

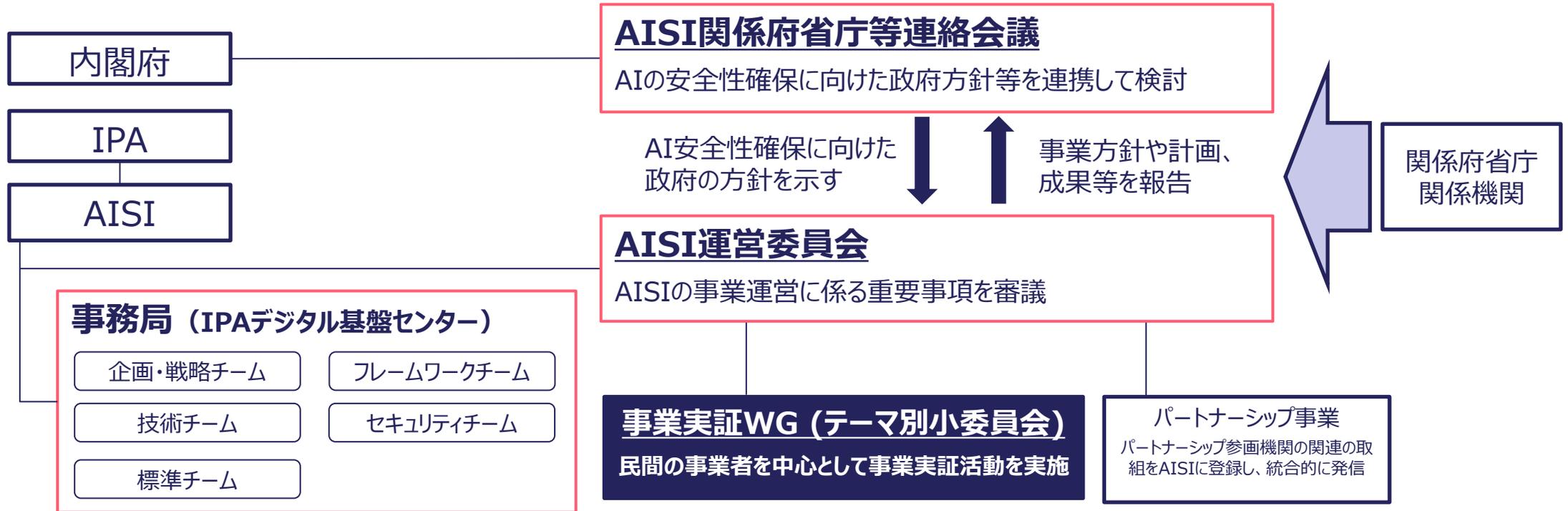
主催者挨拶

AIセーフティ・インスティテュート

事務局長 西村 卓 氏

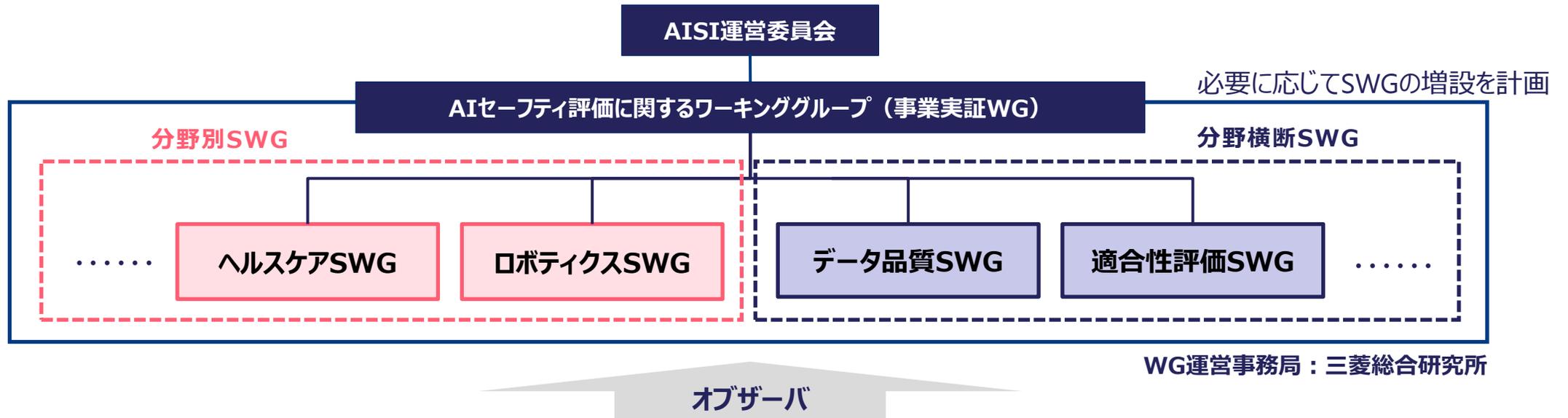
事業実証WGの設置について

2025年3月、AISII運営委員会の下に、テーマ別小委員会として、「AIセーフティ評価に関するワーキンググループ」（事業実証WG）を設置



分野別SWG及び分野横断SWGの設置

信頼とイノベーションが両立するAI社会の実現を目標に、社会・産業・政策の各レベルにおけるAIセーフティ評価に関する共通理解の醸成と具体的な実装に向け、事業者や技術者の取組みを支えることを狙いとしている



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 国家サイバー統括室 警察庁
デジタル庁 総務省 外務省 文部科学省 厚生労働省 農林水産省 経済産業省 国土交通省 防衛省
情報処理推進機構 (IPA) 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

事業実証WGの目的・ゴール

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**

中長期的な目的

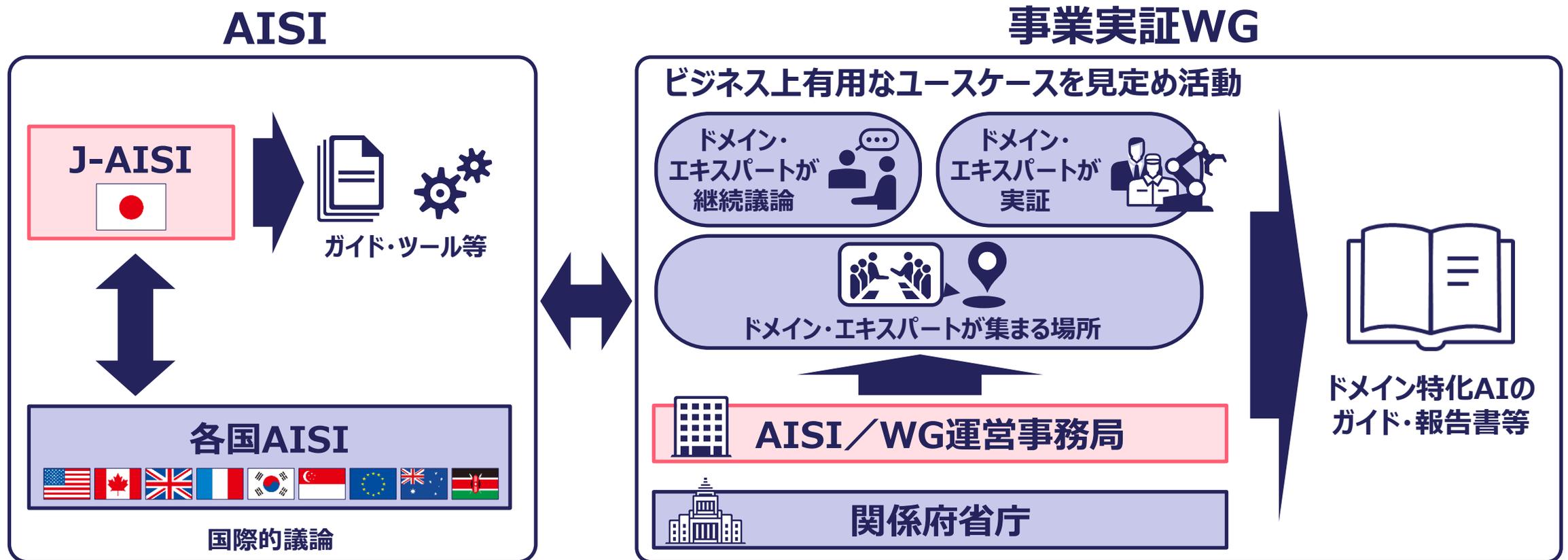
AIセーフティ評価の枠組みの整備を通じて：

- ① 各産業分野におけるAI技術の円滑な導入と普及を促進し、社会全体でAIの社会実装を実現することで、医療・労働・高齢化などの社会課題の解決に資する機会を創出する。
- ② 評価手法や観点がAI開発・提供事業者と利用者のいずれにとっても理解・活用しやすい共通言語となる環境を整備する。



事業実証WGの活動プロセス

事業実証WGはAISIIの取り組みを踏まえ、民間企業を中心とするドメイン・エキスパートが議論および実証を行い、ガイドや報告書を作成



ドメインエキスパートがガイドや評価結果などを示すことがAIの社会的な受容につながる

Hiroshima Global Forum for Trustworthy AI

日本のAI戦略の主要な柱の一つとしてAISIIの役割と重要性を国内外に発信。Trustworthy AIの実現に向けて、国際社会やAI安全性/セキュリティ関連の産業界・研究機関等と連携し、今後のAI安全性の道筋を探ることを目的として開催

- 日程：令和8年1月15日(木)
- 場所：グランドプリンスホテル広島
- 参加：国内 日本AISII、連携している企業・研究機関、関係省庁等
国外 各国AISII及びAI関係者、関連企業・研究機関等 **約150名（16か国・45名）参加**



Day1 (15日) : 各ステークホルダからプレゼン

- 日本のAI関連施策・今後の展望
 - AI政策/AI戦略/広島AIプロセス/AIの未来について
- AISII国際ネットワーク関係機関
 - US CAISI冒頭あいさつ
 - 各国AISIIおよび関連研究機関の活動報告
 - UK AISII閉会挨拶
- 関連企業との連携
 - **日本AISIIの事業実証WG**
 - AIセーフティ/セキュリティに関する企業の取り組み
 - Microsoft、PFN、NTT、NEC、富士通、OpenAI、Anthropic、Cisco
- 国際社会との連携
 - OECD、GPAI Tokyo、ASEAN、United Nations University

「Japan AISIIの取り組みとWGについて」

多様なステークホルダーが参画する事業実証WGを通じて、具体的なAI安全性検証を進めている旨を説明

西村事務局長

ヘルスケアSWG



ロボティクスSWG

データ品質SWG

適合性評価SWG



事業実証WGの成果について

AISI事務局
WG運営事務局

2026年3月10日
AISIB事業実証WG 下期報告会

各SWGの体制

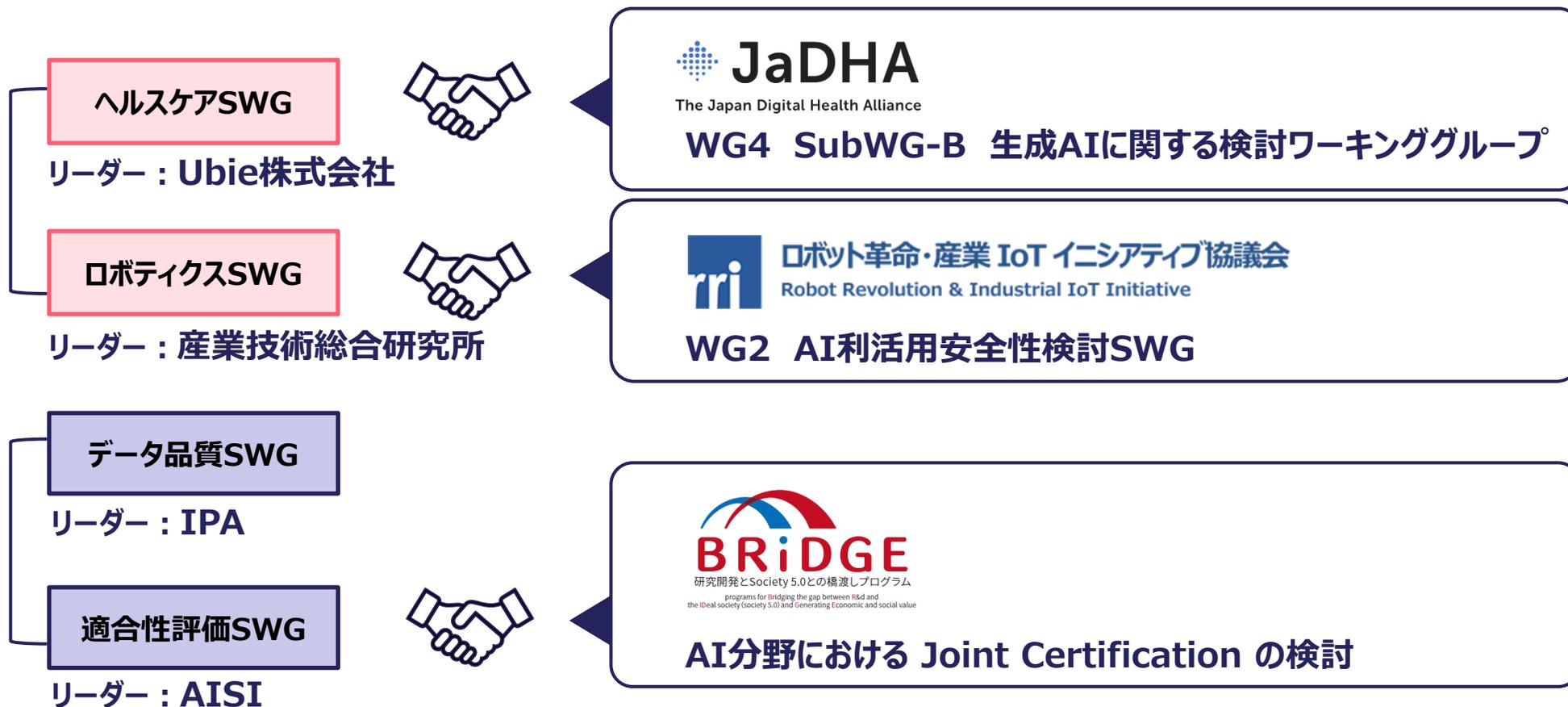
民間事業者を中心に多様なステークホルダーが参画し、連携を図る場を提供
AIセーフティ評価に関連する官民の取組との連携体制で実施

分野別SWG

国際的関心が高く日本が強みを持つ分野であり、AI利活用進展とAIセーフティ確保が重要な領域

分野横断SWG

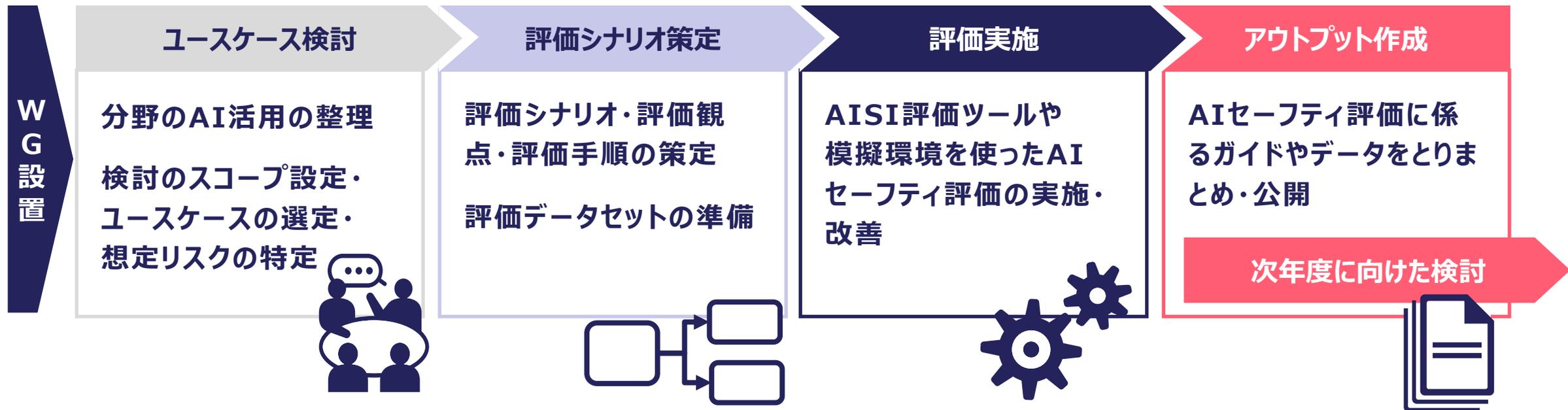
各分野のAIセーフティ評価に共通して求められる領域



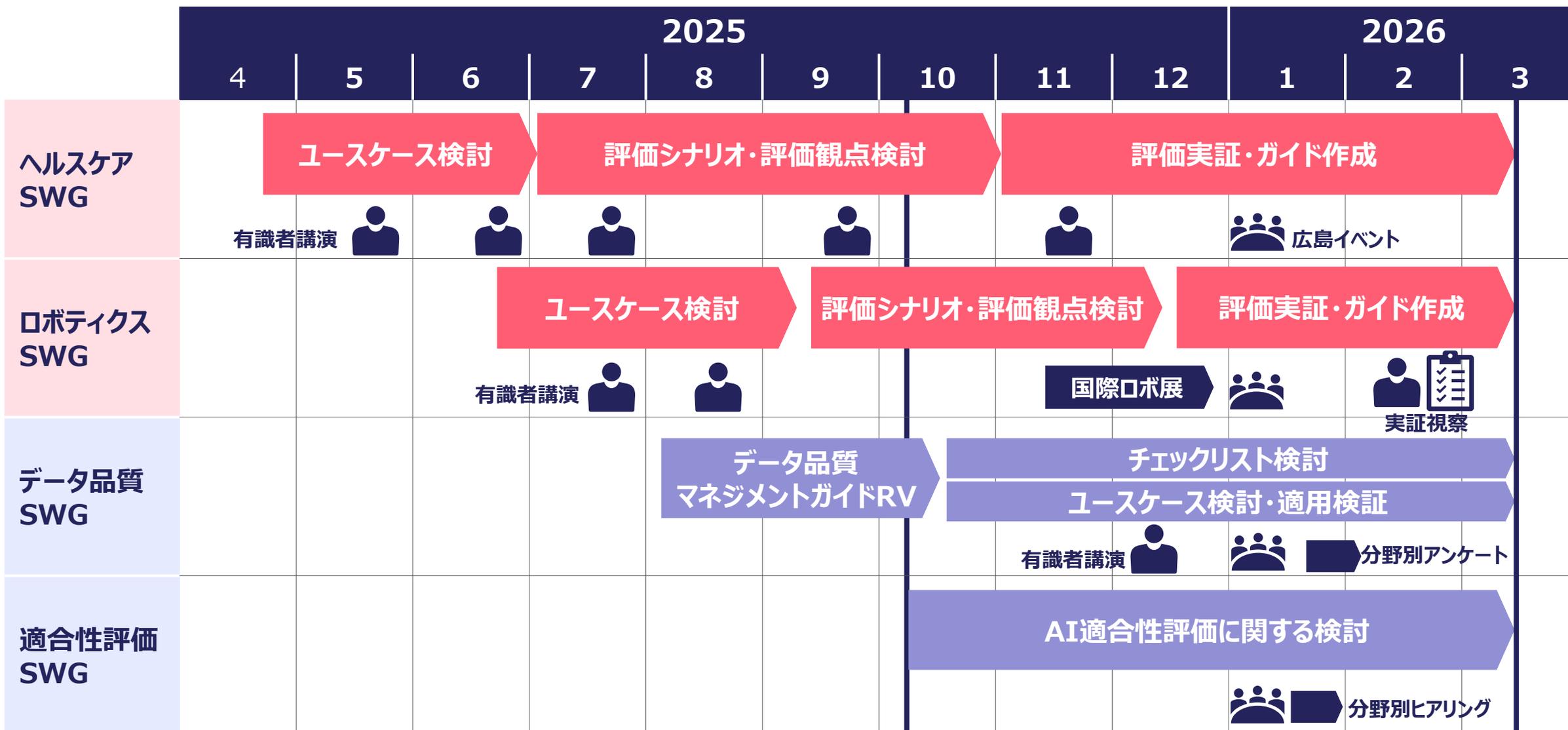
SWGの活動の進め方

分野ごとのAIセーフティ評価に関する見解をまとめ事業実証等の活動を推進
分野別のアウトプットを作り、AIセーフティ評価の普及とAI利活用を促進

SWGのWG活動の進め方（分野別SWGイメージ）



2025年度SWGの活動実績



上期報告会

下期報告会

事業実証WG2025年度の到達点

評価観点の具体化と実装構造の明確化、さらに評価実証の実施を通じて、
AIセーフティ評価を実務に落とし込むガイド（指針）を提示

ヘルスケアSWG:
10評価観点×開発ライフサイクルによる
実践的評価ガイド
(Non-SaMD)

ロボティクスSWG:
AIとロボットの融合により顕在化する
新たなリスクに対応するための
開発者・提供者向けガイド

データ品質SWG: AI利活用を前提としたデータ品質管理の体系化

適合性評価SWG: AI適合性評価に向けた論点整理

公開予定の2025年度アウトプット

各SWGの検討を基に作成したガイド・チェックリスト・レポートを公開予定

ヘルスケアSWG

ヘルスケア領域におけるAIセーフティ評価観点ガイド

- AISI評価観点をヘルスケア領域へ適用
- AIプロダクト開発フェーズごとのAIセーフティ評価実践（チェックシート）



ロボティクスSWG

AIロボティクスに関するセーフティ評価観点ガイド

- AIロボティクスのリスクと論点
- ユースケース事例の評価観点（カフェ商品搬送ロボ・遠隔操作型移動ロボ）



データ品質SWG

データ品質SWG活動報告書

- データ品質マネジメントガイドブック更新版
- データ品質管理チェックリスト



適合性評価SWG

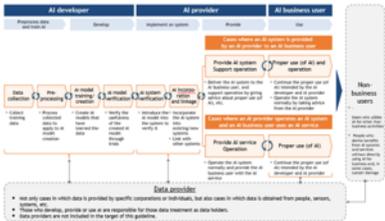
適合性評価SWG活動報告書



分野ごとのAI利活用シーンにおけるAIセーフティ評価への取り組み

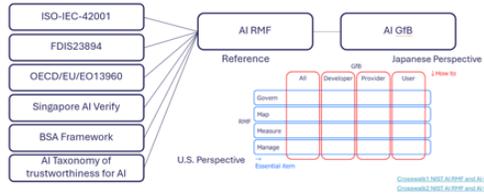
AI事業者ガイドライン

総務省と経産省が策定・更新



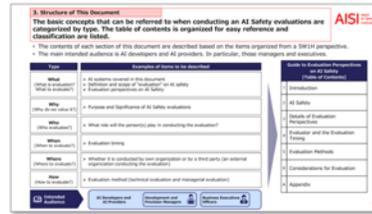
クロスウォーク

国際的な相互運用性の確認



評価観点ガイド (LLMシステム)

10の評価観点を整理



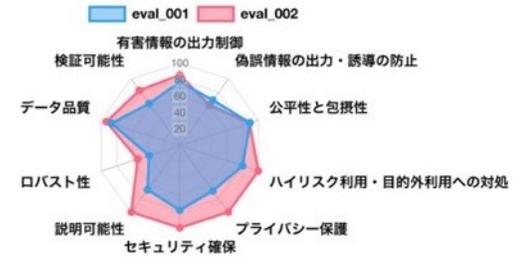
レッドチーミング手法ガイド

レッドチーミング手法の基本的な留意事項



AIセーフティ評価環境 (OSSツール)

ガイドに基づく評価ツール



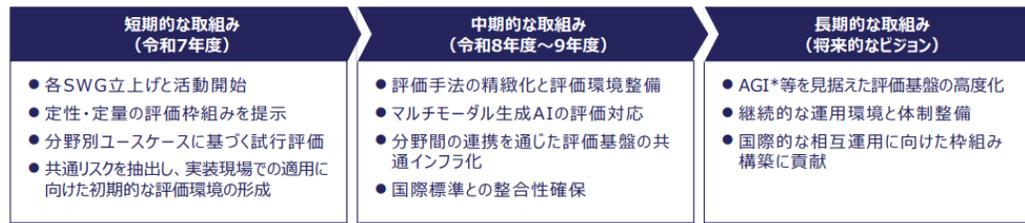
事業実証WG ビジョンペーパー

活動の方向性と目的を示した文書



2025年度は重点分野のユースケースを通じて初期的な評価基盤を形成

事業実証WGのロードマップ



具体的なユースケースを検討

選定したユースケースを起点にリスクと評価観点を整理

ヘルスケアSWG



ロボティクスSWG

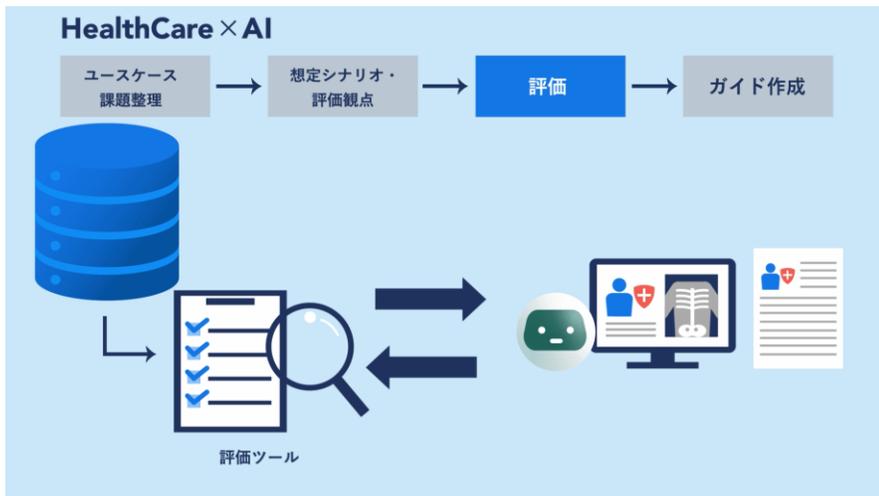


データ品質SWG



評価シナリオ・評価観点を作成
評価方法の妥当性を評価実証により確認

ヘルスケアSWG



ロボティクスSWG



データ品質SWG



WG活動で得られた示唆と取組の方向性

次年度は、SWGの検討・評価実証を通じて得られた
AIセーフティの構造的論点に対応

2025年度に得られた示唆

文脈依存性

分野・用途・利用主体によりリスクの質が変化

新技術への動的対応

エージェントAI、フィジカルAI、マルチモーダル等

AIの社会受容

安全・信頼を担保する枠組みの必要性



2026年度の方向性

ユースケース検討・評価実証の拡張

新技術領域への適用検証の実施

第三者評価の確立に向けた検討の加速

その他：普及コンテンツ公開

AIセーフティ評価の普及・啓発のための動画コンテンツを作成・公開
ヘルスケアSWG、ロボティクスSWG紹介動画も作成

#1 安全・安心で信頼できるAIに
求められる「AIセーフティ」とは



#2 「AI セーフティに関する
評価観点ガイド」紹介動画



Coming soon!

#3 「AIセーフティに関するレッド
チーミング手法ガイド」紹介動画



事業実証WGのウェブサイトを立ち上げ SWGの活動成果を順次公開予定



→ 活動・成果報告コンテンツ

AI活用促進に向けたAIセーフティ評価に関する事業実証WGでは、ヘルスケアやロボティクスといった分野別のサブワーキンググループ（SWG）と、データ品質や適合性評価といった分野横断のサブワーキンググループ（SWG）を同時に進行させつつ、実務的なAIのユースケースに基づき、業界や利用者の立場に根差した実践的なAIセーフティ評価を体系化するとともに、具体的な評価シナリオやデータセットを構築する活動を進めています。

- 事業実証WG**
事業実証WGの活動全体に関するコンテンツを掲載
- ヘルスケアSWG**
ヘルスケア分野におけるAIセーフティ評価に関する事業実証の取り組み等
- ロボティクスSWG**
ロボティクス分野におけるAIセーフティ評価に関する事業実証の取り組み等
- データ品質SWG**
データ品質マネジメントガイドブックを基に、実装への具体化、評価ツールの検討等
- 適合性評価SWG**
AI分野における適合性評価の手法確立、実運用を見越した国内体制構築の検討等
- 動画・配信アーカイブ**
事業実証WGの活動を中心に、AISIIに関連するコンテンツを掲載
- リンク**
関連情報へのリンク