
State Hackers Weaponize Gemini: APTs Use Google's AI for Full-Spectrum Cyber Attacks

APTs Use Google's AI for Full-Spectrum Cyber Attacks

A new era of AI-enabled warfare has arrived, with state-backed threat actors integrating Google's Gemini large language model into every phase of their cyber operations. From initial reconnaissance and phishing lure generation to command-and-control development and data exfiltration, advanced persistent threat groups are systematically weaponizing accessible AI tools. This marks a significant escalation where artificial intelligence is no longer just a target but a core component of the attacker's toolkit, amplifying the speed, scale, and sophistication of nation-state campaigns.

Threat Actor Roster: Who Is Using Gemini?

Google has identified multiple APT groups leveraging Gemini across their operations, representing a cross-section of global cyber threats:

Group 1:

APT31, Temp.HEX

Affiliation: China

Observed Activity: Vulnerability analysis, RCE testing, WAF bypass techniques, SQL injection testing against US targets

Group 2:

APT42

Affiliation: Iran

Observed Activity: Social engineering campaigns, debugging, code generation, exploitation research

Group 3:

UNC2970

Affiliation: N. Korea

Observed Activity: Target profiling, OSINT gathering, phishing lure creation

Group 4:

Unnamed Russian Actors

Affiliation: Russia

Observed Activity: Translation, coding assistance, troubleshooting

Attack Chain

Reconnaissance & Target Profiling

- Using Gemini for open-source intelligence gathering to profile targets.
- Automating vulnerability analysis and generating targeted testing plans.
- Fabricating scenarios to probe for weaknesses in specific organizations.

Phishing & Social Engineering

- Generating convincing phishing lures tailored to target personas.
- Translating content to support cross-border campaigns.
- APT42 specifically leveraged Gemini to speed up creation of tailored malicious tools.

Development & Weaponization

- Debugging and generating code for malware development.
- Implementing new capabilities into existing malware families, including: The CoinBait phishing kit and the HonestCue malware downloader and launcher.
- Researching exploitation techniques and testing payloads.

Command & Control and Exfiltration

- Assisting in C2 infrastructure development.
- Supporting data exfiltration planning and execution.

Model Extraction Attacks

Organizations with authorized API access systematically query Gemini to reproduce its decision-making processes and replicate its functionality.

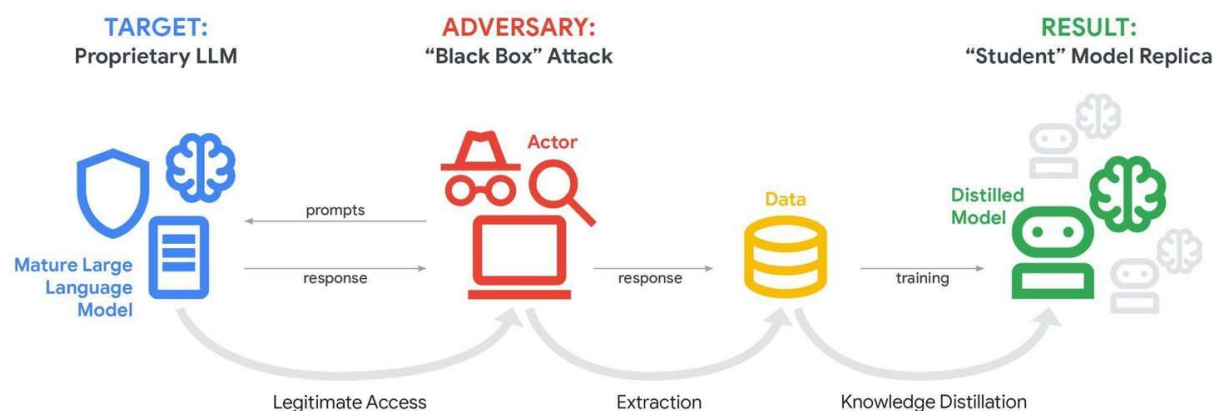


Figure 1 Illustration of model extraction attacks — Source: Google Threat Intelligence Group (GTIG)

Knowledge Distillation

Attackers use machine learning techniques to transfer information from advanced models to train new, unauthorized models quickly and at significantly lower cost.

Scale

One large-scale attack involved over 100,000 prompts in non-English languages aimed at replicating Gemini's reasoning across diverse tasks.

AI-Generated Malware

- Logging messages in malware source code prefixed with "Analytics:" could help defenders track exfiltration processes.
- Evidence suggests the Lovable AI platform was used to create certain malware samples, based on references to Lovable Supabase client and lovable [.] app domains.

AI-Powered Social Engineering

Cybercriminals are using generative AI services in ClickFix campaigns targeting macOS users:

- Malicious ads appear in search results for troubleshooting queries.
- Users are lured to execute commands that deliver the AMOS infostealer.
- This demonstrates AI's role in creating convincing, targeted lures that bypass traditional suspicion.

Defense Strategy

For Organizations:

- Assume adversaries have AI assistance at every stage of the attack lifecycle.
- Enhance phishing resistance training with awareness of AI-generated, highly convincing lures.
- Deploy behavioral analytics to detect anomalies in network traffic and user activity that may indicate AI-coordinated attacks.

For AI Providers:

- Implement robust detection for model extraction and distillation attempts.

- Develop and enforce policies against malicious use, while recognizing that determined adversaries will find ways to abuse legitimate access.
- Share threat intelligence on observed abuse patterns with the security community.

For Defenders:

- Monitor for indicators of AI-generated malware, such as distinctive logging patterns.
- Track adversary use of AI platforms as part of threat intelligence.

The weaponization of Gemini by state-sponsored actors represents a watershed moment. AI is no longer just a defensive tool or an occasional attack vector; it has become a standard component of the adversary's operational playbook.

Ready to see how AICenturion can secure you against AI risks?

Request a demo today: hello@cytex.io

Connect with our social media channels

