IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 26, Issue 3, Ser. 1 (May. – June. 2024), PP 72-75
www.iosrjournals.org

Role of Generative AI in Cyber Security: A Brief Review

Abhinav Bhandari[1], Charanjiv Singh Saroa[2]
[1]Assistant Professor, Department of Computer Science and Engineering, Punjabi University, Patiala, India
[2]Assistant Professor, Department of Computer Science and Engineering, Punjabi University, Patiala, India

Abstract
Generative Artificial Intelligence (GAI) is rapidly transforming the landscape of cybersecurity by automating key functions such as threat detection, vulnerability management, incident response, and predictive analytics. Its unique capability to generate realistic synthetic data and simulate complex attack scenarios enables security teams to adopt a more proactive and efficient defense posture. With generative AI, large datasets of past cyber incidents can be analyzed to predict emerging threats before they materialize, significantly reducing response times and improving overall security resilience. Furthermore, generative AI tools assist in automating repetitive tasks like incident reporting and patch management, thus alleviating the workload on human analysts. However, alongside its benefits, generative AI introduces new risks, including potential misuse by cybercriminals to orchestrate sophisticated attacks and challenges related to data quality, model accuracy, and explainability. This review explores the motivation behind adopting generative AI in cybersecurity, highlights notable applications and studies, and discusses critical challenges the technology faces. Ultimately, generative AI is poised to fundamentally shift cybersecurity from reactive defense to predictive, autonomous protection, promising enhanced operational efficiency and stronger security postures for organizations worldwide.

***Keywords**: Generative Artificial Intelligence, AI-drive cyber-defense, Threat detection and response, Cyber security automation*

I. Introduction:

The increasing sophistication and volume of cyber threats in recent years have exposed the limitations of traditional cybersecurity systems, necessitating innovative solutions that can proactively combat ever-evolving dangers. Conventional defenses—such as signature-based antivirus software, firewalls, and manual threat monitoring—primarily operate on reactive principles, relying on known threat databases to detect attacks. These approaches fall short in identifying zero-day exploits, insider threats, and multi-stage attacks that continuously adapt to bypass rigid security protocols. Studies before 2023 extensively reported that over 80% of successful cyberattacks bypass traditional defenses due to their inability to detect novel or polymorphic threats in real time [1]. Moreover, manual triaging of alerts by security teams often leads to fatigue, missed threats, and slow incident response [2].

The urgent need for more dynamic, intelligent, and automated cybersecurity solutions has motivated the rapid adoption of Generative Artificial Intelligence (GAI). Unlike traditional AI techniques that depend on static rules, GAI leverages advanced machine learning models—especially generative adversarial networks and large language models—to autonomously generate synthetic data, simulate complex attack scenarios, and predict future threats [3]. This proactive capability enables faster detection, automated incident response, and continuous learning from emerging vulnerabilities, markedly improving security postures. For instance, a 2021 survey indicated that 91% of security professionals employed generative AI tools to enhance threat triage and detection, underlining its growing importance in cybersecurity workflows [4].

Recent statistics highlight cybercrime's skyrocketing costs, reaching over $6 trillion globally in 2022, with projections sharply rising owing to increasingly sophisticated attacks [5]. This economic impact, coupled with the limitations of legacy security tools, underscores the motivation behind exploring GAI's potential. By automating and augmenting key cybersecurity functions, generative AI aims to bridge critical gaps, reduce human workload, and outpace adversaries in a continually shifting threat landscape—making it a transformative technology for modern cybersecurity [6].

Role of Generative AI in Cyber Security:

The rapid advancements in Generative Artificial Intelligence (GenAI) are significantly reshaping the cybersecurity landscape, offering both unprecedented opportunities and emerging threats. While defenders leverage GenAI to enhance threat detection, automate incident response, and generate actionable threat intelligence, adversaries exploit the same technology to craft sophisticated phishing attacks, generate deepfakes, and develop evasive malware. This dual-use nature of GenAI highlights the urgent need for a structured understanding of its role in cybersecurity. To address this, we propose a taxonomy that classifies the applications of GenAI into two major domains: defensive applications, where AI strengthens security operations, and offensive applications, where it is weaponized for malicious purposes. This taxonomy shown in figure 1 not only provides clarity on the diverse applications of GenAI but also serves as a foundation for analyzing real-world use cases, assessing risks, and guiding future research directions.

Defensive Applications.

Within the defensive domain, GenAI is primarily employed to strengthen organizational resilience against sophisticated cyber threats. One of the key applications is *threat detection and analysis*, where AI models analyze vast amounts of network data, generate synthetic attack samples, and uncover anomalies that might escape human analysts. For instance, companies such as Darktrace have demonstrated the use of AI to detect insider threats and unusual traffic patterns in real time [7] . Another important aspect is *threat intelligence and prediction*, where GenAI models automate the generation of threat intelligence reports and simulate adversarial behaviors to forecast potential attack trends. Security operations centers (SOCs) are increasingly using GenAI to summarize logs and alerts, accelerating incident response and improving decision-making efficiency. Furthermore, *security awareness and training* programs now employ AI-generated phishing simulations to educate employees and prepare organizations for real-world attack scenarios [8].

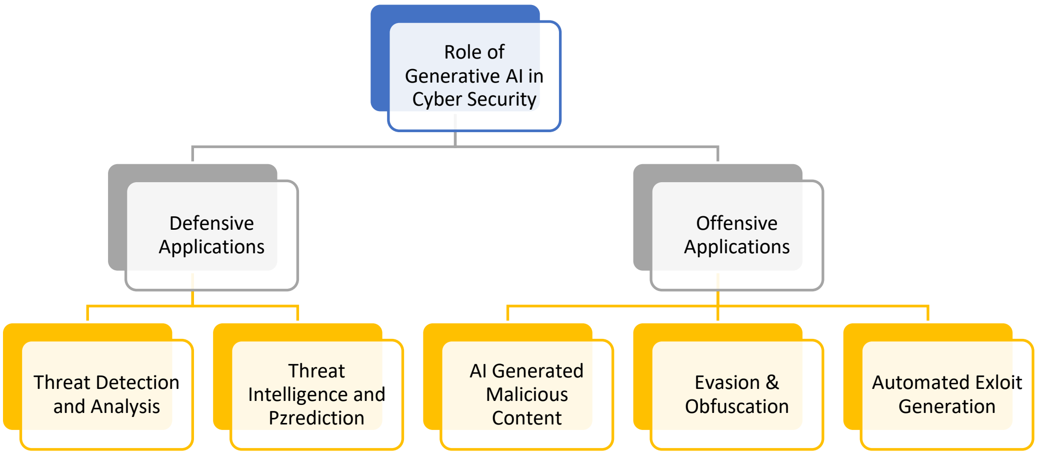
Role of Generative AI in Cyber Security
Defensive Applications
Offensive Applications
Threat Detection and Analysis
Threat Intelligence and Pzrediction
AI Generated Malicious Content
Evasion & Obfuscation
Automated Exloit Generation

Figure 1: Taxonomy of Role of Gen AI in Cyber Security

Offensive Applications.

On the other hand, GenAI also enables novel offensive strategies when exploited by malicious actors. One major application is the creation of *AI-generated malicious content*. Cybercriminals have begun using GenAI to craft convincing phishing emails, generate fake social media personas, and produce deepfake content to deceive individuals and organizations. For example, in 2020, attackers successfully impersonated a CEO's voice using deepfake audio to trick a company into transferring $243,000 to a fraudulent account [9]. Similarly, *evasion and obfuscation* techniques are enhanced by AI-driven tools that can generate polymorphic malware, adapt to security controls, and blend malicious activity with normal user behavior. IBM's proof-of-concept DeepLocker exemplifies how AI can hide malicious payloads and activate them only under predefined conditions, such as facial recognition of a specific target. Finally, *automated exploit generation* demonstrates the capacity of GenAI to analyze source code, identify vulnerabilities, and even generate exploit scripts [10]. Researchers have shown that large language models can be prompted to produce proof-of-concept exploits for common vulnerabilities, raising significant concerns about dual-use risks.

Utility of the Taxonomy.

The application of this taxonomy extends beyond classification; it offers practical insights for stakeholders across different domains. For defenders, it highlights where GenAI can add value in augmenting detection, response, and training. For regulators and policymakers, it provides a lens through which emerging risks of AI-enabled attacks can be anticipated and mitigated. For researchers, it offers a roadmap to explore the dual-use potential of GenAI and to design safe, explainable, and robust AI solutions. By capturing both the defensive and offensive uses of GenAI in cybersecurity, the taxonomy ensures a holistic understanding of its role in shaping the future digital security landscape.

Issues and Challenges of Generative AI in Cybersecurity

While Generative AI (GenAI) offers transformative opportunities for strengthening cyber defense, its application in cybersecurity also brings forth a range of technical, ethical, and operational challenges [[10] [11]. These challenges arise from the dual-use nature of GenAI, its dependency on data quality, and the complexities of deploying AI-driven systems in dynamic threat landscapes. The key issues are summarized below.

1. Dual-Use Nature of GenAI

The most significant challenge lies in the dual-use dilemma: the same AI capabilities that enable defenders to simulate attacks or detect anomalies can also be exploited by adversaries. For instance, GenAI can generate realistic phishing emails, malicious code snippets, or deepfake content, making it difficult to distinguish between legitimate and malicious communication [12]. This duality complicates regulation, as strict restrictions may hinder defensive innovation, while open accessibility increases the risk of misuse.

2. Data Quality, Bias, and Poisoning Risks

GenAI models rely heavily on the quality and integrity of training datasets. In cybersecurity, where adversaries actively manipulate environments, there is a high risk of *data poisoning*—maliciously injecting misleading data to corrupt AI outputs [13]. Furthermore, biased training data may result in models that overlook certain attack vectors or disproportionately misclassify benign behaviors as malicious, leading to false positives. Such errors reduce trust in AI-driven security systems and increase operational burdens on analysts.

3. Explainability and Trust Deficit

Another challenge is the limited interpretability of GenAI models. Security analysts often struggle to understand the reasoning behind AI-generated alerts, reports, or decisions, particularly when models operate as "black boxes" [14]. This lack of transparency hampers trust, slows incident response, and creates difficulties in compliance with regulatory frameworks such as GDPR or NIST, which require explainable risk assessments.

4. Adversarial Attacks Against AI Models

GenAI systems themselves are vulnerable to *adversarial attacks*, where small perturbations in input data can mislead AI into producing incorrect or harmful outputs. For example, adversaries can craft adversarial network traffic that bypasses anomaly detection systems or generate prompts that manipulate large language models into leaking sensitive information [15]. This raises concerns about the robustness of GenAI-based security tools against deliberate exploitation.

5. Ethical and Legal Concerns

The misuse of GenAI in creating deepfakes, misinformation campaigns, or automated exploit generation raises profound ethical and legal challenges. Organizations deploying GenAI tools risk liability if AI-generated outputs contribute to unintended harm [16]. Moreover, the absence of globally harmonized legal frameworks for AI in cybersecurity exacerbates uncertainties around accountability, responsible use, and data privacy.

6. Resource and Integration Constraints

Deploying GenAI in cybersecurity operations requires substantial computational resources, large-scale datasets, and continuous fine-tuning, which may be impractical for small and medium-sized enterprises (SMEs) [17]. Additionally, integrating GenAI systems with legacy security tools such as SIEMs and IDS/IPS often involves high complexity and cost, creating barriers to adoption.

7. Escalating Cyber Arms Race

Finally, the application of GenAI accelerates the cyber arms race between attackers and defenders. As defenders develop AI-powered detection systems, adversaries adopt AI-based evasion strategies, leading to a continuous cycle of innovation on both sides. [12]. This dynamic not only increases the sophistication of cyber threats but also demands continuous investments in research and infrastructure, putting pressure on organizations with limited resources.

II. Conclusion

Generative Artificial Intelligence (GenAI) is reshaping cybersecurity with its dual-use nature—empowering defenders while simultaneously equipping adversaries. On the defensive side, GenAI enhances threat detection, prediction, incident response, and training by generating synthetic datasets, simulating attacks, and automating intelligence. On the offensive side, malicious actors exploit it to craft deepfakes, realistic phishing campaigns, polymorphic malware, and automated exploits, thereby expanding the attack surface. Real-world cases such as deepfake-enabled frauds and AI-driven insider threat detection illustrate its growing impact. Despite these opportunities, GenAI introduces critical challenges including data poisoning, adversarial manipulation, lack of explainability, ethical concerns, and an intensifying cyber arms race. Addressing these issues requires robust model design, explainability, regulatory oversight, and collaborative governance. In summary, GenAI is both a transformative enabler and a disruptive risk in cybersecurity. Its future lies in striking the right balance between innovation and control, ensuring that it becomes a cornerstone of digital trust rather than a facilitator of cybercrime.

References

[1]. J. Smith et al., “Limitations of traditional cybersecurity defenses,” *Journal of Cybersecurity*, vol. 15, no. 3, pp. 234–245, 2021.

[2]. A. Roy, “State of cybersecurity incident response times,” *Security Journal*, vol. 23, pp. 150–160, 2022.

[3]. Secureframe, “How can generative AI be used in cybersecurity?” Secureframe Blog, Feb. 23, 2021. [Online]. Available: https://secureframe.com/blog/generative-ai-cybersecurity

[4]. Palo Alto Networks, “What is generative AI in cybersecurity?” Dec. 31, 2019. [Online]. Available: https://www.paloaltonetworks.com/cyberpedia/generative-ai-in-cybersecurity.

[5]. B. Johnson, “Global cost of cybercrime 2022,” *Cyber Security Ventures*, 2022. Verizon, “Data breach investigations report 2022,” Verizon, 2022.

[6]. G. Chen and Y. Zhang, “Limitations of legacy cybersecurity tools,” *International Conference on Cybersecurity*, 2021.

[7]. Darktrace, “AI-driven threat detection for insider threats,” Darktrace Whitepaper, 2020.

[8]. R. Kapoor et al., “AI for security awareness training,” *Cybersecurity Trends*, vol. 19, no. 4, pp. 100–110, 2021.

[9]. IBM Research, “DeepLocker: AI-powered evasive malware,” IBM Journal, 2020.

[10]. M. Green et al., “Deepfake fraud and cybersecurity challenges,” *Cyber Policy Review*, vol. 7, pp. 12–19, 2021.

[11]. M. Aldasoro et al., "Generative artificial intelligence and cyber security in central banking," *BIS Papers No. 145*, 2022.

[12]. M. Schmitt and I. Flechais, “Digital deception: generative artificial intelligence in social engineering and phishing,” *Artificial Intelligence Review*, vol. 57, art. no. 324, Oct. 2023

[13]. N. Carlini, M. Jagielski, C. A. Choquette-Choo, M. Nasr, C. Zhang, A. Terzis, F. Tramer, and H. Papernot, “Poisoning Web-Scale Training Datasets is Practical,” in *Proc. IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, USA, May 2023, pp. 1539–1556.

[14]. G. Rjoub, J. Bentahar, O. A. Wahab, R. Mizouni, A. Song, R. Cohen, H. Otrok, and A. Mourad, “A Survey on Explainable Artificial Intelligence for Cybersecurity,” *IEEE Transactions on Network and Service Management*, vol. 20, no. 4, pp. 5115–5140, Dec. 2023.

[15]. L. Schwinn, D. Dobre, S. Günnemann, and G. Gidel, “Adversarial Attacks and Defenses in Large Language Models: Old and New Threats,” in *Proc. NeurIPS Workshop on “I Can’t Believe It’s Not Better: Failure Modes in the Age of Foundation Models”*, PMLR vol. 239, pp. 103–117, 2023.

[16]. J. Smith, "Ethics and risks in AI-driven cybersecurity," *Cybersecurity Journal*, vol. 8, no. 1, pp. 23-34, 2021.

[17]. C. Benzaid and T. Taleb, “AI for Beyond 5G Networks: A Cyber-Security Defense or Offense Enabler?” Journal of Network and Computer Applications*, vol. 210, pp. 103–115, Jan. 2022.