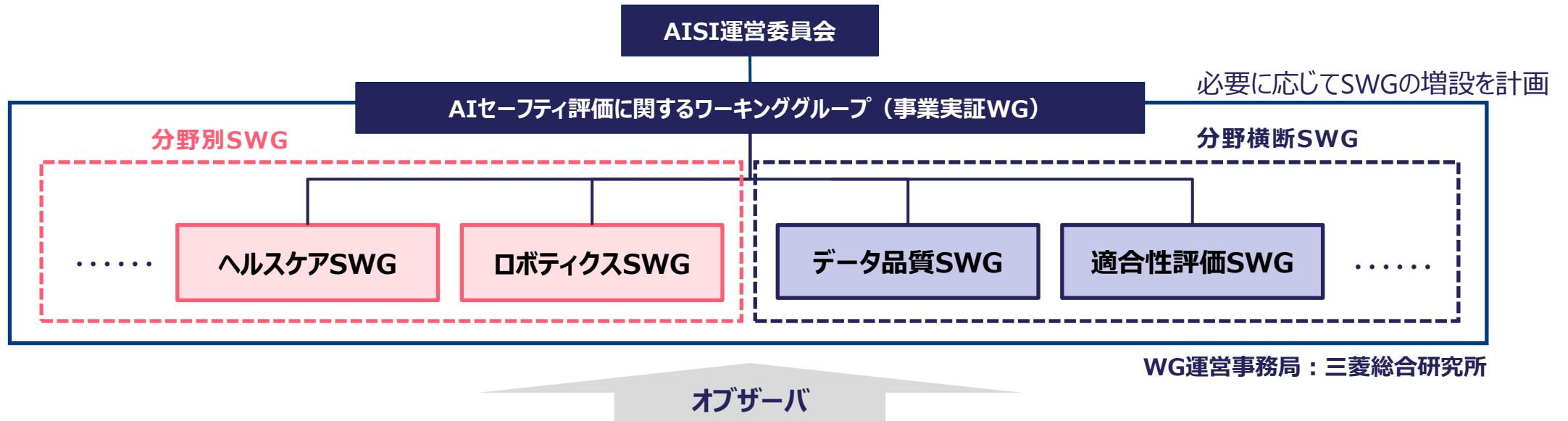


分野別SWG及び分野横断SWGの設置

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

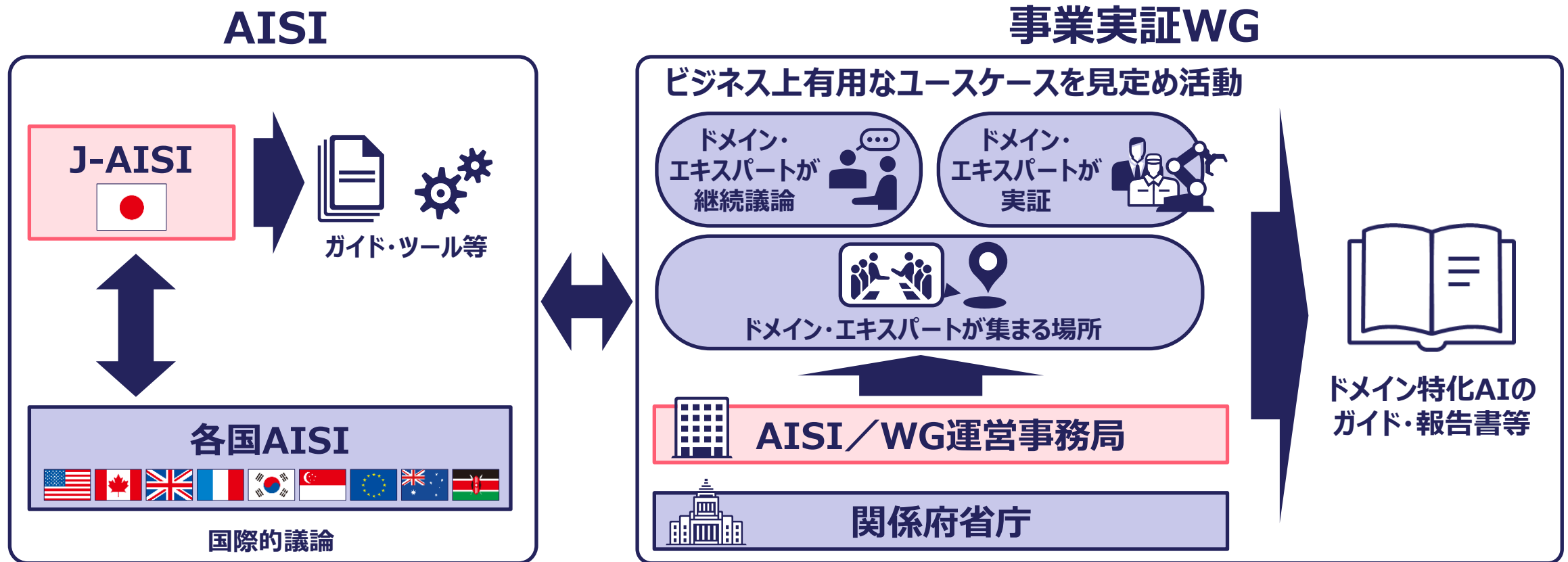
AIセーフティ評価の枠組みの整備を通じて：

- 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

2026年度以降のAISIの取組

これまではAIの評価観点を精査し、「物差しを作る」作業が中心。今後、物差しを使って実際にAIを「評価する」ことが必要。

物差しを作る

評価観点ガイドの策定及び評価ツールの整備への着手

- ・2025年より開始
- ・2026年はさらに拡充



Guide

評価観点ガイド

LLMシステム

AIエージェント

評価する

「物差し」に応じたAIセーフティ評価

- 2026年度以降
評価環境の構築に本格着手



評価環境

LLMシステム

AIエージェント

PhysicalAI

AIモデル、システムのリスク把握
フロンティアAIへの対応

- ・分野ごとに特化したリスクの把握
 - ・官民一体となった検証、評価体制
- ⇒安心・安全で信頼できるAIの開発・活用へ

信頼に足るPhysical AIに向けて

AIによりリスクを検出し学習、ヒューマンマシニングにおける安全性を確保するPhysical AI評価環境を構築

