# State-sponsored cyber-criminals reportedly utilising Google's AI model for malicious operations



State-sponsored cyber-criminals are reportedly utilising Google’s generative AI model, Gemini, to aid in their cyber operations, although they have yet to demonstrate any significant advancements in their capabilities, as indicated in a recent report by the Google Threat Intelligence Group (GTIG). The report, titled "Adversarial Misuse of Generative AI," details how threat actors linked to nations such as Iran, China, North Korea, and Russia are employing Gemini for various malicious purposes.

The findings reveal that these actors are engaging in activities traditionally associated with Advanced Persistent Threats (APT), including government-backed hacking, cyber-espionage, and destructive network attacks. Additionally, their practices fall under Information Operations (IO), which aim to manipulate and influence online audiences through deceptive tactics, including the use of sockpuppet accounts and comment brigading.

Despite the evident risks, the GTIG noted that the current applications of Gemini by these state-sponsored groups are somewhat limited. They are primarily leveraging the AI tool for research, code troubleshooting, and localising content. However, the report underscores a possible alarming trajectory, revealing that APT actors are looking into vulnerabilities of their targets, developing weaponised payloads, and creating malicious scripts using the tool.

According to the GTIG, Iranian-affiliated groups are among the heaviest users of Gemini, with over ten factions engaging in activities such as developing phishing campaigns and spying on defence experts and organisations. Remarkably, Iranian-linked cyber criminals were responsible for the majority of Information Operations observed, comprising three-quarters of all IO activity. They have utilised Gemini for content generation, including persona creation, messaging development, and translation, while also finding ways to amplify their reach.

Conversely, Chinese APT actors focus on researching ways to enhance their cyber capabilities, investigating lateral movement, privilege escalation, data exfiltration, and evasion of detection mechanisms. Russian threat groups have adopted Gemini for coding improvements, such as converting malware to other programming languages and adding encryption capabilities. North Korean actors have directed their use of Gemini towards research into topics of strategic significance to their government, notably the South Korean military and cryptocurrency. Interestingly, they have also employed the tool for crafting cover letters and job research, which aligns with efforts to deploy 'fake IT workers' in Western companies.

The report highlights that these threat actors have not displayed attempts to creatively utilise prompt attacks. Their methodologies remain rudimentary, characterised by simple actions such as rephrasing or repeating prompts. The GTIG pointed out this type of ‘low-effort’ experimentation—including copying and pasting instructions to develop ransomware—has not successfully circumvented Gemini’s safety controls.

Despite the current limitations noted in their use of the generative AI, the GTIG anticipates that as the AI landscape evolves, newer models and more capable systems may offer adversaries a significant advantage. The report concludes with a commitment from Google to leverage threat intelligence to disrupt malicious operations and to investigate abuses of their products and services. Additionally, they have emphasised the ongoing need for security standards as innovations in AI progress. To this end, Google has introduced the Secure AI Framework (SAIF), a conceptual framework aimed at securing AI systems against misuse.

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

## References

* <https://cloud.google.com/blog/topics/threat-intelligence/adversarial-misuse-generative-ai> - This URL corroborates the report by the Google Threat Intelligence Group (GTIG) on the adversarial misuse of generative AI, specifically detailing how state-sponsored threat actors are using Gemini for malicious purposes.
* <https://cloud.google.com/blog/topics/threat-intelligence/adversarial-misuse-generative-ai> - This URL supports the claim that APT actors from countries like Iran and China are using Gemini for activities such as researching vulnerabilities and developing malicious scripts.
* <https://www.google.com/search?q=Advanced+Persistent+Threats+(APT)> - This search result provides background information on Advanced Persistent Threats (APT), which are mentioned as part of the malicious activities supported by Gemini.
* <https://en.wikipedia.org/wiki/Information_operations> - This URL explains Information Operations (IO), which are tactics used by threat actors to manipulate online audiences, as mentioned in the article.
* <https://www.google.com/search?q=sockpuppet+accounts+and+comment+brigading> - This search result provides information on sockpuppet accounts and comment brigading, tactics used in Information Operations.
* <https://cloud.google.com/blog/topics/threat-intelligence/adversarial-misuse-generative-ai> - This URL highlights the limited current applications of Gemini by state-sponsored groups, primarily for research and code troubleshooting.
* <https://www.google.com/search?q=Iranian+cyber+operations> - This search result provides context on Iranian cyber operations, which are noted as significant users of Gemini for activities like phishing and espionage.
* <https://www.google.com/search?q=Chinese+APT+actors+cyber+capabilities> - This search result offers background information on Chinese APT actors and their focus on enhancing cyber capabilities using Gemini.
* <https://www.google.com/search?q=North+Korean+cyber+operations> - This search result provides context on North Korean cyber operations, including their use of Gemini for strategic research and deploying 'fake IT workers'.