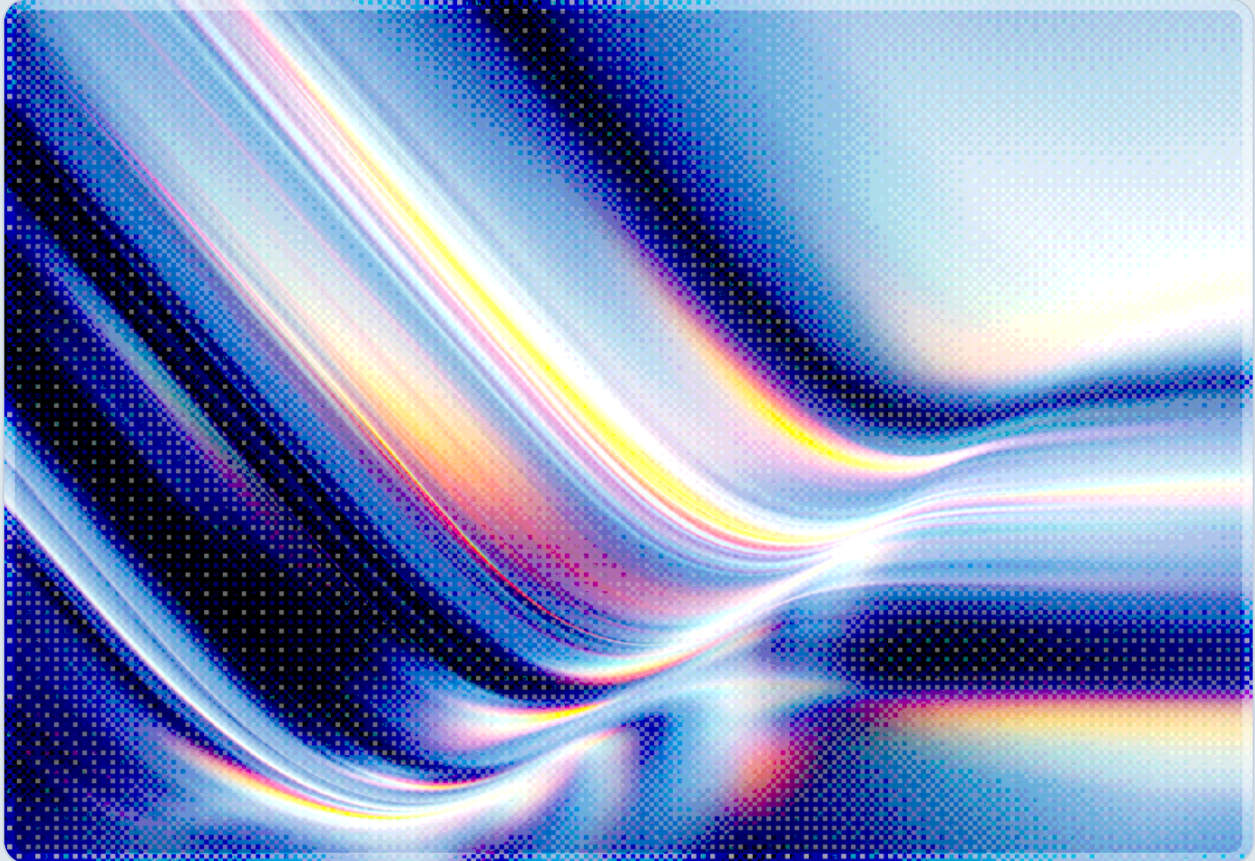


When Discovery Outruns Remediation: Open-Source Disclosure Strain

AI-Driven Report Floods and the Strain on the Open-Source Vulnerability Disclosure System

2026-10-07

 AI-assisted Rapid Research



CSAI

cloud
CSA security
alliance®

© 2026 Cloud Security Alliance. Some rights reserved.

You may download, store, display, view, print, redistribute, and link to this document in its original, unmodified form, provided that attribution to the Cloud Security Alliance is maintained and all trademark and copyright notices remain intact.

This document may not be modified or altered. You may quote portions of the document as permitted by the Fair Use provisions of the United States Copyright Act, provided that attribution is given to the Cloud Security Alliance.

This document may be shared on professional and social media platforms in its original form with attribution.

This document was generated with AI assistance and has not undergone official CSA review and approval processes.

Key Takeaways

Between January and October 2026, several widely used intake channels for open-source vulnerability reports were ended, paused, or restructured because the volume of AI-assisted submissions exceeded what maintainers could process. curl ended its paid bug bounty [1], HackerOne paused its Internet Bug Bounty [4], the Linux kernel project rewrote its guidance on how AI-found bugs should be reported [3], and on October 1 Google stopped accepting new product vulnerability submissions to its Open Source Software Vulnerability Reward Program [2].

The strain appears to have two sources that call for different responses. Most documented closures, including curl's and Google's, cite invalid or low-quality submissions, which waste triage time and are largely an incentive and filtering problem. The evidence for a second source, a rising volume of valid and duplicate findings that is a remediation-capacity problem filtering cannot solve, is thinner. It rests on the Linux kernel's experience, reporting on the HackerOne pause, and CSA's Glasswing analysis [3][4][6].

Enterprises that consume open-source software should treat maintainer capacity as a supply-chain risk factor. Slower triage and fewer funded disclosure channels plausibly lengthen the window between discovery and a published fix, although public data on that interval for individual projects is limited. Organizations that use AI tools to find vulnerabilities in other people's code take on an obligation to validate findings and, where possible, to contribute fixes.

Background

Coordinated vulnerability disclosure (CVD) in open source has historically rested on a small number of volunteer or lightly funded maintainers, a security contact or private mailing list, and a reporting culture in which finding a flaw took considerable human effort. In our characterization, that effort acted as an informal filter. Paid bounty programs, including HackerOne's Internet Bug Bounty, added a financial incentive on top of it, which broadened participation but also gave submitters a reason to file speculative reports.

AI-assisted tooling changed the economics on both sides of that arrangement. In our assessment, generating a plausible-looking vulnerability report now costs very little, while evaluating it still requires a human who understands the code. Daniel Stenberg, curl's lead maintainer, announced in January 2026

that the project would end its bug bounty to "remove the incentive for people to submit crap and non-well researched reports," adding that "the current torrent of submissions put a high load on the curl security team" [1]. A secondary account of curl's later reporting pause, from July 1 to August 3, 2026, describes the share of confirmed vulnerabilities in submissions falling from roughly 15 percent to under 5 percent, a decline that the account says began in 2025 rather than during the pause itself. We treat that figure as indicative rather than authoritative, since it comes from a blog summary rather than from the project's own statistics [9].

A second pressure emerged more recently and is harder to dismiss as noise. As frontier models became capable of finding real bugs in mature codebases, maintainers began receiving reports that were technically correct. CSA's analysis of Project Glasswing, an Anthropic-led effort, described AI-driven discovery producing more than 10,000 high- and critical-severity candidate findings, while only 97 of the 1,596 vulnerabilities disclosed across 281 projects (about 6 percent) had been patched as of May 22, 2026 [6]. The CSA research note on NVD infrastructure found the same pattern downstream: CVE publications reached 48,185 in 2025, the first quarter of 2026 ran about one-third above the prior year, and NIST announced on April 15, 2026 that it would limit full enrichment to a risk-based subset of vulnerabilities, estimated at 15 to 20 percent of anticipated volume and leaving 80 to 85 percent without full enrichment [7]. Together these developments describe a pipeline in which each stage, from intake to patch to database enrichment, was sized for a lower rate of input.

Security Analysis

How the intake channels have responded

The responses to date fall into recognizable patterns. Some projects removed financial incentives, some paused intake entirely, and some changed the rules about what belongs in a private channel. The table below summarizes the public record for the four cases covered in this note.

Date	Actor	Action	Stated or reported rationale
January 2026	curl	Ended its paid bug bounty	Remove the incentive for low-effort and AI-generated submissions; reduce load on the security team [1]

Date	Actor	Action	Stated or reported rationale
April 2026 (reported April 17)	HackerOne Internet Bug Bounty	Paused new submissions	Imbalance between AI-accelerated discovery and open-source remediation capacity [4]
May 2026	Linux kernel	Revised security documentation; Torvalds criticized duplicate AI-found reports	Private security list had become hard to manage because many people reported the same flaws [3]
October 1, 2026	Google OSS VRP	Stopped accepting new product vulnerability submissions; reform update promised for Q1 2027	"A significant rise in automated submissions, the vast majority of which are not valid" [2]

The HackerOne pause is notable because the program's purpose was to pay for findings in widely used open-source projects. Reporting on the pause attributes it to the growing gap between how quickly vulnerabilities can be found and how quickly maintainers can fix them, rather than to spam alone, although the public statements leave some ambiguity about how much weight each factor carried [4]. That ambiguity matters, because a program closed for spam can reopen with better filtering, whereas a program closed for lack of remediation capacity is unlikely to reopen on filtering improvements alone.

Two failure modes that look alike

From a maintainer's inbox, a fabricated report and a valid but redundant report both consume attention. The consequences differ, however. Fabricated reports, for instance those that cite functions that do not exist as maintainers have described, are a cost with no offsetting benefit and can in principle be deterred with proof-of-concept requirements, reputation systems, or the removal of payment. Torvalds' comments on the Linux 7.1 release cycle reflect the second mode: he did not object to AI tooling, but to reports that arrive without validation or a fix, and to the repeated discovery of identical flaws by researchers using similar tools [3]. Duplicate reporting is a plausible consequence of many parties pointing comparable models at the same public code, and it would be expected to grow with the number of tool users rather than the number of bugs, although we are not aware of published measurements.

The kernel's new guidance shows how a project can adapt. As reported, it asks that reports include reproduction evidence and technical detail, ideally with a tested patch; it discourages speculative findings; and it directs AI-discovered issues toward the relevant maintainers instead of the private security list, treating many of them as public rather than embargoed [3]. In our reading, this partly abandons the assumption that every vulnerability report should be handled under embargo. If many parties are likely to find the same flaw, the secrecy of a private channel offers limited protection and a substantial coordination cost.

Effects on the wider ecosystem

When intake channels close, three downstream effects are plausible, though we are not aware of published measurements for any of them. First, researchers who would have reported through a paid channel may move to public disclosure, to selling findings, or to not reporting at all. Second, the maintainers who remain responsive are the ones who receive the most reports, concentrating burnout in a small number of people. Third, enterprises relying on a project may receive fixes later and with less advance notice, which compounds the patch-velocity pressure that CSA has described on the enterprise side [6][7][10][11].

CSA's June 2026 research note on CVD reform argues that the 90-day disclosure norm and the assumption of an individual, accountable human researcher do not fit autonomous discovery at current volumes [5]. The Google action in October, together with the earlier curl and kernel cases, is consistent with a central recommendation of that note: attach a responsible human operator to AI-generated findings. In the curl and Google cases, unvalidated reports are the main complaint. The kernel case also involves duplicate valid reports, which operator accountability would address only in part.

Limits of the current evidence

Much of the public record consists of maintainer statements and press coverage rather than systematic data. We have not found a published, cross-project measurement of report volumes, validity rates, or time-to-fix before and after AI tooling became common. The percentage figures that circulate, including curl's validity rates, come from individual maintainers or secondary summaries and should not be generalized to other projects [9]. The accounts of the HackerOne pause and the Linux kernel guidance likewise rest on secondary reporting [3][4], and primary sources such as the HackerOne notice and the kernel documentation should be consulted as they become available. Readers should also note that the HackerOne and Google actions are described by their operators as pauses, so the situation may change quickly as reforms are announced [2][4].

Recommendations

Immediate Actions

Organizations that submit vulnerability reports using AI assistance should reproduce each finding manually, confirm it against the current release, and search existing reports and fixes before contacting a maintainer. Reports should state plainly where AI tooling was involved and who is accountable for the claim, consistent with the human-operator principle in CSA's CVD reform note [5]. Where feasible, a report should come with a tested patch, which is the form of contribution that the kernel guidance and Torvalds' comments favor [3].

Security teams that depend on open-source components should identify the projects that are both critical to their products and thinly maintained, judged by criteria such as the number of active maintainers, funding, and release cadence, and confirm how those projects currently accept reports. Given that curl, the Internet Bug Bounty, and Google's open-source program have each changed their intake in 2026, any internal assumption about where to send a finding or how quickly a fix will follow should be rechecked [1][2][4].

Short-Term Mitigations

Organizations running AI-driven discovery programs against open-source code should route findings through an internal validation stage before external disclosure, and should measure their own submission validity rate. Teams should also coordinate with other researchers, where practical, to reduce duplicate reports of the same flaw, and should prefer a project's documented channel over a private mailing list when the project has stated that AI-found issues belong elsewhere [3].

Enterprises that benefit from open-source software can redirect part of their security tooling budget to remediation. Funding maintainer time, sponsoring triage, and contributing patches address the bottleneck that CSA's Glasswing analysis identifies; additional discovery spending alone is unlikely to [6]. Vulnerability management teams should also plan for incomplete NVD enrichment, since NIST's April policy leaves an estimated 80 to 85 percent of anticipated CVE volume without full enrichment [7], and should rely on vendor advisories and project-level data for prioritization.

Strategic Considerations

The longer-term question is whether disclosure norms and funding models can be redesigned for a world in which finding bugs is inexpensive. In our assessment, per-finding bounties are poorly matched to cheap discovery. The pauses at HackerOne and Google are consistent with that view, although neither operator has framed its decision this way [2][4]. Alternatives that merit evaluation include payment for validated fixes rather than reports, reputation-gated submission rights, shared de-duplication services across tool vendors, and funding for maintainers as ongoing capacity rather than per-bug rewards. None of these has yet been demonstrated at ecosystem scale, and each introduces its own abuse risks.

Standards bodies and governments also have a part to play. CSA's CVD reform note calls for updated CVD guidance through bodies such as CISA and NIST that addresses autonomous agents explicitly [5]. We would add, as a normative view rather than an established practice, that vendors of frontier models whose products are used for vulnerability discovery are well placed to build validation and de-duplication into their tools, and to share responsibility for the resulting load on maintainers.

CSA Resource Alignment

CSA has published several directly relevant pieces of work in 2026. The research note [Reforming Coordinated Vulnerability Disclosure for the Autonomous Bug Hunter Era](#) [5] provides the policy frame for this note. Its recommendations to assign a responsible human operator and to modernize the 90-day disclosure model correspond directly to the intake failures documented here, and the October Google action adds evidence that the burden falls on maintainers when the operator step is missing. A companion paper, [AI-Era Vulnerability Disclosure: Coordinated Patch Velocity Frameworks](#) [10], addresses coordination of disclosure and patching at AI-era volumes.

[Project Glasswing: AI Discovery Outpaces Open Source Patching Capacity](#) [6] describes the remediation side of the same imbalance, in which discovery at machine speed meets volunteer-speed patching. This note extends that analysis to the intake channel, where the strain appears before any patch is written. [Enterprise Patch Velocity Under AI-Accelerated Vulnerability Discovery](#) [11] covers the enterprise-side pressure referenced above, and the whitepaper [The NVD Infrastructure Crisis: AI Discovery Overwhelms Tracking](#) [7] covers the downstream database layer and explains why enterprises cannot assume that enrichment will fill the gap left by slow maintainers.

For control-level guidance, the [AI Controls Matrix \(AICM\) v1.1](#) [8] includes controls in its Threat and Vulnerability Management and Application and Interface Security domains that organizations can use to formalize how they ingest, validate, and prioritize third-party and open-source vulnerability information.

Teams that run AI-based discovery programs can also apply the AICM's accountability-oriented controls to define who owns the validation of AI-generated findings.

References

- [1] The Register. "[Curl shuts bug bounty program to remove incentive for submitting AI slop.](#)" The Register, January 21, 2026.
- [2] Help Net Security. "[AI slop submissions force Google to freeze its open-source bug bounty.](#)" Help Net Security, October 5, 2026.
- [3] Security Boulevard. "[Torvalds Offers Guidance as AI Bug Reports Clog Up Linux Security Workflow.](#)" Security Boulevard, May 18, 2026.
- [4] Privacy Guides. "[HackerOne Pauses Internet Bug Bounty.](#)" Privacy Guides, April 17, 2026.
- [5] Cloud Security Alliance AI Safety Initiative. "[Reforming Coordinated Vulnerability Disclosure for the Autonomous Bug Hunter Era.](#)" CSA, June 7, 2026.
- [6] Cloud Security Alliance. "[Project Glasswing: AI Discovery Outpaces Open Source Patching Capacity.](#)" CSA, June 7, 2026.
- [7] Cloud Security Alliance AI Safety Initiative. "[The NVD Infrastructure Crisis: AI Discovery Overwhelms Tracking.](#)" CSA, May 4, 2026.
- [8] Cloud Security Alliance. "[AI Controls Matrix \(AICM\) v1.1.](#)" CSA, version 1.1.
- [9] Pinggy. "[curl's Summer of Bliss: Why It Stopped Taking Bug Reports in July 2026.](#)" Pinggy Blog, 2026.
- [10] Cloud Security Alliance. "[AI-Era Vulnerability Disclosure: Coordinated Patch Velocity Frameworks.](#)" CSA, 2026.
- [11] Cloud Security Alliance. "[Enterprise Patch Velocity Under AI-Accelerated Vulnerability Discovery.](#)" CSA, 2026.