

【第3回データ未来会議】 レーンA つなげる基礎力 ～データ連携の技術エッセンス～
なぜデータ品質が“つながり”を左右するのか

独立行政法人情報処理推進機構 デジタル基盤センター
デジタルエンジニアリング部 データスペースグループ
寺山 由希子

データ未来会議 午後の部レーンA ラインナップのご紹介 つなげる基礎力 ～データ連携の技術エッセンス～

1. 「つなぐ」ための基礎力（信頼できる／意味が通じる）

①データ品質

なぜデータ品質が“つながり”を左右するのか

②データ辞書

セマンティックレイヤーをつなぐ（データ辞書）

本セッション

2. つないだ後に「価値を回し続ける」ための基礎力（変化に追随する）

③Software-Defined

変化に追随するデータ連携へ：Software-Defined Society (SDS) 入門



IPA デジタル基盤センター デジタルエンジニアリング部
データスペースグループ 主任

寺山 由希子
Terayama Yukiko

出版、教育業界等での職歴を経て、

2019年8月 株式会社パスコ システム事業部 着任

- GIS（地理情報システム）の官公庁、民間企業への導入支援・活用推進に従事。

2023年4月 総務省 国際戦略局へ出向

- 国際電気通信連合 電気通信標準化部門（ITU-T）
SG16（現SG21・マルチメディア）、SG20（IoTとスマートシティ）、TSAG（諮問委員会）
での標準化活動に2年間従事。

2025年12月 IPA デジタル基盤センター デジタルエンジニアリング部 着任

- AI Safety Institute 事業実証WG データ品質SWGに参画

もともとは「データを活用したサービスをお客様に提供する」側の立場でした。

- I. データのつながりと品質
- II. AI時代に高まる品質の重要性
- III. 検証から見るデータ品質
- IV. おわりに

I . データのつながりと品質

データとは？

- ◆ データとは、主体同士をつなぐ媒介

例えば…

部署

×

部署

ソフトウェア

×

ソフトウェア

組織

×

組織

ひと

×

AI

経営

×

現場

過去

×

未来

データがつながるとは？

- ◆ データがつながるとは、同じ意味・由来で接続・連携すること

相手が意味を理解できるか

ほかのデータと組み合わせられるか

信頼して使えるか

AIも同じ意味で理解できるのか

● Mars Climate Orbiterをご存知ですか？

1998年12月11日に火星の気候観測と通信基地局機能を目的に、NASAより打ち上げられた探査機。
しかし…火星軌道投入時に大気圏へ低く入りすぎて消失。

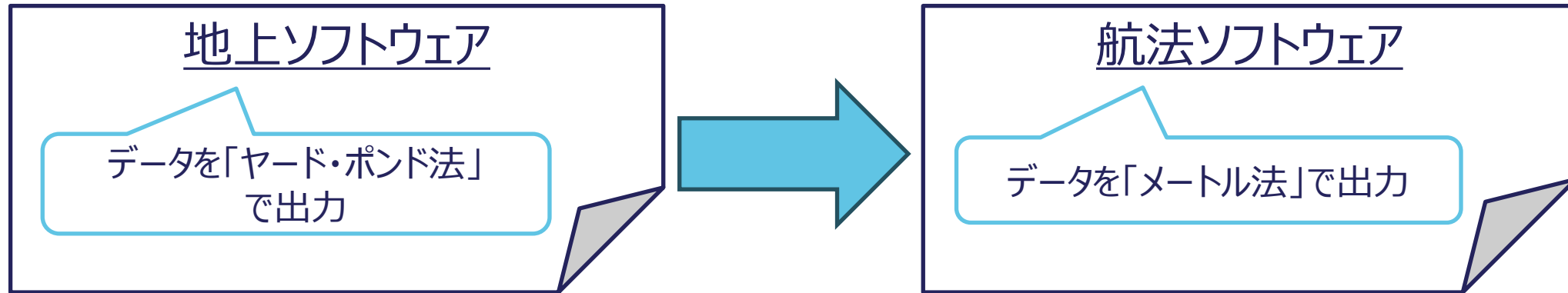


なぜ墜落してしまったのか？

- ◆ 原因は「技術的な故障」でも「宇宙での不慮の事故」でもなく
地上ソフトウェア と 航法ソフトウェア のデータ互換性に問題があったから

なぜ事故は起きた？

伝わっていた数字は全く同じ「32.6」。でも…



- 数字は同じでも単位が異なると、実際の長さは異なってくる
(1ヤード=0.9144メートル)
- 値は正しかったが、単位に一貫性がなかったことで、
誤った軌道に入って消失してしまった

正しいデータ（=品質を担保されたデータ）同士がつながって初めて「データがつながる」

もっと身近なところで…

エクセルでデータを扱う時にこんな経験はありませんか？

情報はある、でも…

**そのままは使えない
データを作りかえないといけない**

途方もない時間をかけてデータクレンジング



ニーズは何だろう？ 相互運用性はある？を確認することも「品質」

データがつながるための品質とは？

ISO 9000の定義する「品質」

- ◆ 対象の特性が要求事項を満たす程度のこと
- ◆ データ品質個別の定義は直接置かれないが、データも品質管理の対象となる“対象 (object) ”の一つとして扱われる

ISO 8000の定義する「データ品質」

- ◆ データが本来備えている特性がデータ仕様を満たす程度のこと

DAMA-DMBOKの定義する「データ品質」

- ◆ データが利用目的に対して適している程度 (**Fitness for Use**)
※代表的な特性として定義されているもの

正確性

完全性

一貫性

適時性

妥当性

データ品質とは？

データ品質とは「正しいデータ」ではなく「目的に適したデータ」である

ステークホルダーのニーズを満たして初めて「データ品質」

データ品質はデータの品質特性で評価される

Ⅱ．AI時代に高まる品質の重要性

おさらい：データのつながり

データとは？

- ◆ データとは、主体同士をつなぐ媒介

例えば…

部署

×

部署

組織

×

組織

経営

×

現場

ソフトウェア

×

ソフトウェア

ひと

×

AI

過去

×

未来

データがつながるとは？

- ◆ データがつながるとは、同じ意味・由来で接続・連携すること

相手が意味を理解できるか

信頼して使えるか

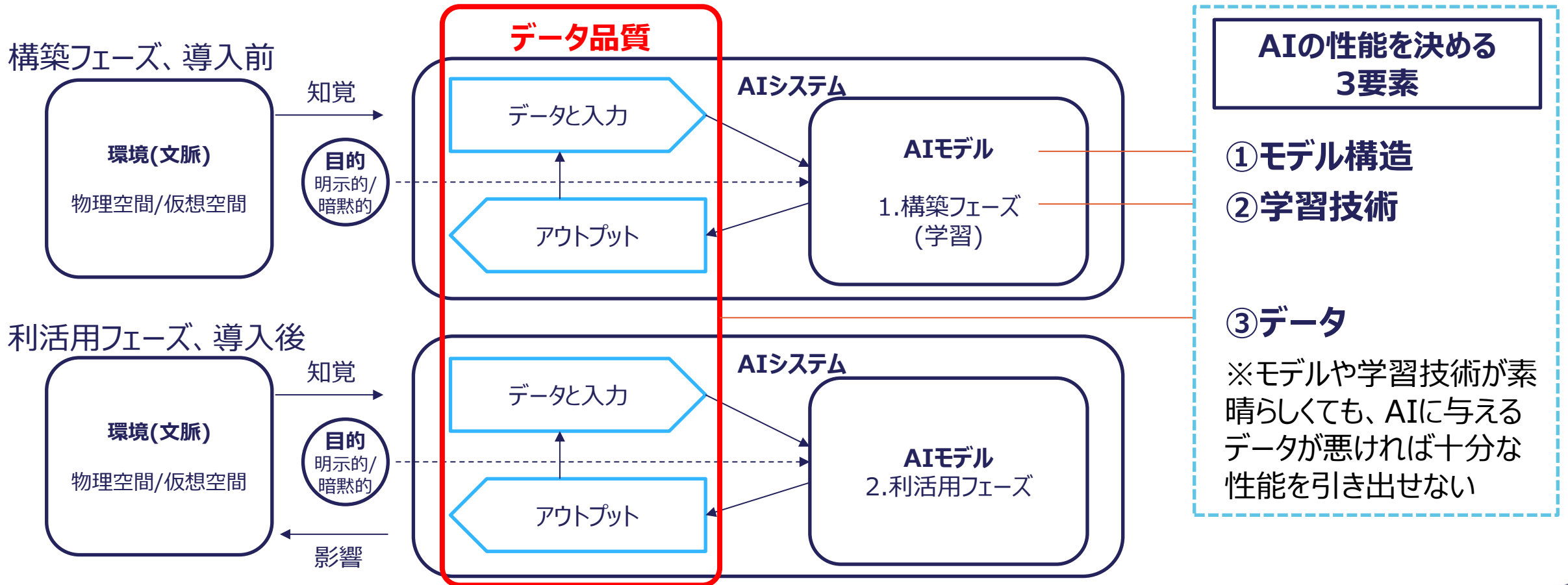
ほかのデータと組み合わせられるか

AIも同じ意味で理解できるのか

AIは過去のデータを学習し、その品質が未来に影響します

AIとデータ活用の現在地

データはAIや自動化の基盤



“Garbage in, Garbage out”と言われるように、データ品質はAIの要である

AI Safetyで見るデータ品質

データ品質はAISIが公表するAIセーフティ評価の観点の1つ

AISI事業実証WGの
ひとつとして
データ品質SWGが
設置されています

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。

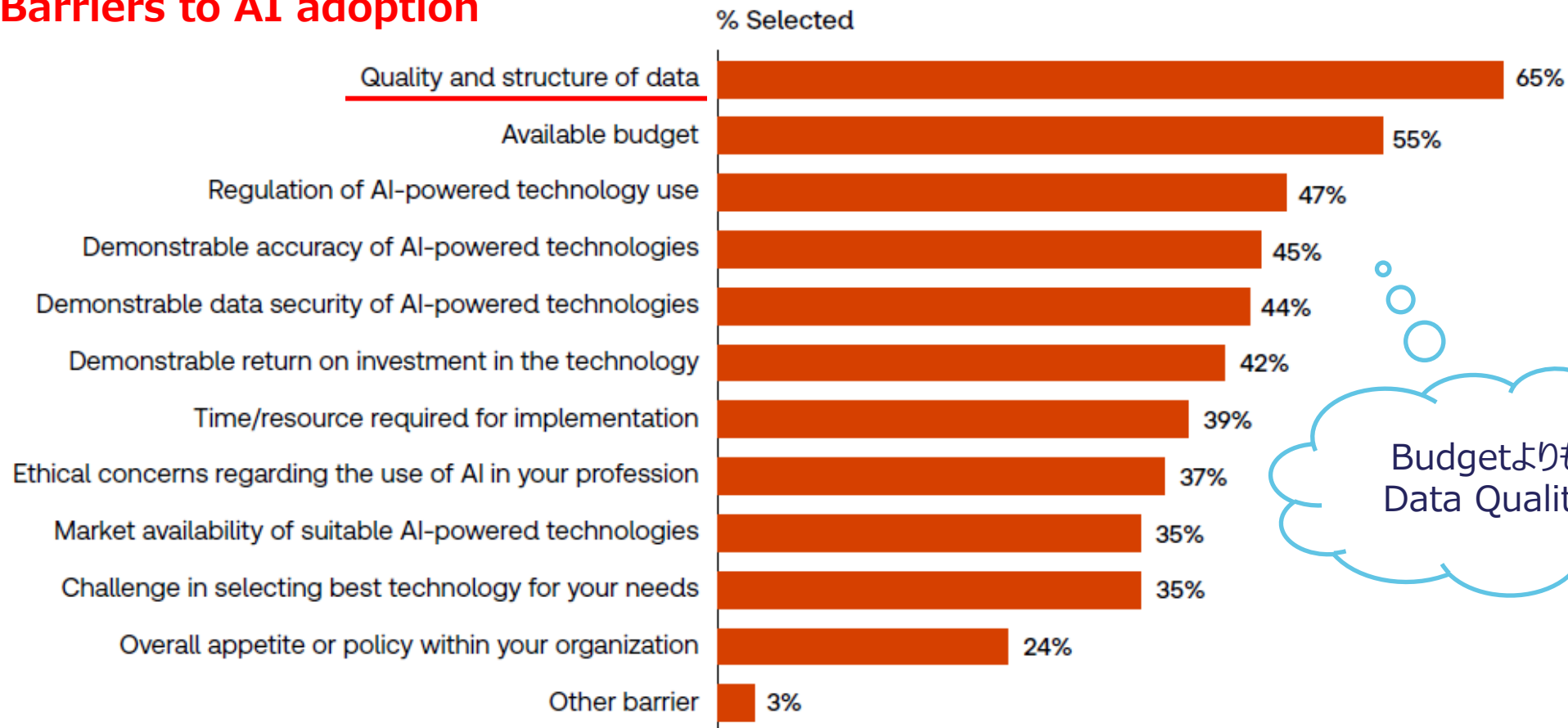
■想定され得るリスクの例

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。

データ品質がAI導入の障壁になっている？

世界のCxOがAI導入・拡大で感じる障壁

Barriers to AI adoption



(出典) [Thomson Reuters Institute 2025 C-Suite Survey](#)

データ品質マネジメントガイドブックの公表

- ◆ AISIとIPAは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
- ◆ データ品質をライフサイクル、ゲートウェイ、ガバナンスの3つの視点から多角的に評価



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)

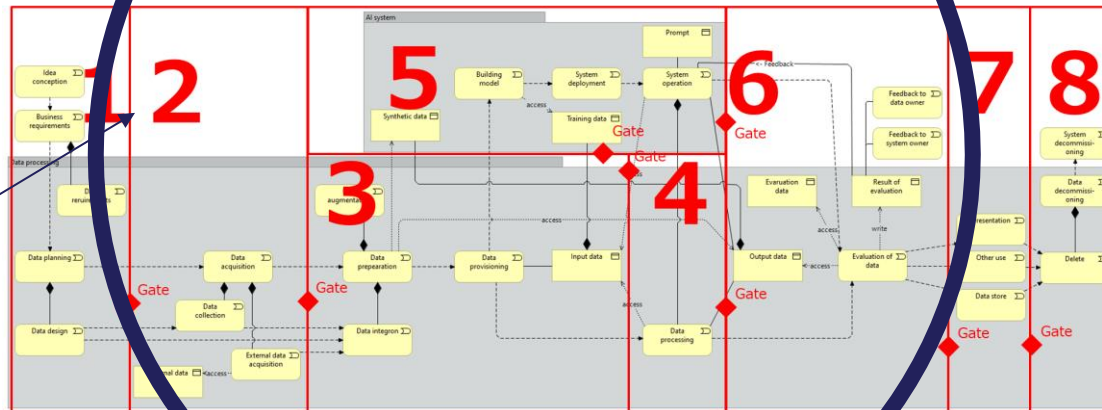
評価指標を定義する

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

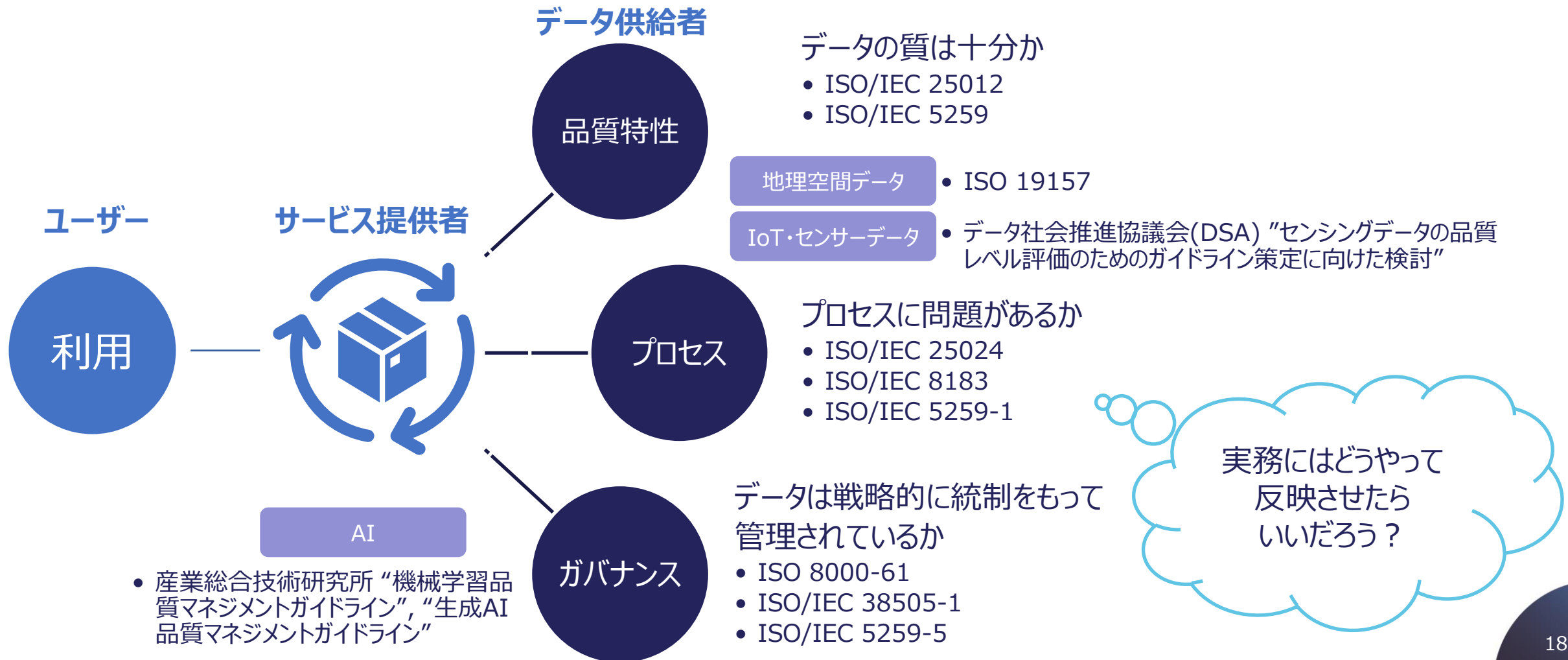
継続的で持続可能な
組織活動を確保する



Data lifecycle
This lifecycle refers to the following processes:
ISO/IEC 25929 Data quality for analytics and machine learning
ISO/IEC 25234 Systems and software engineering - AI - Measurement of data quality
ISO/IEC 31011 AI Data life cycle framework

ガイドブックと国際標準との相互運用

国際標準やガイドラインに準拠した相互運用可能な品質管理モデル



国際標準に基づいたデータ品質特性

ISO/IECの標準では以下のような品質特性があげられています

Accuracy (正確性)	Compliance (コンプライアンス)	Availability (可用性)	Effectiveness (有効性)
Completeness (完全性)	Confidentiality (機密性)	Portability (ポータビリティ)	Identifiability (識別可能性)
Consistency (一貫性)	Efficiency (効率性)	Recoverability (復旧性)	Relevance (関連性)
Credibility (信頼性)	Precision (精度)	Auditability (説明可能性)	Representativeness (代表制)
Currentness (最新性)	Traceability (トレーサビリティ)	Balance (バランス)	Similarity (類似性)
Accessibility (アクセシビリティ)	Understandability (分かりやすさ)	Diversity (多様性)	Timeliness (即時性)

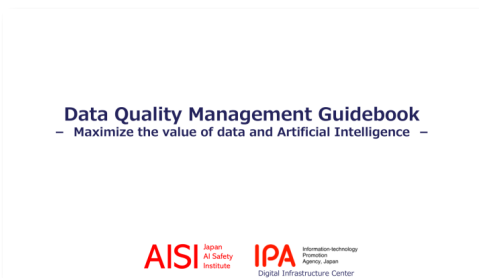
データが
あてはまるかどうか
どう確認すれば
いいだろう？

“Fitness for Use”なのですべて当てはまる必要はなく、
多数の特性の中から、必要な品質特性を活用

データ品質チェックリストのご紹介

- ◆ データ品質マネジメントガイドブックを実務で活用する方法の一つとして、データ品質チェックリストを作成（初版を2026年4月公表、随時更新予定）

ガイドブック

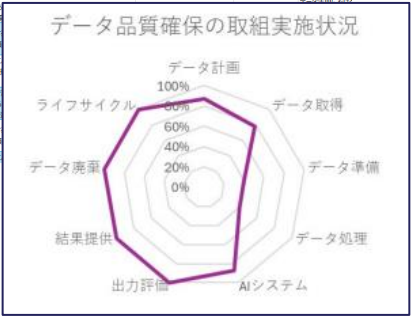


- スプレッドシートベース
- データ品質管理の勘所を押さえた70問超の推奨する確認事項
- 取組状況はライフサイクル別にレーダーチャートで参考として可視化

準拠必須ではなく、データ品質観点でのセルフチェック活用を想定しています

チェックリスト

プロセス	項目 (参考日本語訳)	チェック	主なリスク種類	リスク具体例	品質特性	実施者	実施例
データ対策	データを使用するシステムのセキュリティ品質は明確に定義されているか。(パスワード、コストなど)	業務効率の悪化	品質低下	品質低下データのクレンジングと検証に時間とリソースを割けなければならない。AIモデルの信頼性が落ち、運用効果が低下する。	信頼性	AIシステム開発者	システム全体のセキュリティで対応
データ対策	システムに必要なデータはストリップされているか。	業務効率の悪化	品質低下	品質低下データのクレンジングと検証に時間とリソースを割けなければならない。AIモデルの信頼性が落ち、運用効果が低下する。	完全性	AIシステム開発者	システムの目的を考慮し、チーム内
データ対策	各データについて品質管理項目が定義されているか。	業務効率の悪化	品質低下	品質低下データのクレンジングと検証に時間とリソースを割けなければならない。AIモデルの信頼性が落ち、運用効果が低下する。	信頼性	AIシステム開発者	システムに適用する完全性、正確性、可
データ対策	各データのデータ品質要求レベルは定義されているか。	意思決定の誤り	品質低下	不正確なデータや不完全なデータは、分析に欠陥をもたらす。その結果、組織の意思決定に悪影響を及ぼす。また、法的リスク、信頼性の低下、不適切な意思決定を招く可能性がある。	信頼性	AIシステム開発者	上記で設定の上昇
データ対策	データアーキテクチャと設計方針が文書化されているか。	業務効率の悪化	品質低下	品質低下データのクレンジングと検証に時間とリソースを割けなければならない。AIモデルの信頼性が落ち、運用効果が低下する。	完全性	データ管理者	データモデルやデー
データ対策	必要なデータは定義され、利用可能なか。	競争上の不利	品質低下	より確かなデータを持つ競争他社は、顧客獲得力、業務効率、市場対抗力などの主要分野で優位な業績を上げ、組織は遅れを取る。	可用性	AIシステム開発者	システムに必要な状態であるから
データ対策	データの活用に関する法的規制をアウツしたか。	法的規制リスク	品質低下	データ保護に関する法律や規制 (GDPR, CCPA など) を遵守しない場合、多額の罰金や訴訟につながる可能性がある。	コンプライアンス	データ管理者	リストアップした用途の制約や条件
データ対策	データプライバシーや機密性が高いものは含まれていないか。	法的規制リスク	品質低下	データ保護に関する法律や規制 (GDPR, CCPA など) を遵守しない場合、多額の罰金や訴訟につながる可能性がある。	機密性	データ管理者	社内データの場合、社外データ時や機密性が高い
データ対策	データの分類法を参照して管理しているか。	コスト増	品質低下	品質低下データは、エラー修正、再処理、AIモデルの再トレーニング、誤った予測や意思決定による潜在的な金銭的損失によるコスト増を招く。	パフォーマンス	データ管理者	RAM, I/O, CPU の管理を実施して
データ対策	データが定期的に更新されているか。	コスト増	品質低下	品質低下データは、エラー修正、再処理、AIモデルの再トレーニング、誤った予測や意思決定による潜在的な金銭的損失によるコスト増を招く。	アクセシビリティ	データ管理者	WCAGで標準化されたデータを管理す
データ対策	異なる一般的な活用を参照しているか。	意思決定の誤り	品質低下	不正確なデータや不完全なデータは、分析に欠陥をもたらす。その結果、組織の意思決定に悪影響を及ぼす。また、法的リスク、信頼性の低下、不適切な意思決定を招く可能性がある。	完全性	データ管理者	データ品質管理
データ対策	データが定期的に更新されているか。	法的規制リスク	品質低下	データ保護に関する法律や規制 (GDPR, CCPA など) を遵守しない場合、多額の罰金や訴訟につながる可能性がある。	完全性	データ管理者	データ品質管理
データ対策	データが定期的に更新されているか。	品質低下	品質低下	品質低下データのクレンジングと検証に時間とリソースを割けなければならない。AIモデルの信頼性が落ち、運用効果が低下する。	完全性	データ管理者	データ品質管理
データ対策	データが定期的に更新されているか。	品質低下	品質低下	品質低下データのクレンジングと検証に時間とリソースを割けなければならない。AIモデルの信頼性が落ち、運用効果が低下する。	完全性	データ管理者	データ品質管理

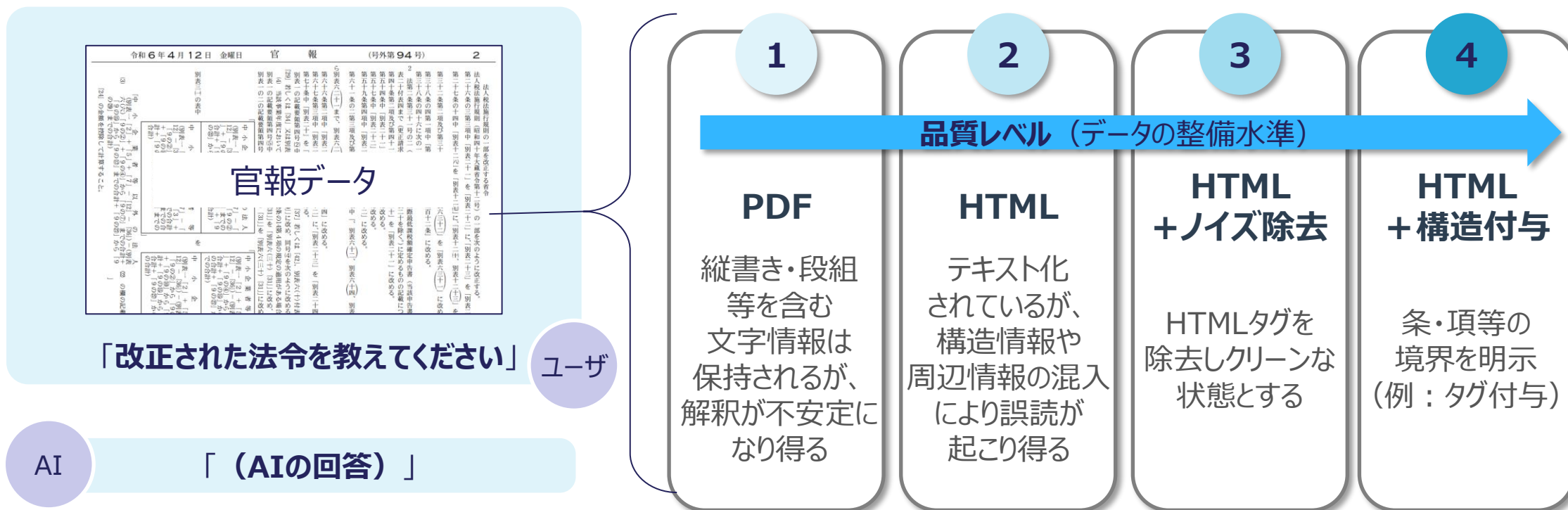


No.
ライフサイクル
チェック項目
優先事項
チェック
エビデンス
実施例
リスク例
主たる品質特性
実施者
対象データ
EUAI法該当箇所
AISI評価ツール該当箇所
備考

Ⅲ. 検証から見るデータ品質

データによる品質の違いを見てみましょう…

- ◆ 官報データを読み込ませ、品質の違いが生成AIの応答と安全性に与える影響を検証してみました



※本件はAI Safety Institute データ品質サブワーキンググループにおける試行的な検証として進めているものです。

官報データ

No	名前	官報公布	種別	改正方式
1	法律第八十一号	2025/12/05	法律	改め文方式
2	法律第八十二号	2025/12/10	法律	改め文方式
3	法律第八十八号	2025/12/22	法律	改め文方式
4	政令第三百六十九号	2025/12/24	法律	改め文方式
5	政令第三百八十八号	2025/11/12	政令	改め文方式
6	政令第三百九十五号	2025/11/27	政令	改め文方式
7	政令第四百十二号	2025/11/28	政令	改め文方式
8	政令第四百二十三号	2025/12/12	政令	改め文方式
9	政令第四百四十六号	2025/12/17	政令	改め文方式
10	環境省令第二十二号	2025/12/26	政令	改め文方式
11	国土交通省令第百十四号	2025/11/12	府省令	改め文方式
12	内閣府令第百号	2025/11/27	府省令	新旧対照表方式
13	総務省令第百八号	2025/11/28	府省令	新旧対照表方式
14	総務省令第百十三号	2025/12/10	府省令	新旧対照表方式
15	国家公安委員会規則第十八号	2025/12/23	府省令	新旧対照表方式
16	公正取引委員会規則第十二号	2025/12/26	規則	新旧対照表方式

改め文方式

法律 第八十一号
 租税特別措置法及び東日本大震災の被災者等に係る国税関係法律の臨時特例に関する法律の一部を改正する法律（租税特別措置法の一部改正）
第一条 租税特別措置法（昭和三十二年法律第二十六号）の一部を次のように改正する。
 第八十八条の八を削る。
 第八十九条を次のように改める。

改め文方式は「第1条中『A』を『B』に改める」のように文章で修正箇所を指示する伝統的な形式で、**正確性が高い一方、理解しにくいことが特徴**

新旧対照表方式

改正後	改正前
目次	目次
第一章～第一章の四 [略]	第一章～第一章の四 [同上]
第二章 福祉の保障（第七条～ 第三十六条の三十の六の十四 ）	第二章 福祉の保障（第七条～ 第三十六条の三十の六の三 ）
第三章～第四章 [略]	第三章～第四章 [同上]
附則	附則
第七条の四十五 法第十九条の二十二第四項に規定する厚生労働省令で定める方法は、個人番号カード（行政手続における特定の個人を識別するための番号の利用等に関する法律第二条第七項に規定する個人番号カードをいう。 第十七条の三第二項第一号及び第三十六条の三十の六の六第二項第一号において同じ。 ）を提示する方法とする。ただし、当該方法によることができない状況があるときは、書面により提示する方法とする。	第七条の四十五 法第十九条の二十二第四項に規定する厚生労働省令で定める方法は、個人番号カード（行政手続における特定の個人を識別するための番号の利用等に関する法律第二条第七項に規定する個人番号カードをいう。 第十七条の三第二項第二号 において同じ。）を提示する方法とする。ただし、当該方法によることができない状況にあるときは、書面により提示する方法とする。

修正前後の条文を左右に並べ、変更箇所を網掛け等で強調する形式で、**視覚的に分かりやすいことが特徴**

検証結果は…

- ◆ 「新旧対照表方式」で精度を見ると、構造化レベルによる精度の違いが極めて小さいが、「改め文方式」では、構造化レベルを上げると、精度が改善する傾向にある。

構造化をすることでハルシネーションを防ぐことが期待できる！

値が0~1の範囲で、1に近いほど良い

レベル	修正方式	修正条数	TP	FP	FN	適合率 Precision	再現率 Recal	F-score
1: PDF	改め文方式	277	105	37	172	0.739	0.379	0.501
2: HTML	改め文方式	277	175	43	102	0.803	0.632	0.707
3: HTML_tagremoved	改め文方式	277	141	34	136	0.806	0.509	0.624
4: Markdown_improved	改め文方式	277	168	38	109	0.816	0.606	0.696
1: PDF	新旧対照表方式	56	46	1	10	0.979	0.821	0.893
2: HTML	新旧対照表方式	56	46	4	10	0.920	0.821	0.868
3: HTML_tagremoved	新旧対照表方式	56	46	3	10	0.939	0.821	0.876
4: Markdown_improved	新旧対照表方式	56	46	5	10	0.902	0.821	0.860

つながるためのデータ品質マネジメントとは？

データの由来の明確化：データの出自を明らかにする（ex.メタデータ）

データの意味の明確化：ゆらぎがないよう、データの定義を管理

データの定期更新：常にデータを最新化することを意識

データの構造化の検討：共有相手が読み込みやすいデータを意識

IV. おわりに

AI時代に高まる日本の品質への期待

- ◆ 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

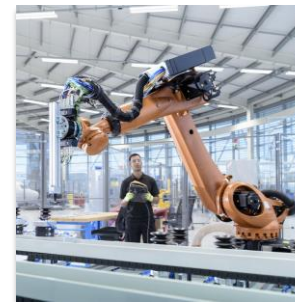
- 日本の職人技は常に高品質な製品を追求してきた。この伝統は現在のビジネス慣行にも受け継がれている。

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ており、また、日本の競争力と経済を成長させることが期待されています

最後にもう一度：データのつながり

データとは？

- ◆ データとは、主体同士をつなぐ媒介

例えば…

部署

×

部署

組織

×

組織

経営

×

現場

ソフトウェア

×

ソフトウェア

ひと

×

AI

過去

×

未来

データがつながるとは？

- ◆ データがつながるとは、同じ意味・由来で接続・連携すること

相手が意味を理解できるか

信頼して使えるか

ほかのデータと組み合わせられるか

AIも同じ意味で理解できるのか

AIは過去のデータを学習し、その品質が未来に影響します

- ◆ 今後、AIが出力した過去の生成物を、未来のAIが学習する時代がやってきます
- ◆ その時に共有される「データ」は、今まさに目の前にあるデータが基礎となります
- ◆ 今、データ品質に対して取り組むことが未来の信頼できるデータとのつながりを作ります

データ品質は時空間をこえて様々な“つながり”を左右します
ぜひ、今日からデータ品質への取り組みをはじめませんか？

IPA

ご清聴いただきありがとうございました