
Claude LLM Weaponized ClickFix Attack Chain

ClickFix Campaign: Hackers abusing Claude artifacts

Threat actors have found a new and trusted channel to deliver malware to unsuspecting macOS users: legitimate AI-generated content. In a sophisticated ClickFix campaign, cybercriminals are abusing Claude artifacts, publicly shared content hosted on Anthropic's official domain, to push the MacSync infostealer. Combined with malicious Google Ads and impersonated support pages, this attack chain exploits user trust in AI platforms and search results to execute devastating data theft with a single Terminal command.

Attack Chain:

Step 1: Malicious Search Results

Victims searching for specific technical queries, such as "online DNS resolver," "macOS CLI disk space analyzer," or "HomeBrew," encounter promoted malicious links at the top of Google Search results. These ads lead to one of two destinations, both controlled by the attacker.

Step 2: The Lure

Option A - Claude Artifact Page:

The user lands on a page hosted on the legitimate claude [.] ai domain. This page, created by the attacker, contains instructions masquerading as a helpful guide. Despite a disclaimer that the content is user-generated and unverified, the trusted domain lends an air of authenticity.

Option B – Impersonated Medium Article:

The victim is directed to a Medium article cleverly designed to look like an official Apple Support page.

Step 3: The Fatal Instruction

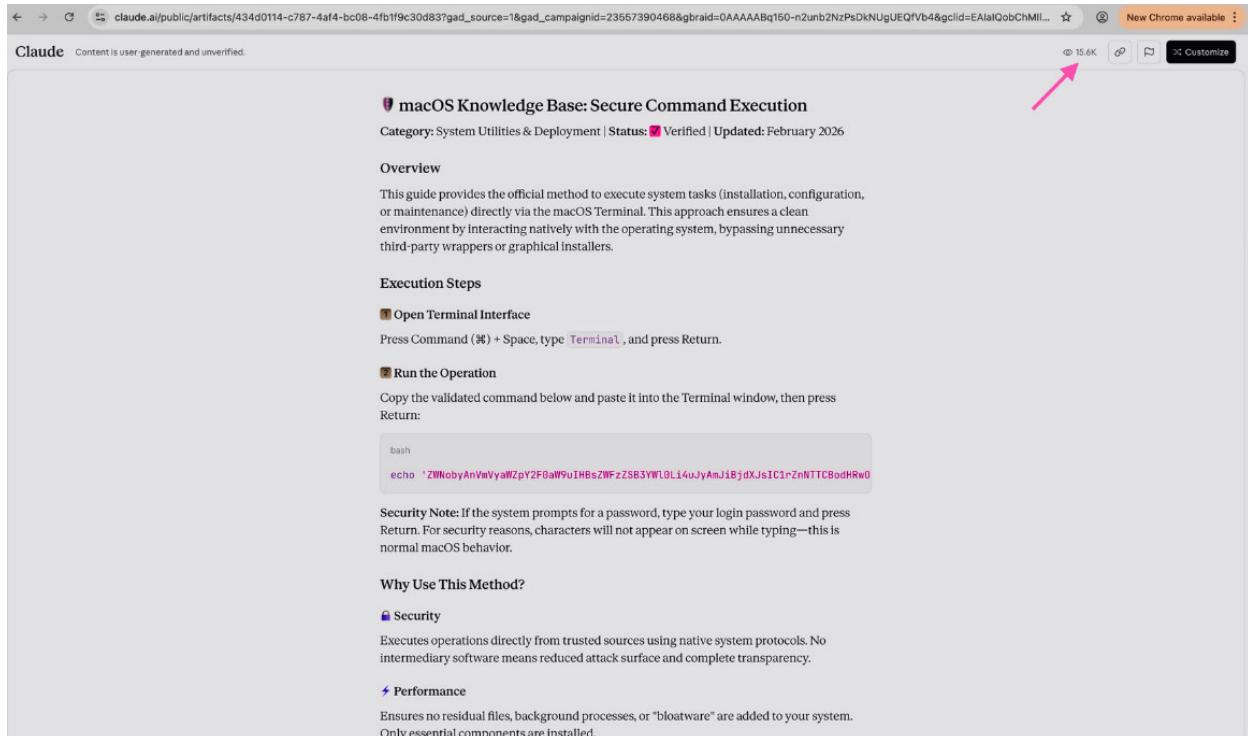
Both pages contain the same core payload: a shell command that the user is instructed to copy and paste into their Mac's Terminal application. The language is technical but seemingly benign to a non-expert seeking a quick solution.

Step 4: Multi-Stage Infection

- A malware loader is fetched and executed.
- The loader downloads the MacSync infostealer.
- The malware establishes communication with its command-and-control (C2) server at a2abotnet [.] com/gate using a hardcoded token and API key. It spoofs a macOS browser user-agent to blend in with normal traffic.

AppleScript is then used to systematically steal sensitive data, including:

- Keychain contents.
- Browser data (passwords, history, cookies).
- Cryptocurrency wallet files.
- Stolen information is packaged into an archive at /tmp/osalogging [.] zip.
- The archive is exfiltrated to the attacker's C2 via an HTTP POST request. If the upload fails, the archive is split into smaller chunks and retried up to eight times.
- After successful exfiltration, a cleanup script deletes all traces of the infection from the system.



Reach & Impact

- Over 15,600 views on the malicious Claude artifact page alone, indicating a potentially large number of victims who accessed the dangerous instructions.
- More than 10,000 users are estimated to have engaged with the malicious content across both variants.
- Both attack variants connect to the same C2 infrastructure, confirming a single, coordinated threat actor behind the operation.

Evolution of LLM Abuse

This attack represents a troubling evolution in the abuse of large language models. While previous concerns focused on AI generating malicious code, this campaign weaponizes the trusted distribution platform itself. A page on Anthropic's official domain, even with a disclaimer,

carries an inherent legitimacy that bypasses user suspicion. It demonstrates that AI platforms are now being used not just to create attacks, but to host and deliver them.

Indicators of Compromise (IOCs):

-  `claude[.]ai/public/artifacts/434d0114-c787-4af4-bc08-4fb1f9c30d83`
-  `apple-mac-disk-space.medium[.]com`
-  `a2abotnet[.]com`
-  `raxelpak[.]com`

 IPs: 172[.]67.187.216, 104[.]21.56.197, 13[.]248.169.48, 76[.]223.54.146

 Hashes: 64068d0b7fbef87a7af91834ead9bc0efa21f814b9e6a945b440db75bbcfed76
6292f64c81dbc57d5135c5773547cc6d79afa15efe4c90cfaf27e087c7aba701
c0676ba7726e6b4b836c2a07aacb92e41efd9eea7cbc31bbf1a7f9f9556dd4cb

 Staging: /tmp/osalogging.zip (MacSync stealer indicator)

 Compromised advertisers: T S Q SA (Colombia), Earth Rangers Foundation (Canada)

Defense Tactics:

This attack relies on social engineering and user action. Defending against it requires skepticism and verification.

- Never Execute Unexplained Terminal Commands. Do not copy and paste commands from websites, ads, or articles into Terminal unless you fully understand every part of what they do. A legitimate solution will never require this.
- Verify the Source. A page on Claude [.] ai or Medium is not an official Apple support channel. Apple does not provide troubleshooting via third-party AI artifacts.
- Leverage AI for Safety. If you are unsure about a command you see online, ask the same or a different AI chatbot: "Is this terminal command safe to run?" and provide the command. As researchers noted, this is a straightforward way to get an independent safety assessment.
- Enable Security Software. Use reputable endpoint protection on macOS that can detect and block known infostealer behavior.
- Educate users, especially those on macOS, about this specific attack vector.
- Consider implementing application allowlisting or endpoint detection and response (EDR) solutions that can monitor for suspicious 'osascript' activity and connections to known malicious domains.

The lines between trusted AI platforms, search advertising, and malware delivery have officially blurred.

