AISI
Japan AI Safety Institute

AI セーフティ年次レポート 2024

National status report of AI Safety in Japan 2024

2025 年 2 月 5 日

AIセーフティ・インスティテュート（AISI）事務局

目次

1　AIセーフティ・インスティテュート（AISI）の設置およびこれまでの取り組み

1．1　基本情報

AISI は、安全、安心で信頼できる AI の実現に向けて、AI セーフティに関する評価手法や基準の検討・推進を行うための機関である。AISI は、「AI セーフティ」を通じて、AI の利活用におけるリスクへの対応とイノベーション促進を両立させていく。

AI セーフティ

人間中心の考え方をもとに、AI 活用に伴う社会的リスク※を低減させるための安全性・公平性、個人情報の不適正な利用等を防止するためのプライバシー保護、AI システムの脆弱性等や外部からの攻撃等のリスクに対応するためのセキュリティ確保、システムの検証可能性を確保し適切な情報提供を行うための透明性が保たれた状態。

※社会的リスクには、物理的、心理的、経済的リスクも含む

1)　AISI の目的（役割）

- 政府への支援として、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の支援を行う。
- 日本における AI セーフティのハブとして、産学における関連取組の最新情報を集約し、関係企業・団体間の連携を促進する。
- さらに、 他国の AI セーフティ関係機関と連携することで、国際的なコンセンサスを確立する。

2)　AISI の特殊な状況

AI は、これまでの技術分野では考えられないほど技術変化が速い最先端の分野であり、その影響範囲は技術だけにとどまらず、社会や雇用、文化など広範にわたる。

各国 AISI の置かれている状況はそれぞれ異なり、AI セーフティの対象範囲も、AI システム自体のセキュリティから、エコシステム全体のガバナンス、法律などの整備から、知的財産や個人情報への配慮など多岐にわたっている。米国では、さらに安全保障などへも対象が広がっている。

月や四半期で製品が置き換わるほど技術開発スピードが速く、分野横断かつ、グローバルな取り組みであり、各国や国際機関のレポートも毎月のよう

に公表されている。こうした中で、日本を含め世界中で一つ一つの取り組みに、これまで以上に時間がかかり、まさに、アジャイル思考で走りながら考えている状況である。従来の年単位で変化する安定した技術に対する対応ではない、全く新しい技術政策の推進の在り方が求められている。数が多いだけではなく、中の質となるような技術の繋がりのチェックも含めて、対応が求められる。

AI セーフティに関する国内横断的やグローバル連携の取り組みは、これまで前例となるものがなく、各国でも政策と技術を融合した体制の作り方を模索している。また、各国が相応の資金を投じる中で、各国の体制の作り方は異なり、その体制の違いが調整をより一層難しくしている反面、AISI は政府機関の一部として活動を行うことで、国として AI セーフティに関する政策を直接推進できる。

各国の政策と技術の推進体制

米国：国務省と商務省と米国国立標準技術研究所（NIST）がチームで取り組みを推進

英国：府省内に AISI を作り、政策と技術の意思決定を一本化

欧州委員会：法律を所管するとともに、利活用も一体で機構内のチームで推進

日本：内閣府と関係省庁が連携して AI の安全性確保に向けた政府方針等を AISI に指示し、AISI はそれに基づいて安全性評価に係る調査、基準等の検討や、実施手法の検討等を行う。

1.2　主要なイベントとアクション

1)　AISI

2024 年 2 月に 10 関係府省庁、5 政府系関係機関が共同して AI セーフティ・インスティテュートを設立した（体制は後述）。AISI は政府の AI セーフティに関する取り組みを支援する機関である。ガイドラインなどの基準を整備するとともに、関係各所と情報交換し、幅広く情報収集していることから、日本国内における AI セーフティに関する情報のハブとなる役割が期待されている。さらに、AI セーフティに関しては、国際的な取り組みが非常に速いスピードで進んでいるが、国際的な場への参加と貢献も期待されている。

情報提供と透明性確保の一環として、設立と同時に Web サイトを設置し、2024 年 9 月に全面リニューアルした。AISI の活動状況を報告するとともに、

成果物の公開を行っている。政策面での内閣府へのリンク、研究開発面での人工知能研究開発ネットワーク（AI Japan R&D Network）へのリンク、AI 利活用に関する情報処理推進機構（IPA）へのリンクなども提供し、AI セーフティに関する情報のハブ機能を実現している。

AISI web サイト

https://aisi.go.jp/

2) AI 事業者ガイドライン

2024 年 4 月に、総務省、経済産業省が、従来の AI ガイドラインを統合するとともに最新化を図り、AI 事業者ガイドラインとして公表した。このガイドラインを参照し、金融業界が生成 AI ガイドラインを策定したり、後述する AI セーフティに関する評価観点ガイドが作られる等、AI セーフティの展開の基盤となっている。日本語版に加えて国際連携の観点から英語版も提供している。

現在は AISI も共同事務局として参加し、2024 年 11 月の 1.01 版に続き継続的な更新に取り組んでいる。

https://www.ipa.go.jp/disc/committee/expert-group-on-aigfb.html

3) AI セーフティ・サミット

2024 年 5 月にソウルで AI セーフティに関するサミットが開催された。AI セーフティの議論を深めるとともに、AI 開発におけるイノベーション促進及び AI の恩恵の公平な享受について議論が行われ、首脳級の成果文書として「安全、革新的で包摂的な AI のためのソウル宣言」が採択された。

https://www.mofa.go.jp/mofaj/files/100672518.pdf

日本をはじめ各国 AISI は、併設された AI Global Forum に参加し、意見交換を行うとともに、2 国間会議などを実施し、今後の国際協調の在り方について意見交換を行った。

4) 日米クロスウォーク

2024 年 4 月と 9 月に、日本の AI 事業者ガイドラインと NIST の AI リスク・マネジメント・フレームワークとのクロスウォーク（用語の相互参照、コンセプトの相互参照）の結果を公表した。これは AISI 設立前の 2024 年 1 月から実施してきたものである。

NIST の AI リスク・マネジメント・フレームワークは、各国のガイドラインや国際標準などとのクロスウォークを実施済であり、NIST の AI リスク・マネジメント・フレームワークを介して各国のガイドラインとの相互参照が可能

となっている。日米のクロスウォークについては、他国に先駆けコンセプトの相互参照まで行うことで、AI リスクマネジメントに関する日米の相互運用性を確認できている。これを礎に、AI セーフティに関する評価観点ガイドや AI セーフティに関するレッドチーミング手法ガイドを策定するアプローチをとった。

クロスウォーク１
https://aisi.go.jp/effort/effort_information/240430/
クロスウォーク２
https://aisi.go.jp/effort/effort_information/240918_1/

5) AI 制度研究会

2024 年 8 月に、政府が AI の法規制や制度整備を検討するために設置した有識者会議である。主な議題は、生成 AI を含む技術進展への対応、国際指針との調和、柔軟な制度設計などで、2024 年 12 月に「中間とりまとめ（案）」が提出され、意見募集が実施された。AISI 所長が、AI 制度研究会の座長代理を務め、検討を進めている。

https://www8.cao.go.jp/cstp/ai/ai_kenkyu/ai_kenkyu.html

6) AI セーフティに関する評価観点ガイド

2024 年 9 月に、AI 事業者ガイドラインをベースに、AI セーフティに関する評価観点のガイドを作成した。AI システムの開発や提供に携わる者が AI セーフティ評価を実施する際に参照できる基本的な考え方を提示している。日本語版に加えて国際連携の観点から英語版も提供している。

AI セーフティに関する評価観点ガイド
https://aisi.go.jp/effort/effort_framework/guide_to_evaluation_perspective_on_ai_safety/

7) AI セーフティに関するレッドチーミング手法ガイド

2024 年 9 月に、AI セーフティの評価手法の一つであるレッドチーミング手法についてのガイドを作成した。AI システムの開発や提供に携わる者が、対象の AI システムに施したリスクへの対策を、攻撃者の視点から評価するためのレッドチーミング手法に関する基本的な考慮事項を示している。日本語版に加えて国際連携の観点から英語版も提供している。

AI セーフティに関するレッドチーミング手法ガイド
https://aisi.go.jp/effort/effort_framework/guide_to_red_teaming_methodology_on_ai_safety/

8) 国連 Global Compact Leaders Summit

2024 年 9 月に、国連未来サミットの機会に開催された国際会議であるリーダーズ・サミットは、持続可能な開発と繁栄を促進するため、リスクに対応しつつ、原則に基づくアプローチで企業とつながる重要なフォーラムである。AI の責任ある開発と利用を推進する AISI への期待が高まりを受け、「Rights & Bytes」セッションで、人権と AI における実行可能な一歩を示し、参加者に行動を促すべく、AISI 所長が各国政府 AI セーフティ関係者や事業者を含め多様な関係者と議論を実施した。

https://events.unglobalcompact.org/leaderssummit24

9) CEATEC

2024 年 10 月に、幕張メッセで行われた CEATEC（Combined Exhibition of Advanced TEChnologies）で AISI 所長が「AI 社会の今後について」と題して講演するとともに CEATEC 25 周年の特別企画「AI for ALL」特設ブースで展示と講演を実施した。また、国内の AI 推進の各団体との AI をテーマとしたネットワーキング・イベントにて AISI 所長が基調講演を行うとともに関係団体と意見交換を実施した。

https://www.ceatec.com/ja/conference/detail.html?id=2570

10) シンガポール AI Safety Red Teaming Challenge

2024 年 11 月に、シンガポールで実施された国際レッドチーミング・チャレンジに主要参加国として参加した。本イベントは多文化・多言語に焦点を当てた AI セーフティに関するもので、日本からは、国立情報学研究所（NII）、東京大学、京都大学、日本ディープラーニング協会（JDLA）も参加し、大規模言語モデルに対する評価を実施した。

https://www.imda.gov.sg/activities/activities-catalogue/singapore-ai-safety-red-teaming-challenge
https://www.tc.u-tokyo.ac.jp/blog/wp-content/uploads/2024/10/Singapore-Report-English.pdf

11) AISI 国際ネットワーク（International Network of AI Safety Institute）

英国、欧州委員会、オーストラリア、カナダ、ケニア、シンガポール、大韓民国、日本、フランス、米国の 10 カ国/地域による、AI セーフティの技術的な内容を検討するための国際協力枠組みである。

2024 年 11 月にサンフランシスコで第一回会議を開催し、ミッションステートメントを採択するとともに、「Track1：合成コンテンツのリスク低減」

「Track2:基盤モデルの共同テスト」「Track3;先進的AIシステムのリスクアセスメント」に関する議論を実施した。多言語の評価手法を検討するセッションでは、シンガポールとともに日本が共同議長を務めた。
ミッションステートメント
https://aisi.go.jp/effort/effort_international/241120/

1.3 AISI関連のその他の活動と成果

1) NIST AI リスク・マネジメント・フレームワーク及びプレイブックの日本語訳

2024年7月に, AIセーフティに関する国内での理解を深めるため、米国に限定することなく国際的な場面においてリスクマネジメントの重要文書として用いられる、NISTが策定したAIリスク・マネジメント・フレームワークについて、商務省や国務省など米国政府にも確認をとるプロセスを踏み、日米の定まった解釈として日本語訳の作成を行った。

https://aisi.go.jp/effort/effort_information/240704/

2) セキュリティ調査レポート

新しい技術としてAIが業務利用されつつある状況の中、国内外のセキュリティ脅威やリスクの実態を把握すべく、AISI及びIPAは調査を行い、2024年5月に「米国におけるAIのセキュリティ脅威・リスクの認知調査」、7月に「AI利用時のセキュリティ脅威・リスク調査報告書」、12月に「AIに起因する選挙リスクとAIガバナンス米国調査」の結果を公表した。

https://aisi.go.jp/effort/effort_technology/technology/

3) AI政策動向調査

2024年5月に米国におけるAI政策の最新動向の把握を目的として、「米国におけるAI政策最新動向調査」を公表した。特に米国国立標準技術研究所（NIST）の取り組みを中心に、NISTのAIに係るこれまでの主な取組、NISTにおけるAI関連の研究概要、大統領令（第14110号）によるNISTへの要求事項、NIST傘下の米国AIセーフティ・インスティテュート（USAISI）と米国AIセーフティ・インスティテュート・コンソーシアム（AISIC）の概要などを示した。

https://aisi.go.jp/effort/effort_technology/technology/

4) DX を推進する人材向けの「DX 推進スキル標準」に生成 AI に関する補記などを追加

2024 年 7 月に、IPA は AI のリスク対応も含む、デジタルスキル標準の改訂を実施した。生成 AI の特性や活用例、新技術への対応方法などを補記し、共通スキルリストの学習項目を追加・変更している。デジタルスキル標準は、ビジネスパーソン全員が必要とする「DX リテラシー標準」と、DX 推進に求められるスキルや役割を定義した「DX 推進スキル標準」の 2 つから成りたっている。

https://www.ipa.go.jp/pressrelease/2024/press20240708.html

5) アクティビティ・マップとターミノロジーのチャレンジ

アクティビティ・マップは、AI セーフティに関する活動が急速に変化・進化を続ける中で、見落とされている領域や活動間の相関関係の全体像を可視化する。日本は、主要文献のベンチマーキングを踏まえ、包括的なアクティビティ・マップと関連ターミノロジーを整備している。日本が主導するこの取り組みは、AI セーフティにおける国際的な連携の基盤をさらに強化し、持続可能で信頼される AI 社会の実現に貢献することが期待される。(2025 年 2 月公表予定)

6) AI システムのためのデータ品質ガイドの検討

AI を適切かつ安全に活用するには、データの品質マネジメントは欠かせない。そのため、既存のデータ品質に関する国際標準やガイドラインを活用しながら、企業にとって有用で実践的かつ、実務的なガイドの検討を行った。データ品質は、前述の AI セーフティ評価観点ガイドにおける評価観点の一つになっている。評価観点の一つに加えた。この中で合成コンテンツなど AI に特化したデータの取り扱いについても検討を行っている。(2025 年 2 月公表予定)

7) AI 認証のための Joint Certification の検討

ISO/IEC JTC 1/SC 42 - Artificial intelligence において、製品に対する適合性評価に関するハイレベル・フレームワーク規格を ISO における適合性評価規格を開発する上層委員会（CASCO）と共同開発することが決定された。この開発のなかで、新しい認証の枠組みである Joint Certification（組織認証および AI 製品・サービス認証を組み合わせて行う新たな方法）の検討も並行して開始されると考えられる。そのため、その対応方針を検討した。

8) 政府による適切な AI の調達・利用の在り方等に関する検討

デジタル庁等を中心に進められている政府による適切な AI の調達・利活用の在り方等に係る検討に参画し、生成 AI のリスクへの対応に関して提案等を行うなど議論に貢献した。

政府は、AI の調達・利活用ルールについて 2025 年春を目途に、ガイドラインを策定予定である。

9) 国際活動

各国 AI 関係者と対面やオンラインで意見交換をするとともに、講演や研修講師などを実施している（計 14 回）。

主な会議相手

- 英国、欧州委員会、オーストラリア、カナダ、ケニア、シンガポール、大韓民国、日本、フランス、米国、マレーシア、マルタとの意見交換
- 海外 AI 関連企業との意見交換
- JICA 研修（アルメニア、イエメン、ウガンダ、グレナダ、ザンビア、トンガ、バヌアツ、ヨルダン）、ASEAN-Japan Cybersecurity Capacity Building Centre（AJCCBC）と JICA の共同研修（AJCCBC セミナーは、ASEAN 加盟国の政府関係者と重要インフラ（Critical Infrastructure）の関係者を対象とし、100 名以上が参加）

10) 国内普及活動

国内での AI セーフティに関する啓発活動として、日本経済団体連合会（経団連）、AI ガバナンス協会、東京都中小企業知的財産シンポジウム、LLM シンポジウム、AI 品質マネジメントシンポジウムなどで講演、パネルディスカッションへ参加した（計 60 回）。

1．4 AISI の体制

1) 全体

内閣府と協力し、10 関係府省庁、5 関係機関が連携する枠組みを構築した。AISI 関係府省庁等連絡会議及び運営委員会を設置し、方針の検討を行っている。

【関係府省庁】

内閣府（科学技術・イノベーション推進事務局）、国家安全保障局、
内閣サイバーセキュリティセンター、警察庁、デジタル庁、総務省、
外務省、文部科学省、経済産業省、防衛省

【関係機関】
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

2) 日本 AI セーフティ・インスティテュートパートナーシップ

2024 年 8 月に、AI セーフティに関して国内の協力体制を構築するため、5 関係機関（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）とパートナーシップ協定を締結した。AISI の活動を効果的に推進するため、AI セーフティに関して知見を有する関係機関と AISI が共同で研究及び調査などを実施し、国内外への情報発信等を行っている。

https://aisi.go.jp/effort/effort_information/240806/
https://aisi.go.jp/assets/pdf/20241021_shiryo1.pdf
（第 3 回 AISI 運営委員会 資料。10p-36p が AISI パートナーシップ協定の内容）

3) 事務局

事務局は IPA 内に置き、5 つのチームで構成される。

■ 戦略・企画チーム
- 政府機関との調整
- 民間活動との調整
- 国際調整
- プロジェクト管理
- 総合企画

■ フレームワークチーム
- 調査
- ガイドライン整備
- フレームワーク的国際調整

■ 技術チーム
- 調査
- アクティビティ・マップ整備
- 技術的検証
- 技術的国際調整

■ セキュリティチーム
- 調査
- セキュリティ技術的国際調整

■ 標準チーム

- 国際標準の動向調査（ISO/IEC JTC1/SC42（Artificial intelligence）での活動含む）
- 標準的国際調整

2025年1月時点で常勤、非常勤含め30人程度である。AIセーフティに求められるスコープが広がる中、米国AISI約60人、英国AISI約100人、EU AI Office約60人などと比較して、少数の専門家集団でAIセーフティの最先端の活動に取り組んでいる。

１．５　AISIの特殊な状況による課題

AISIはAIセーフティに関する技術的な専門チームとして組織しているが、利用者視点で考えれば、AIセーフティだけを深堀するのではなく、利活用も含めた方が理解しやすく、一体化した取り組みへの要望がある。また、AI事業者ガイドラインの別添 3.AI 開発者向け「適切なデータの学習」において、「データに個人情報、機密情報、著作権等の権利又は法律上保護される利益に関係した問題が生じ得る情報が含まれていないか、確認を実施」と記載されるなどデータ品質への対応も求められている。2025年2月にはAIアクション・サミットが開催されるなど、国際的な議論においても一体で扱う方向に変わってきている。AIセーフティとして技術だけでは解決しない分野への対応を関係者と検討する必要がある。

AISIは、対府省庁、技術者、海外と幅広いステークホルダーへの対応が求められる中、国内のパートナーと連携し、AI分野の人材確保に向けた積極的な取り組みを進めている。しかし、国内外でAI人材が不足する現状において、公務員の待遇制約も影響し、厳しい状況である。

また、関係機関が多いうえ、バーチャル組織のため異なる情報管理ポリシーや環境下での運用も求められ、ガバナンス体制が重層的で、臨機応変な機動的な対応が難しい。

さらに、OECDは2024年11月に公表した「Assessing potential future Artificial Intelligence risks, benefits and policy imperatives」で将来のAIリスクの可能性として10のリスクを挙げている。そこでは技術的リスクのほかに「ガバナンスメカニズムや組織が、速く変化するAIの改革についていけなくなる」というリスクを指摘している。

2　今後に向けて

2.1　考慮すべき新たな要素

1.5 の課題に加え、今後に向けて考慮すべき要素には以下のようなものがある。

1) AI セーフティに関する評価手法や基準

1-1) AI 制度研究会の検討結果

研究会の検討結果により、多様な組織の今後の AI セーフティの体制や推進方法に影響を与えると考えられる。例えば、AI 利用のルールや基準を明確にすることで、AI セーフティや倫理性が向上し、透明性が高まることが期待される。規制の適用は負担を伴う可能性があるが、信頼性の向上により技術発展や投資促進が進み、特定分野での競争力の強化が期待される。また、個人情報保護やバイアスの制御といったリスク対策が強化されることで、市民が安心して AI を利用できる環境が整い、AI リテラシーの向上も期待される。さらに、国際的な基準との整合性が図られることで、国際協力や市場参入が容易になり、社会全体の AI 時代に適応する取り組みが進むと考えられる。

1-2) 認証制度の動向

ISO/IEC 42001 は、組織内における人工知能マネジメントシステム（AIMS）の確立、実施、維持及び継続的改善に関する要求事項を規定した国際規格である。AI ベースの製品やサービスを提供または利用する事業者向けに設計されており、AI システムの責任ある開発と利用を保証する。

現在、ISO/IEC 42001 に基づく認証制度の開始に向けて、認証審査機関向けの要求事項を定める ISO/IEC 42006 の開発が進められている。こうした国際的な動きは、企業や組織がグローバル市場で競争力を維持する上で重要であり、日本国内でも対応を進める必要がある。認証の動向を注視しつつ、関連すると考えられる Joint Certification の検討においても適切な対応を行うことが求められている。

2) 技術的な進展

AI エージェント、ロボットへの組み込みや対話型生成 AI の登場でパーソナルユースも含め利活用のすそ野が広がっている。AI に関する新技術や用途の拡大により、新たな AI セーフティの課題が顕在化する可能性がある。また、AI に関する新技術により、AI セーフティのスコープ、観点、評価手法が変化

する可能性もある。これらの技術的な進展を踏まえた、AISI 発行ガイド類の更新が必要になるなど、様々な領域で AISI の活動の重要性が高まる。

3) 他国の AI セーフティ関係機関連携

3-1) AISI 国際ネットワークの取り組み

AISI 国際ネットワークの今後の方向性は、2025 年 2 月にパリで開催される AI アクション・サミットでの議論が重要な役割を果たす。現在のテーマを引き続き発展させるのか、それとも新しいテーマを設定するのかで進むべき道が変わる可能性がある。

3-2) 国際機関などによる取り組み

G7、および OECD や国連などの国際機関でも AI に関する議論が進められており、その検討内容や推進体制、具体的な方法が国内外に影響を与えると考えられる。これらの議論は、AI の倫理的利用や規制の在り方、国際協力の強化といったテーマを中心に展開されており、日本を含む各国の政策や体制に大きな影響を及ぼす可能性がある。特に、国際的な合意や基準が形成されることで、国内の AI 戦略や規制の方向性が決まる重要な要素となる。

4) 民間企業との連携

英国や米国の AISI は、民間企業と連携しながら、AI モデルの評価や評価ツールの開発に取り組んでいる。この取り組みは、単に技術の精度を追求するだけでなく、AI の信頼性や安全性を確保するためのものである。また、これらの評価ツールは産業界や社会全体での実用性を考慮し、長期的な影響やリスクにも対応できるよう、戦略的な視点で設計されている。

2.2 今後の取り組み方向

現状見えている AISI の主たる課題と考慮すべき新たな要素を踏まえて、以下の取り組みによる AI セーフティの実現を通じて、AISI の設立目的に資する活動の充実と高度化を図っていく。

1) AI セーフティに関する調査、評価手法の検討や基準の充実・高度化

- AI に関する新たな技術と脅威が次々に出てくる国際的な状況の中で、AISI から公表した評価観点ガイドやレッドチーミング手法ガイドについてより実態に合ったものとすべく、調査及び改訂していく。
- AI セーフティ評価の活動を広く一般に普及させることを目指して、自動

化／省力化するための技術に関する実現可能性の検証を実施していく。
- 共通技術分野であるデータ品質や Joint Certification に重点的に取り組む。

2) 技術的な進展への対策

- 汎用性の高いマルチモーダル基盤モデルに対象を拡大し実用化の近い最先端のユースケース/技術動向の調査を実施する。
- AISI が発行した「評価観点ガイド」および「レッドチーミング手法ガイド」の実践による評価手法／項目の詳細化により、ガイドラインを改訂する。
- 公助として教材の提供などを通じて社会全体の AI セーフティ向上に取り組んでいく。

3) 他国の AI セーフティ関係機関連携

AISI 関連のトップレベルの連携や各国との意見交換を通じて、国際的な連携・協調を推進していく。例えば、先進的 AI システムのリスクアセスメントにおける重要なポイントに関する共同声明を基に、具体的なアクションへのブレイクダウンや評価手法の開発を、各国の AISI と引き続き協力して進めていく。

4) AISI の特殊な状況への対策と、民間企業との連携

- AISI は、政策、技術、国際調整のキープレイヤーを確保することで、先進的な取り組みを推進し、持続可能な体制を構築する。その一つが官民人材交流で、民間からの出向者が AISI 内で活動しているが、これを強化するとともに、体制強化を図るべく人材募集を引き続き行っていく。
- AISI は、迅速かつ的確なアクションを行うため、AISI 関係機関の多様性を活かし、政策と技術を横断的に扱う新たな情報共有基盤を構築する。共有された情報を各関係者が主体的に利用するプル形にしていく等の対応を行う。
- 2025 年の早期段階で、AI セーフティに関する技術の開発・実証の推進に向けて、民間事業者中心で各業種における具体的な課題に対する検討を進めるためのワーキンググループを設置していく。産業、ヘルスケア、防災など重点的な分野での応用を推進していく。AI セーフティの確保に向けたガイドライン、評価ツール、評価データセットなどを策定・開発し、日本における AI セーフティの情報ハブ機能の拡充に貢献する。