

The End of Unrestricted AI: Why Claude Fable 5 Was Just Forced Offline

10 MIN · YOUTUBE · [HTTPS://WWW.YOUTUBE.COM/WATCH?V=B3JLSJ0IOZS](https://www.youtube.com/watch?v=B3JLSJ0IOZS)

<https://www.youtube.com/watch?v=b3jlsj0i0zs>

SUMMARY

The U.S. government's recent order to take Anthropic's Fable 5 offline marks a significant shift in how frontier AI models are treated, transitioning them from software products to controlled national security assets. This unprecedented action affects not only foreign access but also raises concerns about compliance and operational risks for companies like Anthropic, which operate globally.

- The U.S. government has blocked foreign access to Anthropic's advanced models, Fable 5 and Mythos 5, due to national security concerns.*
- The order affects foreign governments, companies, individuals, and even foreign nationals within the U.S., leading Anthropic to shut off access for all users.*
- The situation highlights a lack of transparency and a clear technical standard for government interventions in private companies.*
- The term "foreign nationals" complicates compliance for AI companies that operate globally, effectively acting as a shutdown for the model.*
- Anthropic has previously collaborated with the government on cybersecurity, suggesting a potential for negotiated access solutions.*
- The incident signals a shift in the AI landscape where access to powerful models will be governed by national security considerations.*
- Future AI model launches will require careful consideration of governance, compliance, and access quality, not just model performance.*
- The speaker anticipates that Fable 5 will return with more stringent access protocols, emphasizing the importance of equitable access to AI technology for all.*

I'm filming this from a plane because this is unprecedented. Anthropic just took Fable 5 offline after a US government order and they did not want to wait until it was back on the ground. The headline version is pretty simple. The US government just moved to block foreign access to Anthropic's most advanced models, Fable 5 and Mythos 5. The order reportedly covers foreign governments, foreign companies, foreign individuals, and even foreign nationals inside the United States. That last part is what turns this from a normal export control story into something much stranger because if Anthropic has foreign national employees or contractors or customers an infrastructure exposure, the practical way to comply is really to shut off the models for everyone. And that is what Anthropic says it has to do. So the first point I want to make in the middle of what is an unprecedented moment is this. We should not view this as normal or expected. I think it's very important that we don't. I think this is the first real test of what happens when frontier models are treated less like software products and more like controlled national security assets. And I think there are at least three layers here.

First, obviously the safety layer. Second, there's a legal fig leaf that's going on. And third, there's a business reality, which is why I don't think this lasts. On the safety layer, I want to be careful. The public record is still quite thin. The government has not laid out a technical finding in public. I'm sure they will not. Reporting says the concern may involve a jailbreak pathway connected to fable and to mythals. Anthropic appears to be arguing that this is very narrow, that this is not universal, and that this does not justify the shape of the intervention in any way whatsoever. My read is slightly different. I actually think anthropic may be directionally right that a jailbreak path against one frontier model is not just about that one model.

If a method works against fable, the default assumption should probably be that it will work against any other advanced system until proven. That does not mean that every model is equally favorable. It means the burden ought to shift. In frontier AI, you don't get to say this only worked on one model and then stop. These systems are trained differently and wrapped differently and filtered differently and deployed differently. But they also share enough structure that a real attack pattern is evidence about the class of model, not just the instance. So I am not at all dismissing the safety concern here. But I think the stronger critique is probably about the process. A verbal report or a vague claim or a narrow jailbreak pathway is simply not enough to justify this kind of sweeping order without a real technical process. And Anthropic is arguing that. And I think they're right. If the government is going to intervene at this level in a private company's affairs, there needs to be a transparent statutory path, a clear technical standard, and some way for the company to respond to the actual evidence. Otherwise, this precedent gets very ugly very fast. Because then any frontier model can be frozen on the basis of a claim the public cannot expect from a process no one can audit under a standard that no one can apply consistently. And that's not safety governance. That's simply the exercise of discretionary power. The second layer is the foreign national piece. And this is where I think the order starts to look a little bit like a fig leaves. The phrase foreign nationals sounds narrow.

It sounds like the government is saying we're not trying to ban the product. We're just trying to control who has access to it under existing US law. But for a modern AI company, that line is almost impossible to draw polymer practice. Anthropic doesn't run like a sealed US-on lab from the 1940s, right? It sells globally. It employs globally. It has enterprise customers with global workforces. It runs through cloud infrastructure and support systems and compliance systems, APIs, teams, vendors, and review processes that are all at least in part international. So if the world says no foreign nationals can touch the model, including inside the US, that's not a surgical restriction. It's a shutdown button with export control language wrapped around it. And I suspect everybody involved knows that. Anthropic is going to have to pull broad access back, which is what they did, because the operational risk of accidental non-compliance is extremely high. You cannot casually promise that no foreign national anywhere inside your customer base can access the model if you are distributing the model to hundreds of millions of people. Not on a Friday night, not with civil and possibly criminal penalties hanging over you and not for a live frontier model. So the rule may sound like a targeted national security restriction, but in practice it's effectively a forced pause on the strongest frontier model we have in the world right now. And that is why I think the foreign national language is effective. It gives the government a way to say it didn't ban fable for everybody. But the way anthropic deploys these systems means not for foreign nationals is effectively not for limited. The third piece here is why I think this probably gets resolved very quickly. Anthropic and the US government have had their disagreements in the past, but they've also shown they can work together. The Mythos and Project Glossing story was explicitly about giving trusted cyber defenders and

infrastructure providers access to dangerous capability before the wider public audit in order to shore up the defenses of the wider internet. That means the underlying relationship is not simply that Anthropic refuses and the government brought story over. It's actually a template for negotiated access. There's a story here around running trusted programs together.

There's already a template for saying the capability is real, the risk is real, and the answer is not necessarily a permanent public break. Even Anthropic's posture seems to be temporary. They're saying they're working to restore access. The company is not acting like Fable is dead. It's acting like the access regime broke, which is certainly true. And that's probably the right way to understand this. As frustrating as it is to me personally, because I love Fable. I've been using Fable. It's incredible. This is not really a story though about whether Fable is good. I'll cover that another day. Fable, by the way, is extremely good. I have a full review coming and I pulled it for now because the excess story changed right under our feet as I was getting rid of the road.

But my underlying view of the model hasn't changed. I think Fable is probably the best model in the world right now. And the reason I was excited about it was not just that it answered harder questions. It was that it could carry longer work. It could take messy source material and build real artifacts and preserve context and get work to a reviewable state in a way that felt meaningfully different, a big step forward. And that's still true. And the ban complicates that story. A model this capable is now no longer just a product launch. And we probably won't ever view it that way again after this favored moment. The model launches from here on out are going to become deployment questions. Who can use it? under what rapper, with what safeguards, with what audit trail, with what fallback behavior, and who decides when the risk is too high. I've been saying and others have been saying that models are going to be coming under national security scrutiny. This is what that looks like.

And that's the real Fable story for me. A few days ago, as I was looking at Fable, the question I kept asking myself was, what work do we get to ask the model to do now that the model is this good? How do we expand our imaginations? Now, my question is, who's going to be allowed to ask that? I think both questions matter, but for today, my advice is pretty simple. Take a breath. Don't overreact and declare that flavor was gone. Don't underreact and pretend that this was nothing and this was just another policy hiccup. It was much more than that. It's the first time a frontier model has been rolled back as a big deal. I think the right read is frontier model access is becoming, whether we like it or not, a policy surface. If your workflow depends on one model, one lab, one country's regulatory mode, and one access contract, you do not have a stable operating plan, you've got a dependency. And that doesn't mean to stop using the best model out there.

It means know what that model is for, keep your alternatives warm, and don't build critical work on the assumption that the frontier tier will always be available on yesterday's terms. Look, my guess is that Fable does come back. I think soon. Look, my guess is that Fable does come back, and I think it comes back soon, probably with more process around trusted access, probably with more explicit compliance language, maybe some modified guardrails behind the scenes, some reporting obligations that Anthropic has commercial reasons to figure out so that they can get this model back online. Customers have real reasons to want it back, too. And these are large enterprise customers that have been using Fable. and the government has reasons not to look like it just kneecapped an American frontier AI lab when the United States has been very vocal about wanting to

lead him. So I do think this gets resolved but the warning shot for today matters a lot to all of us. The next era of AI is not just going to be about model quality. We have to stop thinking one-dimensionally. We have to think about access quality and governance quality as well. We have to ask ourselves from now on whether a lab can ship a model that is powerful enough to be on the frontier, tightly governed enough to be allowed by the state and useful enough for customers to really benefit from. This is the line table is sitting on right now and I am really counting the days until we get this figured out and we can get favorable back into everyone's hands. It's a powerful model and I'm excited to reveal it with you. I think this model matters.

I think the fight to ensure intelligence access for all of us matters a lot. If the future is the intelligence economy, we all need access to these models. It's a big big deal and we should be asking for that access. We should be demanding that access. It's not I don't think it's fair or right for only large corporations to have access to frontier models like this. I think all of us deserve it. And I do think despite all of the hiccups and the headaches today, despite the fact that Fable 5 is offline right now, this is a temporary thing. I think the government and Anthropic are going to get this figured out. and I'm excited to post that Fable 5 review when the time comes. In the meantime, keep an eye on your model dependencies. Make sure that you understand how to deploy alternate models for your workflows and subscribe here for more updates. I will literally film from the prime to bring this to you. Cheers. And I hope I don't have to fill more from the prone today.

We will see.