Shajila Beegam M K, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 16, Issue 3, March 2026, pp 01-08

RESEARCH ARTICLE OPEN ACCESS

The Dual-Use Nature of Generative AI in Cybersecurity: A Review of Threats and Defenses

Shajila Beegam M K*, Shyjila P A**
**(Department of Computer Engineering, Govt. Polytechnic College, Vechoochira*
*** (Department of Computer Engineering, Govt. Women's Polytechnic College, Kayamkulam*

ABSTRACT
Generative AI technologies, such as large language models and synthetic media systems, present a profound dual-use dilemma in the cybersecurity domain. While these systems enable innovation by automating threat detection, generating synthetic datasets, and supporting rapid incident response, they also create new avenues for malicious exploitation, including phishing, malware generation, and deepfake-enabled fraud. This paper critically examines the paradoxical role of generative AI in cybersecurity and introduces the Cyber-AI Paradox taxonomy, a framework that categorizes both threats and defenses to clarify how legitimate applications can be repurposed for harmful ends. Sector-specific case studies in finance, healthcare, and critical infrastructure are analyzed to illustrate the dual-use tension and to highlight emerging vulnerabilities. Building on these insights, the paper identifies key research gaps—such as the absence of standardized risk assessment frameworks, limited interdisciplinary collaboration, and insufficient transparency in commercial AI systems—and offers strategic recommendations. By situating generative AI within the broader discourse on dual-use technologies, this study underscores the need for collective action among government, academia, and industry to balance innovation with security and ensure responsible deployment.

Keywords - Adversarial Generative AI (AdvGenAI), Agentic Identity Governance, Digital Forensic Attribution, Impostor Bias, Proactive Automated Defense

---

Date of Submission: 12-03-2026 Date of acceptance: 25-03-2026

I. INTRODUCTION

Artificial intelligence (AI) has become a transformative force across industries, with generative AI—including large language models (LLMs), image generators, and synthetic media systems—standing out for its ability to produce human-like text, visuals, and audio. These capabilities offer significant benefits for cybersecurity defense, such as automating vulnerability detection, generating synthetic datasets for training, and supporting rapid incident response. At the same time, the versatility of generative AI introduces profound risks. Its dual-use nature means that tools designed to enhance resilience can also be weaponized. Malicious actors exploit generative models to craft convincing phishing campaigns, generate polymorphic malware, and produce deepfake content for fraud or disinformation.

However, the same versatility introduces profound risks. The **dual-use nature** of generative AI means that tools designed to enhance resilience can also be weaponized. Malicious actors exploit generative models to craft convincing phishing campaigns, generate polymorphic malware, and produce deepfake content for fraud or disinformation. This paradox—where innovation and exploitation coexist—poses urgent challenges for researchers, policymakers, and practitioners.

Recent scholars frames this as the "dual-use dilemma of generative AI in cybersecurity", emphasizing the need to balance offensive threats with defensive capabilities. For example, Korimilli et al. [1] argue that generative AI must be studied as both a tool for strengthening cyber defenses and a vector for novel attack strategies. Similarly, IEEE research highlights how generative AI can enhance detection while simultaneously lowering the barrier for sophisticated cyberattacks [2]. Desai describes it as a "double-edged sword," noting its role in enabling phishing, deepfakes, and automated malware alongside its defensive potential [3].

To address this challenge, the paper introduces the **Cyber-AI Paradox Taxonomy**, a structured framework for categorizing both defensive and offensive applications of AI. Unlike the general notion of dual-use, which simply

highlights the coexistence of beneficial and harmful outcomes, the taxonomy systematically pairs positive applications—such as automated vulnerability detection, synthetic data generation, and resilience simulations—with their malicious counterparts, including phishing automation, polymorphic malware creation, and deepfake-enabled fraud. By mapping these contrasts, the taxonomy makes the dual-use dilemma more concrete and actionable, offering researchers and practitioners a tool to analyze risks, anticipate misuse, and design countermeasures while still harnessing the defensive potential of generative AI.

This paper critically examines the dual-use nature of AI systems, showing how technologies that foster innovation and societal benefit can also be exploited for malicious purposes. Its major contributions are as follows:

1. Introduces the Cyber-AI Paradox taxonomy to systematically categorize dual-use threats and corresponding defense mechanisms.
2. Evaluates sector-specific case studies to highlight research gaps and propose actionable recommendations.
3. Demonstrates that addressing dual-use risks requires coordinated efforts among government, academia, and industry to safeguard against misuse while enabling responsible innovation.

II. THE DUAL-USE DILEMMA IN CYBERSECURITY

The dual-use dilemma in cybersecurity reflects the paradox that generative AI can both strengthen defenses and empower attackers. Defensively, AI enables automated vulnerability detection, patch generation, synthetic data creation, and resilience testing. Yet, these same capabilities can be misused to automate phishing campaigns, generate polymorphic malware, and produce deepfake content for fraud and impersonation. This coexistence of innovation and exploitation underscores the urgent need for governance frameworks and collective action to maximize benefits while mitigating risks.

A. TAXONOMY OF AI-DRIVEN CYBER THREATS

We categorize modern threats into three macro-dimensions:

- **Offensive Automation:** Use of LLMs to generate high-fidelity phishing emails and automated vulnerability discovery.
- **Deepfake & Identity Fraud:** The rise of synthetic media used for executive impersonation and bypassing biometric MFA.
- **Adversarial AI Attacks:** Direct attacks on AI models themselves, such as **Data Poisoning** and **Model Inversion** [4] [5].

B. EVOLUTION OF DEFENSIVE FRAMEWORKS

The defensive side has shifted toward *Predictive Analytics*. Key advancements include:

- **Zero-Trust Identity First:** 81% of organizations have now adopted ZTA to combat identity-centric threats.
- **AI-Native SOCs:** Automated triage and forensics that reduce *dwell time* from 280 days to near-zero.
- **Crypto-Agility:** Migration toward Post-Quantum Cryptography (PQC) standards (FIPS 203, 204) to prevent "harvest now, decrypt later" attacks.

III. LITERATURE REVIEW

The evolving intersection of artificial intelligence (AI) and cybersecurity has been widely examined across diverse scholarly contributions, each highlighting the dual-use nature of emerging technologies.

Stevens [6] underscores the centrality of AI-driven anomaly detection in modern cybersecurity, noting how human–machine hybrids enhance efficiency but reduce transparency. Governance challenges arise as algorithms embed political and socioeconomic arrangements, with gaps in accountability across civilian and military domains. Similarly, Grotto & Dempsey [7] emphasize vulnerabilities unique to AI/ML systems—such as evasion, data poisoning, and model replication—arguing for tailored disclosure and management programs to integrate AI risks into broader cybersecurity safeguards.

Aramide [8] highlights the escalating arms race between adversaries exploiting AI for manipulation and defenders deploying AI-based detection systems, calling for resilient and ethical models. Umakor [4] advances this perspective by proposing generative AI–driven simulations to anticipate adversarial threats in critical infrastructure, embedding ethics into technical defenses while identifying governance fragmentation as a key gap. Adewale [9] further stresses the rise of AI-powered cybercrime, emphasizing automation, adaptability, and

deception, and urging global collaboration and stronger regulation.

Blauth [10] develops a typology of AI misuse, distinguishing vulnerabilities in AI models from malicious uses such as misinformation and autonomous weapons, illustrated through case studies like Cambridge Analytica. Anwar Mohammed [11] explores AI's paradoxical role, where defensive anomaly detection coexists with offensive deepfakes and botnets, highlighting ethical dilemmas. Nachaat Mohammed [12] surveys adoption trends in AI/ML for intrusion detection and malware analysis, noting both opportunities and persistent concerns over bias and transparency. Kurtovic [13] examines generative AI's dual role in innovation and cybercrime, stressing risks to data integrity and national security. Anderson [14] focuses on GANs, showing their dual-use in generating malware and strengthening defenses, and advocates for security-by-design governance.

Ndibe [15] emphasizes AI and LLMs in automated code reviews and threat intelligence, while warning of adversarial risks and insecure AI-generated code. Andrioni [16] explores GenAI in autonomous systems, balancing benefits in resilience with risks such as prompt injection and data leakage. Chagant [17] demonstrates GAI's predictive potential in intrusion detection, reporting high accuracy but stressing ethical governance. Ahi [18] surveys LLMs and GenAI, projecting sharp increases in AI-generated malware and recommending watermarking and cross-industry collaboration. Korimilli [1] investigates LLMs in red- and blue-team contexts, proposing governance frameworks like the Dual-Use Risk Index (DURI).

Macron [19] examines deepfakes as both innovation and threat, highlighting detection tools and the need for stronger legal frameworks. Pramod [20] explores GenAI's dual-edged role in both offense and defense, stressing safeguards and governance. Desai [3] reviews GenAI's transformative potential in content generation, urging ethical guidelines and collaboration. Hadji [21] analyzes governance challenges of dual-use digital technologies, advocating for robust GRC frameworks and international cooperation. Ozbay [22] presents a PRISMA-guided review of LLM-enabled cybersecurity, identifying innovations but persistent gaps in robustness and governance. Seghid [23] consolidates fragmented taxonomies into a unified framework of AI misuse across nine domains, calling for proactive, multi-stakeholder governance.

Mahto [24] explores generative AI in dynamic threat intelligence, demonstrating real-time applications in malware analysis and phishing detection, while addressing ethical concerns. Goswamy [25] provides a theoretical analysis of AI's dual-use risks, proposing explainable AI and fairness-aware algorithms within global governance frameworks. Kely [26] analyzes AI-driven social engineering, introducing the UM-AISE model to explain realism, personalization, and automation in attacks, and calling for enforceable auditability. Aluthman [5] surveys GAN-based intrusion detection systems, reviewing architectures and applications while noting challenges such as training instability and dual-use risks. Finally, Virtozu [27]examines AI's dual role in cybersecurity, emphasizing explainable AI, resilient architectures, and global cooperation to balance innovation with responsibility.

Across these studies, three recurring themes emerge:

1. **Technical Dual-Use** – AI technologies simultaneously strengthen defenses and expand attack surfaces.
2. **Governance & Ethics** – Persistent gaps in transparency, accountability, and regulation demand proactive frameworks.
3. **Research Gaps** – Empirical validation, standardized evaluation, and socio-technical integration remain underdeveloped.

Together, this body of literature underscores the urgent need for *multidisciplinary, globally coordinated approaches* to ensure AI contributes to resilience and security rather than instability.

IV. METHODOLOGY

We conducted this review following the PRISMA- (Preferred Reporting Items for Systematic reviews and Meta-Analyses) protocol. This ensures a rigorous, reproducible selection of literature from the 2020–2026 period. PRISMA Diagram of the selected studies is portrayed in Figure 1.

4.1 Search Strategy and Data Sources

A multi-stage search was executed across major academic databases, including IEEE Xplore, ACM Digital Library, PubMed, and ScienceDirect and Websites. We utilized a Boolean search string designed to capture the intersection of AI and Cybersecurity:

("Generative AI" OR "LLM" OR "Machine Learning") AND ("Cybersecurity" OR "Threat Detection") AND ("Zero Trust" OR "ZTA")

4.2 Inclusion and Exclusion Criteria

To maintain relevance to the 2026 landscape, the following criteria were applied:

- **Inclusion:** Peer-reviewed journals, reports and conference proceedings (2020–2026)
- **Exclusion:** Papers focusing on pre-transformer AI models (pre-2020), non-English publications, and studies without empirical or validated framework data.

4.3 Data Extraction and Synthesis

Following the PRISMA protocol, we screened 152 articles across IEEE, ACM, PubMed, and ScienceDirect. After applying inclusion/exclusion criteria, 26 studies were analyzed under three pillars:

1. Technical architecture
2. Governance and policy
3. Agentic identity research

A PRISMA flow diagram as shown in Figure 1 illustrates the selection process. Non-academic sources were excluded to strengthen credibility.

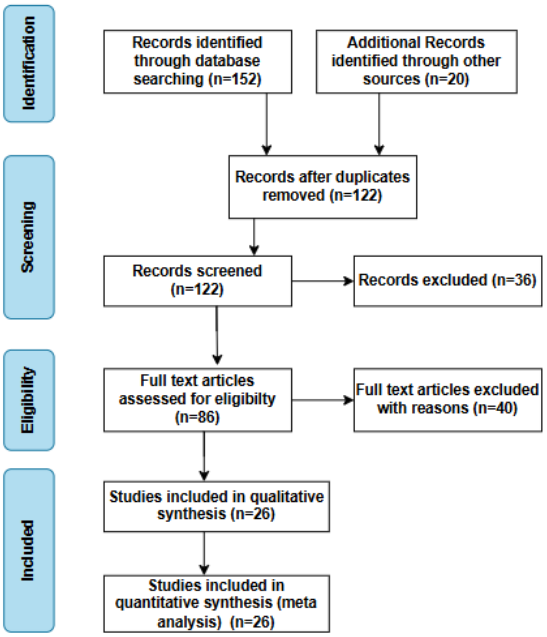

Figure 1: PRISMA Diagram for the selected studies

V. SECTOR-SPECIFIC CASE STUDIES

5.1 Healthcare: Securing the Internet of Medical Things (IoMT)

The healthcare sector remains the most targeted industry in 2026 due to the high value of Protected Health Information (PHI) [28].

- **The Challenge:** IoMT devices (e.g., insulin pumps, pacemakers) often lack the processing power for traditional encryption, making them "weak links" for lateral movement.
- **AI Intervention:** Modern frameworks now utilize *Edge-AI* to monitor device telemetry. For instance, an AI agent at the hospital's network edge can detect a 5% deviation in a pacemaker's communication frequency—a sign of a DDoS attempt—and isolate the device in milliseconds without interrupting clinical workflows.

5.2 Financial Services: Fighting *Agentic* Fraud

In 2026, financial institutions have moved beyond simple credit scoring to defending against **Agentic AI**—AI bots designed to autonomously find and exploit banking API vulnerabilities [29].

- **The Threat:** Attackers now use "Swarms" of AI agents to perform high-speed credential stuffing and KYC (Know Your Customer) bypass using real-time deepfake video.
- **The Defense:** Tier-1 banks like JPMorgan Chase have implemented *Multi-Agent Systems (MAS)*. In this setup, specialized AI agents collaborate: one analyzes transaction patterns, another interprets real-time regulatory shifts, and a third assesses the "Trust Score" of the user's device. This collaborative approach has reduced fraud-related losses by an estimated 40% compared to 2024 levels.

VI. COMPARATIVE ANALYSIS OF THREAT VS DEFENSE

The dual-use nature of Generative AI (GenAI) in cybersecurity between 2020 and 2026 represents a *cat-and-mouse* paradigm where the same models used to automate defense are simultaneously weaponized by adversaries. Current research highlights a shift from static, reactive security models to proactive, AI-driven ecosystems that prioritize real-time adaptation and mobile intelligence. A comparative analysis of Threat Vs Defense is given in Table 1.

VII. KEY RESEARCH TRENDS

- **Proactive vs. Reactive Paradigms**

Research in 2020–2026 emphasizes a move away from reactive *human-in-the-loop* updates toward *proactive automated defense*. For instance, the *GenSQLi framework* demonstrates how Generative AI can be used to synthesize novel SQLi attacks in a controlled environment to stress-test and automatically optimize Web Application Firewalls (WAFs) [30].

- **The Emergence of *Impostor Bias***

A critical theme in recent literature is the erosion of digital trust, termed "Impostor Bias." This phenomenon describes a generalized distrust in digital evidence caused by high-fidelity deepfakes [31] [32]. Malicious actors have shifted from simple misinformation to high-stakes fraud, such as using AI-generated speech to mimic CEOs for multi-million dollar unauthorized transfers [31] [32].

- **Resilience and Hybrid Threats**

As AI capabilities outpace governance, research is focusing on "systemic stresses" and polycrisis [33]. This has led to the development of the European Democracy Shield, which aims to consolidate fragmented tools into a coherent strategy against hybrid operations, including Foreign Information Manipulation and Interference (FIMI) [34].

- **Advancements in Deepfake Forensics**

While GANs and Diffusion Models (DMs) have lowered the barrier for creating realistic forgeries, the forensics community is countering with attribution techniques. Rather than just binary detection (real vs. fake), 2025 research prioritizes identifying the specific generative model or pipeline responsible for a piece of content [31] [32].

TABLE 1: Comparative Analysis: Threats vs. Defenses

Aspect	Defensive Research	Offensive Research
Automation	Use of LLMs (e.g., GPT-4o) to automatically generate WAF rules, blocking up to 99% of SQLi attacks [15] [18] [1] [22].	Automation of SQL injection (SQLi) payloads through in-context learning to bypass security filters [15] [18] [1] [22].
Media Integrity	Development of DCT-based deepfake detectors and "DNA-Det" for architectural fingerprinting of AI models [31].	Adversarial "gray-box" attacks designed to manipulate DCT features, evading state-of-the-art detectors [31].
Infrastructure	Integration of "Edge AI" and Federated Learning (FL) in IoT to detect threats while preserving data privacy [35].	Exploitation of the expanded attack surface in IoT/6G networks through AI-driven "impostor bias" and social engineering [36].
Information	Legislative frameworks like the EU's Digital Services Act (DSA) and FIMI Protocols to counter hybrid threats [34].	Rapid creation of deceptive content and disinformation to destabilize democratic processes and erode public trust [34].

VIII. RESEARCH GAPS

Current research indicates that while technical capabilities for both GenAI-driven attacks and defenses have expanded, several **critical research gaps** remain. These gaps represent "blind spots" where current academic and industrial solutions fail to provide long-term security.

1. The Detection Asymmetry Gap

Detecting AI-generated content is challenging, but the difficulty goes beyond identification. Differentiating a benign AI-assisted code snippet from a malicious polymorphic script is increasingly complex, and current research focuses mainly on detection. The real gap lies in attribution and intent: we lack forensic frameworks to trace an AI-generated attack back to a specific model or its unique *adversarial fingerprint*. Without such tools, establishing legal accountability for malicious use of generative AI remains out of reach.

2. Machine Identity & Autonomous Agent Governance

By 2026, the rise of *Agentic AI* systems capable of acting independently has outpaced security research, creating a new non-human identity problem. Currently, 42% of machine identities such as bots and agents hold privileged access, yet most security programs still define *privileged users* exclusively as humans. This mismatch exposes a critical gap: there are no standardized protocols for *AI Authentication*, meaning we lack reliable ways to verify that an autonomous agent is truly who it claims to be and operating within its intended parameters. Without such safeguards, these agents remain vulnerable to manipulation, including prompt injection attacks, undermining both trust and accountability in autonomous systems

3. Longitudinal Robustness of Defensive LLMs

Recent studies highlight the short-term effectiveness of AI defenses, often reporting success rates such as blocking 99% of SQL injection attacks within a week. Yet adversaries exploit *drift* and continuous adaptation, gradually eroding these protections over time. This reveals a critical gap: there is little longitudinal research examining how defensive AI models degrade when exposed to evolving adversarial environments over months or years. Without such sustained analysis, our understanding of AI resilience remains limited, leaving defenses vulnerable to slow, adaptive attacks.

4. The *Trust Decay* & Human Factors Gap

While technical defenses against GenAI threats are advancing, research into their psychological and societal impact remains underdeveloped. A key issue is the *Impostor Bias*, where widespread distrust of digital content—driven by deepfakes—creates either alert fatigue or a breakdown in organizational communication. This highlights a critical gap: the absence of *Behaviorally-Grounded Models* that integrate technical safeguards with human psychology. Such models are essential to counter social engineering attacks that no longer rely on obvious cues like poor grammar, but instead exploit trust and perception at a deeper level

5. Hardware-Level and *Edge* AI Security

As AI shifts from massive data centers to 6G-enabled edge devices like IoT sensors and smartphones in 2026, a new physical gap has emerged. Most GenAI security research still assumes high-compute cloud environments, leaving resource-constrained devices overlooked. The challenge is clear: how can we deploy robust, real-time AI defenses on low-power edge hardware that cannot support massive language models? This lack of research into lightweight, resource-efficient security frameworks exposes a critical vulnerability, as edge devices become increasingly central to everyday digital infrastructure.

IX. OPEN CHALLENGES AND RESEARCH NEEDS

9.1 Challenges

- **Explainability (XAI):** Security analysts often cannot interpret *why* an AI flagged a packet, leading to trust issues.
- **Regulatory Fragmentation:** Compliance with the EU AI Act and India's DPDP (2025) [37] requires automated audit trails that most systems currently lack .
- **Human-in-the-Loop:** As AI becomes "agentic" (acting independently), defining the boundary for human intervention remains a legal and technical hurdle (ORF, 2026) [38].

9.2 Research Needs

Technical Research Needs

- **Attribution Frameworks**: Develop forensic methods to trace malicious AI-generated code or content back to specific models, architectures, or training datasets.
- **Longitudinal Defense Studies**: Move beyond short-term trials to study how defensive AI systems degrade under adversarial drift and continuous adaptation over months or years.
- **Lightweight Edge Security**: Design resource-efficient AI defenses for IoT and 6G-enabled edge devices that cannot run large models but still require real-time protection.
- **AI Authentication Protocols**: Establish standards for verifying the identity and integrity of autonomous agents, ensuring they operate within intended parameters and resist prompt injection.

Human & Societal Research Needs

- **Behaviorally-Grounded Models**: Integrate psychological insights into technical defenses to counter social engineering attacks that exploit trust rather than obvious cues.
- **Impostor Bias Mitigation**: Study how deepfake-driven distrust affects organizational communication and design interventions to reduce alert fatigue.
- **Legal & Ethical Accountability**: Explore frameworks for assigning responsibility when generative AI is misused, balancing innovation with liability.

Policy & Governance Research Needs

- **Dual-Use Risk Taxonomies**: Create structured classifications of benign vs. malicious applications of generative AI to guide regulation and oversight.
- **Cross-Disciplinary Standards**: Develop interdisciplinary protocols that bridge cybersecurity, law, and ethics for handling AI misuse.
- **Transparency & Disclosure**: Investigate requirements for watermarking, provenance tracking, or mandatory disclosure of AI-generated content in sensitive domains.

X. CONCLUSION

The systematic review of cybersecurity research from 2020 to 2026 confirms that Generative AI has moved beyond being a "supplementary tool" to becoming the *primary engine* of both cyber-offense and defense. While frameworks like *GenSQLi and Edge AI* have significantly lowered the *Time-to-Detect* (TTD), the emergence of *Agentic AI* and *Impostor Bias* has created new vulnerabilities that traditional signature-based security cannot address.

The *Cat-and-Mouse* game has shifted from a battle of algorithms to a battle of *autonomous ecosystems.* Success in this era is no longer defined

by blocking every attack, but by building *systemic resilience*—the ability of a network to autonomously absorb, adapt to, and recover from AI-driven disruptions.

Recommendations

- **For Researchers (Focus: *Agentic* Defense):** Shift research from simply identifying AI-generated content to modeling AI intent. Priorities include developing forensic attribution (fingerprinting) to hold attackers accountable and optimizing Small Language Models (SLMs) for secure, decentralized 6G edge computing.
- **For Policy Makers (Focus: Identity Standards):** Standardize Agent-to-Agent (A2A) protocols to ensure autonomous systems are authenticated when accessing critical infrastructure. This includes leveraging frameworks like the 2025 European Democracy Shield to create a unified front against AI-driven disinformation.
- **For Industry Practitioners (Focus: Zero-Trust AI):** Transition to a "Zero-Trust" posture where every AI output is treated as a potential threat vector. This involves prioritizing Machine Identity Management (verifying the "who" behind an agent) and utilizing continuous AI-led Red-Teaming to automate defense hardening**.**

REFERENCES

[1] S. K. Korimilli, M. H. Rahman, G. Sunkara, M. M. Haque Mukit, and A. Al Hasib, "Dual-Use of Generative AI in Cybersecurity: Balancing Offensive Threats and Defensive Capabilities in the Post-LLM Era," *Available SSRN 5437776*, 2025.

[2] M. R. Islam, *Generative AI, cybersecurity, and ethics*. John Wiley & Sons, 2024.

[3] P. Desai, "A comprehensive study on Generative AI as a double-edged sword in the digital security landscape."

[4] M. F. Umakor, "Threat modelling for artificial intelligence governance: integrating ethical considerations into adversarial attack simulations for critical infrastructure using generative AI," *World J Adv Res Rev*, vol. 15, no. 2, pp. 873–890, 2022.

[5] M. Alauthman, N. Aslam, A. Al-Qerem, A. Aldweesh, and P. Sureephong, "Generative Adversarial Networks for Intrusion Detection Systems: A Comprehensive Survey of Applications, Challenges, and Research Directions," *Arab. J. Sci. Eng.*, pp. 1–25, 2026.

[6] T. Stevens, "Knowledge in the grey zone: AI and cybersecurity," *Digit. War*, vol. 1, no. 1, pp. 164–170, 2020.

[7] A. J. Grotto and J. Dempsey, "Vulnerability disclosure and management for AI/ML systems: A working paper with policy recommendations," *ML Syst. a Work. Pap. with policy Recomm. (November 15, 2021)*, 2021.

[8] O. O. Aramide, "AI-Driven Cybersecurity: The Double-Edged Sword of Automation and Adversarial Threats," *Int. J. Humanit. Inf. Technol.*, vol. 4, no. 04, pp. 19–38, 2022.

[9] T. Adewale, "Artificial Intelligence in Cybercrime: Unveiling the Emerging Landscape of Intelligent Threats," 2022.

[10] T. F. Blauth, O. J. Gstrein, and A. Zwitter, "Artificial intelligence crime: An overview of malicious use and abuse of AI," *Ieee Access*, vol. 10, pp. 77110–77122, 2022.

[11] A. Mohammed, "The Paradox of AI in Cybersecurity: Protector and Potential Exploiter," *Balt. J. Eng. Technol.*, vol. 2, no. 1, pp. 70–76, 2023.

[12] N. Mohamed, "Current trends in AI and ML for cybersecurity: A state-of-the-art survey," *Cogent Eng.*, vol. 10, no. 2, p. 2272358, 2023.

[13] H. Kurtović, E. Šabanović, A. A. Almisreb, M. A. Saleh, and N. Ismail, "Exploring the Dark Side: A Systematic Review of Generative AI's Role in Network Attacks and Breaches," in *Conference of Recent Trends and Applications of Soft Computing in Engineering*, 2024, pp. 27–51.

[14] K. Anderson, "Generative Adversarial Networks (GANs) for Cybersecurity: Dual Role in Attack Simulation and Defense Mechanism Development," 2024.

[15] O. S. Ndibe and P. O. Ufomba, "A review of applying AI for cybersecurity: Opportunities, risks, and mitigation strategies," *Appl. Sci. Comput. Energy*, vol. 1, no. 1, pp. 140–156, 2024.

[16] M. Andreoni, W. T. Lunardi, G. Lawton, and S. Thakkar, "Enhancing autonomous system security and resilience with generative AI: A comprehensive survey," *IEEE Access*, vol. 12, pp. 109470–109493, 2024.

[17] K. C. Chaganti, "Leveraging Generative AI for Proactive Threat Intelligence: Opportunities and Risks," *Authorea Prepr.*, 2024.

[18] K. Ahi and S. Valizadeh, "Large Language Models (LLMs) and Generative AI in Cybersecurity and Privacy: A Survey of

Dual-Use Risks, AI-Generated Malware, Explainability, and Defensive Strategies," in *2025 Silicon Valley Cybersecurity Conference (SVCC)*, 2025, pp. 1–8.

[19] T. Macron, "Generative AI and Cybersecurity: Analyzing the Dual Use of Deepfake Technology for Threats and Defensive Measures," 2025.

[20] S. P. Ugale and A. Sinha, "Generative AI in Cybersecurity: Threats, Vulnerabilities, and Protective Measures," in *2025 International Conference on Future Technologies (ICFT)*, 2025, pp. 1–8.

[21] M. Hadji-Janev, "Navigating the Governance, Risk, and Compliance of Dual-Use Digital Technologies: Balancing Innovation with Security," *Spectr. Dual-Use Technol. Unforeseen Risks Versus Returns*, pp. 229–260, 2025.

[22] R. Özbay, M. Çelebi, and U. Yavanoğlu, "The AI-Cybersecurity Nexus: How Large Language Models Are Reshaping Threat Intelligence and Digital Defense," *IEEE Access*, 2026.

[23] N. Seghid, F. Iqbal, K. Al-Room, and Á. MacDermott, "Emerging Threats in AI: A Detailed Review of Misuses and Risks Across Modern AI Technologies," *Front. Commun. Networks*, 2026.

[24] M. K. Mahto, "Dynamic Threat Intelligence: Leveraging Generative AI for Real-Time Security Response," *Gener. Artif. Intell. Next-Generation Secur. Paradig.*, pp. 107–136, 2026.

[25] A. Sarkar, S. S. Goswami, and S. K. Sahoo, "AI-Powered Threats and Solutions: A Theoretical Analysis of Risks, Governance, and Ethical Safeguards," *Appl. Res. Adv.*, pp. 1–23, 2026.

[26] G. Kely, S. Sérgio, M. Gomes, and M. Silvestre, "AI-Powered Social Engineering: Emerging Attack Vectors, Vulnerabilities, and Multi-Layered Defense Strategies," *Computers*, vol. 15, no. 2, p. 128, 2026.

[27] I. Virtosu, "Guardians or threats? The double-edged role of artificial intelligence in cyber security," in *International Conference on Machine Intelligence & Security for Smart Cities (TRUST) Proceedings*, 2025, vol. 2, pp. 61–73.

[28] S. Sengan, C.-S. Shieh, and M.-F. Horng, "A hybrid blockchain based deep learning model for multivector attack detection in internet of things enabled healthcare systems," *Sci. Rep.*, 2026.

[29] H. K. Sriram and B. M. Bharath M, "Beyond Automation: Exploring the Potential of Agentic AI in Risk Management and Fraud Detection in Banks," *Available SSRN 5275557*, 2025.

[30] V. Babaey and A. Ravindran, "Gensqli: A generative artificial intelligence framework for automatically securing web application firewalls against structured query language injection attacks," *Futur. Internet*, vol. 17, no. 1, p. 8, 2024.

[31] L. Li, "Research on deep fake detection technology of multimedia information based on deep learning," in *International Conference on Wireless, Optical Communication, and Information Engineering (WOCIE 2025)*, 2025, vol. 13973, pp. 17–23.

[32] I. Amerini *et al.*, "Deepfake media forensics: Status and future challenges," *J. Imaging*, vol. 11, no. 3, p. 73, 2025.

[33] M. Lawrence, T. Homer-Dixon, S. Janzwood, J. Rockström, O. Renn, and J. F. Donges, "Global polycrisis: the causal mechanisms of crisis entanglement," *Glob. Sustain.*, vol. 7, p. e6, 2024.

[34] S. Eskens, "Computer Law & Security Review : The International Journal of Technology Law and Practice The role of the Regulation on the transparency and targeting of political advertising and European Media Freedom Act in the EU ' s anti-disinformation strategy," *Comput. Law Secur. Rev. Int. J. Technol. Law Pract.*, vol. 58, no. August, p. 106185, 2025, doi: 10.1016/j.clsr.2025.106185.

[35] M. Rahmati and A. Pagano, "Federated learning-driven cybersecurity framework for iot networks with privacy preserving and real-time threat detection capabilities," in *Informatics*, 2025, vol. 12, no. 3, p. 62.

[36] H. Pennanen, T. Hänninen, and O. Tervo, "6G : The Intelligent Network of Everything," no. November, pp. 1–101, 2024, doi: 10.1109/ACCESS.2024.3521579.

[37] S. D. Kishwar, J. S. Sahani, and S. Tyagi, "Navigating India's Draft DPDP Rules 2025: Implementation challenges in protecting children's personal data," *J. Data Prot. Priv.*, vol. 8, no. 2, pp. 144–164, 2026.

[38] "AI in Modern Warfare: India's Strategic Challenges and Opportunities." [Online]. Available: https://www.orfonline.org/expert-speak/ai-in-modern-warfare-india-s-strategic-challenges-and-opportunities.