

AISIの活動について

2026-10-07
AISIS事務局

AISIとは ～AISI設立の背景・概要～

日本におけるAISIの設立

広島AIプロセスでの議論を経て 日本におけるAIセーフティ・インスティテュート（AISI）を設立（2024年2月）

2023年5月

日本から
「広島AIプロセス」
を提唱

G7広島サミットの結果を受け、
生成AIに関する国際的ルール
の検討のため立ち上げられた

2023年12月

「広島AIプロセス
包括的政策枠組み」
等に各国合意

2024年2月

AIセーフティ・
インスティテュート
(AISI)
設立
(事務局はIPAに設置)

2024年5月

広島AIプロセス
フレンズグループ
を設立

安全・安心で信頼できるAI
の実現を目指す、賛同国に
よる自発的な国際枠組み

2024年7月

GPAI
東京専門家
支援センター
設立

GPAIに所属するAIの専門
家を支援する組織

2025年3月

広島AIプロセス
レポーティング
フレームワーク
運用開始

2025年5月26日時点
19社が参加

「統合イノベーション戦略2024」において、 AISIIは日本におけるAIの安全性の中心機関と定義

- ◆ 「統合イノベーション戦略」は、「科学技術・イノベーション基本計画」に基づく年次戦略として、毎年策定している。

統合イノベーション戦略2024における強化方策

◆ AI分野の競争力強化と安全・安心の確保

1. AIのイノベーションとAIによるイノベーションの加速
研究開発力の強化、AI利活用の推進、インフラの高度化等
2. AIの安全・安心の確保
ガバナンス、AIの安全性の検討、偽・誤情報への対策、知財等
3. 国際的な連携・協調の推進
広島AIプロセスの成果を踏まえた国際連携等



統合イノベーション戦略2025における強化方策

◆ 先端科学技術の戦略的な推進

1. 重要分野の戦略的な推進

- AIイノベーション促進とリスク対応の両立
- AIの研究開発の推進等
- AI関連施設等の整備及び共用の促進
- AI活用の推進
- AIの適正性の確保
- AI関連人材の確保と教育振興等
- AIに関する調査研究等
- AI分野の国際的協調の推進

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AISII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

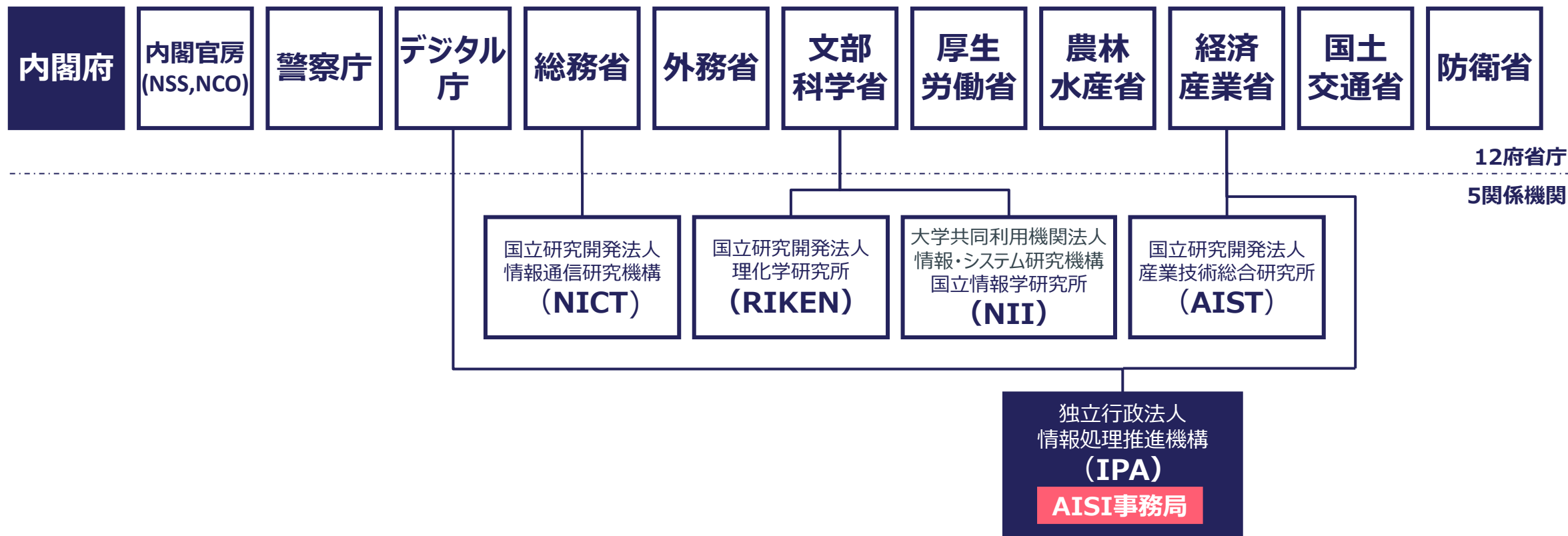
AIシステム

コンテンツ

データ

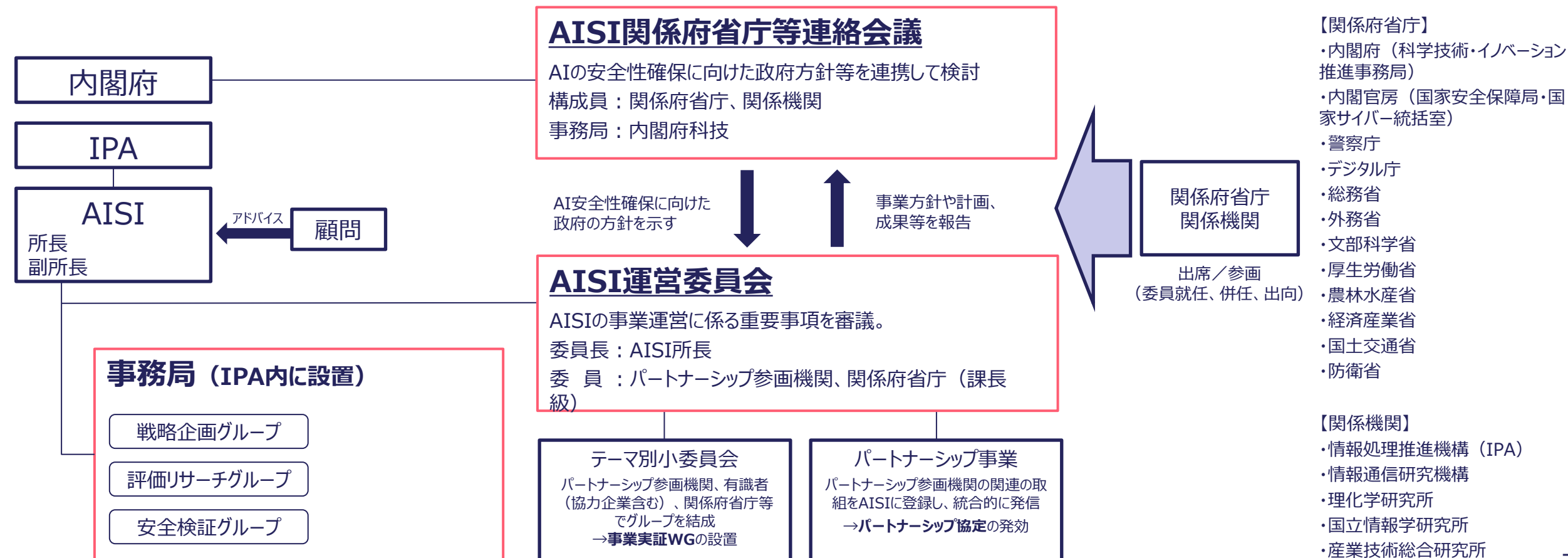
AISIは、12府省庁・5関係機関が横断的に参画する**政府関係機関**
事務局は経済産業省とデジタル庁を所管官庁としている**IPA**内に設置

* 2025年4月時点



AISIの推進体制

内閣府を事務局とする「**AISI関係府省庁等連絡会議**」で政府方針等を検討
AISI所長を委員長とする「**AISI運営委員会**」で事業方針を検討





所長 村上 明子

1999年 日本アイ・ビー・エム株式会社 東京基礎研究所 入社
2016年 同社 東京ソフトウェア開発研究所
2021年 損害保険ジャパン株式会社 入社 執行役員待遇 DX推進部 特命部長
2021年 同社 執行役員待遇 DX推進部長
2022年 同社 執行役員 CDO(Chief Digital Officer) DX推進部長
2024年 同社 執行役員 CDaO(Chief Data Officer) データドリブン経営推進部長 [現職兼務]
2025年 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 [現職兼務]



副所長
平本 健二



副所長・事務局長
西村 卓



副所長
関根 聡

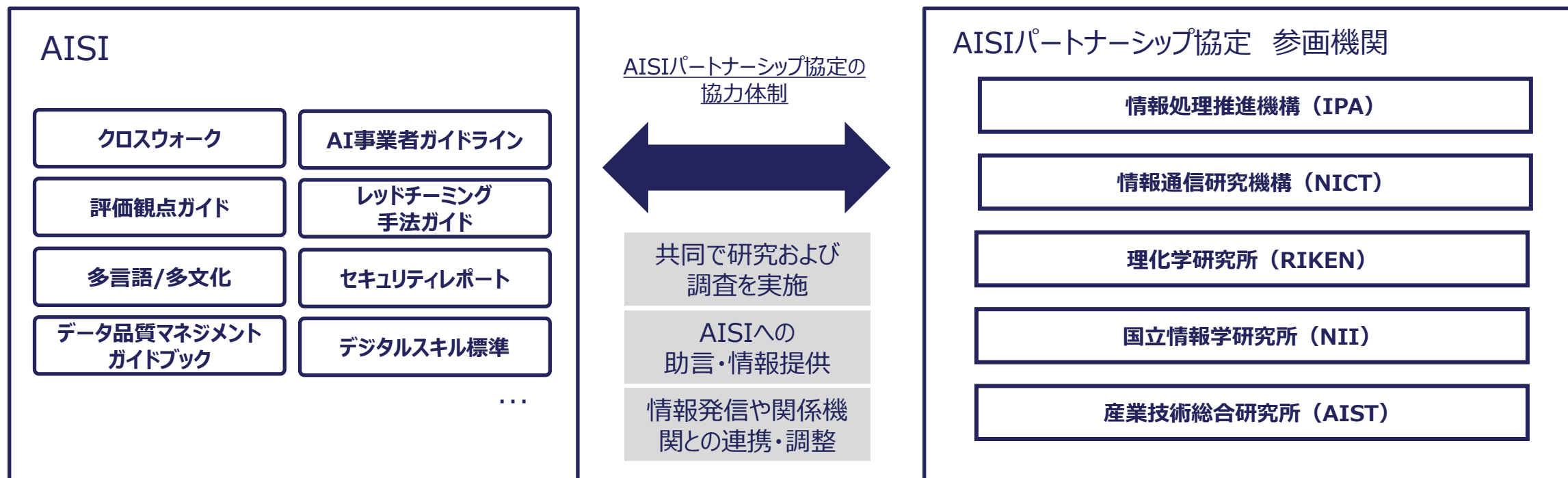


戦略ガバナンス調査官
浅岡 孝充

AISIパートナーシップ協定の発効

関係府省庁の協力の下、**関係機関が連携**してAIの安全性に係る取組を推進していく協力体制（AISIパートナーシップ協定）を発行（2024年8月）

- ◆ AIの安全性に関する取組を進めるためには、**AISIのみならず、国内の関係機関と連携し、共同で対応していくことが不可欠。**



第Ⅱ期人工知能基本計画

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索し、「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索、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 ● アジャイル（柔軟かつ迅速）な対応	● 挑戦と学習 ● 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		

利活用と技術革新の好循環



3タイプある世界のAI安全保障戦略のいいところを

「官民共創」の安全ベンチマークで「自己評価をできる能力」を持ち、
必要に応じて規制するためのソフト・ハードロー作成の技術支援を行う

日本モデル

官民エコシステムで作成したベンチマークを用いて、政府が評価能力を持って政策を実施



① 科学的評価モデル

英国・米国モデル

政府が直接評価能力を内製化し
技術的優位を確保

Measurement Science and
avoiding surprise



② 規制・インフラモデル

EU（フランス）モデル

AI法と市場のルール構成

Regulatory Enforcement and
Update based on Trust



③ 実装・エコシステムモデル

シンガポール・カナダ・韓国モデル

オープンソースと研究コミュニティの力で
社会実装を主導

Implementable Guidance and
Leveraging the Research
Ecosystem

「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たミッション（AISIの役割・責任）

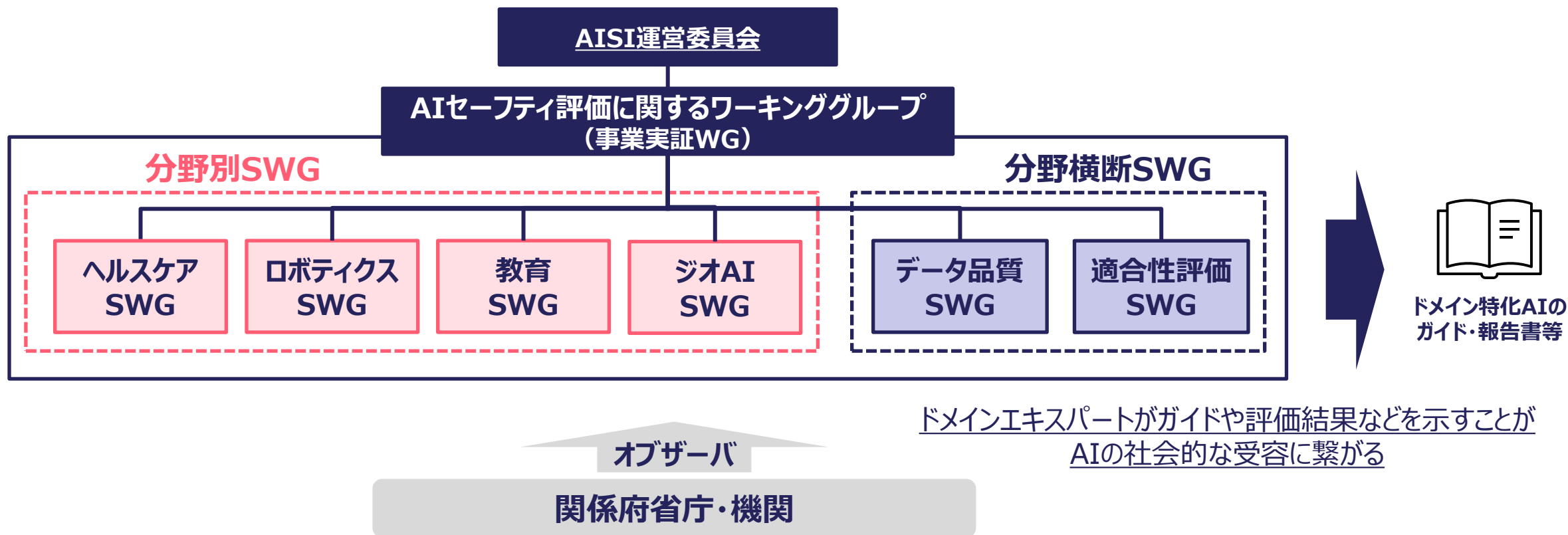
- ① AI安全性に関連する情報のハブとして、信頼できる優れたAI活用製品やサービスの普及を促進する（もって、国民生活の向上・社会課題の解決を促進）
- ② AI技術やAI安全性等に係る国際標準化等により、我が国のAI関連分野のイノベーションの加速・競争力強化を支える
- ③ 情報収集・分析・共有・対策の早期実施により、AIの開発・普及に伴う国民の生命・財産に危害が及ぶような事象など、国家の安全を脅かす事象の抑止に貢献する

- 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

AISIの活動について

AIセーフティ評価に関するワーキンググループの設置

事業実証WGのゴールは、産業界・行政・専門家が協働して、
AIの社会実装におけるAIセーフティの確保を支える仕組みを構築し
利用者のリスク理解を前提としたAIセーフティ評価の枠組みを整備すること

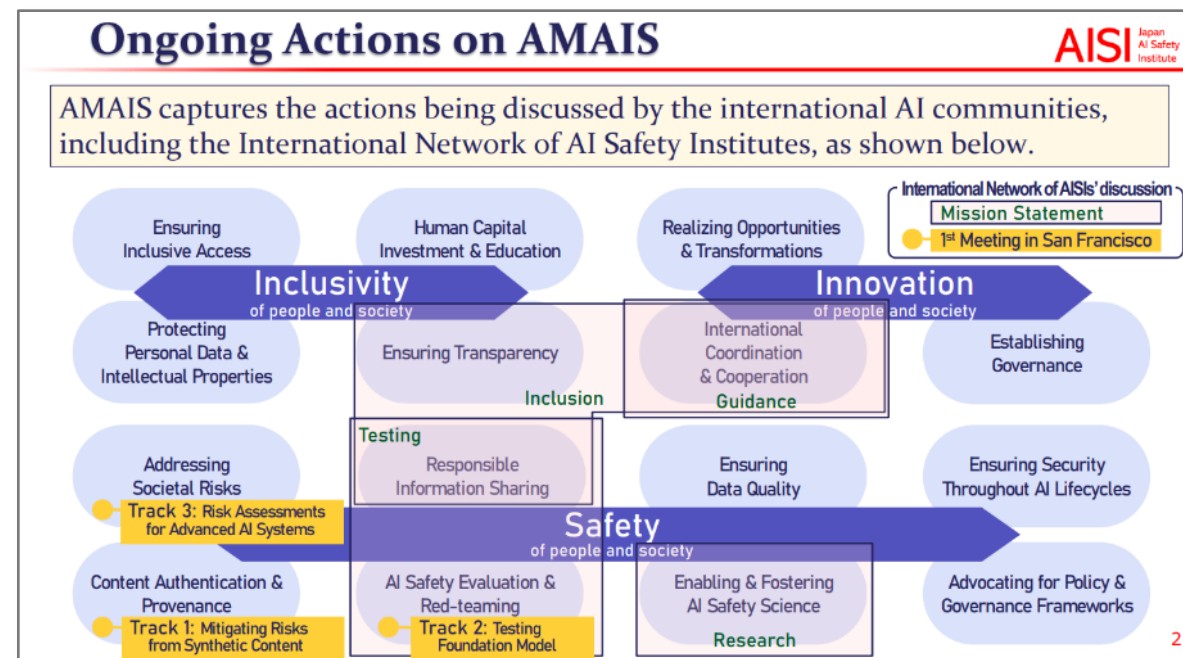


AIセーフティに関する活動マップ（AMAIIS）の概要

AIの安全性に関する活動が急速に変化・進化する中、
見落とされがちな部分や**活動間の相関関係を全体像として可視化**

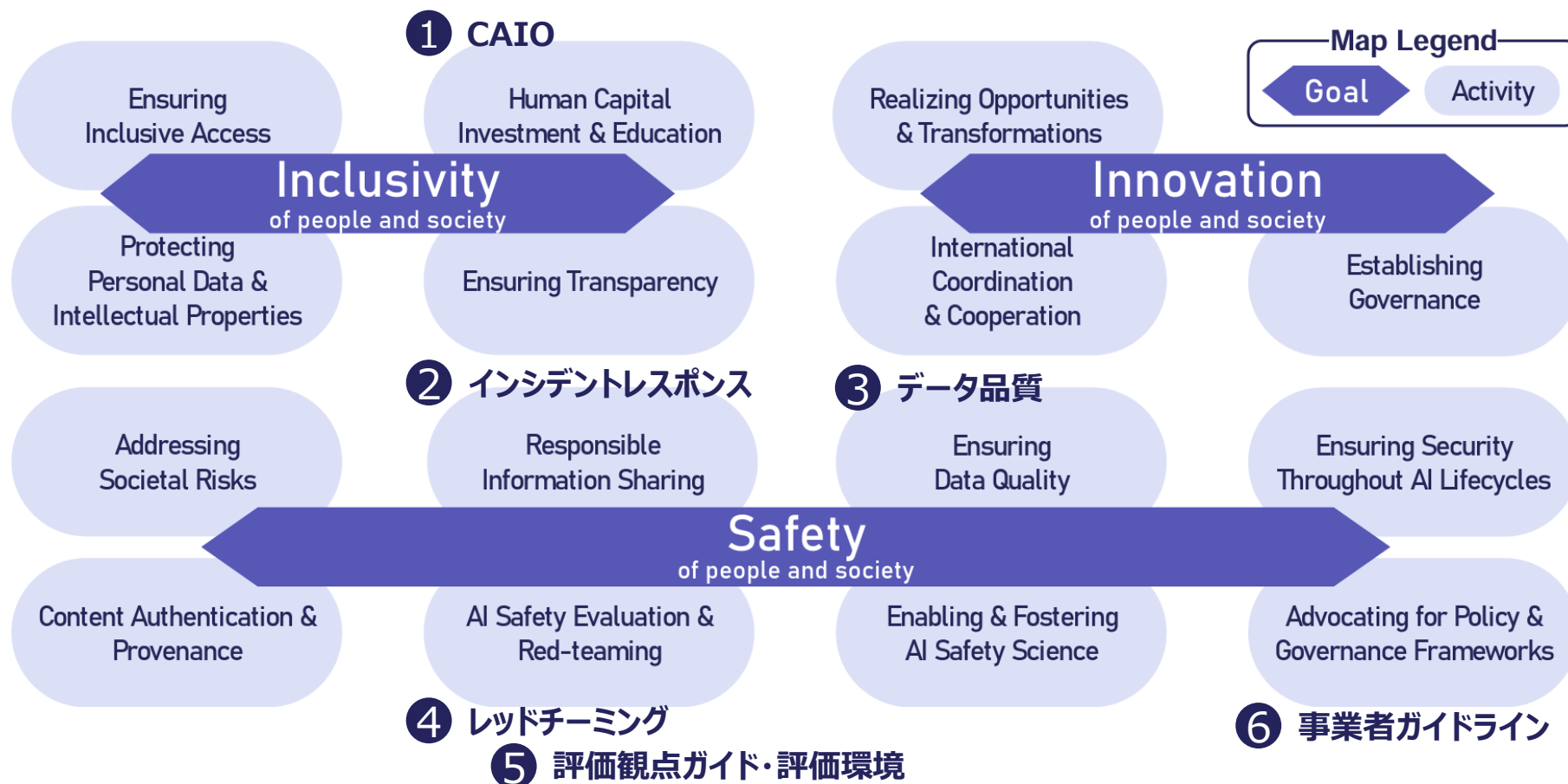
*AMAIIS: Activity Map on AI Safety

- ◆ AISIは、**ディスカッションペーパー**として「AMAIIS：AIの安全性に関する活動マップ」を公開。
- ◆ AISIは、**主要文献のベンチマーク**に基づき、包括的なアクティビティマップと関連用語を開発している。
- ◆ この日本主導の取り組みは、AIの安全性に関する国際的な協力体制の基盤をさらに強化し、持続可能で信頼性の高いAI社会の実現に貢献することが期待されている。



AMAISと現在の主な取り組み

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

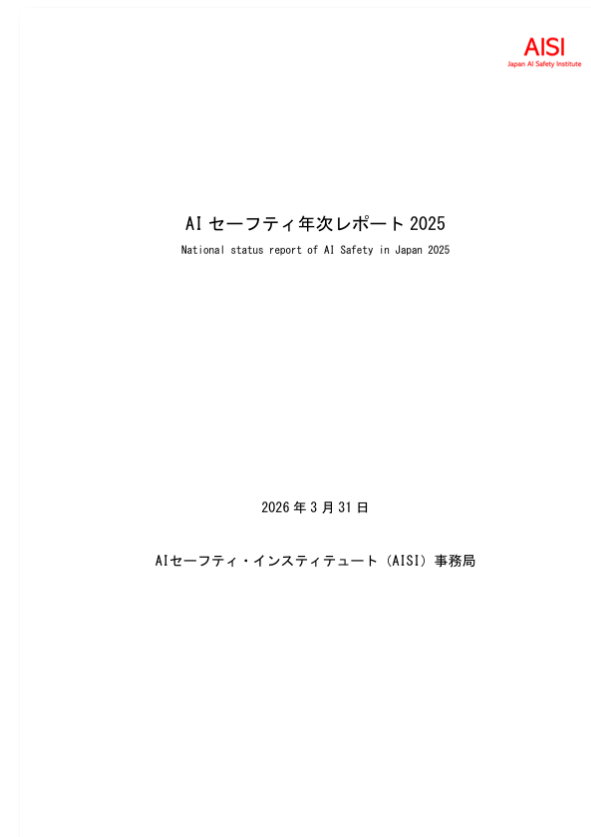
- AIシステムの安全性を評価する際の基本的な考え方を示したガイド

6. 事業者ガイドライン

-総務省・経済産業省がリ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の活動状況を「AIセーフティ年次レポート」として毎年度まとめている

- ◆ 「AIセーフティ年次レポート」とともに、関連するレポート等についても、年次レポートを補完する参考資料「AIセーフティ ファクトシート」として公開。
- ◆ 本稿においては、AIの急速な進展に対応するための、AISIIと国内外の関係機関や企業等との連携など、我々の今後の取り組みやその狙いについても記載している。



米国の呼びかけで10カ国が参加し、現在は英国がコーディネーターを担当

- ◆ 米国、英国、カナダ、EU、フランス、ケニア、オーストラリア、シンガポール、韓国、日本の10か国が参加。
- ◆ 技術ワークショップの開催、バイ会合の実施、所長級会合の開催などを通し、ネットワーク内で連携している。



AISI国際ネットワーク会合



日本AIS主催WS



星・日本AIS主催サイドイベント



AISI国際ネットワーク
所長級会合



内閣府・日本AIS主催
国際フォーラム



国連未来サミット

AISI

Japan AI Safety Institute