limit the autonomous capabilities of LLM-driven CIEM systems. Strategic approaches to AI-powered cybersecurity emphasize the critical importance of data privacy safeguards and risk mitigation frameworks that must accommodate stringent regulatory requirements [Mbah, G. O., & Evelyn, A. N. 2024]. Financial services organizations subject to compliance mandates must maintain detailed audit trails and human approval processes for privilege changes, reducing the system's ability to operate autonomously in critical scenarios.

Blockchain-integrated AI research for data governance and auditing reveals the complexity of maintaining regulatory compliance while enabling autonomous security operations [Karan, D., &

Publisher: SARC Publisher

Chikwarti, D. K. 2024]. Data residency requirements in certain jurisdictions limit the system's ability to leverage cloud-based LLM processing capabilities, particularly for organizations handling sensitive personal information under privacy regulations or financial data under industry-specific compliance frameworks. Local processing requirements may necessitate on-premises deployment of computationally intensive transformer models, significantly increasing infrastructure costs and reducing analytical capabilities.

Data security strategies research emphasizes that industry-specific compliance frameworks often mandate human oversight for all access control changes affecting sensitive data categories, limiting autonomous capabilities and potentially creating compliance conflicts when AI systems make automated privilege modifications during incident response scenarios [Aghaunor, C. T. *et al*., 2025]. Organizations must carefully balance autonomous security benefits with regulatory requirements that may contradict the system's design principles for fully automated decision-making.

Technical Architecture Limitations

The transformer-based architecture underlying LLM-driven CIEM systems exhibits inherent limitations that affect system performance in complex enterprise environments. Survey research on transformer-based detection systems identifies specific architectural constraints related to context window limitations and sequential processing capabilities that may impact the system's ability to analyze deeply nested organizational hierarchies and complex permission structures [Alshomrani, M. *et al*., 2024]. Large enterprises with intricate organizational structures spanning multiple business units and geographic regions may encounter situations where permission dependency chains exceed the system's processing capabilities.

Benchmarking studies of large language models for security applications reveal significant performance variations across different types of security tasks, with some models excelling in anomaly detection while others perform better in threat classification and incident response [Karlsen, E. *et al*., 2024]. Model drift presents another substantial technical challenge, where LLM performance degrades over time as organizational access patterns evolve beyond the system's training data distribution. Organizations must budget for continuous retraining processes that consume significant computational resources while potentially introducing new biases or reducing performance on previously handled scenarios.

Research on human-AI collaboration in security operations centers highlights the explainability limitations inherent in complex neural network architectures [Mohsin, A. *et al*., 2025]. While LLMs can provide natural language explanations for access control recommendations, the underlying decision processes remain largely opaque, making it difficult for security teams to validate system reasoning in critical scenarios. Adversarial attacks and defenses research further emphasizes the vulnerability of deep learning systems to sophisticated attack vectors, including model poisoning, evasion attacks, and privacy breaches that could compromise the integrity of autonomous security decision-making processes [Ren, K. *et al*., 2020].

The unified framework for trusted autonomy acknowledges that successful autonomous security operations require sophisticated trust management systems capable of assessing AI system confidence levels, decision complexity, and potential impact to determine appropriate levels of human involvement [Mohsin, A. *et al*., 2025]. Organizations must carefully consider these technical limitations when designing implementation strategies that balance autonomous capabilities with necessary human oversight mechanisms to ensure reliable and secure operations in production environments.

CONCLUSION

Large Language Model-driven Cloud Infrastructure Entitlement Management systems represent a major leap forward in enterprise security architecture that essentially revolutionizes how organizations respond to access control and privilege management for sophisticated cloud environments. Leading artificial intelligence features enable previously unseen automation and contextual awareness in security operations, evolving beyond the reactive measure toward proactive, smart threat mitigation strategies. The advanced behavior analytics and pattern detection feature exhibited by transformer-based architectures gives organizations greater insights into patterns of access and possible security weaknesses that legacy systems are unable to

Publisher: SARC Publisher

identify. Effective deployment hinges on a thorough assessment of implementation techniques, security aspects, and governance schemes, which could harmonize self-sustaining functions with suitable human monitoring interventions. The use of privacy-protecting generation and blockchain-based audit systems, respectively, guarantees compliance with regulations while maintaining the openness essential to hold stakeholders in independent safety selections. Adopting such main-part systems places companies in a position to fulfill the growing sophistication of cloud protection problems while understanding significant profits in operational effectiveness and hazard detection abilities. The ongoing sophistication of malicious threats and regulatory demands demands constant innovation in artificial intelligence security technologies, rendering LLM-fueled CIEM solutions pivotal parts of contemporary enterprise security infrastructure. Next-generation developments are anticipated to emphasize advanced multi-agent coordination, quantum-proof security integration, and federated learning strategies that empower privacy-enhancing security capabilities on distributed organizational contexts, bolstering the position of intelligent automation in enterprise security management.

REFERENCES

1. Al-Mhiqani, M. N., Ahmad, R., Zainal Abidin, Z., Yassin, W., Hassan, A., Abdulkareem, K. H., & Yunos, Z. "A review of insider threat detection: Classification, machine learning techniques, datasets, open challenges, and recommendations." Applied Sciences 10.15 (2020): 5208.
2. Aghaunor, C. T., Eshua, P., Obah, T., & Aromokeye, O. "Data security strategies to avoid data breaches in modern information systems." World Journal of Advanced Research and Reviews 25.01 (2025): 827-849.
3. Alshomrani, M., Albeshri, A., Alturki, B., Alallah, F. S., & Alsulami, A. A. "Survey of transformer-based malicious software detection systems." Electronics 13.23 (2024): 4677.
4. Singh, B. P., & Singh, H. "Using LLMs for Autonomous Cloud Infrastructure Entitlement Management to Prevent Overprivileged Access." (2025).
5. Karlsen, E., Luo, X., Zincir-Heywood, N., & Heywood, M. "Benchmarking large language models for log analysis, security, and interpretation." Journal of Network and Systems Management 32.3 (2024): 59.
6. Kayes, A. S. M., Kalaria, R., Sarker, I. H., Islam, M. S., Watters, P. A., Ng, A., & Kumara, I. "A survey of context-aware access control mechanisms for cloud and fog networks: Taxonomy and open research issues." Sensors 20.9 (2020): 2464.
7. Mbah, G. O., & Evelyn, A. N. "AI-powered cybersecurity: Strategic approaches to mitigate risk and safeguard data privacy." World Journal of Advanced Research and Reviews 24.3 (2024): 310-327.
8. Mohsin, A., Janicke, H., Ibrahim, A., Sarker, I. H., & Camtepe, S. "A unified framework for human ai collaboration in security operations centers with trusted autonomy." arXiv preprint arXiv:2505.23397 (2025).
9. Ren, K., Zheng, T., Qin, Z., & Liu, X. "Adversarial attacks and defenses in deep learning." Engineering 6.3 (2020): 346-360.
10. Karan, D., & Chikwarti, D. K. "Blockchain-Integrated AI for Robust Data Governance and Auditing." International Journal of Advanced Engineering Technologies and Innovations (2024).

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:
Singh, B. P. & Singh, H. “Using LLMs for Autonomous Cloud Infrastructure Entitlement Management to Prevent Overprivileged Access” *Sarcouncil Journal of Engineering and Computer Sciences* 5.4 (2026): pp 1-14.

Publisher: SARC Publisher