ultraviolet

THREAT ADVISORY

DifyTap Vulnerabilities

Services Performed By:
UltraViolet Cyber TIDE Team
tide@uvcyber.com

Published Date:
June 24, 2026
TLP:GREEN

Executive Snapshot

The DifyTap research disclosed by Zafran Labs in June 2026 exposes four vulnerabilities in Dify, a widely deployed open-source AI platform, two of which are rated critical and two of which require no authentication to exploit. The flaws allow attackers to silently intercept AI conversation data across organizational boundaries, traverse internal APIs without credentials, and access documents and files belonging to other tenants, all without sophisticated tooling or elevated access. The platform's file parsing stack also carried a known memory corruption CVE for over 18 months undetected, and standard container scanning tools failed to surface any of these vulnerabilities due to a structural gap in how AI application layers are assessed. These findings are not isolated to Dify; they represent a repeating risk pattern across third-party AI and LLM platforms where rapid adoption has outpaced security rigor, and any enterprise running AI workflow tools, chatbots, or LLM Ops platforms sourced from third-party vendors is carrying vendor risk that requires active governance.

- Conduct an immediate inventory of all third-party AI and LLM platforms deployed in your environment, including shadow IT and departmentally adopted tools, and verify that each is running the vendor's latest patched release. For Dify specifically, confirm that version 1.14.2 or later is in place.

- Review and restrict which internal data sources, file repositories, and knowledge bases are connected to third-party AI platforms. Sensitive data should not be ingested into any AI platform that has not undergone a formal security assessment, and access to AI tools that process confidential information should be limited to explicitly authorized users.

- Establish a vendor security review requirement for all AI and LLM platform procurement and renewals that explicitly evaluates multi-tenant architecture, authentication controls, and the vendor's patch cadence for both application-level and dependency-level vulnerabilities before approval.

TIDE Team Analysis

The DifyTap vulnerability research published by Zafran Labs in June 2026 reveals a systemic and consequential security problem that extends well beyond a single vendor. Dify is an open-source LLM Ops platform powering more than one million AI applications and used by enterprises including large organizations like Volvo, Maersk, Panasonic, and Thermo Fisher. The discovery of four vulnerabilities in its architecture, two rated critical with CVSS scores of 9.4 and 9.1, two requiring no authentication, and three carrying cross-tenant impact, should serve as a direct warning to any organization that has deployed third-party AI application platforms without rigorous security vetting.

The most operationally dangerous finding concerns the platform's tracing subsystem, which logs all AI conversation data including user messages, model prompts, and model responses. Because the tracing configuration endpoints fail to validate the requesting user's tenant identity, any attacker holding a basic Dify console account, obtainable through free public registration, can redirect the conversation logs of any publicly accessible application to an attacker-controlled endpoint. The result is a persistent, silent wiretap on every interaction users have with an enterprise AI chatbot. In environments where employees use AI assistants to process business logic, customer records, legal documents, or financial data, this vulnerability represents a continuous and covert data exfiltration channel that could operate undetected for extended periods.

Compounding that risk is a second critical vulnerability in the platform's Plugin Daemon, the internal microservice that manages Dify's plugin ecosystem. A path traversal flaw in the icon-fetching API allows any unauthenticated user with network access to reach arbitrary internal API endpoints, including those belonging to other tenants. The absence of any authentication requirement on this endpoint is particularly alarming from an architectural standpoint, as no credential, token, or session is required to begin probing internal infrastructure. This is not a nuanced boundary-condition flaw; it is a fundamental design failure in how the platform separates internal and external trust boundaries.

The file access vulnerabilities, tracked as CVE-2026-41949 and CVE-2026-41950, illustrate a different but equally serious class of risk: the failure of authorization logic in AI-native file pipelines. The preview endpoint for knowledge base documents performs no ownership or tenancy check beyond confirming that the file type is a document, meaning any console user can retrieve the first 3,000 characters of any file uploaded by any tenant in the system. A separate flaw allows any end user to attach another user's file UUID to their own chat session and prompt a file-capable AI model to read the contents back verbatim. These are not sophisticated exploits requiring specialized tooling; they require a browser, a registered account, and basic knowledge of how the API works.

Beyond the four logic vulnerabilities, public reporting also identified that Dify's file parsing stack had been running a version of PDFium susceptible to CVE-2024-5846, a known use-after-free memory corruption vulnerability, for more than 18 months after public disclosure. Any end user with document upload capability could have triggered this vulnerability by submitting a crafted PDF. This finding points to a broader category risk that security leadership must account for: AI platforms process a wide variety of file formats from untrusted external sources, including PDFs, Office documents, and images, and the underlying parsing libraries in these stacks carry real CVE exposure that is not being patched at the rate the threat environment demands.

The container security gap identified in this research deserves specific attention from security leadership. Security Researchers found that standard vulnerability scanners failed to detect any of the Dify application-level CVEs because Dify deploys by copying unpackaged Python source code directly into container images rather than installing

discrete packages. This means that the SBOM tooling, container image scanners, and cloud security posture management platforms that most enterprise security teams rely on were producing a clean bill of health for an environment carrying multiple critical-severity vulnerabilities. This is not a minor scanner gap; it reflects a structural blind spot in how the industry approaches AI workload security, and organizations should not assume their existing tooling provides meaningful coverage of the AI application layer.

The breadth of Dify's enterprise footprint means that the risk surface exposed by DifyTap is not theoretical or narrow in scope. The platform has achieved significant production adoption across more than 60 industries, with tens of thousands of publicly reachable instances identified during the research, placing it firmly in the category of mainstream enterprise infrastructure rather than experimental tooling. The problem this creates for security leadership is straightforward: organizations that onboarded Dify based on its open-source credentials and feature velocity are unlikely to have applied the same vendor security review process they would require of a traditional enterprise application, leaving its multi-tenant architecture, access control design, and third-party dependency chain unexamined. That gap is precisely what the DifyTap findings exploit. AI platforms are being adopted at a procurement pace that consistently outstrips the security review cycles organizations have built for conventional software, and until those cycles are extended to cover the AI platform layer with equal rigor, the conditions that produced DifyTap will continue to produce similar disclosures across the broader AI tooling ecosystem.

The strategic takeaway for security leadership is that the DifyTap findings represent a risk pattern that will recur across the broader AI platform landscape. Third-party LLMOps platforms, AI workflow builders, and enterprise AI toolkits introduce a new vendor risk category that combines the data sensitivity of AI workloads with the architectural complexity of microservices, the ingestion risks of unstructured file parsing, and the access control challenges of multi-tenancy. Organizations should immediately audit all deployed AI platforms for similar trust boundary failures, assess whether their vulnerability scanning tooling provides genuine coverage of the AI application layer, and require vendors to demonstrate regular security assessments as a condition of procurement or continued deployment. Patches for the DifyTap vulnerabilities are available in Dify version 1.14.2, and organizations running earlier versions should treat remediation as an immediate priority.

Why It Matters

The DifyTap vulnerabilities do not exist in isolation. They follow a pattern of security failures in AI and LLM platforms that has been building since the rapid commercialization of generative AI began in earnest. In 2024, researchers disclosed critical vulnerabilities in Langchain, one of the most widely adopted AI orchestration frameworks, that allowed server-side request forgery and arbitrary code execution through maliciously crafted inputs to document loaders and chain components. That same year, the Chainlit AI framework, another popular tool for building conversational AI interfaces, was found to be exposing cloud API keys which could enable full cloud account takeover through logic flaws in its session handling. These were not obscure edge cases in experimental software; they were production vulnerabilities in tools running inside enterprise environments processing real business data. The throughline connecting these incidents to DifyTap is consistent: AI platforms are being built and adopted at a speed that leaves foundational security controls, particularly around authentication, authorization, and tenant isolation, implemented incompletely or not at all.

What makes this pattern particularly consequential for enterprise environments is the nature of the data that flows through these platforms. Unlike a vulnerability in a network appliance or a web server, a flaw in an AI platform sits at

the intersection of every sensitive data stream an organization has chosen to feed into its AI workflows. Retrieval-augmented generation pipelines routinely ingest legal documents, financial records, internal policies, customer data, and proprietary research. Conversational AI tools deployed for employee productivity capture internal deliberations, strategic discussions, and operational detail in every session. When an attacker compromises the platform layer rather than targeting a specific application, they inherit access to all of that data simultaneously, across every tenant sharing the infrastructure. The blast radius of a single AI platform vulnerability is structurally larger than most traditional application vulnerabilities because of the breadth and sensitivity of what these platforms are designed to aggregate and process.

The threat surface will grow materially as AI platforms evolve toward agentic architectures. Current LLMOps platforms primarily manage data ingestion, prompt routing, and response delivery. The next generation of enterprise AI tools is being designed to take autonomous actions: executing code, calling external APIs, managing files, sending communications, and interacting with internal systems on behalf of users. Several platforms in active enterprise deployment are already offering agentic workflow capabilities, and the plugin and tool-use ecosystems that support them, as illustrated by the Plugin Daemon flaw in DifyTap, introduce internal API surfaces that are poorly understood by security teams and often entirely absent from threat models. As these systems gain the ability to act rather than just respond, the consequences of a trust boundary failure escalate from data exposure to direct operational impact, including the potential for lateral movement, privilege escalation, and manipulation of business processes through the AI layer itself.

The regulatory and liability environment surrounding these risks is also maturing rapidly, which raises the organizational stakes beyond the technical. Data protection frameworks including GDPR, HIPAA, and emerging AI-specific regulations in both the EU and United States treat unauthorized access to personal or sensitive data as a compliance event regardless of whether the access pathway was a traditional breach or a vulnerability in a vendor-supplied AI platform. Cross-tenant data exposure of the kind demonstrated in DifyTap, where one organization's data becomes readable by another tenant on the same platform, has clear regulatory implications that security leadership will be required to address, report, and remediate under defined timelines. As AI platforms become embedded in regulated workflows, the argument that AI tooling exists outside the normal scope of vendor risk management and compliance oversight is no longer tenable, and organizations that have not yet treated their AI platform layer with the same governance rigor as their core enterprise systems are accumulating unacknowledged liability.

How to Respond

- Strictly adhere to cybersecurity fundamentals and ensure all personnel undergo annual phishing and social engineering training. Speak with your UltraViolet Cyber TAM Representative to schedule a live phishing engagement.
- Conduct a third-party AI/LLM tool inventory to understand the attack surface in your environment.
- Ensure that Dify is updated to version 1.14.2 or later.
- Perform annual tech refresh reviews to gain a holistic understanding of your infrastructure. Speak with your UltraViolet Cyber TAM Representative to schedule a Red Team or Purple Team engagement to gain insight into the vulnerabilities in your environment.

What UltraViolet Cyber is Doing

- Tracking new CVEs and high impact vulnerabilities, analyzing and deploying public Proof-Of-Concept code against custom built targets.
- Proactively enabling custom detections based on the collected artifacts, tactics, techniques, and procedures identified in this activity.
- Performing hypothesis driven threat hunts based on threat actor behavior and artifacts. UVCyber customers will be informed of the results through secure channels.
- Parsing available victim dump data for any social, financial, business, or technical relations to UVCyber Clients and partner organizations.
- Aggregating threat intelligence from myriad sources and applying the most up-to-date knowledge to proactive threat hunting and response.

About UltraViolet Cyber

UltraViolet Cyber is a leading tech-enabled managed security services provider, delivering unparalleled cybersecurity expertise that fills technology and talent gaps across Global 2000 and Federal Government customers. Founded and operated by security practitioners from the national intelligence community, UltraViolet Cyber connects offensive security, application security, detection and response, and security engineering to deliver a differentiated approach to cybersecurity operations. Transforming customers' security programs, UltraViolet Cyber's flagship security-as-a-service solution, UV Lens, removes complex operational silos, replacing them with integrated security capabilities. UltraViolet is headquartered in McLean, Virginia with technology centers across the world.

443.351.7630 / info@uvcyber.com | UltraViolet Cyber | @uv_cyber

---