2026 年 6 月 17 日　全 8 頁

米国政府はなぜ最先端 AI を停止させたのか

最先端 AI モデルへの輸出規制措置が示す AI 統治の転換点

経済調査部　AI アナリティックリサーチ室　主任研究員　新田 尭之
AI アナリティックリサーチ室　主任研究員　田邉 美穂

[要約]

■ 2026 年 6 月 12 日、米国政府は国家安全保障上の懸念を理由に、米 Anthropic 社が 6 月 9 日に発表した新たな AI モデル「Claude Fable 5」「Claude Mythos 5」に関して、外国籍者のアクセスを全面禁止する輸出規制措置を発出した。この措置を受け、同社は全ユーザーへの両モデルの提供を即時停止した。6 月 17 日現在、両モデルへのアクセスは停止されたままであり、提供再開の時期は明らかにされていない。

■ 本件の主な背景として、AI の性能が安全保障に直結する水準に近付いているとの危機感がトランプ政権にあることが指摘できる。安全保障を左右し得る技術が、米国政府の管理が及ばない新興の民間企業主導で開発・提供されている構造に対し、統制を確保しようとする意図が今回の措置に表れたとみられる。さらに、米 Anthropic 社が自律型兵器への利用を拒否してきた同政権との政治的緊張も、こうした動きを加速させた一因とみられる。

■ 本件は、民間主導で発展してきた先端 AI モデルそのものへのアクセスを米国政府が輸出管理を通じて直接制限した初の事例であり、AI 統治の転換点と位置づけられる。今後の規制の方向性は、他社モデルへの波及の有無、司法判断の帰趨、AI に対する米国世論の変化、中国との AI 開発競争の均衡などによって左右されよう。

■ 各国政府においても、最先端 AI モデルへのアクセスが政策判断によって制約され得るリスクが顕在化したことを受け、自国で AI 基盤やモデルを含む AI システムを構築・運用するソブリン AI の重要性が再認識されつつある。日本政府においても、産業応用に加え、先端 AI モデル自体の確保・開発体制の強化について、改めて検討が求められる可能性がある。

■ 日本企業にとっては、特定の AI モデルへの依存が政治的要因により一夜にして断たれるリスクが現実のものとなった。複数ベンダーを併用するマルチモデル戦略の構築や代替モデルへの切り替え手順の整備など、「政治的要因によるモデルアクセス遮断」を想定した AI の業務継続計画（BCP）の策定が急務であろう。

株式会社大和総研 丸の内オフィス　〒100-6756 東京都千代田区丸の内一丁目 9 番 1 号 グラントウキョウ ノースタワー

1. はじめに

2026 年 6 月 12 日、米国政府は国家安全保障上の懸念を理由に、米 Anthropic 社が 6 月 9 日に発表した新たな AI モデル「Claude Fable 5」「Claude Mythos 5」に関して、外国籍者のアクセスを全面禁止する輸出規制措置を発出した。米国内外を問わずすべての外国籍者のアクセスが禁止され、その対象には同社の外国籍従業員も含まれる。その結果、同社は法令遵守のため、全ユーザーへの両モデルの提供を即時停止した[1]。

4 月の Mythos 発表以降、AI に対する米国の放任主義は転換点を迎え、政府による一定の介入や規制強化に向かうのではないかとの見方が広がっていた[2]が、今回の輸出規制はそれを裏付けた。本レポートでは、規制の経緯と構造を整理し、AI 統治の転換点としての意味を分析する。

2. 公開からわずか 3 日で停止された新たな AI モデル

新モデルの概要

米 Anthropic 社は 2026 年 4 月、ソフトウェアの脆弱性発見やサイバー攻撃手法の構築において「最も熟練した人間を除くすべてを上回る（大和総研翻訳）」とするモデル Claude Mythos Preview を発表した。同社はその能力の危険性を理由に一般公開を見送り、米 NVIDIA 社等を含むサイバー防御コンソーシアム（企業連合）「Project Glasswing」を通じ、約 15 カ国・約 200 組織に限定提供した。

6 月 9 日、同社は Claude Mythos Preview の後継モデルである Claude Mythos 5 を発表した。さらに、このモデルと同一の基盤モデルに安全策（セーフガード）を追加した Claude Fable 5 を一般公開した[3]（**図表 1**）。

図表 1　Claude Mythos 5 と Claude Fable 5 の概要

	Claude Mythos 5	Claude Fable 5
発表/公開日	2026年6月9日 ※同年4月7日に発表されたClaude Mythos Previewの後継・アップグレード版	2026年6月9日
基盤モデル	Mythos Previewで示された能力水準を受け継ぐモデル	Claude Mythos 5と同一の基盤モデルに、一般利用向けの安全策を組み込んだモデル
提供先	・当初は、Project Glasswing参加組織など、審査済みのサイバー防御組織・重要インフラ関係組織に限定 ・Project Glasswingは、2026年6月上旬時点で約15カ国・約200組織に拡大	・当初は有料サブスクリプションおよび従量課金のAPI経由で一般提供
安全策	・Claude Fable 5と同一基盤だが、Project Glasswing向けにはサイバー関連の安全策を解除 ・提供先を審査済み組織に限定することでリスクを管理	・サイバーセキュリティ、生物・化学、モデル蒸留に関する高リスク用途では、下位モデルへ切り替えて回答
輸出規制措置の影響	米国政府の輸出管理指令を受け、両モデル共に全顧客向けアクセスを停止	

（出所）各種報道および米 Anthropic 社公表資料より大和総研作成

[1] Anthropic “Statement on the US government directive to suspend access to Fable 5 and Mythos 5”, June 12, 2026.
[2] 例えば、The Economist “America wakes up to AI’s dangerous power”, April 16, 2026.
[3] Anthropic “Claude Fable 5 and Claude Mythos 5”, June 9, 2026.

具体的には、Claude Fable 5 ではサイバーセキュリティや生物・化学、モデル蒸留[4]に関するユーザーの要求を受けた場合、下位モデル（Claude Opus 4.8）に自動的に切り替わる設計となっている。一般利用の 95%以上ではこの自動切り替えは発動せず、実質的に Claude Mythos 5 と同等の性能が得られるという。

今回の規制措置に至るまでの経緯

報道によれば、今回の規制措置に先立ち、6 月 11 日に米 Amazon 社の CEO アンディ・ジャシー氏が米財務長官スコット・ベッセント氏らに対し、Claude Fable 5 のジェイルブレイク（安全策の迂回手法）される可能性についての懸念を共有した[5]。

翌 6 月 12 日、米国政府は米 Anthropic 社に「国家安全保障上の懸念」を理由に 90 分以内のモデル停止を要求した。これに対し、米 Anthropic 社は停止に応じず協議を試みたものの合意には至らなかった。その数時間後、米商務省から輸出規制措置の書簡が届き、既述の全面的なアクセス禁止に至った[6]（**図表 2**）。6 月 17 日現在、両モデルへのアクセスは停止されたままであり、提供再開の時期は明らかにされていない。

図表 2　今回のモデル提供停止に至る経緯

日付	出来事
2026/2	・米国防総省は米Anthropic社に対し、Claudeの国防利用をめぐり、米国内の大規模監視および人間の関与を伴わない完全自律兵器への利用を禁じる制限の見直しを求めた。 ・米Anthropic社はこれを受け入れず、トランプ大統領は連邦政府機関に同社技術の利用停止を指示した。 ・米国連邦調達庁（GSA）も米Anthropic社を米国政府向けAI基盤および調達枠組みから除外すると発表した。
2026/3	・米国防総省は、米Anthropicを国家安全保障上の「サプライチェーンリスク」に指定。 ・米Anthropicは、同指定は同社の利用制限方針への報復であり、法的根拠を欠くとして連邦裁判所に提訴した。
2026/4	・米Anthropic社は、サイバーセキュリティ能力が極めて高い未公開モデル「Claude Mythos Preview」を発表した。 ・同社は悪用リスクを踏まえ、一般公開は行わず、重要ソフトウェアの防御を目的とする「Project Glasswing」の参加組織に限定提供することとした。
2026/6/9	・米Anthropic社は、Mythos級モデルに安全策を組み込んだ一般向けモデル「Claude Fable 5」と、限定提供モデル「Claude Mythos 5」を発表した。
2026/6/11	・報道によると、米Amazon社によるClaude Fable 5の安全性検証で、サイバー攻撃に悪用され得る情報を引き出せる可能性が指摘され、同社のCEOアンディ・ジャシー氏がその懸念をホワイトハウス関係者に共有した。 ・米Anthropic社は「汎用的なジェイルブレイク」には当たらないと反論した。
2026/6/12	・米国政府が、外国籍者によるFable 5およびMythos 5へのアクセス停止を求める輸出規制措置を発出した。 ・米Anthropic社は、対象者を技術的に切り分けることが困難として、全ユーザー向けに両モデルのアクセスを停止した。

（出所）各種報道および米 Anthropic 社公表資料より大和総研作成

[4] 既存の AI モデルの出力を教師データとして軽量モデルを訓練する技術を指す。
[5] Axios “They screwed us”: Personality clashes sent Anthropic’s models offline”, June 15, 2026.
[6] Axios “How Amazon and the White House ended Anthropic’s Fable”, June 13, 2026.

3. 安全保障に直結するほど進化したAIと政治的緊張が公開停止の背景か

前述の経緯を踏まえ、ここでは規制の背景を探る。

公表されている情報を見る限り、規制の理由は釈然としない。なぜならば、米 Anthropic 社が Claude Fable 5 の公開前に、米国政府、英国 AI Security Institute (AISI)、複数の民間第三者機関、社内チームと合計数千時間に及ぶ安全性検証を行っていたためである。それにもかかわらず、上記の輸出管理措置を受け、Fable 5 および Mythos 5 の全顧客向けアクセス停止を余儀なくされた。

同社によると、米国政府から国家安全保障上の懸念の具体的内容は示されず、「限定的かつ非普遍的なジェイルブレイクの可能性」を口頭で通知されたにすぎないという。そこで示された手法は、モデルに特定のソフトウェアのソースコード全体を読ませてソフトウェア上の欠陥の修正を求めるものであり、同等の能力は他社の公開モデルでも広く利用可能だとしている。そのため、今回の措置が Fable 5 や Mythos 5 に固有のリスクに基づくものではないと主張[7]している。

それでは、米国政府が今回の規制に至った背景には何があるのだろうか。もっとも、前述の通り、米国政府から国家安全保障上の懸念の具体的な内容は十分に開示されておらず、リスクの実態は外部からは必ずしも明らかではない。そのため、政府側が公表していない情報や、より広範な安全保障上の観点に基づき判断した可能性も否定できない。こうした前提を踏まえつつ、以下では考えられる二つの要因を挙げる。

第一に、AI の性能が安全保障に直結する水準に近付いているとの危機感がある。米 Anthropic 社によれば、Claude Mythos Preview は、主要 OS や主要ブラウザなどで、開発者に知られていなかったゼロデイ脆弱性[8]を数千件特定したとされ[9]、悪用されれば銀行から病院に至る重要インフラを脅かしかねない。こうした安全保障を左右し得る技術が、米国政府の管理が及ばない新興の民間企業主導で開発・提供されている構造に対し、統制を確保しようとする意図が今回の措置に表れたとみられる。

第二に、米 Anthropic 社とトランプ政権の政治的緊張も無視できない。2026 年 2 月、米国防総省が完全自律兵器への利用制限の見直しを求めたのに対し、同社がこれを拒否したことを契機に、トランプ大統領は連邦政府機関に同社技術の利用停止を指示し、連邦調達庁も同社を政府向け調達枠組みから除外した。3 月には国防総省が同社を「サプライチェーンリスク」に指定し、同社はこれを利用制限方針への報復として提訴するに至っている。

こうした一連の対立の延長線上にある措置だとすれば、安全保障上の判断なのか、政治的報復なのかは判然としない。いずれにせよ重要なのは、米国政府が輸出管理を通じて、計算資源

---

[7] Anthropic “Statement on the US government directive to suspend access to Fable 5 and Mythos 5”, June 12, 2026.
[8] ソフトウェアに新たな欠陥が発見されてから対策が確立されるまでの間に存在する脆弱性を指す。詳細は大和総研 WORLD「ゼロデイ脆弱性」（2026 年 6 月 15 日アクセス）を参照
[9] Anthropic “Project Glasswing: An initial update”, May 22, 2026.

や半導体だけでなく、先端AIモデルそのものへのアクセスを直接制限する手段を用いた点である。これは、先端AIモデルの公開・提供をめぐる政府関与が、事実上の事前審査に近付き得ることを示している。

4. 今後のシナリオと日本政府・企業への示唆

今後想定される三つのシナリオ

今後は三つのシナリオが想定される（**図表 3**）。

図表 3　今後想定される三つのシナリオ

シナリオ	概要
A:今回限定の一時的な措置	・国家安全保障上の懸念と米Anthropic社固有の事情が重なった個別措置として処理される ・安全策の追加、政府との技術的協議、対象者の切り分け等により、Fable 5／Mythos 5の提供が条件付きで再開される可能性 ・他社への直接波及は限定的だが、米国製最先端モデルへの依存リスクを再認識する契機となる
B:高リスク・最先端モデル全般に規制拡大	・Mythos級、または高度なサイバー能力を持つモデルについて、政府との事前協議・安全性評価・提供先管理が求められる方向に進む ・米OpenAI社、米Google社等の米国製の最先端モデルにも、同様の評価・制限が適用される可能性 ・米国製モデルへの海外アクセスや外国籍者の利用が制限されるリスクが高まり、複数モデル・複数クラウド・非米国モデルを含む代替手段の確保が重要になる
C:体系的な輸出管理制度の構築	・AIモデルの能力評価、提供先管理、外国籍者アクセス、モデルウェイト・API提供などに関するルールが、法令・行政規則・安全性評価枠組みとして整備される ・予見可能性は高まるが、制度設計・国際調整・企業側の実装には時間を要する ・個別企業を狙い撃ちするような恣意的規制のリスクは低下する一方、制度対応コストは上昇する

（出所）大和総研作成

第一のシナリオ A（今回限定の一時的な措置）は、国家安全保障上の懸念と米 Anthropic 社固有の事情が重なった個別措置として処理されるケースである。安全策の追加、政府との技術的協議、対象の切り分け等により、Claude Fable 5／Claude Mythos 5 の提供が条件付きで再開される可能性がある。この場合、米 OpenAI 社などの他の AI 企業への同様の措置の波及は限定的だが、日本政府や企業にとっては、米国製の最先端モデルへの依存リスクを再認識する契機となろう。

第二のシナリオ B（高リスク・最先端モデル全般に規制拡大）は、Mythos 級、または高度なサイバー能力を持つモデルについて、政府との事前協議・安全性評価・提供先管理が求められる方向に進むケースである。米 OpenAI 社、米 Google 社等の米国製の最先端モデルにも、同様の評価・制限が適用される可能性がある。米国製モデルへの海外アクセスや外国籍者の利用が制限されるリスクが高まり、複数モデル・複数クラウド・非米国モデルを含む代替手段の確保が重要となる。

第三のシナリオ C（体系的な輸出管理制度の構築）は、AI モデルの能力評価、提供先管理、

外国籍者アクセス、モデルウェイト・API 提供に関するルールが、法令・行政規則・安全性評価枠組みとして整備されるケースである。予見可能性は高まるが、制度設計・国際調整・企業側の実装には時間を要する。個別企業を狙い撃ちするような恣意的規制のリスクは低下する一方、AI モデルを開発・提供する企業にとっては制度対応コストが上昇する。

現時点ではシナリオ A（米 Anthropic 社限定の措置）の蓋然性が最も高いとみられるが、今回の規制の論拠が同社固有の事情にとどまらない以上、シナリオ B（最先端モデル全般への規制拡大）やシナリオ C（体系的な輸出管理制度の構築）への移行余地は残されている。

シナリオを左右する 4 つの要因

いずれのシナリオに向かうかは、以下の 4 つの要因に左右されると想定される。

第一に、他の AI 企業への波及の有無だ。既述の通り、ジェイルブレイクは最先端モデルに共通する技術的課題であり、同一の論理で米 OpenAI 社や米 Google 社のモデルにも規制が及ぶ可能性は排除できない。もっとも、米 Anthropic 社が規制対象となった一因は自律型兵器への利用拒否にあり、政府の要求に応じる企業には適用されないとの見方もある。

第二に、米 Anthropic 社の訴訟の帰趨だ。同社は 2026 年 3 月に米国防総省が行ったサプライチェーンリスク指定に対して提訴し、連邦裁判所から暫定差止命令を得た実績がある [10]。今回の輸出規制にも司法判断が下される可能性があり、その結果は AI 規制全体の予見可能性を左右することになろう。

第三に、AI に対する米国世論の変化だ。Pew Research Center が 2025 年 6 月に実施した世論調査によると、米国成人の 50%が AI の普及に「懸念」を示し、「期待」と回答した 10%を大きく上回った [11]。さらに米キニピアック大学が 2026 年 3 月に実施した世論調査では 70%が AI は雇用を減らすと回答しており [12]、データセンター建設への反対運動も各地で生じている。AI に懸念を持つ有権者層が広がりつつある中、AI 規制の強化は政治的に合理的な選択肢となりつつある。

第四に、中国との AI 開発競争との均衡だ。米国は半導体輸出規制を通じて中国の AI 能力を抑制してきたが、自国の最先端モデルへの規制強化はこの競争力を損なうリスクを内包する。規制が長期化し米 Anthropic 社の開発が停滞すれば、効率性の向上やオープンソースの活用で急速に追い上げる中国勢との技術格差は縮小しかねない。一方で、規制なき公開も安全保障上のリスクを招く。米国政府は AI 覇権の維持と安全保障の確保という二律背反に直面しており、

[10] CNBC “Anthropic sues Trump administration over Pentagon blacklist”, published March 9, 2026, last updated March 10, 2026.
[11] Pew Research Center “Key findings about how Americans view artificial intelligence”, March 12, 2026.
[12] Quinnipiac University Poll “The Age of Artificial Intelligence: Americans’ AI Use Increases While Views On It Sour”, March 30, 2026.

その綱引きの帰結が規制の最終的な姿を決定づけることになろう。

各国の AI 戦略に与える影響

今後のシナリオの展開によっては、各国の AI モデル開発をめぐる戦略にも影響が及ぶ可能性がある。これまで生成 AI は開発国以外でも広く利用することが可能であったが、今回の措置は、最先端の AI モデルへのアクセスが、安全保障上の観点を背景とした政策判断によって制約され得る可能性を示した点で、従来の前提に一定の見直しを迫るものといえる。今回の報道後には、英国の AI・オンライン安全担当相であるカニシュカ・ナラヤン氏が、この措置を自国 AI 産業への投資拡大の契機とすべきだとの考えを示すなど [13]、各国政府においても、AI を外部に依存するリスクや、自国で AI 基盤やモデルを含む AI システムを構築・運用するソブリン AI の重要性が改めて意識される状況が見られる。

日本においては、2025 年 12 月に閣議決定された「人工知能基本計画」にて、国産の AI 基盤モデルの開発力強化が官民連携で推進されるべき重要課題として位置づけられている [14]。一方で、その具体的な施策を見ると、生成 AI の基盤モデル単体の高度化にとどまらず、製造業データの活用やロボット基盤モデルの研究開発など、AI を現実世界の制御と結びつける「フィジカル AI」への展開に重点が置かれているようにみえる。これは汎用モデル開発で米中に後れを取る現状を踏まえた資源集中戦略と整理できるが、最先端 AI モデルへのアクセスが政策的に制約され得る可能性が顕在化した今、先端 AI モデル自体の開発・確保体制の強化についても改めて検討が求められよう。

日本企業に求められる対応

日本企業にとっても対岸の火事ではない。日本の大手銀行などは Claude Mythos Preview のアクセス確保や検証を進めていたが、今回の停止により、導入途上の業務プロセスの見直しを迫られている。特定の AI モデルへの依存が、技術的障害ではなく政治的要因によって一夜にして断たれるリスクが現実のものとなったわけだ。

とりわけ深刻なのは、既述したように米国の AI 規制が予見可能性を欠いている点だ。安全性評価や関係機関との事前検証を経て提供されたモデルであっても、政府判断により公開から数日で停止され得るのであれば、企業や個人にとって安定的な利用は保証されない。前述のシナリオ B が現実化すれば、米 OpenAI 社や米 Google 社のモデルを含め、米国製 AI 全般へのアクセスが不安定化しかねない。日本企業が業務の中核に据えつつある AI モデルの供給が、米国の政治力学に左右される構造的脆弱性は看過できないであろう。

---

[13] TIME “Anthropic Pulls Its Most Powerful AI Models After U.S. Bars Foreign Access”, June 14, 2026.

[14] 内閣府「人工知能基本計画 ～『信頼できる AI』による『日本再起』～」（令和 7 年 12 月 23 日閣議決定）

対応としては、AI の業務継続計画（BCP）に「政治的要因によるモデルアクセス遮断」を新たなリスク項目として組み込むことが急務であろう。具体的には、複数ベンダーのモデルを併用するマルチモデル戦略の構築、利用中モデルの停止時に代替モデルへ切り替える「AI フェイルオーバー計画」の策定、そして機密性の高い業務については自社で管理ができるオンプレミスやプライベートクラウドでのモデル運用が検討に値しよう。高性能な AI モデルへのアクセスが技術的・商業的条件のみで決まる時代は終わりつつあるとみるべきであり、今後の動向を注視する必要がある。

以上