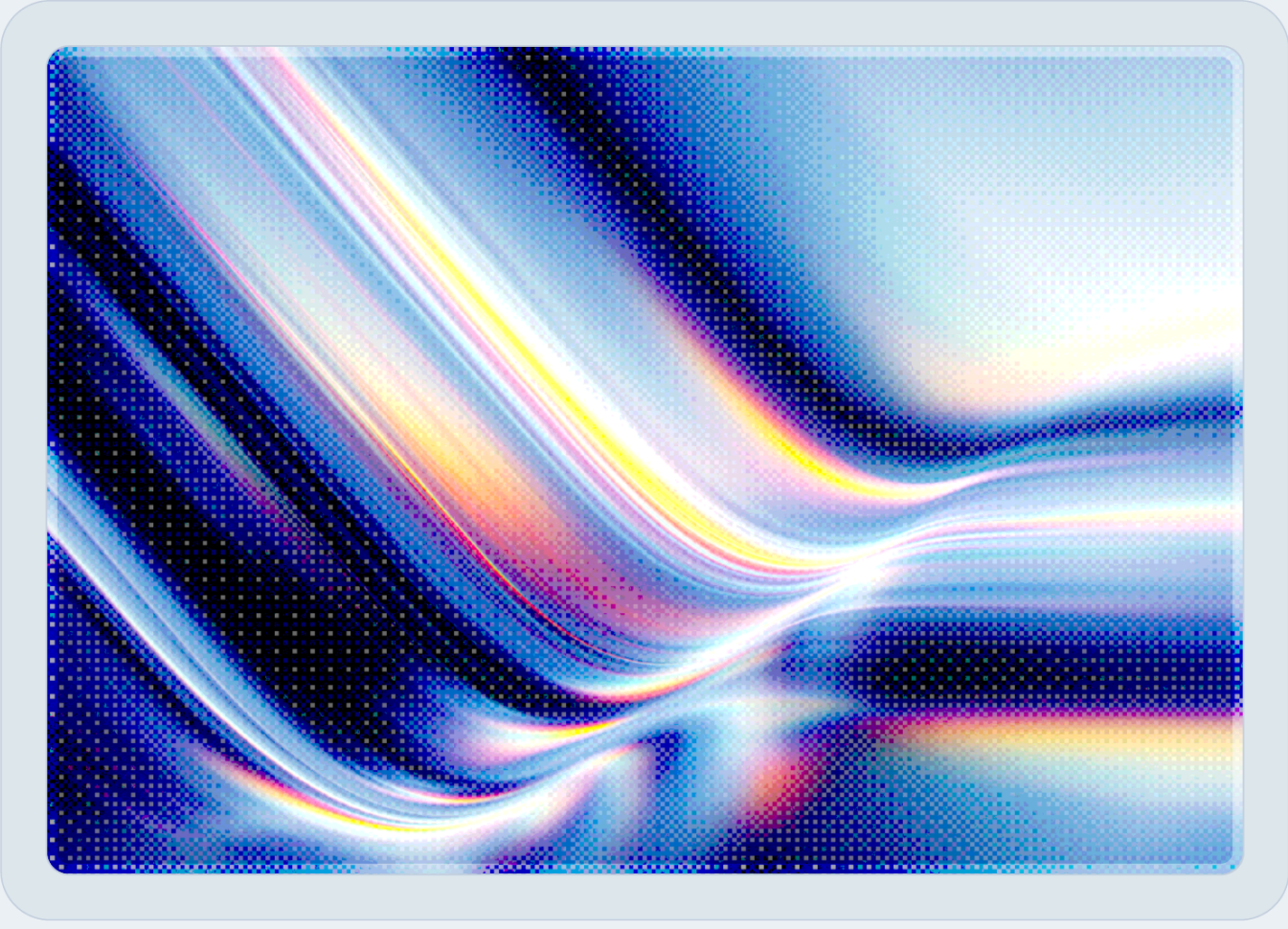
CSAI Foundation | Cloud Security Alliance

State-Sponsored Exploitation of the Langflow AI Platform

CVE-2025-34291: How a CORS Chain Became an APT Initial Access Vector

2026-05-24

AI-assisted Rapid Research



CSAI

cloud security alliance®

Key Takeaways

- CVE-2025-34291 is a critical (CVSS v4.0: 9.4) chained vulnerability in Langflow versions 1.6.9 and earlier that enables account takeover and remote code execution through a browser-based attack requiring no prior authentication. [1][11]
- The vulnerability combines three weaknesses in sequence: an overly permissive CORS configuration, a session cookie configured with `SameSite=None`, and a built-in code-execution endpoint — none of which is critical in isolation, but together constitute a full system compromise. [2]
- MuddyWater, an Iran-nexus APT group subordinate to Iran's Ministry of Intelligence and Security (MOIS), actively exploited this vulnerability for initial access into target networks, according to a March 2026 Ctrl-Alt-Intel analysis. [3]
- CISA added CVE-2025-34291 to its Known Exploited Vulnerabilities catalog on May 21, 2026, mandating federal remediation by June 4, 2026 under Binding Operational Directive 22-01. [4][7][10]
- Because Langflow workflows routinely connect to LLM providers, cloud storage, vector databases, and internal APIs, successful exploitation exposes credentials stored directly in Langflow workspace configurations — which in many deployments include API keys and tokens for every connected downstream service — transforming a single platform compromise into a cascading downstream breach. [2]
- Organizations should immediately set `LANGFLOW_CORS_ALLOW_CREDENTIALS=False` or restrict `LANGFLOW_CORS_ORIGINS` to trusted hosts, and should audit all credentials stored in or accessible through Langflow workspaces. [2]

Background

Langflow is a low-code visual platform for building and deploying agentic AI workflows and retrieval-augmented generation (RAG) applications. It allows developers, AI operations teams, and non-traditional engineering roles to compose multi-step agent pipelines through a drag-and-drop interface,

with integrated support for major LLM providers including OpenAI, Anthropic, and HuggingFace. With over 145,000 GitHub stars and growing adoption among enterprise teams, Langflow has emerged as one of the more widely adopted open-source platforms for agentic AI workflow composition. [5]

Langflow's appeal as a productivity platform is inseparable from what makes it an attractive adversarial target. A single Langflow workspace typically aggregates credentials for every downstream service its workflows touch — LLM API keys, cloud provider tokens, vector database credentials, internal API authentication, and automation platform secrets. This architectural concentration means that a Langflow instance is not merely a workflow engine; it is effectively a centralized credential store for an organization's entire AI integration surface. An adversary who achieves account takeover on a Langflow deployment does not gain access to one service — they gain an inventory of the connected ecosystem and the keys to traverse it.

CVE-2025-34291, disclosed against versions up to and including 1.6.9, turns that credential concentration into a directly exploitable attack path requiring nothing more than getting a logged-in user to visit an attacker-controlled webpage. Active exploitation was observed beginning January 23, 2026. [9] CISA's KEV listing on May 21, 2026 confirms that exploitation was sufficiently widespread and documented to warrant mandatory federal response. [4][7] The MuddyWater attribution, published by Ctrl-Alt-Intel in March 2026, elevates this from a commodity exploitation event to a state-sponsored initial access operation. [3]

Security Analysis

The Vulnerability Chain

CVE-2025-34291 is not a single design error but the combination of three weaknesses that interact across Langflow's authentication architecture. Understanding how they chain together is important for evaluating both the severity and the mitigation options.

The first weakness is Langflow's CORS policy. The application was configured with `allow_origins='*'` and `allow_credentials=True`. This combination represents a misconfiguration: the CORS specification treats the wildcard origin `*` as explicitly incompatible with credential inclusion, but certain implementation configurations can allow an attacker-controlled origin to successfully include credentials in cross-origin requests. The result is that a page served from an adversary's domain can make authenticated requests to a victim's Langflow instance as if it were operating from the same origin. [2]

The second weakness is the session token's cookie configuration. Langflow's `refresh_token_lf` cookie was set with `SameSite=None; Secure`, which is the most permissive cross-origin cookie policy available. Cookies marked `SameSite=None` are transmitted with all cross-origin requests, including those initiated from malicious third-party pages. This setting is typically reserved for specific cross-site authentication flows where it is architecturally necessary; its presence in Langflow's default session handling created the mechanism by which an attacker-controlled page could harvest live session tokens. [2]

The third element is Langflow's built-in code execution capability, which is a deliberate product feature rather than a vulnerability in isolation. Langflow allows Python code to be executed as part of workflow definitions — this is core to its flexibility as an agent composition platform. However, once an attacker possesses a valid access token obtained through the first two weaknesses, they can authenticate to these execution endpoints and run arbitrary code on the host. Account takeover effectively enables remote code execution: an attacker with a valid session token can authenticate to Langflow's workflow execution endpoints and run arbitrary Python code on the host with no additional vulnerability needed. [2]

An attacker exploiting this chain does not need to brute-force credentials, exploit a memory safety bug, or maintain persistent network access to the target. The entire attack proceeds through the victim's own browser: the attacker lures a logged-in Langflow user to a malicious page, the page silently issues cross-origin requests that harvest the session token, and the attacker then uses that token to execute code on the Langflow server. The victim's browser acts as an unwitting relay. [2]

Cascading Compromise and the AI Credential Aggregation Problem

The downstream consequences of a successful Langflow compromise extend well beyond the platform itself. In a typical enterprise Langflow deployment, the workspace configuration contains credentials for every service the organization's agents connect to: API keys for OpenAI, Anthropic, and other model providers; storage and database credentials for vector stores and document repositories; authentication tokens for internal APIs, enterprise SaaS platforms, and messaging services; and in many cases, cloud provider credentials with broad permissions over infrastructure. [2]

This aggregation is not an oversight — it is a deliberate aspect of how workflow orchestration platforms function. The convenience of composing multi-step agent pipelines through a unified visual environment requires that the platform hold the credentials necessary to authenticate against each downstream service. The security implication is that the platform becomes a single point of credential exposure: compromise the orchestration layer, and the attacker obtains a ready-made inventory of the organization's AI integration surface along with the authentication material to traverse it.

This pattern is structurally distinct from the classic credential sprawl problem of individual secrets scattered across an organization. Langflow concentrates credentials intentionally, and operators who have not audited what is stored in their workspaces may have no awareness of the blast radius a platform compromise carries. Organizations that have stored long-lived API keys in Langflow — rather than using short-lived credentials with minimal scoping — face potential exposure across an entire ecosystem of connected services from a single exploitation event.

MuddyWater's Exploitation Campaign

MuddyWater is an Iranian state-sponsored threat actor assessed as a subordinate element of Iran's Ministry of Intelligence and Security. The group is also tracked as Static Kitten, Mango Sandstorm, Earth Vetala, Seedworm, and TA450 across different intelligence vendors. MuddyWater is known for espionage campaigns against government agencies, telecommunications companies, defense contractors, and critical infrastructure, with a persistent geographic focus on Middle Eastern targets and a demonstrated capability to rapidly operationalize public exploit code. [3]

In the March 2026 Ctrl-Alt-Intel analysis, researchers documented MuddyWater's exploitation of CVE-2025-34291 as an initial access vector against target networks across Israel, Jordan, Egypt, the United Arab Emirates, Portugal, and the United States. [3] A concurrent Help Net Security report documented the group deploying new backdoor families against U.S. critical sector organizations during the same operational period. [6] The Ctrl-Alt-Intel analysis identified a campaign infrastructure involving multiple command-and-control servers and a novel botnet using Ethereum smart contracts for C2 communication — a finding that, if confirmed by additional sources, would represent a novel evasion technique in MuddyWater's operational toolkit — suggesting a sophisticated and actively maintained operational capability rather than opportunistic exploitation. [3]

This attribution is based on a single commercial threat intelligence report and has not been independently confirmed by government advisories at the time of writing. The campaign TTPs described are consistent with previously documented MuddyWater operations, and MuddyWater's subordination to MOIS is independently established through prior US Cyber Command, CISA, FBI, and NCSC attributions. CVE-2025-34291 demonstrates that AI orchestration platforms are viable high-value APT initial access targets — a threat model enterprises should actively incorporate into AI governance. A successful Langflow compromise in a target organization does not merely yield a foothold on one server — it yields the credentials, workflow logic, and integration topology of the organization's AI operations, which can inform follow-on collection objectives and lateral movement into connected cloud environments.

The timing of exploitation onset (January 23, 2026) relative to CISA's KEV listing (May 21, 2026) reflects a four-month gap between confirmed active exploitation and formal federal guidance. The gap in this case illustrates a risk for organizations that rely solely on KEV as a prioritization trigger — state-sponsored threat timelines may precede formal guidance by months. [4]

Recommendations

Immediate Actions

Organizations running any version of Langflow up to and including 1.6.9 should treat this as an active incident until they have either patched or implemented compensating controls. The highest-priority step is to disable credential-bearing cross-origin requests by setting the `LANGFLOW_CORS_ALLOW_CREDENTIALS` environment variable to `False`. Where cross-origin access is operationally required, restrict the `LANGFLOW_CORS_ORIGINS` setting to an explicit allowlist of trusted hostnames rather than permitting wildcard origins. [2]

Simultaneously, teams should audit Langflow workspaces for stored credentials and rotate any API keys, service account tokens, cloud provider credentials, or database passwords that may have been accessible through the platform during the exploitation window. Rotation should be treated as mandatory for any organization that cannot conclusively rule out exploitation — the attacker's primary objective in using CVE-2025-34291 for initial access was to harvest credentials for downstream services, not simply to compromise the Langflow host itself. [2]

Langflow instances exposed to the internet without network-layer access controls should be taken offline or placed behind a VPN or identity-aware proxy until patched. The browser-based attack vector for CVE-2025-34291 means that internal users accessing an internet-exposed Langflow instance are automatically at risk from any attacker who can serve them a malicious page, with no requirement for direct network access to the Langflow host.

Short-Term Mitigations

Upgrade to Langflow version 1.7.0 or later, which addresses the CORS and session cookie misconfiguration underlying CVE-2025-34291. [2] Teams should verify the upgrade resolves the specific cookie and CORS settings, not merely update the application version, and should confirm that `SameSite=None` is no longer the default configuration for session tokens in the deployed instance.

Implement network-layer segmentation for Langflow deployments. AI orchestration platforms that serve as credential aggregators for downstream services should not be directly accessible from the public internet or from untrusted internal network segments. Access should be gated through authenticated network controls — VPN, identity-aware proxy, or zero-trust network access (ZTNA) — that add a layer of protection independent of application-layer vulnerability status.

Establish a credential inventory for all Langflow workspaces. Document which external services, APIs, and cloud providers each workspace has access to, and assess the blast radius if that workspace were compromised. Use this inventory to drive scope prioritization for rotation, access scoping reduction, and migration to short-lived credentials where possible.

Strategic Considerations

CVE-2025-34291 is a case study in the credential aggregation risk specific to AI orchestration platforms. Unlike conventional enterprise software, which typically authenticates against one or two downstream services, agentic workflow platforms are architecturally designed to hold credentials for many services simultaneously. Security teams should apply the same threat model to Langflow deployments, and to analogous platforms such as n8n, Flowise, and Dify, that they would apply to a privileged secrets management system: least-privilege access controls, mandatory audit logging, systematic credential rotation, and heightened detection monitoring.

The exploitation of CVE-2025-34291 demonstrates that AI orchestration platforms are viable APT initial access targets — a threat model enterprises should actively incorporate into AI governance for any similar platform. Access to Langflow — and similar platforms — should be treated as access to a highly privileged credential broker, not as access to a productivity tool. This means applying elevated access controls, enforcing MFA, reviewing all stored credentials against the principle of least privilege, and ensuring that operator activity within these platforms is logged to a centralized security information and event management (SIEM) system with detection rules for anomalous workflow creation or code execution. For high-risk or sensitive deployments, privileged access workstation (PAW) controls may be warranted; at minimum, enforce MFA and dedicated session management for all Langflow operators.

Organizations should also evaluate whether agentic workflow platforms should hold long-lived credentials at all. Where cloud provider integrations are involved, migration to short-lived credentials obtained through workload identity federation — where supported by the cloud provider, such as AWS IAM Roles Anywhere, GCP Workload Identity Federation, or Azure Managed Identities — removes the static credential exposure that makes a Langflow compromise so consequential. For LLM provider API keys and other services that do not yet support workload identity patterns, scoping keys to minimum necessary permissions and rotating them on a scheduled cadence reduces the window of exposure following a platform breach.

CSA Resource Alignment

CVE-2025-34291 maps directly to threat categories addressed by several CSA frameworks, and organizations responding to this vulnerability should use these resources to inform both immediate response and longer-term governance.

CSA's MAESTRO framework, a seven-layer threat modeling architecture for agentic AI systems, addresses the class of vulnerabilities represented by this CVE at multiple layers. [8] Layer 3 (Agent Frameworks) is the direct impact layer — MAESTRO specifically identifies the agent framework as a trust boundary where authentication weaknesses can propagate through the entire agent stack. Layer 4 (Deployment Infrastructure) governs the CORS and session configuration weaknesses underlying the vulnerability. Layer 7 (Agent Ecosystem) captures the cascading downstream credential exposure risk, reflecting MAESTRO's recognition that agentic orchestration platforms serve as integration hubs whose compromise radius extends to every connected service within the broader agent ecosystem. Organizations should use MAESTRO's Layer 3 and Layer 7 threat catalogs to scope their incident response and credential rotation priorities.

The AI Controls Matrix (AICM), CSA's control framework for AI systems and the superset of the Cloud Controls Matrix for AI-specific governance, provides control categories directly applicable to the CVE's root causes. Access control and identity management controls in the AICM address the insufficient authentication boundary that allowed cross-origin credential extraction. Secrets and key management controls address the credential aggregation pattern that amplifies the blast radius of a platform compromise. Organizations should use AICM as a reference baseline for evaluating whether compensating controls are in place for other AI platform deployments beyond Langflow.

CSA's Zero Trust guidance provides the architectural framing for the network segmentation and identity-aware access controls recommended above. Treating Langflow as a privileged service requiring explicit authentication and strict access controls — rather than a developer-accessible productivity tool — aligns with zero-trust principles and eliminates the broad network exposure that enables browser-based exploitation.

CSA's Security Trust Assurance and Risk (STAR) program provides a structured mechanism for cloud service providers and AI platform vendors to publish security assessments. Organizations evaluating Langflow or similar platforms for enterprise deployment should review available vendor security documentation and, where STAR assessments are available, use them to validate that baseline controls meet enterprise requirements before exposing these platforms to production workflows.

References

[1] NIST National Vulnerability Database. "CVE-2025-34291 Detail." NVD, 2025–2026.

[2] Obsidian Security. "CVE-2025-34291: Critical Account Takeover and RCE Vulnerability in the Langflow AI Agent & Workflow Platform." Obsidian Security Blog, 2026.

[3] Ctrl-Alt-Intel. "MuddyWater Exposed: Inside an Iranian APT Operation." Ctrl-Alt-Intel Research, March 2026.

[4] CISA. "Known Exploited Vulnerabilities Catalog." U.S. Cybersecurity and Infrastructure Security Agency, May 21, 2026.

[5] GitHub. "langflow-ai/langflow." GitHub, 2026.

[6] Help Net Security. "Iran-linked APT Targets US Critical Sectors with New Backdoors." Help Net Security, March 6, 2026.

[7] The Hacker News. "CISA Adds Exploited Langflow and Trend Micro Apex One Vulnerabilities to KEV." The Hacker News, May 2026.

[8] Cloud Security Alliance. "Agentic AI Threat Modeling Framework: MAESTRO." CSA Blog, February 6, 2025.

[9] CrowdSec. "CVE-2025-34291 Exploited in the Wild: LangFlow AI Framework Under Fire." CrowdSec VulnTracking Report, January 26, 2026.

[10] Security Affairs. "U.S. CISA adds Trend Micro Apex One and Langflow to its Known Exploited Vulnerabilities Catalog." Security Affairs, May 2026.

[11] GitHub Advisory Database. "Langflow versions up to and including 1.6.9 – GHSA-577h-p2hh-v4mv." GitHub, 2026.

