iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
Peer-Reviewed | Open Access | International Journal
https://ijetrm.com/

ARTIFICIAL INTELLIGENCE FOR STRENGTHENING NATIONAL CYBERSECURITY: ADVANCED THREAT DETECTION AND AUTOMATED RESPONSE SYSTEMS

A comprehensive review of how AI-driven detection and response systems are reshaping the U.S. cybersecurity landscape

Muhammad Awais
Dept. of Computer Science Cumberland University, Tennessee, United States
mawais27@students.cumberland.edu
Muhammad Ayaz
Department of Electrical and Computer Engineering,
Pak-Austria Fachhochschule Institute of Applied Sciences and Technology,
muhammad.ayaz@paf-iast.edu.pk

ABSTRACT
The cybersecurity threat landscape confronting U.S. government agencies, critical infrastructure operators, and private enterprises has grown more complex, more automated, and more adversarial in nature. Attackers now use artificial intelligence (AI) to scale phishing campaigns, generate adaptive malware, and orchestrate multi-stage intrusions with minimal human oversight, while defenders increasingly rely on the same class of technology to detect anomalies, triage alerts, and automate incident response at machine speed. This article examines the dual-use nature of AI in the cybersecurity domain, tracing its evolution from rule-based detection to behavioral and generative AI-assisted defense. It surveys the architecture of modern AI-enabled Security Operations Centers (SOCs), reviews emerging U.S. federal guidance including NIST's draft Cybersecurity Framework Profile for Artificial Intelligence and CISA's agentic AI security guidance and evaluates measurable outcomes such as breach-cost reduction and faster detection-to-containment cycles. The discussion also addresses the risks introduced by AI itself: shadow AI, prompt injection, hallucinated detections, and compromised autonomous agents. The article concludes that AI is neither a silver bullet nor an optional upgrade, but a structural shift in how national cybersecurity must be governed, staffed, and audited going forward.

Keywords
Artificial Intelligence, Cybersecurity, Threat Detection, Automated Incident Response, Security Operations Center (SOC), National Infrastructure Protection, Agentic AI, NIST Cybersecurity Framework

1. INTRODUCTION
Cybersecurity has always been an arms race, but the terms of that race changed fundamentally once both attackers and defenders gained access to machine learning and generative AI. For decades, national cybersecurity strategy in the United States centered on perimeter defense, signature-based antivirus tools, and human analysts manually reviewing alerts inside a Security Operations Center (SOC). That model is buckling under the sheer volume and sophistication of modern threats. Nation-state actors, ransomware syndicates, and increasingly capable independent hackers are using AI to write malware variants that evade signature detection, to generate phishing emails indistinguishable from legitimate correspondence, and to automate reconnaissance across thousands of targets simultaneously.

At the same time, AI has become indispensable to the defensive side of the equation. Behavioral analytics engines can flag a compromised credential the moment it is used in a way that deviates from a user's normal pattern something a static, rule-based system would never catch. Automated response platforms can isolate an infected endpoint, revoke a session token, or block a malicious IP address within seconds of detection, long before a human analyst could act. According to industry-wide estimates, roughly 97% of organizations now use or plan to deploy AI enabled cybersecurity solutions, and defenders continue to expand automation into every layer of the security stack, from cloud workload protection to identity governance.

This shift is not confined to the private sector. National level cybersecurity agencies, including the Cybersecurity and Infrastructure Security Agency (CISA), the National Security Agency (NSA), and the National Institute of Standards and Technology (NIST), have all issued guidance in the past year addressing how AI should be secured, how it should be used defensively, and how it changes the risk calculus for critical infrastructure. The result is a

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
Peer-Reviewed | Open Access | International Journal
https://ijetrm.com/

rapidly evolving policy and technical environment in which "AI for cybersecurity" is no longer an emerging niche but a central pillar of national digital resilience.

The urgency behind this shift is not abstract. Critical infrastructure sectors energy, water, transportation, healthcare, and finance have each experienced a steady rise in digitally enabled disruption over the past several years, and the operational technology (OT) environments underlying much of this infrastructure were often designed decades before modern cybersecurity, let alone AI-driven defense, was a design consideration. Retrofitting AI based monitoring onto these legacy environments is technically difficult, but the alternative leaving them dependent on outdated, static detection is increasingly untenable given the pace at which AI assisted attackers can now probe for weaknesses. This tension between legacy infrastructure and modern, adaptive threats is a recurring theme throughout the remainder of this article.

1.1 Research Problem

Despite widespread AI adoption in security operations, organizations report a persistent paradox: even as AI improves detection speed and accuracy, breach rates and financial losses from cyber incidents continue to climb. This raises a core question is AI genuinely strengthening national cybersecurity, or is its defensive value being offset by AI-powered attacks and new AI-specific vulnerabilities? Understanding this tension is essential for policymakers, security leaders, and technologists designing the next generation of national cyber defense.

1.2 Objectives

This article aims to:

Trace the evolution of AI-based threat detection from static, rule-based systems to adaptive, behavior-driven models.

Examine the architecture and operational impact of automated incident response systems.

Review current U.S. federal policy and framework development around AI and cybersecurity.

Present quantitative evidence on the measurable benefits and residual risks of AI-driven defense.

Identify the emerging risks AI itself introduces into the national security ecosystem.

1.3 Significance

As critical infrastructure power grids, water systems, financial networks, and healthcare systems becomes more digitally interconnected, the stakes of getting AI-driven cybersecurity right extend well beyond individual companies. A well governed AI security posture can meaningfully reduce national exposure to catastrophic, cascading cyber incidents; a poorly governed one can introduce new single points of failure. This makes rigorous, evidence based analysis of AI's role in cybersecurity a matter of national interest.

2. LITERATURE REVIEW

2.1 From Signature Based to Behavioral Detection

Traditional intrusion detection relied heavily on signature matching: a system compared incoming traffic or files against a database of known malicious patterns. This approach is fast and low cost, but structurally incapable of catching novel or polymorphic threats (Ahmad et al., 2021; Lansky et al., 2021). The security research community has, over the past decade, documented a steady shift toward behavioral and anomaly-based detection, in which machine learning models establish a baseline of "normal" activity for a user, device, or network segment, and flag deviations from that baseline regardless of whether the specific attack pattern has been seen before.

This shift matters enormously for identity-based attacks, which have grown sharply as attackers increasingly rely on stolen credentials rather than malware to move through networks a technique that evades traditional file-based detection entirely. Research indicates AI-driven credential theft techniques have expanded rapidly, reinforcing the argument that detection strategy must center on behavior rather than static signatures (Moustafa et al., 2023).

Complementing these advances in cyber threat detection, recent research on critical energy infrastructure has applied machine learning and artificial neural networks to establish behavioral baselines and improve system resilience in operational technology environments. Ayaz et al. (2025a) demonstrated adaptive ANN-based voltage control strategies that enhance power-grid resilience under disturbances, while related work optimized voltage regulation using ANN-PID controllers for improved stability and disturbance rejection (Ayaz et al., 2025b). Additional contributions include AI-enabled energy load forecasting for smart-grid management (Arsalan et al., 2025), wavelet-transform time-frequency analysis for fault detection and classification in transmission systems (Baig et al., 2025), and machine-learning robust control solutions for power-electronic converters (Ayaz et al., 2024). Collectively, these studies illustrate how AI-driven behavioral modeling and anomaly-sensitive control in the energy sector provide both operational resilience and the foundational telemetry patterns that modern cybersecurity defenses can leverage for detecting cyber-physical attacks against national critical infrastructure.

2.2 The Rise of the AI Augmented SOC

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
Peer-Reviewed | Open Access | International Journal
https://ijetrm.com/

A second body of literature focuses on the transformation of the Security Operations Center itself. Historically, SOC analysts were flooded with alerts, the majority of which turned out to be false positives, leading to well documented "alert fatigue." AI-driven triage systems address this by scoring and prioritizing alerts based on contextual risk, correlating signals across multiple tools, and surfacing only the incidents most likely to represent genuine threats (Kinyua & Awuah, 2021). Industry data suggests AI-augmented SOCs detect threats roughly 50% faster and reduce analyst triage workload by around 60%, freeing scarce human expertise for complex investigations rather than repetitive alert review.

Natural-language interfaces layered on top of these systems exemplified by tools such as CrowdStrike's Charlotte AI and Microsoft Security Copilot further lower the barrier to effective threat investigation, allowing analysts to query complex telemetry without mastering an underlying query language. This lowers the expertise threshold required to operate a modern SOC, which is significant given the well-documented, chronic shortage of skilled cybersecurity professionals in the United States.

This transformation also changes what a SOC analyst's job actually looks like day to day. Rather than manually correlating logs across a dozen disconnected tools, analysts increasingly work through a single AI-mediated interface that has already performed the correlation and ranked the results by likely severity. Proponents argue this allows a smaller team to cover more ground without sacrificing coverage quality; critics counter that it risks producing a generation of analysts who are skilled at interpreting AI output but less practiced at independently reconstructing an attack chain from raw telemetry a skill that remains essential when an AI system's own output is wrong, incomplete, or itself the target of manipulation. Both perspectives appear in the literature, and neither is fully resolved by the data currently available.

2.3 Automated and Autonomous Response

A parallel research thread examines automated incident response Security Orchestration, Automation, and Response (SOAR) platforms that execute predefined playbooks (isolating a host, disabling an account, blocking an IP range) without waiting for human sign-off on routine, well-understood threat patterns (Alsmadi & Xu, 2019; Rathore & Kolhe, 2026). The literature is broadly consistent in finding that automation compresses the time between detection and containment, which is the single variable most strongly correlated with total breach cost. Data from IBM's long-running Cost of a Data Breach research indicates organizations with extensive AI and automation in their security operations experience average breach costs of roughly $3.62 million, compared with $5.52 million for organizations without such capabilities a difference of nearly $1.9 million per incident. The same research links automation to a reduction in breach identification and containment time of roughly 80 days on average.

2.4 The Offensive Use of AI

A growing and increasingly urgent strand of the literature addresses how attackers exploit the same underlying technology. Generative AI has proven highly effective at producing convincing phishing content: research on AI-enabled phishing detection techniques (Basit et al., 2021; Salloum et al., 2022; Thakur et al., 2023) documents that AI-generated phishing can achieve substantially higher click-through rates than traditional phishing at a fraction of the production cost, and a large majority of social-engineering attacks observed by security vendors are now assessed to be AI-assisted in some form. Beyond phishing, researchers have documented AI-assisted vulnerability discovery, adaptive malware capable of altering its own behavior to evade detection, and deepfake audio used for executive impersonation and fraud (Gupta et al., 2023).

A related and increasingly important thread concerns the entire attack chain, rather than any single technique in isolation. Earlier generations of cyberattacks typically required a human operator to manually pivot between reconnaissance, exploitation, and exfiltration, each stage introducing delay and opportunity for detection. Current-generation, AI-orchestrated attacks increasingly compress these stages, using generative models to draft convincing pretexts, automated scanning to identify exploitable weaknesses, and scripted follow-on actions to exfiltrate data all with minimal human involvement after the initial launch. This compression of the attack timeline is one of the primary reasons defenders have had to adopt equally automated response capabilities: a human-speed response process is structurally unable to keep pace with a machine-speed attack process.

2.5 Agentic AI as an Emerging Risk Category

The most recent literature emerging strongly in 2025 and 2026 identifies a distinct and rapidly escalating risk category: autonomous AI agents deployed inside enterprise and government environments. Because agentic systems can take actions (executing code, calling APIs, moving data) with limited human oversight, a compromised agent represents a fundamentally new class of insider-like threat. Leading AI developers, including OpenAI and Google DeepMind, have publicly flagged agentic AI security as a top near-term safety concern, and

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
Peer-Reviewed | Open Access | International Journal
https://ijetrm.com/

researchers have demonstrated that compromised agents can exfiltrate data, escalate privileges, and move laterally across a network without any human interaction in the loop. Security researchers estimate that a substantial majority of current enterprise security stacks are not yet equipped to detect this behavior, representing a significant and largely unaddressed gap in national cyber defense readiness.

2.6 Gaps in the Literature

Three gaps recur across this body of research. First, most quantitative benefit estimates (breach-cost reduction, detection-speed improvement) come from vendor-sponsored or vendor-adjacent research, and independent longitudinal validation remains limited. Second, governance and audit frameworks for AI security tools themselves are still immature relative to the pace of deployment most organizations report using generative AI or agentic AI in security operations well ahead of having formal governance controls in place (Malatji & Tolah, 2024). Third, there is limited empirical research quantifying how AI-driven attacks and AI-driven defenses interact dynamically over time, as opposed to static point-in-time comparisons. This article's discussion section returns to these gaps when considering implications for U.S. national cybersecurity policy.

3. METHODOLOGY

This article uses a narrative and quantitative synthesis methodology, drawing on publicly available industry research, U.S. federal agency publications, peer-reviewed academic literature, and trade-press sources published primarily between 2024 and 2026. The approach follows three stages:

1. Source identification Publications were drawn from recognized peer-reviewed journals (IEEE Access, Information Fusion, Applied Artificial Intelligence, Computers & Security), cybersecurity research organizations and vendors (IBM, CrowdStrike, Fortinet, Capgemini, Abnormal Security), U.S. federal agencies (NIST, CISA, NSA), and specialized cybersecurity policy and news outlets. Only sources published within the last 24 months were used for statistical claims, to ensure currency given how fast this field is moving.

2. Data categorization : Following best practice highlighted in recent industry analyses, statistics were labeled by type (AI-specific adoption data, general breach-cost benchmarks, SOC automation metrics, and threat-intelligence findings) so that AI-specific claims are not conflated with broader, non-AI cybersecurity statistics.

3. Cross-referencing : Where multiple sources reported overlapping figures (e.g., market size projections), a representative range or midpoint is presented rather than a single unverified figure, and discrepancies between sources are noted explicitly rather than silently reconciled.

This is a review-and-synthesis methodology rather than an original empirical study; as such, it does not involve primary data collection, human subjects, or a sampled population. Its "replicability" lies in the transparency of its sourcing and categorization approach, which future researchers can update as new reports are published.

4. RESULTS

4.1 Market Growth as an Adoption Indicator

Market sizing is an imperfect but useful proxy for how aggressively the AI-cybersecurity sector is scaling. Estimates vary by methodology and region, but the direction is consistent across every source reviewed: rapid, sustained growth.

Table 1. Global AI-in-Cybersecurity Market Size Estimates (Selected Sources, 2023–2030)

Year	Estimated Market Size (US$ Billions)	Source Basis
2023	$22.4B – $24.3B	Statista / industry composite
2025	$28.5B – $34.1B	MarketsandMarkets / industry composite
2026	~$25.5B	MarketsandMarkets
2030 (projected)	$93.75B – $134B	Statista / industry composite (range reflects differing methodologies)
2031 (projected)	~$50.8B	MarketsandMarkets (narrower scope estimate)

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
Peer-Reviewed | Open Access | International Journal
https://ijetrm.com/

Note: Figures diverge because different research firms scope "AI in cybersecurity" differently (e.g., including or excluding adjacent categories like identity management or cloud security posture management). The consistent signal across all sources is a compound annual growth rate in the low-to-high teens, reflecting sustained enterprise and government investment.

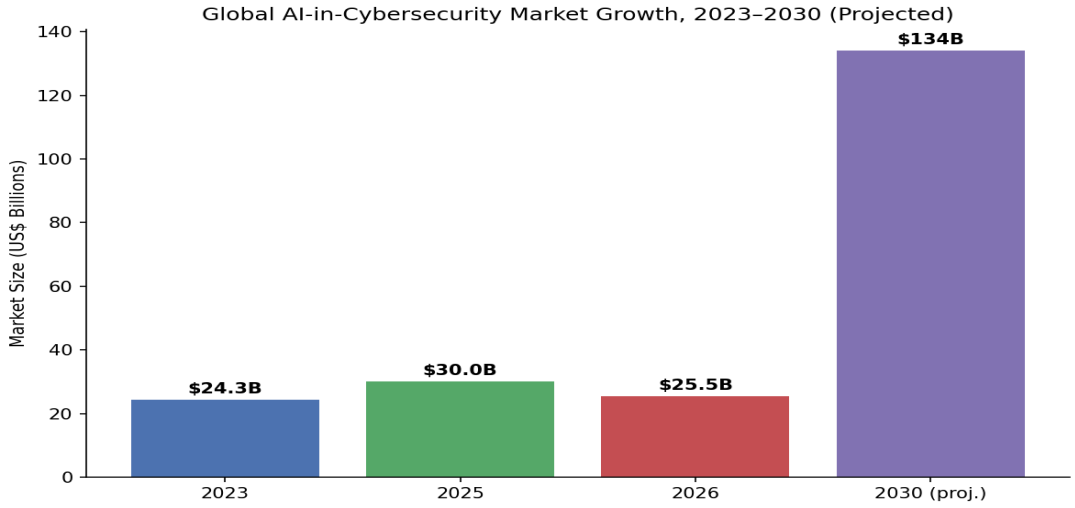

Figure 1. Illustrative trajectory of AI-in-cybersecurity market size, blending mid-range estimates across multiple industry sources. Actual figures vary by research methodology; see Table 1 for source ranges.

4.2 Measurable Defensive Impact

The most consistently cited and independently corroborated result across the literature concerns breach cost and response time. Organizations with mature AI and automation deployments in their security operations report materially better outcomes than those without.

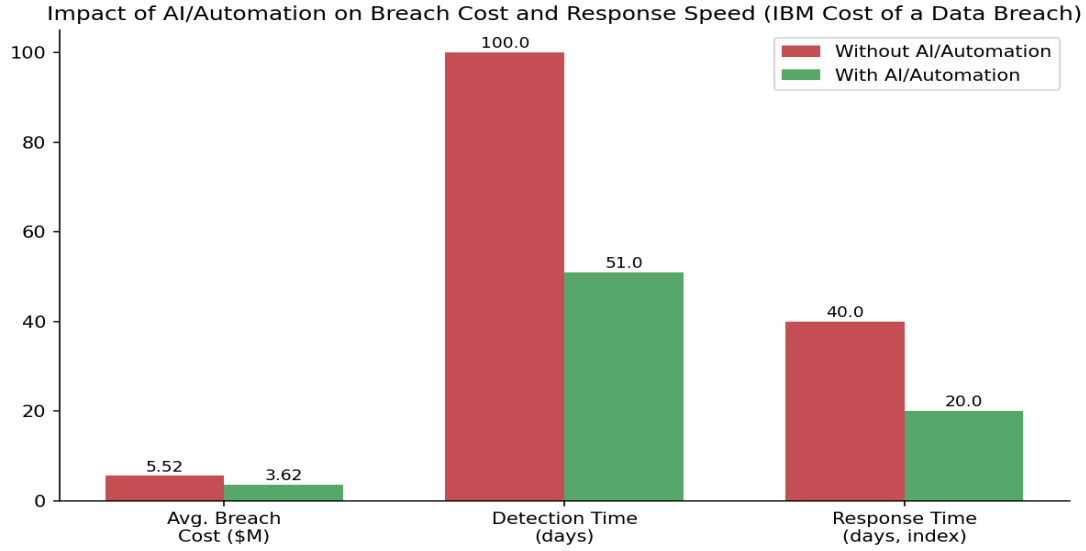

Figure 2. Comparative impact of AI/automation on average breach cost (US$ millions) and detection/response speed, based on IBM Cost of a Data Breach research. Response-time figures are presented as an indexed comparison reflecting the reported ~80-day reduction in combined identification and containment time.

Key findings from this line of research include:

Organizations with extensive AI and automation report average breach costs of approximately $3.62 million, compared to $5.52 million for organizations with little or no such capability.

AI-augmented SOCs report threat detection roughly 50% faster, alongside a 60% reduction in analyst triage workload.

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
Peer-Reviewed | Open Access | International Journal
https://ijetrm.com/

Combined AI-driven detection and response is associated with roughly 80 fewer days between breach occurrence and containment, compared to traditional, largely manual workflows.
Organizations with mature AI security deployments report detection accuracy approaching 95%, achieved with smaller analyst teams than traditional SOC staffing models required.

4.3 The Offensive Side: AI-Enabled Threats

The same underlying technologies are accelerating the offense. Reported findings include:
More than 80% of social-engineering attacks observed by specialized vendors are now assessed to involve some degree of AI assistance.
AI-generated phishing content has been found to achieve click-through rates in the range of 54%, at a small fraction of the cost of traditional phishing campaign production.
Credential-based (identity-focused) attacks as opposed to malware-based intrusions have grown sharply, a trend directly linked to attackers' use of AI for large-scale credential harvesting and validation.
Banking-sector cyberattacks have been reported to have risen by as much as 280%, reflecting the disproportionate targeting of the financial sector by AI-augmented threat actors.

Figure 3 summarizes the threat categories that security leaders identified as top concerns for 2026.

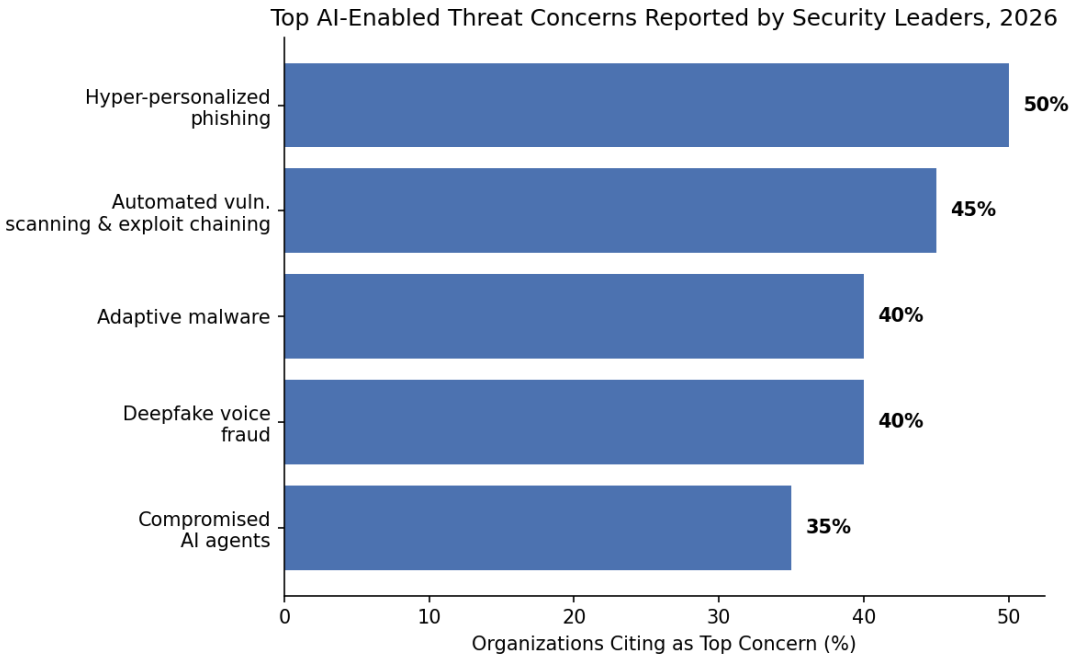

Figure 3. Share of surveyed security leaders citing each category as a top AI-enabled threat concern, based on industry survey data (State of AI Cybersecurity 2026 report and related commentary).

4.4 U.S. Federal Policy and Framework Development

Results from the policy review show a marked acceleration in federal guidance activity through late 2025 and into 2026:
In May 2025, the NSA's AI Security Center, CISA, and the FBI, together with allied international agencies, issued joint guidance on securing the data used to train and operate AI systems.
In December 2025, NIST released a preliminary draft of its Cybersecurity Framework Profile for Artificial Intelligence ("Cyber AI Profile"), organized around three focus areas — Secure (securing AI systems themselves), Defend (using AI to enhance cyber defense), and a third area addressing how organizations can proactively counter AI-enabled attacks — mapped against the six functions of the existing NIST Cybersecurity Framework 2.0 (Govern, Identify, Protect, Detect, Respond, Recover).
In February 2026, NIST's Center for AI Standards and Innovation (CAISI) launched an AI Agent Standards Initiative, the first U.S. government program focused explicitly on interoperability and security standards for autonomous AI agents.
On May 1, 2026, CISA and the NSA, together with counterpart agencies from Australia, Canada, New Zealand, and the United Kingdom, jointly published "Careful Adoption of Agentic AI Services" — the first coordinated

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
Peer-Reviewed | Open Access | International Journal
https://ijetrm.com/

multinational guidance addressing agentic AI risk specifically, defining five risk categories (privilege escalation, design and configuration failures, behavioral misalignment, structural brittleness, and accountability gaps) and calling for cryptographically verifiable, short-lived credentials for AI agents operating in sensitive environments. The EU AI Act, which took effect on August 2, 2026, includes provisions specifically addressing AI systems used in cybersecurity contexts, adding an international regulatory dimension that U.S. multinational organizations must also account for.

This cluster of activity indicates that U.S. federal guidance is shifting from general AI risk management toward specific, technical treatment of AI as both a defensive tool and an attack surface in its own right.

5. DISCUSSION

5.1 AI Is Winning Some Battles, Not the War

The results present a genuine paradox rather than a simple success story. AI driven detection and automated response demonstrably shorten the time between compromise and containment, and they measurably reduce the financial cost of the breaches that do occur. At the same time, overall breach frequency and the sophistication of attacks have not declined if anything, they have grown, driven substantially by attackers using the same class of technology. This is consistent with a broader pattern in cybersecurity history: defensive gains are rarely permanent, because they reshape attacker incentives and tactics rather than eliminating the contest altogether. The appropriate interpretation of the data, then, is not that "AI is failing" but that AI has raised the operational tempo on both sides simultaneously, making the quality of an organization's or a nation's AI governance the decisive variable, not mere AI adoption.

5.2 Behavioral Detection as a Structural Necessity, Not an Option

The literature is unambiguous that the shift from signature-based to behavioral and identity-centric detection is not a marginal improvement but a structural requirement going forward. As attackers rely increasingly on stolen, valid credentials rather than malware, detection systems that only inspect files or known indicators of compromise are structurally blind to a growing share of real-world intrusions. National cybersecurity strategy including sector-specific guidance for critical infrastructure operators should treat behavioral, identity-aware detection as baseline infrastructure rather than an advanced or optional capability.

5.3 Automation Compresses Risk, But Concentrates It

Automated response systems produce clear, measurable benefits in reducing containment time and total breach cost. However, this benefit comes with a structural trade-off: automation concentrates decision-making authority into a smaller number of systems, and a flaw, misconfiguration, or compromise in an automated response system can cause harm at machine speed, just as it can prevent harm at machine speed. This is precisely the concern reflected in CISA and NIST's growing focus on agentic AI risk privilege escalation, behavioral misalignment, and accountability gaps are not hypothetical concerns but the direct consequence of granting autonomous systems the ability to act without a human in the loop. Effective national cybersecurity policy must therefore pair automation adoption with proportional investment in monitoring, auditing, and constraining the automated systems themselves a form of "securing the securer."

In practice, this means treating an automated response platform with the same rigor traditionally reserved for the systems it protects: change management for its playbooks, least-privilege scoping for its credentials, comprehensive audit logging of every action it takes, and a tested rollback procedure for when it acts on a false positive. Several of the frameworks discussed in Section 4.4 particularly the CISA/NSA agentic AI guidance's emphasis on cryptographically verifiable, short-lived credentials are best understood as early, formalized attempts to codify exactly this kind of discipline at a national level, rather than leaving it to individual organizations to improvise inconsistently.

5.4 The Governance Gap

Perhaps the most consequential finding from the policy review is the sheer pace of guidance development relative to deployment. The rate at which organizations are adopting generative and agentic AI in security operations has significantly outpaced the rate at which formal governance frameworks have matured. NIST's Cyber AI Profile was still in draft and public-comment status as of early 2026, and the CISA/NSA agentic AI guidance was only issued in May 2026 long after autonomous agents had already been deployed across large portions of the private sector. This lag matters because it means many of the AI security tools currently protecting national infrastructure are operating without a settled, authoritative technical standard for how they should be configured, credentialed, or audited. Closing this governance gap should be treated as a national security priority in its own right, not merely a compliance exercise.

5.5 Workforce Implications

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
Peer-Reviewed | Open Access | International Journal
https://ijetrm.com/

A secondary but important theme is the changing shape of the cybersecurity workforce. Natural-language interfaces and AI-assisted triage lower the technical barrier to conducting sophisticated threat investigation, which has the potential to meaningfully ease the U.S. cybersecurity talent shortage. However, this same capability also raises the risk of over-reliance on AI-generated conclusions by less experienced analysts who may lack the background to critically evaluate a flawed or hallucinated detection. National workforce strategy should treat AI literacy and critical evaluation of AI-generated security findings as a core competency, not an optional add-on to traditional security training.

5.6 Sector-Specific Considerations

The benefits and risks of AI-driven cybersecurity are not evenly distributed across sectors, and national policy should account for this variation rather than treating "critical infrastructure" as a monolithic category. Financial services, for example, have both the resources and the regulatory pressure to adopt AI-driven fraud and intrusion detection quickly, which likely explains the disproportionate rise in reported attacks against banking targets noted in Section 4.3 attackers are, in effect, targeting the sector with the most valuable data and the most mature (and therefore most interesting to defeat) defensive AI. Healthcare systems, by contrast, often operate a mix of modern IT and older, harder-to-patch medical devices, where deploying AI based monitoring is technically feasible but automated response is far riskier: automatically isolating a compromised device that is actively supporting patient care carries a very different risk calculus than isolating a compromised laptop in a corporate office. Energy and water utilities present yet another profile, where operational technology networks are often intentionally isolated from the internet, limiting the immediate applicability of cloud-based AI security tools and requiring purpose-built, on-premises alternatives.

This variation argues for a layered national policy approach: baseline requirements and shared frameworks such as the NIST Cyber AI Profile that apply broadly, paired with sector-specific implementation guidance that accounts for differences in automation risk tolerance, legacy infrastructure constraints, and the practical consequences of a false positive. A one-size-fits-all mandate for automated response, for instance, would likely be appropriate for a cloud-native financial services firm and inappropriate for a hospital network, even though both fall under the broad umbrella of "critical infrastructure."

5.7 Toward Measurable National Resilience Metrics

A further implication of this review concerns measurement itself. Much of the quantitative evidence available today breach cost, detection time, response time is measured at the level of individual organizations, reported voluntarily, and often aggregated by vendors with a commercial stake in favorable results. There is comparatively little public, standardized reporting on how these individual improvements aggregate into national resilience: whether faster average detection times are actually reducing the frequency or severity of cascading, cross-sector incidents, as opposed to simply improving outcomes within isolated organizational boundaries. Building toward standardized, sector-level reporting — potentially coordinated through CISA in partnership with sector-specific regulators — would allow policymakers to evaluate whether the substantial, ongoing investment in AI-driven cybersecurity is translating into measurable national-level risk reduction, rather than relying on the kind of organization-level anecdotes and vendor statistics synthesized in this review.

5.8 Limitations

This review has several limitations. First, much of the quantitative evidence originates from cybersecurity vendors with a commercial interest in demonstrating AI's effectiveness, and independent, vendor-neutral longitudinal studies remain comparatively scarce. Second, figures such as market-size projections vary considerably by methodology and should be read as directional rather than precise. Third, because this is a fast-moving field, some of the policy developments discussed here (particularly NIST's Cyber AI Profile, which remained in draft status as of this writing) may have changed in scope or finalized form by the time of publication. Readers should verify the current status of any specific framework before relying on it for compliance purposes.

6. CONCLUSION

Artificial intelligence has become structurally embedded in both offensive and defensive cybersecurity operations, and this dual-use dynamic is now a defining feature of the U.S. national cybersecurity landscape. The evidence reviewed here shows that AI-driven detection and automated response systems deliver real, measurable benefits faster detection, faster containment, and materially lower breach costs while simultaneously enabling more scalable, more convincing, and harder-to-detect attacks. The net effect is not a resolved victory for either side but

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
Peer-Reviewed | Open Access | International Journal
https://ijetrm.com/

an elevated operational tempo that rewards organizations and agencies with mature AI governance and penalizes those that adopt AI tools without commensurate oversight.

For U.S. national cybersecurity policy specifically, four priorities emerge from this analysis: first, treat behavioral and identity-centric detection as baseline infrastructure across critical sectors rather than an advanced or optional upgrade; second, pair any expansion of automated and agentic AI response capability with equally serious investment in monitoring, credentialing, and constraining those systems; third, close the current gap between AI deployment speed and governance framework maturity, particularly around the newly identified risks associated with autonomous AI agents; and fourth, adapt implementation guidance to reflect genuine differences in automation risk tolerance across sectors, rather than applying a single national standard uniformly to environments as different as a cloud-native bank and a regional water utility.

None of this suggests that AI adoption in cybersecurity should slow down the cost, speed, and accuracy benefits documented throughout this review are real and substantial, and adversaries show no sign of slowing their own use of the technology. What the evidence does suggest is that the organizations, sectors, and national frameworks that treat AI governance as seriously as they treat AI adoption will be the ones that convert these technical capabilities into durable resilience, while those that adopt AI tools without corresponding oversight risk simply relocating their vulnerabilities rather than reducing them. Continued research particularly independent, longitudinal studies less reliant on vendor-sponsored data, and standardized sector-level reporting of the kind discussed in Section 5.7 will be essential to validate these trends as the underlying technology and threat landscape continue to evolve.

REFERENCES

1) Ahmad, Z., Khan, A. S., Shiang, C. W., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies, 32*(1), e4150. https://doi.org/10.1002/ett.4150
2) Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. *Frontiers in Computer Science, 3*, Article 563060. https://doi.org/10.3389/fcomp.2021.563060
3) Alsmadi, I., & Xu, H. (2019). Security automation in cybersecurity operations. *Computers & Security, 87*, Article 101568. https://doi.org/10.1016/j.cose.2019.101568
4) Basit, A., Zafar, M., Liu, X., Javed, A. R., Jalil, Z., & Kifayat, K. (2021). A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommunication Systems, 76*(1), 139–154. https://doi.org/10.1007/s11235-020-00733-2
5) Catal, C., Giray, G., Tekinerdogan, B., Kumar, S., & Shukla, S. (2022). Applications of deep learning for phishing detection: A systematic literature review. *Knowledge and Information Systems, 64*(6), 1457–1500. https://doi.org/10.1007/s10115-022-01672-x
6) Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. *IEEE Access, 11*, 80218–80245. https://doi.org/10.1109/ACCESS.2023.3300381
7) Karlzén, H., & Sommestad, T. (2023). Automatic incident response solutions: A review of proposed solutions' input and output. In *Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES '23)* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3600160.3605066
8) Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion, 97*, Article 101804. https://doi.org/10.1016/j.inffus.2023.101804
9) Kinyua, J., & Awuah, L. (2021). AI/ML in security orchestration, automation and response: Future research directions. *Intelligent Automation & Soft Computing, 28*(2), 527–545. https://doi.org/10.32604/iasc.2021.016240
10) Lansky, J., Ali, S., Mohammadi, M., Mohammed, A. H., Rashidi, S., Hosseinzadeh, M., & Rahmani, A. M. (2021). Deep learning-based intrusion detection systems: A systematic review. *IEEE Access, 9*, 101574–101599. https://doi.org/10.1109/ACCESS.2021.3097247
11) Malatji, M., & Tolah, A. (2024). Artificial intelligence (AI) cybersecurity dimensions: A comprehensive framework for understanding adversarial and offensive AI. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00427-4

iJETRM
INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGY RESEARCH & MANAGEMENT (IJETRM)
Peer-Reviewed | Open Access | International Journal
https://ijetrm.com/

12) Moustafa, N., Koroniotis, N., Keshk, M., Zomaya, A. Y., & Tari, Z. (2023). Explainable intrusion detection for cyber defences in the internet of things: Opportunities and solutions. *IEEE Communications Surveys & Tutorials, 25*(3), 1775–1807. https://doi.org/10.1109/COMST.2023.3280465
13) Ofusori, L., Bokaba, T., & Mhlongo, S. (2024). Artificial intelligence in cybersecurity: A comprehensive review and future direction. *Applied Artificial Intelligence, 38*(1), Article 2439609. https://doi.org/10.1080/08839514.2024.2439609
14) Rathore, P., & Kolhe, K. (2026). Integrating automation and orchestration in security incident handling: A review of SOAR frameworks and platforms. In M. Singh et al. (Eds.), *Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences* (Mechanisms and Machine Science, Vol. 185). Springer. https://doi.org/10.1007/978-3-031-95963-9_38
15) Rios-Campos, C., Paz, S. C. V., Vilema, G. O., Díaz, L. M. R., Zambrano, D. P. F., Zambrano, G. M. M., Viteri, J. D. C. L., Vara, F. E. O., Vallejos, P. A. A., Llontop, R. F. G., & Anchundia-Gómez, O. (2024). Cybersecurity and artificial intelligence (AI). *South Florida Journal of Development, 5*(8), Article e4276. https://doi.org/10.46932/sfjdv5n8-021
16) Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022). A systematic literature review on phishing email detection using natural language processing techniques. *IEEE Access, 10*, 65703–65727. https://doi.org/10.1109/ACCESS.2022.3183083
17) Santhosh Kumar, S. V. N., Selvi, M., & Kannan, A. (2023). A comprehensive survey on machine learning-based intrusion detection systems for secure communication in internet of things. *Computational Intelligence and Neuroscience, 2023*, Article 8981988. https://doi.org/10.1155/2023/8981988
18) Thakur, K., Ali, M. L., Obaidat, M. A., & Kamruzzaman, A. (2023). A systematic review on deep-learning-based phishing email detection. *Electronics, 12*(21), Article 4545. https://doi.org/10.3390/electronics12214545
19) Zeadally, S., Adi, E., Baig, Z., & Khan, I. A. (2020). Harnessing artificial intelligence capabilities to improve cybersecurity. *IEEE Access, 8*, 23817–23837. https://doi.org/10.1109/ACCESS.2020.2968045
20) Achuthan, K., Ramanathan, S., Srinivas, S., & Raman, R. (2024). Advancing cybersecurity and privacy with artificial intelligence: Current trends and future research directions. *Frontiers in Big Data, 7*, Article 1497535. https://doi.org/10.3389/fdata.2024.1497535
21) Ayaz, M., Baig, U., & Awais, M. (2025a). Enhancing power grid resilience with adaptive ANN-based voltage control strategies. In *2025 International Conference on Emerging Technologies in Electronics, Computing, and Communication (ICETECC)* (pp. 1–6). https://doi.org/10.1109/ICETECC65365.2025.11070231
22) Ayaz, M., Baig, U., Dawood, M., & Awais, M. (2025b). Optimizing voltage control in power systems: An artificial neural network approach for improved performance and stability. In *2025 International Conference on Emerging Power Technologies (ICEPT)* (pp. 1–6). https://doi.org/10.1109/ICEPT66058.2025.11036228
23) Arsalan, M., Ayaz, M., Ali, Y., & Baig, U. (2025). AI-enabled energy load forecasting for smart grid management. *International Journal of Scientific Research and Engineering Development, 8*(6).
24) Baig, U., Ayaz, M., & Baig, D. E. Z. (2025). Advanced time-frequency analysis for fault detection and classification in power transmission systems using wavelet transform. In *2025 International Conference on Emerging Power Technologies (ICEPT)* (pp. 1–6). https://doi.org/10.1109/ICEPT66058.2025.11036201
25) Ayaz, M., Rizvi, S. M. H., Rathi, L., Haq, M. A. U., & Ejaz, S. (2024). Machine learning based robust control solutions for buck converters: ANN and fuzzy logic methods. In *2024 21st Bhurban Conference on Applied Sciences and Technology (IBCAST)*.

