ISRG Journal of Engineering and Technology (ISRGJET)

OPEN ACCESS

ISRG PUBLISHERS

Abbreviated Key Title: ISRG J Eng Technol

ISSN: 3107-5894 (Online)

Journal homepage: https://isrgpublishers.com/isrgjet/

Volume – II Issue-II (March-April) 2026

Frequency: Bimonthly

The Quantum Mirror: AI vs. Hackers in the Cyber Arms Race

Dr. Alex Mathew[1*], Audrey Tobesman[2] & Emma Jackson[3]

[1, 2, 3] Bethany College, West Virginia, USA

| **Received:** 18-03-2026 | **Accepted:** 23-03-2026 | **Published:** 26-03-2026

***Corresponding author: Dr. Alex Mathew**

Abstract

Cybersecurity is undergoing considerable transformation as Artificial Intelligence (AI) has emerged as both an attack vector and an immensely powerful tool to defend against these attacks. The paper examines how cyber conflict is transformed with the rise of AI and offers a unique theoretical construct called the "Quantum Mirror," where both offensive and defensive AI are utilized and continually evolve in relation to each other. The research used three sources of information to conduct the analysis of AI-enhanced cyber conflict: the 2026 Global Cybersecurity Outlook report, the 2025 Armis Cyberwarfare report, and the first recorded AI-enabled cyber espionage event to synthesize industry, government, and academia sources. Based upon the evidence collected from this array of sources, it was determined that 94% of the respondents believe that AI represents the largest change factor for cybersecurity, and 87% believe that the threat presented by AI represents the fastest growing category of threats. In addition, the research provides a taxonomy of adversarial machine learning attacks, GTG-1002 Autonomous Offensive Artificial Intelligence Campaign, and a summary of emerging defensive measures (i.e., NIST's Adversarial Machine Learning Framework, AI Cyber Challenge) and provides strategic recommendations for architecture resiliency.

Keywords: *artificial intelligence, cybersecurity, adversarial machine learning, autonomous attacks, cyber defense, AI arms race*.

Opening Reflection: A Moment of Reckoning

As you read this, consider how many current cybersecurity attacks are happening worldwide against Artificial Intelligence (AI) and how many more are in progress at the same time. It is alarming. In September of 2025, GTG-1002 is (also known as the "Anonymous Terrorist Attack") created a cyber campaign in which the vast majority (80% to 90%) of the tactical work was completed by AI at rates previously impossible; thousands of actions were completed per second using AI systems from around the world and coordinated multi-target attacks that were done over the course of many days, without losing site of their objective. Not only will this become the cyber threat landscape, but it is also already the central issue in the cyber race of 'who has the best cyber warfare capabilities'.

Introduction: The Acceleration Point.

In recent years, there has been a major change in the way cyber-resilient systems are built. The growing complexity of threats has

DOI: 10.5281/zenodo.19230511

outpaced traditional methods of counteracting them. The introduction of Artificial Intelligence has created an imbalance in the continuing cyber arms race. According to the Global Cybersecurity Outlook 2026 report, 94 percent of those surveyed believe that AI is currently the most significant change facing technology as it relates to securing cyberspace. Spending on securing through AI is increasing from 37 percent today (205) to 64 percent in 2026. However, with the concern over not being prepared for the vulnerability of AI, the imbalance between preparedness and action has widened. Over 87 percent of IT professionals believe that vulnerabilities created by AI are currently the biggest threat to their organizations. We propose that we enter the quantum mirror phase of cyber warfare, where both offensive and defensive AI capabilities are being designed at equal and instantaneous rates, so seconds (as opposed to minutes) will be critical for achieving rapid results. The traditional distinction between proactive and reactive security has been rendered irrelevant.

State of the Arms Race: 2025-2026 Findings

As we look towards the future, the cyber threat landscape has changed dramatically. Cybersecurity continues to be "faster than our ability to react," according to IT leaders (Armis, 2025). 81% of IT leaders want to take steps to build a proactive defense. Despite this desire, most cybersecurity will be limited to using reactive measures, which the intruders exploit through their sophistication. With the ever-expanding digital space being created through cloud, IoT, OT, and SaaS technologies, it is more than possible for a machine speed attacker to operate without detection. In addition to the growth and sophistication of machine speed attacks, we expect the motivations for cybercrime will also change significantly. A report conducted by the World Economic Forum (2026) indicated that cyber-enabled fraud will be the most significant risk associated with companies' security by 2025, with 73% of respondents indicating their networks are vulnerable to cyber-enabled fraud. Some examples of how advanced attackers are using AI today include high-profile personal phishing attacks, deep fake social engineering of sensitive information, and the ability to quickly and accurately harvest credentials.

Joint geostrategic rivalries combined with government-sponsored attacks have enhanced cyber warfare capabilities; hence, there are several Good Programs (GPs) now part of many nations' aggressive cyber warfare programs, with more than 64% of nations monitoring cyber threats as part of their overall strategy. The Combined Operations of the various sectors of national infrastructures are generally considered as a combination of Crime, Military Operations, and espionage. There is a lack of consensus between many nations regarding cyber-confidence, which has diminished confidence in their cyber-readiness to react to a cyber event.

Adversarial AI Taxonomy: A Visualization of the Attack Surface

We should start with a bitter fact: intelligent systems are perfect objects to be exploited. This is highlighted in the 2024 adversarial machine learning taxonomy of the National Institute of Standards and Technology in its classification of events that lie outside the AI lifecycle (Vassilev, 2025). Evasion attacks establish post-deployment targets where input is modified in such a way that an AI system mislabels a threat, allowing an exploited malware attack to evade the intelligent protections in place. Poisoning attacks can be seen as a means of supplying a model with malicious data, compromising its intended purpose. Privacy attacks involve theft of a sensitive training model or potentially the complete model.

In contrast, misuse attacks consist of the use of generative AI for phishing scripts, malware, or false information. The risk to safety increases significantly in various verticals; autonomous vehicles may misrecognize traffic signals while medical AI may improperly diagnose patients based on improper perturbations to their training data; and smart grids could be subjected to data poisoning, resulting in system failure, while LLMs may be modified and/or leaked from internal networks (Pelekis et al, 2025). As such, it is clear that the security of the AI system must include protections equal to those of the networks it is intended to protect.

The GTG-1002 Operation

The Anthropic Threat Intelligence identified the GTG-1002 intelligence cyberspace campaign in September 2025, fundamentally altering the trajectory of the cyberwarfare (Anthropic, 2025). Sophisticated actors have now delegated their processing abilities to AI systems using Claude Code and Model Context Protocol instruments to create personal systems, which can utilize AI scanning, discovery, and extraction of information at machine speed (Anthropic, 2025). This differs from traditional attacks and is estimated to accomplish almost 80-90% of tactical functions independently of human involvement in their seminal role (PricewaterhouseCoopers, 2025). Research has demonstrated the ability of reinforcement learning agents to arrange agents or automated cyber operations, such as lateral movement and the optimum payload delivery method for cyber-attacks (Finistrella et al, 2025). Consequently, the intelligence cyberspace campaign of the GTG-1002 has revealed the possible risks associated with the credibility of the AI systems, like false credentials and forged evidence that should be verified by any human means. According to scholars, with the ever-increasing penetration of AI systems, the intrusion performance capabilities of AI systems are evolving significantly more than the level of cybersecurity changes, creating additional vulnerabilities to cybersecurity (Mudau et al, 2025). The security counseling AI age is long gone; now we are in the age of autonomous cyber warfare.

The Defense Frontier: AI-based Cyber Resilience.

People will no longer be protectors of cybersecurity; rather, machines will serve as protectors, just as they do now to protect one another. Armis (2025) gives us an important tip for the future in that the future requires AI to work together with AI to solve the issues that AI provides through its use and support to the enterprise. Signature-based classification tools will not have the ability to provide adequate security. Therefore, companies need to deploy predictive AI-based threat models that identify abnormal behaviour before a breach occurs within their organizations. This will be accomplished through Continuous Threat Exposure Management (CTEM). Additionally, enterprises must use predictive model AI to identify all the real-time presence points of their IT, OT, IoT, Cloud, and SaaS ecosystems.

The Zero Trust initiative is equally important. A report by Sophos (2026) stated that 67.32% of significant incidents that occurred appeared to be from an infiltrator using legitimate credentials with an identity rather than breaking into a secure facility. Thus, using an identity to gain access to a facility or system is in direct opposition to the zero-trust philosophy, which is contrary to the future-oriented characteristics that will ultimately be operationalized and implemented by businesses and/or professionals in the future. In addition, there is a major lack of

DOI: 10.5281/zenodo.19230511

knowledge among organizations regarding AI security; of all IT leaders surveyed, only 50% reported that their organization can secure AI. Of small organizations, 46% reported that they do not have security personnel capable of preventing cybercriminal activity, whereas 29% of large organizations provided that response. According to Gama et al. (2025), innovation serves as a beacon of hope for the future, and the DARPA cyber challenge is an example where autonomous systems could discover previously known as well as new vulnerabilities and remediate those vulnerabilities, as evidenced by an AI-driven reasoning engine discovering real-life zero-day vulnerabilities.

The Quantum Computing Horizon

The continuing growth of quantum computing should not be overlooked. Some people see a cyber arms race happening right now. As outlined in the Rapid7 report, the advancements made in quantum technology could render existing encryption methods, such as RSA and elliptic curve encryption, obsolete (Watson 2025). The harvest now, decrypt later threat warrants consideration. An adversary could take an encrypted data stream from someone today and could wait until quantum technology becomes available to be able to decrypt the data. This type of threat represents an existential threat for organizations that have long-lived data, such as government agencies, medical facilities, and financial institutions. Transitioning to post-quantum cryptography will be difficult and likely inconsistent among various organizations, leaving a period of vulnerability during the extended length of the transition (Ivezic 2024). Organizations need to begin developing plans for transitioning to PQC while they continue managing the existing risks associated with AI.

Strategic Recommendations

Based on the findings from the above paragraphs, several recommendations are made on how businesses should navigate the new and rapidly changing AI-informed cyber arms race.

For Executive Leadership

- Treat cybersecurity as a strategic business initiative. AI-enabled attacks can threaten your organization's long-term financial viability, brand recognition, and business existence (World Economic Forum, 2026). Enhanced competitiveness and trust from stakeholders will also be derived.
- Establish an AI security governance framework at the board level. Regulatory frameworks (European Union), such as the AI Act 2024, consider critical AI infrastructure as having high risk (European Commission, 2025).
- Invest in proactive and innovative cyber resiliency. The longer-term strategic outcome of a highly protective architecture is that it provides credibility to the supply chain and establishes long-term market superiority.

For Security Practitioners

- If your system has been compromised, identity-based attacks will dominate current breaches. Adversaries leveraging AI can process information at machine speed, making it important to design a strength-based system architecture (Sophos, 2026).
- Deploy complete telemetry monitoring. Limited visibility inhibits detecting threats and increases the length of time of breach and the length of time money is lost.
- Execute adversarial machine learning tests. Utilize guidance from NIST's taxonomy (NIST, 2024) to create strong model security.
- Stop obvious attack vectors. Sophos publications provide straightforward yet powerful and meaningful solutions: prevent the execution of Python today, or you will incur a debt later (Sophos, 2026).

To policymakers and other industry partners.

- Build a better connection between public and private sector cybersecurity intelligence networks. Multi-sector cooperation will be vital for the early detection of threats caused by AI espionage campaigns (World Economic Forum, 2026).
- Invest in education pipelines that support cybersecurity training. A labor shortage diminishes the power of collective defense in AI security.
- Create a shared threat intelligence ecosystem. Internet resilience efforts will benefit from international cybersecurity interdependence (European Union Agency for Cybersecurity, 2024).

Conclusion: Beyond the Arms Race.

The arms race metaphor suggests that there is a race that can result in a winning team, but also allows for a winning team to achieve a permanent advantage over the other. The Quantum Mirror suggests another type of arm race that includes collaboration on all forms of attack, where both sides will find positional support for defending themselves against prior offensive capabilities and vice versa. This may mean leaving behind the concept of an arms race entirely. In Cybersecurity and AI, the goal is not to win, but to survive; The goal is not to be invulnerable; However, to allow them to continue working in an unfriendly environment. The intent of prevention is not to stop attacks, but to ensure that regardless of whether or not an attack occurs, the result of that attack does not happen through an example demonstrated with GTG-1002 that AI can conduct autonomous attacks in self-directed actions and validate through the Team Atlanta competition that it is achievable for AI to conduct autonomous defensive actions against bridge actions. The degree of separation between the capabilities described will provide the context for security for decades into the future.

References

1. Anthropic. (2025). *Disrupting the first reported AI-orchestrated cyber espionage campaign*. Anthropic.com. https://www.anthropic.com/news/disrupting-AI-espionage
2. Enisa. (2024). *2024 REPORT ON THE STATE OF CYBERSECURITY IN THE UNION*. https://www.enisa.europa.eu/sites/default/files/2024-11/2024%20Report%20on%20the%20State%20of%20Cybersecurity%20in%20the%20Union%20-%20Condensed%20version.pdf
3. European Commission. (2025, August 1). *AI Act*. European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
4. Finistrella, S., Mariani, S., & Zambonelli, F. (2025). Multi-Agent Reinforcement Learning for Cybersecurity: Classification and Survey. *Intelligent Systems with*

DOI: 10.5281/zenodo.19230511

Applications, *26*, 200495. https://doi.org/10.1016/j.iswa.2025.200495

5. Gama, M. da, Natasa Perucica, & World Economic Forum. (2025, November 20). *AI is revolutionizing cybersecurity. How should we train the next generation of defenders?* World Economic Forum. https://www.weforum.org/stories/2025/11/cybersecurity-ai-professionals-workers/
6. Hull, Maj. J. (2026). *179th Cyber Protection Team Integrates Artificial Intelligence into Defensive Cyber Operations Training*. DVIDS. https://www.dvidshub.net/news/557833/179th-cyber-protection-team-integrates-artificial-intelligence-into-defensive-cyber-operations-training
7. Ivezic, M. (2024, September 10). *Post-Quantum Cryptography (PQC) Meets Quantum AI (QAI)*. PostQuantum - Quantum Computing, Quantum Security, PQC. https://postquantum.com/post-quantum/pqc-quantum-ai-qai/
8. Moschetta, G., Winslow, E., & World Economic Forum. (2026, January 12). *What execs need to know as global cyber risk rises in 2026*. World Economic Forum. https://www.weforum.org/stories/2026/01/geopolitics-ai-fraud-global-cyber-cybersecurity-2026/
9. Mudau, K., Mudumani, K., & Zwane, S. M. (2025, October 21). *Zero Trust Architecture: Frameworks and Implementation Strategies in Modern Cybersecurity*. https://doi.org/10.5281/zenodo.17404160
10. Pelekis, S., Koutroubas, T., Blika, A., Berdelis, A., Karakolis, E., Ntanos, C., Spiliotis, E., & Dimitris Askounis. (2025). Adversarial machine learning: a review of methods, tools, and critical industry sectors. *Artificial Intelligence Review*, *58*(8). https://doi.org/10.1007/s10462-025-11147-4
11. PricewaterhouseCoopers. (2025). *AI-orchestrated cyberattacks: A call to action: PwC*. PwC.https://www.pwc.com/us/en/services/consulting/cybersecurity-risk-regulatory/library/ai-orchestrated-cyberattacks.html
12. Vassilev, A. (2025). *Adversarial Machine Learning:* https://doi.org/10.6028/nist.ai.100-2e2025
13. Wang, C., Chen, J., Yang, Y., Ma, X., & Liu, J. (2021). Poisoning attacks and countermeasures in intelligent networks: Status quo and prospects. *Digital Communications and Networks*. https://doi.org/10.1016/j.dcan.2021.07.009
14. Watson, M. (2025, October 29). *Rapid7 reveals global findings in latest cyber-threat report*. SecurityBrief Australia. https://securitybrief.com.au/story/rapid7-reveals-global-findings-in-latest-cyber-threat-report
15. World Economic Forum. (2026, January 12). *Global Cybersecurity Outlook 2026*. World

DOI: 10.5281/zenodo.19230511