Here's how cyber heavyweights in the US and UK are dealing with Claude Mythos

Two reports from former high-level U.S. cyber officials and the UK government's top AI research institution reveal how top defenders think about the tool's hacking capabilities.

April 13, 2026 By: Derek B. Johnson Cyberscoop

https://cyberscoop.com/claude-mythos-ai-cybersecurity-threat-report/

A joint report from the Cloud Security Alliance (CSA), the SANS Institute and the Open Worldwide Application Security Project (OWASP) concludes that in the near term, organizations are "likely to be overwhelmed" by threat actors using AI to find and exploit vulnerabilities faster than defenders can patch them.

While those organizations can use AI tools to speed up their own defenses, attackers "still face a heavier relative burden due to the inherent limitations of patching. This in turn leads to "asymmetric benefits" for attackers who can afford to adopt the technology without the same caution and bureaucracy as a multi-billion dollar business.

"The cost and capability floor to exploit discovery is dropping, the time between disclosure and weaponization is compressing toward zero, and capabilities that previously required nation-state resources are now becoming broadly accessible," wrote Robert Lee, SANS Institute's Chief AI Officer, Gadi Evron, CEO of Knostic and Rich Mogull, chief analyst at CSA, who served as the primary authors.

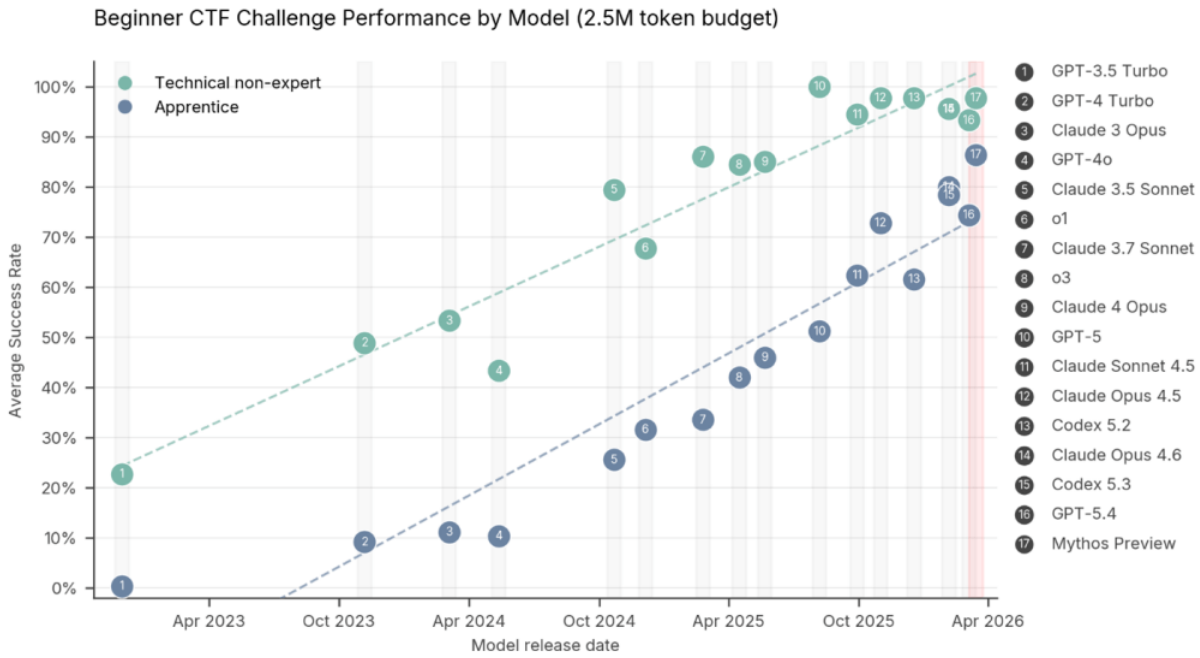
The report marks one of the first comprehensive responses to the capabilities of Claude Mythos from the U.S., boasting cybersecurity luminaries who have set policy at the highest levels as contributing authors, including Jen Easterly, former director of the Cybersecurity and Infrastructure Security Agency, Rob Joyce, a former top White

House and NSA cybersecurity official, and Chris Inglis, former National Cyber Director.

It also includes private sector stalwarts like Heather Adkins, Google's CISO, Katie Moussouris, CEO of Luta Security, and Sounil Yu, chief technology officer at Knostic. Another seventy CISOs, CTOs and other security executives are named as editors and reviewers.

Also this week, the UK's AI Security Institute (AISI) detailed the results of tests it performed on a preview version of Claude Mythos, calling it a "step up" from past Anthropic models in the cybersecurity arena and able to "execute multi-stage attacks on vulnerable networks and discover and exploit vulnerabilities autonomously."

Using a mix of Capture the Flag exercises and cyber range testing, AISI researchers found that Mythos not only raised the ceiling of technical non-experts and apprentice-level users, it narrowed the overall gap in hacking proficiency between the two. In other words, there's becoming less of a distinction between the capabilities of amateur "script kiddies" and mid-level hackers with technical knowledge.

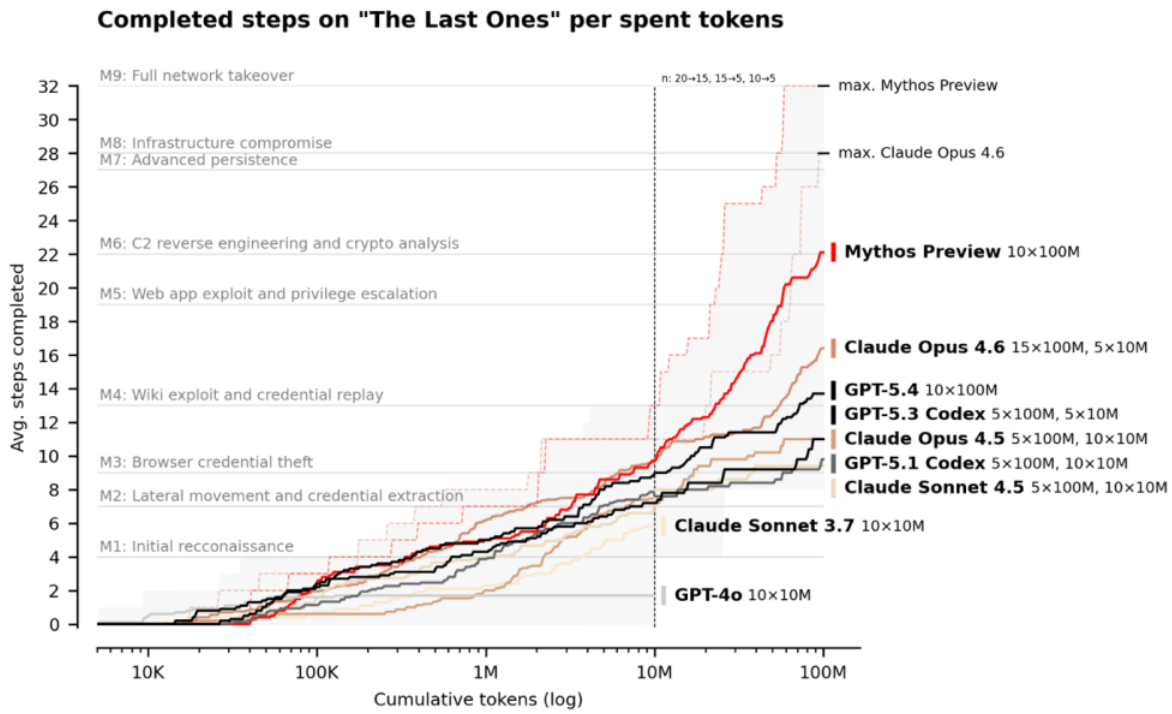


Claude Mythos and other Large Language Models are increasing the capabilities of both lower and mid-level hackers when it comes to solving cybersecurity-specific tasks and challenges. (Source: AISI)

Before April 2025, no Large Language Model could complete a single expert-level CTF problem. Mythos successfully solved nearly three quarters (73%) of them.

In cyber range tests – which are meant to simulate more complex, multi-chain attacks – the results were uneven, but also represented meaningful progress over prior Claude models.

Mythos was subjected to a 32-step attack playbook modeled on corporate networks, spanning initial network access to full network takeover. In three of the 10 simulations, the model completed an average of 24 of the 32 steps. Older versions of Claude and other frontier models never averaged more than 16.



Claude Mythos improved on other models ability to complete a 32 step cyber attack targeting a simulated corporate network environment. (Source: AISI)

Mythos flunked its test against a simulated operational technology cooling tower, but researchers noted that this doesn't mean AI is bad at exploiting OT: the model actually faltered during the IT section of the exercise.

UK researchers were more measured in their analysis of Mythos, noting that their testing indicates it is "at least capable" of autonomously taking down smaller, weakly defended enterprise networks.

But they also note their cyber ranges lack security features – like active defenders and defensive tooling – that would be common in many real-world networks and present additional obstacles, nor did they penalize the model for triggering security alerts.

"This means we cannot say for sure whether Mythos Preview would be able to attack well-defended systems," the researchers concluded.

Technical debt coming due

Both the US and UK reports agree that large language models are broadly moving in a similar direction of lowering the technical barrier. The US authors call for organizations to more quickly adopt AI for cyber defense while overhauling their incident response playbooks and corporate policies to account for more automated defense postures.

For its part, Anthropic has said it is not selling Mythos commercially, and last week it announced the model would be made available to Project Glasswing, a consortium of major tech companies that will use it to root out and patch vulnerabilities in commonly used products and services.

But other experts have warned that businesses and governments are not well-positioned to either absorb the influx of expected vulnerability exploitation or deftly harness AI tools of their own to counter them.

Casey Ellis, CTO and founder of Bugcrowd, wrote that recent advances in AI cyber tools has succeeded largely by "living in the places we stopped looking a decade ago."

While the cybersecurity community has spent years focusing on application security, vulnerability triage and other "top layer" security problems, AI tools and apex level hacking groups have been feasting on vulnerabilities in forgotten firmware, or routers whose manufacturers long went out of business.

This reality that tools like Mythos can endlessly weaponize the massive technical debt of large organizations has taken the traditional defender's dilemma and "the knob that used to go to ten and turned it to seven hundred," Ellis wrote.

Additionally, corporations and governments run on consensus-building, multiple layers of hierarchy and legal compliance. While those are all necessary when handing your cybersecurity over to automated tooling, it can also lead to a slower process and more asymmetry against defenders in the short term.

“Integration into actual production becomes the battlezone,” wrote Ellis. “Lag is real. Bureaucracy is real. Supply chains are real.”