CYBERSECURITY INTELLIGENCE BRIEF · APRIL 2026

AI-Powered Vulnerability Discovery: Claude & Mozilla Firefox

22	14	~20%	271	181×
CVEs issued (Opus 4.6)	High-severity bugs fixed	Of 2025 Firefox criticals	Vulns patched (Mythos FF150)	More exploits vs. Opus 4.6

BACKGROUND & PARTNERSHIP

In late 2025, Anthropic's Frontier Red Team deployed Claude Opus 4.6 against Mozilla Firefox's sprawling C++ codebase as a controlled security evaluation. The two-week collaboration — initiated after Claude autonomously flagged a use-after-free vulnerability in Firefox's JavaScript engine just 20 minutes into analysis — marked the beginning of a new paradigm in software security research.

PHASE 1: CLAUDE OPUS 4.6 + FIREFOX 148

Over two weeks in January 2026, Claude Opus 4.6 scanned nearly 6,000 C++ files and submitted 112 unique vulnerability reports to Mozilla. Of these, 22 were formally issued as CVEs — surpassing the number reported from any single source in any month of 2025. Mozilla fixed all high- and moderate-severity issues in Firefox 148 (released February 2026).

- 14 high-severity CVEs — ~20% of all Firefox high-severity bugs patched in 2025
- 7 moderate-severity CVEs; 1 low-severity
- First bug found in just 20 minutes; 50+ crashes surfaced before human review
- Novel logic errors detected that traditional fuzzers completely missed
- Claude generated proposed patches alongside each vulnerability report

EXPLOITATION CAPABILITY TESTING

Anthropic tasked Opus 4.6 with developing working exploits from its findings — spending roughly $4,000 USD in API credits across several hundred attempts. The model succeeded in only two cases, both "crude" exploits requiring sandboxing to be disabled. The key asymmetry revealed:

- Discovery cost is an order of magnitude lower than exploit development
- Defenders gain a meaningful time advantage for patch prioritization
- One successful exploit targeted CVE-2026-2796 (CVSS 9.8) — a JIT miscompilation in Firefox's WebAssembly component

PHASE 2: CLAUDE MYTHOS PREVIEW + FIREFOX 150

In April 2026, Anthropic unveiled Claude Mythos Preview — a frontier model so capable it is being withheld from general public release. Mozilla received early access and applied it to Firefox. The results dwarfed Phase 1:

- 271 vulnerabilities identified and fixed in Firefox 150 (April 2026)
- Mythos produced 181 working Firefox exploits vs. only 2 for Opus 4.6 — a 90× leap
- Chains of 4+ vulnerabilities assembled into full sandbox-escape exploits autonomously
- Zero-days found across every major OS and web browser
- Uncovered a 27-year-old OpenBSD flaw and a 16-year-old FFmpeg vulnerability missed by every prior automated tool

PROJECT GLASSWING — DEFENSIVE COALITION

Recognizing the existential dual-use risk, Anthropic launched Project Glasswing — a restricted consortium granting Mythos Preview access to 40+ critical-infrastructure organizations including AWS, Apple, Broadcom, Cisco, CrowdStrike, Google, JPMorgan Chase, Linux Foundation, Microsoft, NVIDIA, and Palo Alto Networks. Anthropic committed up to $100 million in API credits and $4 million in direct donations to open-source security organizations.

KEY IMPLICATIONS FOR AI IN CYBERSECURITY

- Discovery cost collapse: AI slashes the resource threshold for finding critical bugs — once gatekept by elite researchers
- Defender-first window: AI currently finds far more than it can exploit; asymmetry favors rapid patching
- Patch velocity imperative: Security teams must orient toward 24–48 hour remediation cycles as AI-speed discovery becomes the norm
- Dual-use risk is real: The same capabilities that harden software can be weaponized; access governance is as critical as the models themselves
- Open source exposure: Legacy C/C++ codebases carry decades of latent vulnerabilities now surfaceable at machine speed — supply-chain audits are urgent
- Regulatory horizon: EU AI Act cybersecurity rules (effective Aug 2026) increase compliance burden for AI-assisted security tooling

Jan 2026	Feb 2026	Mar 2026	Apr 7, 2026	Apr 21, 2026
Opus 4.6 begins Firefox analysis	22 CVEs filed; Firefox 148 ships	Public disclosure; Claude Code Security	Mythos & Project Glasswing announced	Firefox 150 ships; 271 Mythos fixes

1. The Hacker News — Anthropic Finds 22 Firefox Vulnerabilities, thehackernews.com/2026/03/anthropic-finds-22-firefox.html
2. Security Affairs — Claude Opus AI discovers 22 Firefox bugs, securityaffairs.com/189131/ai/... 3. InfoQ — Claude AI Firefox Vulnerability, infoq.com/news/2026/03/claude-ai-firefox-vulnerability/
4. The Hacker News — Claude Mythos Finds Thousands of Zero-Day Flaws, thehackernews.com/2026/04/anthropics-claude-mythos-finds.html 5. Mozilla Blog / Reddit r/linux — Firefox 150 / 271 Vulnerabilities, reddit.com/r/linux/1srxt82