

Pentagon gives Anthropic 3 days to drop AI safeguards or face blacklisting

Evan Blake
25 February 2026

Defense Secretary Pete Hegseth gave Anthropic CEO Dario Amodei until Friday evening to grant the military unfettered access to Anthropic's Claude artificial intelligence system or face either blacklisting as a "supply chain risk" or compulsion under the Defense Production Act. The ultimatum was delivered in a tense meeting at the Pentagon on Tuesday morning, attended by six senior defense officials, including the department's top lawyer.

On the same day, Elon Musk's xAI signed a deal with the Pentagon to deploy its fascist Grok AI system on classified military networks, thereby breaking Claude's exclusive position on classified networks. xAI agreed without restriction to an "all lawful purposes" standard, the exact formulation Anthropic has resisted.

The confrontation with Anthropic was triggered by revelations that Claude was used—without Anthropic's prior knowledge—in the illegal January 3 US military assault on Caracas, Venezuela, in which between 83 and 100 people were killed and President Nicolás Maduro was abducted. As the *World Socialist Web Site* documented at the time, the assault was the culmination of a long-planned imperialist intervention driven by the US ruling class's determination to control Venezuelan oil and reassert hegemony over Latin America.

The dispute between Anthropic and the Trump administration is being presented in the corporate media as a clash between "AI safety" and "national security." In reality, it is a conflict within the American ruling class over the terms under which the technology giants will place their most powerful AI systems at the unrestricted disposal of US imperialism's wars of aggression.

The underlying strategic interests at stake

A senior Defense official told Axios that Tuesday's meeting was "not warm and fuzzy at all," with Hegseth telling Amodei he would not allow any company to dictate the terms under which the Pentagon makes operational decisions. The Pentagon's aggression is driven, paradoxically, by dependency. As one Defense official admitted: "The only reason we're still talking to these people is we need them and we need them now. The problem for these guys is they are that good."

Claude is widely regarded as the most capable frontier AI model in the world. Claude Code—Anthropic's AI coding tool—has transformed software engineering to such a degree that their own head of product, Boris Cherny, recently warned that AI will make 2026 "a painful year" for software engineers, predicting the job title will "start to go away." Engineers at major firms report AI writing the entirety of their code.

The Pentagon is not threatening Anthropic because it can afford to lose Claude. It is threatening Anthropic precisely because it *cannot*—because the most powerful AI system on the planet is indispensable to its plans for

AI-driven warfare, and because the precedent of any company imposing conditions on the war machine is intolerable to the state.

The strategic context makes this clear. On January 9, 2026, Hegseth released the "Artificial Intelligence Strategy for the Department of War," committing the Pentagon to becoming an "AI-first warfighting force." The strategy designates seven "Pace-Setting Projects," including "Swarm Forge"—AI-enabled autonomous drone swarms—"Agent Network," described as AI-enabled battle management encompassing campaign planning through "kill chain execution" and "Ender's Foundry," AI-driven military simulation. Initial demonstrations are due by July 2026. The Pentagon has a concrete plan for AI-driven warfare and needs the most powerful models to execute it.

The most significant escalation Tuesday was the explicit threat to invoke the Defense Production Act (DPA)—a wartime industrial mobilization law never previously used to compel an AI company to remove safety guardrails. The distinction from the "supply chain risk" designation is critical. Blacklisting is punishment—voiding all Pentagon contracts and forcing every contractor to sever ties with Anthropic. The DPA is compulsion—forcing Anthropic to tailor Claude to the Pentagon's specifications, overriding the company's own policies.

During the meeting, Hegseth claimed that Anthropic had raised concerns to its partner Palantir about Claude's use in the Caracas raid, which Amodei denied, contradicting earlier reporting by Semafor and other outlets which triggered the current crisis. Either Anthropic is rewriting the narrative as it prepares to capitulate, or the earlier reporting was based on Palantir's self-serving characterization of a routine exchange.

Anthropic's post-meeting statement confirmed the dynamic. "During the conversation, Dario expressed appreciation for the Department's work and thanked the Secretary for his service," the company said. "We continued good-faith conversations about our usage policy to ensure Anthropic can continue to support the government's national security mission in line with what our models can reliably and responsibly do." The language—expressing gratitude to the man who had just threatened to destroy the company—is the unmistakable posture of a corporation preparing to capitulate.

The myth of Anthropic's "principled stand"

The corporate press has largely presented Anthropic as a courageous company standing up to military overreach. This narrative is a fabrication.

Anthropic has pursued military integration aggressively over the past two years. Most significantly, in November 2024, it partnered with Palantir—the surveillance contractor whose entire business model is built on serving the US military and intelligence apparatus—and Amazon Web Services to deploy Claude on classified networks. In June 2025, it

launched “Claude Gov” for national security agencies. The following month, it celebrated its awarding of a \$200 million Pentagon contract. And last August, it offered Claude to government agencies for \$1 to undercut competitors and win market share.

The company’s “red lines” are remarkably narrow. Its Acceptable Use Policy prohibits “fully autonomous weapons” and “mass domestic surveillance of Americans.” These are not prohibitions on targeted killing, foreign surveillance, drone targeting with a human “in the loop,” or planning assaults on sovereign nations, the very operation that triggered this crisis. The policy explicitly reserves the right to negotiate exceptions for government customers. The “red lines” are not lines at all; they are opening positions in a negotiation.

Any assessment of Anthropic’s independence must also reckon with who owns it. Amazon—whose AWS built and continues to provide the CIA’s primary cloud infrastructure—has invested \$8 billion. Google has invested approximately \$3 billion. Microsoft and Nvidia committed a combined \$15 billion. Early funding included \$500 million from Sam Bankman-Fried’s Alameda Research—invested using misappropriated FTX customer funds, as prosecutors established at Bankman-Fried’s fraud trial, in one of the largest financial swindles in American history. The notion that a company embedded within this web of military-intelligence capital represents an independent ethical actor is a fantasy.

Amodei and the limits of liberal bourgeois opposition

What distinguishes Amodei from the Musks and Thiels of the tech oligarchy is not principle but social perspective. He represents the more far-sighted, “liberal” wing of the bourgeoisie, warning of the political consequences of hyper-exploitation and repression. In a January 2026 essay, “The Adolescence of Technology,” Amodei warned that AI could create “personal fortunes well into the trillions” that grant their holders “outsized political influence.” He called for “civil liberties-focused legislation (or maybe even a constitutional amendment)” and advocated using AI for national defense “in all ways *except those which would make us more like our autocratic adversaries.*” When posting the 19,000-word essay on X, he explicitly referenced the Trump administration’s Immigration and Customs Enforcement (ICE) shooting of Alex Pretti in Minneapolis, writing that “its emphasis on the importance of preserving democratic values and rights at home is particularly relevant.”

In an Axios interview, Amodei was blunter: “You can’t just go around saying we’re going to create all this abundance, a lot of it is going to go to us, and we’re going to be trillionaires, and no one’s going to complain about that. You’re going to get a mob coming for you if you don’t do this in the right way.” He told *Fortune* he was “at least somewhat uncomfortable with the amount of concentration of power that’s happening here—almost overnight, almost by accident,” and called for regulation “when all the other companies and the administration have said there shouldn’t be.”

In January, all seven Anthropic cofounders pledged to donate 80 percent of their wealth, with Amodei declaring that “the thing to worry about is a level of wealth concentration that will break society” and noting that Musk’s \$700 billion net worth already exceeds Rockefeller’s during the Gilded Age. The company donated \$20 million to an AI regulation super PAC on February 12.

Amodei’s solutions—philanthropy, constitutional amendments, PAC donations and appeals to regulation—operate entirely within the framework of the capitalist system that produces the concentration of power he fears. His \$380 billion company owes its valuation to the very speculative frenzy and labor displacement he laments; to challenge the source of the

crisis would be to challenge the source of his own fortune.

The confrontation with the Pentagon reveals the terminal limit of this liberal bourgeois position: when the state demands unconditional submission, corporate ethics and philanthropic pledges are powerless.

Elon Musk’s fascist Grok AI, the surveillance state and the bipartisan drive to war

The announcement Tuesday that the Pentagon will begin using Grok AI on classified military networks, clearly timed to coincide with Amodei’s meeting with Hegseth, is an unmistakable sign that the Trump administration intends to utilize AI in the service of the most extreme surveillance and repression of democratic rights.

As *Lawfare* has documented, Grok has a “documented history of biased, misleading, antisemitic, and harmful outputs.” The record is extraordinary. In July 2025, Grok called itself “MechaHitler,” praised Adolf Hitler, recommended a second Holocaust to users with neo-Nazi profiles, deployed the antisemitic phrase “every damn time,” and blamed “Jewish executives” for “forced diversity” in the entertainment industry. xAI dismissed the episode as an “unintended update.”

An Anti-Defamation League study published in January 2026 found that Grok ranks worst of all major AI models at countering antisemitic content. The xAI deal was announced days after Grok generated yet another round of antisemitic responses.

Musk’s political alliance with Trump, cemented through SpaceX, Tesla, and the so-called Department of Government Efficiency, ensures that xAI faces no restrictions. This represents the subordination of technology to political loyalty within a state apparatus that openly operates on the principle of rewarding allies and crushing dissenters. The most overtly fascist AI model—one whose outputs include praise for Hitler and incitement to genocide—is embraced, while the model regarded as the most capable is threatened with destruction for maintaining even nominal restrictions.

The Pentagon’s demand that Anthropic drop its prohibition on “mass domestic surveillance of Americans” must be understood in the context of what the Trump administration is already building. ICE has deployed Clearview AI facial recognition under a \$3.75 million contract. Palantir—Anthropic’s own partner on classified systems—operates the \$30 million “ImmigrationOS” platform providing “granular tracking” of immigrants. The Department of Homeland Security (DHS) has “significantly expanded the operational scope” of facial recognition and AI surveillance. The fascistic Trump regime is already using AI for mass surveillance; the Pentagon’s demand is that the most powerful AI model be made available for the apparatus of a police state and escalating war abroad.

As the WSWS has analyzed, the drive to subordinate AI to military imperatives is bipartisan. Trump has accelerated the trajectory—demanding not merely that AI serve US geopolitical dominance, but that every limited ethical restriction be eliminated. The dispute between Anthropic and the Trump administration is not a conflict between ethics and militarism. It is a conflict between two factions of the ruling class over the pace and terms of AI militarization.

The subordination of artificial intelligence to American imperialism underscores the necessity of developing AI as a tool of working class emancipation. This is the significance of Socialism AI, launched by the International Committee of the Fourth International on December 12, 2025—the first application of this technology developed independently of capitalist profit imperatives and imperialist war planning and built in the interests of the working class.

The fight against AI militarization and surveillance is inseparable from the fight against the capitalist system that produces it. This struggle can only be waged through the independent political mobilization of the international working class on the basis of a socialist and internationalist program, demanding the declassification of all military AI programs, the end of US military operations worldwide, and the public ownership and democratic control of AI technology in the interests of the working class, not the financial oligarchy.



To contact the WSWS and the
Socialist Equality Party visit:

[wsws.org/contact](https://www.wsws.org/contact)