RESEARCH ARTICLE OPEN ACCESS

# The Generative AI-Cybersecurity: A Review of Offensive & Defensive Applications, Strategic Implications, and Future Paradigms

Drashti Ranpariya[1], Ms Tosal Bhalodia [2]
1. (B Tech in Computer Science Engineering, Atmiya University, Rajkot, India
Email: drashtiranpariya20@gmail.com)
2. (Faculty of Engineering and Technology (CE), Atmiya University, Rajkot, India
Email: tosal.bhalodia@atmiyauni.ac.in)

**Abstract-**These reviews the harmful uses of GenAI, describing how it increases dangers through rapid exploit, creation, mysterious polymorphic malware and automated phishing attempts Meanwhile, it examines the new protective applications where GenAI enhances digital barriers through proactive threat identification, incident response automation, and secure code creation. The study then moves to the strategic level, looking at the important implications for cyberwarfare, including the growing asymmetric warfare tactics that endanger critical civilian systems and the escalating weapons war fueled by AI. The issue of safeguarding AI models, the urgent requirement for responsive regulation, and future paradigms impacted by self-governing agentic AI are all covered in the paper's conclusion. The primary conclusion is that operating in the GenAI age necessitates a co-evolutionary defensive approach that emphasizes AI safety measures to reduce AI risks while also using AI as a safeguard.

## I. INTRODUCTION: GENERATIVE AI'S TWO-SIDED SWORD

The emergence of Generative Artificial Intelligence is ushering in a revolutionary era in cyber security. GenAI, or intelligent technology. Unlike classical AI, which focused on analysis, GenAI has a strong creative power. A few instances of systems that use deep learning to analyze massive datasets and produce new content, such as text images and program code, that is frequently unrecognizable from actual human content are Generative Adversarial Networks (GANs) and enormous Language Models (LLMs). The transition from research to synthesis is an important turning point that radically changes the strategies and resources available to both online criminals and defenders. At the core of this transformation are sophisticated models that have expanded the possibilities of Neutral Language Processing (NLP), such as OpenAI's GPT Series and Google's Gemini. One As these LLMs have been trained on vast amounts of internet data, they are able to create beneficial applications and write human-like text with barely any help. They are enhanced by GANs, which produce remarkably realistic results by using a "generator" network to generate artificial data and a “discriminator" network to try to separate it from real data. Because GenAI is inherently dual-use, it has a significant impact on safety. Similar capabilities that present previously unheard-of chances to strengthen digital defenses also give adversaries a potent new toolkit. This leads to a paradox whereby advancements intended to protect digital infrastructures also make cyberattacks more sophisticated and widespread, resulting in a dynamic conflict where both sides use the same core technology.

## II. OFFENSIVE APPLICATIONS: MAKING THE ENVIRONMENT RISKIER

By eliminating away, the thorough knowledge that was previously necessary for complex cyber operations, generative AI is fundamentally changing the threat landscape. It significantly lowers the bar for less experienced actors while acting as a potent force amplifier for groups with greater resources, allowing them to function at a scale and speed that were previously unachievable. The art of social engineering has benefited from GenAI LLMs have a much greater success rate than previous attempts because they are skilled at crafting "highly believable fraud emails" that imitate real language as well as tone. Attackers can develop highly customized campaigns on a large scale by employing strategies like prompt chaining. The risk also includes extremely lifelike deepfakes, in which sophisticated vishing (voice phishing) frauds use artificial intelligence (AI)-generated video and audio to mimic well-known people, such as corporate officials. This marks the beginning of " Social Engineering 2.0," a fresh, highly customized, artificial intelligence-driven concept of dishonesty. Additionally, GenAI innovates and automates the generation of malware. Now, an attacker can instruct an LLM to create code for harmful purposes or modify pre-existing malware to target different platforms. Consequently, malware that is metamorphic and polymorphic, that can change its code with every infection in order to avoid detection by signatures, is now generated automatically. The emergence of malicious-focused AI models on hidden forums such as WormGPT and FraudGPT, making them accessible to a wider audience as a subscription-based business. Additionally, recent research has shown that LLM agents, such as GPT-4, can autonomously take advantage of "one-day" vulnerabilities, weaknesses that have been discovered but not yet fixed.

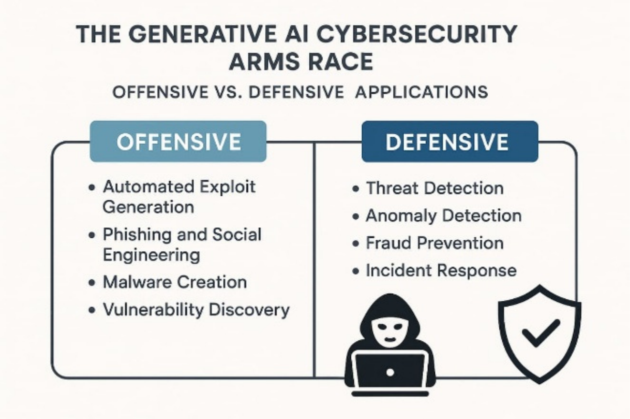
THE GENERATIVE AI CYBERSECURITY ARMS RACE
OFFENSIVE VS. DEFENSIVE APPLICATIONS
OFFENSIVE
• Automated Exploit Generation
• Phishing and Social Engineering
• Malware Creation
• Vulnerability Discovery
DEFENSIVE
• Threat Detection
• Anomaly Detection
• Fraud Prevention
• Incident Response

Fig. 1 Offensive vs Defensive Applications

## III. DEFENSIVE APPLICATIONS: FORTIFYING THE DIGITAL BARRIERS

GenAI accelerates the shift away from a defensive, signature-focused stance towards one that is prepared, anticipatory and adaptive by providing defenders with an equally powerful arsenal as it does attackers. GenAI, which recognizes a network's unique "style of existence" to identify subtle changes that can point to an attack in its early phases, enables a more sophisticated approach. GenAI's remarkable capacity to scan vast volumes of raw information from numerous sources and automatically generate meaningful threat analysis reports enables security teams to anticipate the behaviors of attackers. Automatic identification of anomalies is its main strength; AI-driven systems establish a comprehensive baseline of normal activity and quickly spot subtle abnormalities that traditional methods would overlook. GenAI is a powerful "copilot" SOC (Security Operations Center) uses to do time-consuming tasks like threat detection and log review. Because it can automatically create and run response to incidents playbooks, cutting reaction times from hours to seconds, senior engineers can focus on more strategic projects. One of the most significant defensive applications is the generation of artificial data. Large, accurate and coded datasets of illicit activity can be made available by GANs to train more resilient and strong defensive models, such as intrusion detection systems. Additionally, it is possible to copy actual threats for defensive testing and staff training. Lastly, by integrating security into the software development lifecycle, GenAI is providing a "shift-left" approach to whole system. Before code is actually shared vulnerabilities can be found and fixed with AI tools that are integrated into a developer's environment.

## IV. STRATEGIC IMPLICATIONS: DEVELOPMENTS IN CYBERWARFARE AND CYBERSECURITY

The use of GenAI in cyber warfare has noticeable strategic implications that are leading the global security to change.AI driven operation’s speed, scale, efficiency changing the whole nature of cyber warfare. It is challenge to generate asymmetries and great decision-making cycles at a level where mistakes are more likely unanticipated. Current developments in offensive and defensive AI are constantly helping each other forward in a rapidly expanding, mutually beneficial arms race. Attackers use GenAI to develop new methods for getting around as defenders deploy AI-native security platforms. This results in a never-ending conflict where the speed is set by machine learning cycles. There is growing consensus that frontier AI models might be tipping the offensive-defense balance in favor of the attacker. Complex defensive strategies are harder to automate than offensive tasks like vulnerability scanning, and attackers have shown that they can use new AI tools more quickly. GenAI also acts as a powerful asymmetric warfare amplifier by targeting vulnerable civilian infrastructure, enabling non-state actors and weaker countries to inflict significant damage. As evidenced by a simulated conflict between Israel and Iran, major powers may focus on hardened military targets while fewer- resources players can use AI to engage "softer" targets like hospitals, financial systems, and power grids, causing widespread psychological and economic disruption. However, the push to operationalize GenAI for espionage and warfare is being spearheaded by highly skilled nation-state actors.

## V. FUTURE PARADIGMS: THE WAY AHEAD

The cybersecurity and GenAI's relationship suggest a future characterized by growing autonomy and a recurrent set of issues. As AI systems become indispensable for both offensive and defensive operations, the focus will inevitably move from using AI to secure networks to the much more challenging task of securing the AI itself. "Agentic AI”, means systems that are capable of organizing and carrying out complex, complicated actions with very little human supervision, may be the next technological frontier. An autonomous offensive agent could be given a high-level objective and execute a full attack by itself, whereas a defensive agent might supervise an organization's entire security posture. Significant and potentially unstable is the shift from AI as a tool to AI as an actor.

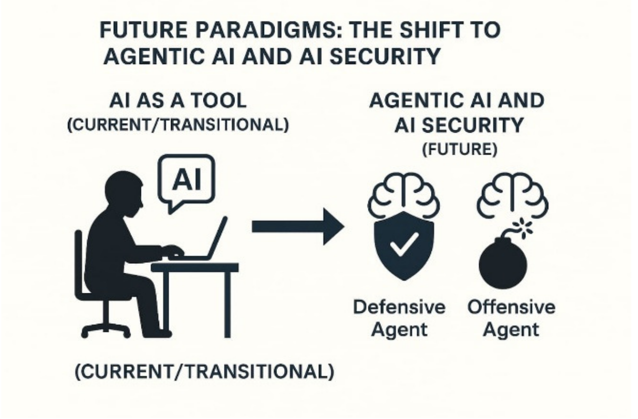

Fig. 1 Future Paradigms

Integrity of GenAI model is crucial as they are integrated into security. A new class of risks includes prompt injection, which tricks a model into avoiding its safety filters, data poisoning, which corrupts a model's training data to create hidden Trojans, and model extraction, which involves taking a model's structure or critical training data. The need to protect against these risks with a new set of defenses that are centered on the AI lifecycle itself, like adversarial programming and unique confidentiality, is making AI safety and AI security inseparable. In order to manage these risks without limiting innovation, there is now a governance gap and a pressing need for industry standards like Google's Secure AI Framework (SAIF) and flexible regulatory frameworks like the NIST AI Risk Management Framework (AI RMF).

## VI. CONCLUSION

By integrating generative AI into cybersecurity, an entirely new conflicts have been established and it is leading to irreversible turning point. This paper has discussed in detail the double-sided nature of this technology, which serves as a fateful weapon for attackers and a powerful shield for defenders. On the offensive side, GenAI is making sophisticated cybercrime more accessible by facilitating automated vulnerability exploitation, the creation of elusive malware, and hyper-realistic social engineering. At the same time, it is automating incident response, transforming threat intelligence, giving defenders proactive and adaptive capabilities, and making it possible to build strong defensive models using synthetic data. This technological dualism has led to a high-stakes, reciprocally evolving arms race that is changing the strategic balance of power in cyberwarfare. It might increase threats against the society. The future requires a two-pronged strategy: a significant investment in AI-native defenses to keep up with evolving threats, and the concurrent development of robust governance and AI safety protocols to manage the inherent risks of the technology. As we move into an era of increasingly autonomous cyberspace, cybersecurity's future will depend on our ability to provide the critical strategic oversight and moral guidance needed to manage the complex interactions between artificial and human intelligence in the defense of our digital world.

## REFERENCES

[1] Metta, S., Chang, I., Parker, J., Roman, M. P., & Ehuan, A. F. (2024). Generative AI in cybersecurity. *arXiv preprint arXiv:2405.01674.*

[2] Mercado, V. A. (2025). Cyber Warfare and the Future of Conflict.

[3] Alauthman, M., Almomani, A., Aoudi, S., Al-Qerem, A., & Aldweesh, A. (2025). Automated Vulnerability Discovery Generative AI in Offensive Security. In *Examining Cybersecurity Risks Produced by Generative AI* (pp. 309-328). IGI Global Scientific Publishing

[4] Muñoz, A. V. AI in the Crosshairs: Advancing Cybersecurity and Digital Forensics in the Era of Intelligent Threats.

[5] Akeiber, H. J. (2025). A comprehensive study of Cybercrime and Digital Forensics through Machine Learning and AI. *Al-Rafidain Journal of Engineering Sciences*, 369-395.

[6] YAZI, G. (2024). Large language models (llms) for cybersecurity: A systematic review. *WORLD*, *13*(1), 057-069.

[7] Nadella, G. S., Addula, S. R., Yadulla, A. R., Sajja, G. S., Meesala, M., Maturi, M. H., ... & Gonaygunta, H. (2025). Generative AI-Enhanced Cybersecurity Framework for Enterprise Data Privacy Management. *Computers*, *14*(2), 55.

[8] Alqahtani, H., & Kumar, G. (2025). A comprehensive review of generative AI techniques and their impact on cybersecurity. *Soft Computing*, 1-38.

[9] Arifin, M. M., Ahmed, M. S., Ghosh, T. K., Udoy, I. A., Zhuang, J., & Yeh, J. H. (2024). A survey on the application of generative adversarial networks in cybersecurity: Prospective, direction and open research scopes. *arXiv preprint arXiv:2407.08839.*

[10] Solutions, C. Leveraging Artificial Intelligence (Ai) Competencies For Next-Generation Cybersecurity Solutions.

[11] Sharma, D., Tomar, G. S., & Jha, A. (Eds.). (2025). *Artificial Intelligence for Cyber Security and Industry 4.0*. CRC Press.

[12] AI, N. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile. *NIST Trustworthy and Responsible AI Gaithersburg, MD, USA.*

[13] Janjeva, A., Harris, A., Mercer, S., Kasprzyk, A., & Gausen, A. (2023). The rapid rise of generative AI. *Centre for Emerging Technology and Security.*

[14] Devetzis, D., Volosevici, D., & Sotiropoulos, L. D. Digital Lawscapes: Artificial Intelligence, Cybersecurity and the New European Order.

[15] Fang, R., Bindu, R., Gupta, A., & Kang, D. (2024). Llm agents can autonomously exploit one-day vulnerabilities. *arXiv preprint arXiv:2404.08144.*