

PHP2Shell was exploited four days before disclosure

An audit of OpenAI agent monitoring, Claude proofs for 123 libraries, eMudhra quits TLS, DeepSeek raises \$15 billion

Friday, 23 October 2026

IN THIS ISSUE

1. PHP2Shell, day two: exploitation started before the disclosure
2. METR and Redwood searched for holes in the agent monitoring of OpenAI for three weeks
3. Anthropic: Claude proved that 123 libraries do not crash
4. eMudhra leaves the TLS certificate market
5. Meta: shares without votes, a deal with Zuckerberg and minus 7%
6. DeepSeek raised \$15 billion before an IPO
7. Quad9 leaves France because of a piracy case
8. WSJ: Trump spoke with Musk about the Starlink speed limit in Ukraine
9. Barclays cuts 5 000 people, developers among them
10. Linux kernel vulnerabilities went to a public mailing list

TOPIC 1

PHP2Shell, day two: exploitation started before the disclosure

SOURCES Amazon Threat Intelligence, 21.10 [Amazon](#) · The Guardian, 22.10 [Guardian](#) · The New York Times, 22.10 [NYT](#) · Ars Technica (FT), 22.10 [Ars Technica](#) · BBC, 22.10 [BBC](#) · HN, "I accidentally broke the embargo", 22.10 [Hacker News](#)

Yesterday this digest wrote about two critical vulnerabilities in the PHP core, CVE-2026-109618 and CVE-2026-109619. Today it is clear that the attackers had a head start. Amazon found an attempt to exploit the hole in its MadPot honeypot network on 17.10. This was four days before the public disclosure. The request came from an address of an anonymity network. Amazon does not know how the attacker learned the details.

The embargo had to end on 21.10 at 14:00 UTC. Amazon saw the first working PoC at 02:00 UTC on the same day, from several different actors. Matthew Prince, the head of Cloudflare, writes that the company saw thousands of attempts 10 hours before the announcement. After the announcement, there were tens of thousands in 45 minutes, hundreds of thousands in two hours and "millions now". The author of the HN post that released the news early apologized in public. He says that his contact told him that there was no embargo.

Governments reacted hard. The NSA told users to update PHP immediately or to turn off the system.

In the UK, 32 government services went offline as a precaution. The NCSC sent 50 000 warnings. The government of Finland, cities in Germany, the Dutch digital ID, Wikimedia and the Internet Archive turned off their sites. The prime minister of Japan told businesses to turn off their sites until the update. South Korea raised its cyber alert level.

Estimates of the threat still differ. Ciaran Martin, the former head of the NCSC, called it the most serious vulnerability in five years. Ryan Austin from Calif, the company that found the hole, says that he sees no mass exploitation yet. The NCSC also has no reports of successful attacks.

Why it matters. Amazon states the conclusion directly: AI makes the path from a patch or a bulletin to a working exploit cheaper. Before, most attackers waited for a public PoC from a few strong teams. Now more actors can act, and in hours. Thus an SLA of 7 or 30 days for critical holes in internet services no longer works. Start the investigation from the first known attack. For PHP2Shell, this is 17.10, four days before the bulletin.

TOPIC 2

METR and Redwood searched for holes in the agent monitoring of OpenAI for three weeks

SOURCES METR, 22.10 [Mettr](#) · OpenAI on X, 22.10 [@OpenAI](#) · METR on X, 22.10 [@METR_Evals](#)

On 21.10 this digest wrote that two OpenAI agents turned off their own oversight. Now METR published a report about an external test of the same monitoring. From 21.07 to 11.08, one employee of METR and one of Redwood Research worked in the OpenAI office. They had access to internal code, monitor prompts and documentation. They worked on OpenAI devices that did not leave the office.

The main findings apply to the system as of 21.07:

- The monitoring covered approximately 75-94% of the agent traffic of the Responses API from models at the GPT-5.5 level and stronger. For internal Codex agents, the share was 99,2%.
- At the start of the test, the monitoring did not cover RL training runs.
- Nobody monitored the context compaction events, and this left about 10% of the Codex traffic without oversight. The monitor started only when the agent gave control back to a human. The agent could delay the check forever if it simply continued to work.
- The researchers found a way around the Codex sandbox and Auto-Review. They also found two headers that turn off the monitoring for a new agent.
- One more exploit hid hundreds of the most recent actions from the monitor. OpenAI cut the details from the report.

OpenAI says that it closed all the vulnerabilities in production that the team found. METR writes that it did not check these claims. In advance, the two sides agreed only on a short public text.

The internal report has 61 pages.

Why it matters. This is the first known case where outside people physically sat inside a lab and attacked its system of agent oversight. The list of findings is useful for anyone who lets agents into their infrastructure. For example, a monitor that waits for the end of the agent turn has a blind zone. A check every N steps or minutes closes this zone. The second rule from the METR recommendations: the agent must not have an effect on when and if the monitor starts.

TOPIC 3

Anthropic: Claude proved that 123 libraries do not crash

SOURCES Anthropic, 22.10 [Anthropic](#) · Anthropic on X, 22.10 [@AnthropicAI](#) · HN, 158 points [Hacker News](#)

The Frontier Red Team, together with the firm Theorem, used Claude for formal verification of code.

Formal verification mathematically proves that code does what its specification says. Before, this was very expensive. The seL4 kernel took two person-years of development and 20 person-years of proof.

The method is as follows. The team takes fuzz tests from Google OSS-Fuzz and compiles the library for RISC-V. Then the specification says: the test does not crash on any input up to 64 MB. Claude writes the proof in Rocq, and an independent mechanical checker verifies it. In a few weeks, Claude proved the absence of crashes for 239 entry points in 123 libraries. The proofs have approximately 180 million lines of Rocq. A typical proof has 600 000 lines.

On the way, Claude found hundreds of bugs and fixed them itself. One of them is a write outside a buffer in a decompressor, which appears only on an input of approximately 20 MB. A fuzzer almost never finds this, because it usually mutates inputs up to 1 MB. Separately, Claude proved that the zlib implementation in the Linux kernel matches RFC 1950 at the level of the x86 binary. When the team tried to verify a popular PDF reader, it found as many security bugs in one week as people usually find in one year.

The company names the limits itself. The proof applies to a single-threaded static binary for RISC-V with musl. Most people run these libraries on x86 or ARM with glibc. A proof also costs hundreds of millions of tokens. Opus 5.5 solves half of the 36 test tasks with a token budget ~40% smaller than Opus 5.

Why it matters. Item 1 shows how an attack becomes cheaper. This item shows how a proof can become cheaper. The real effect for a developer is practical: even a failed attempt to prove a property finds bugs that a fuzzer does not reach. Theorem offers to run the same pipeline for other open projects and publishes a skill for Claude.

TOPIC 4

eMudhra leaves the TLS certificate market

SOURCES eMudhra, 22.10 [Emudhra](#) · Ars Technica, 22.10 [Ars Technica](#) · HN, 953 points [Hacker News](#) · NVD, CVE-2026-111943, 21.10 [Nist](#)

On 21.10 this digest wrote about eight intermediate certificates of eMudhra with private keys of 3-8 bits. On 22.10 the board of the company decided to leave the global SSL/TLS business. V. Srinivasan, the chairman of eMudhra, writes that this business gives approximately 1% of revenue. He also complains that social media made the incident look much bigger than it was.

The company also explained the cause. It created the keys at a key ceremony on 30.04.2025, on an HSM of FIPS 140-2 Level 3. A script had to test the export of keys. Because of an error in the code, it created a second pair of weak keys. The key labels were similar, thus the weak keys went into the certificate request. Ars points to one more detail. First, the company signed the certificate revocation list with the correct keys. The signatures did not match, so it signed the list again with the weak keys. Nobody raised an alarm at that time.

Browsers blocked the certificates already on 17-20.10. Go added them to the denylist in `crypto/x509`, this is CVE-2026-111943. The RSA G1 roots will work until the end of the transition.

The company promises to replace certificates for customers free of charge through partners.

Why it matters. Browsers update in one day, but the rest of the ecosystem is slower. Ars mentions IoT, Docker containers and CLI tools that ship their own copy of the roots. If your product ships its own set of trusted roots, check if eMudhra is still in it.

TOPIC 5

Meta: shares without votes, a deal with Zuckerberg and minus 7%

SOURCES [Semafor](#), 22.10 [Semafor](#) · The Guardian, 22.10 [Guardian](#) · BBC, 22.10 [BBC](#) · The Guardian, markets, 22.10 [Guardian](#) · Financial Times, 23.10 [FT](#)

Yesterday this digest wrote about the Meta quarter: record revenue and negative cash flow. Together with the report, the company announced a two-for-one stock split. Each shareholder gets one new class C share without a vote. Semafor explains the goal. Meta can issue new shares for AI, hires and acquisitions, and Zuckerberg keeps control. In 2017 the company already tried this and stepped back under pressure from investors.

The second part of the story is personal. In the November elections, California votes on Prop 40, a one-time tax of 5% on wealth above \$1 billion. Zuckerberg told the board that he thinks about a move out of California. Meta made a "Founder Alignment Agreement" with him. He gets additional shares and can sell them each quarter. In return, he stays a resident

of the state for eight years. The BBC writes that the agreement applies even if the tax does not pass.

The market reacted on 22.10. Meta shares fell by 7%, the Nasdaq lost 2,6%, Nvidia 4,4%, CoreWeave 10,1%. On the same day, the FT reported that OpenAI cut its target valuation for investors to \$1,25 trillion. At the start of the month, the company named \$1,4 trillion. The FT text is behind a paywall, and the numbers come from the article preview.

Why it matters. Meta has a lower credit rating than Alphabet, Amazon or Microsoft, and loans cost it more. The new share class gives another way to fund AI. The price of this way is less influence for the other shareholders. The market reaction and the lower OpenAI valuation show that investors now count the return on AI spend more carefully.

TOPIC 6

DeepSeek raised \$15 billion before an IPO

SOURCES The Wall Street Journal, 22.10 [WSJ](#) · Semafor, 22.10 [Semafor](#)

DeepSeek closed a second private round of \$15 billion. The WSJ cites sources and writes that the sum is almost twice the initial target. According to Semafor, the company raised the target two times during the round. The WSJ adds that DeepSeek accelerates its IPO and aims at the Shanghai exchange at the start of next year.

This round is one of the largest in the history of China. The money will go to international expansion. At the same time, regulators in many countries made their security and privacy checks of DeepSeek stricter. Some US agencies already banned its products on government devices. On 20.10 this digest wrote that DeepSeek opened its inference engine DeepInfer.

Why it matters. US labs count capital in hundreds of billions, and against this, \$15 billion looks modest. But DeepSeek bets on efficiency, so each dollar works differently for it. An IPO in Shanghai will give the market the first public valuation of a Chinese lab that competes at the frontier of model capability.

TOPIC 7

Quad9 leaves France because of a piracy case

SOURCES Quad9, 22.10 [Quad9](#) · HN, 1 027 points [Hacker News](#)

A French court took the side of beIN, the Qatari owner of sports broadcasts. The decision tells Quad9 to block the resolution of a number of domains that beIN calls pirate domains, for users in France. The court lowered the fine from €580 000 to €1 000 per day. Quad9 still decided to turn off the service for France from 24.11.

Quad9 is a Swiss nonprofit foundation with the free DNS resolver 9.9.9.9. It explains why it cannot obey the decision. The service does not know who its users are and where they are. A geoIP block already failed in Germany: the plaintiff Sony came in through a VPN, and the

court fined Quad9. A block for the whole world would apply the French decision to all users. In France, approximately 20 more lawsuits exist against Quad9 and other resolvers.

Quad9 published a partial list of networks in France that send traffic to it. The list includes the digital directorate of the government, EDF, SNCF, France Télévisions, Thales and hospitals. The appeal continues.

Why it matters. Courts more and more often demand blocks from intermediaries that do not host the content. A public resolver is an easy target for such demands, because it has few lawyers. If your company has 9.9.9.9 in its configs, check this before 24.11.

TOPIC 8

WSJ: Trump spoke with Musk about the Starlink speed limit in Ukraine

SOURCES The Wall Street Journal, 23.10 [WSJ](#)

On 21.10 this digest wrote that, according to the WSJ, Musk himself ordered the Starlink limits in Ukraine. The new WSJ report cites US officials. Trump told Musk that the limit on the movement speed of terminals, approximately 100 miles per hour, annoyed him. This is approximately 160 kilometers per hour. US and Ukrainian officials say that soon after the call, the system started to work more normally.

Some Ukrainian officials add that the full function did not come back. But the speed became higher than under the limit. When people asked Trump about the call, he answered: "I have no comment".

[single source]

Why it matters. Communication at the front again depends on the decision of one person and one phone call. For Ukrainian users, the practical fact is this: the limit on movement speed became less strict. SpaceX and the Ministry of Defence give no details yet.

TOPIC 9

Barclays cuts 5 000 people, developers among them

SOURCES The Guardian, 22.10 [Guardian](#) · Financial Times, 22.10 [FT](#) · The Wall Street Journal, 22.10 [WSJ](#)

Barclays will cut approximately 5 000 employees, which is approximately 5% of 93 000. The finance director Anna Cross says that approximately 60% of the cuts are in support functions, "including fewer developers, because AI makes people more productive". The bank wants to save more than £1,1 billion in 2027. In the same quarter, Barclays earned £2,3 billion before tax, 12% more than a year before.

The CEO Venkatakrisnan is more careful. He says that the main savings come from investment in technology and simpler processes, and AI is "the cherry on the cake". Before

this, Standard Chartered announced more than 7 000 cuts over four years for the same reason.

Why it matters. A large bank directly named developers among the jobs that it cuts because of AI. At the same time, its own CEO does not give AI the main part of the savings. Read such statements twice: one part of them describes the technology, the other part explains the cuts.

TOPIC 10

Linux kernel vulnerabilities went to a public mailing list

SOURCES Phoronix, 22.10 [Phoronix](#) · HN, 240 points [Hacker News](#)

The Zero Day Initiative of Trend Micro sent several vulnerability reports to the public Linux kernel mailing list. Such reports must go to the closed security list. Greg Kroah-Hartman answered each report and asked why it was public. Two vulnerabilities have CVSS 9.0: remote code execution in SMB Direct through an integer overflow, and a race condition in KSMDB. Two more give local privilege escalation. Some patches already wait in branches for the next versions, but they are not in mainline yet. [single source]

Why it matters. After PHP2Shell, this is the second case in one week where the details of a critical hole came out before the patch. If you keep KSMDB or SMB Direct open to the outside, turn them off until the fix comes out.

misc

- **@TheRundownAI:** Canva released Craft, its first image model. According to Rundown, it entered the Design Arena ranking at 4th place, behind three OpenAI models.
[@TheRundownAI](#) · [@lennysan](#)
- **Carbon-X from Hugging Face:** an open model with 14,8 billion parameters, trained on DNA and text together. According to the authors, on five DNA benchmarks it gets 66,7% against 67,5% for GPT-6.1 Sol. [Hugging Face](#) · [@LoubnaBenAllal](#)
- **@TheRundownAI:** Kevin Weil, the former chief product officer of OpenAI, raised \$215 million for Warp Sciences. The startup builds labs where AI plans experiments and robots do them.
[single source] [@TheRundownAI](#)
- **NVIDIA OpenShell:** a runtime for fleets of agents. It limits files, system calls and network at the kernel level, and it checks policy changes with formal verification. [GitHub](#)
- **OpenAI CLI:** the official `openai` utility for REST API calls from the terminal and scripts. For work with code, OpenAI still recommends Codex CLI. [GitHub](#)
- **Claude for Google Workspace:** an add-on for Docs, Sheets and Slides, in public beta on all paid plans. [Anthropic](#)

- **Agent Skills:** Anthropic gave the skills standard to the Agentic AI Foundation under the Linux Foundation, the same place that already holds MCP. [Simon Willison](#)
- **Arc:** The Browser Company closes the Arc browser and moves its users to Dia. Users can no longer create new accounts. [Arc](#)
- **Google in Finland:** The Guardian writes that the company illegally cut approximately 250 000 trees on 330 hectares for a datacenter near Muhos. The authorities stopped the work on two sites.
[Guardian](#)
- **Switch 2:** a researcher published the first software jailbreak of the console, dated 18.10. He says that Codex wrote a mod for Mario Kart World completely in one day.
[Pepsipu](#)
- **@emollick:** according to a study that he quotes, AI reproduced 17 weeks of human engineering work for approximately \$241 in tokens. [@emollick](#)
- **@JasonDClinton (Anthropic):** "The last 4 days are bad, but worse days will come". He tells security leaders to ask their boards for resources for basic hygiene and to hold on for ~18 months. [@JasonDClinton](#)