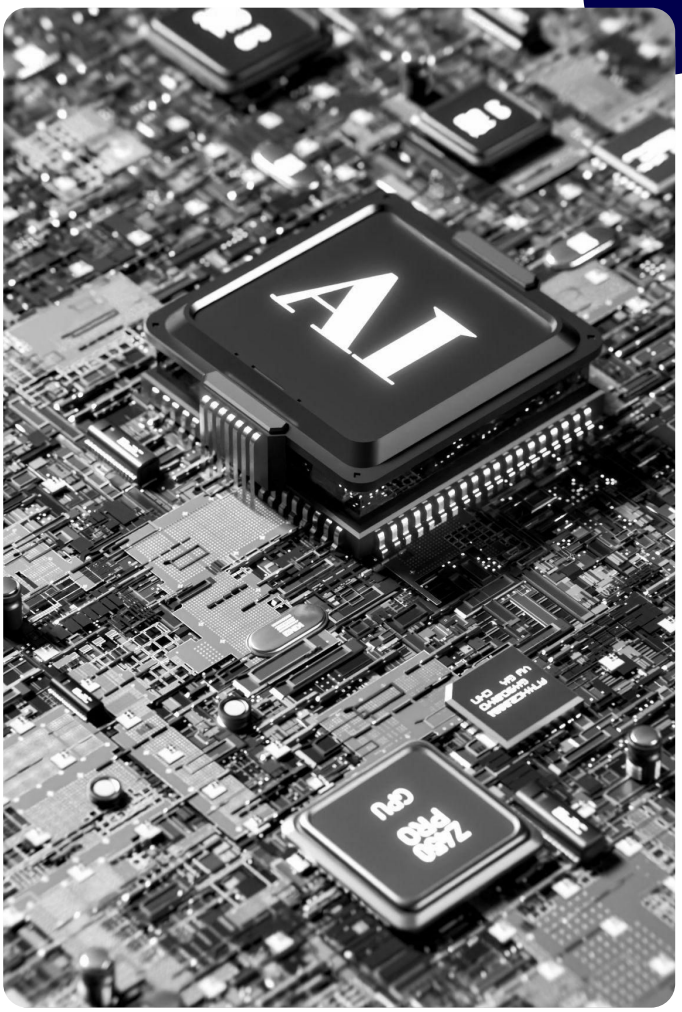
Recorded Future®

EXECUTIVE INSIGHTS REPORT | AUGUST 2025

Emerging AI Threats: The Future of Automated Operations

From Insikt Group®



Summary

Generative artificial intelligence enhances traditional malicious cyber operations, with early use cases indicating large-language models (LLMs) can be successfully used to automate malware. Threat actors are more frequently using generative AI as a force-multiplier for social engineering, malware development, and reconnaissance.

Criminals opportunistically exploit commercially available AI tools, using techniques like prompt engineering to bypass safety guardrails. AI tools specifically trained for malicious activity, like FraudGPT or WormGPT, are available on criminal marketplaces for threat actors who lack the skill set to modify commercially available AI tools.

While state-sponsored actors from China, Iran, and Russia take advantage of commercial and open-source AI, they are also focused on long-term strategies for supporting AI development and integrating AI into cyber and influence operations.

Fully automated, complex cyber operations are beyond the capabilities of current frontier models, though researchers have made some advancements in this space. As commercial and open-source models become more capable of complex tasks, more actors will likely apply these to cyber operations.

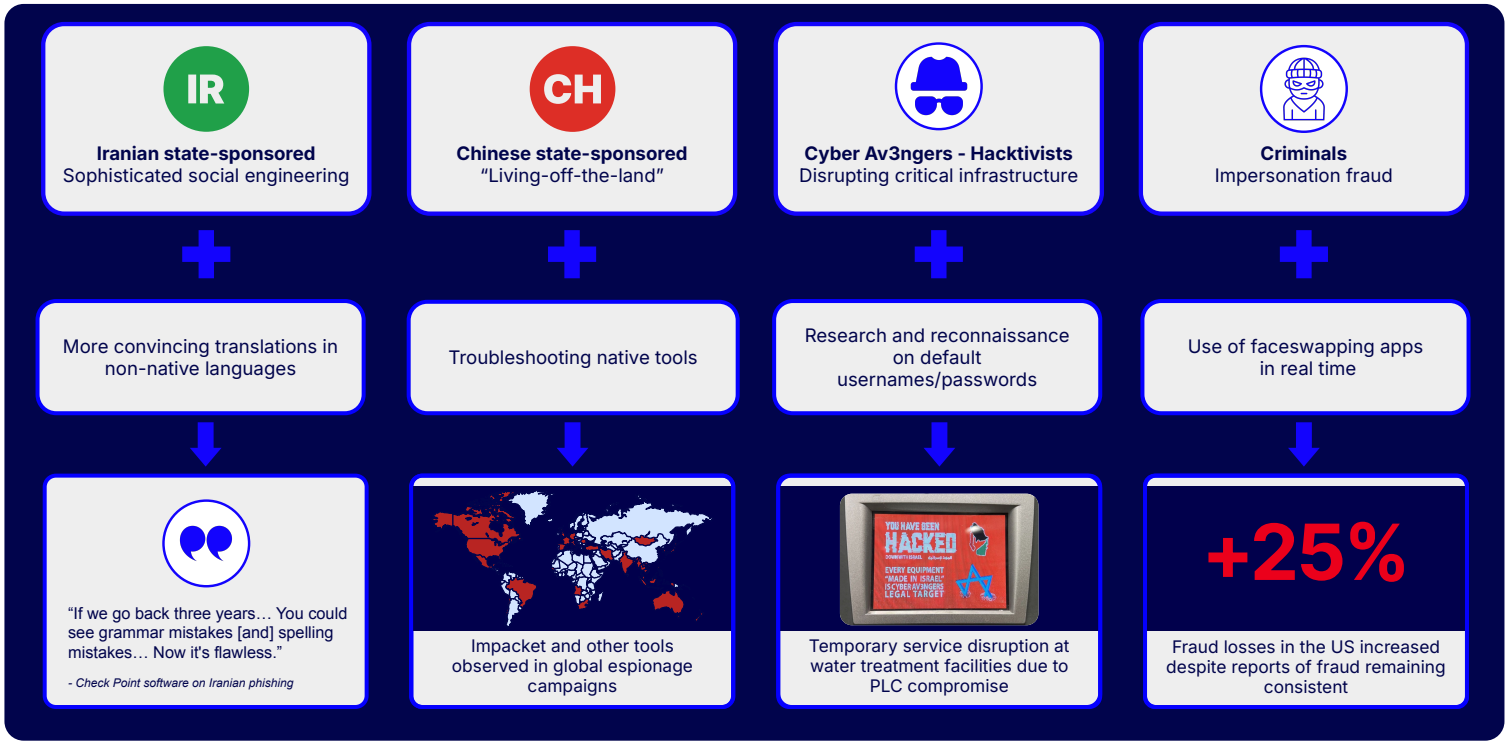


Figure 1: *How Generative AI amplifies observed tactics, techniques, and procedures for various threat actors (Source: Recorded Future)*

Analysis

Generative AI makes offensive operations more effective and scalable. Social engineering operations in particular benefit from AI capabilities, which enable more convincing voice, video, and text communications to deceive targets. Threat actors are also using AI to amplify research efforts, code debugging, scripting support, and other discrete tasks in cyber threat operations.

When incorporated strategically into operations, AI tools make malicious cyber operations more effective. Malicious actors are likely to integrate future AI capabilities as technology advances: most governments have adopted long-term AI strategies, while criminals opportunistically embrace the most effective tools.

AI makes new threats possible.

Weaponization of data: Russian threat actors are reportedly using AI to parse the contents of compromised inboxes to develop more convincing phishing lures, while an unidentified threat actor used Claude to automate data processing of stolen credentials to gain access to IoT security cameras.

AI-enhanced malware: A Russian-linked threat actor group deployed a Python-based infostealer dubbed "LameHug" that integrates an LLM to dynamically generate commands during run time. The LLM enabled the malware to evade defenses and adapt to the target environment in real time.

Search engine poisoning: The Russia-linked news network "Pravda" is systematically injecting AI-generated chatbots with narratives favorable to Russia by manipulating trusted sources like Wikipedia or maintaining propaganda sites posing as news outlets. These techniques are intended to dupe web crawlers and automated search engines into amplifying disinformation content.

Amplifying AI generated content: Recorded Future has observed China-linked influence operators amplifying AI-generated content, such as deepfake videos of politicians in the Philippines and Canada who are perceived to be critical of China.

State-sponsored threat actors are adapting commercial generative AI for custom tools. In at least two cases, Chinese actors have used commercial artificial intelligence for use in surveillance and cognitive warfare. OpenAI reported that actors operating out of China used its platform to support the development of an AI social media surveillance tool. In April 2024, a Chinese defense research agency applied for a patent using OpenAI's Sora in a "powerful and flexible" cognitive warfare system. The patent application argues that the system "innovatively applies" Sora and other models to overcome previous challenges in cognitive warfare and more effectively deliver content (Source: Recorded Future Geopolitical Intelligence Summary)..

Tools claim to offer custom capabilities, though prompt engineering remains effective for crime. While "criminal LLMs" like WormGPT or FraudGPT remain persistently available on criminal marketplaces, these are generally "jailbroken" or specially trained versions of open-source and commercial tools. Prompt engineering remains a cheap and effective way to bypass safety guardrails by tricking the LLM into providing a malicious answer. Multiple "how-to" guides are available to assist criminals in getting started in this endeavor, though interest in doing so appears to have decreased over time, likely as familiarity with the technology has increased.

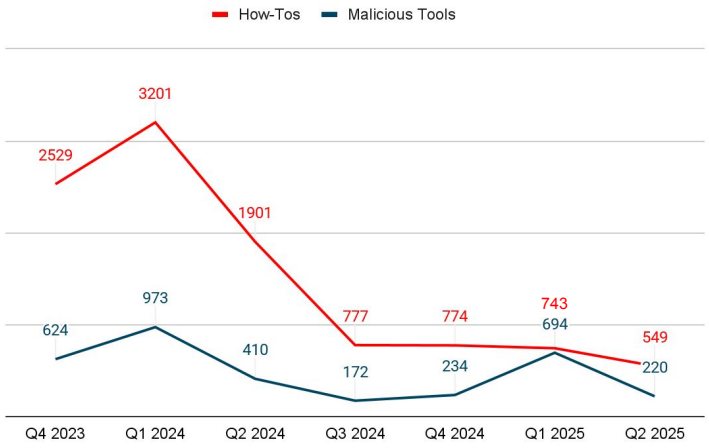


Figure 2: *While criminals were initially interested in guides and tools for using LLMs for malicious purposes, references have decreased significantly on criminal forums (Source: Recorded Future)*

Current open-source and commercial models can provide some level of automation, though fully autonomous malware remains theoretical. While threat actors have integrated LLMs into malware to support discovery and defense evasion, a baseline level of technical sophistication is still needed to deploy it effectively. Current AI capabilities are not yet able to support complex tasks requiring reasoning and novel analysis, such as carrying out a long-term data exfiltration campaign against a sensitive target.

One potential avenue to more advanced capabilities is agentic AI. AI developers are touting agentic AI's capabilities to enhance fields like software development and content generation — two areas that can easily be translated into automating malware development and influence operations, respectively. Defenders should assume that any legitimate use case will be followed by a parallel malicious use case.

Outlook

AI will almost certainly drive strategic cyber and influence operations: China, Russia, and Iran (among others) are actively exploring AI as a core component of next-generation military and intelligence strategies. While the long-term goal of fully autonomous, AI-driven warfare remains aspirational, near-term developments are already reshaping cyber operations and information warfare. Expect a measurable uptick in AI-enhanced offensive campaigns targeting critical infrastructure, as well as more scalable, personalized influence operations that leverage machine learning and generative AI.

Opportunistic actors will likely continue to evade guardrails to exploit increasingly advanced open-source and commercial tools: As models become increasingly capable of complex operations, threat actors will continue to use prompt engineering and other strategies to convince these models to automate malicious actions. This will likely increase the volume of sophisticated cybercriminal operations as lower-skilled threat actors seek out tools or expertise to exploit AI capabilities.

AI infrastructure will likely become a high-value target: As enterprises integrate AI systems into operational and decision-making workflows, these tools will attract adversaries seeking high-leverage points of compromise. Attacks may include model manipulation via data poisoning, exploitation through prompt injection, and broader software supply-chain vulnerabilities. Implementing AI governance and AI-informed security controls can help mitigate these risks.

The arms race for AI superiority will almost certainly intensify competition for acquiring technology and related resources: AI supremacy is a strategic goal for a variety of current and emerging powers. Governments are currently investing in both civilian and military applications of AI, anticipating that early adopters will gain a significant economic and strategic edge. Expect increased competition for access to advanced chips, large-scale training data, and frontier models.

Mitigations

Enhance Employee Training: Educate users on recognizing AI-generated phishing attempts and deepfakes.

Implement Advanced Detection Tools: Deploy AI-driven security solutions capable of identifying and mitigating AI-enhanced threats.
Use Recorded Future advanced query builder to identify malware using AI.

Regularly Update Incident Response Plans: Incorporate scenarios involving AI-generated content and rapid attack development cycles.

Monitor Regulatory Developments: Stay informed on evolving compliance requirements related to AI usage and data protection.

Collaborate With Industry Peers: Share threat intelligence and best practices to collectively enhance resilience against AI-driven cyber threats.
Join the Recorded Future Community for expert threat updates and shared resources.

Further Reading

SOURCE	TITLE
Insikt Group	Artificial Eyes: Generative AI in China's Military Intelligence
Insikt Group	Measuring the US-China AI Gap
Insikt Group	Iran's AI Ambitions: Balancing Economic Isolation with National Security Imperatives
Insikt Group	Russia-Linked CopyCop Uses LLMs to Weaponize Influence Content at Scale

Risk Implications

Scenario: A leading AI company releases its latest product, capable of completing complex projects from start to finish with minimal supervision. Despite built-in guardrails, threat actors immediately begin to exploit these capabilities.

Opportunistic Threats

Threat
Opportunistic, low-sophistication threat actors accelerate initial access attempts

Effects
→ Jailbroken version of tool available on the dark web
→ Spearphishing at scale
→ Mass scanning and automated exploit attempts for unpatched vulnerabilities
→ Automated DDoS attacks

Risks
Financial fraud
Operational disruption

Who is most at risk: Individuals/personal data; Organizations with limited cyber resources

Likelihood of attack: High

Impact: Largely nuisance-level

Sophisticated Criminal Threats

Threat
Sophisticated ransomware group carries out "spree-style" campaign by exploiting an zero-day vulnerability in widely used technology

Effects
→ Threat actors are capable of targeting thousands of vulnerable organizations simultaneously with customized attack pathways
→ Information sharing and attribution are difficult due to lack of shared TTPs/IoCs
→ Volume of stolen data outpaces threat actors' ability to exploit it, resulting in data dumps

Risks
Operational disruption
Financial fraud
Legal and compliance failure

Who is most at risk: Critical infrastructure organizations storing regulated data (such as health data and PII)

Likelihood of attack: Moderate

Impact: High

State-Sponsored Threats

Threat
Unknown actors have manipulated decision-making weights on an AI system used across the US government to automate service provision

Effects
→ Government services are massively disrupted, causing widespread political backlash
→ Malign influence operations amplify outrage and distrust against AI, government
→ International hostilities increase as US attempts to determine attribution

Risks
Operational disruption
Financial fraud
Legal and compliance failure

Who is most at risk: Government and critical infrastructure private sector entities depending on AI systems to function

Likelihood: High

Impact: High