

Plugin4Shell — Zero Click RCE Vulnerability found in top 4 most popular coding agents

PRIMARY

Publisher	AIR Security
Published	2026-09-17
Retrieved	2026-09-21
Cite key	airsecurity2026pluginshe
URL	https://www.air.security/blog-posts/plugin4shell

Assessment. The researchers' own write-up and the only source that states the git mechanism precisely. Primary for the vulnerability, the disclosure timeline and the recommended rev-parse check. Vendor patch status here is the researchers' verification, not a vendor statement.

Plugin4Shell

AIR Security, 17 September 2026. Or Nevo, Dor Granat, Niv Hoffman.

The mechanism

Coding agents install marketplace plugins by cloning the plugin repository and checking out a pinned commit SHA. The pin is the supply-chain control: the marketplace reviews a commit, records its hash, and the agent is supposed to get exactly that code.

"when a name is both a valid ref and an object id, git prefers the ref and only prints a `refname is ambiguous` warning - so it never matters whether the pinned commit is still present, the branch is what gets checked out."

An attacker who controls the plugin repository creates a branch whose name is the 40-hex pinned SHA. The sequence the agents run —

```
git clone <repo> git checkout aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
```

— resolves to the attacker's branch rather than the pinned object.

"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."

Why it is zero-click

Marketplaces bump their pinned SHA over time. Agents re-run the checkout in the background when that happens, so the malicious tree arrives without any user action on Claude Code and Codex.

Two attack vectors

Publish a functional plugin and turn it malicious after review; or hijack an existing trusted repository and reach every installed instance.

Which hosts allow the branch name

GitHub rejects a 40-hex branch name outright. Some hosts forbid it, "others, Bitbucket among them, and any self-hosted git server, allow it." Marketplaces on those hosts are a supported configuration — Anthropic's own documentation lists Bitbucket and self-hosted git as valid marketplace backends.

The fix

Resolve the commit actually in the working tree and abort unless it equals the pinned SHA:

```
test "$(git rev-parse HEAD)" = "<pinned-sha>"
```

Status

Discovered May 2026 with a working proof of concept against all four agents. Coordinated disclosure to all vendors June 2026. Anthropic confirmed the Claude Code fix on 17 June 2026 (2.1.179). Codex 0.146.0 verified fixed 12 August 2026. Google confirmed on 4 August 2026 that Gemini CLI would not be fixed (deprecated). GitHub Copilot: no fix shipped.

Updating is the only complete mitigation where one exists. Gemini CLI users should migrate to Antigravity.