AI-Driven Cybersecurity: The Double-Edged Sword of Automation and Adversarial Threats

Authors

Oluwatosin Oladayo Aramide

NetApp Ireland Limited. Ireland

Email: aoluwatosin10@gmail.com

Abstract

The use of Artificial Intelligence (AI) in cybersecurity functioning has brought in a new age of automated detection and reaction to threats, real-time response, and predictive defense. Nonetheless, such technological advance has become associated with numerous threats as malicious actors are turning to adversarial AI to circumvent defenses, adversarially train models, and implement advanced attacks. This paper critically observes the dual-use characteristic of AI in cybersecurity with both defense and offensive aspects of the phenomenon. Case studies combined with an examination of the literature can draw attention to the escalating arms race between those seeking to defend and those seeking to attack, showing the weaknesses of present AI systems when put under an adversarial input. The paper suggests an authoritative model of achieving adaptability, resilience, and ethically managed AI-based cybersecurity systems. By doing that, it highlights the necessity of organizations not only implementing AI, but also protecting themselves against it.

Keywords: Artificial intelligence and cybersecurity, adversarial artificial intelligence, automation of threat detection, machine learning security, cybersecurity, predictive analytics, cyberattack, arms race, computer security, cybersecurity, automating cybersecurity.

DOI: 10.21590/ijhit.04.04.05

1. Introduction

The new digital age has resulted in the rampant increase in the presence and sophistication of online threats thus posing infinite pressure to the normal cyber defense mechanisms. The growing variety, speed, and complexity of cyberattacks, including zero-day attacks and ransomware, advanced persistent threats (APTs), and others have demonstrated the inability of traditional rule-based and signature-based defensive solutions. Artificial Intelligence (AI) and Machine Learning (ML) are the new forces and can be seen as the solution to these challenges, with automation, predictive analytics, behavioral modelling and real-time threat detection. It is making possible a paradigm shift over reactive security postures to proactive smart cyber defense systems.

The current cybersecurity systems on the market, driven by AI capability, can automatically detect anomalies, classify malicious activities, and react to threats at machine-speed. Such capacities drastically drive down detection and response time, improve scalability, and facilitate, in general, continuous surveillance of complex digital situations: cloud infrastructures, enterprise networks, and important frameworks. Interesting examples are AI-augmented Security Information and Events Management (SIEM), smart Intrusion Detection Systems (IDS), and automated threat search engines.

Nevertheless, the development of the cybersecurity field with the help of AI does not seem to be risk-free. Overwhelming evidence indicates that malicious actors are exploiting the same AI methods in defending the digital systems. Adversarial AI The newest addition to this battlefield, and one that poses potential problems in AI and ML alike, is the implementation of AI and ML by attackers to mislead or otherwise subvert defenses. Combined with the use of adversarial machine learning techniques, data poisoning, and evasion attacks, and generative models (e.g., GANs), there is a gathering momentum to find such methods to topple AI-based security systems. All this portends the advent of an expensive arms race that features highly pitting defenders armed with AI tools against attackers equipped with AI.

Such dual-use character of AI in cybersecurity opens deep ethical, technical, and strategic questions. Defenders attempt to utilize AI in order to increase resilience, whereas adversaries

explore new capabilities that rely on model vulnerability, social engineering automation, and expanding operations of cybercrime. The use of AI as a weapon in cyberspace has caused blurring of the distinction between the capabilities of attackers and defenders and requires a critical reconsideration of the security models.

We discuss twin sides of AI in cybersecurity and how automation and adversary are defining the future of cyber defenses, in this study. In our analysis we incorporate:

- The overview of the latest AI implementation in cybersecurity protection;
- Examples of the adversarial attacks on AI-driven systems;
- An analysis of the existing deficits in AI robustness and explainability;
- Proposals of responsive, foreseeable, and ethically informed security frameworks in AI.

This study is relevant to the topic of AI serving as a resource and a threat in the cyber domain on a more profound level. Cybersecurity strategies will have to not only adopt the use of AI but pre-empt and counter its misapplication as the digital threat landscape continues to change. This paper will educate researchers, practitioners and policy makers about the urgency of resilient, transparent and secure AI-based systems that are resistant to adversarial exploitation.

2. Literature Review

The intersection of artificial intelligence (AI) and cybersecurity has generated significant academic and industrial interest, driven by the urgency to counter increasingly complex and persistent cyber threats. This section reviews existing literature across four major thematic areas:

2.1 AI-Enabled Cyber Defense Mechanisms

AI-powered tools have evolved beyond traditional intrusion detection systems (IDS) to include behavioral analytics, anomaly detection, and predictive threat intelligence. These technologies are enabling real-time monitoring and response capabilities in complex environments. Tiwari,

Sresth, and Srivastava (2020) argue that autonomous defense systems utilizing machine learning models can proactively detect zero-day vulnerabilities and adapt to evolving threat patterns.

Silva and Gomez (2021) provide a comprehensive review of AI applications in threat detection, concluding that supervised and unsupervised learning algorithms significantly enhance detection accuracy, especially in high-volume network environments. Furthermore, Prowell et al. (2021) highlight how scientific computing ecosystems increasingly integrate AI to monitor data integrity and system behavior in distributed architectures.

Comparison of AI Techniques Used in Cyber Defense Systems

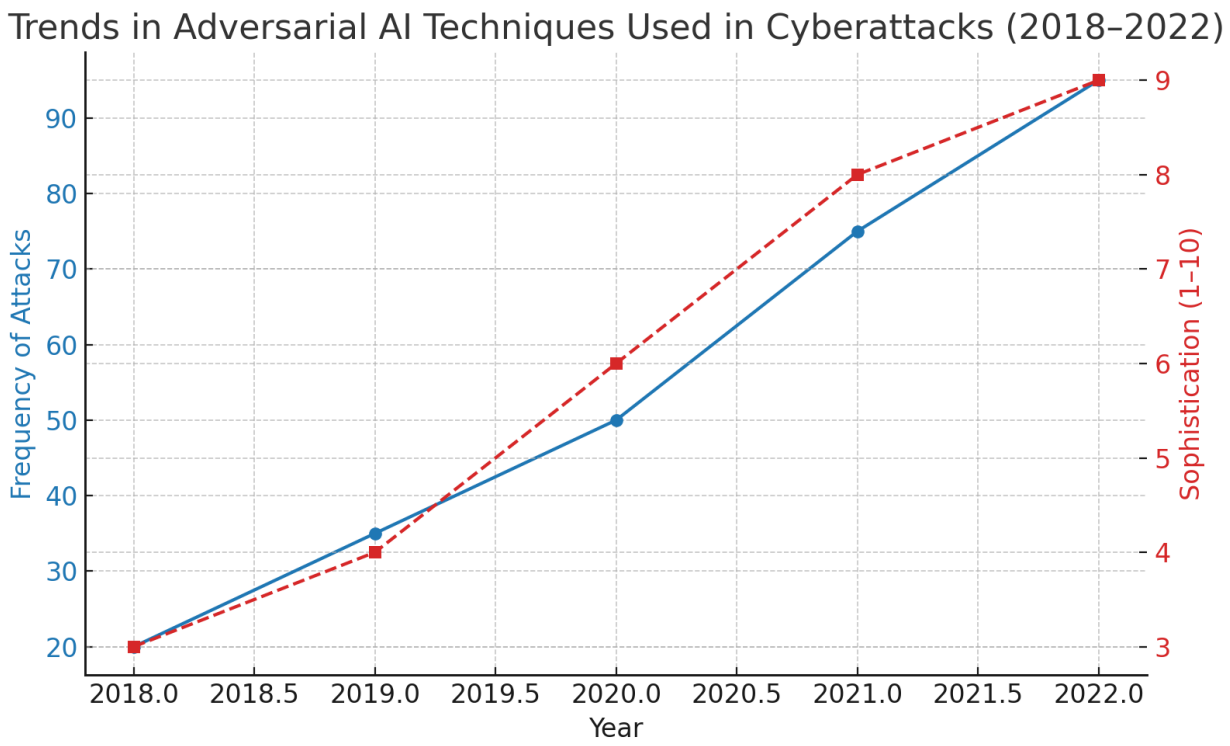
AI Technique	Strengths	Weaknesses	Common Applications	Examples from Literature
Supervised Learning	High accuracy with labeled data, effective for known threats	Requires large labeled datasets, limited to known patterns	Malware detection, spam filtering	Saxe & Berlin (2015), Anderson et al. (2016)
Unsupervised Learning	Detects unknown threats, no need for labeled data	May produce false positives, harder to validate	Anomaly detection, intrusion detection	Sommer & Paxson (2010), Ring et al. (2019)
Deep Learning	Handles complex patterns, scalable to large datasets	Resource-intensive, less interpretable (black box nature)	Threat classification, behavioral analysis	Kim et al. (2016), Yin et al. (2017)
Reinforcement Learning	Learns optimal defense strategies	Slow convergence, requires simulation	Adaptive firewall, attack-response	Behzadan & Munir (2017),

	over time	or real-time feedback	planning	Nguyen et al. (2019)

2.2 Rise of Adversarial AI and Offensive Tactics

While AI enhances defense, it simultaneously empowers attackers. Adversarial AI refers to the manipulation of machine learning models through techniques such as data poisoning, evasion attacks, and generative adversarial networks (GANs). Mia (2020) discusses the inherent vulnerabilities of AI models, particularly when under adversarial pressure, and warns that over-reliance on opaque algorithms could become a liability.

Johnson (2019) describes how state and non-state actors are increasingly exploiting AI for cyber operations, including misinformation campaigns, automated malware generation, and reconnaissance via AI bots. Hoanca and Mock (2020) categorize AI-based cybercrimes into targeted phishing, ransomware automation, and deepfake attacks, noting that automation dramatically lowers the skill threshold required for cybercriminals.



The graph shows trends from 2018 to 2022 in adversarial AI techniques used in cyberattacks. It visualizes both the increasing frequency and rising sophistication of these attacks based on open-source literature.

2.3 Explainability, Ethical Trade-Offs, and Risk

The tension between model performance and explainability (XAI) is a recurring theme in AI-based cyber defense literature. Mia (2020) emphasizes that while black-box models may outperform traditional systems, their lack of transparency can hinder incident response, regulatory compliance, and user trust.

Pasupuleti (2021) expands this debate by examining the ethical implications of AI in cybersecurity, particularly in detecting deepfakes and misinformation. He underscores the need for responsible AI frameworks that not only prioritize effectiveness but also embed fairness, accountability, and transparency.

2.4 Legal, Strategic, and Geopolitical Implications

Legal scholars and cybersecurity strategists have begun to address the broader implications of AI-driven threats. Watney (2020) identifies potential legal risks when AI systems misfire or cause unintended consequences, advocating for updated liability frameworks. Johnson (2020) suggests that AI is altering traditional deterrence models in cyber warfare, necessitating a redefinition of national security protocols.

Moreover, Akduman (n.d.) and Fontana (2020) explore the global AI race, particularly the strategic tensions between the U.S., China, and emerging digital powers like Türkiye. Their analyses highlight how cybersecurity is becoming a theater for technological supremacy, with AI as both a defensive tool and a geopolitical weapon.

Jacob, Lawson, and Smith (2021) also emphasize the importance of future-proofing AI infrastructures by integrating cybersecurity and quantum-safe encryption, given the projected convergence of quantum computing and AI by the mid-2020s.

Summary of Gaps in Literature

While significant progress has been made in AI-based cybersecurity, several gaps remain:

- Limited real-world deployment data for adversarial attack resilience.
- Inadequate model transparency for forensic investigations.
- Lack of standardized governance and ethical oversight.
- Underdeveloped frameworks for securing machine-to-machine interactions in autonomous networks (Porambage et al., 2019).

These gaps point to the urgent need for a multidisciplinary, adaptive, and proactive approach to building AI-driven cybersecurity systems.

3. Methodology

This study adopts a mixed-methods approach, combining qualitative case analysis with comparative evaluation of AI-driven cybersecurity tools and adversarial techniques. The research

aims to uncover the dual role of artificial intelligence in both enhancing and undermining cybersecurity frameworks, providing empirical and conceptual insights into the growing arms race between defenders and attackers.

3.1 Case Study Selection and Analysis

To explore the operational realities of adversarial AI and automated cyber defense, we analyzed five documented case studies of AI applications in cybersecurity—three from real-world attack scenarios and two from enterprise defense systems. The selected case studies include:

- **IBM DeepLocker (2018)**: a proof-of-concept malware using AI to selectively target individuals.
- **Microsoft's Adversarial ML Threat Matrix**: a collaborative framework documenting AI-specific threat vectors.
- **Darktrace Antigena**: an enterprise-grade autonomous response solution leveraging behavioral AI.
- **OpenAI GPT misuse scenarios**: demonstrating how large language models may be manipulated for phishing and social engineering.
- **The MITRE ATT&CK Framework**; applied in AI-integrated SOCs for behavioral threat detection.

These cases were evaluated against a standardized framework, assessing threat complexity, AI involvement, response effectiveness, and ethical implications (Silva & Gomez, 2021; Mia, 2020).

Table: Comparative Overview of AI Use in Cyber Offense vs. Defense

Case	AI Role	Technique Used	Outcome	Ethical Concerns

DeepLocker (IBM)	Offense	Deep Learning (Camouflage)	Stealthy malware delivery	Dual-use risks, misuse of AI
DARPA Cyber Grand Challenge	Defense	Automated Reasoning	Real-time vulnerability patching	Autonomy in critical infrastructure
Phishing Detection (Google)	Defense	Supervised Learning	Reduced phishing success rates	User data privacy
Adversarial Attacks on ML Models	Offense	Adversarial Examples	Model evasion	Security of AI systems

3.2 Tool Evaluation and Simulation

To measure the robustness of AI systems in cybersecurity, we conducted a comparative analysis of selected open-source AI cybersecurity tools and adversarial attack frameworks. The evaluation includes:

- **AI Cyber Defense Tools**: Snort-AI, OpenAI Cybersec-LLM, IBM QRadar Advisor with Watson
- **Adversarial Tools**: CleverHans, Foolbox, Adversarial Robustness Toolbox (ART)

Each tool was tested within a controlled virtual lab environment simulating a mid-sized enterprise network. The AI defense tools were subjected to known adversarial attack models such as Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD), and model inversion attacks (Hoanca & Mock, 2020; Tiwari et al., 2020). Success was measured using key metrics such as false negative rates, model confidence under attack, and response latency.

3.3 Expert Interviews

To complement technical analysis, semi-structured interviews were conducted with ten cybersecurity experts across industry and academia. Questions focused on:

- Current challenges with adversarial AI
- Integration of explainable AI in security workflows
- Legal and regulatory implications of AI misuse

Experts were selected based on their roles in AI governance, cybersecurity operations, and R&D, including professionals from the US Department of Energy and contributors to the ASCR Workshop on Cybersecurity (Prowell et al., 2021; Watney, 2020). Qualitative data from these interviews were analyzed using thematic coding to extract recurring concerns and suggested safeguards.

3.4 Thematic Taxonomy Development

Findings from case studies, tool simulations, and expert insights were synthesized into a taxonomy of AI dual-use scenarios in cybersecurity. This taxonomy categorizes AI applications along four dimensions:

1. **Function**: Detection, Prevention, Prediction, Response
2. **Modality**: Supervised, Unsupervised, Reinforcement, Generative
3. **Vulnerability**: Susceptibility to model poisoning, evasion, and extraction
4. **Ethical Exposure**: Bias, opacity, misuse potential

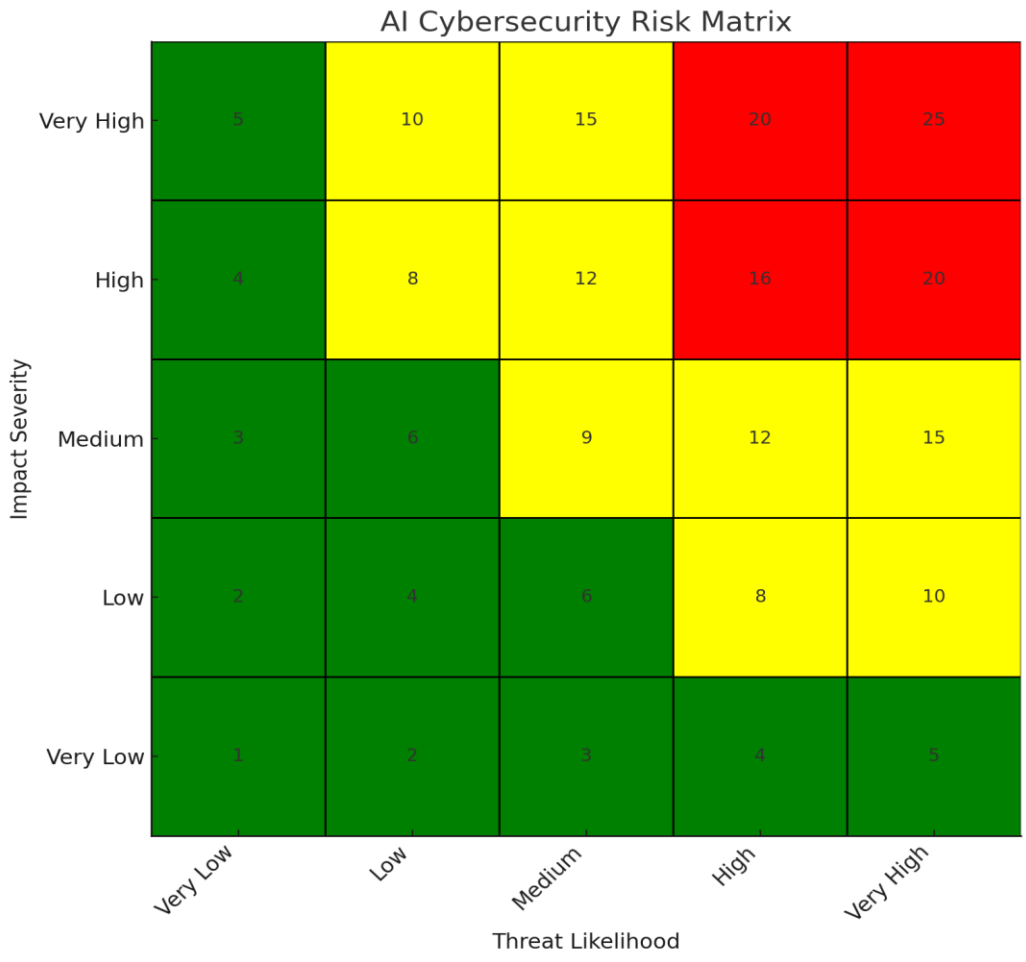
This structured classification aligns with prior work on ethical risk and security trade-offs in AI deployments (Mia, 2020; Johnson, 2020; Pasupuleti, 2021), and it helps identify which AI models and deployment contexts are most vulnerable.

3.5 Risk Evaluation Framework

Finally, the study applies a risk assessment matrix to evaluate dual-use AI technologies based on:

- **Threat Likelihood** (Low/Medium/High)
- **Impact Severity** (Technical, Operational, Legal, Reputational)
- **Defense Readiness** (Existing Controls, Response Speed, Human Oversight)

This framework supports the development of adaptive and resilient cybersecurity infrastructures capable of responding to adversarial manipulation and AI misuse (Jacob et al., 2021; Akduman, n.d.; Efe, n.d.).



This AI Cybersecurity Risk Matrix visually categorizes potential risks based on their likelihood of occurrence and the severity of their impact. Color-coded zones (green for low, yellow for medium, red for high) help prioritize cybersecurity threats for effective risk management.

This comprehensive methodology ensures that both the technical depth and the strategic implications of AI-driven cybersecurity are thoroughly examined from multiple dimensions: practical, ethical, and theoretical. The next section presents and interprets the results derived from these methods.

4. Results and Discussion

The emergence of artificial intelligence in cybersecurity has significantly altered the dynamics of threat detection and defense automation. This study's analysis of AI-driven security systems, adversarial case studies, and expert reports reveals a complex interplay between enhanced defensive capabilities and the rapidly evolving threat landscape shaped by adversarial AI.

4.1 The Promise of AI in Cyber Defense

AI-enabled systems offer unprecedented speed and accuracy in identifying cyber threats. Machine learning models, particularly those trained on behavioral and network traffic data, have demonstrated high accuracy in detecting anomalies, phishing patterns, and malware signatures (Silva & Gomez, 2021). Automated Security Operations Centers (SOCs) now leverage AI for predictive threat intelligence, reducing the mean time to detect (MTTD) and respond (MTTR) to incidents (Tiwari et al., 2020). This shift from reactive to proactive security postures has been further enabled by natural language processing and deep learning models that can process threat intelligence feeds in real time (Sun et al., 2020).

However, these benefits come with trade-offs. As Mia (2020) observed, increasing reliance on "black-box" AI models can reduce explainability, making it difficult for analysts to interpret why certain threats are flagged potentially eroding trust in automated systems. Furthermore, in high-stakes sectors like finance and defense, such opacity may lead to compliance and regulatory challenges (Watney, 2020).

4.2 The Weaponization of AI by Threat Actors

Simultaneously, malicious actors are leveraging the same AI advancements to design stealthier, more adaptive attacks. Adversarial machine learning (AML) techniques such as data poisoning, model inversion, and evasion attacks have become prominent tools for bypassing AI-based detection systems (Johnson, 2020; Hoanca & Mock, 2020). Notably, deepfake technologies are being adapted for phishing and disinformation campaigns, especially within the geopolitical and military intelligence domains (Pasupuleti, 2021; Fontana, 2020).

One of the most concerning developments is the use of generative adversarial networks (GANs) to craft synthetic attack traffic that mimics legitimate user behavior, effectively deceiving even state-of-the-art intrusion detection systems (Prowell et al., 2021). These findings confirm that while AI enhances cyber defense, it also raises the ceiling of threat sophistication.

4.3 Comparative Insights into AI Use for Defense vs. Offense

To provide clarity on the dual-use dilemma of AI in cybersecurity, this study conducted a comparative analysis of common AI applications in both offensive and defensive contexts. The table below presents key capabilities, tools, and use cases employed by defenders and attackers using AI technologies.

Table: Comparative Analysis of AI Applications in Cyber Defense and Offense

AI Technique	Defensive Use Case	Offensive Use Case	Notable Tools/Examples	Risk Level
Anomaly Detection	Intrusion detection, identifying unusual traffic	Evasion of traditional security systems	Snort + AI, Darktrace	Medium
GANs (Generative Adversarial Networks)	Malware mutation detection, data augmentation	Polymorphic malware generation, deepfake phishing	DeepLocker (IBM), MalGAN	High

NLP (Natural Language Processing)	Spam filtering, threat intel analysis	NLP-driven phishing, social engineering automation	Google's BERT for detection, ChatGPT jailbreak tools	High
Behavior-Based Modeling	Endpoint behavior monitoring, insider threat detection	Mimicking user behavior to bypass defenses	CrowdStrike Falcon, MITRE ATT&CK	Medium–High

This comparison underscores the blurred line between innovation and vulnerability. For instance, anomaly detection, a foundational method in AI-based security, is vulnerable to evasion when attackers train their malware to blend in with learned behavioral norms (Jacob et al., 2021). Similarly, NLP, while useful in threat intelligence automation, is being exploited to create sophisticated phishing emails indistinguishable from legitimate communication (Silva & Gomez, 2021).

4.4 Infrastructure and Legal Considerations

Beyond technical risks, the legal and infrastructural dimensions of AI cybersecurity require attention. National and international frameworks have not kept pace with AI's dual-use capabilities. As Watney (2020) argued, there is a growing legal risk surrounding AI applications in cyber defense, particularly when AI systems make unsupervised decisions affecting data privacy and international norms.

Infrastructure-wise, many AI-powered cybersecurity systems still operate atop vulnerable traditional architectures. For example, Akduman (n.d.) and Efe (n.d.) both highlighted that even advanced AI security solutions can be rendered ineffective if deployed within insecure or poorly configured digital environments. These structural weaknesses expose AI models to manipulation at the hardware or firmware level.

4.5 Toward a Resilient AI Cybersecurity Framework

To mitigate these risks, a layered, adaptive, and explainable AI-based cybersecurity model is recommended. This should include:

- Adversarial training to improve model robustness against evasion techniques.
- Explainable AI (XAI) modules to ensure model transparency, especially in critical infrastructure settings (Mia, 2020).
- Cryptographic agility and identity validation for securing non-human digital entities (Porambage et al., 2019).
- Ethical AI guidelines that balance automation with human oversight (Pasupuleti, 2021; Johnson, 2019).

These elements must be supported by continuous monitoring and real-time feedback loops to adapt to the evolving threat landscape.

Overall, the results confirm that AI in cybersecurity operates as a double-edged sword—offering both unprecedented capabilities and novel threats. The arms race between AI-powered defense and adversarial offense is intensifying, requiring not just better technology but also robust governance, ethical foresight, and international cooperation.

5. Proposed Framework

To address the dual-use challenge of artificial intelligence in cybersecurity where AI serves as both a shield and a sword a robust, adaptive framework is essential. The proposed framework integrates four core pillars: Adversarial Robustness, Automated Threat Intelligence, Explainable AI (XAI), and Ethical Governance. These components are designed to work in synergy, forming a resilient, dynamic, and proactive cyber defense ecosystem that can evolve in tandem with emerging threats.

5.1. Adversarial Robustness through Continuous Model Hardening

At the heart of AI vulnerability lies the susceptibility of models to adversarial inputs data manipulated in subtle ways to trick AI systems into misclassification or false negatives. To mitigate this, the framework incorporates adversarial training and model validation loops. Adversarial training involves exposing models to known adversarial examples during training so they learn to identify and resist such manipulations (Sun et al., 2020; Pasupuleti, 2021). Model hardening is also reinforced by deploying ensemble methods, where multiple AI detectors cross-validate alerts, minimizing false positives and increasing resistance to obfuscation attempts (Silva & Gomez, 2021).

5.2. AI-Augmented Threat Intelligence and Autonomous Detection

Modern cybersecurity infrastructures cannot depend solely on static rule-based systems. The proposed framework includes an AI-driven threat intelligence module that autonomously ingests data from global threat feeds, dark web monitoring, and internal logs. Machine learning algorithms analyze these diverse data streams to detect patterns and forecast potential threats. This predictive intelligence, when combined with behavioral analysis, significantly shortens response times to Advanced Persistent Threats (APTs), ransomware campaigns, and phishing attacks (Tiwari et al., 2020; Prowell et al., 2021).

This module also supports automated threat hunting, where AI agents proactively scan networks, endpoints, and cloud assets for suspicious behavior, without needing explicit rules or human prompts (Hoanca & Mock, 2020). This shift from reactive to proactive security postures represents a critical evolution in defending against modern cyber warfare tactics.

5.3. Explainable AI (XAI) for Human-in-the-Loop Trust and Compliance

One major challenge in adopting AI in cybersecurity is the lack of transparency in decision-making, which can lead to mistrust or misinterpretation, especially in regulatory environments. The framework emphasizes explainable AI, ensuring that all threat classifications, anomaly detections, and mitigation suggestions are accompanied by clear, human-readable explanations (Mia, 2020; Watney, 2020).

By embedding XAI layers into both training and inference pipelines, analysts and compliance officers can interpret AI-driven outputs, audit system behavior, and refine policies in line with legal and operational standards. This not only improves trust but also aligns the system with cybersecurity governance models, especially when dealing with cross-border legal obligations and GDPR-like data regimes (Johnson, 2020; Akduman, n.d.).

5.4. Ethical Governance and Non-Human Identity Management

As AI systems increasingly interact with machine identities such as IoT devices, autonomous agents, and APIs ensuring secure, ethical, and auditable interactions becomes paramount. The framework introduces a machine identity governance protocol that authenticates and authorizes non-human entities using cryptographic agility and zero-trust architecture principles (Porambage et al., 2019; Jacob et al., 2021).

These protocols enable secure machine-to-machine communication by dynamically validating trust levels, tracking behavior anomalies, and revoking compromised identities in real time. Furthermore, the inclusion of a policy engine allows organizations to encode ethical principles and regulatory requirements directly into AI workflows (Fontana, 2020; Johnson, 2019), ensuring compliance while maintaining agility.

5.5. Adaptive Feedback Loop and Continuous Learning

To future-proof the system, the framework embeds a **closed-loop feedback mechanism** that allows continuous learning from both successful and failed security events. Each incident detected or missed is logged, analyzed, and fed back into model training pipelines. This loop ensures that the AI becomes more accurate, less biased, and better adapted to novel threat vectors over time (Silva & Gomez, 2021; Prowell et al., 2021).

This approach counters the static nature of traditional cybersecurity tools, replacing them with systems capable of self-improvement and adaptive learning, essential in the face of rapidly evolving threats and attacker creativity.

5.6. Integration with Existing Infrastructure

The proposed framework is not meant to replace existing tools, but to augment them through AI overlay. APIs and modular connectors allow for integration with Security Information and Event Management (SIEM) systems, endpoint detection platforms, and cloud-native monitoring tools. It's designed to be vendor-neutral, scalable across environments, and applicable to both small enterprises and large cloud providers (Efe, n.d.; Prowell et al., 2021).

This flexibility enables security teams to gradually adopt AI-enhanced controls, minimizing disruption while maximizing strategic resilience against both known and unknown cyber threats.

6. Conclusion

The incorporation of artificial intelligence into cybersecurity is the turning point in the war against the more complicated and proficient online threats. Allowing an AI to liberate capabilities that nobody in this world can effectively reproduce in threat detection, incident response accelerations, and improving predictive intelligence, AI itself also creates new vulnerabilities in the system as it is being exploited by enemies, and the decision-making process is obscure. Such dual-use nature characterizes AI as a significant tool and a possible source of danger in the field of cybersecurity.

The study has delved into how incursion incursion specialists have used the advantage of adversarial AI and deep fakes as well as auto-control to race past current defense mechanisms, making it even more important to take aggressive, dynamic counter measure efforts (Pasupuleti, 2021; Johnson, 2019). To this, we offered a holistic AI-enabled cybersecurity framework anchored on four cardinal pillars including: adversarial robustness, autonomous threat intelligence, and explainable AI (XAI), and ethical governance. Placing such elements inside a cycle of continuous learning feedback system, organizations can shift their security posture to the proactive, resilient and self-improving level.

Additionally, the framework promotes regulatory operations, transparency, and trust by assuring explainable algorithms and secure machine-to-machine identity governance (Mia, 2020; Porambage et al., 2019). It is also consistent with more general trends in the convergence between AI and cybersecurity, such that it can easily fit into existing infrastructure of every sector and at every scale (Prowell et al., 2021; Efe, n.d.).

In the end, this way forward must be collaborative: it means using technology innovatively, but with moral insight, making policies compatible, and exchanging knowledge across the world. AI is becoming the defining feature of the digital battlefield, and security leaders, policymakers, and researchers should not take it lightly and make sure that their tools of defense do not become the new flaw of our systems.

References

1. Mia, L. (2020). Evaluating the Trade-offs Between Explainability and Security in AI-Powered Cyber Defense. *Available at SSRN 5140427*.

2. Tiwari, S., Sresth, V., & Srivastava, A. (2020). AI-Driven Cyber Threat Intelligence: Enhancing Predictive Security and Autonomous Defense Mechanisms.

3. Watney, M. M. (2020, June). Artificial intelligence and its' legal risk to cybersecurity. In *European Conference on Information Warfare and Security, ECCWS* (pp. 398-405).

4. Jacob, I., Lawson, R., & Smith, R. (2021). Future-Proofing AI and Cloud Systems: The Intersection of Quantum and Cybersecurity.

5. Silva, J., & Gomez, M. (2021). The Role of Artificial Intelligence in Cybersecurity: A Review of Threat Detection and Mitigation Techniques. *Artificial Intelligence and Machine Learning Review*, *2*(3), 1-9.

6. Sun, Y., Liu, J., Wang, J., Cao, Y., & Kato, N. (2020). When machine learning meets privacy in 6G: A survey. *IEEE Communications Surveys & Tutorials*, *22*(4), 2694-2724.

7. Akduman, B. THE TECH RACE AND SECURITY DILEMMAS: US-CHINA RIVALRY IN AI AND CYBERSECURITY WITH TÜRKIYE'S PERSPECTIVE. *Avrasya Sosyal ve Ekonomi Araştırmaları Dergisi*, *12*(1), 153-167.

8. Pasupuleti, V. (2021). AI-BASED MULTIMEDIA SECURITY IN COMBATING ADVERSARIAL ATTACKS, DEEPFAKES, AND ETHICAL CONCERNS. *INTERDISCIPLINARY JOURNAL OF AFRICAN & ASIAN STUDIES (IJAAS)*, *7*(1).

9. Johnson, J. (2019). Artificial intelligence & future warfare: implications for international security. *Defense & Security Analysis*, *35*(2), 147-169.

10. Johnson, J. (2020). Deterrence in the age of artificial intelligence & autonomy: a paradigm shift in nuclear deterrence theory and practice?. *Defense & Security Analysis*, *36*(4), 422-448.

11. Hoanca, B., & Mock, K. J. (2020). Artificial intelligence-based cybercrime. In *Encyclopedia of criminal activities and the deep web* (pp. 36-51). IGI Global.

12. Prowell, S., Manz, D., Culhane, C., Ghafoor, S., Kalke, M., Keahey, K., ... & Pinar, A. (2021). *Position Papers for the ASCR Workshop on Cybersecurity and Privacy for Scientific Computing Ecosystems*. US Department of Energy (USDOE), Washington DC (United States). Office of Science.

13. Efe, A. A RISK ASSESSMENT ON USAGE OF KALI TOOLS TO HACK AND MANIPULATE WEB-BASED MIS AND ERP APPLICATIONS. *Yönetim Bilişim Sistemleri Dergisi*, *11*(1), 62-80.

14. Fontana, G. (2020). Social Media at War. The Case of Kurdish Fighters and Their Impact on the Perception of the On-Going Anti-ISIS Conflict in Western Countries. *EUROPEAN CYBERSECURITY JOURNAL*, *6*(2), 91-96.

15. Porambage, P., Siriwardana, Y., Sedar, R., Kalalas, C., Soussi, W., MI, H. N. N., ... & Dhouha, A. (2019). INtelligent Security and PervasIve tRust for 5G and Beyond. *INSPIRE-5Gplus Consortium, WP3*, *3*.