

産業データ・AI政策について

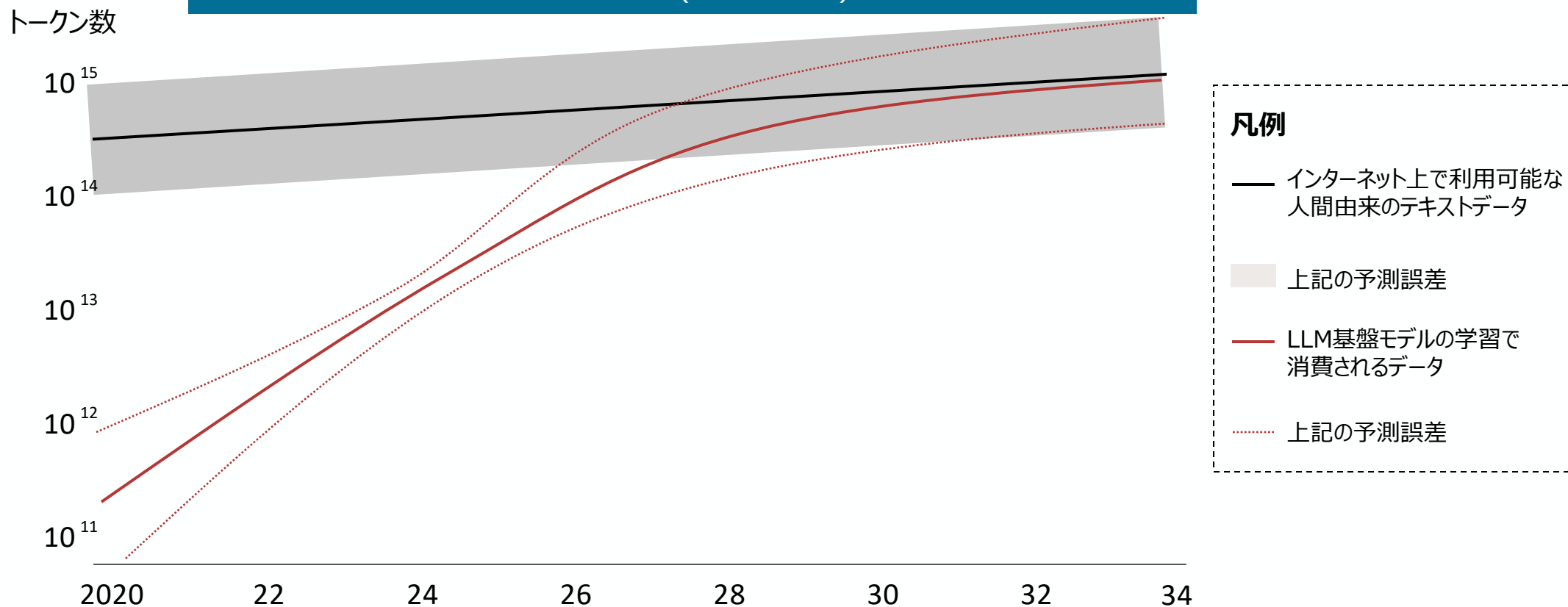
2026年3月

商務情報政策局 情報経済課

AIの学習データ枯渇問題

- 経済・産業活動のデジタル化が進展し、データそのものの価値やデータ利活用のニーズが高まる中、生成AI等の登場がこの動きを更に加速化。あらゆる産業の競争力がデータによって規定される時代に入っている。
- 一方で、これまでインターネット上の大量のテキストデータを学習し、あらゆる場面で活用されつつある生成AIも、昨今では目前に迫っている「学習データの枯渇」が大きな問題になっている。

LLM基盤モデル学習におけるインターネット上のテキストデータ利用の予測
2022-34年(予測値ベース)



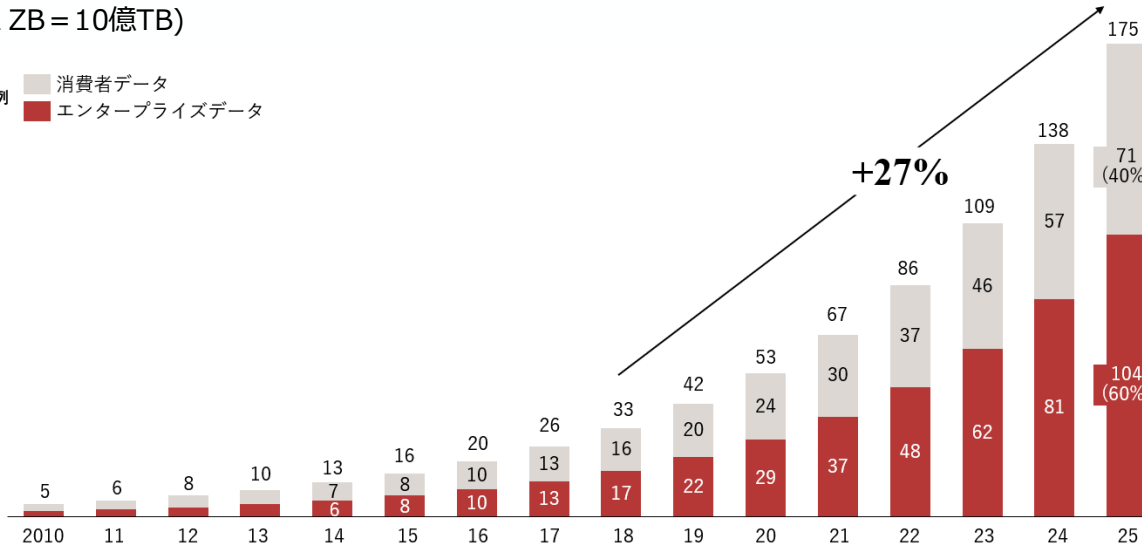
今後は企業内データの利活用が鍵

- 今後は、全世界で流通するデータの**6割を占める企業内データ（≒エンタープライズデータ）**の利活用が**産業戦略上の焦点**となり、**生成AIの高度化・マルチモーダル化のためにも重要**。
- 特に、エンタープライズデータの**2割以上を占める製造分野**は、**製造業に強みを有する我が国にとってデータ活用のポテンシャルが非常に高い**。

年間のデータ量（世界中で創出・取得・複製・消費されるデータ量）の推移 2010-25年（予測値ベース）

ゼタバイト
(1 ZB = 10億TB)

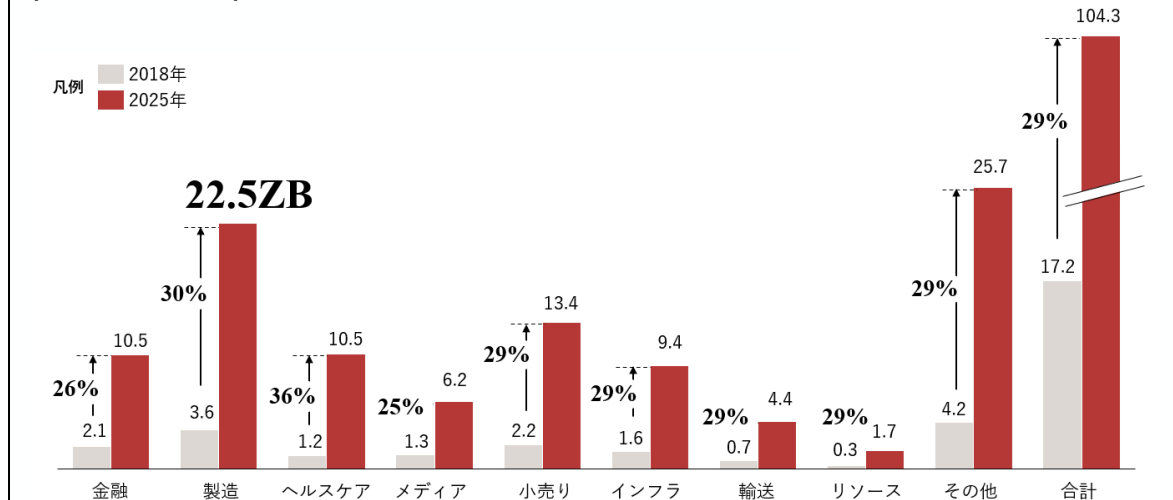
凡例
■ 消費者データ
■ エンタープライズデータ



エンタープライズデータ量の比較 2018年 対 2025年（予測値ベース）

ゼタバイト
(1 ZB = 10億TB)

凡例
■ 2018年
■ 2025年



(出典) The Digitization of the World From Edge to Core - IDC

(注) 左図：IDCレポート内では、2018年に32ZB、2025年に175ZBのみ定量推測データが公開。上記グラフではCAGRを算出後、同一ベースで成長すると仮定し、2010年以降のデータを算出。

右図：2025年のデータについて、全体・金融・製造・ヘルスケア・メディアはレポート記載のCAGRをもとに算出し、その他産業は左記CAGRの平均値から算出。

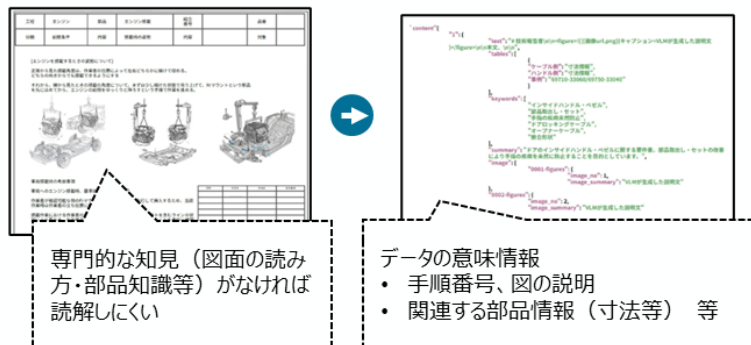
データ基盤を巡る技術動向

- AI時代の到来に対し、**AI学習・利用やデータ連携等が容易な形式にデータを精製する技術や、分散管理されたデータ資源の連携を信頼ある形で、利用者目線で柔軟に、スケーラブルに実現するためのアーキテクチャ（データスペース）**に関する検討が国際的に進展。
- こうした技術・アーキテクチャをサービスとして実現し、産業界においてAI時代の価値の源泉となるデータ資源の開発やデータ流通の実現を推進していくことが、産業競争力強化においても重要に。

データ精製（AI-Ready化等）

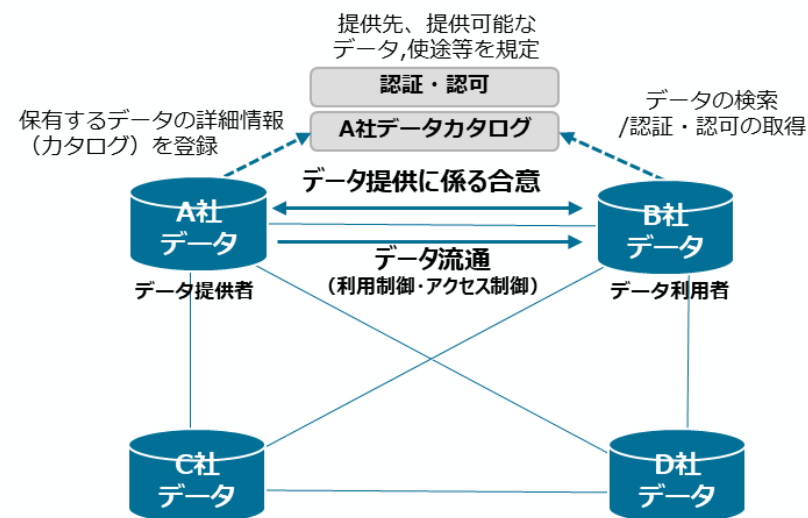
- データを意味・関係性が整理された、AIが理解しやすい形式に加工・整理するプロセス。

例：手順書



データ連携（データスペース※）

- 管理者・管理ポリシーが異なるデータ資源を利用制御によって分散型で安全に連携利用できるようにするための仕組み。



※欧州を中心にデータスペースのコンセプトが提案される一方、米国においてはデータメッシュという類似のコンセプトも提案されている。

ロボティクスにおけるAI（フィジカルAI）の発展

- 従来のロボティクス（単一作業特化型ロボット）は、ティーチングプレイバック（人間がロボットに直接動作を覚えさせる制御手法）が主流。各動作にティーチングが必要なため、導入コストが高く、環境変化への柔軟性がない。
- 近年は、**VLA（Visual Language Action Model）、模倣学習、強化学習等の活用**を通じて、**実データとシミュレーションデータも含めた大量のデータをAIに学習させ、自律性や汎用性を高めたAIロボティクスの開発が加速**。

従来の制御（ティーチングプレイバック）

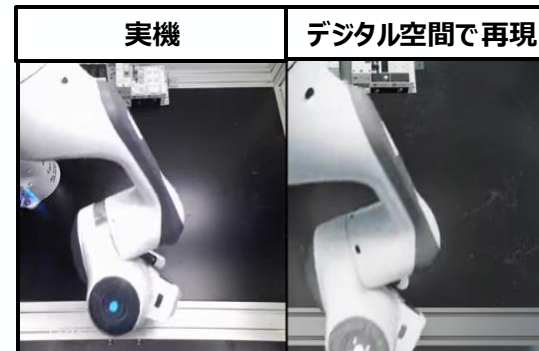
- 従来のティーチングプレイバックは、定型作業に強く動作の精度は高い。
- 一方、**教示時間が数週間や場合によっては数か月/ 1ライン**かかり、**初期導入コストが高い**。
- また、**段取り替えが発生する場合、都度教示が必要**で、環境変化への柔軟性がない。



近年のAIを活用した学習・制御（VLA・VLM）

- **ロボット基盤モデル（VLA/VLM）の進化**で、従来の**教示時間が大幅に短縮**。（数週間から数分間に短縮したとの見方も存在。）
- **新たな部品・生産ラインに柔軟かつ高速に対応が可能**に。

センサーデータを基に物理空間上の実体をデジタルツインで再現

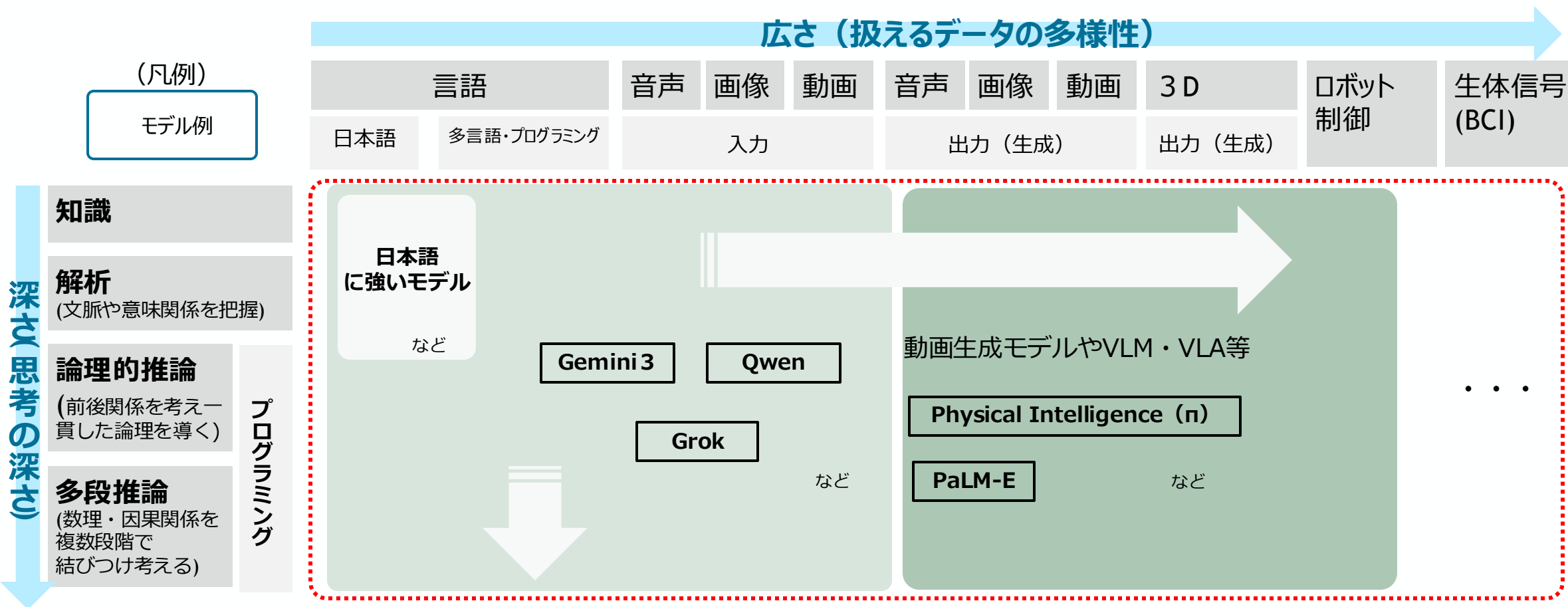


デジタルツイン上で生成された3Dモデリング（フォークリフト）



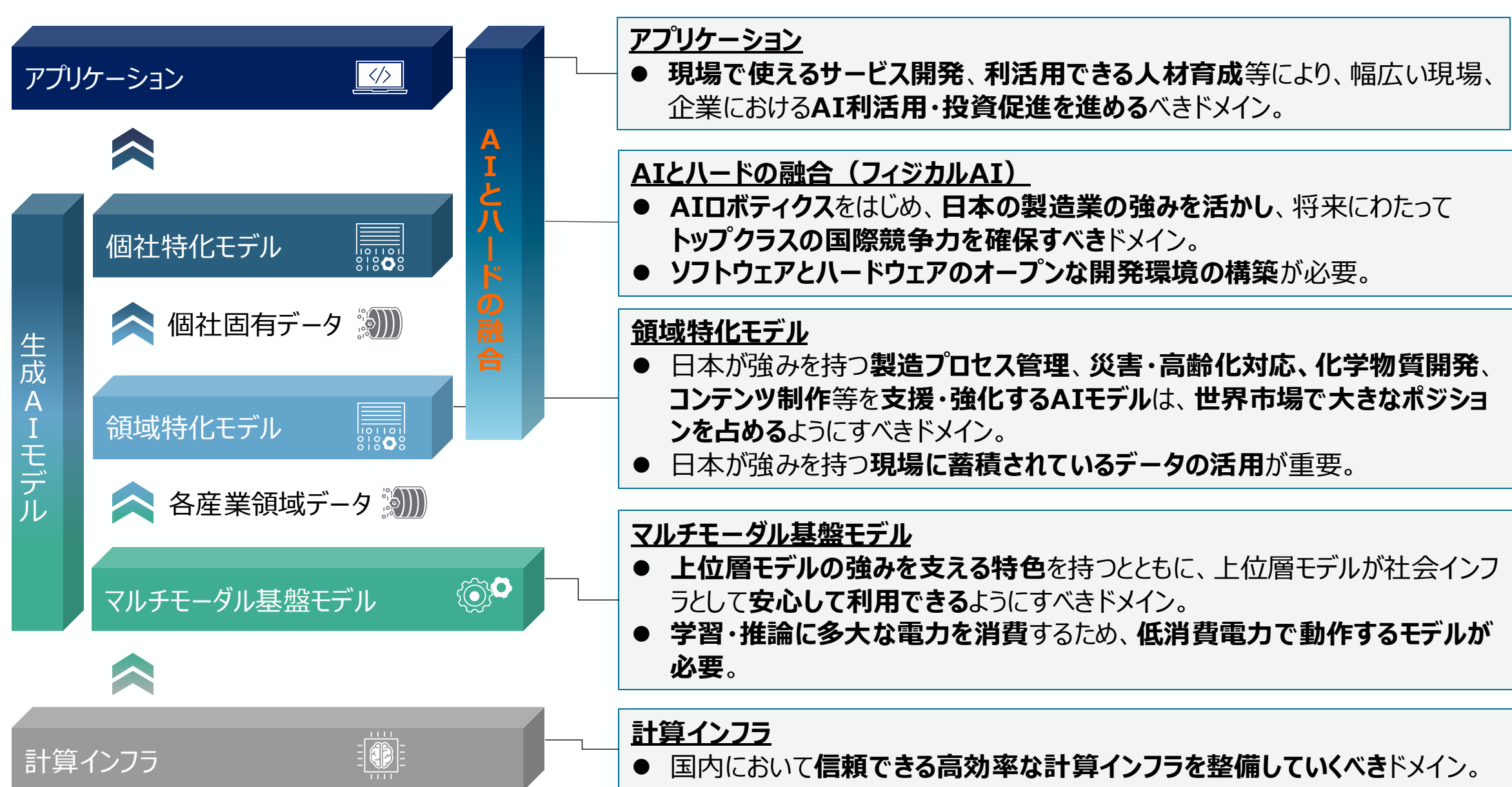
AIのマルチモーダル化を巡る動向

- AIは、言語のみならず、**画像・音声等多様なデータを広く扱う**マルチモーダル化に向けた開発が激化。
- また、**文脈・意図・因果関係を踏まえて深く思考することも、開発の重要な要素**になっている。



知識の深さに相当するベンチマーク例。知識：JAQKET, NIILC (常識応答), JCommonsenseQA, pfgem-bench (日本語知識)、解析：llm-jp-eval (Wikipedia読解・要約推論)等、論理的推論：MMU-Pro, 等、多段推論：AIME 2025, Humanity's Last Exam等

AIサプライチェーンの各ドメインの考え方



フィジカルAIの重要性とマルチモーダル基盤モデルとの関係

- フィジカルAIは、工場の自律制御・最適化、ロボットの自律制御、自動運転などを実現するための基盤となるもの。
- フィジカルAIの開発では、言語に留まらず、多様なデータを扱うマルチモーダル基盤モデルが不可欠。



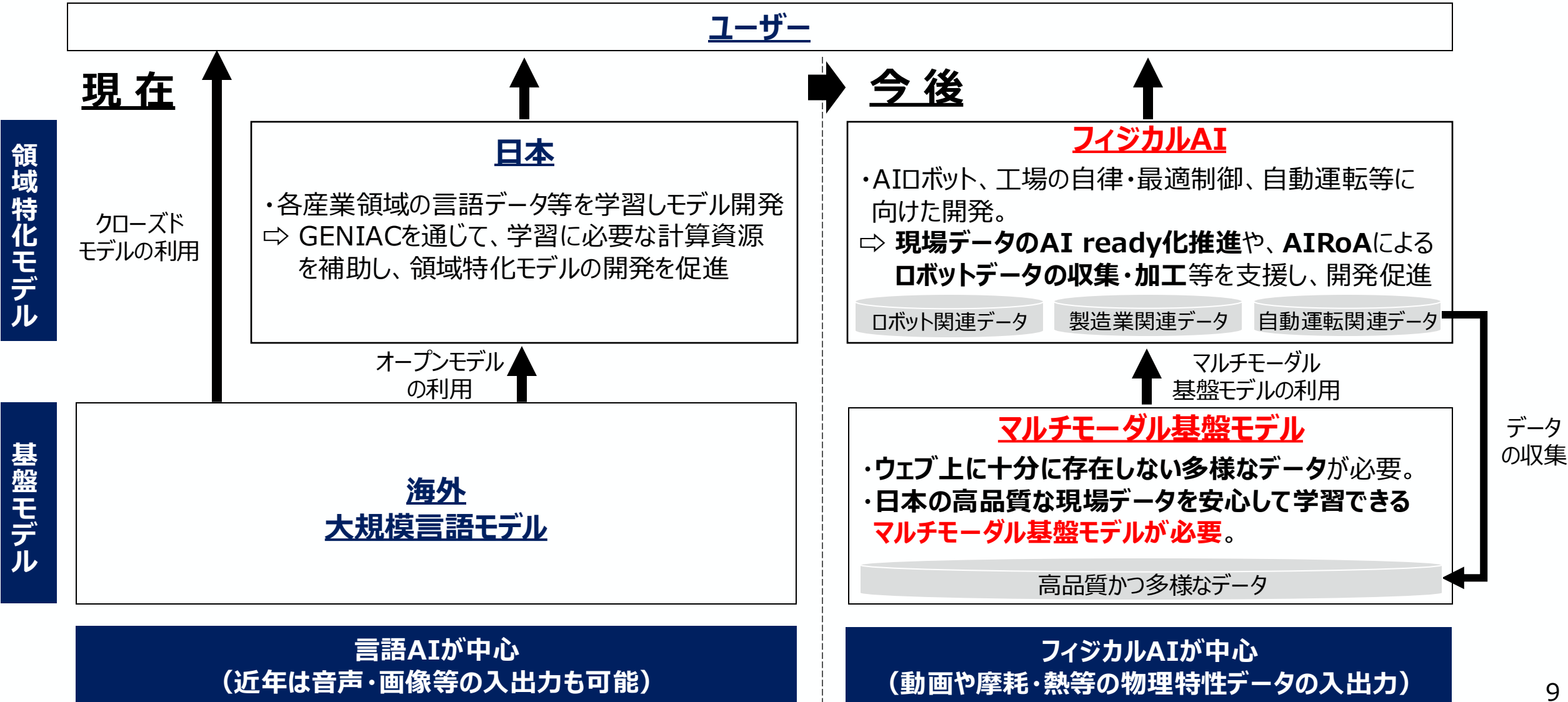
フィジカルAI

マルチモーダル基盤モデル

（言語だけでなく、音声・画像・動画・センサーデータなど多様なデータを扱うことが可能）

国産のマルチモーダル基盤モデルの開発

- まずは、日本企業で一般的に活用されるオープンモデルと同程度の基本性能のモデルを開発。それ以降はAIロボットや工場の自律・最適制御、自動運転等を念頭に、扱えるデータの多様性や思考の深さをステップ・バイ・ステップで獲得する方針。



AI-Ready化による精度の向上

- 現場の生データをAI-Ready化していくことで、精度が大幅に改善されることが、ビッグテックや学术界の論文からも 多数報告されている。
- 既に製造業内でもAI Readyなデータを整備し、AIモデル開発を目指す先進的な取組も存在。

ビッグテックや学术界の論文

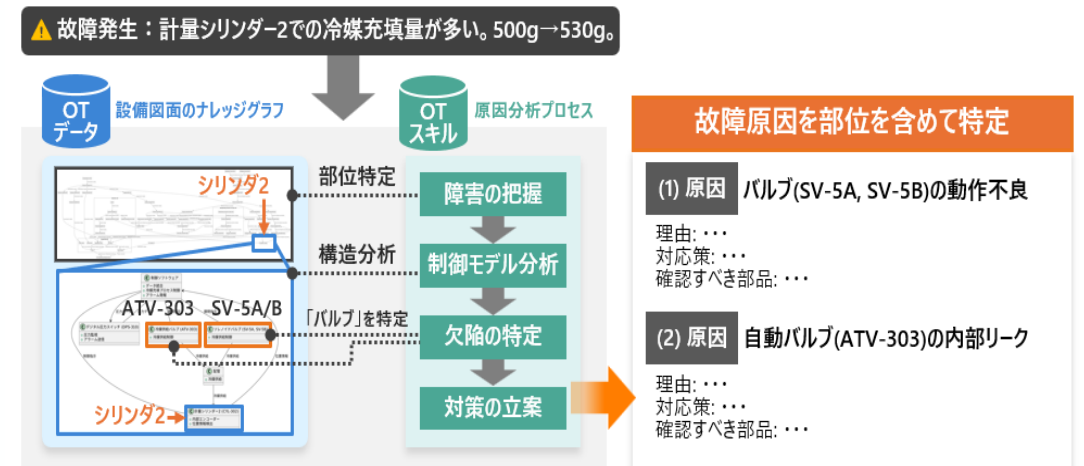
企業	タイトル	アプローチ概要	精度改善
OpenAI社	DevDay: "A Survey of Techniques for Maximizing LLM Performance" *1	RAGにおけるモデルへ与えるデータの改善をさまざま実施することで大幅な改善が見られた。例えば、チャンクサイズの最適化、リランキング、メタデータ付与、クエリ拡張、プロンプトエンジニアリング、データ取得のための外部ツールの活用など	53ポイント改善 (45% -> 98%)
Anthropic社	Contextual Retrieval in AI Systems *2	チャンクの前後文脈を追加するなどのコンテキストデータを最適化する (Contextual Retrieval) ことで、RAG性能の向上	取得失敗率改善 (5.7% → 1.9%)
カーネギー大などの共著	LumberChunker: Long-Form Narrative Document Segmentation *3	物語系長文 (GutenQA) で、動的に"話の切れ目"を見つけて分割するなどし、情報検索精度を大きく改善	7.37%改善

*1 : <https://www.youtube.com/watch?v=ahnGLM-RC1Y>
 *2 : <https://www.anthropic.com/engineering/contextual-retrieval>
 *3 : <https://arxiv.org/pdf/2406.17526>

製造業の事例

設備故障診断AIの開発

- 日立とダイキンが共同で、工場の設備故障診断AIエージェントの試験運用を開始
- 「OTデータ」と、設備故障原因分析プロセスである「OTスキル」を組み合わせたAIで、**ダイキンの一般的な保全技術者と同等以上の故障診断を実現**

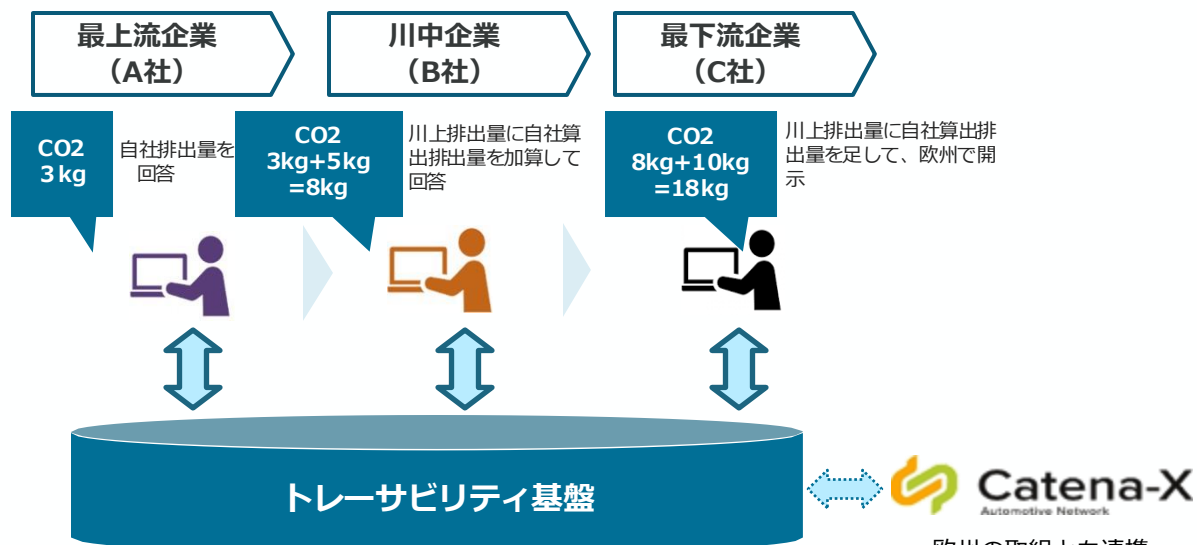


企業や業界を越えたデータ連携①

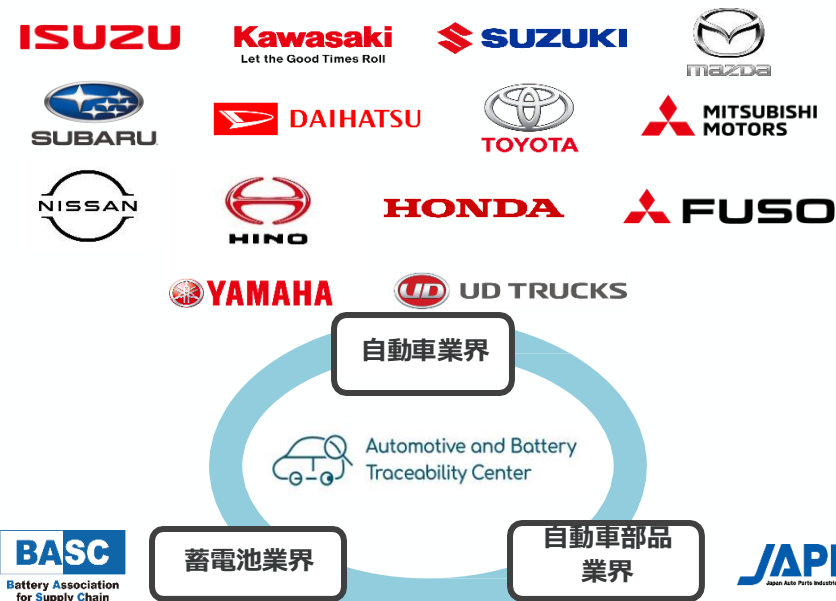
- データ連携を通じて新しい価値を生み出す企業間連携の取組を「ウラノス・エコシステム」と名付け、官民で連携し推進。
- まずは、具体的な取組として、CO2排出量の管理などを実現するための、自動車・蓄電池のデータ連携基盤を構築。この取組をモデルとして、今後、化学物質管理などの他分野での展開や、国際連携を推進

CO2排出量管理などのための、自動車・蓄電池のデータ連携基盤を構築

システム運営の担い手として、各業界団体が共同で「一般社団法人自動車・蓄電池トレーサビリティ推進センター(ABtC)」を設立

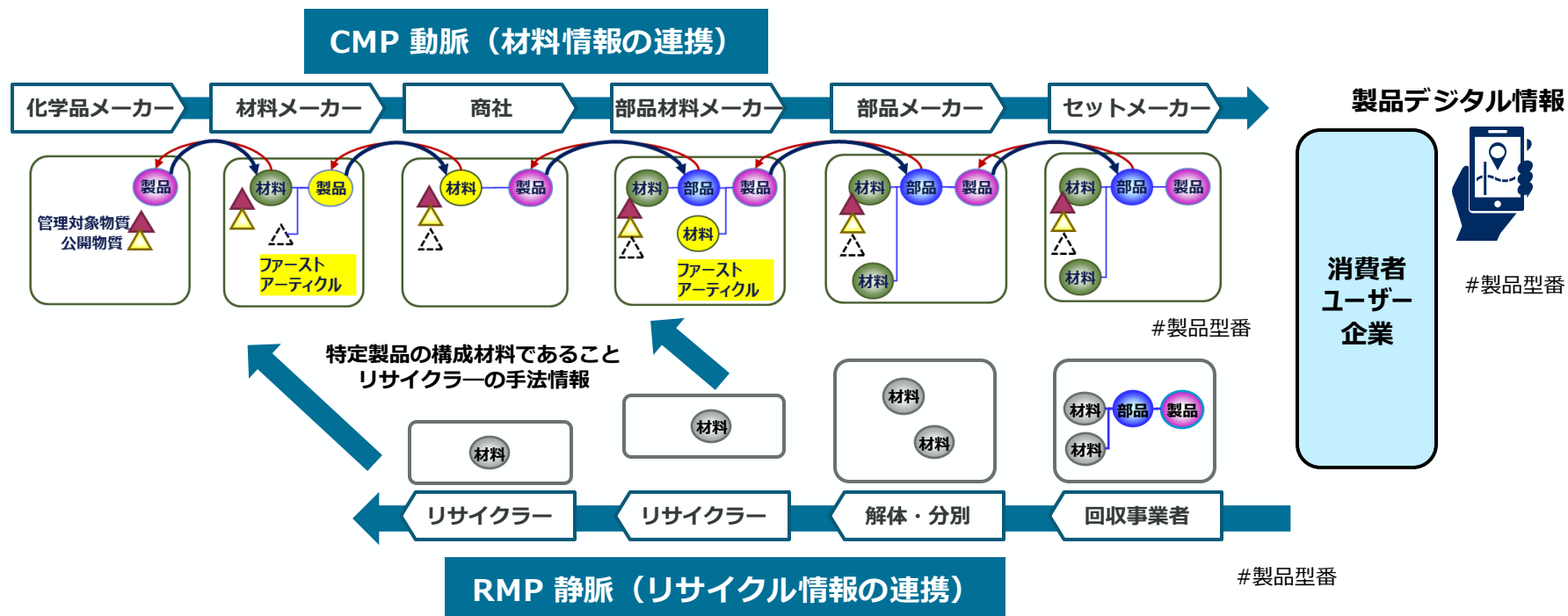


※企業秘密等の情報は厳重に保護



企業や業界を越えたデータ連携②

- 諸外国による規制や囲い込みの動きに適切に対応し、日本企業が諸外国の制度・PFに依存することなく我が国として対等な関係性を確保するための取組が必要。特に、化学物質管理等の資源循環分野での欧州規制等に対応することが急務。
- 今後は、新たに製品含有化学物質のトレサビシステム（CMP）構築し、国際相互運用性の確保等を目指していく。



- CMPタスクフォース | CMP構想の進捗について（次世代化学物質情報伝達システムについて）
https://www.meti.go.jp/shingikai/sankoshin/hoan_shohi/chemicals/pdf/002_08_00.pdf を基に経済産業省作成

データスペース技術に関する取組

- 信頼を伴ったデータ流通・活用環境としての「データスペース」技術に関する日本発の共通仕様を「Open Data Spaces」と位置付け、OSS（オープンソースソフトウェア）としての技術開発や国際標準化等を、IPAを事務局としてワンチームとして推進していくことを2025年10月に発表。

産業イニシアティブ
と推進主体



データスペース技術コンセプト
と推進主体・関連団体

技術仕様と
参照/準拠する活動



推進

Open Data Spaces

推進主体/事務局
経済産業省



連携

推進主体/事務局
独立行政法人情報処理推進機構



技術仕様
ODS-RAM、ODP

参照/
準拠

連携

一般社団法人
データ社会推進協議会



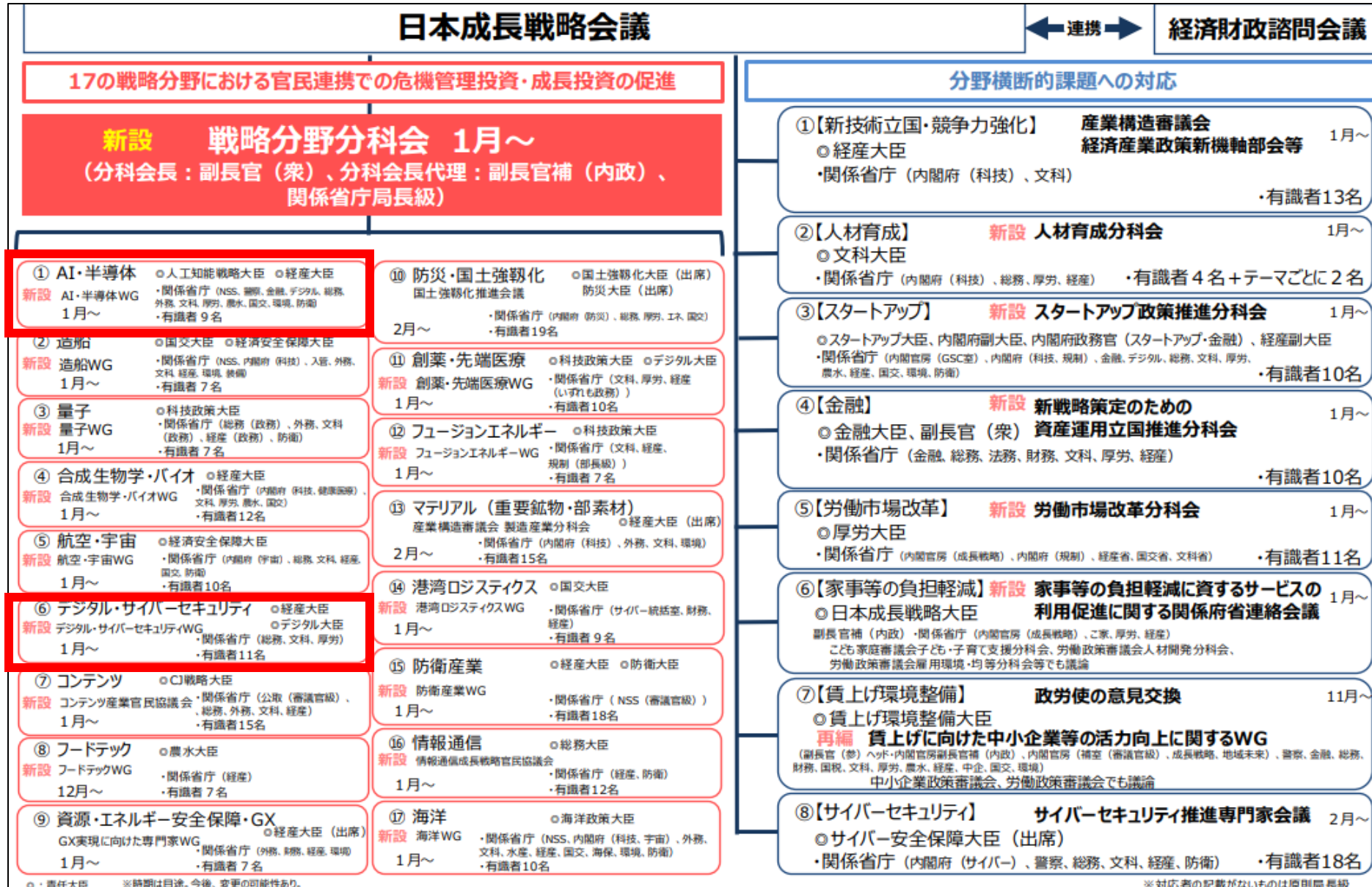
DATA-EX関連活動

ロボット革命・産業IoT
イニシアティブ協議会



RRI WG1
産業データ連携関連活動

(参考) 日本成長戦略の検討体制



○：責任大臣 ※時期は目途、今後、変更の可能性あり。

※対応者の記載がないものは原則局長級

(参考) デジタル・サイバーセキュリティの全体像

