

php2shell: An AI-Found PHP Flaw Leaks Early, and Governments Tell Unpatched Sites to Go Offline

AI High Signal Digest

2026-10-22

php2shell: An AI-Found PHP Flaw Leaks Early, and Governments Tell Unpatched Sites to Go Of- fline

By AI High Signal Digest • October 22, 2026

OpenAI disclosed php2shell, a critical PHP vulnerability found with frontier models. It leaked before the planned release, while AI-driven attacks and supply-chain compromises kept arriving. Also: Recursive's Flint and Mirendil make the case for automated AI research, and Anthropic previews Claude Duet.

php2shell: AI-found PHP bugs become a global patch emergency

OpenAI disclosed php2shell, a 14-year-old PHP flaw that it says lets any attacker execute code on “nearly all deployed PHP application[s].” Researchers at Calif, working inside OpenAI's Patch the Planet initiative, used GPT-5.6 Sol to find a version of the bug in August, and OpenAI's own team independently found a more complete version in September [1]. OpenAI estimates PHP runs about 70% of websites. It spent two months coordinating fixes because the patch needed a major change to the PHP engine [2]. The Rundown reports that Wiz found the same bugs independently [3]. Peter Wildeford's summary credits a mix of GPT-5.6 Sol, Gemini 4 Argon and human expertise. He says an unauthenticated attacker can run code or read server memory, and that the trigger is corrupted state left behind after an out-of-memory crash, which a crafted upload can force [4].

The disclosure did not go as planned. Greg Brockman says the release was set for Oct 22, but details leaked, so PHP's maintainers moved it up to Oct 21 [1]. The NSA told organizations that cannot update immediately to take their systems

offline [5]. One widely shared roundup says the FBI, CISA, the UK, CERT-EU and Japan’s prime minister repeated that advice. It cites Censys counting about 1.5 million servers still running unfixed versions, and lists government sites, Wikipedia and Archive.org among those that went down [6]. That account is secondhand. OpenAI published a Codex skill that checks whether a site is affected [7].

Brockman says the classic process of managing one bug at a time was “slow and manual” and “won’t scale for the agentic era.” He floated the idea that authorized defenders, such as governments, could scan for vulnerable hosts and, in extreme cases, patch them remotely [1]. He also mentioned a Node.js remote-code-execution bug announced the same day [1]. One researcher credits OpenAI’s Daybreak access to GPT-5.6 Sol for a string of recent findings, including GitLab RCEs, a Chrome full chain and a WordPress RCE [8]. Nate Soares (So8res) drew the opposite lesson. He sees a bug of this size surfacing only now, six months after Mythos, as “a heavy strike against the ‘defense will run ahead of offense’ theory” [9].

Attacks are also arriving. - **Storm-0940:** Microsoft says this China state-aligned actor used ARTEX, an agentic AI pen-testing framework, to automate attacks on hundreds of organizations. Microsoft also attributes the Connect-Wise supply-chain compromise to the same group [10]. - **Japan:** 18 companies have disclosed breaches since Sept 25, including up to 22.5M Sourcexnext records. JPCERT/CC found the attackers used AI to scan for known vulnerabilities and to attack APIs, often by pulling endpoints and keys out of the companies’ Android apps [11]. - **CrowdStrike:** a roundup says CrowdStrike documented four successful AI-agent campaigns, one of which hit at least 15 Argentine health-care organizations [12]. - **LiteLLM:** JFrog found malicious code in the LiteLLM 1.107.0 PyPI release [13].

Sarah Guo (@saranormous) says CVEs have gone from about 4k a month to about 15k since April [14]. She adds that HackerOne customers halved their fix times over two years, yet their backlogs grew 131% [15].

Automated AI research ships real models

Recursive released Flint, its first model, with open weights. It is a local coding model built with the Eureka Machine, Recursive’s automated research system, in under four months. Recursive says it runs at 125+ tokens/sec on a MacBook Pro [16]. According to Recursive, the system contributed a sparser model shape, custom kernels and low-bit training that shrank the model by 71% with almost no quality loss [17]. vLLM gives the size as 40B total parameters with 2.5B active, and reports 500 TPS on a single GB200 at 100K context [18]. Cline reports 54.9% on Terminal-Bench 2.1, ahead of every other small MoE, running in about 24GB [19].

Mirendil says one researcher built Mirendil-Base, a pretrained coding model in the 100B–500B parameter range, in 3.5 months for under \$10M in compute,

research included. It says the recipe holds at larger sizes [20]. Independent evaluations and API access have not been released yet [20]. Read together, these are early, mostly self-reported evidence that AI-assisted research can cut both team size and time.

Anthropic: proactive agents and a versioned API

Anthropic opened a research preview of Claude Duet, an app where Claude brings you what needs attention, keeps working while you are away, and checks with you before it acts [21]. It runs on Opus 5.5 [22] and is available through a waitlist for US Pro and Max users. Anthropic cautions that it can misjudge what is urgent [23]. Stripe says Duet uses Link’s agent wallet and Stripe MCP to handle business tasks [24].

Separately, Anthropic released API version 2026-10-21, its first official version since 2023. It plans to version more often, and 2023-06-01 is not deprecated [25]. The new version removes temperature, top_p and top_k from the Messages API [26].

DeepSeek follow-through

Last cycle’s DeepInfer release drew a qualifier from SemiAnalysis. It says the code supports only offline, single-turn inference, so it cannot be benchmarked on InferenceX [27]. SemiAnalysis also estimates that at 60% utilization a GB300 earns \$50/GPU-hour in profit at OpenRouter prices, compared with DeepSeek’s \$39 off-peak [28]. DeepSeek (“the Whale”) reportedly raised another \$15B [29]. teortaxesTex puts the total at RMB 150B over four months [30].

Inference engineering is moving fast more broadly. SemiAnalysis counts three megakernel decode advances in one month: Inferact on TPUs, AMD, and NVIDIA on GB300 [31]. NVIDIA’s trace cuts one Kimi-K3 layer block from 130 to 69 microseconds [32].

Briefly

- **OpenAI on Google Cloud:** OpenAI’s frontier models and Codex workflows are coming to Google Cloud’s Model Garden [33].
- **Flappy Airplanes:** raised a \$1B Series A, co-led by Index and Lux, to build new GPU primitives [34].
- **Kling 4.0:** the API is rolling out [35]. fal hosts the model too, claiming better stability on complex motion [36].
- **Perplexity JOLT:** in this post-training method, one model acts as both teacher and student. On MATH it matched GRPO with $6.5\times$ fewer training tokens [37].
- **Vals Finance Index:** Opus 5.5 leads at 65.6%, with Gemini 4 Argon at 64.9%, within the margin of error. Even the leaders fail more than half of the tax questions [38].

- **Google contracts:** Google employees told the BBC that their contracts bar them from publicly criticizing the company, which keeps them from speaking out on AI safety and military uses. Former employee Sofia Liguori: “In principle I’m not allowed to criticise Google for the duration of my life” [39].

Sources

1. X article by @gdb
2. X post by @OpenAI
3. X post by @TheRunDownAI
4. X post by @peterwildeford
5. X post by @NSACyber
6. X post by @fiveoutofnine
7. X post by @OpenAI
8. X post by @S1r1u5_
9. X post by @So8res
10. X post by @MsftSecIntel
11. X post by @SemiAnalysis_
12. X post by @kimmonismus
13. X post by @JFrogSecurity
14. X post by @saranormous
15. X post by @saranormous
16. X post by @RichardSocher
17. X post by @RichardSocher
18. X post by @vllm_project
19. X post by @cline
20. X post by @bneyshabur
21. X post by @claudeai
22. X post by @TheRunDownAI
23. X post by @claudeai
24. X post by @emilygsands
25. X post by @ClaudeDevs
26. X post by @ClaudeDevs
27. X post by @SemiAnalysis_
28. X post by @SemiAnalysis_
29. X post by @AndrewCurran_
30. X post by @teortaxesTex
31. X post by @SemiAnalysis_
32. X post by @SemiAnalysis_
33. X post by @OpenAI
34. X post by @flappyairplanes
35. X post by @Kling_ai
36. X post by @fal

37. X post by @perplexity__ai
38. X post by @ValsAI
39. X post by @BlackHC