

INTELLIGENCE BRIEF

AI Gateways Join the Exploited List as LiteLLM and Puppet Flaws Reach CISA KEV

CISA listed LiteLLM and Puppet Enterprise flaws as exploited on November 18 and 17. Both tools hold fleet-wide secrets, so patch, rotate every key, restrict access, and review logs for the exposure window.

EDITION

Fall/Winter 2026/2027 Edition

TR-02 · Technology risk

PUBLISHED BY

Z7 Solutions LLC

Orlando, Florida

READ ONLINE

z7solutions.com

844-974-8669

01 At a Glance

STATUS

ACTIVE

CISA added BerriAI LiteLLM CVE-2026-40217 to its Known Exploited Vulnerabilities catalog on November 18, 2026, and Puppet Enterprise CVE-2026-115731 on November 17, 2026. [1]

HEADLINE NUMBERS

Dec 2

2026 federal due date for LiteLLM CVE-2026-40217 [1]

Dec 1

2026 federal due date for Puppet CVE-2026-115731 [1]

4

LiteLLM flaws listed in CISA KEV as of November 18 [1]

KEY FACTS

LEAD FLAW

BerriAI LiteLLM CVE-2026-40217, KEV added November 18, 2026, federal due date December 2 [1]

SECOND FLAW

Perforce Puppet Enterprise CVE-2026-115731, KEV added November 17, 2026, federal due date December 1 [1]

LITELLM FIX TARGET

Upgrade to 1.83.14 or later, which closes the full reported chain [5]

PUPPET FIXED RELEASES

Enterprise 2026.0.1, 2025.12.1, 2023.8.13; Core 9.1.3 and 8.22.2 [8]

RANSOMWARE USE

Unknown for both, per CISA [1]

02 What Happened

CISA added two control-plane products to its exploited-vulnerability catalog in two days: BerriAI LiteLLM on November 18 and Perforce Puppet Enterprise on November 17, 2026. [1] An AI gateway and a configuration server share one trait: whoever controls them reaches everything they manage.

- CISA describes the LiteLLM flaw as an unprotected alternate channel that lets remote attackers execute code through bytecode rewriting at the /guardrails/test_custom_code URI. CISA's catalog lists the due date as December 2, 2026. [1]
- The LiteLLM advisory describes a sandbox escape in a custom-code guardrail endpoint, with code running in the proxy process, which runs as root in the default Docker image. Exploitation requires a proxy-admin credential in default configurations. [3]
- Obsidian Security reported that an earlier role-escalation flaw (CVE-2026-47101) let a low-privilege internal user reach admin routes and then this code execution path. It reports that an attacker on the host can read provider keys and every prompt and response passing through the gateway. [5]
- CISA describes the Puppet flaw as deserialization in report-processing code that lets a client with any CA-signed agent certificate run code on the Puppet Server as the service user. If the CA automatically signs requests without strong verification, an unauthenticated network attacker can obtain such a certificate. [1]
- Perforce addressed the Puppet issue on November 3, 2026, and CISA cited reports of active exploitation. Public detail on how it is being used is limited. [8]
- Neither CISA entry names victims or an attacker. CISA lists known ransomware campaign use as unknown for both. [1]

03 How It Works

01 A control plane holds shared trust

LiteLLM is an open-source gateway that fronts more than 100 model providers and adds virtual keys, spend tracking and guardrails. [9] A single instance can hold master keys, provider keys, database strings and logged prompts. [6]

02 A privileged feature runs code

LiteLLM's guardrail test endpoint compiles and runs submitted code; a regex deny-list was reported as bypassable by manipulating bytecode. [5]

03 Access to that feature is the gate

The advisory requires a proxy-admin credential by default, while Obsidian describes earlier chains that let lower-privileged users reach admin routes. [3, 5]

04 Code runs inside the gateway

Execution happens in the proxy process, which runs as root in the default Docker image, so environment secrets such as the master and salt keys are within reach. [3, 5]

05 Puppet follows the same pattern

A crafted report submitted by an agent with a CA-signed certificate is deserialized by the server, and permissive auto-signing removes the authentication barrier. [1]

04 Who Is at Risk

Risk follows where these tools run and who can reach them, including instances IT never deployed.

- Organizations running self-hosted LiteLLM proxies in the affected range, especially with the admin interface or API reachable from untrusted networks. [2, 3]
- Teams whose gateway holds many provider keys and logs prompts, which may include source code, personal data and internal documents. [6]
- Puppet Enterprise, Puppet Core and OpenVox users on versions before the fixed releases, particularly where the CA auto-signs certificate requests. [1, 8]
- Z7 analysis: LiteLLM is open source and easy to stand up, so developer and data teams may run proxies outside IT's asset inventory. No source in this brief measures how often that happens.

Sectors in scope: Organizations running self-hosted AI gateways or LLM proxies, Teams using Puppet Enterprise, Puppet Core or OpenVox for configuration management, Federal civilian agencies bound by BOD 26-04 deadlines.

05 What Changed Since the Last Edition

EACH CHANGE WITH ITS SOURCE

01 November 17, 2026: CISA added Puppet Enterprise CVE-2026-115731 to the KEV catalog with a December 1 federal due date.

SOURCE 1

02 November 18, 2026: CISA added LiteLLM CVE-2026-40217 with a December 2 federal due date. NVD's record, modified November 19, carries the KEV fields.

SOURCE 2

03 This is not CISA's first LiteLLM listing: the catalog also holds CVE-2026-42208 (SQL injection, added May 8), CVE-2026-42271 (command injection, added June 8) and CVE-2026-59822 (MCP authentication flaw, added September 2).

SOURCE 1

04 In May, Sysdig reported probing of LiteLLM tables holding virtual keys and provider credentials about 36 hours after the SQL injection advisory, with no follow-up use of keys observed.

SOURCE 7

06 What to Do Now

IN PRIORITY ORDER, HIGHEST IMPACT FIRST

01 DO FIRST **Find every AI gateway and orchestration server, including ones IT did not deploy.**

Z7 recommends searching container registries, cloud workloads, code repositories and network listeners for LiteLLM images, packages and the /guardrails paths, and asking developer teams directly, because unmanaged instances cannot be patched.

02 **Patch LiteLLM to 1.83.14 or later and Puppet to a fixed release.**

NVD shows LiteLLM fixed from 1.83.10, while Obsidian and the Cloud Security Alliance say 1.83.14 or later closes the full reported chain. Puppet fixes are Enterprise 2026.0.1, 2025.12.1 and 2023.8.13, and Core 9.1.3 and 8.22.2. [2, 5, 6, 8]

03 **Rotate every upstream provider key, every virtual key, the master and salt keys, and database credentials.**

A code-execution compromise exposes configured provider keys and conversation content, and the Cloud Security Alliance advises rotating stored provider keys regardless of confirmed exploitation. [5, 6]

04 **Restrict gateways to internal networks behind authentication, and block the guardrail test endpoint at the reverse proxy until patched.**

Advisories recommend limiting admin ports to trusted hosts, keeping the master key away from regular consumers, and blocking POST /guardrails/test_custom_code. [3, 4, 6]

05 **Run the gateway as a non-root user and limit who holds proxy-admin.**

Execution as root in the default image raises impact, and Obsidian advises reviewing proxy_admin holders and any registered custom code guardrails. [3, 4, 5]

06 **Turn off Puppet CA auto-signing or require strong verification, and review signed certificates.**

CISA states that auto-signing without strong verification lets an unauthenticated attacker obtain a valid certificate. [1]

07 **Review logs for the exposure window and keep them for investigation.**

The Cloud Security Alliance advises monitoring proxy logs for unexpected subprocess execution and unusual API key usage, and CISA's required action references forensic triage. Z7 recommends starting at least from the first affected release or earliest exposure. [1, 6]

08 **Put AI infrastructure under the same patch SLA, owner and monitoring as other servers.**

Four LiteLLM flaws have reached CISA's exploited catalog since May 2026, which shows these tools face the same pressure as other internet-reachable servers. [1]

07 Indicators

Indicators the sources publish, as they print them. Hunt with them, and check the source before you block.

BEHAVIOR	Unexpected subprocess execution on the LiteLLM host or unusual API key usage. [6]

08 Sources

[1] **Known Exploited Vulnerabilities Catalog (kev-data feed), entries for CVE-2026-40217, CVE-2026-115731, CVE-2026-42208, CVE-2026-42271 and CVE-2026-59822** CISA · Catalog version 2026.11.18, released November 18, 2026. https://raw.githubusercontent.com/cisagov/kev-data/develop/known_exploited_vulnerabilities.json

[2] **CVE-2026-40217 detail (API record)** NIST National Vulnerability Database · Published April 10, 2026; modified November 19, 2026. https://nvd.nist.gov/vuln/detail/CVE-2026-40217

[3] **Sandbox escape in custom-code guardrail (GHSA-wxxx-gvqv-xp7p)** BerriAI LiteLLM GitHub security advisory · May 7, 2026. https://github.com/BerriAI/litellm/security/advisories/GHSA-wxxx-gvqv-xp7p

[4] **X41-2026-001: LiteLLM custom code guardrail** X41 D-Sec · April 8, 2026. https://www.x41-dsec.de/lab/advisories/x41-2026-001-litellm/

[5] **Breaking LiteLLM: From Low-Privilege User to Admin and RCE** Obsidian Security · As published by the source. https://obsidiansecurity.com/blog/litellm-privilege-escalation-rce

[6] **LiteLLM AI Gateway: KEV-Listed Attack Chain Enables Full Takeover** Cloud Security Alliance AI Safety Initiative · June 17, 2026. https://labs.cloudsecurityalliance.org/research/csa-research-note-litellm-ai-gateway-attack-chain-20260617-c/

[7] **U.S. CISA adds a flaw in BerriAI LiteLLM to its Known Exploited Vulnerabilities catalog** Security Affairs · May 11, 2026. https://securityaffairs.com/191964/security/u-s-cisa-adds-a-flaw-in-berriai-litellm-to-its-known-exploited-vulnerabilities-catalog.html

[8] **Hackers exploiting critical Puppet RCE** The Hacker News · November 18, 2026. https://thehackernews.com/2026/11/hackers-exploiting-critical-puppet-rce.html

[9] **BerriAI litellm repository** GitHub · Accessed November 20, 2026. https://github.com/BerriAI/litellm

NEXT STEP

What does this mean for your security?

Bring your questions about AI gateway and orchestration infrastructure, your environment and the controls you rely on. Talk through the most useful next step with Z7 Solutions LLC.

01 / Exposure

Bring one question about how this threat could affect your environment.

02 / Priorities

Bring your current safeguards, open gaps and budget questions.

ALSO IN THIS EDITION

Post-quantum deadlines set, attack estimates fall
More from the edition · z7solutions.com/z7-cyber/quantum-era/

November 2026 Threat Watch
More from the edition · z7solutions.com/z7-cyber/november-2026-threat-watch/

PHP2Shell needs patching and investigation
More from the edition · z7solutions.com/z7-cyber/php2shell-emergency-patching/

CALL 844-974-8669

EMAIL sales@z7solutions.com

BOOK A SECURITY CONVERSATION Book a Security Conversation ↗

Z7 Solutions™, Z7-Claw™ and Z7-Guard™ are trademarks of Z7 Solutions LLC.