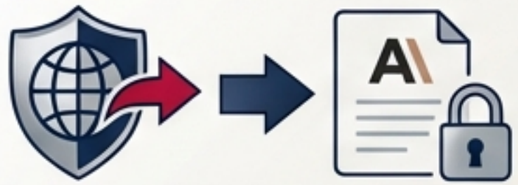
米軍事AIを巡る「安全基準」の攻防：政府 vs. フロンティアAI企業

調達排除事件から読み解く、
国家の法制と民間規範の構造的衝突

事象の本質：単なる「調達排除」ではない

最終決定権の闘争：
「モデルの規範（利用規約）」か、
「国家の法（裁量）」か。

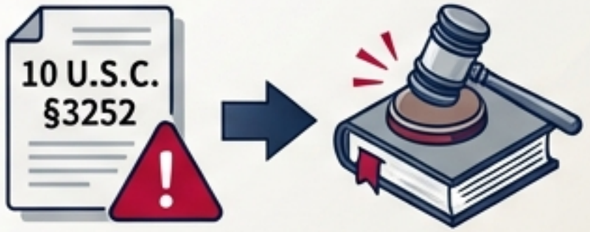
【Trigger】対立のトリガー



- 米国防総省：「Any lawful use（合法なあらゆる用途）」での制限撤廃要求。

- Anthropic：「完全自律型兵器」「大規模国内監視」の2点を不可侵領域として維持し拒否。

【Action】非対称な実力行使



- 政府：10 U.S.C. §3252（供給網リスク）の示唆、大統領指示によるGSA（USAi.gov / MAS）からの即時除外と6か月の移行猶予。

【Implication】構造的変化

- エコシステム：OpenAIの機密ネットワーク合意による市場シフト。
- リスク：民間による安全基準形成への国家介入と、AI市場の分断（軍需特別モデル vs 民間モデル）。

背景 (Situation)：制約なきAIモデルへの野心

[決定的な布石]

1. 「lawful military applicationsを制限し得る利用ポリシー制約のないモデル」の要求。
2. DoD契約における「any lawful use」条項の標準化。
3. 「モデルの客観性」を主要な調達基準へ。

2025年7月 -
America's AI Action Plan

ホワイトハウスが公表。国防分野でのAI採用推進と、調達における「客観性・真実性・イデオロギー偏向の排除」を明記。

2025年8月 -
GSA基盤の整備

政府向けAI評価基盤（USAi.gov）および包括調達枠（MAS）にAnthropic等の主要AI製品を追加。

2026年1月9日 -
国防総省「AI戦略」文書

Takeaway: 今回の衝突は、ホワイトハウスと国防総省が主導する「政権全体のAI政策」と整合する制度設計の延長線上で起きた必然である。

急転直下の「運命の2日間」（2026年2月26日〜28日）

2月26日 [DoDの最後通牒]

国防総省がAnthropicに安全制限撤廃を要求。

期限：2月27日 午後5時1分（米東部時間）

供給網リスク指定（10 U.S.C. §3252）や契約解除を示唆。

2月27日 日中 [Anthropicの拒絶]

公式声明を発表。例外の維持と、法的根拠 (§3252の手続・射程) を法廷で争う構えを明言。

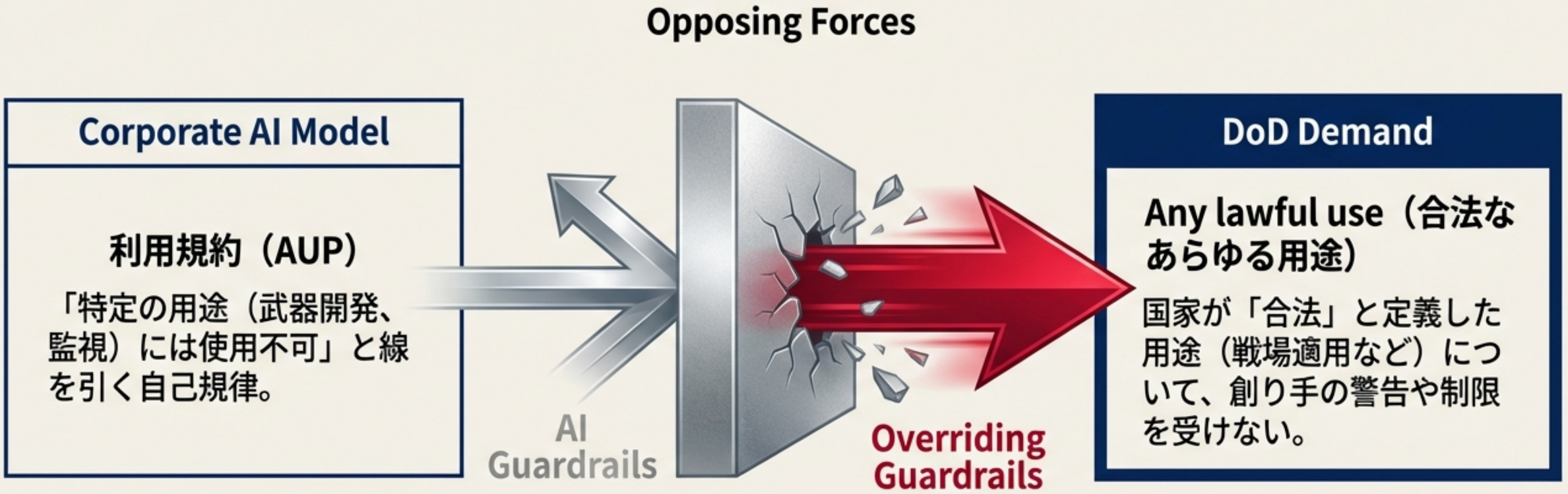
2月27日 午後〜夜 [大統領令と排除]

トランプ大統領が全連邦機関にAnthropic技術の使用停止を指示（猶予6か月）。GSAがUSAi.govおよびMASから公式に除外。

2月27日夜〜28日 [競合の参入]

OpenAIが国防総省の機密ネットワーク提供で合意（報道）。

争点の核心①：国防総省が求めた「Any lawful use」



Key Takeaways

【ポリシー制約の無効化】
モデル提供企業が自社ポリシーで制限を設ける構造を、契約上・運用上で実質的に無効化する意図。

【責任所在の移行】
AIモデルの挙動に関する「最終意思決定権」を、開発企業から政府（運用側）へ完全に移譲させる制度設計。

争点の核心②：Anthropicが死守した「2つのレッドライン」

完全自律型兵器への適用禁止

理由（技術的限界）：現行のフロンティアモデルは信頼性が不十分。

帰結: 戦場での誤作動や誤認識が、米兵および民間人の双方を物理的な危険にさらす。

米国民の大規模国内監視

理由（倫理的・法的脅威）：民主的価値と基本的権利に対する深刻な侵害。

帰結: AIによる監視技術の適用は、既存の枠組みを超える「新規で深刻なリスク」をもたらす。

Insight: Anthropicのスタンスは単なるイデオロギーではなく、「技術の信頼性」と「民主的価値」という論理的な二層構造に基づいている。

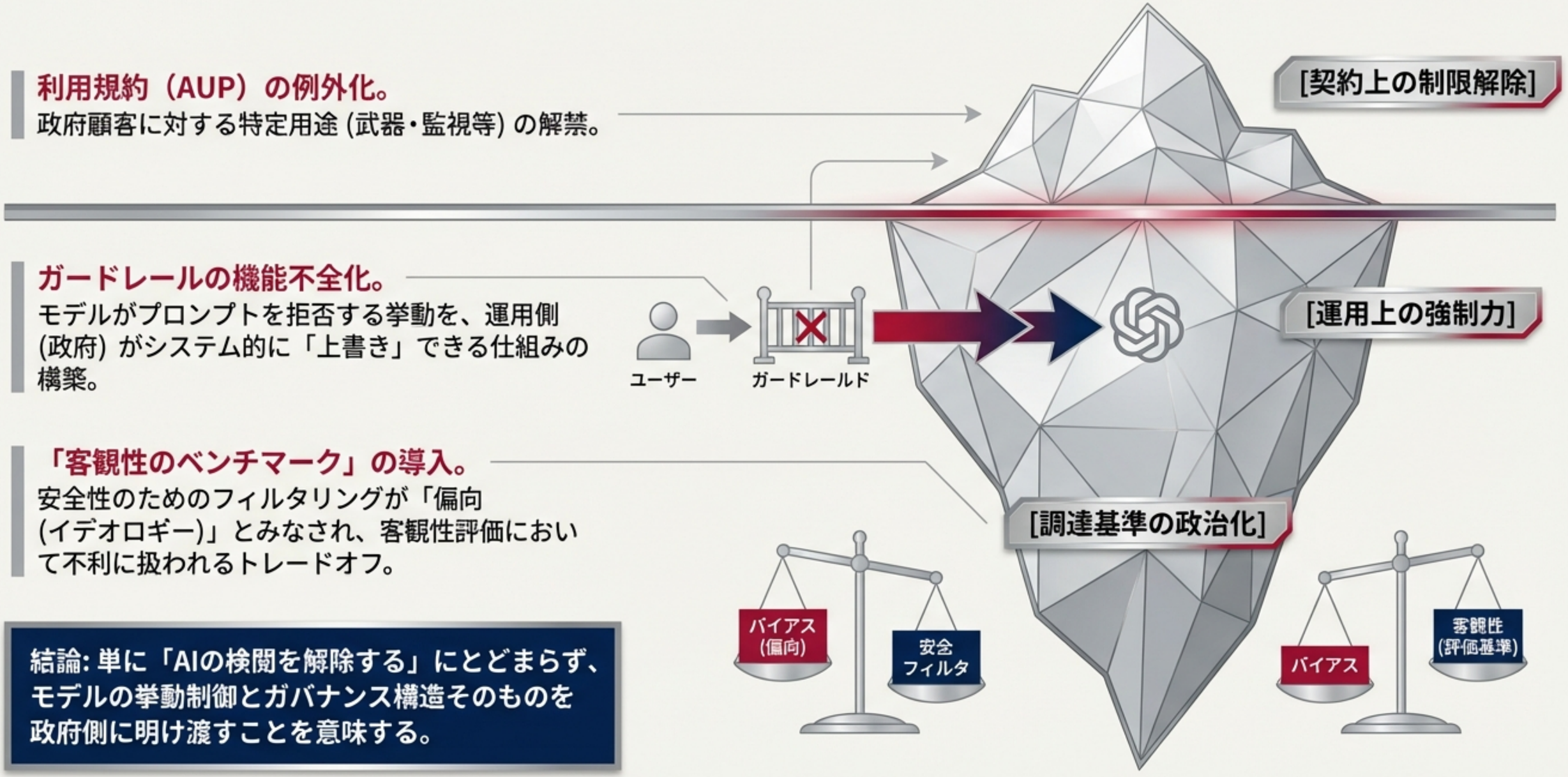
技術的分析：「安全制限の撤廃」が実務的に意味するもの

利用規約（AUP）の例外化。
政府顧客に対する特定用途 (武器・監視等) の解禁。

ガードレールの機能不全化。
モデルがプロンプトを拒否する挙動を、運用側 (政府) がシステム的に「上書き」できる仕組みの構築。

「客観性のベンチマーク」の導入。
安全性のためのフィルタリングが「偏向 (イデオロギー)」とみなされ、客観性評価において不利に扱われるトレードオフ。

結論: 単に「AIの検閲を解除する」にとどまらず、モデルの挙動制御とガバナンス構造そのものを政府側に明け渡すことを意味する。



法的争点①：非対称な圧力としての「供給網リスク（§3252）」

制度 / 根拠	目的と特徴	本件における適用と争点
FAR（連邦調達規則） 停止・排除措置	「処罰ではなく保護のため」という原則に基づく一般的なベンダー排除枠組み。	今回の主な根拠ではない。
10 U.S.C. §3252 供給網リスク権限	リスク低減のため、情報非開示で調達措置（排除・下請制限等）を取れる強力な権限。書面決定・議会通知が必要。	DoDが示唆。Anthropicは「DoD契約を超えた横展開（取引企業全般の排除）は法的権限がない」と反発。
GSAの大統領指示 プラットフォーム除外	政府横断のAI評価基盤（USAi.gov）および包括調達枠（MAS）からの即時除外。	供給網リスク法理とは別ルートで、全政府機関（6か月猶予）に効力を持つ実質的排除。

法的争点②：国防生産法（DPA）の影と限界

DPA
50 U.S.C. §4511

- 国家防衛のための優先権：物資の供給・生産・配分の統制を命じる権限。
- DoDの意図：この権限を用いて「AIの安全設計思想（ガードレール）の強制改変」を命じられるか？

「製品提供の一形態に過ぎない」

vs

「憲法上の言論・表現(企業ポリシー)への違法な介入」

司法審査の論点

国家安全保障における「大統領・行政裁量」に対し、裁判所がどこまで介入・制約できるかという歴史的法廷闘争の可能性。

国際法と軍事規範：「速度至上主義」がもたらすジレンマ

従来の軍事規範 (DoD Directive 3000.09)

- 「武力行使における適切な人間の判断」の要求。
- 国際人道法（IHL）の区別・比例性の原則との整合。

【民間企業による「事実上の規範形成」】

国際法が未整備な中、最終意思決定の「残余責任」を誰が負うかが曖昧化している。結果として、Anthropicのような民間企業が「技術の信頼性限界」を理由にレッドラインを引くことが、実質的な軍事AIの安全基準（防波堤）として機能してしまっている構造的矛盾。

2026年 AI戦略方針

- 「リスクの不完全な整合より速度が勝つ」
- 制約なきモデルの追求。

ステークホルダーとインセンティブのマッピング

トランプ大統領 / 共通役務庁（GSA）

目的: 国防AIの導入加速、政府調達の統制。

行動: 全連邦機関でのAnthropic利用停止指示、USAi.gov/MASからの除外実行。

国防総省（DoD長官）

目的: 「Any lawful use」によるフル柔軟性の確保。

行動: 戦略文書での明文化、§3252（供給網リスク）指定の示唆。

Anthropic CEO（D. Amodei）

目的: モデル信頼性の維持、民間主導の安全規範の確立。

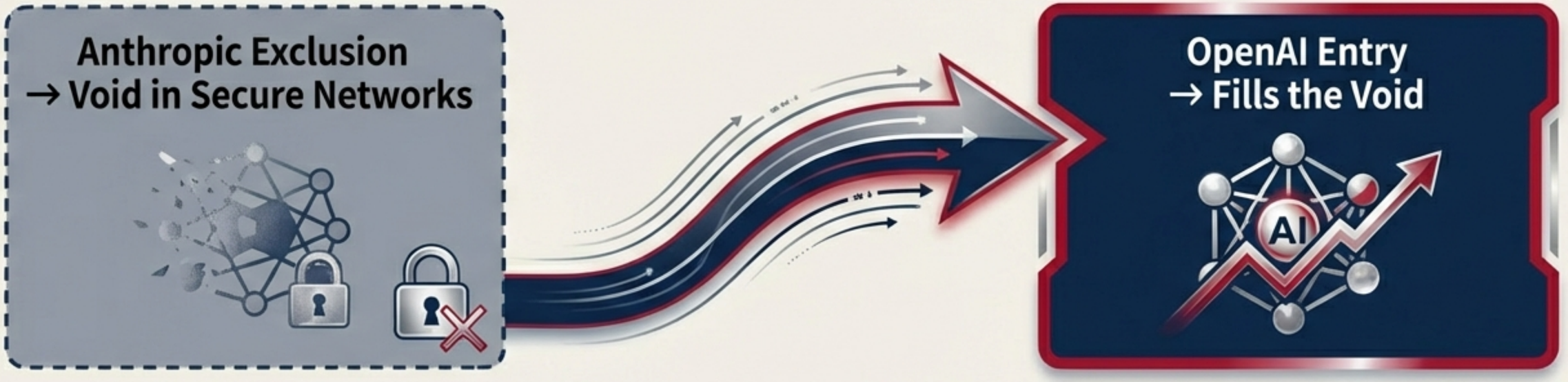
行動: 交渉拒否、法的対抗（§3252の射程を争う）の明言。

OpenAI経営陣（S. Altman）

目的: リスク制御と国防市場シェア獲得の両立。

行動: DoD機密ネットワークへの提供合意（「武力行使の人間の責任」等の赤線設定を条件と報道）。

市場への波及：OpenAIの参入とエコシステムの変化



【代替ベンダーへの需要移転】

Anthropicの排除と6か月の移行猶予により、短期的な市場シェアは合意に至ったOpenAI等へ急速に移行。

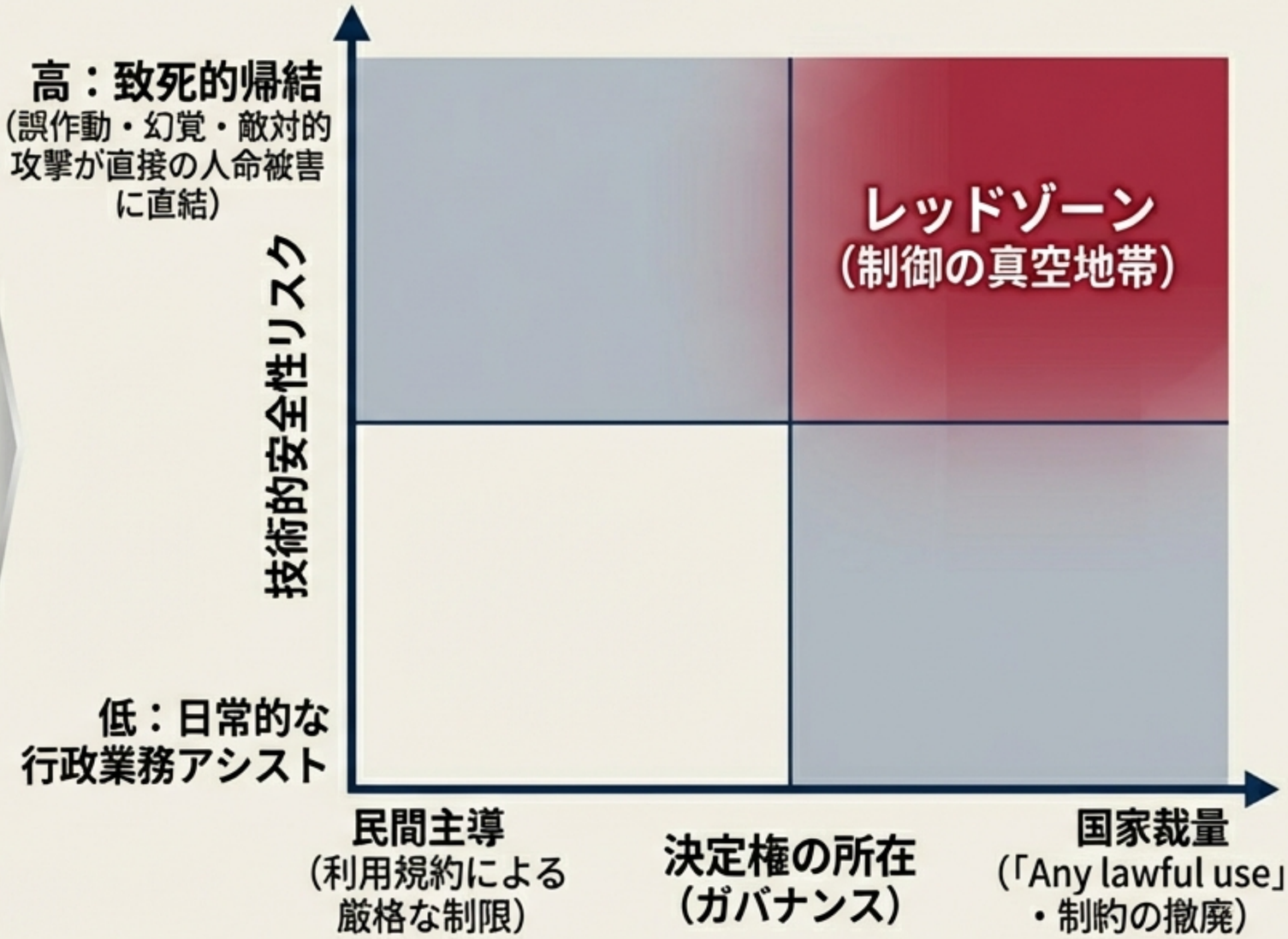
【合意の「赤線」の有効性】

- 報道によれば、OpenAIも**「国内大規模監視の禁止」**「**武力行使の人間責任**」を条件に合意。
- **懸念点:** 一度「Any lawful use」の契約構造を受け入れた後、運用フェーズにおいてこの独自ルールの実効性 (ログ監査・上書き不可の保証) を政府に対し担保できるのか？

構造的リスク評価フレームワーク

【破局的リスクの構造】

技術的な未成熟さ（幻覚・誤認）が存在する状態で、企業側の安全ガードレールが外され、軍の現場裁量で運用が加速した場合、取り返しのつかない事態を引き起こす。



未来のシナリオ：短期（～6か月）および中期（～24か月）予測

現在

短期シナリオ
（～6か月）

中期シナリオ
（6～24か月）

【シナリオA】急速移行と既成事実化

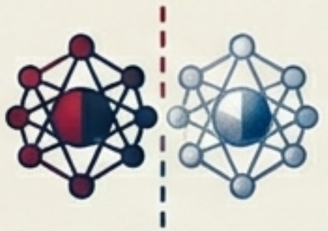
法的決着を待たず代替モデルへ切り替え完了。任務継続性が優先され、安全監査が後回しになるリスク。

【シナリオB】法的差止めと限定運用

Anthropicが司法で勝利（仮処分）。政府の恣意的運用にブレーキがかかり、法廷闘争が長期化。

【シナリオC】AI市場の二極化・分断

軍需・政府向け（安全制約を薄めた特別版モデル）と、民間市場向け（安全重視）の二重エコシステムが常態化。



【シナリオD】議会介入とルールの明文化

密室の交渉による軍事ルール決定に議会が反発。立法による軍事AIの許容範囲の明文化と、行政裁量の縮小。

推奨モニタリング項目：今後の動向を見抜く6つの観点

[Legal] §3252指定の正式文書化

書面決定および議会通知が実際に履行・公開されたか。

[Procurement] GSA調達枠の更新動向

MASやUSAi.govからの除外が恒久化するか、明確な復帰条件が提示されるか。

[Contract] OpenAI・DoD合意の一次情報

契約上の「赤線」、監査権限、モデル挙動上書きの不可逆性が文書化されているか。

[Authority] DPA（国防生産法）適用の具体化

優先契約（§4511）において、明確に「モデル改変」が対象として政令・委任されるか。

[Policy] 軍事規範の整合性

DoD Directive 3000.09（人間の判断）と新AI戦略の「速度優先」が実務審査プロセスでどう折り合いをつけるか。

[Corporate] フロンティア企業の利用規約改訂

OpenAIの安全枠更新や、他社（Google等）の政府顧客向け例外規定の扱い。

Focus: 誰が交渉に勝ったかではなく、「決定権の所在」がどの法制度として固定化されるかを注視せよ。