The Duluth Lede

Grassroots reporting from Duluth — a harbinger and catalyst for change.

TECHNOLOGY / AI

U.S. Government Forces Anthropic to Pull Its Two Most Powerful AI Models

A national-security export order shut down Fable 5 and Mythos 5 over a narrow software-bug demonstration — even as the company's months-long feud with the administration grinds through the courts.

By Bryan Russell · 2026-06-13

On the evening of June 12, the U.S. government did something it had never done before to a publicly released American AI model: it ordered one switched off. Citing national-security export authorities, the Commerce Department directed Anthropic to suspend all access to its two most capable systems, Fable 5 and Mythos 5, for any foreign national — whether overseas, inside the United States, or even working at Anthropic itself. Because that net was impossible to draw around foreign nationals alone, the company says it had no choice but to disable the models for *everyone*.

Anthropic's other models, including Claude Opus 4.8, remain available. But the directive marks the first time a leading AI company has taken a deployed, commercially available model offline at the direct instruction of the federal government. The letter, according to the company, arrived at 5:21 p.m. Eastern and was signed by Commerce Secretary Howard Lutnick with input from the department's Bureau of Industry and Security — the office that polices the export of sensitive technology.

The stated reason is unusually thin

By the government's own account, as relayed by Anthropic, the trigger was a demonstrated method of "jailbreaking" Fable 5 — bypassing its safety guardrails. But the company says the technique it was shown amounted to little: asking the model to read a codebase and flag software flaws, which surfaced a handful of previously known, minor vulnerabilities. Anthropic says the same results are readily produced by other models already on the market, including OpenAI's GPT-5.5, none of which face a comparable shutdown order.

Anthropic is complying, but it is not conceding the logic. The company argues it built unusually aggressive safeguards into Fable — so aggressive that users complained they were overbroad — and red-teamed them with the U.S. government, the U.K.'s AI Safety Institute, and outside groups for thousands of hours before launch. No tester, it says, has found a "universal" jailbreak that broadly defeats the model's protections. Its public position is blunt: a single narrow vulnerability should not be grounds for recalling a model used by hundreds of millions of people, and applying that standard across the industry would halt new model releases everywhere. The company called the episode a misunderstanding and said it is working to restore access.

> CONTEXT
>
> Anthropic's CEO, Dario Amodei, has publicly argued that the government *should* have the power to block unsafe model deployments — but only through a process that is, in his words, transparent, fair, clear, and grounded in technical facts. The company's statement on the shutdown says plainly that this action does not meet that bar. In other words, the standard being applied here is one Anthropic has openly disagreed with even while it has invited regulation.

This did not come out of nowhere

To read the directive in isolation is to miss the story. Anthropic and the Trump administration have been in open conflict since winter. In a February meeting, Amodei and Defense Secretary Pete Hegseth failed to bridge a dispute over how the military could use Claude. Anthropic wanted assurances that its models would not be used to power lethal autonomous weapons or to conduct mass surveillance of American citizens. The Pentagon insisted Claude be available for "all lawful use" and would not accept the carve-outs.

When the two sides missed a February 27 deadline, the response was swift and severe. President Trump ordered every federal agency to immediately stop using Anthropic's technology. Hegseth announced the company would be designated a "supply chain risk" — a label that national-security experts note is normally reserved for contractors tied to foreign adversaries, and that is extraordinary to apply to an American firm. The practical effect bars defense contractors from using Claude in any work touching the military.

Anthropic sued. In a 48-page complaint filed in federal court in California, it called the moves unprecedented and unlawful, arguing the government cannot punish a company for constitutionally protected speech, and filed a separate petition asking the D.C. Circuit to review the supply-chain finding. That litigation remains open. The company has estimated the government's actions could cost it billions in revenue.

And here is the detail worth holding onto: in the months since the feud began, the Pentagon has said Elon Musk's xAI and OpenAI's ChatGPT have been cleared for use on classified systems — even though Anthropic was the *first* AI lab permitted on those networks. The company that drew a hard line on autonomous weapons and domestic surveillance was blacklisted. Competitors that did not draw that line were waved through.

Who is in the room

The backdrop to all of this is a tight relationship between the largest technology firms and the current administration. In March, the White House named the first members of its revived President's Council of Advisors on Science and Technology, or PCAST, an advisory body shaping federal AI strategy. Its initial roster included Meta's Mark Zuckerberg, Google co-founder Sergey Brin, Nvidia's Jensen Huang, Oracle's Larry Ellison, and AMD's Lisa Su, co-chaired by the administration's AI adviser David Sacks and science-policy director Michael Kratsios. The council advises; it does not write law. But the pattern is hard to miss: the incumbents hold seats at the table, while the safety-forward outsider is being sued by the same government and has now had its flagship models pulled.

Analysis

> OPINION — INTERPRETATION, NOT REPORTING
>
> Everything below is my reading of the documented facts above. I separate it deliberately so you can judge the evidence on its own.

You do not need a secret cabal to explain what happened here. The simpler, better-supported account is also the more troubling one: an administration already locked in a retaliation lawsuit with Anthropic reached for an elastic export-control authority, on a pretext its own description renders flimsy, against the one frontier lab that has publicly defied it — while leaving more compliant competitors untouched. That is the shape of selective enforcement.

It is worth being precise about what is established and what is inference. Established: the prior blacklisting, the lawsuit, the autonomous-weapons dispute, the clearance of rivals, the thinness of the stated rationale, the PCAST roster. Inference — mine — is that the timing and the asymmetry suggest the shutdown is better understood as another move in that fight than as the narrow technical safety action it is dressed up as. I cannot prove motive, and I will not pretend to. But the burden of explanation is now on the government, which by Anthropic's account did not even put its specific national-security concern in writing.

If there is a single principle that should outlast this news cycle, it is the one Amodei himself articulated: the state may well need the power to stop a dangerous model, but that power has to run through a process that is transparent, fair, clear, and grounded in technical facts. A late-Friday letter that disables a product for hundreds of millions of users, without a written rationale, applied to one company and not its competitors, is the textbook example of how *not* to do it. Whatever one thinks of Anthropic, that procedural question belongs to everyone.

What to watch next

Anthropic said it would share more technical detail within 24 hours of its statement; that material, and whether it names the specific finding the government relied on, will tell us a great deal. The existing supply-chain litigation is the venue where the broader pattern is most likely to surface, because discovery and the public record do that work without anyone having to allege a conspiracy. And the open question for the rest of the industry is the precedent: if a narrow, widely reproducible vulnerability can pull a model overnight, every frontier provider now operates at the discretion of a single letter.

This is a developing story. Sourcing: Anthropic's public statement of June 12, 2026; reporting by CNBC, NBC News, Bloomberg, Fortune, NPR, and CNN; and Anthropic's March 2026 federal court filings. Corrections and tips are welcome.

Tags: Anthropic · Claude · Fable 5 · Mythos 5 · export controls · Commerce Department · First Amendment · AI regulation · Pentagon

FILED UNDER TECHNOLOGY / AI · 2026-06-13 · THE DULUTH LEDE