

Announcements

Statement on the US government directive to suspend access to Fable 5 and Mythos 5

Jun 12, 2026

The US government, citing national security authorities, has issued an export control directive to suspend all access to Fable 5 and Mythos 5 by any foreign national, whether inside or outside the United States, including foreign national Anthropic employees. The net effect of this order is that we must abruptly disable Fable 5 and Mythos 5 for all our customers to ensure compliance. **Access to all other Anthropic models will not be affected.**

We received the directive from the government today at 5:21pm (ET). The letter did not provide specific details of its national security concern. Our understanding is that the government believes it has become aware of a method of bypassing, or “jailbreaking” Fable 5. We reviewed a demonstration of this specific technique being used to identify a small number of previously known, minor vulnerabilities. These vulnerabilities all appear relatively simple, and we have found that other publicly-available models are able to discover them as well without requiring a bypass.

Anthropic’s posture with respect to Fable’s safeguards, as laid out in our launch [blog](#)

A

≡

is misused for tasks related to cybersecurity (among others). In fact, our safeguards are so strong that many users have complained that they are overly broad.

- In the weeks leading up to the launch of Fable, Anthropic worked with the US government, the UK AISI, multiple private third-party organizations and internal teams to red-team Fable’s safeguards for thousands of hours in total.
- These tests showed that Fable’s safeguards are substantially more effective than those of any previously deployed model.

- No testers have yet been able to find a *universal jailbreak*—a jailbreak method that can very broadly bypass the model’s safeguards, unblocking a wide range of cyber capabilities.
- We suspect that perfect jailbreak resistance is not currently possible for any model provider. Every safeguard used in the industry is vulnerable to *non-universal jailbreaks* (which can elicit *some* cyber information in specific circumstances), and it is likely that universal jailbreaks will eventually be found in the future. We stated this clearly when we released Fable 5.
- Given that perfect jailbreak resistance does not appear to be possible today, Anthropic adopted a *defense in depth* strategy with Fable 5. We aimed to make jailbreaks either narrow (in the case of non-universal jailbreaks) or very expensive to produce (in the case of universal jailbreaks), and to combine this with thorough monitoring to quickly detect and shut down any successful attacks. This is also why Anthropic has required 30-day retention of customer data with Fable—a policy change that carries real costs for us with customers, but that allows us to research and mitigate jailbreaks.
- We stand by this defense in depth strategy. It reduces the risks posed by Fable, making them comparable to the risks of existing models already deployed across the industry.
- We have not even received a disclosure of a concerning non-universal potential jailbreak that led to a harmful result. The potential jailbreaks that have been disclosed to us are either entirely benign responses or are minor findings that provide no Mythos-specific uplift.

To date, the government has only given us verbal evidence of a potential narrow, non-universal jailbreak, which essentially consists of asking the model to read a specific codebase and fix any software flaws. Our understanding is that one potential jailbreak was shared with the government. We have reviewed a report that we believe is the basis of the government's directive and validated that the level of capability displayed there is widely available from other models (including OpenAI’s GPT-5.5), and is used every day by the defenders who keep systems safe. We will share more details over the next 24 hours.

We are complying with the government’s legal directive and are removing access to Fable 5 and Mythos 5 for all users. However, we disagree that the finding of a narrow potential jailbreak should be cause for recalling a commercial model deployed to hundreds of millions of people. If this standard was applied across the industry, we

believe it would essentially halt all new model deployments for all frontier model providers.

As we have stated publicly, we believe the government should have the ability to block unsafe deployments, as part of a statutory process that is transparent, fair, clear, and grounded in technical facts. This action does not adhere to those principles.

We apologize for this disruption to our customers. We believe this is a misunderstanding and are working to restore access as soon as possible.



Related content

Results from the first Anthropic Public Record

[Read more →](#)

TCS and Anthropic partner to bring Claude to regulated industries

We're announcing a partnership with Tata Consultancy Services (TCS). TCS will provide Claude to 50,000 of its own employees across 56 countries; build Claude-powered products for clients in financial services, healthcare, the public sector, and other regulated industries; and join the Claude Partner Network.

[Read more →](#)

DXC will integrate Claude into the systems banks, airlines, and other regulated industries rely on

We're announcing a multi-year global alliance with DXC Technology, one of the world's largest IT services companies.

[Read more →](#)



Products

Claude

Claude Code

Claude Code Enterprise

Claude Cowork

Claude Security

Claude for Chrome

Claude for Slack

Claude for Microsoft 365

Skills

Download app

Pricing

Log in to Claude

Models

Mythos

Fable

Opus

Sonnet

Haiku

Solutions

AI agents

Code modernization

Coding

Customer support

Education

Enterprise

Financial services

Government

[Healthcare](#)

[Legal](#)

[Life sciences](#)

[Nonprofits](#)

[Security](#)

[Small business](#)

[Startups](#)

Claude Platform

[Overview](#)

[Developer docs](#)

[Pricing](#)

[Marketplace](#)

[Regional compliance](#)

[Claude on AWS](#)

[Google Cloud's Vertex AI](#)

[Microsoft Foundry](#)

[Console login](#)

Resources

[Blog](#)

[Claude partner network](#)

[Community](#)

[Connectors](#)

[Courses](#)

[Customer stories](#)

[Engineering at Anthropic](#)

[Events](#)

[Inside Claude Code](#)

[Inside Claude Cowork](#)

[Inside Claude Enterprise](#)

[Inside Claude Security](#)

[Plugins](#)

[Powered by Claude](#)

[Service partners](#)

[Tutorials](#)

[Use cases](#)

Help and security

[Availability](#)

[Status](#)

[Support center](#)

Company

[Anthropic](#)

[Careers](#)

[Policy](#)

[Economic Futures](#)

[Research](#)

[News](#)

[Claude's Constitution](#)

[Claude Corps](#)

[Policy on the AI Exponential](#)

[Responsible Scaling Policy](#)

[Security and compliance](#)

[Transparency](#)

Terms and policies

[Privacy choices](#)

[Privacy policy](#)

[Consumer health data privacy policy](#)

[Responsible disclosure policy](#)

Terms of service: Commercial

Terms of service: Consumer

Usage policy

© 2026 Anthropic PBC

