ALLIANCE PRO

CYBERSECURITY WHITE PAPER

When the Attacker Uses AI:

The Rise of AI-Augmented Threat Actors

How Groups Like GREYVIBE Are Using Generative AI to Supercharge Cyberattacks—and What Your Organization Must Do Now

Published: June 2026 | Alliance Pro Cybersecurity Division

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

Executive Summary

A new era of cyber warfare is here. Generative artificial intelligence (AI) has crossed from the defender's toolkit into the attacker's arsenal—and the results are alarming. In May 2026, threat intelligence firm With Secure published research on GREYVIBE, a previously untracked threat group actively using ChatGPT, Google Gemini, and Ideogram AI to design phishing lures, build malware, and accelerate attack operations against Ukrainian government, military, and civilian targets. GREYVIBE is not an outlier. It is a preview of what every enterprise faces today.

This white paper examines the GREYVIBE case in depth, situates it within the broader trend of AI-powered attacks, explores the technical and strategic implications for organizations worldwide, and provides a concrete defensive framework that security leaders can act on immediately.

Key Findings at a Glance

- AI-enabled cyberattacks grew 89% year-over-year in 2025 (CrowdStrike 2026 Global Threat Report)
- The average attacker breakout time fell to just 29 minutes, down 65% from the prior year
- 82.6% of all phishing emails now contain AI-generated content (Keepnet/VIPRE 2025)
- AI-generated phishing achieves a 54% click-through rate versus 12% for traditional campaigns
- The global cost of a data breach reached $4.88 million on average in 2024 (IBM 2025)
- Verizon’s 2026 DBIR found the median threat actor used AI across 15 MITRE ATT&CK techniques
- ChatGPT was mentioned in criminal forums 550% more than any other AI model (CrowdStrike 2026)

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

Table of Contents

1	Executive Summary	2
2	The Threat Landscape: How AI Changed Everything	4
3	Case Study: Inside GREYVIBE	6
4	AI Across the Attack Lifecycle	8
5	The Paradox: AI Also Exposes Attackers	10
6	Industry Impact & Risk Sectors	11
7	The Defender’s Blueprint: 10 Actionable Strategies	12
8	Conclusion & Strategic Outlook	15
9	Sources & References	16

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

The Threat Landscape: How AI Changed Everything

For decades, the most dangerous cyberattackers were also the most technically sophisticated. Nation-state actors and elite criminal groups invested years in mastering exploitation frameworks, reverse engineering, and custom malware development. That barrier is gone. Generative AI has democratized offensive cyber capability at a scale and speed that has no historical precedent.

2.1 The Numbers Are Alarming

The data from 2025 and early 2026 paints a stark picture:

89% AI-enabled attack growth (YoY)	**29 min** Average attacker breakout time
82.6% Phishing emails using AI content	**54%** AI phishing click-through rate
28M+ Global AI-driven incidents in 2025	**87%** Organizations hit by AI attacks
$4.88M Average data breach cost (2024)	**Top performers** Orgs that saved $1.9M with AI defense

Sources: CrowdStrike 2026 Global Threat Report; IBM Cost of a Data Breach 2025; Keepnet/VIPRE 2025; Verizon DBIR 2026

2.2 AI as a Force Multiplier for Attackers

The CrowdStrike 2026 Global Threat Report **found that the fastest recorded breakout time in 2025 was an extraordinary 27 seconds**—down from minutes in prior years. AI is not making attackers smarter; it is making them dramatically faster, cheaper, and more scalable.

550%	More mentions of ChatGPT in criminal hacking forums vs. any other AI model (CrowdStrike 2026)

Consider what AI enables an attacker to do in a single afternoon: craft hundreds of hyper-personalized spear-phishing emails; generate functional malware code without writing a single line; produce convincing fake websites, CAPTCHA pages, and decoy documents; and automate post-compromise reconnaissance and data exfiltration. All of this was once the domain of elite teams. Today it is available to any threat actor with an internet connection.

2.3 The Verizon 2026 DBIR: A Landmark Dataset

Published in May 2026, the **Verizon Data Breach Investigations Report** examined 793 threat actors in detail and found the median actor used AI across **15 MITRE ATT&CK techniques**, with some leveraging it across 40–50 techniques. AI-assisted text in phishing emails **doubled year-**

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

over-year. The report also highlighted VoidLink—a malware framework an AI agent assembled in just six days—and PromptLock, described as the first AI-powered ransomware to dynamically generate cross-platform encryption scripts. Verizon framed these developments as a ***point of no return*** for automated threat development.

2.4 The Role of Public AI Platforms

Perhaps the most unsettling development is the use of mainstream commercial AI platforms—ChatGPT, Gemini, Grok—by threat actors. Over 90 organizations in 2025 had legitimate AI tools exploited to generate malicious commands and steal sensitive data (CrowdStrike). In IBM's experiment, AI needed only 5 prompts and 5 minutes to build a phishing campaign as effective as one that took human experts 16 hours. This is not a future threat. It is the current operating environment.

1,265%	Surge in phishing attacks linked to generative AI since 2023 (FBI / Security researchers)

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

Case Study: Inside GREYVIBE

GREYVIBE **represents one of the most comprehensively documented examples of AI-augmented offensive operations to date.** Identified and tracked by WithSecure in May 2026, the group provides a real-world window into how even a moderately skilled threat actor can execute sophisticated, multi-vector campaigns using generative AI as a force multiplier.

3.1 Who Is GREYVIBE?

GREYVIBE is a previously untracked threat group active since at least August 2025. While no definitive attribution has been established, WithSecure researchers identified strong alignment with Russian state interests: Russian-language artifacts across malware samples, operational activity concentrated in Moscow business hours, and targeting almost exclusively aligned with Ukrainian government, military, and civilian infrastructure.

The group exhibits a hybrid threat profile—overlapping with known cybercriminal infrastructure—suggesting possible links to former or active cybercriminal actors operating under state direction or opportunistic alignment.

3.2 Multi-Vector Attack Architecture

GREYVIBE does not rely on a single attack vector. Instead, it deploys a layered approach combining three distinct infection chains:

VECTOR	DESCRIPTION
Spear-Phishing	AI-crafted emails impersonating Ukrainian government agencies. Malicious archives delivered via Google Drive silently execute decoy documents while initiating infection chains using custom loaders (DAYLIGHT, TEASOUP).
Fake CAPTCHA Pages	Deceptive verification pages trick victims into running malicious commands. AI was used to generate Ukrainian-language instructions, making the pages indistinguishable from legitimate verification flows.
Deceptive Websites	Fake ‘adult club’ websites targeting Ukrainian military personnel. Used for both malware delivery (FallSpy for Android, PhantomRelay for Windows) and social engineering via fake Telegram personas.

3.3 The Malware Toolkit

GREYVIBE deployed a suite of custom tools, each serving a distinct operational role:

- PhantomRelay (Windows): A modular Remote Access Trojan (RAT) using WebSockets for command execution. Enables file theft, screenshot capture, and exfiltration of messaging application data.
- FallSpy (Android): Spyware that exfiltrates contacts, real-time GPS location, device metadata, and communication history from mobile devices.
- LegionRelay: A PowerShell-based RAT. AI-generated code patterns were detected in its core modules. Despite its effectiveness, design flaws exposed its backend infrastructure to researchers.

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

- DAYLIGHT / TEASOUP: Custom loaders responsible for staging and executing payloads. AI-generated obfuscation patterns were identified across both.

3.4 AI's Role in GREYVIBE's Operations

WithSecure researchers identified AI involvement across multiple stages of GREYVIBE's attack chain. The use of ChatGPT, Google Gemini, and Ideogram AI is evidenced by characteristic LLM output markers embedded in both code and generated images. AI use was identified in:

- Phishing lure generation (text, imagery, decoy documents)
- Malware component development (loaders, obfuscators, RAT modules)
- Post-compromise activity support (data analysis, exfiltration logic)
- Image and asset generation for fake websites (Ideogram AI)

6 Days	Time for AI to assemble the VoidLink malware framework from scratch (Verizon DBIR 2026)—comparable sophistication to GREYVIBE's toolkit

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

AI Across the Attack Lifecycle

GREYVIBE's use of AI is not unique to any one phase of its operations. This reflects a broader industry trend: AI is now being integrated across the entire cyber kill chain, from initial reconnaissance to post-compromise persistence. Understanding where AI intervenes at each stage is critical for defenders.

PHASE	AI TECHNIQUE	GREYVIBE EXAMPLE
Reconnaissance	AI-assisted OSINT, target profiling, social graph mapping	*Profiling Ukrainian military personnel via Telegram personas*
Initial Access	AI-generated spear phishing, deepfake content, fake verification pages	*Ukrainian-language CAPTCHA pages; government agency impersonation emails*
Execution	AI-written loaders, obfuscators, polymorphic code	*DAYLIGHT/TEASOUP loaders with AI-generated obfuscation*
Persistence	AI-assisted RAT development, WebSocket C2 protocols	*PhantomRelay and LegionRelay with AI-generated core modules*
Exfiltration	AI-optimized data prioritization and compression	*FallSpy mobile exfil; LegionRelay file theft and screenshot capture*
Defense Evasion	AI-generated polymorphic code, reduced code reuse	*Novel obfuscation patterns making traditional signature matching ineffective*

4.1 The Democratization Problem

The strategic implication is profound: **AI has permanently lowered the technical barrier to conducting advanced persistent threat (APT)-level operations.** Groups that previously lacked the skills to develop custom malware, craft convincing multilingual phishing, or build resilient C2 infrastructure can now do all three with commercial AI tools.

This fundamentally changes the defender's calculus. Organizations can no longer assume that their attackers require significant resources or expertise. The threat surface has expanded to include any motivated actor with access to a laptop and an AI subscription.

$30B	Projected global damages from AI-assisted cyberattacks in 2025 (multiple security research firms)

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

The Paradox: AI Also Exposes Attackers

One of the more counterintuitive findings from the GREYVIBE investigation is that the group's reliance on AI introduced significant operational security failures. This paradox—AI as both amplifier and liability—is a critical insight for defenders.

5.1 How GREYVIBE's AI Use Created Exploitable Weaknesses

WithSecure researchers identified that AI-generated code in LegionRelay contained design weaknesses that exposed the RAT's backend infrastructure. These flaws allowed researchers to monitor attacker activity over an extended period—essentially turning the group's own tool into a surveillance mechanism against them.

The LLM markers present in generated images also served as attribution evidence, allowing analysts to link assets across campaigns that would otherwise appear unconnected. AI-generated content carries detectable patterns—stylistic, structural, and statistical—that experienced threat hunters can identify.

5.2 Operational Immaturity in AI-Augmented Groups

GREYVIBE exhibited several signs of operational immaturity that are common in groups that leverage AI to compensate for limited technical expertise:

- Uploading test malware samples to public platforms (VirusTotal), exposing tooling before deployment
- Inconsistent use of infrastructure, leaving traceable overlaps between campaigns
- AI-generated code patterns that, while functional, lacked the hardening of hand-crafted equivalents
- Overreliance on a small number of AI-generated asset styles, creating detectable signatures

5.3 Implications for Threat Intelligence

This paradox has significant implications for threat intelligence programs. Organizations that invest in behavioral analytics, LLM output detection, and AI-assisted threat hunting can potentially identify and track AI-augmented attackers more reliably than traditional signature-based methods allow. The very tools attackers use to scale their operations leave new types of fingerprints.

Key Takeaway for Defenders

AI-generated content is detectable. Investing in LLM output pattern detection, behavioral analytics, and AI-assisted threat hunting can turn an attacker's use of AI into an attribution advantage.

The GREYVIBE case proves that AI adoption by threat actors does not automatically translate to undetectable operations. Operational security discipline remains a human problem—and most threat actors are bad at it.

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

Industry Impact & Risk Sectors

AI-augmented threats are not sector-agnostic. Analysis of 2025 incident data reveals that certain industries face dramatically elevated risk—either due to the sensitivity of their data, the criticality of their operations, or the legacy nature of their security infrastructure.

76% Rise in targeted AI attacks on healthcare (2025)	**High-priority target** AI-driven incidents in government/public sector
56% Rise in AI-enabled attacks in Asia-Pacific	**31%** Rise in AI-assisted espionage in Middle East
54% vs 12% AI phishing vs. traditional click rate	**$1.9M avg** Orgs with AI defense saving per breach

6.1 Sectors Under Greatest Pressure

Based on 2025–2026 incident data, the following sectors face the highest exposure to AI-augmented threats:

- Government & Defense: Primary targets of state-aligned groups like GREYVIBE. AI enables highly convincing impersonation of agencies and officials. Sensitive operational data creates extreme breach consequences.
- Healthcare: A 76% rise in targeted AI attacks in 2025. Patient data commands premium prices on criminal markets. Ransomware disruption has life-or-death consequences.
- Financial Services: AI-powered Business Email Compromise (BEC) fraud logged a 37% rise in 2025 per the FBI IC3 report. $2.77 billion in BEC losses reported from 21,442 complaints.
- Critical Infrastructure (Energy, Utilities): Middle East energy sector reported a 31% rise in AI-assisted espionage campaigns. OT/ICS systems represent catastrophic risk if compromised.
- Technology & SaaS: Supply chain attacks via compromised developer tooling are rising sharply. Source code theft enables downstream attacks at scale.

6.2 The Human Element Remains Central

Despite AI's expanding role, the human element remains the primary attack surface. The Verizon DBIR 2025 found the human element involved in 62% of breaches. AI makes this problem worse, not better: AI-generated phishing eliminates the traditional red flags (poor grammar, generic salutations) that employees are trained to spot. Traditional security awareness training is now insufficient on its own.

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

The Defender's Blueprint: 10 Actionable Strategies

The GREYVIBE case and the broader AI threat landscape demand a fundamental rethink of enterprise cybersecurity. The following ten strategies represent a prioritized, practical framework for organizations seeking to reduce their exposure to AI-augmented attacks.

1. Deploy AI-Powered Security Operations

Traditional SIEM and rule-based detection cannot keep pace with polymorphic, AI-generated threats. Organizations using AI in security operations identified breaches in 148 days and contained them in 42 days, compared to 168 and 64 days respectively for those without (IBM 2025). Deploy AI-driven EDR, SIEM, and SOAR platforms that can detect behavioral anomalies rather than relying solely on signatures. Prioritize solutions with automated triage and response capabilities.

2. Implement Behavioral-Based Phishing Detection

AI-generated phishing has eliminated the traditional markers humans are trained to detect. Security tools must shift to behavioral analysis: does this email's metadata match the claimed sender? Is this domain newly registered? Does the link redirect through unusual infrastructure? AI-generated text detection models can flag LLM-produced lures before they reach inboxes. AI phishing achieves a 54% click-through rate—behavioral detection is the only scalable defense.

3. Mandate Phishing-Resistant MFA Everywhere

The average breakout time of 29 minutes means that stolen credentials alone can enable full compromise before a human analyst responds. Phishing-resistant MFA (FIDO2/passkeys) eliminates the credential theft vector entirely. Deploy it for all privileged accounts and internet-facing applications as an immediate priority.

4. Invest in AI Red Team Exercises

Organizations must test their defenses against AI-generated attacks—not just human-crafted ones. Conduct quarterly AI red team exercises that simulate AI-generated spear phishing, CAPTCHA-based delivery, and polymorphic malware. Measure detection rates and time-to-respond specifically against these scenarios. Assume your current defenses will fail the first time.

5. Establish an AI Governance Policy

The Arctic Wolf 2025 Human Risk Report found that 60% of IT leaders admit to feeding confidential data into public GenAI tools. Shadow AI use is a major and underappreciated attack surface. Establish clear policies governing which AI tools employees may use, what data categories may be processed, and what controls are required. Implement Data Loss Prevention (DLP) tools with AI platform awareness.

6. Modernize Security Awareness Training

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

AI has made traditional phishing awareness training obsolete as a primary defense. The goal is no longer to teach employees to spot bad grammar. Programs must now focus on: reporting culture (any suspicious communication, regardless of quality), verification protocols (out-of-band confirmation for any financial or credential-related requests), and continuous simulation using AI-generated attack scenarios that mirror real threat actor capabilities.

7. Monitor for LLM Output Signatures in Threats

AI-generated content leaves detectable patterns: stylistic consistency, structural repetition, and statistical regularities in word choice and sentence construction. Threat intelligence teams should build or acquire detection capabilities specifically tuned to identify AI-generated phishing lures, malware code, and attacker infrastructure assets. This capability turned GREYVIBE's AI use into an attribution advantage for WithSecure.

8. Accelerate Zero Trust Architecture Adoption

Zero Trust Architecture (ZTA)—"never trust, always verify"—is the architectural response to a world where perimeter-based security is insufficient. In an AI-attack environment, where credentials may be compromised quickly and lateral movement is rapid, continuous verification of every access request is essential. Prioritize micro-segmentation, least-privilege access, and continuous authentication for all sensitive systems.

9. Build a Threat Intelligence Program

The GREYVIBE group was identified through consistent infrastructure overlaps, tooling reuse, and behavioral signatures. Organizations that subscribe to commercial threat intelligence feeds and contribute to information-sharing communities (ISACs, CERTs) gain early warning of emerging AI-augmented threat groups. Threat intelligence directly informed the GREYVIBE detection and should be a standing budget line item.

10. Prepare an AI-Specific Incident Response Plan

Most incident response plans were written for a pre-AI threat environment. They assume human-speed attacks and human-detectable indicators. Organizations must develop AI-specific playbooks that account for: sub-30-minute breakout times; AI-generated malware with no known signatures; multi-vector campaigns that launch simultaneously across email, mobile, and web; and the possibility of AI-orchestrated post-compromise persistence.

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

Conclusion & Strategic Outlook

GREYVIBE is not the last AI-augmented threat group. It is the first one we caught documenting itself clearly enough to study. **For every GREYVIBE, there are dozens of groups operating with greater sophistication and operational security.**

The convergence of generative AI and offensive cyber operations represents a structural, permanent shift in the threat landscape. The barriers to entry for advanced cyberattacks have collapsed. The speed of attacks has accelerated beyond human-response thresholds in some scenarios. The scale of operations that a single threat actor can sustain has multiplied.

At the same time, AI is the most powerful defensive tool the security industry has ever had. Organizations using AI-powered security operations achieve measurably better outcomes: faster detection, faster containment, lower breach costs. The challenge is that defenders must adopt AI at pace with or ahead of attackers—and most organizations are currently behind.

The Strategic Imperative for 2026 and Beyond

1. Treat AI-augmented attacks as the baseline, not the exception. All threat models must account for AI-enabled adversaries.
2. Invest in AI-powered defense as a board-level priority. The $1.9M average saving per breach for AI-enabled organizations makes the ROI case compelling.
3. Build detection capabilities specifically for AI-generated content across phishing, malware, and infrastructure.
4. Close the governance gap on internal AI use before attackers exploit it through data leakage and shadow AI.
5. Treat incident response planning as a living document—update it quarterly as the AI threat landscape evolves.

The organizations that thrive in this environment will be those that match the attacker's AI adoption with equal or greater investment in AI-powered defense, governance, and human capability. The window to act is now, before the next GREYVIBE—one with better operational security—goes undetected for months or years.

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.

Sources & References

All statistics and findings cited in this white paper are drawn from primary research reports, peer-reviewed threat intelligence, and authoritative industry sources. Full citations are provided below.

1. WithSecure Labs. (2026, May). GREYVIBE Threat Intelligence Report. https://labs.withsecure.com/publications/greyvibe
2. CrowdStrike. (2026). 2026 Global Threat Report. https://www.crowdstrike.com/en-us/global-threat-report/
3. Verizon. (2026, May). 2026 Data Breach Investigations Report (DBIR). https://www.verizon.com/business/resources/reports/dbir/
4. IBM Security. (2025). Cost of a Data Breach Report 2025. https://www.ibm.com/reports/data-breach
5. Cybersecurity News. (2026, May 30). GREYVIBE Hackers Leverage ChatGPT and Google Gemini to Fuel Cyberattacks. https://cybersecuritynews.com/greyvibe-hackers-chatgpt-and-google-gemini/
6. Security Boulevard. (2026, May). The AI Governance Gap: Verizon's 2026 DBIR Shows Attackers Scaling AI While Employees Leak Data Through It.
7. Global News / CrowdStrike. (2026, February). Cyberattacks are hitting faster with AI fuelling an 89% jump.
8. AllAboutAI. (2025). AI Cyberattack Statistics 2026: What the Data Warns Us About. https://www.allaboutai.com/resources/ai-statistics/ai-cyberattack/
9. SQ Magazine. (2025). AI Cyber Attacks Statistics 2025. https://sqmagazine.co.uk/ai-cyber-attacks-statistics/
10. The Network Installers. (2025). AI Cyber Threat Statistics: The 2025 Landscape. https://thenetworkinstallers.com/blog/ai-cyber-threat-statistics/
11. Brightside AI. (2025). AI-Generated Phishing vs Human Attacks: 2025 Risk Analysis. https://www.brside.com/blog/ai-generated-phishing-vs-human-attacks-2025-risk-analysis
12. StrongestLayer. (2026). AI-Generated Phishing: The Top Enterprise Threat of 2026. https://www.strongestlayer.com/blog/ai-generated-phishing-enterprise-threat
13. StationX. (2026). Phishing Statistics 2026: Latest Attack Data & Trends. https://app.stationx.net/articles/phishing-statistics
14. Arctic Wolf. (2026, March). Who's Winning the AI Arms Race? https://arcticwolf.com/resources/blog/who-is-winning-the-ai-arms-race-threat-actors-or-cybersecurity-defenders/
15. Sting RAI. (2026). AI Cybersecurity Threats 2026 Defender Reference. https://www.stingrai.io/blog/ai-cybersecurity-threats-2026
16. Industrial Cyber. (2026). Verizon DBIR finds vulnerability exploitation overtakes stolen credentials. https://industrialcyber.co/reports/
17. DeepStrike. (2025). AI Cyber Attack Statistics 2025. https://deepstrike.io/blog/ai-cyber-attack-statistics-2025
18. FBI Internet Crime Complaint Center (IC3). (2025). 2024 Internet Crime Report.
19. Microsoft Security. (2025). Microsoft Cyber Signals Issue 9: AI-Powered Fraud. https://www.microsoft.com/en-us/security/blog/
20. Gartner. (2026). Global Information Security Spending Forecast. https://www.gartner.com/en/newsroom/

© 2026 Alliance Pro Cybersecurity Division. All rights reserved.