

人工知能（AI）技術の利用と 消費者問題に関する専門調査会

令和8年10月22日（木）

内閣府 科学技術・イノベーション推進事務局

人工知能政策推進室



世界で最もAIを開発・活用しやすい国に向けて

- AI利活用で、日本の長年の課題である、**人口減少**、国内への**投資不足**、**賃金停滞**を解決
健康・医療、防災を含む**安全・安心な国民生活**、**安全保障**や平和構築にも貢献
- 日本のAI産業を振興することで、日本社会の持つ**潜在力の発揮を実現**、**デジタル赤字抑止**に貢献し、**国外市場への展開**も期待
- 技術進歩に伴い変動する**リスクに適時適切**に対応し、**人間中心のAIを堅持**
- **AIを基軸として、新たな経済発展と安全・安心な社会を構築**

主なメリット：自律的に業務を実行する「エージェント型AI」、現実世界でロボット等を動かす「フィジカルAI」、といった近時の技術進歩で、多様な可能性が拡大

効率化・
生産性向上
(自動化、最適化)

新事業・
新市場創造
(創薬、新素材)

社会課題解決
(農業、医療、介護)

包摂的成長
(中小企業、公共
サービス高度化)

生活の質の向上
(病気の早期発見、
自動運転)

イノベーション促進

イノベーションの促進とリスク対応の両立

リスク対応

主なリスク：AIの開発・利用の進展で、誤判断、ハルシネーション、サイバーセキュリティといったAIの有する技術的リスクから「人との協働」に関する社会的リスクへ拡大

差別・偏見の助長

犯罪への利用

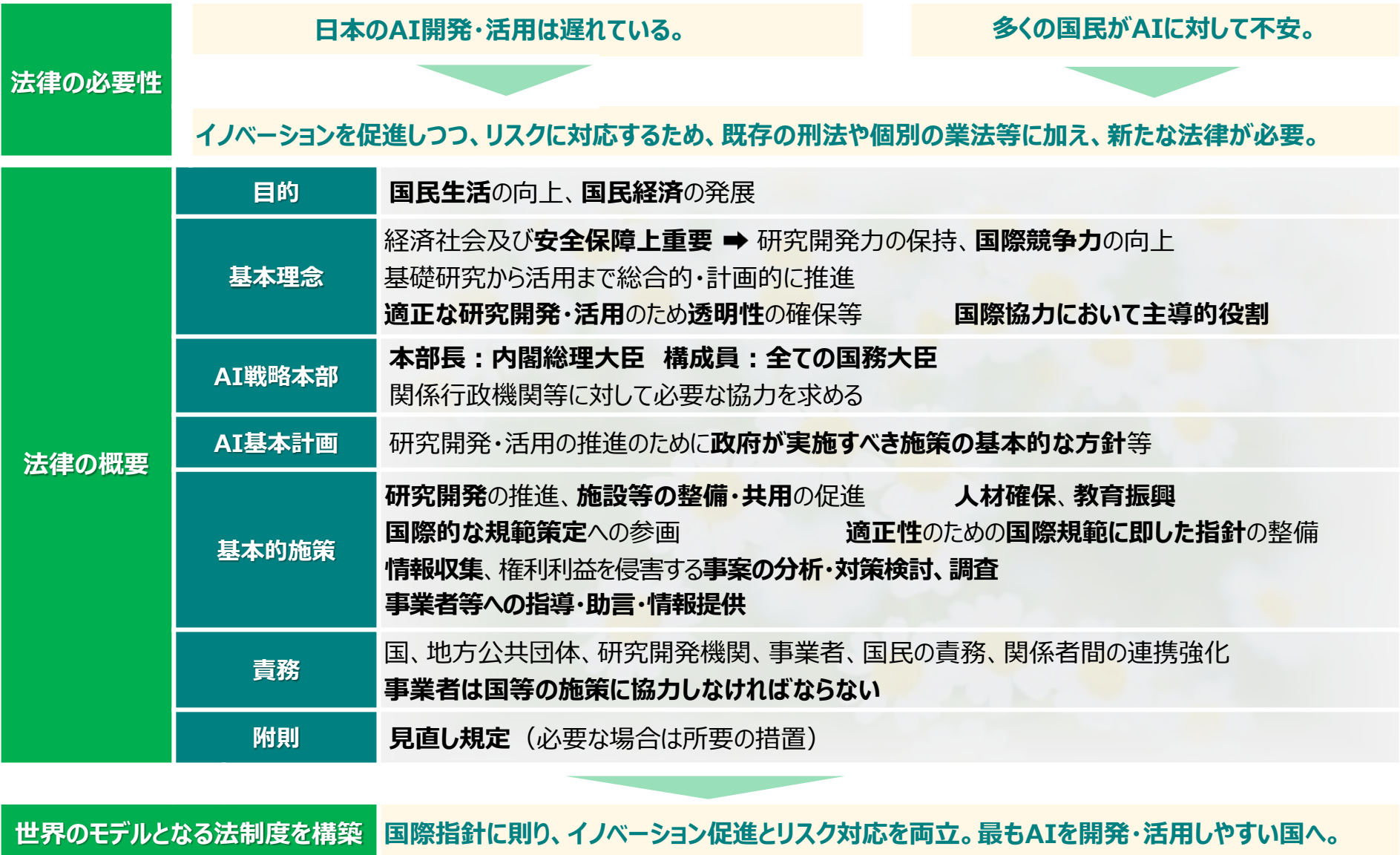
プライバシー・
財産権の侵害

偽・誤情報の拡散

雇用・経済不安

人工知能関連技術の研究開発及び活用の推進に関する法律（AI法）の概要

（2025年5月28日成立、6月4日公布・一部施行、9月1日全面施行）



第Ⅱ期人工知能基本計画（概要）

～日本A X、より強く、より豊かに～

令和8年7月14日 閣議決定

基本構想

- ◎急伸するエージェント型A Iに適応し、計算資源や電力から、制度・政策まで日本のA I実装能力を抜本強化
- ◎A I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「A Iトランスフォーメーション」（A X）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A Iが協働する社会」を率先して実現
- ◎「信頼できるA I」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

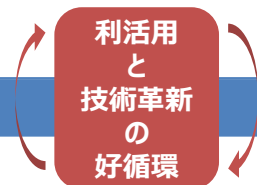
政府一丸でA X推進に必要な体制を構築、あらゆる制度や運用を先導見直し

1. A I利活用の加速的推進「A Iを使う」：エージェント型A I等の積極適用、政府における率先導入

- 政府・自治体でのA Iの徹底した利活用 ガバメントA I「源内」の進化と普及、地方の伸び代を成長に転換（地域A X）
- A I利活用促進による社会課題解決、新事業創出 バーティカルA I・フィジカルA I中心に社会実装加速
- データ戦略、更なるA I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. A I開発力の戦略的強化「A Iを創る」：A I実装能力を高め、「A I主権」を確保する

- A I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A I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A I基盤モデル等の開発 日本の文化等を踏まえたA I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 バーティカルA I（領域別戦略）、フィジカルA I（A Iロボティクス、自動運転）、AI for Science



3. A Iガバナンスの主導「A Iの信頼性を高める」：A I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
- A S E A N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A I共創に向けA Iサミットの早期日本開催

4. A I社会に向けた継続的変革「A Iと協働する」：人が意思決定への責任を持ちながら、A Xで日本を駆動

- A I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A X推進（取組可視化等）、雇用への影響を調査研究
- A I社会における制度・枠組みの検討・実証 規制・制度改革、人とA Iの責任関係、知財保護・利活用につながる透明性確保
- A I時代の人材の育成・確保 A I実装人材育成・確保、全ての国民のA Iに関するリテラシー向上、リ・スキリング支援
- A I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A Iの役割分担の探究

➡ 本基本計画は当面毎年変更、適切なベンチマーク（KPI等）を設定しモニタリング

■ A I 技術の進展

- A I は業務を支援するツールから、意思決定・実行を担う主体へと進化。
- エージェント型 A I が急速に伸長、世界各国で、産業・安全保障含めた広範な分野で業務全体を自律的に回す基盤に。
- エージェント型 A I をいかに社会に実装していくのか、その在り方は経済力、防衛力、技術力といった国力に直結。

■ 我が国の現状と課題

- 世界では、エージェント型 A I を始め先端技術に関する利活用と開発がますます加速。
- 日本でも、現場実務での A I 利活用は加速、開発投資も伸長しているが、組織全体や中小企業、地域での活用に課題。
- 計算資源、電力、人材、データ、制御・管理や制度・政策まで、総体としての A I 実装能力強化が必要。

■ 日本の勝ち筋

- 我が国が持つ現場の力を A I の実装能力としていくために、「バーティカル A I」、「フィジカル A I」に A I 戦略の重点を置き、我が国の課題を解決、独自の価値を生み、世界の課題解決にも寄与。
- A I エコシステム全体の中で、戦略的自律性及び戦略的不可欠性を確保し、開かれた「A I 主権」を確立。同志国等と連携しつつ、戦略的領域において自律性を強化し、耐遮断性・運用能力を確保。現場やハードウェアの力で優位性を創出。
- 産業や行政、国民生活の課題を解決、国力を強化するため、あらゆる組織で、A I を前提として意思決定や業務の進め方を見直す「A I トランスフォーメーション (AX)」を推進。

■ 複雑化・深刻化するリスクへの対応

- A I がもたらすリスクは、エージェント型 A I を始め技術革新が進む中で、複雑化かつ深刻化。
- A I 法等の「制度による対応」に加え、「技術による対応」、「組織管理による対応」に統合的に取り組みながら、「信頼できる A I 」体现のため、「責任あるアジャイル・ガバナンス」に能動的に取り組む。

■ 社会全体での A X に向けて

- 制度や政策について A I を前提として再構築することを含めて A I の実装能力を抜本的に強化し、エージェント型 A I についても日本の「信頼性」を再現する。
- 人と社会との関係を再定義し、判断責任を始め人が創造すべき価値は何かを探究、人が人として生きるための社会の枠組みの構築や「人間力」の増進に率先して取り組む。
- 「信頼できる A I 」で日本を駆動する A X に社会を挙げて取り組む。

(参考) A I 施策の方向性: A I 利活用の加速的推進 (A I を使う)

主な拡充内容

- エージェント型 A I の導入や国内で開発された A I モデルの積極活用も含めたガバメント A I を推進。「源内」のオープンソース化の取組を通じて自治体の A I 導入支援を進める。
- 優良なユースケースの横展開などにより、A I による自治体業務の構造変革を目指す。
- 地方の伸び代を成長に転換するため、中堅・中小企業における A I 導入、A I を活用した地域密着型の新規事業の立ち上げや先進的な地域課題解決プロジェクトの創出を支援。
- 医療、金融、教育、防災等の各分野において、エージェント型 A I も含めたバーティカル A I ・フィジカル A I を核とした A I 開発・実証・導入・社会実装を促進。バーティカル A I として領域別戦略策定。官民で投資集中。
- 防衛力の変革に向けて、データと A I を最大限に活用する。
- 地理空間情報のデータ連携・流通環境の構築を進め、「ジオ A I」を推進。
- 政府・準公共・産業・研究分野における各データについて、データ連携基盤を構築。
- データ精製技術に関するミドルウェアや、データスペース技術の研究開発・実証を支援。データ精製・データ連携を中核的に担う国内プラットフォームサービスを育成。
- 社会全体の A X を推進する観点から、府省庁横断で、法制度・ガイドライン・運用ルールについて抜本的な見直しを行う。

主な拡充内容

- 国内で開発・製造されるA I インフラ (データ、地域共生にも配慮したデータセンター、計算基盤、半導体) の生産能力拡大及び供給能力拡大を推進し、サプライチェーンを強化。
- 信頼性確保にも留意しつつ、トップ人材を含めたA I 研究者・開発者を確保、A I 駆動型の研究に積極的に取り組むことができるよう支援制度を充実。
- 「AI for Scienceの推進に向けた基本的な戦略方針」を踏まえ、AIの基礎研究を含め新たな挑戦や国際連携等を通じた先導的研究、研究インフラの構築、「オープン・アンド・クローズ戦略」を踏まえた研究データの戦略的管理・利活用等、AI for Scienceの取組を推進。
- 我が国独自の価値を持ち、国際競争力あるA I システムの創出も視野に入れ、暗黙知が豊富な現場へのバーティカルA I の開発・実装・輸出を促進。
- 産業用ロボットや自動車産業で培った我が国サプライチェーンの強みを活用し、多用途ロボットの国産ロボットメーカーやシステム・インテグレーターの機能の強化・育成。
- 日本が世界を主導できるよう、官民双方においてA I ロボティクスの実装を含めたフィジカルA I の研究開発及び実証を戦略的かつ統合的に促進。
- より高度な自動運転技術導入に向けた研究開発実証及び本格的普及に向けた取組を推進。

主な拡充内容

- A I 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、A I 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A I 利活用時の留意点等を整理するほか、広くA I 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A I 関連の知見を集約し、A I 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A I 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。
- A I S I を中心として、A I 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。
- A I に関する標準化等に関してA I S I 国際ネットワーク等を活用によるA I ガバナンスの構築を主導。各国と連携し、高性能A I 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A I イノベーションの結節点となるよう、A I サミットを早期に日本で開催し、A I での国際的な協調・協力モデルを主導。

主な拡充内容

- 中堅・中小企業や地方部を含め、各産業においてA Xを進め、A I を融合させた新たな産業構造を生み出す。
- エージェント型 A I も前提に、A I 利活用時における権利侵害や損害が発生した場合などの責任分界について、継続的に検討。
- 適切な知的財産の保護と利活用につながる透明性を確保。コンテンツホルダーへの対価還元等を推進。
- A I が働き方や雇用に与える影響を踏まえて、リ・スキリングの推進を含めた包括的な対策に能動的に取り組む。
- 初等中等教育段階の生成 A I の利活用に関するガイドライン改訂。教育分野に特化した A I の実証研究等を通じ、学校教育における安全・安心で主体的な A I の利活用を推進。
- A Xを進める上で中核となる A I 実装人材など、A I 人材を質・量の両面から育成・確保。
- 全ての国民が A I 適正活用人材となれるよう、教職員に向けた研修機会の提供の充実等を通じ、初等中等教育段階での情報活用能力を抜本的向上。
- 人と A I の協働による新たな社会を追求するため、A I の普及に伴う心理的影響、社会構造の変化も踏まえつつ、創造力や思考力を始めとした、判断責任を果たす人的主体性を中核に「人間力」向上を図る。

人工知能関連技術の研究開発及び活用の適正性確保に関する指針（概要）

令和7年12月19日 AI戦略本部決定

本指針の 位置付け

- AI法第13条に基づき、信頼できるAIの実現に向けて、国際的な規範の趣旨に即して策定
- 全ての主体におけるAIの研究開発及び活用の適正な実施に係る自主的かつ能動的な取組を促す

適正性確保に関する基本的な考え方

AIに関わる全ての主体

- ◎国 ◎地方公共団体
- ◎研究開発機関
- ◎活用事業者 ◎国民

適正性確保に必要な主要要素

- 人間中心 ●公平性 ●安全性 ●透明性
- アカウントビリティ ●セキュリティ
- プライバシー・個人情報 ●公正競争
- AIリテラシー ●イノベーション

適正性確保のための基本方針

- ① リスクベースでのアプローチ
- ② ステークホルダーの積極的な関与
- ③ 一貫通貫でのAIガバナンスの構築
- ④ アジャイル（柔軟かつ迅速）な対応

各主体が特に取り組むべき事項

活用事業者、研究開発機関

- AIガバナンスを構築・運用。
- ステークホルダーとの信頼関係の構築に向けて透明性を確保。
- 技術を用いて十分な安全性を確保。
- データの重要性を踏まえデータ保有者等のステークホルダーへ配慮。

国、地方公共団体

- AIを積極的かつ先導的に活用。
- 各主体がAIの責任ある利用ができるようリテラシーの向上を図る。
- AIガバナンスの在り方を検討。
- 行政の信頼性を確保するため、アカウントビリティを果たす。

国民

- 人間中心の原則に基づき、倫理、法令、人権等の様々な課題を理解しAIの責任ある利用者としての自覚を持って行動。
- AIリテラシーを能動的に身に付け、AIを適切に利用。

AIリスクへのアジャイルな対応（イメージ）

- 事業者・国民の自主的かつ能動的な取組を促す **AI法第13条に基づく指針（AI指針）を整備。**
- AI指針やAI事業者ガイドライン等を基にした、**AIセーフティ・インスティテュート（AISI）**による **AIモデルの技術的評価**の実施、当該評価も踏まえたAIがもたらすリスクに係る **AI法第16条に基づく調査研究**、各府省庁ガイドライン等を**有機的に組合せて、アジャイルに対応。**

AI法第13条 指針

評価・調査を踏まえた
アジャイルな見直し

AIの適正性確保のための自主的かつ
能動的な取組を促す基本的な考え方

各府省庁ガイドライン等

分野ごとの具体的な対策等

ガイドライン等を基とした
評価・調査

AISIによるAIモデル技術的評価

AIの適正性に係る評価

AI法第16条 調査研究等

国民の権利利益の侵害事案について実態把握

適正性の評価・調査

積極的協力

自主的取組の促進

実態を踏まえた
アジャイルな見直し

事業者・国民による自主的取組

政府の取組に積極的に協力し、
自分事として適正性確保に取り組む

政府による支援

リスク対策の技術開発等

AIセーフティ・インスティテュート (AISI) について

- AIが社会に与える影響が拡大するなかで、その**安全性と信頼性を専門的かつ中立的な立場で検証する「公的な第三者機関」**である**AIセーフティ・インスティテュート(AISI)**が必要不可欠に。
国際的にAIガバナンスの重要性が高まる中、**AI安全性サミット**（2023年11月、英国）を契機に「**AI安全性**」を具現化するための議論が進み、英・米はAIセーフティ・インスティテュートを設置。
- 我が国も2024年2月に日本の**AIセーフティ・インスティテュート**を設置。A I 安全性の知見のハブとして、**国内外の関係機関との協力しながら、安全性評価のための基準・ツール、ガイダンスを作成・提供**。
- 昨今のサイバー性能が高い高性能 A I の登場を踏まえ、A I 安全性のみならず、**A I 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、そのための体制を強化**※。

※Anthropic Mythos等の高性能AIのAIモデルの詳細な技術評価が可能である英国AISI（約200名人以上（うち専門家は90人））をベンチマークに、日本AISIでもA I セキュリティの体制強化（令和8年9月時点：61人（専門家40人））を図り、米中のサイバー性能が高い高性能 A I モデルの評価を試験的に実施してきている



エイシー 日本のAISIの概要

名 称

（日本語）AIセーフティ・インスティテュート
（英 語）Japan AI Safety Institute（略称 J-AISI）

主要な実績

- AIセーフティに関する評価観点ガイド、レッドチーミング手法ガイド等の公開。
- AISI国際ネットワーク(米・EU等の主要国AISI関連機関10カ国が参加※)に参加し、AIの共同テストに関するトラックをリード。

関係機関

（※）米、豪、加、EU、仏、日、ケニア、韓、シンガポール、英内閣府、AI戦略本部関係省庁（情報通信研究機構、理化学研究所、国立情報学研究所、産業技術総合研究所、情報処理推進機構含む）



村上 明子(AIセーフティ・インスティテュート所長)
(SOMPO ホールディングス株式会社
執行役員常務 グループChief Data Officer)
2024年の設置時より所長として活動

AISIの役割とスコープ

AIの**安全・安心な活用**が促進されるよう
官民の取組を支援することがAISIの役割

役割

- 主に3つの役割を担う。

政府への支援

- AIセーフティに関する調査、評価手法の検討や基準の作成等

日本におけるAIセーフティのハブ

- 産学における関連取組の最新情報の集約
- 関係企業・団体間の連携促進
- 国研等の関係研究機関との連携

国際協調の主導

- AISI国際ネットワーク間の連携
- その他国際機関との連携

AIの開発や利用をする者が
AIのリスクを正しく認識
できる仕組みの構築

+

ガバナンス確保などの
必要となる対策を
ライフサイクル全体で実行
できる仕組みの構築

↔

国内・国際的
な関係機関

イノベーションの促進と
ライフサイクルにわたるリスクの緩和を両立する枠組みを実現

スコープ

- AIによる以下の事象や検討事項の中で、諸外国や国内の動向も見ながら柔軟にスコープを設定し取組を進めていく。

社会への
影響

ガバナンス

AIシステム

コンテンツ

データ

AISIの現在の主な取組

実務に適用可能な評価手法やガイドを体系的に整備することを重視



1. Chief AI Officerガイドブック

- 主に民間事業者がCAIOを設置・運用する際の標準的な実務指針を提供

2. インシデントレスポンスアプローチブック

- AIシステム特有のリスクに起因するインシデントに対応するための新たな枠組みとしてアプローチブックを公開

3. データ品質マネジメントガイドブック

- データ品質を持続的に確保するため何をすべきか整理したガイド

4. レッドチーミング手法ガイド

- レッドチーミング手法について基本的な留意事項を示したガイド

5. 評価観点ガイド・評価環境

- AIシステムの安全性を評価する際の基本的な考え方を示したガイド

6. 事業者ガイドライン

- 総務省・経済産業省がビングドキュメントとして更新の検討を取りまとめ公表

●「AISIのパーパス：信頼できるAIの提供に必要な情報や技術を有し、日本を世界で最もAIを開発・活用できる国にする」

- パーパス実現のためAI基本計画等を踏まえて、**ミッション（AISIの役割・責任）を整理**
- ① AI安全性に関連する情報のハブとして、信頼できる優れたAI活用製品やサービスの普及を促進する（もって、国民生活の向上・社会課題の解決を促進）
- ② AI技術やAI安全性等に係る国際標準化等により、我が国のAI関連分野のイノベーションの加速・競争力強化を支える
- ③ 情報収集・分析・共有・対策の早期実施により、AIの開発・普及に伴う国民の生命・財産に危害が及ぶような事象など、国家の安全を脅かす事象の抑止に貢献する

- ミッション達成に必要なAISIが行うべき4つの活動
 - （1）AIの**安全性に関する物差し（ベンチマーク）**を作る
 - （2）**AIモデル・システム**を自ら評価する能力を持つ
 - （3）AI関連情報の収集・分析・提供を**率先して行う**
 - （4）**AI安全性に関して**国際協調を主導し、**国際競争力**を高める

AIセーフティ・インスティテュート（AISI）の活動状況



1. アンソロピック社との協力覚書（MoC）締結（2025年10月）

- 米国アンソロピック社のダリオ・アモデイCEOの初来日に伴い、高市総理への表敬はじめ、日本政府と米・アンソロピック社との間で、信頼できるAIエコシステムについて意見交換を実施。
- このMOCは、**AI評価の科学を進展させる**ことを目的とし、**AIモデルの能力、限界、潜在的风险を企業が評価するための手法**に関する情報交換とベストプラクティスの共有を行うことへの**合意**である。
- AI技術の進展や将来の動向に関しても**継続的に意見交換**を進め、国際的に適用可能な共通基準の検討を支援し、**責任あるイノベーションを加速**させる。



出典：[令和7年10月29日 アモデイ・アンソロピック社CEOによる表敬 | 総理の一日 | 首相官邸ホームページ](#)

2. OpenAI社との協力覚書（MoC）締結(2026年5月)

- 米国OpenAI社のジェイソン・クウォンCSOが来日し、**AISIとの間で**、AIの安全性評価等の協力に係る覚書（MOC）の署名を実施。
- このMOCは、AIの**安全性評価**および**ベンチマーク、サイバーセキュリティ**に関する協力の可能性を探ることを目的とし、業界別の安全性評価や評価手法に関する知見、ベストプラクティスの共有について意見交換を行うことへ**合意**である。
- AIの安全性評価やサイバーセキュリティに関する継続的な対話**を通じて、国際的に適用可能な**安全性評価の在り方やベンチマークの課題**について検討を進める。



出典：[ジェイソン・クウォンCSOの大臣表敬訪問の様子 - 内閣府](#)