Una falla de seguridad crítica que afecta a Langflow ha comenzado a ser explotada activamente en menos de 20 horas tras su divulgación pública, lo que pone de manifiesto la rapidez con la que los actores maliciosos convierten nuevas vulnerabilidades en herramientas de ataque.

El defecto de seguridad, identificado como CVE-2026-33017 (puntuación CVSS: 9.3), corresponde a un caso de ausencia de autenticación combinada con inyección de código que podría derivar en ejecución remota.

«El endpoint POST /api/v1/build_public_tmp/{flow_id}/flow permite construir flujos públicos sin necesidad de autenticación,» según el aviso oficial de Langflow sobre la vulnerabilidad.

«Cuando se proporciona el parámetro opcional data, el endpoint utiliza datos del flujo controlados por el atacante (que incluyen código Python arbitrario en las definiciones de nodos) en lugar de los datos almacenados en la base de datos. Este código se pasa a exec() sin ningún tipo de sandbox, lo que provoca ejecución remota sin autenticación.»

La vulnerabilidad impacta a todas las versiones de esta plataforma de inteligencia artificial de código abierto hasta la 1.8.1 inclusive. Actualmente ha sido corregida en la versión de desarrollo 1.9.0.dev8.

El investigador de seguridad Aviral Srivastava, quien descubrió y reportó el problema el 26 de febrero de 2026, señaló que es diferente de CVE-2025-3248 (CVSS: 9.8), otro fallo crítico en Langflow que explotaba el endpoint /api/v1/validate/code para ejecutar código Python arbitrario sin autenticación. Según la Cybersecurity and Infrastructure Security Agency, este último ya está siendo explotado activamente.

«CVE-2026-33017 se encuentra en /api/v1/build_public_tmp/{flow_id}/flow,» explicó Srivastava, añadiendo que la causa raíz proviene del uso del mismo exec() que en CVE-2025-3248 al final de la cadena.

«Este endpoint está diseñado para no requerir autenticación porque sirve flujos públicos. No

se puede simplemente añadir autenticación sin romper esta funcionalidad. La solución real es eliminar completamente el parámetro data del endpoint público, de modo que solo se ejecuten los datos almacenados en el servidor y nunca definiciones proporcionadas por atacantes.»

Una explotación exitosa permitiría a un atacante enviar una sola solicitud HTTP y lograr ejecución de código arbitrario con todos los privilegios del proceso del servidor. Con ese acceso, podría leer variables de entorno, acceder o modificar archivos para introducir puertas traseras o eliminar datos sensibles, e incluso obtener una shell remota inversa.

Srivastava comentó a *The Hacker News* que explotar CVE-2026-33017 es *«extremadamente sencillo»* y puede activarse mediante un comando curl preparado para el ataque. Basta una única solicitud HTTP POST con código Python malicioso en el payload JSON para conseguir ejecución remota inmediata.

La empresa de seguridad en la nube Sysdig indicó que detectó los primeros intentos de explotación en entornos reales dentro de las 20 horas posteriores a la publicación del aviso el 17 de marzo de 2026.

«No existía código público de prueba de concepto (PoC) en ese momento,» señaló Sysdig. *«Los atacantes crearon exploits funcionales directamente a partir de la descripción del aviso y comenzaron a escanear internet en busca de instancias vulnerables. La información exfiltrada incluía claves y credenciales, lo que permitió acceder a bases de datos conectadas y potencialmente comprometer la cadena de suministro de software.»*

También se ha observado que los atacantes pasaron de escaneos automatizados a usar scripts personalizados en Python para extraer datos de «/etc/passwd» y desplegar una carga útil de segunda fase alojada en «173.212.205[.]251:8443». La actividad posterior desde esa misma dirección IP apunta a una operación completa de recolección de credenciales, que incluye la obtención de variables de entorno, enumeración de archivos de configuración y bases de datos, así como la extracción del contenido de archivos .env.

Esto sugiere que el atacante ya contaba con una planificación previa, preparando malware para desplegarlo en cuanto detectara un objetivo vulnerable. *«Se trata de un atacante con un kit de explotación listo que pasa de validar la vulnerabilidad a desplegar la carga útil en una sola sesión,»* destacó Sysdig. Por ahora, no se conoce quién está detrás de los ataques.

El intervalo de 20 horas entre la publicación del aviso y la primera explotación coincide con una tendencia creciente: el tiempo medio hasta la explotación (TTE) se ha reducido de 771 días en 2018 a apenas horas en 2024.

De acuerdo con el informe Rapid7 Global Threat Landscape Report 2026, el tiempo medio desde la divulgación de una vulnerabilidad hasta su inclusión en el catálogo KEV de CISA bajó de 8.5 días a cinco días en el último año.

«Esta compresión del tiempo representa un desafío serio para los defensores. El tiempo medio que tardan las organizaciones en aplicar parches es de aproximadamente 20 días, lo que significa que permanecen expuestas durante demasiado tiempo,» añade el informe. *«Los atacantes monitorean las mismas fuentes de avisos que los defensores y desarrollan exploits más rápido de lo que la mayoría de las organizaciones puede evaluar, probar e implementar parches. Las organizaciones deben replantear por completo sus programas de gestión de vulnerabilidades para adaptarse a esta realidad.»*

Se recomienda a los usuarios actualizar a la versión corregida lo antes posible, auditar variables de entorno y secretos en cualquier instancia pública de Langflow, rotar claves y contraseñas de bases de datos como medida preventiva, vigilar conexiones salientes hacia servicios inusuales y restringir el acceso de red mediante reglas de firewall o proxies inversos con autenticación.

La actividad observada en torno a CVE-2025-3248 y CVE-2026-33017 demuestra que las cargas de trabajo de inteligencia artificial se han convertido en objetivos prioritarios para los atacantes debido a su acceso a datos valiosos, su integración en la cadena de suministro de software y la falta de controles de seguridad robustos.

«CVE-2026-33017 [...] demuestra un patrón que está dejando de ser una excepción para convertirse en la norma: vulnerabilidades críticas en herramientas open source populares son explotadas en cuestión de horas tras su divulgación, a menudo incluso antes de que exista código público de prueba de concepto,» concluyó Sysdig.