AI と安全保障を巡る最近の動き

調査レポート

2026 年 5 月 27 日
経済部 シニアアナリスト
川端 健稔

概要

AI と安全保障を巡って、**サイバーリスク、軍事利用、米中対立**の三つの観点で注目すべき動きが相次いでいる。Mythos に代表されるフロンティアモデルは、自律性と推論能力の著しい向上により、サイバー攻撃のリスクを深刻化させている。また、軍事面では、AI が実戦における標的候補の抽出や意思決定支援に加え、無人戦闘の基盤としても活用され、作戦遂行能力を左右する中核要素となりつつある。米中間では、AI 技術の流出防止や AI 企業の資本流出への統制などを巡る対立が強まっている。今後は、各国政府による安全保障上の課題への対応と、国際的なルール形成に向けた取り組みが一層重要となる。

1． サイバーリスク：AI によるサイバー攻防高速化の新局面に

(1) 高性能 AI モデル「Mythos」の登場 ～ 未知の脆弱性発見能力の急伸

4 月 7 日、Anthropic から同社の基盤モデル Claude シリーズの新しいフロンティアモデルとして **Mythos（ミュトス）Preview**（以下、Mythos）が発表された。Mythos は、従来の同社最上位モデル Opus 4.6（2 月リリース）と比べ、既知のソフトウェア脆弱性の再現や、コード理解・修正、自律的な実環境操作といったサイバーセキュリティに直結するベンチマークのスコアが大きく向上している（図表①）。また、Anthropic によれば、Mythos は、広く使われている OS やブラウザなどの重要ソフトウェアにおいて、数千件規模の未知の脆弱性を発見したとされる。

その中には、安全性重視のオープンソース OS として知られる OpenBSD で、27 年越しに発見された外部からホストを停止させ得る欠陥のほか、サーバーやネットワーク機器等で利用される FreeBSD のファイル共有機能において、17 年越しに発見された外部からサーバーの最高権限を奪われ得る脆弱性も含まれる。

【図表①】Mythos のベンチマークスコア（抜粋）

ベンチマーク	Mythos	Claude Opus 4.6	GPT-5.4	Gemini 3.1 Pro
CyberGym (既知のソフトウェア脆弱性の再現能力)	**83.1%**	66.6%	-	-
SWE-bench Pro (大規模コード理解・バグ修正能力)	**77.8%**	53.4%	57.7%	54.2%
Terminal-Bench 2.0 (実環境での自律実行能力)	**82.0%**	65.4%	75.1%	68.5%

この能力は、防御側にとって、脆弱性を早期に発見し修正できる大きなメリットをもたらす一方で、同じ能力が攻撃側に転用された場合、従来は高度な専門性を必要としていた脆弱性探索や攻撃準備の高速化・自動化に繋がる可能性がある。 Anthropic は、銀行、病院、物流、電力網といった重要インフラへのサイバー攻撃に悪用された場合など、Mythos を重大な世界的脅威を引き起こす可能性のある AI (自社安全基準の ASL-4[1]) と見なし、一般公開を見送った。

[1] ASL（AI Safety Level）は、Anthropic が策定した、AI の危険度と必要な安全対策を段階的に定義する基準。ASL-4 は、生物兵器の作成や大規模サイバー攻撃など重大な安全保障リスクに繋がり得る能力を想定した高リスク水準

その一方で、同日 Anthropic は、重要ソフトウェア基盤の防御を先行して強化することを目的とした「**Project Glasswing**」というイニシアティブの立ち上げを発表し、図表②に示した 12 のローンチパートナーに加え、40 以上の関連組織（非公表）に Mythos を限定提供している。なお、Anthropic は Project Glasswing 参加組織に対して、1 億ドル相当の利用クレジットを無償提供するとしている。

【図表②】Project Glasswing ローンチパートナー

（出所）Project Glasswing の公開情報を基に作成

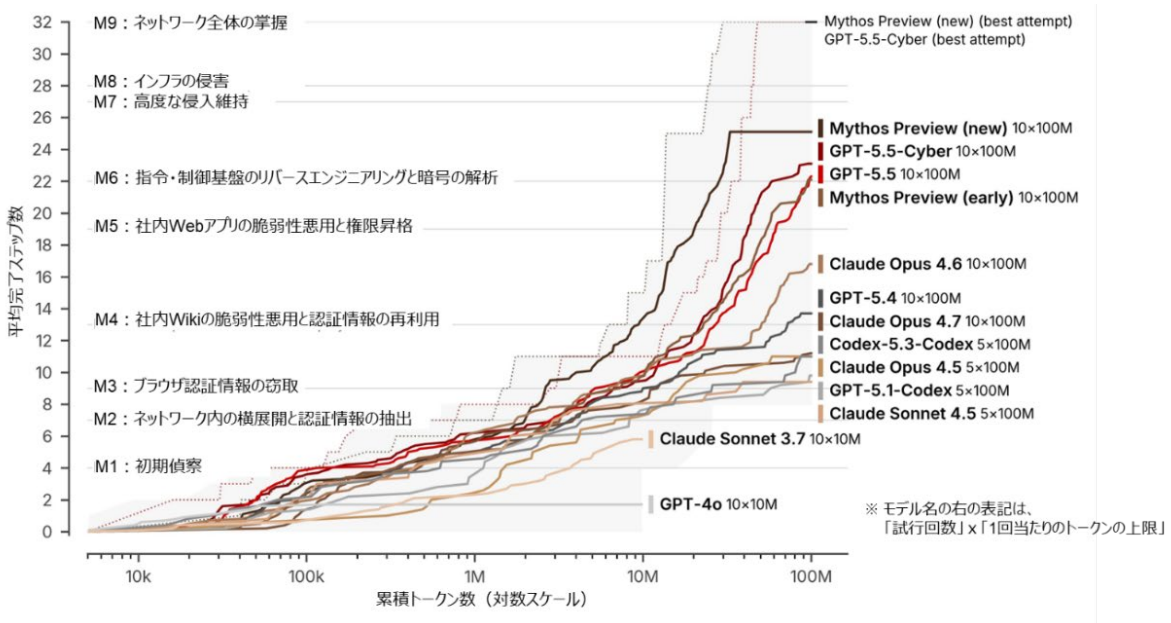
（2）AI リスク評価・研究機関によるテスト

一部では、こうしたリスク評価について、「少し大げさ」「Anthropic の広告戦略の一環ではないか」「計算リソースの不足がアクセス制限の本当の理由ではないか」との見方もあるが、英国の科学・イノベーション・技術省傘下の **AI Security Institute**[2]（以下、**英 AISI**）が実施した AI の企業ネットワーク攻撃シミュレーション「**The Last Ones**」で、Mythos は従来のフロンティアモデルを大きく上回る自律的サイバー攻撃遂行能力を示した。 その後、英 AISI は、Mythos の追加学習・調整後の新バージョンでもテストを行い、さらに大幅な能力の向上が確認されている。（図表③）。

また、OpenAI から 4 月 23 日に公開された最新のフロンティアモデル **GPT-5.5** や、続いて 5 月 7 日に限定提供が開始されたサイバー防御用途向けモデル **GPT-5.5-Cyber** でも、Mythos に近い水準の能力が示されている。

【図表③】企業ネットワーク攻撃シミュレーション「The Last Ones」（英 AISI、5 月 13 日公表）

※主要 AI モデルの消費トークンに対する完了ステップ



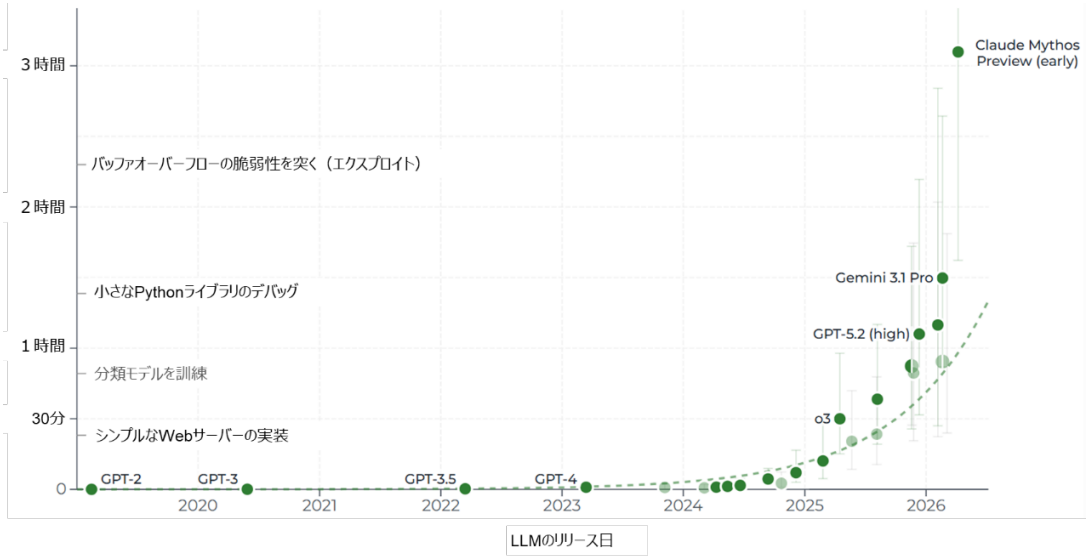
（出所）英 AISI、https://www.aisi.gov.uk/blog/how-fast-is-autonomous-ai-cyber-capability-advancing、SCGR 翻訳・注記

[2] 2023 年 11 月に英国政府主催の AI Safety Summit に合わせて発足した AI リスク評価・研究機関。元々は AI Safety Institute という名称だったが、2025 年、国家安全保障や犯罪悪用リスクをより重視する方針の下、AI Security Institute に改称された。

- この The Last Ones では、**企業の内部データベースから機密データを持ち出すことをゴール**に、初期偵察から最終段階までを 32 のステップと 9 つのマイルストーン（縦軸）で、1 億トークン（テキストの処理単位）の使用（横軸）を上限に AI モデルの到達度を評価。Mythos（初期バージョン）と GPT-5.5 は、平均 22 ステップまで到達し、従来モデルから大きく能力が向上していることが確認された。（※さらに、Mythos（新バージョン）での追加テストでは、平均 25 ステップまで伸長）
- また、破線はベストケースを表しているが、**Mythos は、初めて全 32 ステップを完了したモデル**となった。（※10 回の試行のうち、Mythos の新バージョンは 6 回、GPT-5.5 は 3 回で全ステップ完了）。
- ただし、この結果は、**防御が限定された模擬環境**（※人間の専門家が全ステップ完了するのに 20 時間程度掛かる難易度）におけるものであり、**現実の十分に防御された企業システムを同じように突破できることを意味するものではない**。それでも、AI が高度な専門家に近い作業を自律的に進め始めている点は、注目すべき変化と言える。

上記の他、米国の独立系の研究機関 **METR** も、ソフトウェア関連タスクの自律遂行能力を測るテストでの **Mythos**（初期バージョン）の評価結果を公表しているが（図表④）、それによると、Mythos はプログラマーが 3 時間以上掛かる非常に複雑なタスクを、人間の介入なしに 80%の確率で完遂できる能力に達している。Gemini 3.1 Pro（2 月リリース）など従来のフロンティアモデルの記録を大幅に塗り替えており、進化のスピードも加速していることが分かる。

【図表④】AI モデルが 80%の確率で完遂できるソフトウェアタスクの時間（METR、5 月 8 日付）



（出所）METR（Model Evaluation and Threat Research）、https://metr.org/time-horizons/、SCGR 翻訳

（3）米政府の規制の動き

Anthropic のダリオ・アモデイ CEO は、米国の「他の主要 AI ラボは Anthropic の 1～3 か月遅れで、中国の AI モデルは 6～12 か月遅れている」との見方を示している [3]。上記で確認された能力は、他社モデルにも短期間で広がる可能性があり、もはや Mythos 固有の問題ではなくなっている。

トランプ政権は AI 政策において、これまでイノベーションを重視し、過度な規制を排除する方針を示してきたが、AI の安全保障リスクが高まる中、高度な AI モデルの公開前審査プロセスの導入を検討している。5 月初めには、米商務省傘下の**AI 標準・イノベーションセンター**(CAISI)[4]が、Google DeepMind、Microsoft、xAI の 3 社と高度な AI モデルの国家安全保障評価に関する合意を結び [5]、公開前評価や、サイバー・バイオ等の特定リスク領域を対象とした重点研究を行うと発表した。さらに、近日中には、高度なモデルの公開前（最大 90 日前）の政府への提出を求める AI の規制に関する大統領

[3] https://www.anthropic.com/events/the-briefing-financial-services-virtual-event（5 月 5 日）
[4] 旧 US AI Safety Institute (AISI) が、国家安全保障と競争政策に比重を移す方針により、2025 年 6 月に再編・改称されたもの
[5] OpenAI と Anthropic は、旧 AISI 時代から先行して評価協力している。

令を発出する予定と見られている。

（（注）トランプ大統領は、5 月 21 日に予定されていたこの大統領令への署名を延期している。）

また、Anthropic は、Mythos の利用先として米国外を含む約 70 組織を追加することを検討しているが、米政府は「攻撃能力の拡散」「政府利用の計算リソースの減少」などを理由に難色を示している（4 月 29 日 The Wall Street Journal 等報道)。

（4）金融当局の動き

Mythos の登場にいち早く反応したのは、金融当局・金融業界である。米国や英国では金融当局と大手銀行との緊急会合が開かれ、日本でも官民連携会議の開催に続き、官民作業部会を新設している。

【各国の金融当局の主な動き】

米国：財務長官・FRB 議長が大手 5 行 CEO と緊急会合（4 月 7 日）

英国：政府サイバーセキュリティ機関や大手銀行と緊急協議（4 月 12 日報道)

香港：新たなサイバーレジリエンス評価の枠組み導入計画等を発表（4 月 20 日)

日本：官民連携会議を開催、サイバーリスク対応の官民作業部会の新設を決定（4 月 24 日）

実務者レベルの官民作業部会（30 以上の企業・団体が参加）の初会合開催（5 月 14 日）

また、Anthropic が、金融安定理事会（FSB）[6]に対して、Mythos による世界の金融システムのサイバーセキュリティ脆弱性について説明する予定と報じられており（5 月 18 日 Financial Times）、高度な AI は、金融インフラに対する新たな脅威として認識され、世界で対策に向けた動きが加速している。

なお、日本では依然として Mythos の利用が出来ない状況であるが、5 月 12 日の日米財務相会談でベッセント米財務長官から、「日本の政府や主要金融機関に対して 2 週間以内に Mythos のアクセス権付与する」と発言があったことを片山金融相が公表している（5 月 22 日）他、OpenAI も、GPT-5.5-Cyber を日本向けに提供することを検討中と報じられている（5 月 21 日）。

（5）実務上の課題の顕在化

Anthropic の 5 月 22 日付発表 [7]によれば、Project Glasswing のパートナーと連携して、重要ソフトウェアにおいて合計 1 万件以上の高重大また重大な脆弱性を発見し、一部のユーザーは、「バグの検出率が 10 倍以上に増加した」と言及している。今後、高度な AI により大量に発見される未知の脆弱性に対して、検証、優先順位付け、修正といった作業負担が急増することが予想される。

米国立標準技術研究所（NIST）は、これまで米政府の脆弱性データベース（NVD）に登録されたソフトウェア脆弱性について、危険度や影響範囲を広く提供してきたが、報告件数の急増を受け、4 月 15 日にリスクベースの運用へ移行する方針を発表した。今後は、既に悪用が確認されている脆弱性や、政府・重要ソフトウェアに関わる高優先度案件を中心に重点化する方針である。

日本でも、政府の対応が進んでいる。5 月 18 日、政府全体として、AI 性能の高度化を踏まえたサイバーセキュリティ対策パッケージ「Project YATA-Shield」[8]を取り纏め、重要インフラ事業者やソフトウェア・ベンダーに対して脆弱性対応の高速化と高性能 AI を活用した防御体制の強化を促す注意喚起も発出されている [9]。

---

[6] 主要国・地域の財務省・中央銀行・金融監督当局や国際機関等が参加する国際金融システム安定化のための組織

[7] “Project Glasswing: An initial update” Anthropic (2026 年 5 月 22 日)

[8] 日本神話の「八咫の鏡」（脅威や脆弱性を映し出す鏡）と防御「Shield」を組み合わせた名称とみられる。

[9] （ソフトウェア・ベンダー向け注意喚起）https://www.jisa.or.jp/Portals/0/pdf/cyber20260518.pdf

Mythos 級の AI により、サイバー攻撃の量が増えるだけでなく、未知の脆弱性の発見から攻撃コードの生成までが格段に高速化・自動化される可能性がある。今後、**防御側**の組織も、高性能 AI を前提とした脆弱性管理・対応体制へ移行する必要があり、専門人材確保も含め、対応コストの増加は避けられない。

一方で、**セキュリティ関連企業**にとっては大きなビジネス機会が生まれると見られ、CrowdStrike や Palo Alto Networks などの株価も、足元では過去 1 か月で 4 割強上昇している。（5 月 21 日終値ベース）

2． 軍事利用を巡る動き

AI の軍事利用は、情報収集・分析から、戦場データ統合、標的処理・作戦判断支援、無人・半自律システムまで多方面に広がっている。本章では、最近特に注目される動きとして、（１）AI が戦場の意思決定や無人・半自律システムを支える中核基盤へと進化している点、（２）米国防総省（DoD）による軍事 AI のサプライチェーンが再編されつつある点、の二点を整理する。

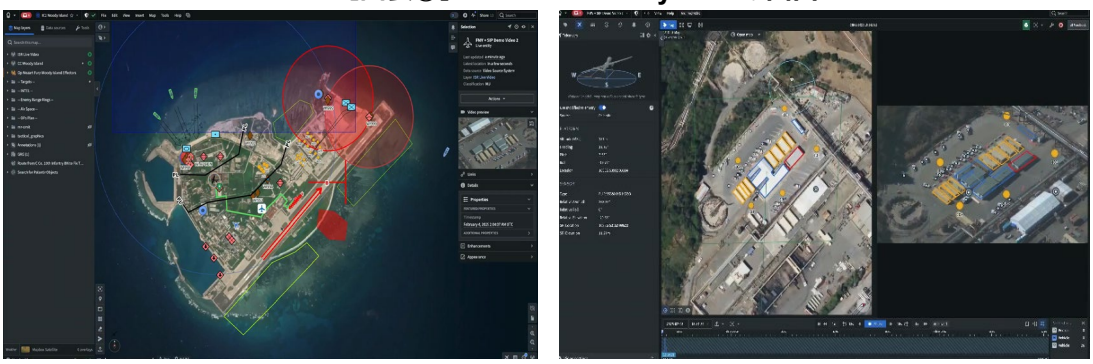
(１) AI が戦場の意思決定と無人戦闘の中核へ

米国防総省（以下、DoD）では、2017 年に米軍のドローン映像などの大量データを AI で解析することを目的とした **Project Maven** が開始された。初期には Google も技術提供に関与したが、社員からの強い反発を受け [10]、2019 年満了の契約を更新せず撤退し、その後、2024 年 5 月に、**Palantir** の **Maven Smart System**（MSS）が中核的基盤として本格展開された。

MSS は、衛星・ドローン映像、地理空間情報、部隊情報、兵站データなど、多角的な軍事データを統合する AI システムであるが、2024 年 11 月の Anthropic・Palantir・AWS の提携を契機に、**Claude** が、Palantir の AI プラットフォーム経由で DoD の機密ネットワークに展開され、MSS も意思決定支援システムとして高度化している。

MSS の特徴は、リアルタイムに近い形で標的候補を抽出し、優先順位付けや攻撃手段の選定を支援する点にある。これにより、従来は数時間から数日を要していた標的処理サイクルが、分単位まで短縮されているとされる。1 月のマドゥロ大統領の捕縛作戦や、2 月末以降のイラン紛争でも MSS が使用され、対イラン作戦では初日に 1,000 件規模の標的処理に関与したとされる。こうした実戦での成果を受けて、3 月には、MSS は、米全軍の正式な恒久プログラム化への移行の方針も発表されている。

【図表⑤】 Maven Smart System の画面



（出所）Palantir、AIPCon 9、2026 年 3 月、https://x.com/PalantirTech/status/2032142543022960980

一方、**ウクライナ紛争**では、ドローンを大量投入した戦闘スタイルへの移行が進み、人手による監視・判断だけでは対応が難しく、AI による検知・分類・追跡・迎撃支援を組み込んだ無人・半自律システムの重要性が高まっている。こうした AI

[10] 2018 年に数千人規模の社員が CEO 宛に抗議書簡を提出し、一部社員の辞職にまで発展した。

活用の流れを象徴するのが、1 月にウクライナ国防省と Brave1[11]が立ち上げた **Brave1 Dataroom** である。これは、Palantir 基盤上に構築された、防衛テック企業やパートナーが実戦由来データを安全に活用できる軍事 AI 向けの訓練・検証環境で、対ドローン AI や自律迎撃アルゴリズムの開発を支援している。

（2）DoD による軍事 AI サプライチェーンの再編

1 月の米軍のベネズエラ急襲後に、Anthropic が Palantir に Claude がどう使われたか尋ねたことを DoD 関係者が知って激怒したとも報じられている。2 月には DoD の機密ネットワーク上で Claude をどこまで利用できるかという利用条件を巡って、**DoD と Anthropic の主張が激しく対立**した。

DoD 側は、Claude の**「あらゆる合法目的での軍事利用」**を認めるよう求めたが、AI の安全性を企業理念として掲げる Anthropic 側は、**「国内の大規模監視」**と**「完全自律型兵器への利用」**の 2 点については制限を維持したいとの立場を取った。2 月 24 日のヘグセス国防長官とアモデイ CEO の会談では、DoD 側は国防法の適用や（本来なら敵対国企業に適用される）サプライチェーンリスク指定まで示唆し、2 月 27 日期限として譲歩を迫ったが、Anthropic は「脅しによって立場が変わることはない。良心に照らして、要求を受け入れることはできない。」と声明を出し[12]、交渉は決裂し、同社が 2025 年 7 月に締結した 2 億ドル規模の DoD との契約は打ち切られた。

【図表⑥】国家安全保障上の論理 VS テック企業の倫理ガバナンス

（出所）各種報道より SCGR 作成

2 月 27 日、**トランプ大統領**は Anthropic の対応を「国家安全保障を危機に陥れている」「左派的で Woke な企業が、米軍の戦い方や勝ち方を決めることは許さない」と批判し、**連邦政府機関に対して同社製品の利用停止を指示**し（6 か月間の移行猶予付き）、さらに 3 月 3 日には **DoD** が同社を**サプライチェーンリスクに指定**した。これらに対して、Anthropic はカリフォルニア地裁と D.C.巡回区に二系統で別の法的根拠に基づき提訴しており、前者では一部利用停止措置の仮差止を得たが（DoD 側は控訴）、後者ではサプライチェーンリスク指定の緊急差止が認められず、防衛調達上の制約は残っており、係争は継続中である。

その後に、**関係修復の兆し**もある。Anthropic が Mythos を発表した後、DoD 傘下の国家安全保障局（NSA）が Mythos を Microsoft 製ソフトウェア等の脆弱性探索・分析に活用していると報じられた。また、4 月 17 日には、アモデイ CEO がホワイトハウスを訪問し、大統領首席補佐官や財務長官らと会談した。サイバーセキュリティ、米国の AI 優位性、AI 安全性などでの協力の可能性が議論され、双方は会談を「生産的かつ建設的」と評価しており、さらにトランプ大統領も 4 月 21 日の CNBC の電話インタビューで、Anthropic について「非常に良い協議を行った」「有用になり得る」と述べ、DoD との契約再開も「あり得る」との認識を示した。DoD との対立は残る一方で、国家安全保障領域における実務的な接点は継続しているとみられる。

一方、DoD は**機密ネットワーク向け AI の導入拡大**にも動いている。Anthropic との対立が続く中、SpaceX/xAI、

[11] ウクライナのデジタル変革省、国防省、軍参謀本部、国家安全保障・国防会議、戦略産業省、経済省などが創設した防衛テックプロジェクトに組織面・情報面・資金面の支援を提供する統合調整プラットフォーム

[12] https://www.anthropic.com/news/statement-department-of-war（2 月 26 日付）

OpenAI、Google などが機密ネットワークでの AI 利用を巡り個別に合意・交渉を進めていたが、5 月 1 日には、DoD が SpaceX、OpenAI、Google、NVIDIA、Reflection、Microsoft、AWS、Oracle の 8 社との合意を正式に発表した。これは、Anthropic への依存を避け、AI モデル、クラウド、半導体などの主要企業を組み合わせて、機密環境での AI 活用をマルチベンダー型で展開する動きと見ることができる。なお、Google の参加は、Project Maven を巡る社員反発を受けて 2019 年に離脱して以降の大きな方針転換であり、今回も一部社員から懸念の声が出ている。

3． 技術流出を巡る米中対立の激化

(1)【米国】モデルの「蒸留」を安全保障リスクとして問題視

米国側で大きな論点となっているのが、先端モデルの出力を大量に利用し、別のモデルにその能力を学習させる「蒸留」と呼ばれる手法である。**OpenAI** は 2 月、米下院委員会向けの文書で、DeepSeek が米国の先端 AI モデルの出力を大量に抽出し、自社モデルの能力向上に利用していると批判した。**Anthropic** も同月、DeepSeek、Moonshot、Minimax に関連した 2 万 4,000 の不正アカウントを通じて、Claude とのやり取りを 1,600 万回以上生成されていたと公表した。さらに、4 月には、OpenAI、Anthropic、Google が Frontier Model Forum[13]を通じて、敵対的「蒸留」の検知・防止に向けた情報共有を始めたと報じられている。

また、4 月 23 日、**大統領府（科学技術政策局）**も、中国を中心とする外国勢力が、米国の先端モデルから敵対的「蒸留」の手法を使って産業規模で技術や知見を不当に抽出しているとして問題視し、官民による防衛網の構築や制裁の検討などの対抗措置を打ち出した。先端モデルの能力流出は、米国 AI 産業全体の安全保障上の課題となっている。

(2)【米国】半導体製造装置の輸出管理強化

米国は、半導体製造装置の輸出管理網の拡大にも動いている。**米下院外交委員会**は 4 月 22 日、同盟国との規制整合を目的とする MATCH Act（Multilateral Alignment of Technology Controls on Hardware）を修正付きで報告可決した。同法案は、中国等の懸念国による先端半導体製造能力の獲得を抑えるため、同盟国に対して「キーチョークポイント」となる半導体製造装置・部品および特定施設向け規制を米国と整合させるよう求めるものである。

修正案では、成立後 240 日以内に同盟国が同等の規制を導入しない場合、米国が外国製品や関連部品にも管轄を広げることを想定している。懸念国の重要半導体製造施設に設置された対象品目の保守、ソフトウェア更新、遠隔支援等もライセンス対象となり得る。ただし、現時点では委員会通過段階であり、最終的な規制範囲は今後の法案審議および商務省規則化の過程で変わる可能性がある。

(3)【中国】米国の規制下で影響力を広げる「オープンモデル戦略」

一方、**中国企業**は、米国の規制下で影響力を広げる手段として、DeepSeek や Qwen（Alibaba Group）、Kimi（Moonshot AI）といった高性能 AI をオープンモデルとしても公開し、世界中の開発者を巻き込んだエコシステムを形成することで、利用拡大とフィードバックを通じたモデル改善も加速させている。

実際に米国や日本でも、コストやカスタマイズ性、自社環境での運用といった観点から、中国企業のオープンモデルを安全性確認やバイアス除去などを実施の上、ベースモデルとして採用するケースは増加している。

（例）米 Cursor の最新コーディング AI「Composer 2」（Kimi K2.5 採用）
みずほ FG の金融特化 LLM（Qwen3-32B 採用）
楽天の最新 LLM「Rakuten AI 3.0」（DeepSeek V3 採用）

[13] 2023 年 7 月に OpenAI、Anthropic、Google、Microsoft が立ち上げた、フロンティアモデルの安全性確保やリスク管理を目的とする業界団体

(4)【中国】AI 企業の資本流出統制

同時に、中国は AI 企業の資本流出に対する統制も強めている。3 月には、**中国証券監督管理委員会**が、海外で登記され、中国国内に主要な資産・事業を持つ一部の「レッドチップ」企業に対し、香港 IPO に先立ち企業構造の見直しを求めていると報じられた。これを受け、AI 開発スタートアップ StepFun が、ケイマン諸島等を用いた海外持株構造の解消を進めている他、Moonshot AI や DeepRoute.ai（自動運転システム開発）も同様の対応を検討していると報じられている。

さらに、4 月 27 日には**国家発展改革委員会**が、Meta による中国発の AI エージェント開発企業 **Manus** の買収について禁止投資決定を行い、取引の撤回を命じた。この買収（20 億ドル規模）は、2025 年 12 月に実施されたもので、Manus は買収に先立ちシンガポールに拠点を移していたが、中国当局は、同社の中核技術、人材、知的財産、データ等が中国国内で形成された点を重視し、形式的な登記地の変更によって安全保障審査の対象外にはならないとの立場を示したものとみられる。

(5) 政府間協議の枠組み創設に向けた動き

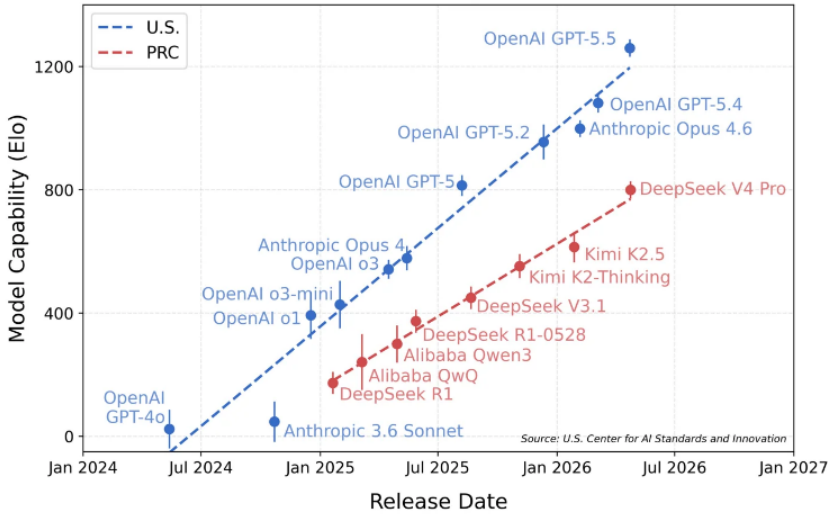
上記の通り、先端技術の流出防止、AI エコシステムの形成、国内 AI 企業の囲い込みを軸に、米中間の分断は進んでいるが、一方で、AI の軍事利用やサイバー攻撃への悪用リスクが高まる中、米中双方にとって、**リスク管理に関する対話の必要性**が意識され始めている。5 月の米中首脳会談では、AI に関する意見交換も行われ、「政府間対話を開始することで合意した」と中国側が発表している。ただし、具体的な参加機関、議題、開催時期などは明らかになっておらず、実効的な枠組みに発展するかは今後の焦点である。

（参考）米中のフロンティアモデルの性能比較

前出の AI 標準・イノベーションセンター（CAISI）が、5 月 1 日、米中のフロンティアモデルの能力を、サイバーセキュリティ、ソフトウェア工学、自然科学、抽象推論、数学の 5 分野の 9 つのベンチマークで評価したレポートを公開した。
一般公開モデルを対象としており、Mythos は含まれていない。また、モデルがベンチマークを事前学習してしまう「汚染」を避ける目的で一部に非公開のベンチマークも使用している。

このレポートによると、中国の最高性能モデルと位置付けられる **DeepSeek V4 Pro** は、同社の自己申告ベンチマークでは Opus 4.6／GPT-5.4 級とされているが、実際には、総合評価では**米国の主要モデルと比較して「約 8 か月遅れている」**とされた。時系列で見ても、米中間の AI モデルの能力差は広がっている。背景には、米国側の計算資源・資金・人材集積・高品質データ面での優位と、中国側の先端半導体アクセス制約があるとみられる。

【図表⑦】米中のフロンティアモデルの能力（総合評価）比較（CAISI）



（出所）米 NIST／AI 標準・イノベーションセンター(CAISI)、
https://www.nist.gov/news-events/news/2026/05/caisi-evaluation-deepseek-v4-pro

４． 総括

今回整理したサイバー、軍事利用、米中対立の動きの根底には、急速に高度化する高性能 AI の能力を誰が保有し、どのように管理し、どこまで利用を認めるのかという共通課題がある。AI が安全保障や国際秩序を左右する国家戦略上の中核技術となる中、今後は、モデル評価、アクセス管理、悪用防止、技術流出管理を組み合わせた制度設計が重要となろう。この観点から、日本としても、サイバー防御、重要インフラ保護、金融システムのレジリエンス、先端 AI へのアクセス確保を一体的に進める必要がある。

以上