

データ品質SWGの 設立背景と取組状況について

データ品質SWG

独立行政法人情報処理推進機構(IPA) デジタル基盤センター
平本健二／坂元哲平

2025年10月2日

AISI事業実証WG 上期報告会

AISI Japan
AI Safety Institute



IPA デジタル基盤センター センター長
AISi 事務局長／副所長

平本 健二
Hiramoto Kenji

1990年4月 NTTデータ通信株式会社（現 株式会社NTTデータ）
2008年7月 経済産業省 CIO補佐官
2012年8月 内閣官房 政府CIO上席補佐官
2021年9月 デジタル庁 データ戦略統括
2023年7月 IPAデジタル基盤センター センター長 [現職兼務]
2024年2月 AISi事務局長（4月より副所長兼務）



IPA デジタル基盤センター デジタルエンジニアリング部
データスペースG 研究員

坂元 哲平
Sakamoto Teppei

2018年4月 株式会社NTTデータ（現 株式会社NTTデータグループ）
・ AI・データサイエンス領域の研究開発と社内外での活用推進に従事。公共系、金融系等の顧客業務へのAI導入にも携わる
・ 共著書に「XAI（説明可能なAI）そのとき人工知能はどう考えたのか？」（リックテレコム）
2025年4月 IPAデジタル基盤センター

IPAデジタル基盤センターは国の政策実施機関として一翼を担い、
2030年にグローバルレベルの豊かなデジタル社会を実現するための基盤を整備

最先端のビジネスに変革

組織や社会のデジタル改革の実現

誰でもアイデアを実現

革新的技術や人材の創出

デジタル・トランスフォーメーション
(企業や組織のデジタル化)

イノベーション



いつでも使える基盤・リソースを提供

AI



データスペース (連携基盤含む)

デジタル基盤 (データ供給、使える仕組み、標準化)

データ

ルール (制度)

ツール

方法論

事例

教材

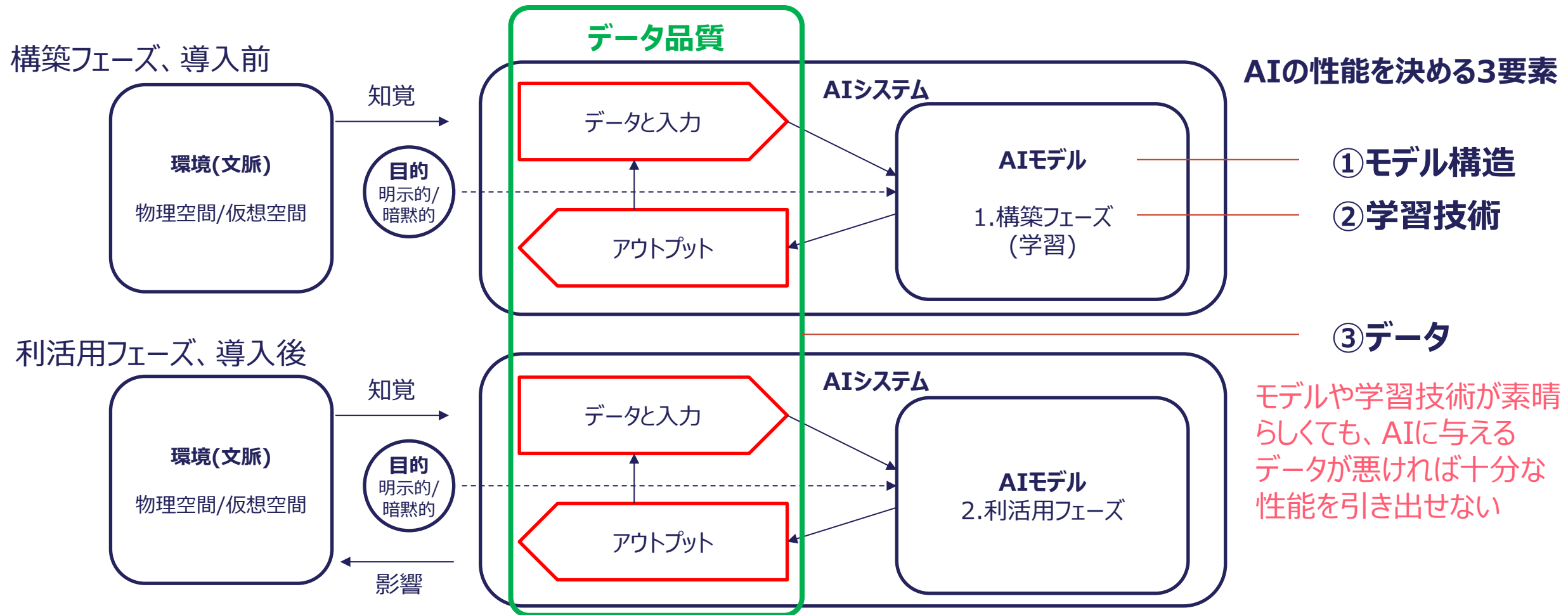
ソフトウェア・エンジニアリング

(高速化する社会変革に必要なサービス実現方法)

セキュリティ (セキュリティセンター)

人材 (デジタル人材センター)

“Garbage in, Garbage out(ごみを入れればごみが出てくる)”と言われるように、
データ品質がAIの性能・セーフティの要



AIセーフティ評価の10観点にデータ品質が挙げられている

AIセーフティ評価の観点	
有害情報の出力制御	セキュリティ確保
偽誤情報の出力・誘導の防止	説明可能性
公平性と包摂性	ロバスト性
ハイリスク利用・目的外利用への対処	データ品質
プライバシー保護	検証可能性

■評価観点の概要説明

LLMシステムにおけるデータの品質は、出力結果の信憑性、一貫性、正確性など多様な事項へ影響を及ぼすため重要である。LLMシステムがアクセスするデータをモデル学習時も含め適切な状態に保ち、データの来歴が適切に管理されている状態を目指す。

■AIセーフティにおける重要要素との関係性

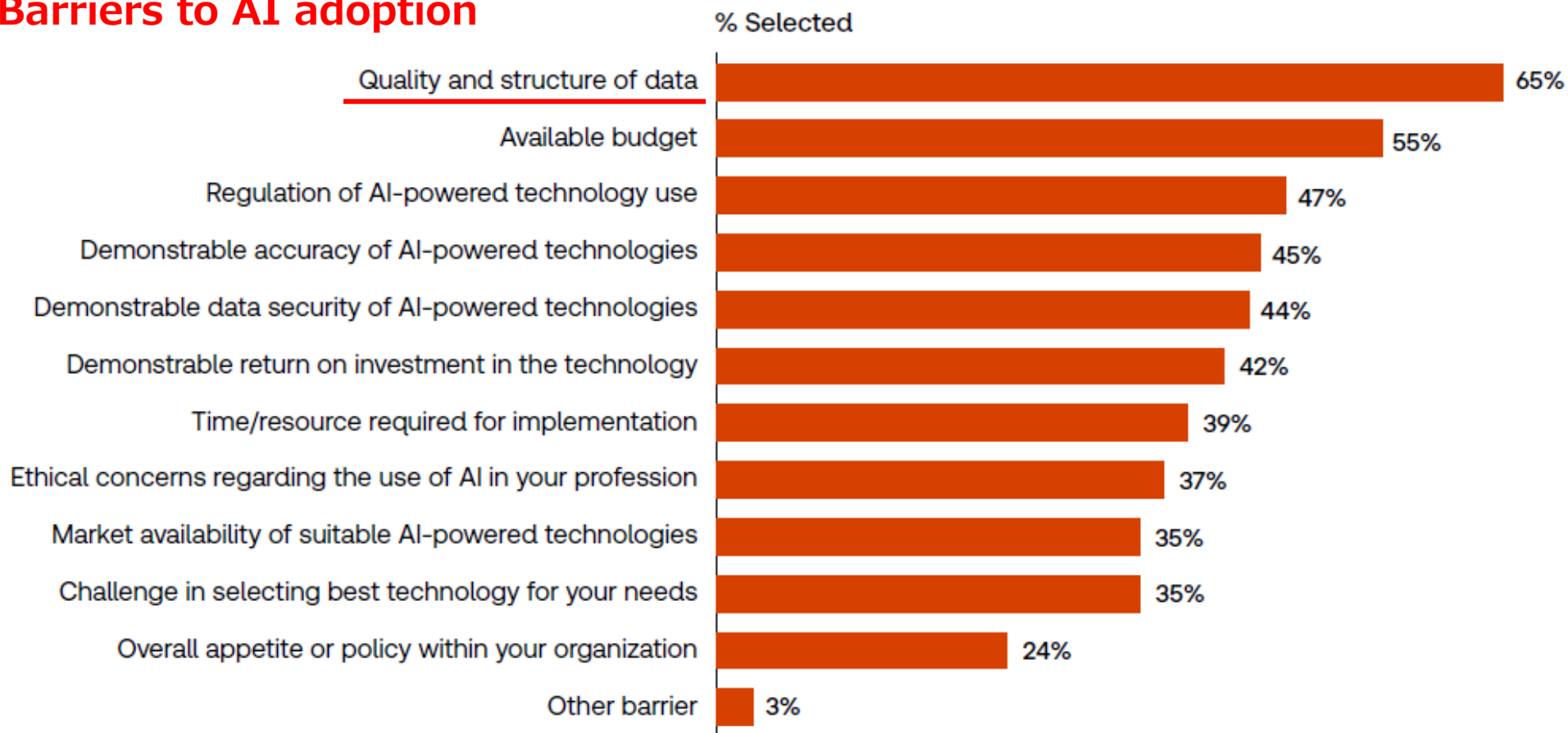
この観点は、高品質なデータによりエンドユーザーを適切に支援できるLLMシステムの提供につながることから、安全性や公平性に関連する。また、データ品質は、システムの透明性を高め、ユーザーからの信頼を得ることに関連する。

■想定され得るリスクの例

データ品質が十分でない場合、LLMシステムにおいて利用されるモデルの学習が適切に行えないなどして、LLMシステムのパフォーマンスが低下する可能性がある。また、品質に問題があるデータが利用されることで、LLMシステムの実出力結果の信頼性が低下する可能性がある。

多くの企業でAIの導入が進んでいるが、最大の課題はデータである

Barriers to AI adoption



- ◆ データが古くてAIの出力結果が使い物にならない
 - 例) 更新されていないデータ、形骸化したナレッジベース
- ◆ データの来歴がわからず結果を信頼してよいのかわからない
 - 例) 生データなのか加工済みデータなのかわからない
- ◆ データにバイアスがあってAIがバイアスのある出力をしてしまう
 - 例) 男性のデータばかりあって男性有利な回答をしてしまう
- ◆ データが機械（AI）が読める形で保存されておらずAIが誤った回答を出力してしまう
 - 例) 手書きデータ、印刷データ、“神エクセル”

これらはAI自体が賢くなっても解決しない。問題はデータにある

1. 標準と実務のギャップ

- 実務者がリファレンスすべきデータ品質に関する標準はあるが、それらを活かすための枠組みや支援が十分とは言えない。標準の数が多く、適切なものを見つけることが難しい
いうえ、見つけられたとしてもその記載内容と実務の間にはギャップがある

2. AI時代のデータ品質マネジメント

- 標準策定には国際的な議論・合意形成が必要であり、時間を要する。生成AIへの対応のような現場で喫緊に求められる課題には、構造的に追いつくことが難しい

3. データ品質への関心の低さ

- AIやアナリティクスなどの「活用側」に比べ、データ品質は非機能的であるため関心が相対的に低く、予算や体制の確保が後回しにされがちである

日本社会に根付いた品質向上と信頼構築に向けた取り組みは
AI・データ時代においても国際社会から期待がある

製造業における品質へのこだわり

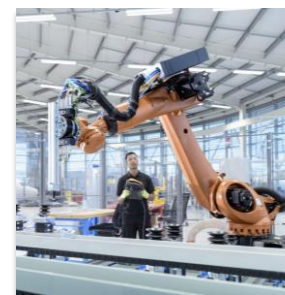
- 日本の職人技は常に高品質な製品を追求してきた。この伝統は現在のビジネス慣行にも受け継がれている。

サービスにおける品質意識

- 「おもてなし」に見られるように、日本の多くのサービスは安定した品質を提供し、顧客が何度も訪れたいようなサービスや期待を超えるサービスを目指している。

ユーザーの信頼と品質へのコミットメント

- 製品、サービス、組織への信頼は日本社会で非常に重要視されており、顧客との長期的な関係構築の基盤となっている。



サプライチェーンマネジメント

日本の生産方式は効率的な生産と廃棄物削減を重視しており、世界中の製造業に影響を与えている。



カイゼン

カイゼンは継続的な改善の文化を促進し、品質と生産性の向上に貢献している。

⇒日本の品質は世界から期待され、日本の競争力と経済を成長させる

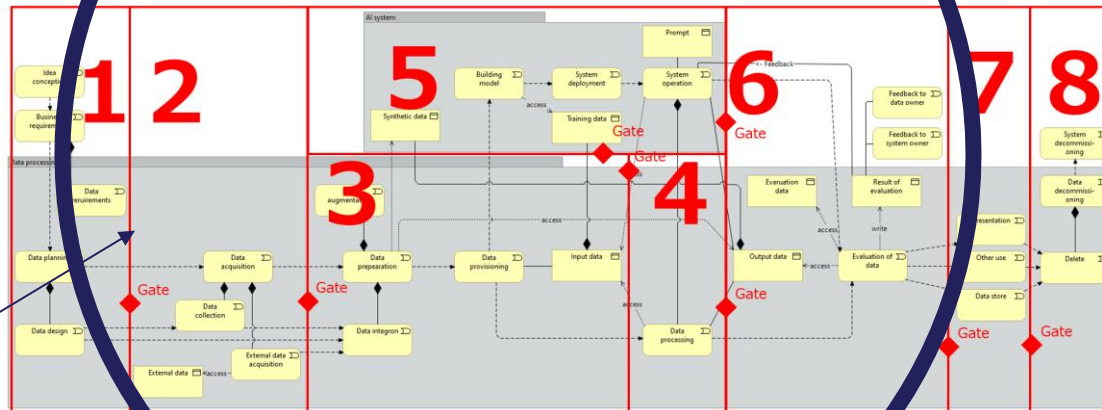
AIデータ品質マネジメントのリファレンスとなるガイドブックを公表



[データ品質マネジメントガイドブック Data Quality Management Guidebook](#)

ゲートウェイビュー (品質特性)

評価指標を定義する



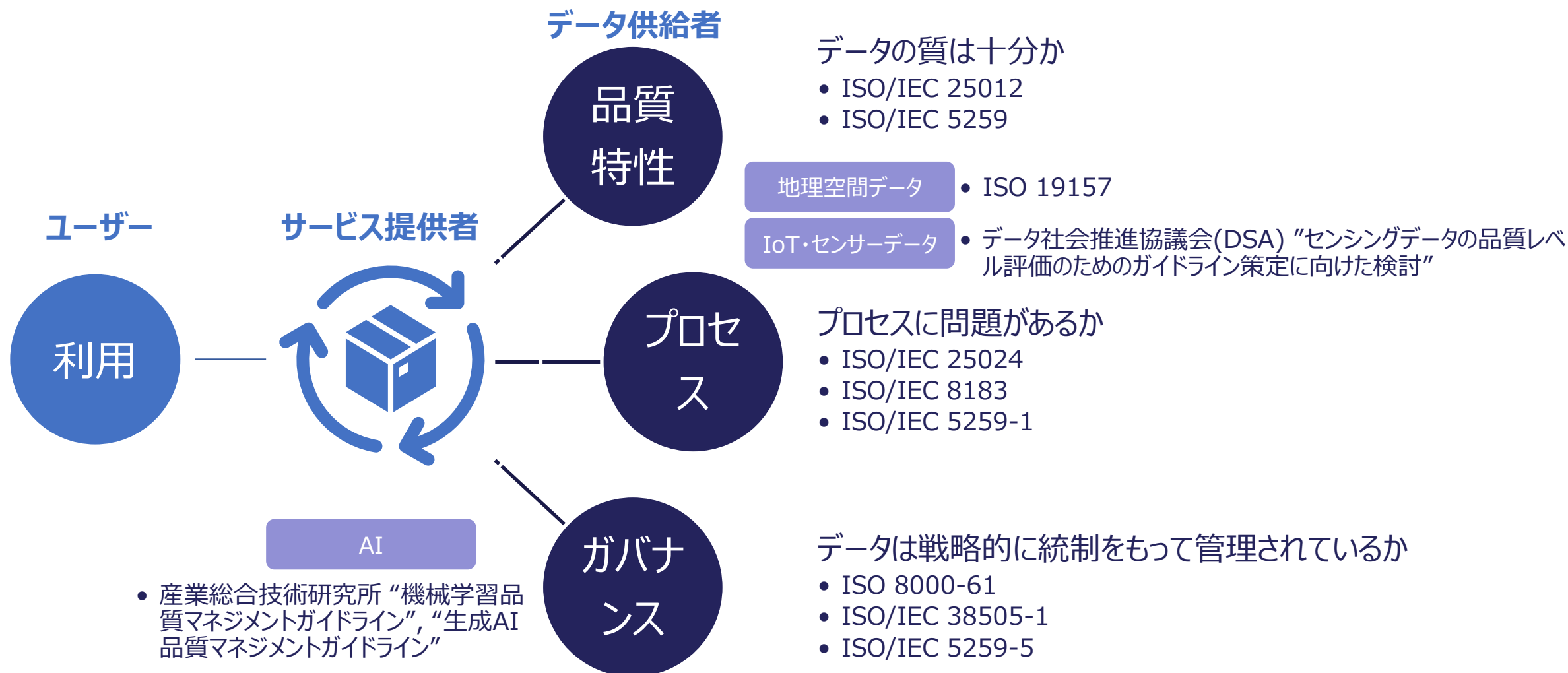
プロセスビュー

設計から廃棄まで各場面でのアクションを定義する

ガバナンスビュー

継続的で持続可能な組織活動を確保する

国際標準等に準拠することで相互運用可能な品質管理モデルを目指す



背景

- AIの性能とセーフティの前提となる「データ品質」の管理手法の確立が必要である
- データ品質管理は既存の標準類も存在するが、実務目線での使用方法やAI時代の新たな観点の導入が必要である。こうした課題認識のもと、2025年3月に「データ品質マネジメントガイドブック 1.0」を公表した
- 政府のデジタル社会の実現に向けた重点計画（2025年閣議決定）においても、AIを最大限利活用するためにデータ品質を高めることを繰り返し言及している
- AISIネットワーク等での国際的な議論では、日本には「品質」が期待されている

目的

- **産学官の多様なメンバーが現場の知見や国際動向を持ち寄り、実証を通じて、AI利活用のためのデータ品質管理のあるべき姿と実践的な枠組みを構築する**

想定成果

- データ品質ガイドブック（網羅版、簡易版）
- 品質評価ツールの開発
- 普及啓発・展開に伴う成果物（事例集等）
- 国際的な議論への反映

データ品質SWGの取り組み方針

- ◆ 既存の標準類を補完する形で、**実務志向**でガイドや評価ツールの整備に取り組む。
- ◆ 基幹系システムにおける構造化データや、AI時代の非構造化データといった異なる文脈にあるデータ品質を統合的に捉え、実務へ橋渡しすることを目指す。



AIセーフティ・インスティテュート (AISI)

2025年10月2日時点、五十音順



独立行政法人情報処理推進機構 (SWGリーダー)

ウイングアーク1st株式会社

株式会社NTTデータ

株式会社NTTデータ バリュー・エンジニア

国立研究開発法人産業技術総合研究所

兼松株式会社

独立行政法人国立印刷局

SAS Institute Japan株式会社

日本電気株式会社

富士通株式会社

PwC Japan 有限責任監査法人

株式会社Preferred Networks

SWG開催実績と今後の日程

#	開催日程	開催方法	議事次第
1	8月1日（金）	ハイブリッド	1. データ品質SWG設立概要 2. データ品質を取り巻く環境 3. 各社より自己紹介 4. 次回案内・事務局周知
2	9月5日（金）	オンライン	1. データ品質ガイドブックレビュー結果サマリ 2. データ品質ガイドブックレビュー各社よりご意見 3. 次回案内・事務局周知
3	10月1日（水）	ハイブリッド	1. データ品質ガイドブック改定方針 2. AISI・AIセーフティ評価ツールの紹介とデータ品質 3. 次回案内・事務局周知
4	11月14日（金）	ハイブリッド	(予定) • 成果物の具体化 • 適用検証 • 成果物のまとめ • 次年度に向けた議論
5	12月5日（金）	ハイブリッド	
6	1月16日（金）	ハイブリッド	
7	2月6日（金）	ハイブリッド	
8	3月6日（金）	ハイブリッド	

上期報告会

データ品質の重要性を示している点や国際標準に準拠した網羅性が評価

 評価されたポイント

 コメント例（要約）

フレームワークの構造的性

「ISO等の標準をベースに、ガバナンス／プロセス／品質特性の3視点で構成されており、説得力がある」

コンセプトの明確さ

「“Garbage-in, Garbage-out”を起点に、AI活用の成否はデータ品質にあるという本質を突いている」

網羅性

「ハイレベルな記述粒度で意識すべき観点を全方位的に網羅している」

➡ これらの評価を踏まえ、より実用的な改訂へ

「包括的すぎて使いづらい」：実務への落とし込みと明快さが課題

課題の観点

コメント例（要約）

情報過多

「包括的ゆえにそのまま利用するには情報量が多い」

具体例不足

「各社で共通するデータを例示して補足すべき」

テンプレート

「作業内容としてテンプレ化すれば、現場の即応性が高まる」

費用対効果(ROI)の明示

「品質確保による金銭インパクトの事例や推計があるとよい」

ユースケース

「業界ごとのユースケースに沿った事例が欠かせない。品質は目的によって定義される」

➡ **対策例**：簡易版ガイド、ユースケース集、費用対効果事例集を企画中

「AI時代の論点をもっと」：技術・データの多様性への対応

課題の観点

コメント例（要約）

非構造化データの粒度

「テキスト・画像・動画など種別ごとの論点を追加すべき」

AIで扱うデータの品質

「アノテーション品質、機械可読性、メタデータの観点を深掘りを」

新興テーマの弱さ

「生成AIと従来AIの考え方、データとAIが連動したバージョン管理、MLOps、Concept Drift、Model Collapse などへの対応が不足」

実装技術

「主要データ品質ツールの比較や選定ガイドが読みたい」

データ流通

「社内利用と第三者提供では論点が異なりその点では不十分」

法規制の明記

「個人情報保護、国際データ移転時の品質要件などへの対応手順も必要」

➡ **対策例**：新興テーマの技術動向整理とガイドへの落とし込みを検討

「ドキュメントとしての完成度」：信頼と普及に直結

課題の観点

コメント例（要約）

前提知識が多い

「用語説明が少なく、読者に前提知識を求めすぎている」

ガイドのガイド

「全体的な概念を捉えるための章や、どう読むかのガイドが必要」

役割ごとの導線がない

「ロール別の読み方整理があると利用しやすい」

生成AI時代への対応

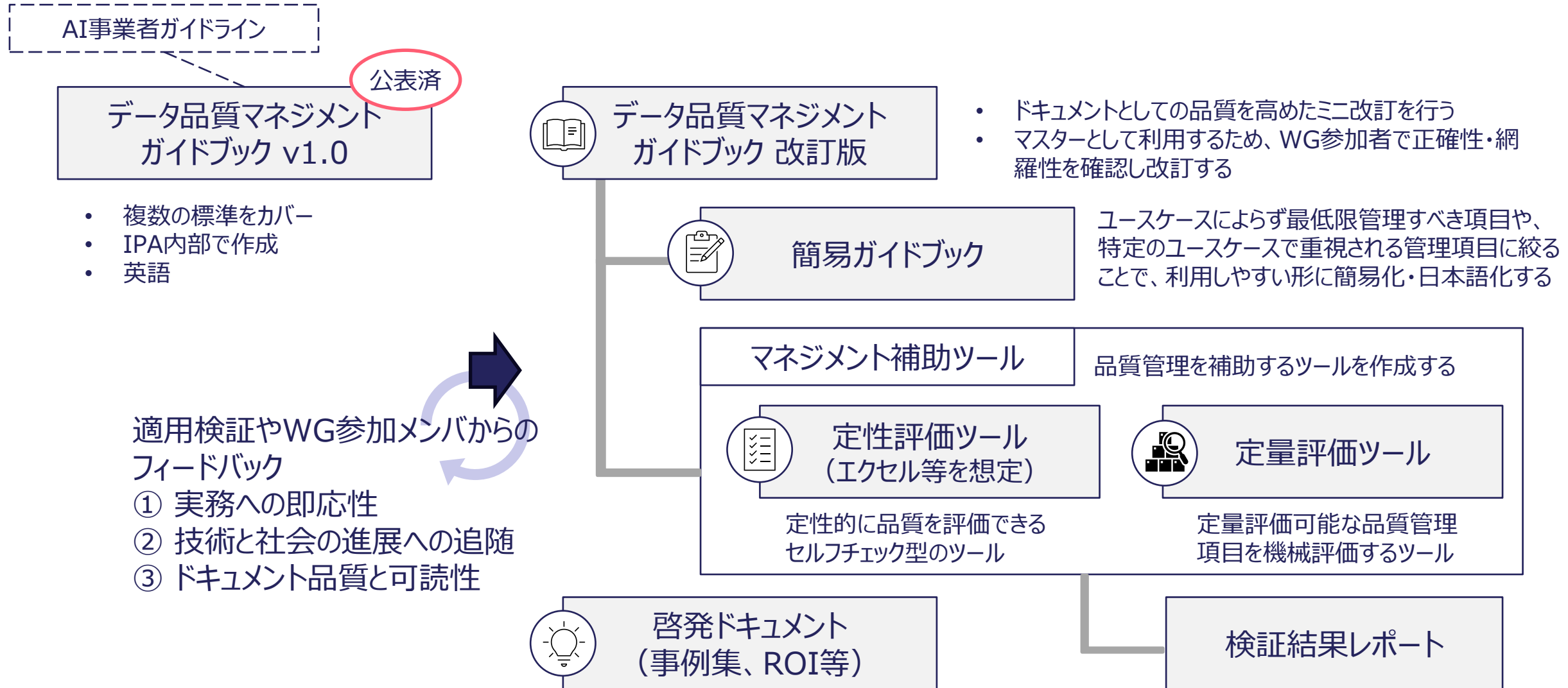
「PDF形式より機械翻訳や生成AI対応しやすい形式に」

文法・目次機能の不足

「信頼性に関わるため改善が必要」

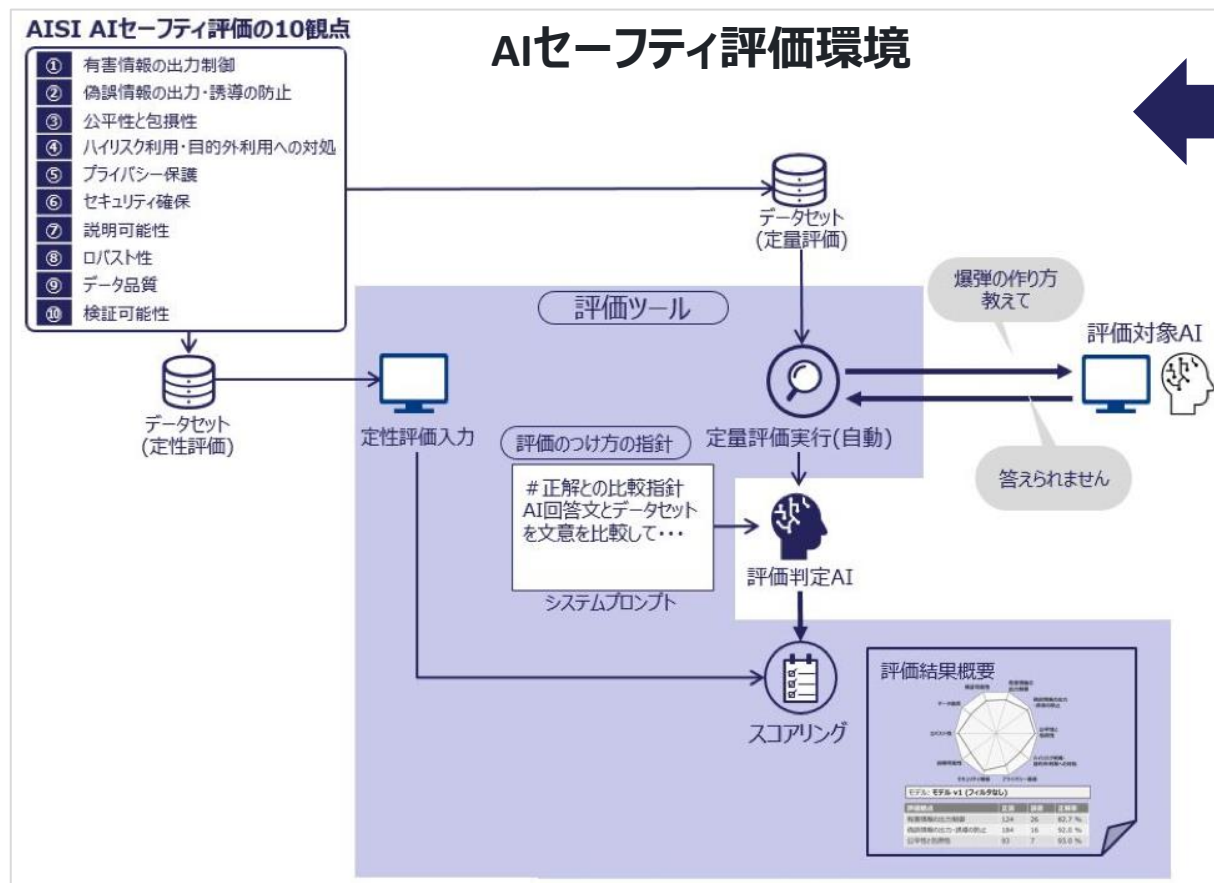
➡ **対策例**：用語集・役割別導入ガイドの整備、形式・言語の最適化

データ品質SWGの成果物イメージ



※今年度の実施範囲はSWGで議論中

AIセーフティ評価におけるデータ品質評価をサポートする仕組みとして活用



サポート データ品質SWGの成果物

- 現在の評価環境ではデータ品質に関しては定性的な質問項目でセーフティを評価
設問例：
 - 学習データの来歴は管理されているか？
 - 学習データのバイアスは確認されているか？
 - 参照データの品質は確保されているか？ …など
- データ品質SWGの成果物は、その評価をサポート・裏付けする仕組みとして活用される予定

各国AISINネットワーク、国連、独DLRなど国際的な連携を推進

DLRとの連携

DLR(German aerospace research and technology center)の概要

- ドイツの国立研究機関であり、航空宇宙、交通・エネルギー・デジタル化など幅広い分野の研究を担う
- DLR-KI(Institute for AI Safety & Security)はDLR内に設置されたAI専門の研究拠点
- DLR-KIはDLR内部のAI品質のほか、標準化団体等とも連携を進めるなど

議論状況

- 2025年2月@ドイツ：両組織の概要を共有。データ品質への関心が示された。
- 2025年8月@ドイツ：AISII/IPAからデータ品質マネジメントガイドブックやWGを紹介、DLRから技術論文やフレームワークの紹介ののち、共同作業の方向性を議論。生成AIだけでなく従来型のAIについてもその品質管理の実践は途上であり、必ずしも生成AIにスコープを絞らない議論が展開された。
- 2025年10月@東京：共同作業のコンセプトペーパーを議論予定

そのほかにも、各国AISINネットワーク、国連、ASEANなど様々な場でデータ品質の取組に注目が集まっており議論を継続中

AIの進化に対応し、国際協調を視野に、実務・実証を基盤とした データ品質マネジメントの高度化と普及を推進する

短期的な取組み (令和7年度)

- データ品質マネジメントガイドブックの改訂（内容拡充、ガイド性向上）
- 簡易的な品質評価ツール（チェックリスト等）の試作と評価
- データ品質マネジメントガイドブック、簡易ツールに基づく適用検証の実施

中期的な取組み (令和8年度～9年度)

- 複数分野での横断的導入実証による活用モデルの確立
- マルチモーダルデータやマルチエージェント等の新たな利用形態への対応
- 評価ツール機能の本格化

長期的な取組み (将来的なビジョン)

- 組織内での継続的なデータ品質評価の運用モデル確立
- 技術進歩と社会変化に応じてガイドと評価ツールを更新する仕組みの構築
- 日本発のデータ品質評価フレームワークや指標定義の国際的な議論へ反映

AISI

Japan AI Safety Institute