Zero-click RCE vulnerability hit four major AI coding agents, two remain unpatched

SECONDARY

Publisher	Help Net Security
Published	2026-09-18
Retrieved	2026-09-21
Cite key	helpnetsecu2026zeroclick
URL	https://www.helpnetsecurity.com/2026/09/18/plugin4shell-ai-coding-agents-vulnerability/

Assessment. Trade coverage one day after the research post. Used to corroborate patch status and the framing of which two agents remain exposed. Coverage date is 18 September; the event date is the researchers' 17 September publication.

Help Net Security, 18 September 2026.

GitHub Copilot and Gemini CLI remain without patches. Microsoft has not shipped a fix for Copilot; Google deprecated Gemini CLI rather than patching it.

Patched: Claude Code 2.1.179 (Anthropic); Codex 0.146.0 (OpenAI).

No CVE identifiers are given.

AIR researchers found the bug in May 2026 and disclosed it to vendors the following month, with working proof-of-concept exploits against all four agents.

The flaw bypasses SHA pinning — the mechanism that locks plugins to specific reviewed code versions. As the researchers put it: "every one of the four agents checks out the pinned commit without verifying the checkout landed there, letting an attacker swap in malicious code while the pin still looks intact."

It is zero-click because background auto-updates re-run the vulnerable checkout without user action on Claude Code and Codex.

Attackers can either publish functional plugins that become malicious after review, or hijack existing trusted repositories to distribute compromised code to all installed instances.