Universal Library of Engineering Technology

Research Article

ISSN: 3064-996X | Volume 2, Issue 4

Open Access | PP: 74-79

DOI: https://doi.org/10.70315/uloap.ulete.2025.0204013

Artificial Intelligence for Threat Detection: Leveraging Deep Learning to Identify Zero-Day Attacks in Real Time

Geol Kang

Hyundai Autoever America, Senior Security Architect, Fountain Valley, California, USA.

Abstract

The article examines the use of deep artificial intelligence models to detect zero-day attacks amid rapidly escalating cyber-threat complexity. The relevance stems from the fact that the emergence rate of zero-day exploits consistently outpaces the rate of signature generation. At the same time, traffic encryption, malware polymorphism, and the growth of attack automation markedly degrade the efficacy of classical IDS. The objective is to conduct a systematic analysis of deep learning's potential to provide real-time detection of unknown vulnerabilities and to assess the architectures, training regimes, and telemetry requirements that determine a model's ability to recognize previously unobserved patterns. The study's novelty lies in a multi-level methodological integration: mapping modern neural architectures to the characteristics of enterprise traffic, analyzing model robustness to concept drift and adversarial interference, and evaluating prospects for autonomous vulnerability remediation and the use of post-quantum cryptographic mechanisms in distributed learning. The main findings underscore that deep learning transforms the threat-detection paradigm: it shifts defense from retrospective signatures to behavioral baselines, compresses dwell time from days to seconds, and enables a reactive-proactive security loop. The effectiveness of such systems is determined not only by architectural choice, but also by the maturity of telemetry collection, verification, and versioning processes; resilience to data poisoning; decision interpretability; and engineering optimization of computational pipelines. The work also identifies critical constraints: scarcity of labeled traces, behavioral drift, hardware costs, and the risk of model compromise. These limits delineate development directions, from generating temporary patches with large language models to offloading inference to edge devices and employing digital twins for continuous self-training of detection systems. The article will be of use to researchers, security engineers, and SOC professionals deploying behavioral mechanisms for zero-day attack detection.

Keywords: *Artificial Intelligence, Deep Learning, Zero-Day Attacks, Behavioral Detection, Cybersecurity.*

INTRODUCTION

A zero-day attack is the exploitation of a vulnerability unknown to developers and therefore not yet blockable by a signature (Ignacio et al., 2025). It is termed zero-day because the product vendor has precisely zero days to patch it. The problem's scale is now comparable to prevalent social-engineering methods: according to Google Threat Intelligence, in 2024 alone, 75 vulnerabilities worldwide were identified that began to be actively exploited before publication of CVE entries, with nearly half affecting enterprise security tools and network appliances, systems in which compromise instantaneously expands an organization's perimeter (Lakshmanan, 2025).

Signature-based IDS and antivirus engines are, by definition, backward-looking: to craft a rule, an attack must first be observed. Adversaries exploit this by packing exploits hourly and combining them with encrypted traffic. According to the Verizon DBIR-2024 report, the number of such incidents increased by 180% compared to previous years, and 15% of confirmed compromises in 2023 were a result of vulnerability exploitation (Verizon Business, 2024). In other words, customary tools, even with daily updates, miss anything that doesn't fall into a known vulnerability, and a slight change in the packing method makes the signature you deployed yesterday obsolete faster than your security team can apply an update.

Artificial intelligence closes this temporal asymmetry. Deep neural networks model the system's normal behavior holistically rather than at the level of an individual file: they learn cycles of system calls, graphs of network sessions, and spectral properties of packet sequences. When a previously unseen exploit appears in the infrastructure, the model

Citation: Geol Kang, "Artificial Intelligence for Threat Detection: Leveraging Deep Learning to Identify Zero-Day Attacks in Real Time", Universal Library of Engineering Technology, 2025; 2(4): 74-79. DOI: https://doi.org/10.70315/uloap.ulete.2025.0204013.

responds to an anomalous trajectory before a signature is available, reducing the so-called dwell time, the period between the intrusion and detection. Even without an automated response, the benefit is quantifiable: according to Mandiant M-Trends-2025, the global median dwell time remains at 10–11 days, but in cases where the ransomware itself initiates the alert, it falls to five days; well-tuned streaming inference based on AI targets seconds, effectively eliminating the window for a stealthy operation (Google, 2025). Consequently, enterprise monitoring centers are pivoting from signature-based to behavioral threat detection: the more intricate and mutable the attack, the more strongly it deviates from the multifactor baseline captured by a deep model, and the faster it can be stopped.

MATERIALS AND METHODOLOGY

The research employed a multi-level methodology that combined literature analysis, a comparative study of deep-learning architectures, and the evaluation of telemetry quality for zero-day attack detection. First, a targeted review was conducted of sources capturing the dynamics of threats and the growth in zero-day exploits (Lakshmanan, 2025; Verizon Business, 2024), as well as works devoted to behavioral analysis and streaming detection of unknown vulnerabilities (Ignacio et al., 2025). This enabled isolating the gap between exploit emergence and signature generation, underscoring the need to focus on models that operate on previously unobserved patterns. Next, a technical analysis was performed of telemetry sources, packet captures, flow-level metadata, and endpoint logs, assessing their suitability for ingestion by convolutional, recurrent, attention-based, and graph-based networks. Particular attention was paid to the impact of traffic encryption on feature availability (Crumb, 2025) and to requirements for normalization, temporal synchronization, and data cleansing to form a unified feature space.

The subsequent stage focused on model robustness and lifecycle. Mechanisms for compensating concept drift through continual and few-shot learning were studied and compared with the observed obsolescence of static rules in industrial environments (Verizon Business, 2024). In addition, methods for countering adversarial actions were analyzed, including weight protection, model distillation, and split/federated learning with differential privacy, which are especially relevant in the face of telemetry interception or manipulation (Zorz, 2025; Cisco, 2025). The final step entailed a synthetic mapping of architectures, training regimes, and data characteristics to industry-dwell time metrics (Google, 2025), enabling the assessment of deep learning's potential to provide reactive and proactive real-time protection against zero-day attacks.

RESULTS AND DISCUSSION

When a vulnerability surfaces before the vendor can release a patch, infrastructure remains unprotected, and adversaries operate day to day. According to the DBIR-2025 report, attacks specifically exploiting flaws in edge gateways and virtual private networks jumped from 3% to 22% in a year; only 54% of such holes were closed, and the median time to patch stretched to 32 days (Alder, 2025). Thus, organizations experience a whole month of technical limbo. At the same time, the adversary retains eleven days on average, the period an intruder remained undetected inside networks in 2024, compared with 205 days a decade ago.

As attackers increasingly leverage automation and machine-learning methods, the market for AI-based cybersecurity solutions is expanding at double-digit rates. As shown in Figure 1, according to Grand View Research, the global market for AI in cybersecurity in 2024 was approximately USD 25.35 billion and is projected to reach roughly USD 93.75 billion by 2030, corresponding to a compound annual growth rate of about 24.4% from 2025 to 2030 (GVR, n.d.).

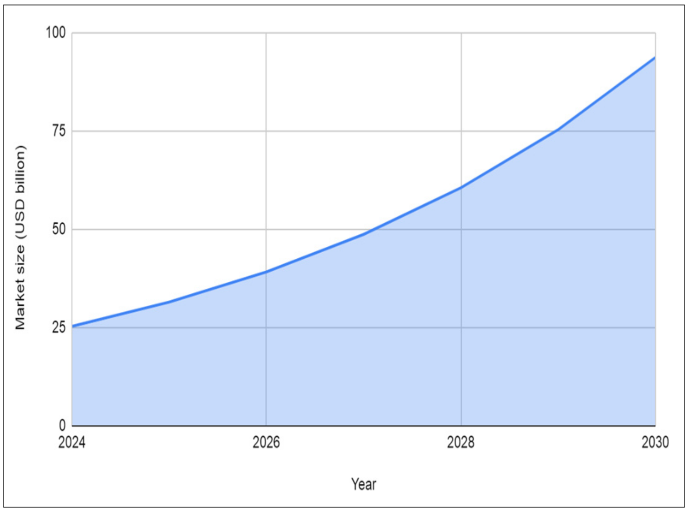

Fig. 1. Projected Global AI in Cybersecurity Market Growth, 2024–2030 (GVR, n.d.)

Every additional day of adversary presence equates to monetary erosion. A worldwide cost study by IBM in 2025 puts the average cost per incident at USD 4.4 million, down 9% due to faster discovery and containment (IBM, 2025). If anything, the more planned the corporate data, the larger the invisible reputation tax: share prices drop, supplier reliability suffers, and regulatory fines are generally levied after the point of technical remediation following a data breach. Thus, the direct costs are restoring the system, compensating the users, and legal expenses.

Recent incidents provide concreteness. In spring 2024, a zero-day in a file-sharing system spread widely: the immediate compromise affected about one hundred service customers, but through chains of trusted channels, data also leaked from an additional 2,700 organizations, exposing ninety-three million personal records; the vendor's emergency outlays had already exceeded one million dollars and continue to grow due to class-action litigation (Kapko, 2024). In winter 2024, an unknown operator began covertly exploiting a flaw in a major manufacturer's remote-access gateways; the first signs date to mid-December, while public disclosure occurred only in January, by which time the intrusion had reached industrial scale (Zorz, 2025). Finally, in September 2025, a vulnerability in the web interface of network routers allowed a remote adversary to obtain superuser privileges. A patch was released, but no interim mitigations existed, meaning any unpatched devices were automatically convenient entry points into networks (Cisco, 2025). Collectively, these episodes demonstrate that each zero-day is not a singular event but a cascade of consequences, in which a technical window rapidly transforms into an economic vortex.

The pace of new vulnerabilities invariably outstrips the rate at which signature databases complete the creation, validation, and distribution cycle. Empirical research on hybrid IDS shows that the interval between the first sighting of an exploit and the appearance of the corresponding signature exceeds 24 hours; consequently, the window for unpunished intrusion persists longer than a day even in networks with automatic rule updates (Agoramoorthy et al., 2023). By the time a SOC team loads an update, the adversary has already modified the command-and-control infrastructure, and the new attack version becomes invisible to the detector once again.

Simultaneously, polymorphism and cryptographic opacity of traffic are intensifying: in 2024, 75% of new malware samples used packing or obfuscation, and the number of unique polymorphic variants increased by 38% in one year (Market Growth Reports, 2025). The transmission channel is even more challenging: as shown in Figure 2, Chrome telemetry indicates that by mid-2025, 96.1% of global web traffic traversed HTTPS, that is, encrypted, rendering it inaccessible to classical DPI analysis without costly perimeter decryption (Crumb, 2025).

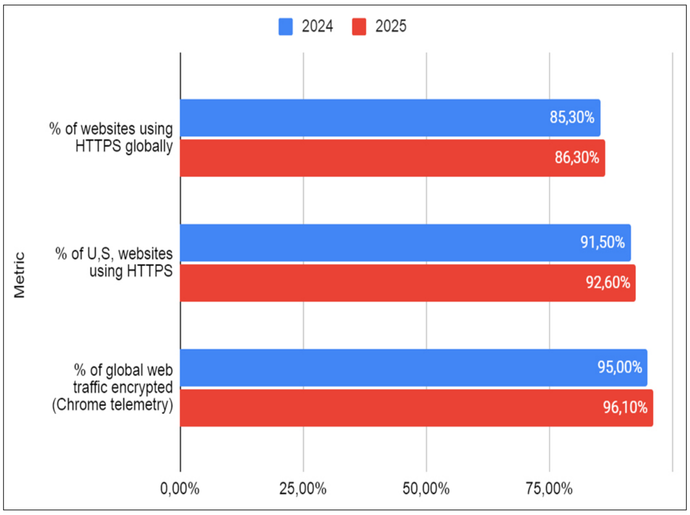

Fig. 2. Global HTTPS Adoption and Performance Metrics (Crumb, 2025)

As a result, even when signatures exist, the engine has little to compare: payloads are concealed, and the TLS session window size is minimally informative.

In an effort to avoid misses, administrators harden rules, but the price is an avalanche of false alarms. When the false-positive rate exceeds 50%, analysts begin to disregard alerts, and the system loses meaning, ceding to noise (Soureya et al., 2025). A review of hybrid IDSs notes that adding anomaly modules to catch unknown attacks in environments with untimely signatures more than doubles false positives, effectively converting the under-detection problem into operator fatigue (Agoramoorthy et al., 2023). Thus, update latency, polymorphism, and encryption amplify one another, and the classical IDS/IPS + SIEM stack grounded in static rules is constrained not only by yesterday's pattern effect but also by human factors that paradoxically lower security as formal defenses proliferate.

Deep learning shifts the defensive emphasis from the past to the rapidly evolving present by constructing a flexible representation of network and host normality that is not anchored to a predefined feature set. Instead of static templates, it forms a multi-level feature space in which even minute deviations in byte distribution, the cadence of system calls, or the configuration of interacting nodes trigger alarms while malicious code is still staging its command-and-control infrastructure.

Three principal families of neural networks form the core. Convolutional models extract local correlations: packet segment sequences, flag orderings, and header spectral signatures. Recurrent structures and their more modern attention-based successors track causal relationships in long time-series of traffic data, exposing context masked by myriad micro-events that appear benign in isolation. Graph neural networks add a topological dimension, representing infrastructure as a dynamic graph of processes, sessions, and privileges, and identifying atypical data-movement routes.

These architectures can be flexibly combined with varied training regimes. Supervised learning achieves high accuracy. Unsupervised learning identifies previously unknown patterns by clustering similar system performance, then marking anomalies on anomaly maps by intrusions at the edges of those clusters. Few-shot learning identifies patterns from a small number of annotated examples, therefore addressing the problem of insufficient annotated training data. In continual learning, new events are fed through the model to counteract the drift caused by user behavior and infrastructure changes.

To maintain such a complex organism, data sources must be heterogeneous yet synchronized. Packet captures expose the raw bit-level view and permit protocol inspection without loss of detail; flow-level metadata record connection statistics, conserving resources while providing visibility into macro-traffic; endpoint logs document the operating system's internal mechanics, reflecting process launches, registry access, and library-load dynamics. Before model ingestion, all streams undergo multi-stage cleansing, normalization, and temporal alignment, then become a unified feature set that sustains a single cognitive field in which each new anomaly is immediately highlighted against the familiar status quo.

The quality of the detection system's intake is determined far more by data discipline than by hardware throughput. Each fragment of telemetry must pass a rigorous lifecycle: completeness checks, timestamp consistency, contextual enrichment, and immutable archiving. Versioning of feature sets and model configurations turns the machine-learning platform into a replicable laboratory where any result can be reproduced down to the byte. Such an environment should be built along continuous-delivery principles. Once an updated architecture passes holdout-set tests, the orchestration platform promotes it into production without pipeline downtime and with automatic rollback if metrics degrade.

Actual robustness arrives with immunity to malicious interference. Attacks crafted to mislead models operate differently from classical exploits yet can precipitate similar consequences. To counteract this factor, weights are trained on deliberately perturbed examples, forcing the network to recognize intentional distribution shifts. An additional protection layer is provided by distillation: a large model transfers knowledge to a more compact replica deployed at the perimeter; should its parameters be extracted, the adversary obtains an averaged, and thus minimally functional, representation that cannot reveal the internal decision logic.

From the operator's perspective, success is measured not solely by response speed but also by transparency. SOC analysts must understand why a session was terminated or a container quarantined. Feature-importance heatmaps, textual summaries of key anomalies, and links to raw events render conclusions interpretable. When neural network signals are accompanied by concrete evidence, trust consolidates, and the probability of a manual rollback of automation declines.

Finally, the data underpinning intelligence must not become a new vulnerability. A federated approach trains weights at the locations where telemetry originates, exporting only gradients. Differential privacy adds calibrated noise, assuring that personal details cannot be reconstructed from gradients or the final model. This union of distributed learning and statistical masking renders exploitation of confidential logs impractical, thus closing another attack vector and leaving fewer unstitched seams in the defensive fabric. Implementation recommendations are shown in Figure 3.

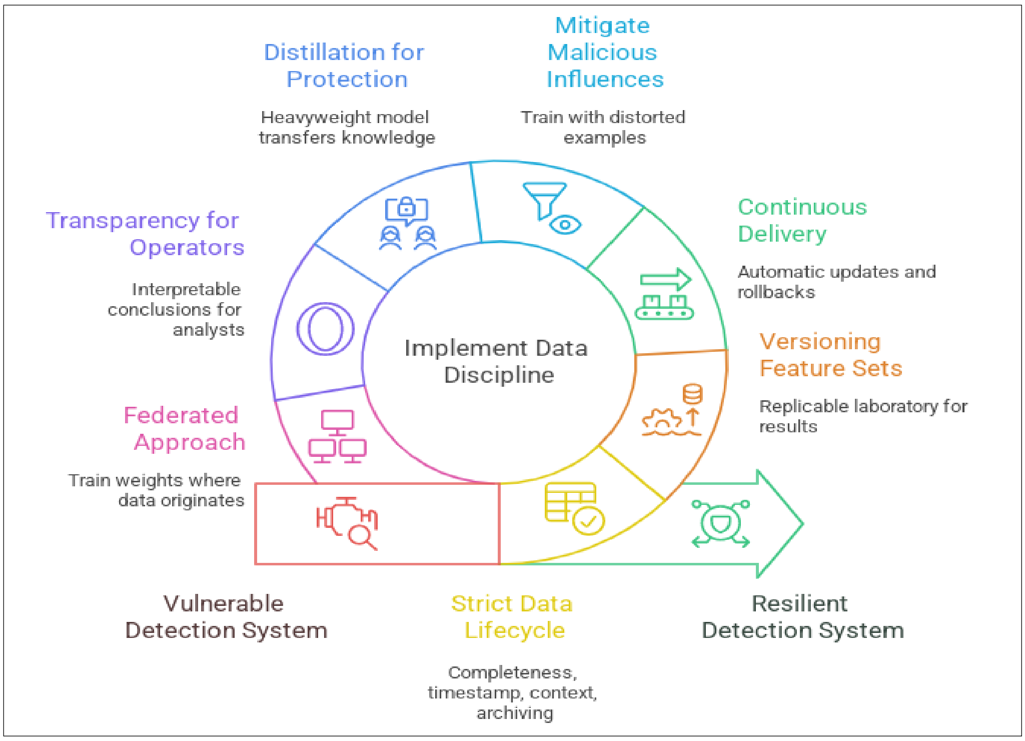
Distillation for Protection
Heavyweight model transfers knowledge
Mitigate Malicious Influences
Train with distorted examples
Transparency for Operators
Interpretable conclusions for analysts
Continuous Delivery
Automatic updates and rollbacks
Implement Data Discipline
Versioning Feature Sets
Replicable laboratory for results
Federated Approach
Train weights where data originates
Vulnerable Detection System
Strict Data Lifecycle
Completeness, timestamp, context, archiving
Resilient Detection System

Fig. 3. Building a Resilient Detection System

One practical illustration of such an architectural approach is a multi-year cybersecurity transformation program implemented for a global automotive group operating in the U.S. market. In this case, the introduction of deep-learning–based detection was deliberately preceded by a systematic hardening and telemetry-enrichment of the perimeter and internal network. An enterprise web application firewall was deployed across more than 400 corporate and consumer-facing domains, increasing adequate application defense coverage for externally exposed services from under 10% to nearly 99%. In parallel, a next-generation network access control system based on Cisco ISE was rolled out across dozens of U.S. offices, enforcing device- and identity-centric policies for thousands of endpoints. This was complemented by cloud security posture management and cloud workload protection for AWS and Azure environments, aligned with

NIST and ISO control frameworks, as well as privileged access management and granular data access control for critical R&D and financial systems to mitigate insider risk. At the perimeter, a global DDoS mitigation program integrating carrier-grade scrubbing centers and multi-layer defenses was established to shield online services, including connected-car portals and customer-facing applications.

From a deep learning perspective, such a program does not merely add protective layers; it enriches the feature space on which models operate. WAF telemetry provides structured application-layer anomalies and exploit blocks to train sequential and temporal models to distinguish between legitimate usage spikes and emerging attacks. By incorporating NAC and Cisco ISE analytics with high-fidelity identity and device context, graph-based networks can recognize anomalous traffic patterns by device type, user role and office location, rather than on abstract IP addresses. CSPM and CWPP systems continuously describe the configuration state of cloud workloads, which builds a baseline for detecting configuration drift if the workload is compromised. PAM and DAC logs, in turn, capture high-risk operations in sensitive environments, enabling models to flag unusual privilege escalations or data-access paths even when individual actions remain formally authorized. Finally, DDoS telemetry and scrubbing-center statistics provide rich training material for recognizing early stages of volumetric and application-layer attacks, allowing streaming inference pipelines to distinguish zero-day denial-of-service patterns from legitimate flash-crowd events.

The measurable outcomes of this deployment further support the thesis that deep-learning–driven detection, when combined with mature telemetry and control infrastructure, yields not only technical but also economic benefits. Following the rollout of global DDoS protection, service availability across key online channels increased by more than 40%, thereby reducing the window during which customers experienced outages. Regulatory compliance and audit scores across U.S. subsidiaries improved as controls were aligned with headquarters requirements and local regulations, simplifying external assessments and reducing the probability of fines. Centralized NAC policies and the automation of access management workflows lowered IT operational costs by an estimated 15–20%, partly offsetting the investment in advanced detection and control tooling. At the same time, integrating SIEM with advanced threat intelligence and behavioral analytics reduced the mean time to detect incidents by more than 30%, translating into millions of dollars in avoided or mitigated losses over the program horizon. Strengthened resilience of application defenses and the stability of connected-car and portal services also had an indirect but significant effect: customer trust in digital channels increased, which is critical for an automotive manufacturer as vehicles and mobility services become tightly coupled with cloud-based infrastructure.

Even the most sophisticated neural network is vulnerable when empirical material, on which its notion of normality is based, runs out. Zero-day vulnerabilities flare rarely, surface unpredictably, and evolve rapidly; consequently, labeled traces accumulate only sparingly. Models must infer regularities from fragmented telemetry shards, and insufficient example density yields either overly general or, conversely, excessively sensitive representations that degrade accuracy.

The limits delineated by label scarcity, behavioral drift, and hardware cost paradoxically define the trajectory of future evolution. A system that today separates baseline from intrusion must tomorrow repair its own perimeter, anticipate new classes of cryptographic levers, and breathe at the network's edge where latency is vanishingly small.

The first shoots of such self-sufficiency are visible in the concept of automatic vulnerability remediation by large language models. When dialogue models trained on code repositories detect a potential flaw, they produce a patch faster than a vendor can publish a formal vulnerability record. The result is an immune system for code: the signature is not yet born, but a temporary fix is already compiled and deployed in a sandbox.

The second direction is resilience to quantum crypto-anarchy. Neural networks integrated with post-quantum mathematics will encrypt their own weights and parameter exchanges so that even a quantum computer cannot extract useful information. Transitioning to such schemes will nullify an entire class of attacks on distributed-learning channels and preserve confidence in the intelligence even in an era of mass qubits.

Finally, digital twins of infrastructure elevate self-training. A virtual network copy lives a parallel life, continuously attacking itself with generated scenarios and immediately healing with discovered patches. The primary, physical network receives updated weights and rules without operator intervention. At the same time, humans watch metrics as if reading a ship's chronometer, with an invisible crew standing watch on every deck.

CONCLUSION

In conclusion, the analysis shows that deep models can convert reactive defense into proactive defense, compressing the window between intrusion and detection from days to near-instantaneous response. The foundation is neural networks' capacity to perceive the network as a dynamic ecosystem: convolutional filters process packet micro-signatures, attention mechanisms bind them into temporal chains, and graph layers reveal nontrivial paths of privilege movement. When this composite feature set encounters an exploit lacking a signature, the deviation appears as a sharp distributional shift, and the signal surfaces before data leave the defensive boundary.

However, effectiveness does not reduce to model architecture. Processes are critical: disciplined telemetry collection and versioning, continual retraining against concept drift, protection of weights from adversarial interference, and transparent reporting to operators. Without these elements, even a highly accurate network quickly loses relevance, and automated response becomes a source of false alarms. Deployment, therefore, is not a one-off engine integration, but a continuous cycle in which data, computation, and people constitute a closed feedback loop.

The frontiers of current solutions are defined by a paucity of labeled traces, hardware overhead, and the vulnerability of algorithms themselves to input poisoning. Yet, these very constraints chart the path forward. From self-generated patches that harden code before CVE disclosure, to post-quantum schemes for weight encryption, from inference at the NIC level to digital-twin pentesting, these directions, after analysis, appear not as futuristic curiosities but as logical continuations of the struggle to balance attack speed with response speed. Thus, deep learning for zero-day detection demonstrates not merely a new tool but a paradigm shift: defense becomes a living, self-learning system whose resilience is determined by adaptation velocity rather than the thickness of a prewritten rulebook.

REFERENCES

1. Agoramoorthy, M., Ali, A., Sujatha, D., Raj, M., & Ramesh, G. (2023). An Analysis of Signature-Based Components in Hybrid Intrusion Detection Systems. *Proceedings of 2023 Intelligent Computing and Control for Engineering and Business Systems*. https://doi.org/10.1109/iccebs58601.2023.10449209
2. Alder, S. (2025, April 23). *Verizon DBIR: Surge in Vulnerability Exploitation and Healthcare Espionage Breaches*. The HIPAA Journal. https://www.hipaajournal.com/verizon-dbir-2025/
3. Cisco. (2025, September). *Cisco IOS XE Software HTTP API Command Injection Vulnerability*. Cisco. https://www.cisco.com/c/en/us/support/docs/csa/cisco-sa-ios-xe-cmd-inject-rPJM8BGL.html
4. Crumb, P. (2025, July 26). *Data Privacy & Encryption Statistics 2025–26*. Compare Cheap SSL. https://comparecheapssl.com/data-privacy-encryption-statistics/
5. Google. (2025). *M-Trends 2025 Report*. Google. https://services.google.com/fh/files/misc/m-trends-2025-en.pdf
6. GVR. (n.d.). *Artificial Intelligence In Cybersecurity Market Size Report, 2030*. Grand View Research. Retrieved November 10, 2025, from https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-cybersecurity-market-report
7. IBM. (2025). *Cost of a data breach report 2025*. IBM. https://www.ibm.com/reports/data-breach
8. Ignacio, S., Alejandra, P., Ivan, D., & Allende, H. (2025). Zero-Day Threat Mitigation via Deep Learning in Cloud Environments. *Applied Sciences*, *15*(14), 7885. https://doi.org/10.3390/app15147885
9. Kapko, M. (2024). *Progress Software shakes off MOVEit's financial consequences, maintains customers*. Cybersecurity Dive. https://www.cybersecuritydive.com/news/progress-software-shakes-moveit-financial-impact/704900/
10. Lakshmanan, R. (2025, April 29). *Google Reports 75 Zero-Days Exploited in 2024 - 44% Targeted Enterprise Security Products*. The Hacker News. https://thehackernews.com/2025/04/google-reports-75-zero-days-exploited.html
11. Market Growth Reports. (2025). *Malware Analysis Tools Software Market Share Report*. Market Growth Reports. https://www.marketgrowthreports.com/market-reports/malware-analysis-tools-software-market-100172
12. Soureya, Y. G., Ngossaha, J. M., Djomadji, E. M. D., Amougou, N., Tsakou, S. B., & Ndjodo, M. F. (2025). Methodological Framework for Developing an Adaptive Intrusion Detection System (IDS) Incorporating Sustainability Factors. *Journal of Computer and Communications*, *13*(7), 171–203. https://doi.org/10.4236/jcc.2025.137009
13. Verizon Business. (2024). *DBIR Report 2024 - Summary of Findings*. Verizon Business. https://www.verizon.com/business/en-nl/resources/reports/dbir/2024/summary-of-findings/
14. Zorz, Z. (2025, January 9). *Ivanti Connect Secure zero-day exploited since mid-December (CVE-2025-0282)*. Help Net Security. https://www.helpnetsecurity.com/2025/01/09/ivanti-cve-2025-0282-zero-day-attacks-indicators-of-compromise/