# New AI-driven social media ads deploy malware to millions in evolving cybercrime wave



A new wave of cybercrime is emerging, propelled by the growing interest in artificial intelligence and enticing digital tools. A malicious group, identified as UNC6032 and allegedly linked to actors in Vietnam, has been exploiting this trend by deploying fake advertisements on social media platforms. These ads promote non-existent AI video generators, luring users to fraudulent websites that impersonate legitimate services such as Luma AI, Canva Dream Lab, and Kling AI. The findings, uncovered by Mandiant, a division of Google that focuses on threat intelligence and security, indicate that malicious ads have reached over two million users since late 2024.

These deceptive ads promise users the allure of AI-driven text- and image-to-video generation. However, clicking these advertisements leads to a perilous outcome. Users who interact with these sites are subjected to a phony video-generation interface, which, upon clicking the "Start Free Now" button, triggers the download of a ZIP file embedded with malware. This malware, once executed, backdoors the user’s device, allowing the attackers to log keystrokes and access password managers, digital wallets, and other sensitive information. Mandiant's investigation also highlights the convoluted methods used by UNC6032 to avoid detection, including the frequent rotation of domains associated with these fake ads, which were posted on compromised accounts as well as on pages created by the attackers.

The scale of this operation underscores significant vulnerabilities inherent in the digital landscape. Mandiant clarified that while the ads garnered wide visibility, this does not necessarily equate to the number of actual victims. Current estimates suggest that the ads on LinkedIn alone generated between 50,000 to 250,000 impressions, with a substantial portion originating from the United States. Despite these alarming figures, Meta, the parent company of Facebook, has taken proactive measures, including actively removing malicious advertisements and blocking related URLs. A spokesperson for Meta remarked on their collaboration with Google to enhance their defences against evolving cyber threats, acknowledging the continuous ingenuity of cybercriminals who frequently adapt their tactics.

Mandiant's exploration into this incident also serves as a stark reminder of the growing role of generative AI in future cybercrime schemes. Researchers predict that the increasing accessibility of AI tools will empower a new breed of information operations and sophisticated social engineering attacks. This phenomenon, driven by the capabilities of AI to generate convincing images, videos, and text, presents formidable challenges for users attempting to discern legitimate content from malicious designs. Analysts estimate that the use of generative AI in phishing and disinformation campaigns could lead to an escalation of targeted attacks, previously achievable only through extensive social engineering.

Among the malware deployed in this campaign is STARKVEIL, a dropper capable of delivering various modular malware families, specifically designed for information theft. For example, one malware type, named GRIMPULL, possesses advanced evasion techniques against virtual machine detection and anti-virus software, employing Tor for command and control connections. Simultaneously, XWORM, a more sinister tool, can execute commands remotely, log keystrokes, and even spread via USB drives. Another named FROSTRIFT aims to establish long-lasting access on the affected machines, seeking out browser extensions tied to password management to further exploit compromised accounts.

Mandiant emphasizes the broader implications of such attacks, noting that the rise of carefully crafted fake “AI websites” has broadened the pool of potential victims beyond digital professionals to everyday users. The trio of analysts, Diana Ion, Rommel Joven, and Yash Gupta, express concern over the significant threat posed by these strategies—digital predators are no longer solely targeting niche markets but rather anyone captivated by the promise of innovative technology. As the digital realm continues to evolve, the potential for misuse becomes ever more concerning, urging users to remain vigilant as these threats proliferate.

## Reference Map:

* Paragraph 1 – [[1]](https://www.theregister.com/2025/05/27/fake_social_media_ads_ai_tool/), [[2]](https://www.itpro.com/technology/artificial-intelligence/mandiant-says-generative-ai-will-empower-new-breed-of-information-operations-social-engineering)
* Paragraph 2 – [[1]](https://www.theregister.com/2025/05/27/fake_social_media_ads_ai_tool/), [[4]](https://www.techrepublic.com/article/mandiant-malware-proliferating/)
* Paragraph 3 – [[3]](https://cloud.google.com/blog/topics/threat-intelligence/threat-actors-generative-ai-limited), [[5]](https://www.securityweek.com/mandiant-offers-clues-to-spotting-and-stopping-north-korean-fake-it-workers/)
* Paragraph 4 – [[1]](https://www.theregister.com/2025/05/27/fake_social_media_ads_ai_tool/), [[6]](https://www.techtarget.com/searchsecurity/news/366617826/Mandiant-links-Ivanti-zero-day-exploitation-to-Chinese-hackers)
* Paragraph 5 – [[1]](https://www.theregister.com/2025/05/27/fake_social_media_ads_ai_tool/), [[7]](https://www.tanium.com/blog/cti-roundup-north-korean-remote-workers-spam-deepfakes/)

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.theregister.com/2025/05/27/fake_social_media_ads_ai_tool/> - Please view link - unable to able to access data
2. <https://www.itpro.com/technology/artificial-intelligence/mandiant-says-generative-ai-will-empower-new-breed-of-information-operations-social-engineering> - Mandiant researchers have warned that generative AI will enable a new wave of personalized social engineering and disinformation campaigns. They anticipate that AI-generated images, videos, text, and audio will be used to create highly convincing phishing attacks and deepfakes, making it more challenging for users to identify malicious content. The researchers also noted that AI could be used to generate large volumes of tailored 'lure' content based on analysis of a victim's social media presence, potentially leading to more effective phishing attacks.
3. <https://cloud.google.com/blog/topics/threat-intelligence/threat-actors-generative-ai-limited> - Mandiant anticipates that information operations actors' adoption of generative AI will vary by media form – text, images, audio, and video – due to factors including the availability and capabilities of publicly available tools and the effectiveness of each media form to invoke an emotional response. They believe that AI-generated images and videos are most likely to be employed in the near term; and while they have not yet observed operations using large language models (LLMs), they anticipate that their potential applications could lead to their rapid adoption.
4. <https://www.techrepublic.com/article/mandiant-malware-proliferating/> - Mandiant has observed a proliferation of malware, with cybercriminals employing common methods to great effect. Notably, groups like Lapsus (UNC3661) and UNC3944 have targeted major corporations using data from underground cybercrime markets and sophisticated social engineering schemes. These incidents underscore the threat posed by persistent adversaries willing to eschew traditional rules of engagement, highlighting the need for organizations to adjust their protections and expectations accordingly.
5. <https://www.securityweek.com/mandiant-offers-clues-to-spotting-and-stopping-north-korean-fake-it-workers/> - An American collaborator assisting fake North Korean IT workers to secure jobs at US companies generated approximately $7 million in revenue over three years, underscoring the profitability of a growing threat with serious nuclear weapons implications. According to fresh documentation from Google’s Mandiant unit, the revenue generated by the fake IT worker scheme can be substantial with a single American facilitator helping to compromise over 60 identities, impacting 300 companies, and generating $6.8 million in illicit revenue between 2020 and 2023.
6. <https://www.techtarget.com/searchsecurity/news/366617826/Mandiant-links-Ivanti-zero-day-exploitation-to-Chinese-hackers> - Mandiant connected the recent zero-day attack against Ivanti Connect Secure VPN appliances to UNC5337, the same China-nexus threat actor that was tied to the exploitation of two Ivanti zero-day flaws one year ago. In a blog post published on Wednesday, Mandiant detailed an attack campaign involving a zero-day vulnerability, tracked as CVE-2025-0282, discovered in Ivanti Connect Secure (ICS), Ivanti Policy and ZTA Gateways. Ivanti disclosed the flaw on Wednesday and warned users that it was being exploited in the wild. Patches are available, and users are urged to apply fixes as Ivanti products have proved to be a popular target for attackers.
7. <https://www.tanium.com/blog/cti-roundup-north-korean-remote-workers-spam-deepfakes/> - In this week’s roundup, CTI investigates a network of global IT workers operating on behalf of the North Korean government. Next up, CTI explores how threat actors abuse legitimate website features to deliver spam. Finally, CTI wraps up with a look at Recorded Future’s analysis of 82 deepfakes identified between July 2023 and July 2024. Mandiant is actively tracking a network of global IT workers who are operating on behalf of the North Korean government. These workers seek employment with organizations across various industries. Their main objectives are to make money, maintain long-term access for potential future exploitation, and potentially conduct espionage. Mandiant tracks the group as UNC5267.