Archives available at journals.mriindia.com

International Journal of Recent Advances in Engineering and Technology

ISSN: 2347 - 2812

Volume 15 Issue 01, 2026

A Proactive AI Framework for Detecting and Mitigating Mutating Malware Generated via Generative AI Models: A Literature Review

[1]Avinash Wasnik, [2]Shweta Meshram

[1]Research Scholar, Department of Computer Management, MES Institute of Management & Career Course (IMCC), Savitribai Phule Pune University, Pune, India

[2]Associate Professor, Department of Computer Management, MES Institute of Management & Career Course (IMCC), Savitribai Phule Pune University, Pune, India

Email: [1]avinash.wasnik@outlook.com, [2]sdm.imcc@mespune.in

Peer Review Information

Submission: 08 April 2026

Revision: 29 April 2026

Acceptance: 11 May 2026

Keywords

Generative AI, Mutating Malware, Proactive Detection, Cybersecurity, Adversarial Learning, Behavioural Analysis

Abstract

The cybersecurity landscape has changed due to the rapid development of generative artificial intelligence (AI), which empower attackers to generate highly adaptive malware that can bypass currently available detection methods. This study reviews current research (2023–2026) on initiative-taking AI-driven strategies done on purpose to identify and get rid of such real time threats. It looks at advancements in behavioural analytics, polymorphic malware, adversarial machine learning, and predictive threat intelligence. The study underlines the need for flexible, intelligence-driven defence mechanisms and draws attention to the diminishing efficacy of signature-based systems. Important discoveries show that to successfully fight AI-generated malware, future cybersecurity solutions must integrate behavioural monitoring, ensemble learning, and anticipatory threat modelling. In the paper's conclusion, research gaps are listed and strategies for developing robust, next-generation detection frameworks are suggested.

Introduction

1. Changing Nature of Cyber Threats

Artificial intelligence is becoming a crucial component of both attack and defence tactics, which is causing a significant shift in cybersecurity [1]. Modern AI-enabled malware can dynamically alter its structure and behaviour, in contrast to traditional malware, which depends on static code patterns. Because these threats can adapt in real time and evade detection methods that rely on predefined signatures, this change creates new difficulties [2].

According to recent developments, attackers are using sophisticated AI models, such as generative adversarial networks and large language models, to automate the creation of malware and improve evasion strategies. These features make it much more difficult to identify and analyse malicious programs by enabling them to create new variations while they are running [3].

2. AI as a Force Multiplier for Attackers

The technical barrier needed to create complex cyberattacks has decreased thanks to generative AI. It is now possible to partially automate tasks like malicious code generation, vulnerability identification, and reconnaissance. Consequently, state-sponsored actors and organized cybercriminals are increasingly incorporating AI into their operations.

AI-driven malware demonstrates a number of sophisticated traits [2]:

- Its internal structure is constantly changing.
- Adaptation to the intended environment

- The capacity to imitate acceptable system behaviour.
- Adapting in real time to defensive measures.
- Making decisions on your own while carrying out

The complexity of detection and response is greatly increased by these features.

3. Research Motivation

The rise of adaptive malware draws attention to the shortcomings of conventional security systems, which rely on recognized attack signatures. These systems are useless against constantly changing threats and zero-day exploits [5]. Proactive methods that can spot suspicious activity before harm is done are therefore becoming more and more necessary.

The purpose of this study is to:

- Examine current developments in malware produced by AI.
- Examine initiative-taking detection methods.
- Analyse detection techniques based on machine learning.
- Explore adversarial strategies used by attackers.
- Point out research gaps and suggest future lines of inquiry.

4. Scope of the Study

The literature released between 2023 and 2026 is the focus of this review. It discusses advances in adversarial tactics, machine learning methods, explainable AI, detection frameworks, and malware evolution. The paper is prepared to give readers a thorough grasp of the problems and innovative solutions in this field.

Evolution of AI-Generated Malware

1. Emergence of Runtime-Generated Malware

A notable advancement in malware design is the ability to generate malicious code during execution rather than relying on pre-written payloads [1]. This approach produces unique instances each time the malware runs, making detection through traditional methods extremely difficult.

Such malware can adapt its behaviour based on the environment, learning from previous detection attempts and modifying its strategy accordingly [2].

2. Key Characteristics

AI-driven malware differs from conventional threats in multiple ways [2]:

- Self-modification: Continuously alters its structure to evade detection.
- Dynamic payloads: Generates unique attack scripts for each target.
- Stealth behaviour: Mimics legitimate processes to avoid suspicion.
- Environmental awareness: Adjusts actions based on system conditions.
- Autonomous execution: Operates with minimal human intervention.

3. AI Across the Attack Lifecycle

Artificial intelligence is now used throughout the attack process [4]:

- Reconnaissance: Automated data collection and target analysis
- Development: Rapid creation of malware variants
- Delivery: Personalized phishing and social engineering
- Post-attack: Optimized lateral movement and data exfiltration

This end-to-end integration increases both efficiency and effectiveness of cyberattacks.

4. Evasion Techniques

Modern malware uses advanced techniques to bypass detection systems, including [6]:

- Code transformation and obfuscation.
- Reordering execution sequences
- Generating adversarial inputs to mislead detection models

Adversarial learning methods, particularly those involving generative models, are used to systematically identify and exploit weaknesses in security systems [8].

Proactive Detection Frameworks

1. Shift Toward Proactive Security

Traditional cybersecurity focuses on responding to known threats. However, evolving attack patterns require systems that can anticipate and prevent attacks before they occur. Proactive frameworks rely on predictive analytics and continuous monitoring rather than static rules [16].

2. Ensemble Learning Approaches

Combining multiple machine learning models improves detection accuracy and resilience. Ensemble methods offer [14]:

- Better generalization across diverse threats
- Reduced impact of individual model weaknesses
- Improved confidence in predictions

- Increased resistance to adversarial attacks

3. Hybrid Analysis Techniques

Effective detection requires combining [15]:

- Static analysis: Examining code without execution.
- Dynamic analysis: Observing runtime behaviour.

This integrated approach provides a more complete understanding of malware characteristics and improves detection performance [16].

4. Behavioural Detection

Behaviour-based systems focus on identifying anomalies rather than known signatures [17]. Indicators include:

- Unusual process behaviour
- Suspicious memory usage
- Abnormal network activity
- Unauthorized file modifications

Machine learning models trained on normal system behaviour can detect deviations that indicate potential threats [18].

5. Predictive Threat Intelligence

Predictive systems analyse historical and real-time data to forecast potential attacks. These systems help organizations:

- Anticipate attack vectors.
- Identify vulnerable assets.
- Prepare defensive strategies in advance.

This approach transforms cybersecurity from reactive defence to strategic planning [19] [20].

Research Gaps and Future Directions

1. Current Limitations

Despite progress, several challenges remain:

- Difficulty scaling advanced models for real-time use.
- Limited robustness against adaptive adversaries
- Lack of large datasets for training
- Rapid evolution of threats causing model degradation.

2. Integration Challenges

Existing systems often fail to effectively combine threat intelligence with detection mechanisms. Improved integration can enhance situational awareness and response prioritization [20].

3. Automated Response Systems

Future frameworks should include automated mitigation capabilities. However, safeguards are necessary to prevent incorrect actions caused by false positives [22].

4. Emerging Technologies

Quantum computing may influence both attack and defence strategies. Research is required to prepare for its impact on cybersecurity systems [23].

5. Standardization Needs

There is a lack of standardized benchmarks for evaluating detection systems. Establishing common datasets and metrics will improve research consistency and comparability [24].

6. Cross-Domain Learning

Future models should generalize across different platforms and malware types. Transfer learning techniques can help achieve this adaptability [25].

7. Human-AI Collaboration

Effective cybersecurity will require collaboration between automated systems and human experts. AI can handle large-scale analysis, while humans provide contextual judgment and strategic oversight [26].

Conclusion

The rise of AI-generated malware represents a significant challenge for modern cybersecurity. Traditional detection methods are no longer sufficient to address threats that continuously evolve and adapt. Proactive approaches that combine behavioural analysis, machine learning, and predictive intelligence offer promising solutions.

However, achieving robust and scalable systems requires addressing key challenges such as adversarial resistance, data limitations, and system integration. Future research must focus on developing adaptive, transparent, and collaborative frameworks capable of staying ahead of evolving threats.

This study provides a foundation for advancing proactive cybersecurity strategies in the era of generative AI, emphasizing the need for continuous innovation and interdisciplinary collaboration.

References

Google Threat Intelligence Group, “GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools,” Cloud Security Blog, 2025. https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools

SASA Software, “Adaptive Malware: Understanding AI-Powered Cyber Threats in

2025," SASA Blog, Dec. 28, 2025. https://www.sasa-software.com/blog/adaptive-malware-ai-powered-cyber-threats/

Sify, "Malware with AI-Powered Code Mutations: Google Sounds the Alarm!," Sify Technology News, Nov. 24, 2025. https://www.sify.com/ai-analytics/malware-with-ai-powered-code-mutations-google-sounds-the-alarm/

Cloudwars, "Cybercriminals Are Operationalizing AI: New Findings from Google Threat Intelligence Group Reveal Escalating Threats," Cloudwars AI, Feb. 24, 2026. https://cloudwars.com/ai/cybercriminals-are-operationalizing-ai

Pillai, R. A., and A. Dhamal, "Revolutionizing Cybersecurity: A Generative AI-Powered Malicious File and URL Detection Framework," International Journal of Innovative Science and Research Technology, vol. 10, no. 12, pp. 1341–1352, 2025.

Werne, J., "AI Agents, Malware Mutations & New Interfaces Shift Attack Surfaces," Jochen Werne Consulting, Jan. 11, 2026. https://jochenwerne.com/cyber-trends-2026-ai-agents-malware-mutations-new-interfaces-shift-attack-surfaces/

Viking Cloud, "5 Ways Generative AI Is Reshaping Cyber Defence Strategies," Viking Cloud Blog, Feb. 15, 2026. https://www.vikingcloud.com/blog/generative-ai-cybersecurity

Huntress, "What is Adversarial AI? Cybersecurity Threats and Defenses," Huntress Cybersecurity 101, Jul. 8, 2025. https://www.huntress.com/cybersecurity-101/topic/adversarial-ai-cybersecurity-threats-defenses

Al-Dujaili, A., et al., "Enhancing Adversarial Examples for Evading Malware Detection Using Memetic Algorithms," International Journal of Computer Network and Information Security, vol. 17, no. 1, pp. 1–15, 2023. https://www.mecs-press.org/ijcnis/ijcnis-v17-n1/v17n1-1.html

Everbridge, "Preparing For The AI-Driven 2026 Threat Landscape," Everbridge Blog, Jan. 25, 2026. https://www.everbridge.com/blog/ai-and-the-2026-threat-landscape/

Nature Publishing Group, "Ensemble Machine Learning for Proactive Android Ransomware Detection," Scientific Reports, vol. 16, Art. no. 38271, Feb. 17, 2026. https://www.nature.com/articles/s41598-026-38271-7

Uplatz, "Predictive Threat Intelligence: A Framework for Proactive Cyber Defence in the AI Era," Uplatz Blog, Sept. 22, 2025. https://uplatz.com/blog/predictive-threat-intelligence-a-framework-for-proactive-cyber-defense-in-the-ai-era/

Pillai, R. A., and A. Dhamal, "A Generative AI-Powered Malicious File and URL Detection Framework," International Journal of Innovative Science and Research Technology, 2025. https://ijisrt.com/assets/upload/files/IJISRT25DEC1341.pdf

Nature Publishing Group, "Ensemble Machine Learning for Proactive Android Ransomware Detection Using Network Traffic Metadata," Scientific Reports, vol. 16, Art. no. 38271, 2026.

International Journal of Computer Science and Information Technology, "Proactive Ransomware Detection with Machine Learning," IJCSIT, vol. 15, no. 6, pp. 87–102, 2024. https://www.ijcsit.com/docs/volume15/volume15issue6/ijcsit2024150603.pdf

International Journal of Applied Computer Technology, "Machine Learning-Based Detection of Malware Threats: A Proactive Approach to Cybersecurity," IJACT, vol. 3, no. 1, pp. 115–128, 2025. https://www.espjournals.org/IJACT/2025/Volume3-Issue1/IJACT-V3I1P115.pdf

SentinelOne, "What is Behavioural Threat Detection & How Has AI Improved It?," SentinelOne Cybersecurity 101, Oct. 16, 2025. https://www.sentinelone.com/cybersecurity-101/threat-intelligence/behavioral-threat-detection/

Webasha, "How is AI Used in Cyber Threat Detection and Real-Time Response to Zero-Day Exploits," Webasha Blog, Jul. 28, 2025. https://www.webasha.com/blog/how-is-ai-used-in-cyber-threat-detection-and-real-time-response-to-zero-day-exploits

Uplatz, "Predictive Threat Intelligence: A Framework for Proactive Cyber Defence in the AI Era," Uplatz Blog, Sept. 22, 2025.

Akitra, "The Role of Threat Intelligence in Proactive Cyber Defence," Akitra Blog, Apr. 3, 2025. https://akitra.com/blog/threat-intelligence-in-proactive-cyber-defense/

Cyble, “Predictive Threat Intelligence | AI Security,” Cyble Knowledge Hub, Feb. 19, 2026.

Uplatz, “Predictive Threat Intelligence: A Framework for Proactive Cyber Defence in the AI Era,” Uplatz Blog, Sept. 22, 2025.

Everbridge, “Preparing For The AI-Driven 2026 Threat Landscape,” Everbridge Blog, Jan. 25, 2026.

Frontiers in Artificial Intelligence, “A Systematic Review on the Integration of Explainable Artificial Intelligence in Intrusion Detection Systems,” Frontiers in AI, 2025.

ArXiv, “Deep Learning-Driven Malware Classification with API Call Sequence Analysis,” arXiv Preprint, Feb. 27, 2025.

International Academy of Computer and Information Sciences, “Improving Cybersecurity through Explainable Artificial Intelligence,” 2025.