

OpenAI Daybreak vs. Anthropic Mythos: Two Rival Visions for AI Cybersecurity

Kabui, Charles

2026-05-22

[Read at ToKnow.ai](#)

OpenAI Daybreak vs. Anthropic Mythos: Two Rival Visions for AI Cyber
Defense-first platform vs. restricted exploit-capable model

- 3 Tiers**
Daybreak access levels from default to red team
- 181**
Working exploits by Mythos vs near-zero for Opus 4.6
- EU Blocked**
Mythos restricted to US partners and allies only

May 22, 2026

ToKnow.ai

OpenAI launched [Daybreak](#), a cybersecurity platform built on GPT-5.5 with three access tiers: a default mode, a [Trusted Access for Cyber](#) tier for verified defenders doing code review and malware analysis, and GPT-5.5-Cyber for authorized red teaming. Daybreak runs through Codex Security, which builds threat models from repositories, traces attack paths, and generates patches in isolated environments. OpenAI committed \$10M in API credits and

signed Cloudflare, CrowdStrike, Palo Alto Networks, and Cisco as partners. The philosophy is defense-first: find vulnerabilities, validate fixes, and build security into software from the start.

Anthropic's [Claude Mythos Preview](#), announced April 7 with [Project Glasswing](#), took the opposite path. Mythos autonomously found thousands of zero-day vulnerabilities across every major OS and browser, including a 27-year-old OpenBSD bug and a FreeBSD remote code execution flaw it exploited without human guidance. Its exploit success rate jumped from near-zero (Opus 4.6) to 181 working exploits on patched Firefox bugs. But Anthropic restricted Mythos to a 12-company coalition and [blocked EU access](#), pushing European banks toward Mistral for alternatives. Security researchers at [AISLE showed](#) that small open-weight models replicated much of Mythos's showcase analysis, arguing "the moat is the system, not the model."

The split reflects a core question: should AI cybersecurity be packaged as accessible enterprise products with guardrails (Daybreak), or gated because the capabilities are too dangerous to distribute widely (Mythos)? Both assume AI will reshape offense and defense. They disagree on who gets to use it.

For background on Anthropic's earlier security moves, see [Claude Security Enters Public Beta, Scans Code for Vulnerabilities with Opus 4.7](#).

Sources:

- [OpenAI Daybreak](#)
- [Scaling Trusted Access for Cyber with GPT-5.5 and GPT-5.5-Cyber \(OpenAI\)](#)
- [Assessing Claude Mythos Preview's Cybersecurity Capabilities \(Anthropic Red Team\)](#)
- [Project Glasswing: Securing Critical Software for the AI Era \(Anthropic\)](#)
- [AI Cybersecurity After Mythos: The Jagged Frontier \(AISLE\)](#)

Disclaimer: For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. Read more: [/terms-of-service](#)