

The vulnerability didn't change. **The attacker did.**

Patch deferrals assumed exploits took skill and weeks. Anthropic's 2026 evaluations show a model doing it in under an hour.

Every deferral assumes a slow attacker

- 01 **Legacy system.** Nobody will bother before the replacement lands
 - 02 **Change freeze.** The window is short next to attacker effort
 - 03 **Vendor rating.** "Exploitation less likely" will hold
 - 04 **Internal only.** We will catch them getting in
 - 05 **Rollout takes weeks.** Exploits take longer
-

21 of 41

patched Chrome V8 flaws turned into full code execution by one model

Anthropic Frontier Red Team. No other model in the same test achieved one. Anthropic expects this capability to be widely available within 6 to 12 months.

<1 hour

from public Firefox patches to the first working exploit

Anthropic, Measuring LLMs' impact on N-day exploits.
Eight working exploits in about 12 hours. Windows privilege escalations at about \$2,000 each.

Exploits finish before the fleet is patched

31 min First Windows proof of concept

6 hrs All 18 proofs of concept

Day 7 Typical fleet 90% patched

Day 43 Median fix for known-exploited flaws

Anthropic, June 2026; Verizon DBIR 2026



Multiple threat actors have effectively established **automated exploit foundries** with AI.

Anthropic threat intelligence report. One breach went from a stolen developer token to full cloud admin control in about three hours.

13 of 14

Windows flaws rated "exploitation unlikely" or "less likely" got a working proof of concept

A risk acceptance that rests on the vendor's rating is resting on a judgement about human effort.

Deadlines set by exposure and exploitability

3 days Known exploited, total control

Plus forensic triage

14 days Known exploited, lower impact or not automatable

Up to 60 Exposed, not yet on the KEV list

3, 14 or 60 days by automation and impact. Issued June 2026; revokes BOD 22-01.

Six changes to the risk register

- 01 Rescore likelihood on exploitability, not attacker effort
 - 02 Give every acceptance an expiry set by exposure
 - 03 Reopen acceptances automatically on triggers
 - 04 Test compensating controls with evidence
 - 05 Measure patch deployment in hours
 - 06 Test your own backlog with the same tools
-

Rescore every accepted risk **this** **quarter**

The bugs in the register are the same bugs. The dates next to them have to change.



Terence Kok · AI & Innovation Strategist

SAVE · COMMENT · FOLLOW