

「ミュトス級 AI」で顕在化した AIに伴うリスクと 日米政府の対応

江夏 あかね

Ⅰ 要 約 Ⅰ

1. 米国の新興企業アンソロピックが 2026 年 4 月 7 日に公表した「クロード・ミュトス・プレビュー」（以下、プレビュー）をきっかけに、人工知能（AI）の歴史が新たな段階に進む兆しが見られている。アンソロピックは、同年 6 月 9 日に「クロード・ミュトス 5」（以下、ミュトス 5）及び「クロード・フェイブル 5」（以下、フェイブル 5）を公表した。ミュトス 5 と同等の性能を有する AI は、「ミュトス級 AI」と称されており、他の AI モデルに比して、サイバーセキュリティ上の弱点や欠陥、すなわちシステムの脆弱性を発見する能力が極めて高い点が特徴とされている。
2. 米国政府による主な対応としては、（1）財務長官及び米連邦準備制度理事会（FRB）議長による米銀大手への注意喚起、（2）トランプ大統領による最先端の AI を事前検証する大統領令への署名、（3）政府によるフェイブル 5 とミュトス 5 の提供差し止め及び許可、が挙げられる。
3. 日本政府による主な対応としては、（1）政府全体としての「Project YATA-Shield」のとりまとめ、（2）金融庁・日本銀行による「フロンティア AI による脅威変化を踏まえた金融機関等の短期的な対応」の要請、（3）内閣官房 国家サイバー統括室（NCO）による「政府機関等のためのガイドライン」の改定、（4）政府による「AI 基本計画」の改定・閣議決定、が挙げられる。
4. ミュトス級 AI 等の先端 AI をめぐる日米の政策面から見た今後の主な論点としては、（1）イノベーション促進とリスク対応を両立する政策を迅速かつ適切に講じ続けられるか、（2）民間セクターが AI ガバナンス体制を自主的・自律的に構築できるように適切に関与・支援できるか、（3）自国の競争力を確保しつつ、サイバーセキュリティや安全保障面で国際的に適切に協調・連携していけるか、が挙げられる。

野村資本市場研究所 関連論文等

- ・江夏あかね「AI ガバナンスの概念と政府・企業・投資家による対応ーリスク調整後の投資パフォーマンスの向上に向けてー」『野村サステナビリティクォーターリー』2024 年秋号。
- ・江夏あかね「AI 法の成立と日欧米の現在地」『野村サステナビリティクォーターリー』2025 年夏号。
- ・江夏あかね「米国における AI 法の策定に向けた大統領令」『野村資本市場クォーターリー』2026 年冬号。

I AI の歴史を新たな段階に進める「ミュトス級 AI」

米国の新興企業アンソロピック¹が 2026 年 4 月 7 日に公表した「クロード・ミュトス・プレビュー」（以下、プレビュー）をきっかけに、人工知能（AI）の歴史が新たな段階に進む兆しが見られている²。アンソロピックは、プレビューの公表後、同年 6 月 9 日に同モデルの後継となる「クロード・ミュトス 5」（以下、ミュトス 5）及びミュトス 5 に安全策を組み込んだ一般向けモデル「クロード・フェイブル 5」（以下、フェイブル 5）を公表した（図表 1 参照）。ミュトス 5 と同等の性能を有する AI は、「ミュトス級 AI」と称されている³。

ミュトス級 AI は、フロンティア AI⁴とも称される最先端の AI モデルの 1 つであり、他の AI モデルに比して、サイバーセキュリティ上の弱点や欠陥、すなわちシステムの脆弱性を発見する能力が極めて高い点が特徴とされている⁵（図表 2 参照）。そのため、各国政府は、ミュトス級 AI がサイバー攻撃等に悪用されるリスクや安全保障の要素も踏まえた対応に迫られている（図表 3 参照）。

本稿では、米国と日本政府におけるミュトス級 AI を始めとした最先端の AI モデルへの対応状況を概観した上で、両国の政策面から見た今後の論点を考察する。

図表 1 アンソロピックによる先端 AI の概要と提供状況

概要			
項目	クロード・ミュトス 5		クロード・フェイブル 5
名称の意味	ギリシャ語由来「神話」		ラテン語由来「寓話」
提供先	米国 100 以上の企業・組織(2026 年 6 月末時点)		企業顧客、有料契約する消費者
用途	サイバー防御や生物学分野の研究		ソフトウェア開発や知識労働の自動化等
安全対策	防御に使うため、安全措置の一部を解除		サイバー攻撃や生物兵器に関連する質問は回答拒否
提供状況			
	クロード・ミュトス		クロード・フェイブル 5
	クロード・ミュトス 5	クロード・ミュトス・プレビュー	
2026 年 6 月 12～30 日	米政府指示で停止 →6 月 26 日に米国内のみ再開	日本含む一部企業や政府機関に限定提供	米政府指示で停止
2026 年 7 月 1 日以降	日本等海外向けは段階的に拡大へ	日本含む一部企業や政府機関に限定提供	7 月 7 日まで:世界の定額プラン契約者に提供 8 日から:従量課金に移行

（注） 2026 年 7 月 1 日時点。日付は米国時間に基づく。

（出所）「アンソロピックの AI『Fable』2 日再開 禁輸解除も定額は 1 週間限定」『日本経済新聞』2026 年 7 月 1 日、より野村資本市場研究所作成

¹ アンソロピックは、オープン AI 幹部のダリオ・アモデイ氏（最高経営責任者〔CEO〕）と妹のダニエラ・アモデイ（社長）らが 2021 年に共同設立したパブリック・ベネフィット・コーポレーション（PBC）。本拠地を米国カリフォルニア州サンフランシスコに置き、安全で信頼できる AI の開発を行っている。同社の中核プロダクトは、対話型 AI モデルのクラウド（Claude）で、用途に応じて複数のモデルを展開している。

² Anthropic, “Project Glasswing: Securing Critical Software for the AI Era,” April 7, 2026; Anthropic, “Assessing Claude Mythos Preview’s Cybersecurity Capabilities,” April 7, 2026.

³ 「ミュトス級 AI とは サイバー攻撃能力が突出」『日本経済新聞』2026 年 6 月 29 日。

⁴ フロンティア AI をめぐって世界的に浸透している定義はない。ただし、英国政府の報告書によると、広範なタスクを実行可能で、現在最も進んだ AI のモデルが有する能力に匹敵、またはそれを超える能力を持つ高性能な汎用型 AI モデル（Department for Science, Innovation & Technology, “Capabilities and Risks from Frontier AI,” October 2023、日本貿易振興機構「英政府、AI セーフティサミットに先立ち、フロンティア AI に関する報告書発表」『ビジネス短信』2023 年 10 月 31 日）。

⁵ アンソロピックは、「ソフトウェアの脆弱性を発見・悪用する能力において、最も熟練した人間を除く全ての人間を凌駕できるレベルのコーディング能力に達している」と説明している（前掲注 2 参照）。

図表 2 ミュトス級 AI の機能の比較

項目	アンソロピック／ クロード・ミュトス 5 (2026 年 6 月 9 日)	オープン AI／ GPT5.5 (2026 年 4 月 23 日)	グーグル／ Gemini (2026 年 5 月 19 日)
長時間の開発作業で バグ修正	80.3%	58.6%	55.1%
パソコン操作での 自律的な開発	88.0%	83.4%	76.2%

(注) Gemini は Gemini3.5 フラッシュ。括弧は発表日。

(出所) 「ミュトス級 AI とは サイバー攻撃能力が突出」『日本経済新聞』2026 年 6 月 29 日、より野村資本市場研究所作成

図表 3 ミュトス級 AI をめぐる日米政府等に関する主な動向

時期	詳細
2026 年 4 月 7 日	アンソロピック、「クロード・ミュトス・プレビュー」の公表とともに、同モデルを利用できる約 50 の企業・組織が参加する「プロジェクト・グラスウィング」の発足を発表
2026 年 4 月 7 日	スコット・ベッセント米財務長官、ジェローム・パウエル米連邦準備制度理事会 (FRB) 議長 (当時)、米大手銀行最高経営責任者 (CEO) と緊急会合を開催し、AI の進化が金融システムにもたらす新たなサイバーセキュリティリスクについて注意喚起
2026 年 4 月 13 日	英国政府の研究機関の AI セキュリティ・インスティテュート (AISI)、プレビューのサイバーセキュリティ能力評価を公表。同モデルが従来モデルを上回る性能を示し、人間の専門家が数日を要する多段階サイバー攻撃を自律的に実行できることを確認
2026 年 4 月 16 日	米国証券協会 (ASA)、プレビューが米国証券取引委員会 (SEC) の市場監視システム「統合監視体制 (CAT)」を攻撃し、金融危機を招く恐れがあるとベッセント米財務長官向けの書簡で指摘
2026 年 4 月 20 日	自由民主党、AI によるサイバー攻撃への対策を政府に要請
2026 年 4 月 24 日	金融庁、「AI 脅威に対する金融分野のサイバーセキュリティ対策強化に関する官民連携会議」を開催
2026 年 5 月 7 日	国際通貨基金 (IMF)、最先端 AI が金融の安定性を脅かす可能性を指摘。安全対策や被害時の復旧を含む総合的な計画策定を金融機関に促すよう、政策当局者に注意喚起
2026 年 5 月 12 日	高市早苗内閣総理大臣、閣僚懇談会で AI を悪用したサイバー攻撃への対策を指示
2026 年 5 月 14 日	金融庁、「AI 脅威に対する金融分野のサイバーセキュリティ対策強化に関する官民連携会議」の作業部会を開催
2026 年 5 月 14 日	自由民主党、高市内閣総理大臣にサイバー対処能力向上等を提言。高度自律型 AI の脅威への対策も特記
2026 年 5 月 18 日	日本政府、「AI 性能の高度化を踏まえたサイバーセキュリティ対策に関する関係省庁会議」を開催。「AI 性能の高度化を踏まえたサイバーセキュリティ対策の強化について—Project YATA-Shield—」を公表
2026 年 5 月 19 日	フランス・パリで開催された G7 財務相・中央銀行総裁会議、AI に関連するサイバーリスクへの具体策をまとめることで合意
2026 年 5 月 22 日	金融庁と日本銀行、金融機関等に対し、「フロンティア AI による脅威変化を踏まえた金融機関等の短期的な対応」について、要請
2026 年 5 月 22 日	アンソロピック、プロジェクト・グラスウィングの状況報告として、1 万件超の深刻な脆弱性を発見した旨を公表
2026 年 6 月 2 日	ドナルド・トランプ大統領、最先端の AI を事前検証する大統領令に署名
2026 年 6 月 2 日	アンソロピック、プロジェクト・グラスウィングの拡大を発表。対象を従来の約 50 組織に加え、新たに日本を含む 15 カ国以上の約 150 組織に拡大
2026 年 6 月 9 日	アンソロピック、「クロード・フェイブル 5」及び「クロード・ミュトス 5」の提供を開始
2026 年 6 月 12 日	米政府、外国籍の個人によるフェイブル 5 及びミュトス 5 へのアクセスを停止するよう求める輸出管理指令を发出
2026 年 6 月 17 日	フランスのエマニュエル・マクロン大統領、主要 7 カ国首脳会議 (G7 サミット) の成果として、AI に関する国際協調の進展を強調
2026 年 6 月 26 日	アンソロピック、米商務省からミュトス 5 の輸出規制を一部解除し、米政府が許可した約 100 の企業及び政府機関に対するアクセス再開の許可を受け、ミュトス 5 の許可された範囲内の提供を再開
2026 年 7 月 1 日	アンソロピック、米商務省から前日にフェイブル 5 及びミュトス 5 の輸出管理規制の解除通知を受け、フェイブル 5 も提供を再開
2026 年 7 月 14 日	日本政府、AI 基本計画を改定し、閣議決定。高性能 AI を活用した取り組みの強化も含めサイバー対処能力を不断に強化する方針を明記
2026 年 7 月 14 日	ホワイトハウス、トランプ大統領が署名した 2026 年 6 月 2 日の大統領令に基づき、サイバーセキュリティの脆弱性をめぐる連携を可能とする一元窓口 (クリアリングハウス) として、「ゴールドイーグル」の創設を公表

(出所) 各種資料、より野村資本市場研究所作成

Ⅱ 米国政府による対応

米国では、プレビュー公開同日の 2026 年 4 月 7 日、スコット・ベッセン特米財務長官及びジェローム・パウエル米連邦準備制度理事会（FRB）議長（当時）が米大手銀行最高経営責任者（CEO）に AI の進化に伴う新たなサイバーセキュリティリスクへの注意喚起を行った⁶。

2026 年 6 月には最先端の AI モデルをめぐり、政府の関与強化を示す動きが複数あった。ドナルド・トランプ大統領は同月 2 日、最先端の AI モデルを提供開始の 30 日前までに検証を行う大統領令に署名した⁷。その 1 週間後となる同月 9 日、アンソロピックはフェイブル 5 及びミュトス 5 の提供を開始した⁸。しかし、アンソロピックは同月 12 日、両モデルが政府による輸出管理対象の指定を受けたとして、提供停止を公表した⁹。

その後、アンソロピックは同月 26 日、政府の許可を受け、米国企業・組織を対象としてミュトス 5 の提供を再開するとした¹⁰。さらに、米商務省が同月 30 日、フェイブル 5 及びミュトス 5 について輸出管理指定を撤回すると通知したことを受け、アンソロピックは同日、フェイブル 5 についても翌日の 7 月 1 日より世界の有料顧客を対象に提供を再開する旨を公表した¹¹。

本章では、米国政府による主な対応として、（1）財務長官及び FRB 議長による米銀大手への注意喚起、（2）トランプ大統領による最先端の AI を事前検証する大統領令への署名、（3）政府によるフェイブル 5 とミュトス 5 の提供差し止め及び許可、について概観する。

1. 財務長官及び FRB 議長による米銀大手への注意喚起

スコット・ベッセン特財務長官とジェローム・パウエル FRB 議長（当時）はプレビューの公表同日の 2026 年 4 月 7 日、コロンビア特別区（ワシントン DC）の財務省本部に米大手銀行 CEO を緊急会議に招集した。会議は、プレビューがサイバーリスクの増大を招く可能性を懸念し、銀行各社がミュトスや同様のモデルによって生じる可能性のある将来的なリスクを認識し、システムを守るための予防措置を講じていることを確認するために開催された。

その後、ゴールドマン・サックス等の金融機関は連邦政府・当局と、プレビューによって露呈した脆弱性への体系的な対応策を協議するとともに、一部でプレビューのアクセス

⁶ 「Anthropic 最新 AI 巡り緊急会合、財務長官と FRB 議長が米銀 CEO を招集」『ブルームバーグ』2026 年 4 月 10 日、「米財務長官と銀行 CEO、アンソロピック AI で緊急会合 金融リスク警戒」『日本経済新聞』2026 年 4 月 10 日、「米財務長官と FRB 議長、アンソロピック最新 AI 巡り銀行幹部に警告」『ロイター』2026 年 4 月 10 日。

⁷ The White House, “Promoting Advanced Artificial Intelligence Innovation and Security: Presidential Actions, Executive Orders,” June 2, 2026.

⁸ Anthropic, “Claude Fable 5 and Claude Mythos 5,” June 9, 2026.

⁹ Anthropic, “Statement on the US Government Directive to Suspend Access to Fable 5 and Mythos 5,” June 12, 2026.

¹⁰ 「アンソロピック、AI『ミュトス』を米 100 社に提供再開へ 米政府許可」『日本経済新聞』2026 年 6 月 27 日。

¹¹ Anthropic, “Redeploying Fable 5,” June 30, 2026.

権も得つつ、アンソロピックやセキュリティベンダーと連携しながら、サイバー及びインフラの耐性の補強を続けている¹²。しかし、2026 年 5 月後半には、FRB と米通貨監督庁（OCC）が大手銀行に対する一部のサイバー関連検査を見合わせていると報じられた¹³。銀行側に分析を進める時間を与える狙いがあるとみられている。

2. トランプ大統領による最先端の AI を事前検証する大統領令への署名

ドナルド・トランプ大統領は 2026 年 6 月 2 日、AI のイノベーション促進及びサイバーセキュリティ強化に関する大統領令に署名した（図表 4 参照）。トランプ政権はそれまで、AI に対する規制に慎重なスタンスを維持していたものの¹⁴、最先端 AI の登場などを受けて、同大統領令では、開発企業による自主的な協力を前提に、最新モデルの提供開始 30 日前までに米政府が安全性に関するリスクをチェックする仕組みを設けることとされた。

図表 4 大統領令 14409 号「先端 AI のイノベーション促進とセキュリティ強化」の概要

本大統領令の目的は、先端 AI 能力がもたらす新たな国家安全保障上の考慮事項に対処しつつ、AI の革新を促進することにある。主な規定は以下のとおり。

第 1 条(目的)

本条では、2025 年 12 月の大統領令の文言を踏まえ、米国は「過度に負担の大きい規制によってイノベーションを抑え込まない」ことを改めて表明する。

第 2 条(米国の先端 AI システムの高度化)

30 日以内に、複数の機関（戦争長官、国土安全保障省、サイバーセキュリティ・インフラセキュリティ庁〔CISA〕等）は、それぞれの連邦システム（戦争省の情報システムを含む）のサイバー防衛を優先しなければならない。国土安全保障長官は、他の関係者と協議の上、以下を実施する拘束力のある運用指令を公表する必要がある。

- (i) 国の重要機能を守るため、連邦民間政府の情報システムのサイバー防衛を拡大し優先すること。
- (ii) AI を活用した防御ツールを強化する連邦のサイバープログラムやサービスを設置または拡充すること。
- (iii) 機関、州・地方当局、重要インフラ事業者（例：地方病院、地域銀行、地方公益事業者）が、適切な場合には対象フロンティアモデルを含むサイバーセキュリティ・ツールやサービスにアクセスしやすくすること。

また 30 日以内に、財務長官は他の関係者と協議し、産業界および重要インフラ事業者との任意協力のもとで AI サイバーセキュリティ・クリアリングハウスを設置する。これは、ソフトウェアの脆弱性スキャンを調整し、重複や競合を解消しつつ、脆弱性の発見・検証を行い、修正パッチの配布と優先順位付けを管理する役割を担う。

さらに 60 日以内に、行政管理予算局(OMB)は国家サイバー局長及び CISA 局長と連携し、先進 AI の脆弱性検出の開発者に向けて振り向け可能な、利用可能かつ関連する資金を有する連邦補助金プログラムがあるかを判断しなければならない。加えて 60 日以内に、人事管理局は、米国テックフォースの情報サイバーセキュリティ専門職の採用・配置ルートを拡大しなければならない。

¹² 「ゴールドマン、Mythos活用ーアンソロピックと連携しAIインフラ補強」『ブルームバーグ』2026年4月14日、「焦点：『ミュトス』登場で米銀はシステム修正に奔走、サービスに影響も」『ロイター』2026 年 5 月 13 日、「『ミュトス』対応で大手銀へのサイバー防御検査を延期ー米金融当局」『ブルームバーグ』2026 年 5 月 20 日。

¹³ 「『ミュトス』対応で大手銀へのサイバー防御検査を延期ー米金融当局」『ブルームバーグ』2026 年 5 月 20 日。

¹⁴ トランプ大統領は 2025 年 1 月 23 日、前政権による大統領令を撤回し、AI をめぐる規制緩和を指示する大統領令を公表した。さらに、同年 12 月 11 日には「AI に関する国家政策枠組みの確立」と題した大統領令に署名した。同大統領令には、AI をめぐる競争力を強化すべく、イノベーションを妨げる恐れのある、過度な負担となる州法を牽制し、連邦政府レベルの AI 法の枠組みを策定すること等が明記されている。さらに、トランプ政権は 2026 年 3 月 20 日、AI 分野における国際競争力の確保及び米国民の繁栄のためにイノベーションの促進を図るとともに、AI が国民生活に与える影響への懸念に対処すべく「AI に関する国家政策枠組み」を公表した。

図表 4 大統領令 14409 号「先端 AI のイノベーション促進とセキュリティ強化」の概要（続き）

第 3 条(保護対象フロンティアモデルの展開)

60 日以内に、各省庁は以下を実施する。

- ・ 高度 AI モデルのサイバー能力を評価し、特定のモデルを本大統領令上の「対象フロンティアモデル」と指定するための、機密扱いのベンチマーク手続きを策定・維持する。判断は、他者と協議の上、国家安全保障局 (NSA) 長官が行う。
- ・ 開発者が以下を実施できる任意の枠組みを設計する。
 - ・ 開発中のモデルが対象フロンティアモデルに該当するかを連邦政府と協議する。
 - ・ 適切な保護措置の下、他の信頼できるパートナーに公開する最大 30 日前まで、対象フロンティアモデルへの早期アクセスを連邦政府に提供する。
 - ・ セキュリティを確保しながらイノベーションを促進し、重要インフラのサイバーセキュリティを強化するため、対象フロンティアモデルへの早期アクセスを有する信頼できるパートナーの選定について、連邦政府と協力する。

本条は、新たな AI モデル、フロンティアモデルを含むモデルの開発、公表、公開、配布に関して、義務的な政府ライセンス、事前承認、許可制度を設けるものではないことを明確にしている。

第 4 条(刑事執行)

司法長官は、AI を用いてコンピュータシステムに違法にアクセスしたり、損害を与えたりする者、またはそのような違法アクセスの際に AI を使用して他の犯罪を助長する者に対し、既存の連邦刑事法(コンピュータ詐欺、電信詐欺、身分詐欺)に基づく取締りを優先しなければならない。

これには、公的または私的な情報システムへの侵害や、その後に犯罪的または違法目的で使用するデータ・情報への不法アクセスに AI エージェントを用いる場合も含まれる。

第 5 条(一般規定)

本条には、一般的な留保規定として、本大統領令が既存法令、権限等に影響を与えないものとされている。

大統領令の公表費用は、戦争省が負担する。

(出所) The White House, “Promoting Advanced Artificial Intelligence Innovation and Security: Presidential Actions, Executive Orders,” June 2, 2026; Maynard Nexsen, “White House Unveils Executive Order Promoting Artificial Intelligence Innovation and Security,” June 4, 2026、より野村資本市場研究所作成

同大統領令の主な柱としては、(1) サイバー攻撃等に応用し得る高度な AI モデルを「対象フロンティアモデル」(Covered Frontier Model)として審査対象と指定、(2) 対象フロンティアモデルの開発企業は提供開始の 30 日前までに米政府が安全性を検証できるよう、情報等を提供、が挙げられる。一方で、AI 開発企業に求める内容はあくまでも自主性に基づくもので、政府が強制的なライセンスや事前承認、許可要件の創設を求めるものとして解釈しないように定めている。併せて、政府と AI 開発業界や重要インフラ運営企業と連携し、ソフトウェアの脆弱性スキャンを調整し、重複や競合を解消しつつ、脆弱性の発見・検証を行い、修正パッチの配布と優先順位付けを管理する組織として「AI サイバーセキュリティ・クリアリングハウス」を 30 日以内に設立する旨が記された¹⁵。

3. 政府によるフェイブル 5 とミュトス 5 の提供差し止め及び許可

アンソロピックが 2026 年 6 月 9 日に提供を開始したフェイブル 5 とミュトス 5 はその 3 日後の同月 12 日に、米政府が外国籍の個人による両モデルへのアクセスの停止を求める輸出管理上の命令を発出したことを受け、技術的に国籍をリアルタイムで正確に判別することが困難であったため、全世界で全ユーザーに対する提供が停止となった。輸出管理上

¹⁵ ホワイトハウスは 2026 年 7 月 14 日、サイバーセキュリティの脆弱性をめぐる連携を可能とする一元窓口(クリアリングハウス)として、「ゴールドイーグル」の創設を公表した(The White House, “White House Launches Gold Eagle Initiative for Unprecedented Cybersecurity Vulnerability Coordination,” July 14, 2026)。

の命令とは、米国輸出管理規則（EAR）¹⁶に基づき、安全保障上で懸念される品目に適用される規制を指していると解釈される。

アンソロピックは、両モデルの提供停止について、政府からの書簡には国家安全保障上で何か懸念されているか具体的な内容が記されていなかったとした上で、フェイブル 5 を回避する手法、いわゆる「ジェイルブレイク」の方法を把握したいと考えているようだと説明した¹⁷。ジェイルブレイクとは、一般的に、AI モデルに本来の安全対策を回避させて、通常は出さないはずの情報や能力を引き出す手法であり、同社の発表ではフェイブル 5 の安全対策をすり抜けて、サイバーセキュリティに関する情報や能力を引き出すことが問題として想定されている。

アンソロピックは、（1）ジェイルブレイクの問題は以前から知られていた軽微な脆弱性に近く、他の公開モデルでも見つかる程度のもので危険性は限定的、（2）どのモデルでもジェイルブレイクへの耐性を 100%にすることは不可能であるため、突破されても被害を小さく抑える多層防御が重要、（3）今般の政府の措置は、仮に業界全体に適用された場合に新しいモデルの公開そのものがほぼ不可能になるほど厳格、との見方を明らかにした。

その後、同社は 6 月 26 日、米政府の許可を受け、米国企業・組織を対象としてミュトス 5 の提供の再開を公表したのに続き、同月 30 日にはフェイブル 5 及びミュトス 5 ともに輸出管理指定の解除の通知を米政府から受けた旨を公表した¹⁸。同社は、危険な利用を検知・遮断する安全性分類器を強化し、必要に応じて安全な別モデルに自動的に切り替える仕組みを導入した。また、誤検知による利用制限の可能性は残るものの、悪用防止を優先する設計とした。併せて、同社は今後、AI モデルの安全対策を回避するジェイルブレイクの深刻度を評価する共通基準の整備や、米政府との事前評価、情報共有、共同研究を通じて、より持続的な安全運用体制を構築する方針とした。そして、フェイブル 5 についても 2026 年 7 月 1 日より世界の有料顧客向けに提供を再開した。

Ⅲ 日本政府による対応

日本政府においても、2026 年 4 月 7 日のプレビューの 11 公表を契機として、急速に対応が進められている。自由民主党は 2026 年 4 月 20 日、高度自律型 AI をめぐる対策の強化を議論し、日本版「プロジェクト・グラスウィング」に取り組む必要性を指摘した¹⁹。

金融庁は 2026 年 4 月 24 日、「AI 脅威に対する金融分野のサイバーセキュリティ強化に関する官民連携会議」を開催し、内閣官房 国家サイバー統括室（NCO）、日本銀行、3 メ

¹⁶ 米国輸出管理規則（EAR）は、軍事用としても非軍事用としても利用可能な「デュアルユース品目」と呼ばれる民生品の輸出に適用される規則で、商務省産業安全保障局（BIS）が管理・規制の執行を管轄している。EAR 規制対象品目は、物品（服、建築資材、回路基板、自動車部品等）、ソフトウェア、技術をはじめ多岐にわたり、米国からの輸出のみでなく、再輸出の取引にも適用される（日本貿易振興機構「米国 貿易管理制度」2025 年 9 月 22 日）。

¹⁷ 前掲注 9 参照。

¹⁸ 前掲注 11 参照。

¹⁹ 高度自律型 AI とは、人間による細かい指示を必要とせず高度で複雑なタスクを実行できる AI システム（自由民主党「高度自律型 AI の対策強化を 最新 AI『ミトス』巡り議論」2026 年 4 月 21 日）。

ガバンク及び東京証券取引所が出席の下、金融の「日本版プロジェクト・グラスウィング」として事務方レベルで議論を深める作業部会の設立等を議論した²⁰。その後、同庁は 2026 年 5 月 14 日に作業部会を開催し、同月 22 日には、日本銀行と連名で「フロンティア AI による脅威変化を踏まえた金融機関等の短期的な対応」（以下、「金融機関等の短期的な対応」）を要請した。

一方、日本政府全体としても、2026 年 5 月 18 日に「AI 性能の高度化を踏まえたサイバーセキュリティ対策に関する関係省庁会議」を開催し、政府全体の対策パッケージとして、「AI 性能の高度化を踏まえたサイバーセキュリティ対策の強化について－Project YATA-Shield－」（以下、「Project YATA-Shield」）を公表している²¹。

さらに、NCO は 2026 年 6 月 12 日、政府機関がサイバーセキュリティ対策を講じる際の運用を定めた「政府機関等の対策基準策定のためのガイドライン（令和 7 年度版）」（政府機関等のためのガイドライン）を一部改定し、高性能 AI の悪用などのサイバー攻撃の高度化・自動化等を踏まえ、脆弱性対策（セキュリティパッチ適用等）の強化等を示した²²。

さらに、政府は 2026 年 7 月 14 日、AI 政策の指針となる「AI 基本計画」を改定し、閣議決定した²³。同計画には、新たに高性能 AI を活用した取り組みの強化も含めサイバー対処能力を不断に強化する旨が記された。

本章では、日本政府による主な対応として、（1）政府全体としての「Project YATA-Shield」のとりまとめ、（2）金融庁・日本銀行による「金融機関等の短期的な対応」の要請、（3）NCO による「政府機関等のためのガイドライン」の改定、（4）政府による「AI 基本計画」の改定・閣議決定、について概観する。

1. 政府全体としての「Project YATA-Shield」の取りまとめ

政府全体としての対策パッケージとして取りまとめられた「Project YATA-Shield」では、実施する施策として、（1）重要インフラ²⁴事業者²⁵等及び政府機関等への対応、（2）脆弱性の発見・修正等の対応、が挙げられている（図表 5 参照）。

²⁰ 金融庁「『AI 脅威に対する金融分野のサイバーセキュリティ対策強化に関する官民連携会議』の開催」『アクセス FSA』2026 年 5 月。

²¹ 内閣官房 国家安全保障局他「AI 性能の高度化を踏まえたサイバーセキュリティ対策の強化について～Project YATA-Shield～」2026 年 5 月 18 日。

²² 内閣官房 国家サイバー統括室「政府機関等の対策基準策定のためのガイドライン（令和 7 年度版）」2025 年 7 月 1 日（2026 年 6 月 12 日一部改定）。

²³ 閣議決定「人工知能基本計画－日本 AX、より強く、より豊かに－」2025 年 7 月 14 日。

²⁴ 重要インフラは、他に代替することが著しく困難なサービスを提供する事業が形成する国民生活及び社会経済活動の基盤であり、その機能が停止、低下または利用不可能な状態に陥った場合に、日本の国民生活または社会経済活動に多大なる影響を及ぼすおそれが生じるものを指す。重要インフラのサイバーセキュリティに係る行動計画では、「重要インフラ分野」として、「情報通信」、「金融」、「航空」、「空港」、「鉄道」、「電力」、「ガス」、「政府・行政サービス（地方公共団体を含む）」、「医療」、「水道」、「物流」、「化学」、「クレジット」、「石油」及び「港湾」の 15 分野を特定している（国家サイバー統括室「重要インフラ対策関連」）。

²⁵ 重要インフラ事業者は、サイバーセキュリティ基本法第 3 条第 1 項に規定する重要社会基盤事業者を指す。国民生活及び経済活動の基盤であって、その機能が停止し、または低下した場合に国民生活または経済活動に多大な影響を及ぼすおそれが生ずるものに関する事業を行う者（国家サイバー統括室「重要インフラ対策関連」）。

1 点目では、4 つの論点（重要インフラ事業者等への注意喚起、金融分野での先行的な取組及び他分野への展開、人材育成支援、政府機関等の情報システムにおける対応）が挙げられた。特に、金融分野等での先行的な取組の実施及び他分野への展開として、金融分野での官民連携の枠組みの設置（2026 年 4 月 24 日）や経済産業省の所管分野における官民での認識共有を図るための意見交換（同年 5 月 1 日）を取り上げ、各重要インフラ分野において、分野特性も踏まえて必要な対応を検討・実施できるよう、官民連携を推進する旨が記された。

2 点目では、5 つの論点（外国政府機関やビッグテック等との更なる連携、ソフトウェア・ベンダへの注意喚起、AI セーフティ・インスティテュート〔AISI〕による技術支援等、技術開発の推進、高性能 AI を活用したサイバー対処能力強化）が挙げられた。

「Project YATA-Shield」の取りまとめ・公表と併せて、重要インフラ事業者等、政府機関等、ソフトウェア・ベンダへの注意喚起も公表・実施された²⁶（図表 6 参照）。具体的には、（1）重要インフラ事業者等に対しては、経営層のリーダーシップの下での対策の実施、（2）政府機関等に対しては、組織トップのリーダーシップの下、対応の徹底を要請、（3）ソフトウェア・ベンダに対しては、高性能 AI も活用しながら脆弱性の早期発見・対応、が掲げられた。

図表 5 Project YATA-Shield の概要

・ フロンティア AI モデルによるサイバーセキュリティ性能が向上する中においても、我が国のサイバーセキュリティが確保されるよう、 政府全体としての対策パッケージを取りまとめ	
基本的な認識・考え方	
重要インフラ事業者等・政府機関等が取るべき対応	・ 発見された脆弱性のパッチ適用やリスク緩和措置 を速やかに実施(リスクベース) ・ 基本的な対策、多層防御の実施、インシデント発生時の備え 等 ※英国・米国政府の注意喚起も参考に
脆弱性を発見する AI の進化	・ Anthropic (Project Glasswing) : Mythos へのアクセスを、 ビッグテックや重要インフラ等 に限定。 ・ OpenAI: GPT-5.5-Cyber へのアクセスを、 一部の認証者 に限定して付与。
実施する施策	
重要インフラ事業者等・政府機関等への対応	① 重要インフラ事業者等への注意喚起等 ② 金融分野での先行的な取組及び他分野への展開 ③ 人材育成支援 ④ 政府機関等の情報システムにおける対応
脆弱性の発見・修正等への対応	① 外国政府機関やビッグテック等との更なる連携 ② ソフトウェア・ベンダへの注意喚起 ③ AI セーフティ・インスティテュート(AISI)による技術支援等 ④ 技術開発の推進 ⑤ 高性能 AI を活用したサイバー対処能力強化

(注) 1. 太字・下線は原文に基づく。

2. YATA : Yielding Advanced Threat Awareness with AI (脅威の可視化) の頭文字。「正確に写す」という八咫鏡 (やたのかがみ) もあるように、「AI 性能の高度化に伴うサイバー脅威を正しく認識し、防御する／対応する」という趣旨。

(出所) 内閣官房 国家サイバー統括室「Project YATA-Shield」2026 年 5 月 18 日、より野村資本市場研究所作成

²⁶ 内閣官房 国家サイバー統括室他「AI 性能の高度化を踏まえたサイバーセキュリティ対策の強化について（重要インフラ事業者等に対する注意喚起）」2026 年 5 月 18 日、内閣官房国家サイバー統括室他「AI 性能の高度化を踏まえたサイバーセキュリティ対策の強化について（ソフトウェア・ベンダに対する注意喚起）」2026 年 5 月 18 日、内閣官房 国家サイバー統括室「AI 性能の高度化を踏まえたサイバーセキュリティ対策の強化について AI 性能の高度化を踏まえたサイバーセキュリティ対策の強化について」2026 年 5 月 18 日。

図表 6 関係機関への注意喚起

・ 対策パッケージ「Project YATA-Shield」の取りまとめ・公表を行い、 <u>重要インフラ事業者等、政府機関等、ソフトウェア・ベンダへの注意喚起</u> を公表・実施	
重要インフラ事業者等	・ <u>経営層のリーダーシップの下での対策の実施</u> → 必要な投資と捉えて、組織の<u>リスクマネジメント</u>として実施 ・ <u>基本的な対策の確実な実施等</u> 英: 基本的な対策、米: 隔離・復旧 が重要 → 資産管理、脆弱性対策、インシデント対応・復旧など(重要インフラ統一基準) → 実施状況の機動的な確認 ・ <u>高速化する脆弱性の発見・修正等への対応</u> → 脆弱性のリスク評価、パッチ適用・リスク緩和措置の速やかな実施
政府機関等	・ <u>組織のリーダーシップの下、対応の徹底を要請</u> ・ <u>基本的な対策の確実な実施等</u> 英: 基本的な対策、米: 隔離・復旧 が重要 → 資産管理、脆弱性対策、インシデント対応・復旧など(政府統一基準) → 実施状況の機動的な確認(各機関・国家サイバー統括室[NGO]による監督) ・ <u>脆弱性対策の強化</u> → パッチ管理・適用の運用設計の見直し、パッチ適用・リスク緩和措置の速やかな実施
ソフトウェア・ベンダ	・ <u>高性能 AI も活用しながら、脆弱性の早期発見・対応</u> ① リリース前のソフトウェア → <u>脆弱性を低減させた上でリリース</u> ② リリース後のソフトウェア → <u>脆弱性の把握、パッチの早期作成、顧客への早期提供</u>

(注) 太字・下線は原文に基づく。

(出所) 内閣官房 国家サイバー統括室「Project YATA-Shield」2026 年 5 月 18 日、より野村資本市場研究所作成

2. 金融庁・日本銀行による「金融機関等の短期的な対応」の要請

金融庁・日本銀行は 2026 年 5 月 22 日、「フロンティア AI」により、短期間に大量の脆弱性を発見され、修正プログラムであるパッチ²⁷を大量に提供される可能性を踏まえて、金融機関に求められる対応を整理し、短期的な対応として要請した。要請においては、(1) フロンティア AI への対応を経営課題として扱うこと、(2) 対応のパッチが提供された場合に備えた体制整備を行うこと、等が盛り込まれている(図表 7 参照)。

図表 7 フロンティア AI による脅威変化を踏まえた金融機関等の短期的な対応の概要

金融機関等においては、短期的には以下に掲げる対応を速やかに講じる必要がある。ただし、これらはあくまでも応急的措置であり、中長期的には脆弱性対応の自動化等への移行に取り組む必要がある。
1. フロンティア AI への対応を経営課題として扱う フロンティア AI がもたらす脅威の変化は、IT・サイバーセキュリティ部門にとどまらない課題であるため、経営トップは全社的な経営課題として扱う必要がある。各業務所管部門、リスク管理部門、IT・サイバーセキュリティ部門、財務部門等関係部門が横断的に連携して対応できるよう、経営トップとしてのコミットメントが不可欠である。
2. 優先的に対応すべきサービス/IT システムを特定する 優先的に対応すべきサービス/IT システムを特定し、リソースを重点的に配分する等、リスクベースで対応する必要がある。

²⁷ パッチは、完成したプログラムに対して、脆弱性等をなくするために後から配布される修正プログラムのこと。メーカーのウェブページ等で提供される(総務省 国民のためのサイバーセキュリティサイト「用語集 | パッチ」)。

図表 7 フロンティア AI による脅威変化を踏まえた金融機関等の短期的な対応の概要（続き）

3. 特定した資産の技術負債^(注1)を解消しておく

優先的に対応すべきサービス／IT システムにおいて、そのソフトウェア構成及びネットワーク構成等を再確認し、脆弱性発見時にパッチ適用対象を即時に特定できる状態を確保しておく必要がある。

4. パッチ適用に係る人的リソースを追加する

今後増加が見込まれる脆弱性への対応力を向上させるため、優先的に対応すべきサービス／IT システムに関連する実施計画を見直すとともに、他の IT システム部門からの支援等により、人的リソースの追加を検討すべきである。

5. ベンダーとの維持保守契約の内容を確認する

本邦金融機関等においては、自組織が所管する IT システムの維持保守をベンダーに委託しているケースが多く見られる。こうした状況を踏まえ、パッチ適用作業が現行の維持保守契約に含まれていること及び役割分担が明確であることを確認する必要がある。

6. パッチ適用プロセスをリスクベースにする

金融機関等においては、発見された脆弱性が自組織のサービス／IT システムに及ぼし得る影響を適切に把握するとともに、攻撃が成立する蓋然性も踏まえた評価を行うことが重要である。その上で、リスクベースで優先順位付けを行い、よりリスクの高い脆弱性から迅速かつ着実に対応すべきである。

7. パッチ適用以外の対策も強化する

優先的に対応すべきサービス／IT システムにおいて、パッチ適用そのものが困難である場合や、パッチ適用に要する期間の短縮が困難である場合には、早期に効果が期待できるクラウド型のウェブアプリケーション防御機能(WAF)等を用いた仮想パッチ^(注2)の適用やボット^(注3)対策の導入等、多層防御の強化を図る必要がある。

8. 優先サービス／IT システムの停止に備える

各種対策を徹底してもサイバー攻撃を防御できない可能性を前提に、サイバー攻撃により IT システムが停止する場合を想定するとともに、優先的に対応すべきサービス／IT システムを能動的に停止させざるを得ない場合についても経営トップはあらかじめ選択肢として検討しておくべきである。

9. 外部との連携を維持・強化する

フロンティア AI に関する情報は短期間に多数公開されることから、自組織のみでの網羅的な把握等は困難と想定される。このため、金融 ISAC^(注4)や各業界団体・コミュニティ、当局等からの情報に積極的にアクセスし、必要な情報の収集に努めるべきである。

- (注) 1. 技術負債（技術的負債）は、ソフトウェア開発等で、短期的な開発効率や納期を優先し、その場しのぎの不完全あるいは低品質な設計やコードを採用した結果、将来の保守や改修に追加的な負担をもたらす状態（「技術的負債」『IT 用語辞典』インセプト）。
2. 仮想パッチは、ソフトウェア・ベンダが提供する本来のセキュリティパッチをサーバー等に早急に適用することが難しい場合の暫定的なソリューション。
3. ボットは、コンピューターに感染し、そのコンピューターを、ネットワーク（インターネット）を通じて外部から操ることを目的として作成されたプログラム（情報処理推進機構「初めての情報セキュリティ対策のしおり」2012 年 1 月 30 日）。
4. 金融 ISAC は、日本の金融機関によるサイバーセキュリティに関する情報の共有及び分析を行い、金融機関の安全性の向上を推進することにより、利用者の安心/安全を継続的に確保することを目的として 2014 年 8 月に設立された。会員は金融機関等により構成されている。

(出所) 金融庁・日本銀行「『フロンティア AI による脅威変化を踏まえた金融機関等の短期的な対応』に係る要請について」2026 年 5 月 18 日、より野村資本市場研究所作成

3. NCO による「政府機関等のためのガイドライン」の改定

NCO は 2026 年 6 月 12 日、「政府機関等のためのガイドライン」を公表した。同ガイドラインは、政府機関等の情報セキュリティ水準を向上させるための統一的な枠組みである政府統一基準群の 1 つである。松本尚 内閣府特命担当大臣（サイバー安全保障）は同ガイドライン公表同日の記者会見にて、前述の「Project YATA-Shield」を踏まえて、セキュリティパッチの迅速な適用が必要な場合に、場合によっては情報システムの運用を一時停止するといったことなど、高性能 AI の悪用リスクに備えた脆弱性対策の強化を盛り込んでいく旨を強調した²⁸（図表 8 参照）。

²⁸ デジタル庁「松本大臣記者会見（令和 8 年 6 月 12 日）」2026 年 6 月 15 日。

図表 8 「政府機関等の対策基準策定のためのガイドライン」（抜粋）

改定事項	改定の主な内容
脆弱性対策(セキュリティパッチ適用等)の強化	・ 高性能 AI の悪用などのサイバー攻撃の高度化・自動化等をふまえ、 ・ 全情報システムについて、セキュリティパッチの適時の適用を前提とした運用設計（パッチマネジメント）を行う。 ・ サイバー攻撃の高度化・自動化等の状況をふまえ、運用設計を見直す。 ・ 迅速な適用の可否を判断した上で、脆弱性対策計画を策定・実施する。 ・ 迅速な適用が必要な場合、情報システムの運用を一時停止することも検討する。

（注） 太字は原文に基づく。

（出所）内閣官房 国家サイバー統括室「「政府機関等の対策基準策定のためのガイドライン（令和 7 年度版）」の一部改定（令和 8 年 6 月 12 日）のポイント」2026 年 6 月、より野村資本市場研究所作成

4. 政府による「AI 基本計画」の改定・閣議決定

政府は 2025 年 12 月 23 日に策定・閣議決定していた AI 基本計画を改定し、新たな計画（第 II 期 AI 基本計画）として 2026 年 7 月 14 日に閣議決定した（図表 9 参照）。

高市早苗内閣総理大臣は、第 II 期 AI 基本計画案を政府の AI 戦略本部で決定した同年 7 月 10 日、（1）高性能 AI は、日本の課題解決と国力強化に直結する一方、サイバー攻撃への悪用など、新たなリスクも懸念されている、（2）日本は、技術革新に向き合い、信頼できる AI で社会全体を駆動する AX（AI トランスフォーメーション）を進めなければならない、（3）本基本計画で「日本 AX」を掲げ、バーティカル AI²⁹とフィジカル AI³⁰の官民投資を、「『強く豊かな日本』投資枠」を活用して強力に推進する、と述べた³¹。

同計画は、政府が総合的かつ計画的に講ずべき施策として 4 つの項目（AI 利活用の加速的推進、AI 開発力の戦略的強化、AI ガバナンスの主導、AI 社会に向けた継続的変革）が挙げられている。特に、AI ガバナンスの主導の項目では、最先端 AI の存在も意識し、（1）「『Project YATA-Shield』に基づき、我が国のサイバーセキュリティを確保。重要インフラ事業者等への対応と脆弱性の発見・修正等に関する対応を実施」、（2）高性能 AI を活用した取組の強化も含めサイバー対処能力を不断に強化」、が掲げられた。

²⁹ バーティカル AI は、領域特化型で、データ、AI モデル、アプリを垂直統合したシステムで、専門性等を有することから、実際に「現場で使える AI」となっている（内閣府 人工知能戦略本部「バーティカル AI 領域別戦略 中間とりまとめ」第 5 回人工知能戦略本部資料、2026 年 7 月 10 日）。

³⁰ フィジカル AI は、身体を持って実世界で動く AI。ロボット、ヒューマノイドを始め、自動運転車、工場内で動く産業機械を制御する AI 等もフィジカル AI のカテゴリーに入る（産業技術総合研究所「フィジカル AI とは？ 科学の目でみる、社会が注目する本当の理由」『産総研マガジン』2026 年 5 月 13 日）。

³¹ 首相官邸「総理の一日 人工知能戦略本部」2026 年 7 月 10 日。

図表 9 第Ⅱ期 AI 基本計画の概要（抜粋）

全体	
項目	内容
基本構想	・ 急伸するエージェント型AIに適応し、計算資源や電力から、制度・政策まで日本のAI実装能力を抜本強化 ・ AIを前提として意思決定や業務の進め方を根底から見直す「AIトランスフォーメーション」(AX)を推進 ・ 人が人として創造すべき価値を探求し、「人間力」を増進しながら、「人とAIが協働する社会」を率先して実現 ・ 「信頼できるAI」で社会全体を駆動し、課題解決と国力強化を探求、日本独自の価値を創出
4つの原則	・ イノベーション促進とリスク対応の両立 ・ 挑戦と学習 ・ アジャイル(柔軟かつ迅速)な対応 ・ 内外一体での政策推進
4つの基本的な方針に基づく施策	政府一丸でAX推進に必要な体制を構築、あらゆる制度や運用を先導見直し
	1. AI利活用の加速的推進「AIを使う」: エージェント型AI等の積極適用、政府における率先導入 ・ 政府・自治体でのAIの徹底した利活用「ガバメントAI「源内」」の進化と普及、地方の伸び代を成長に転換(地域AX) ・ AI利活用促進による社会課題解決、新事業創出 パーティカルAI・フィジカルAI中心に社会実装加速 ・ データ戦略、更なるAI利活用に向けた仕組みづくり 政府・準公共・産業分野でデータ連携基盤構築 2. AI開発力の戦略的強化「AIを創る」: AI実装能力を高め、「AI主権」を確保する ・ AI研究開発・利用基盤の拡充・高度化 国家基盤として計算資源やデータ、安定した電力供給を確保 ・ 日本国内のAI開発力の強化 データセット創出、内外トップ人材確保、モデル開発高度化、評価基盤・テストベッド整備 ・ 信頼できるAI基盤モデル等の開発 日本の文化等を踏まえたAI開発・評価、マルチモーダル基盤モデル開発推進 ・ 日本の勝ち筋の追求 パーティカルAI(領域別戦略)、フィジカルAI(AIロボティクス、自動運転)、AI for Science 3. AIガバナンスの主導「AIの信頼性を高める」: AIエコシステムにおける「信頼」の構築で世界をリード ・ 責任あるアジャイル・ガバナンスの実現 AI法を含めた制度等の能動的かつ不断の見直し、Project YATA-Shield ・ ASEAN等グローバルサウス諸国を含めた国際協調 AISI機能強化、信頼できるAI共創に向け AIサミットの早期日本開催 4. AI社会に向けた継続的変革「AIと協働する」: 人が意思決定への責任を持ちながら、AXで日本を駆動 ・ AIを前提とした産業構造・雇用環境の構築 中小企業、地方含むAX推進(取組可視化等)、雇用への影響を調査研究 ・ AI社会における制度・枠組みの検討・実証 規制・制度改革、人とAIの責任関係、知財保護・利活用につながる透明性確保 ・ AI時代の人材の育成・確保 AI実装人材育成・確保、全ての国民のAIに関するリテラシー向上、リ・スキリング支援 ・ AI時代における人間力の向上 創造力、批判的思考、判断責任など人間力向上、人とAIの役割分担の探究
(参考)AI施策の方向性:AIガバナンスの主導(AIの信頼性を高める)	
主な拡充内容	
・ AI法第16条に基づく調査研究等をより機動的に運用。複雑化かつ深刻化するリスクにより実効的に対応するため、AI法を含めた制度等を能動的かつ不断に見直す。 ・ 国民の権利保護や透明性・公正性の確保するため、行政事務におけるAI利活用時の留意点等を整理するほか、広くAI利活用に伴う権利侵害に関する法的整理なども含め各種ガイドライン等を整備。 ・ 人工知能戦略本部にAI関連の知見を集約し、AI関連の制度や指針等の点検・見直しを促す仕組みを構築。 ・ 「Project YATA-Shield」に基づき、我が国のサイバーセキュリティを確保。重要インフラ事業者等への対応と脆弱性の発見・修正等に関する対応を実施。 ・ 高性能AIを活用した取組の強化も含めサイバー対応能力を不断に強化。 ・ AISIを中心として、AIモデルの評価、トレーサビリティ、ガードレール等の技術的制御、危機時の情報共有などを可能とする技術的評価能力を確立・強化。評価結果が制度・運用改善に接続される仕組みを整備。 ・ AIに関する標準化等に関して AISI国際ネットワーク等を活用によるAIガバナンスの構築を主導。各国と連携し、高性能AIのガバナンスに関する国際的な議論を主導。 ・ 日本が世界のAIイノベーションの結節点となるよう、AIサミットを早期に日本で開催し、AIでの国際的な協調・協力モデルを主導。	

(注) 1. 下線は原文に基づく。

2. 本基本計画は当面毎年変更、適切なベンチマーク（重要業績評価指標〔KPI〕等）を設定しモニタリング。

(出所) 内閣府「人工知能基本計画の概要」2026年7月14日、より野村資本市場研究所作成

IV 今後の論点

ミュトス級 AI を含めた先端 AI をめぐる日米政府の取り組みは、先端 AI の開発企業を抱える米国とそうではない日本という違いはあるものの、（1）安全保障や経済成長の観点から国家の将来に大きな影響がある要素として捉え、政府が迅速かつ包括的に対応策を講じている、（2）民間セクターに一定の関与をしつつ、連携しながら対策を進めている、（3）中央銀行とともに金融機関に対して注意喚起を行っている、といった共通点が見受けられる。

先端 AI をめぐる日米の政策面から見た今後の主な論点としては、（1）イノベーション促進とリスク対応を両立する政策を迅速かつ適切に講じ続けられるか、（2）民間セクターが AI ガバナンス体制を自主的・自律的に構築できるように適切に関与・支援できるか、（3）自国の競争力を確保しつつ、サイバーセキュリティや安全保障面で国際的に適切に協調・連携していけるか³²、が挙げられる。

特に、2 点目の AI ガバナンスは、一般に AI 利用によるリスクを管理・抑制しつつ、最大限の便益を得ることを目的とするものであり、日本政府においても事業者等に対して行動目標を示す等の取り組みを進めている³³（図表 10 参照）。金融市場の観点からも、投資家が評価項目や企業との対話（エンゲージメント）のトピックとして取り上げることも増えてきており、今後もさらに重要性・注目度が高まることが想定される（図表 11 参照）。

図表 10 AI ガバナンス構築に向けた行動目標一覧

分類	行動目標
1. 環境・リスク分析	1-1 便益／リスクの理解
	1-2 AI の社会的な受容の理解
	1-3 自社の AI 習熟度の理解
2. ゴール設定	2-1 AI ガバナンス・ゴールの設定
3. システムデザイン	3-1 ゴール及び乖離の評価及び乖離対応の必須化
	3-2 AI マネジメントの人材のリテラシー向上
	3-3 各主体間・部門間の協力による AI マネジメント強化
	3-4 予防・早期対応による利用者のインシデント関連の負担軽減
4. 運用	4-1 AI マネジメントシステム運用状況の説明可能な状態の確保
	4-2 個々の AI システム運用状況の説明可能な状態の確保
	4-3 AI ガバナンスの実践状況の積極的な開示の検討
5. 評価	5-1 AI マネジメントシステムの機能の検証
	5-2 社外ステークホルダーの意見の検討
6. 環境・リスクの再分析	6-1 行動目標 1-1～1-3 の適時の再実施

（出所）総務省・経済産業省「AI 事業者ガイドライン（第 1.2 版）別添（付属資料）」2026 年 3 月 31 日、より
野村資本市場研究所作成

³² フランスで開催された主要 7 カ国首脳会議（G7 サミット）の閉幕後の記者会見（2026 年 6 月 17 日）にて、フランスのエマニュエル・マクロン大統領は AI に関する成果として、国際協調が進展したことを挙げた上で、今後の方向性として（1）共通基準（スタンダード）の策定、（2）サイバーセキュリティや安全保障分野での情報共有、（3）AI の影響と適切な対応策の共有を提示し、民主主義国間で協力する枠組みを構築していく旨を表明した。2026 年 9 月に閣僚級レベルで進展を確認する予定とされている（「AI 規制の国際的なコンセンサスが広がった G7 サミット、マクロン大統領が国際協調の進展を強調」『ビジネス短信』日本貿易振興機構、2026 年 6 月 23 日）。

³³ AI ガバナンスに関する詳細は、江夏あかね「AI ガバナンスの概念と政府・企業・投資家による対応ーリスク調整後の投資パフォーマンスの向上に向けてー」『野村サステナビリティクォーターリー』2024 年秋号、を参照されたい。

図表 11 AI・サイバーセキュリティに関する要請や評価項目の例

朝日ライフアセットマネジメント

サステナビリティ重点分野の1つである情報セキュリティの KPI は、AI ガバナンスの構築状況、AI リスクの対応状況、偽・誤情報問題に関する危機管理の取り組み状況

アセットマネジメント One

資産運用会社として取り組むべき重要課題を示した「マテリアリティ・マップ」の項目の中にサイバーセキュリティ、AI/デジタルイノベーション

SOMPO アセットマネジメント

「2024 年度に議論した主なテーマ」の1つは、「生成 AI の影響について検討(産業構造の変化、労働者の付加価値に対する影響等)」

第一ライフグループ

責任投資のインテグレーション手法(株式・社債・融資)の評価の視点で、ガバナンス「DX 推進体制」の中に「サイバーリスク・AIの適切な活用体制等」

大和アセットマネジメント

「投資先企業のあるべき経営の姿」としてガバナンスのリスクマネジメントの中に、サイバーセキュリティの堅牢化や社員教育徹底等の対策を十分に講じていること

野村アセットマネジメント

経営トップのコミットメントを示す方針の策定と開示、平時の体制、有事の体制の3要素を特に重要視と記載

三井住友 DS アセットマネジメント

カシオ計算機のエンゲージメント事例で「エンゲージメントテーマ」の1つとして「不正アクセス被害への対応」

三井住友トラスト・アセットマネジメント

ESG 投資ポリシーの「ESG マテリアリティ」のガバナンスの中で AI 等の発達による情報セキュリティ全般に係るリスクに関する企業の管理態勢等を考慮と記載

(注) 機関投資家は 50 音順、社名は現社名。各社のウェブサイト公開している責任投資やサステナビリティ、スチュワードシップ等に関する報告書から抜粋・要約し、QUICK ESG 研究所作成。

(出所) 「AI 対策、ESG 投資の重要課題に 企業に社会的責任—サステナブル投資最前線」『日経ヴェリタス』2026 年 6 月 10 日、より野村資本市場研究所作成