Agile Japan 2026

デジタル赤字と コンテキスト枯渇問題

独立行政法人情報処理推進機構 デジタル＆AIシステムズ・デザインセンター
Open Data Spaces Chief Architect 津田通隆

本資料の無断転載及び複製、AI学習等を禁じます.
お問い合わせ: dadc-info@ipa.go.jp

経歴



津田通隆 Michitaka Tsuda
独立行政法人情報処理推進機構
Open Data Spaces最高設計責任者

- 1997年、奈良県宇陀市生まれ。奈良県立畝傍高等学校、大阪大学卒。
- 大学在学中にエストニア共和国タリン工科大学に留学し、フィンテック事業で起業（代表取締役CEO/CTO）、2023年経済産業省に入省（総合職・政治国際区分）。デジタル地方創生、無人航空機運航管理、AI・産業データ政策分野の担務を経て現職。
- 分散データマネジメント技術のオープンソースプロジェクトの最高設計責任者として、日本、欧米のAgentic AI・データマネジメント業界で活動。
- Forbes JAPAN 30 Under 30 2026 (Science & Technology) 受賞。

X公式アカウント

著書



Why Open Dataspaces: 設計思想とアーキテクチャパラダイム



デジタル経済レポート：データに飲み込まれる世界、聖域なきデジタル市場の生存戦略

データに飲み込まれる世界の到来
DATA IS EATING THE WORLD

SDX
SDV市場を典型としたハードウェアとソフトウェアの主従逆転 (Software Defined X)



写真: Tesla gallery

×

AI革命
大規模マルチモーダルモデル（LMM）を中核としたデータパラダイムの転換

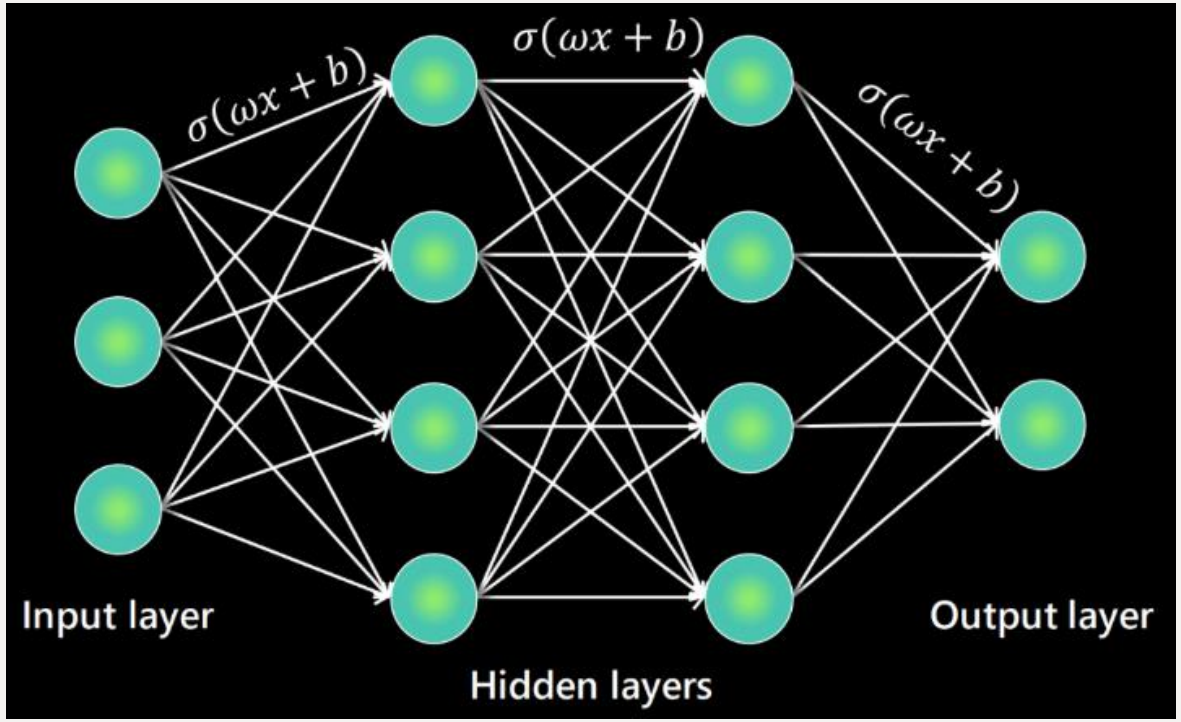


写真: Microsoft AI Tour

技術パラダイムの変化により、「聖域なきデジタル市場」というべき競争環境が到来



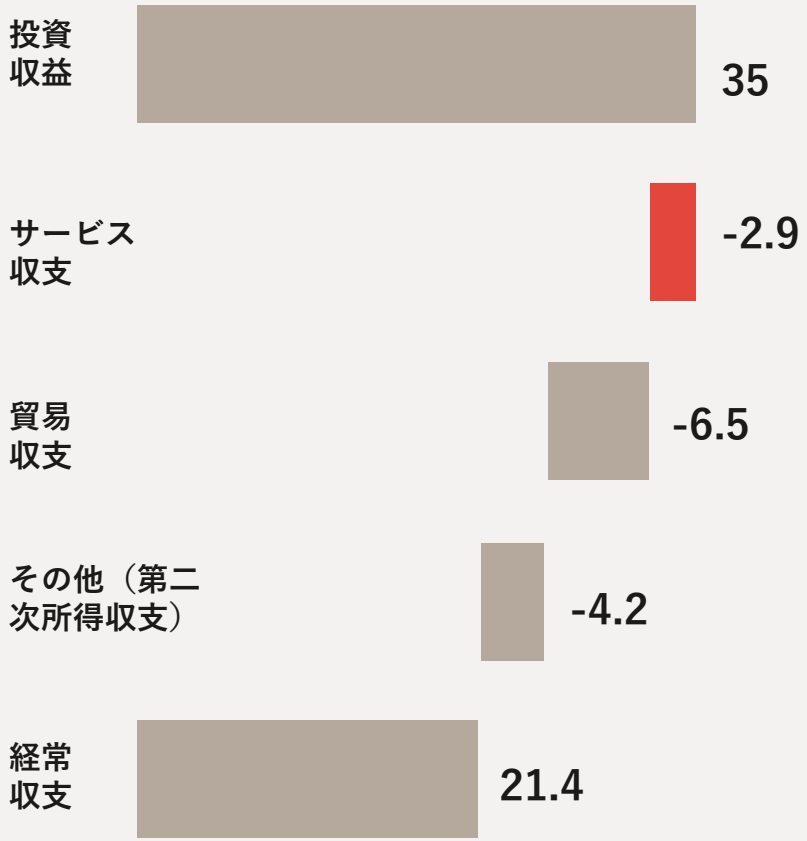
- 製造業等を典型とする伝統産業を含むすべての企業活動の付加価値がソフトウェア、そしてそれを駆動するデータによって規定される。
- ソフトウェアが売れないとハードウェアが売れない。そして、データがなければ価値あるソフトウェアが生み出せず、競争力が維持できない。

出典：経済産業省 若手新政策プロジェクト PIVOT「デジタル経済レポート：データに飲み込まれる世界、聖域なきデジタル市場の生存戦略」
写真：PAKUTASO

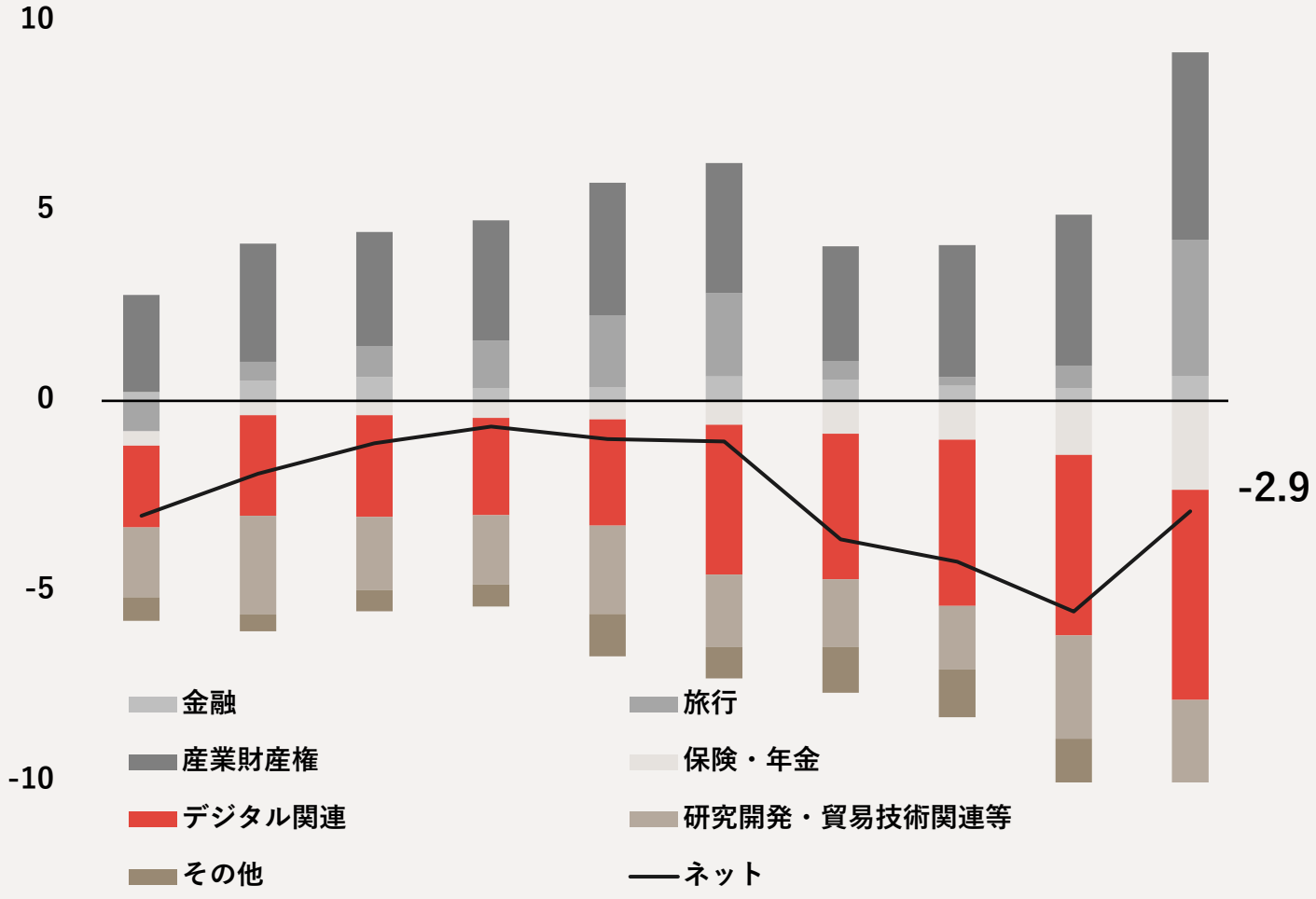
我が国経常収支において最も重要な変数となりつつある「デジタル赤字」（デジタル関連収支の支払超過）

Open Data Spaces

経常収支の構成
2023; 兆円



サービス収支の構成と推移
2014-2023; 兆円

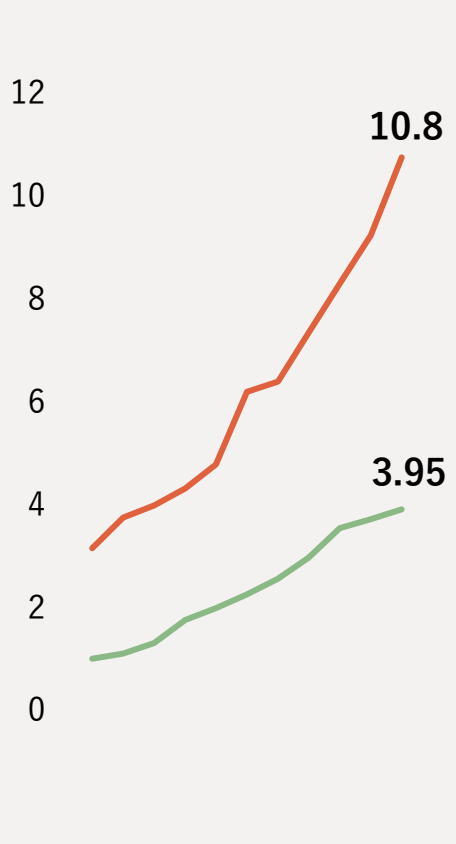


出典：日銀統計資料より経済産業省若手新政策プロジェクトPIVOT作成

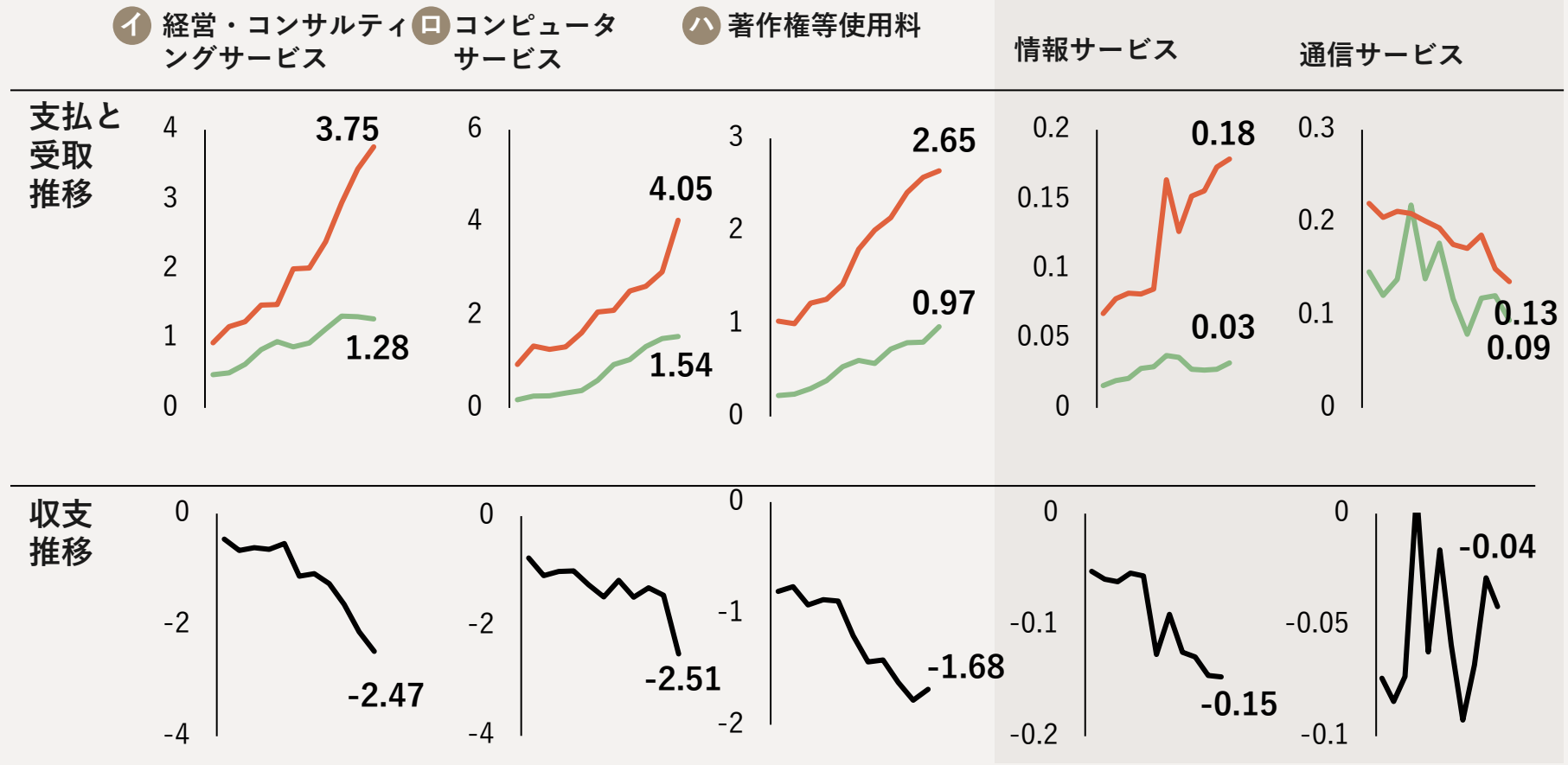
デジタル関連収支の推移とデジタル関連収支を構成する5項目の推移

凡例 支払 受取 収支

デジタル関連収支の推移
2014-2024年; 兆円



デジタル関連収支を構成するサービスの収支
2014-2024年; 兆円



出典：日本銀行 時系列統計データ（2025年4月時点取得）を基に経済産業省若手新政策プロジェクトPIVOT作成

日本銀行モデルとPIVOTデジタル赤字推計モデルの比較

Open Data Spaces

	日本銀行モデル（国際収支関連統計）	PIVOTデジタル赤字推計モデル
目的	対外経済取引・資金フローの動向、対外資産負債残高の構造からみた脆弱性等、各種の経済分析に有用なデータを提供する[1]	再現性、説明力、シナリオ予測性に重点を置き、日本銀行の国際収支関連統計の考え方をベースとしつつも、政府が対処すべき領域の優先順位の検討と政策的打ち手を定量的に評価する
方法論	国際通貨基金が2008年に公表した「国際収支マニュアル第6版」に準拠[2]	事業区分別の各社売上高の積み上げ、送金比率、将来の市場成長性（CAGR）により推計
ソース	外為法に基づく報告書 等	有価証券報告書 等
課題	❶ 国際収支関連統計は外為法に基づき提出された企業の送金情報を計上したものであり、**基礎資料における国際収支項目の割り当ては各社の管理会計に依存**しているため、**企業間でデータ品質や実態反映のばらつきが存在。** ❷ デジタル関連収支には**デジタルに直接関係ない送金項目が算入**されており、データ・ソフトウェアによって生み出された付加価値を適切に反映できていない。 ❸ 国際収支関連統計は個社の機密情報を日本銀行が統計処理の上、マクロ情報として公開しているため、**統計情報からのデジタル赤字の構造問題の分析は難しい。**	解決

1. 日本銀行「国際収支関連統計のFAQ」より引用
2. 日本銀行「『国際収支関連統計（IMF国際収支マニュアル第6版ベース）』の解説」より引用

出典：経済産業省若手新政策プロジェクトPIVOT作成

PIVOTデジタル赤字推計モデルにおける事業区分と国際収支項目との対応関係

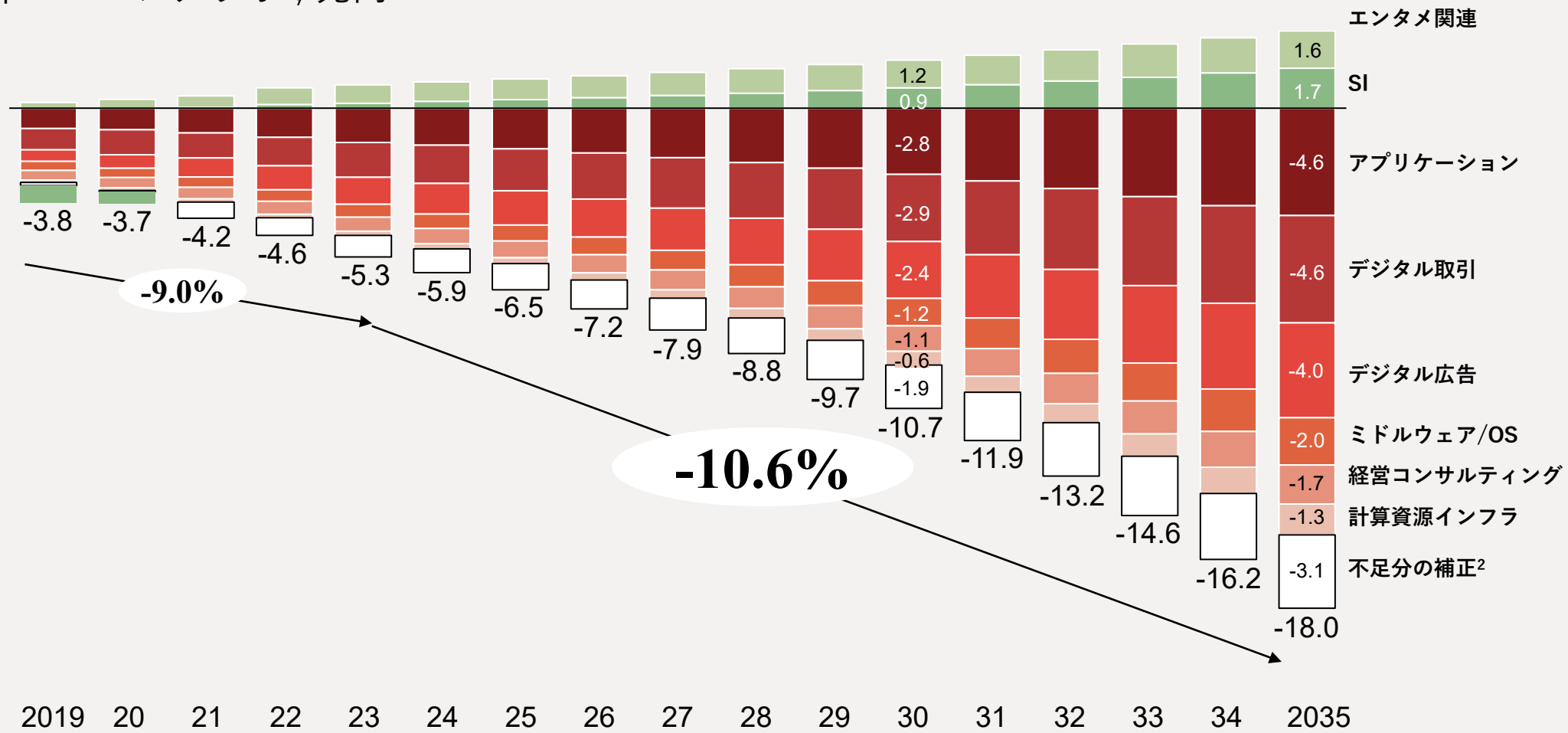
PIVOTデジタル赤字推計モデルの事業区分				国際収支項目との対応関係			定義
				専門・経営コンサルティング	コンピューターサービス	著作権等使用料	
デジタル関連	B2B	❶	経営コンサルティング	✓			ビジネス・戦略コンサルティング、総合コンサルティング、シンクタンクの売上
		❷	SI（システム開発・運用保守）		✓		システム導入、システム開発、システム開発に伴うコンサルティング、ソフトウェア導入支援等による売上
		❸	ソフトウェア：アプリケーション(SaaS含む)		✓	✓	SaaSの利用料、オンプレミス環境またはPaaS上にBYOL※2 形態で導入するソフトウェア販売および保守による売上
		❹	ソフトウェア：ミドルウェア/OS(PaaS含む)		✓	✓	PaaS・DaaS※1の利用料、オンプレミス環境またはIaaS上にBYOL 形態で導入するミドルウェア販売・保守、 OSのライセンス利用料
		❺	計算資源インフラ（DC, IaaS）		✓	✓	IaaSの利用料
	B2C	❻	デジタル広告	✓			検索エンジン、SNS、動画サイト、EC等における広告売上
		❼	デジタル取引（EC, 音楽, 映像等のPF）		✓	✓	EC、音楽配信サービス、映像配信サービス、旅行予約サイト等のプラットフォーム使用料
その他		❽	エンタメ関連			✓	映像、アニメ、音楽の配信サービスによる売上を除いたコンテンツの売上、及びゲーム（オンラインゲーム含む）の売上

※1 DaaS: デスクトップ仮想化をクラウドコンピューティング方式で提供するもの。 ※2 BYOL: Bring Your Own Licenseの略。

出典：R6年度経済産業省委託調査（図表は経済産業省若手新政策プロジェクトPIVOTにて編集及び一部追記）

PIVOTデジタル赤字推計モデルでは、アプリケーションとMSP由来の赤字がさらに拡大し、2035年に18兆円に到達と予測

2023-35年ベースシナリオ[1]; 兆円



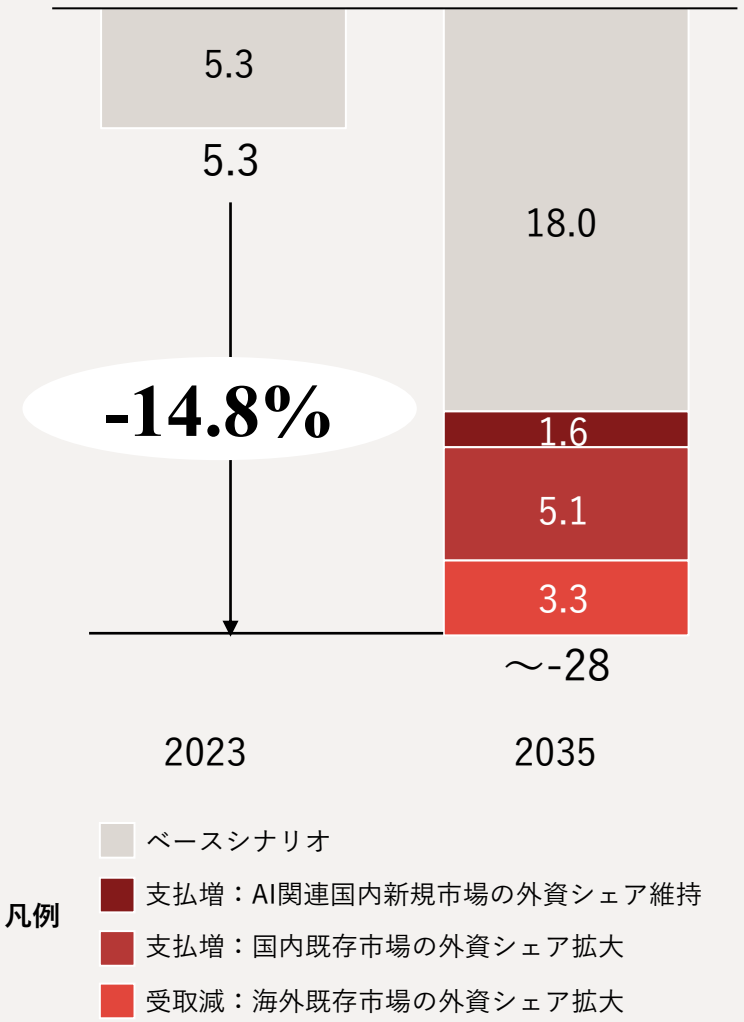
1. 各市場で定義した送金率が最大のケース、かつ、各サービス領域における2022-24の市場CAGRを2035まで引き延ばす形で算出
2. 非デジタル産業における送金が影響している可能性、将来予測は全体の赤字額に比例させる形で算出

出典：経済産業省 若手新政策プロジェクト PIVOT「デジタル経済レポート：データに飲み込まれる世界、聖域なきデジタル市場の生存戦略」

AI革命を踏まえた悲観シナリオでは2035年に28兆円のデジタル赤字に

Open Data Spaces

PIVOTデジタル赤字推計モデルに基づくデジタル赤字の将来予測; 悲観シナリオ; 兆円



2035年のベースシナリオに対する悲観シナリオの差分

兆円

		ベースシナリオとの差分	支払増 AI関連国内市場の外資シェア維持	支払増 国内既存市場外資シェア拡大	受取減 海外既存市場日系シェア縮小
経営コンサルティング		0.1	0.0	0.1	0.0
SI		2.3	0.1	0.7	1.5
ソフトウェア	アプリケーション	1.5	0.4	1.1	0.0
ソフトウェア	ミドルウェア/OS	0.9	0.9	0.0	0.0
計算資源インフラ		0.3	0.2	0.1	0.0
デジタル広告		1.6	0.0	1.0	0.6
デジタル取引		2.5	0.0	2.0	0.5
エンタメ関連		0.8	0.0	0.1	0.7

出典：経済産業省 若手新政策プロジェクト PIVOT「デジタル経済レポート：データに飲み込まれる世界、聖域なきデジタル市場の生存戦略」

サービス収支はソフトウェア産業のみが主眼
聖域なきデジタル市場では、財の貿易収支にも赤字が影響

隠れデジタル赤字
＝デジタル赤字の拡大

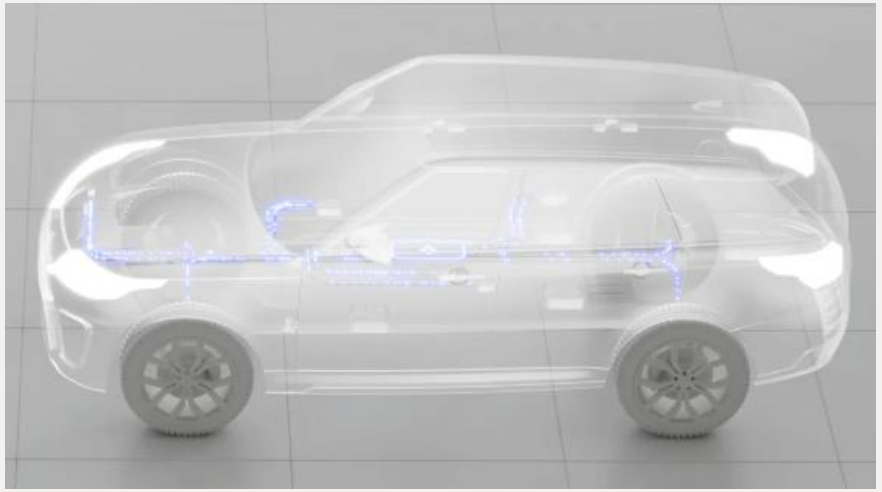


写真: Applied Intuition

自動車産業での事例：
インフォテイメント、ADAS、車載OSの
領域で外製化による開発が加速

≈1.3兆円

＋

SDX赤字
＝貿易収支の減少



写真: Tesla gallery

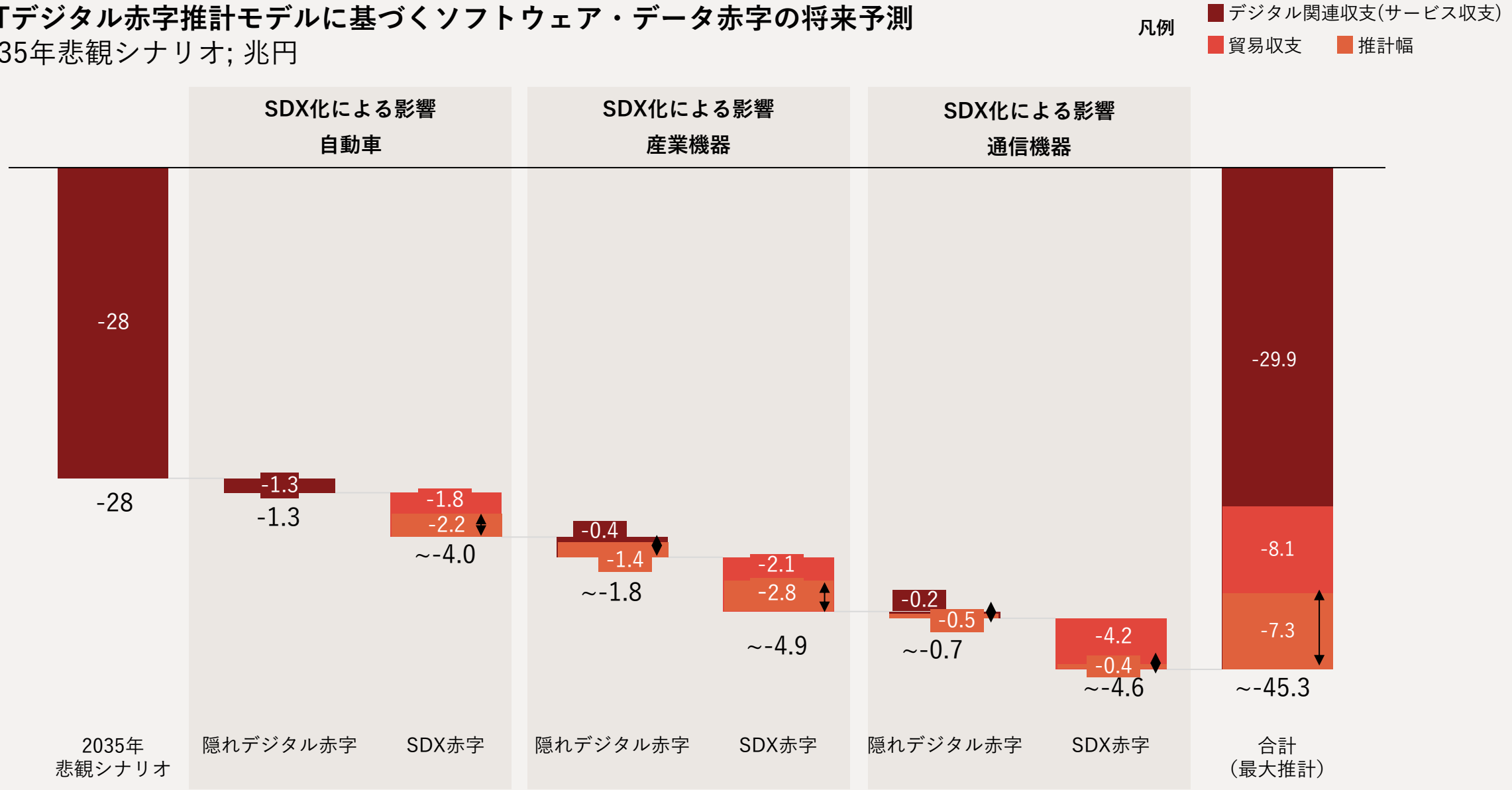
自動車産業での事例：
SDVの波を受け、日系OEMのグローバル
販売台数が減少

≈4.0兆円

広義のデジタル赤字は、最大推計で45.3兆円まで到達しうる可能性

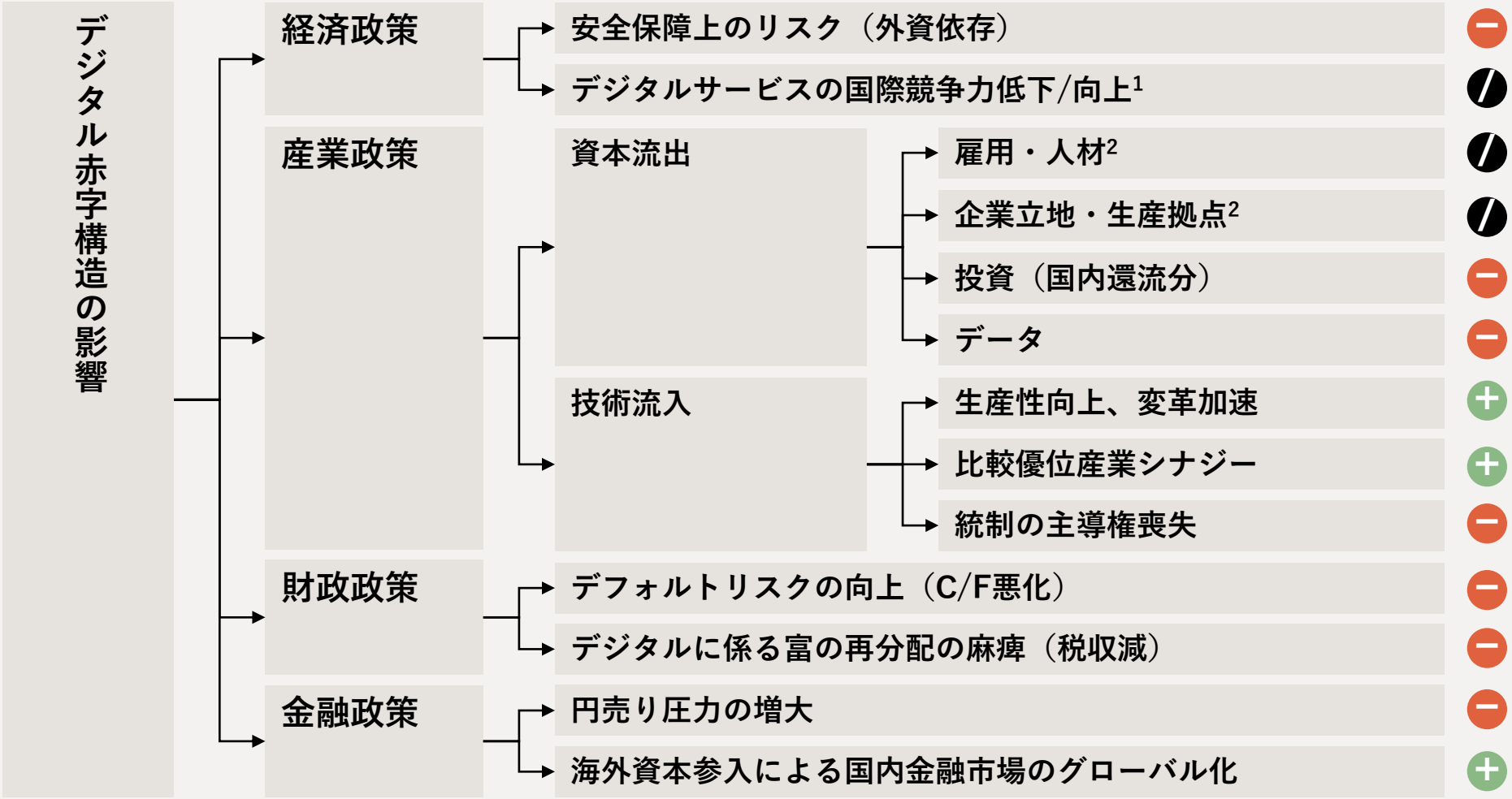
Open Data Spaces

PIVOTデジタル赤字推計モデルに基づくソフトウェア・データ赤字の将来予測
2023-35年悲観シナリオ; 兆円



出典：経済産業省 若手新政策プロジェクト PIVOT「デジタル経済レポート：データに飲み込まれる世界、聖域なきデジタル市場の生存戦略」

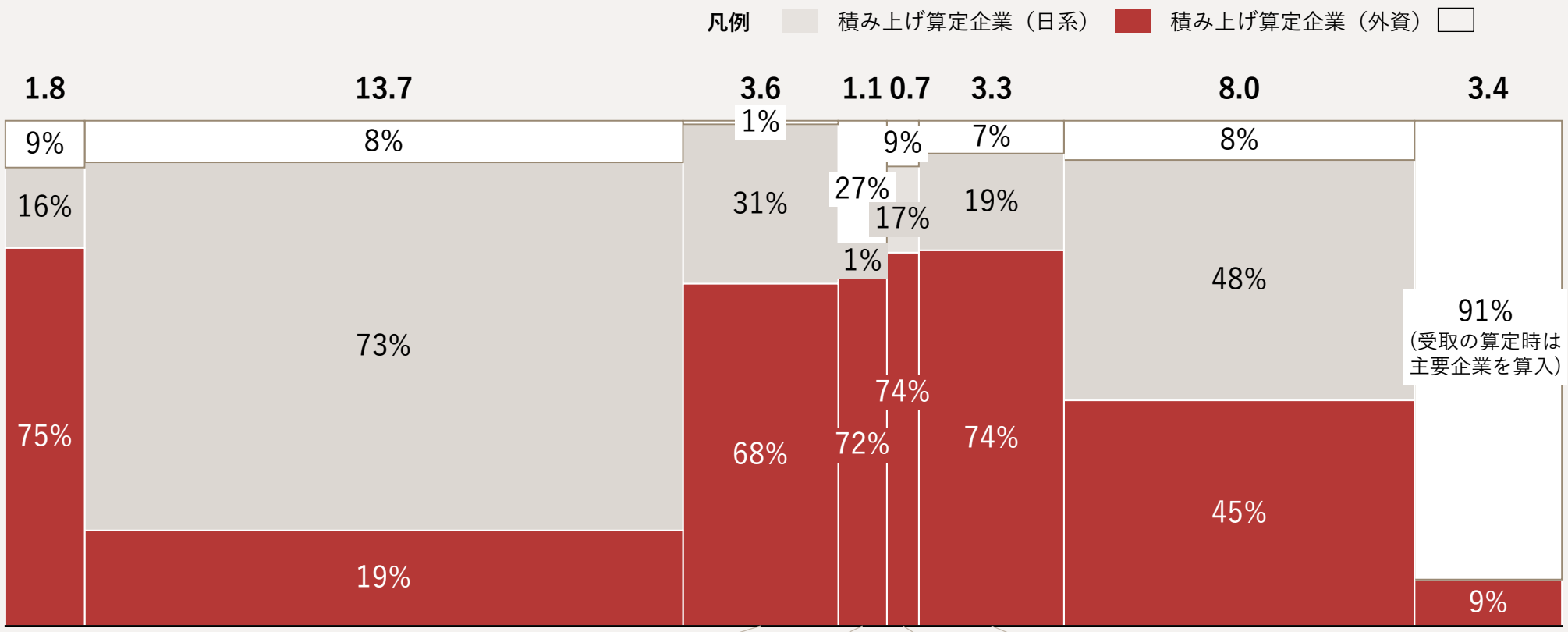
「デジタル赤字」それ自体は問題ではない
しかし、その背景にある「構造」は問題である



出典：経済産業省 若手新政策プロジェクト PIVOT「デジタル経済レポート：データに飲み込まれる世界、聖域なきデジタル市場の生存戦略」

事業構造―「資本効率の差」がデジタル赤字構造の根源

Open Data Spaces



国内市場規模;兆円（2023年）	経営コンサル	SI	アプリケーション	ミドルウェア/OS	計算資源インフラ	デジタル広告	デジタル取引	エンタメ関連
市場CAGR;%（2022-24年）	8.8	3.5	10.2 (12.9)※1	11.5 (22.9)※2	18.1	10.0	9.7	2.1
業界粗利率;%	30-60	20-40	80-90	60-80	50-60	50-80	20-70	個別のビジネスモデルに依存

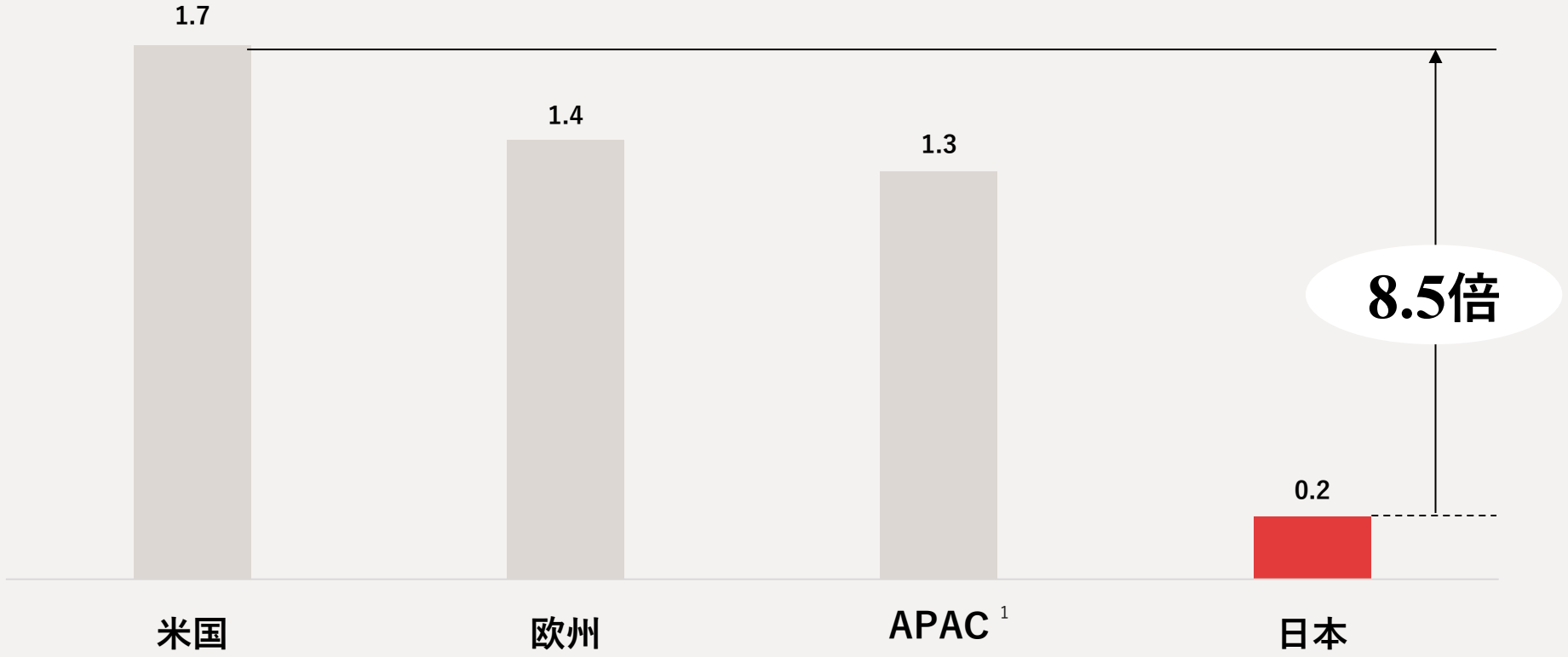
※1SaaSのみのCAGR　※2PaaSのみのCAGR

出典：経済産業省 若手新政策プロジェクト PIVOT「デジタル経済レポート：データに飲み込まれる世界、聖域なきデジタル市場の生存戦略」

Open Data Spaces

市場構造―国内市場への過適合（PMOF）により デジタル産業の原則であるスケーラビリティで勝てない

地域別ICT支出 / 投資額
2022年; 兆ドル



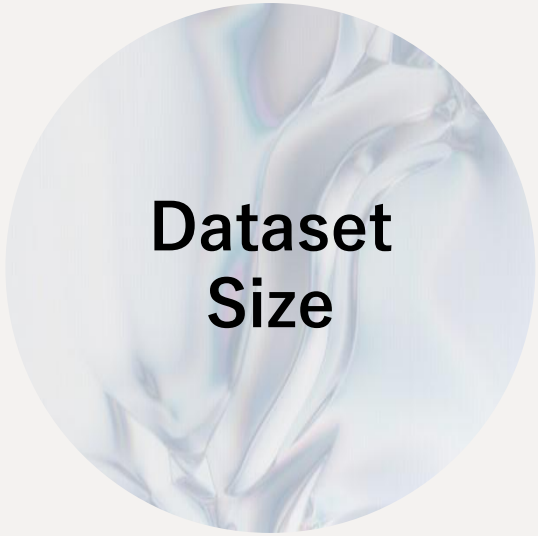
1. APAC：日本及び中国を含まないアジア太平洋地域

出典：経済産業省 若手新政策プロジェクト PIVOT「デジタル経済レポート：データに飲み込まれる世界、聖域なきデジタル市場の生存戦略」

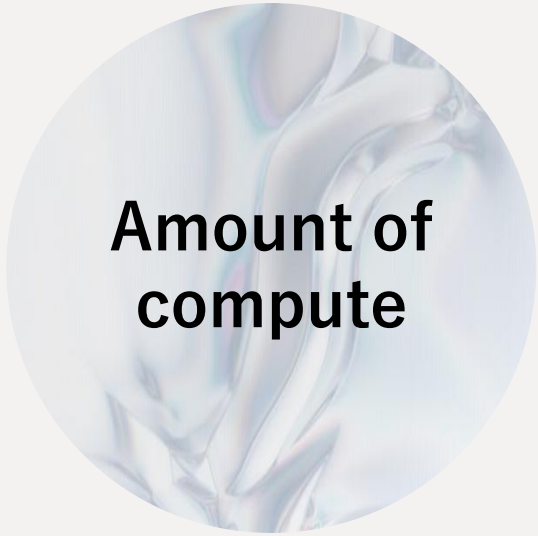
Frontier Labsはモデルの性能を決定する「スケーリング則」を追い続けており、そのボトルネックはデータとコンピュート



ー
（データ量と同程度が
最適配分）



大規模な学習データセット

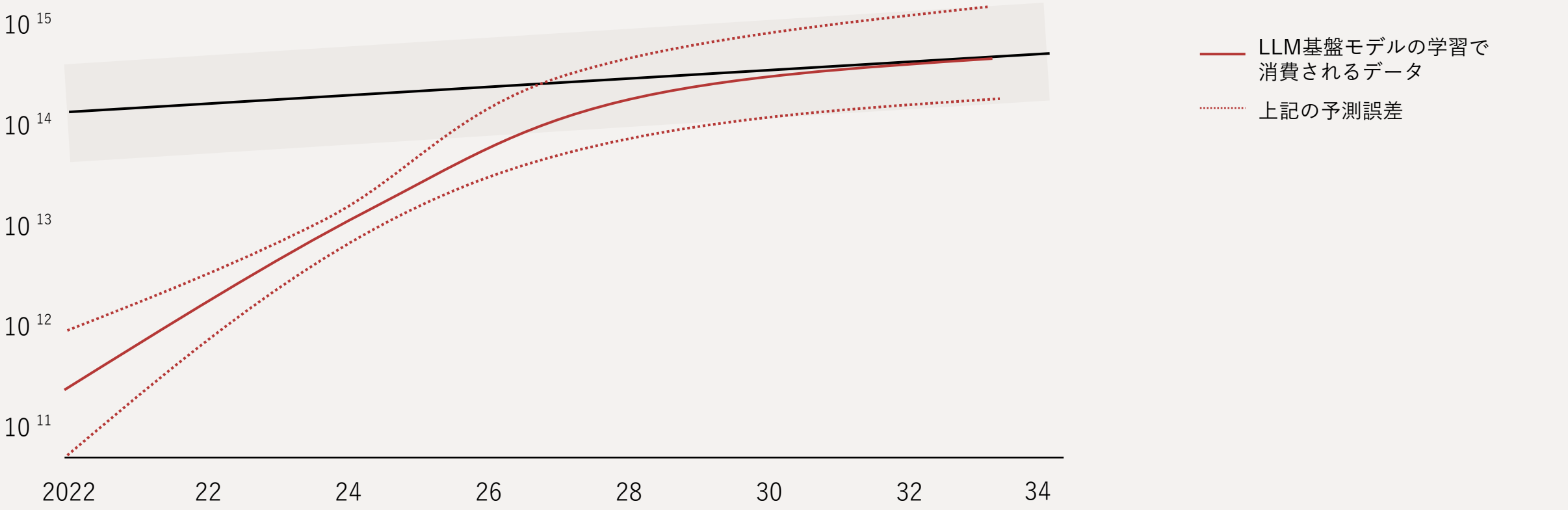


半導体と電力

Open Data Spaces

現在の学習速度ではスケーリング則の前提であるインターネットデータが2026-32年の間に枯渇し、モデルが進化を止める

LLM基盤モデル学習におけるパブリックテキストデータ利用の予測
2022-34年(予測値ベース); トークン数



出典：Epoch AI. “Will We Run Out of Data? Forecasting Dataset Size for Language Models.” Technical Report, June 2024. をもとに発表者作成

クローリングの拒否、「Pay Per Crawl」での課金で、同意・アクセスによる供給キャップも生まれ始めている

拒否・制限

5%

1年でAI向けの収集制限（Robots.txt）が追加されたデータ
2023-24; N=約1.4万ドメイン[1]

遮断・課金

ウェブ全体の約2割のトラフィックを扱うCloudFlareでは、新規ドメインでAIクローラーを規定遮断

「Pay Per Crawl」の課金サービスも開始

出典：1 Data Provenance Initiative “Consent in Crisis” / MIT Technology Review

合成データとモデル崩壊問題：
合成データは代替ではなく、リアルデータの増幅器として機能

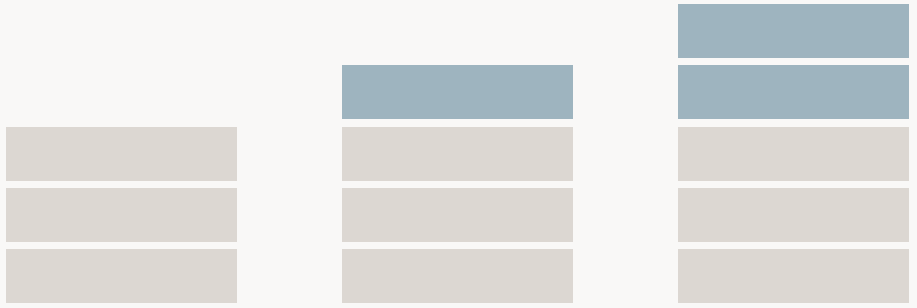
合成データでの
置き換え・再帰学習



前の世代のAIが作ったデータだけで次を学習。まれな事象が消え、分布が痩せていく。

モデルが崩壊する

合成データの蓄積・混合



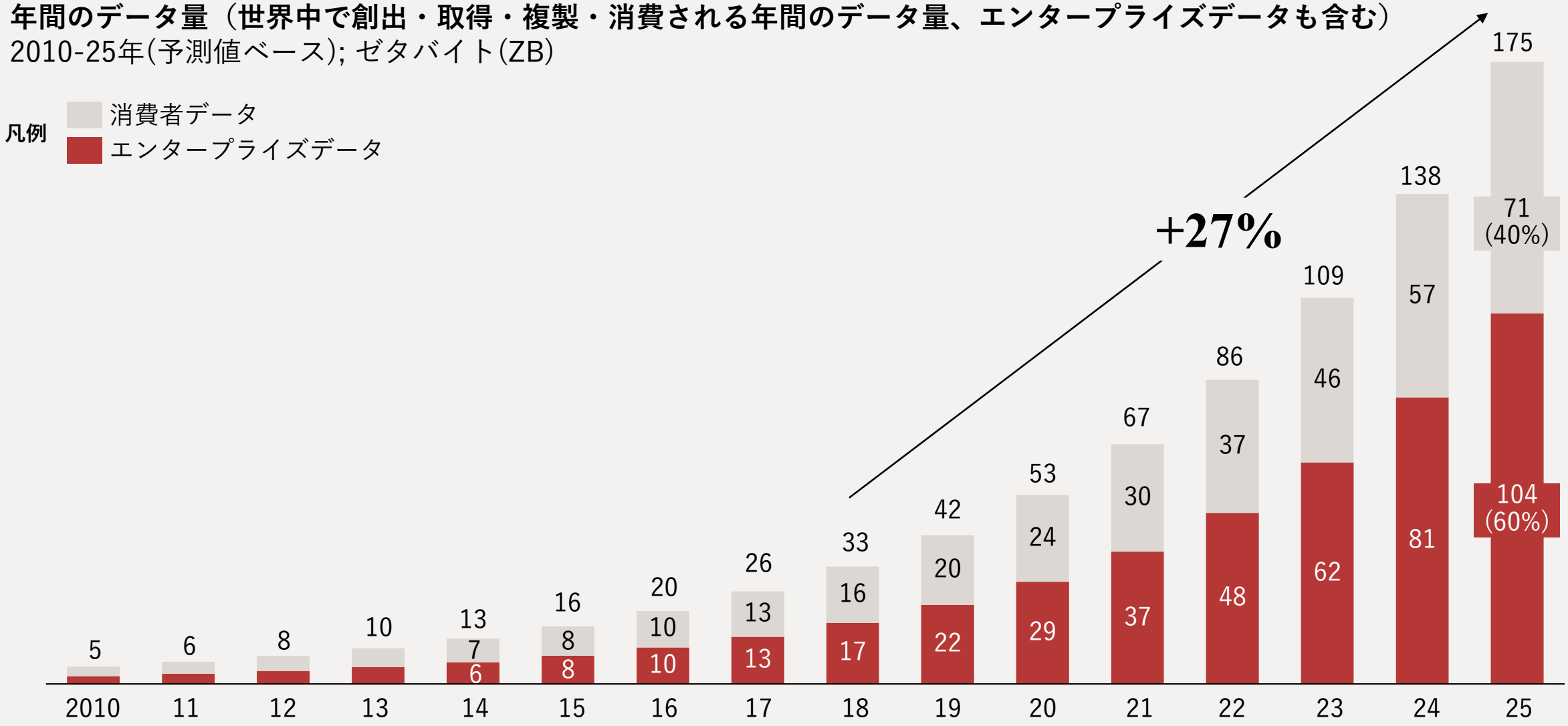
元の実データを残したまま合成データを積み増す。現実のウェブもこの形に近い。

モデルは崩壊しない

年間のデータ量は2025年時点で175ZBに到達し、全体の60%をエンタープライズデータが占めている

Open Data Spaces

年間のデータ量（世界中で創出・取得・複製・消費される年間のデータ量、エンタープライズデータも含む）
2010-25年(予測値ベース); ゼタバイト(ZB)



出典：国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）「ウラノス・エコシステムの実現のためのデータ連携システム構築・実証事業」調査資料※をもとに発表者作成
※The Digitization of the World From Edge to Core – IDC

Open Data Spaces

AIが学習対象とし得る17.5ZBのユニークデータのうち約92%(16ZB)のデータが現時点で活用できていないと推定

グローバルデータスフィア全体のデータ量
2025年(予測値ベース); ゼタバイト(ZB)

		ユニークデータ 10%				複製データ 90%			
		非構造	非構造	構造/非構造	半構造	非構造	非構造	構造/非構造	半構造
		エンタメ画像／動画データ 27%	非エンタメ画像／動画データ 26%	プロダクティビティデータ 27%	組み込みデータ 20%	エンタメ画像／動画データ	非エンタメ画像／動画データ	プロダクティビティデータ	組み込みデータ
消費者データ 40%	公開データ 10%	0.51 ZB	枯渇するデータ 0.08 ZB	0.08 ZB	0.08 ZB	複製データのため、本質的な価値はない			
	制限付き・非公開データ 90%	4.53 ZB	0.65 ZB	0.65 ZB	0.65 ZB				
エンタープライズデータ 60%	公開データ 10%	0.04 ZB	枯渇するデータ 0.36 ZB	0.38 ZB	0.27 ZB				
	制限付き・非公開データ 90%	0.36 ZB	3.22 ZB	3.37 ZB	2.37 ZB				

出典：国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）「ウラノス・エコシステムの実現のためのデータ連携システム構築・実証事業」調査資料※をもとに発表者作成。
※The Digitization of the World From Edge to Core – IDCをもとに推計。

データは「採掘」から「調達・製造」の時代へ

データ調達

専門家データ

専門家を雇い、問題・正解・評価基準を作らせる

MERCOR

物理・金融などの専門家を契約者として集めて回答・思考プロセスデータを収集

scale

データラベリング大手（Metaが買収）

リアルデータ仲介

データの保有者と、AI開発者をつなぐ

Protege

病院やメディア等のデータを仲介

troveo

動画、音声等のデータを仲介

データ製造

データファクトリー

ロボットや模擬職場を動かしてデータを生成する

mecka

一人称視点の身体知能データを収集・製造

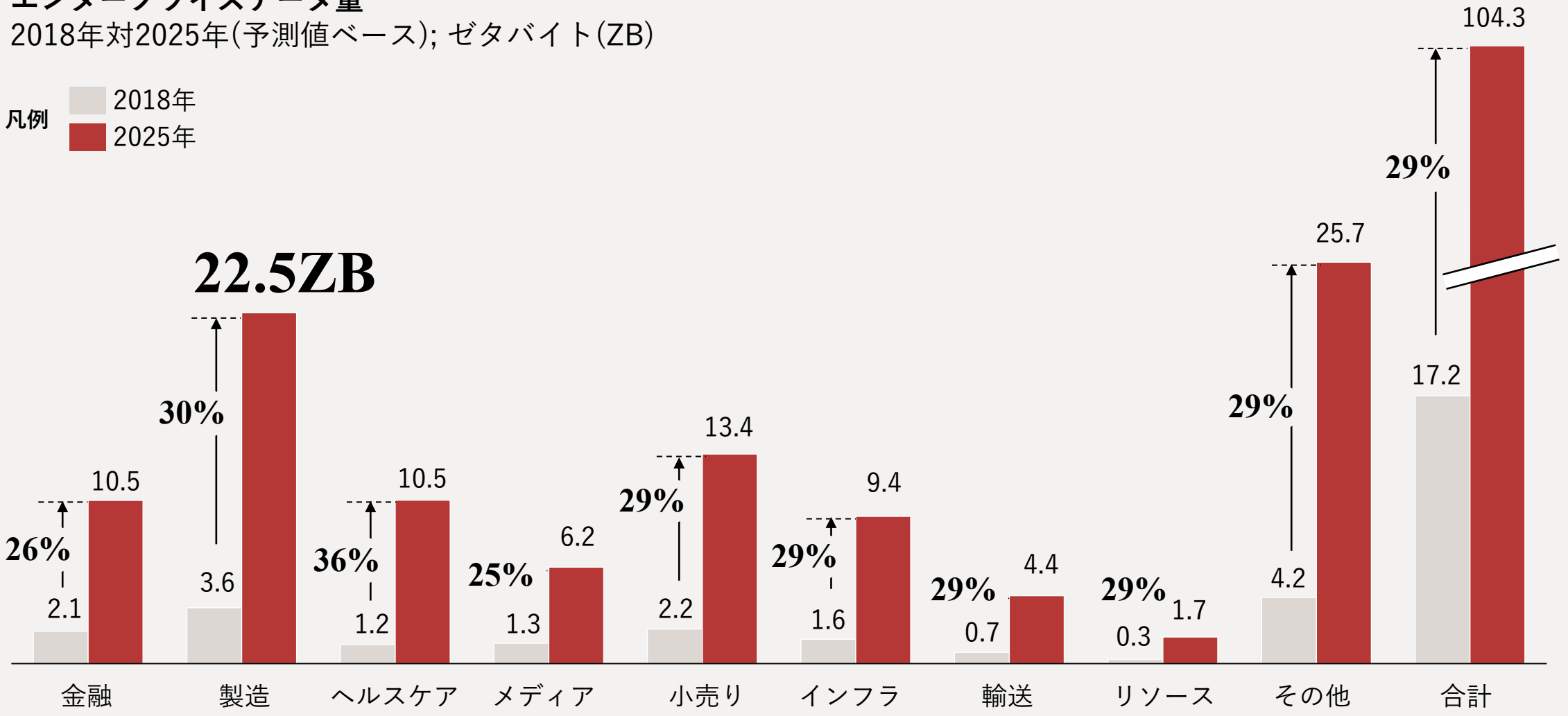
tutor

遠隔オペレーターのチューター形式でのデータ収集・製造

エンタープライズデータ量は、製造分野が全体の22%を占め、金融、小売り、インフラ分野が追従

Open Data Spaces

エンタープライズデータ量
2018年対2025年(予測値ベース); ゼタバイト(ZB)



出典：国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）「ウラノス・エコシステムの実現のためのデータ連携システム構築・実証事業」調査資料※をもとに発表者作成。
※The Digitization of the World From Edge to Core – IDCをもとに、2025年のデータについては全体・金融・製造・ヘルスケア・メディアはレポート記載のCAGR(グラフ掲載データ)をもとに算出、その他産業は左記CAGRの平均値から算出

しかし、LLMアーキテクチャではAIモデルがコモディティ化しつつあり、現場特有の文脈（Domain Context）への需要シフトが起きつつある



大量の学習データが必要

ドメイン固有のコンテキストが必要

なぜ、Agentic AIの社内導入は成果が出ないのか？
問題はモデルではなく、現場の文脈と信頼不足。

AIパイロットの状況

グローバルで300億〜400億ドル投資しているが…

95%[1]

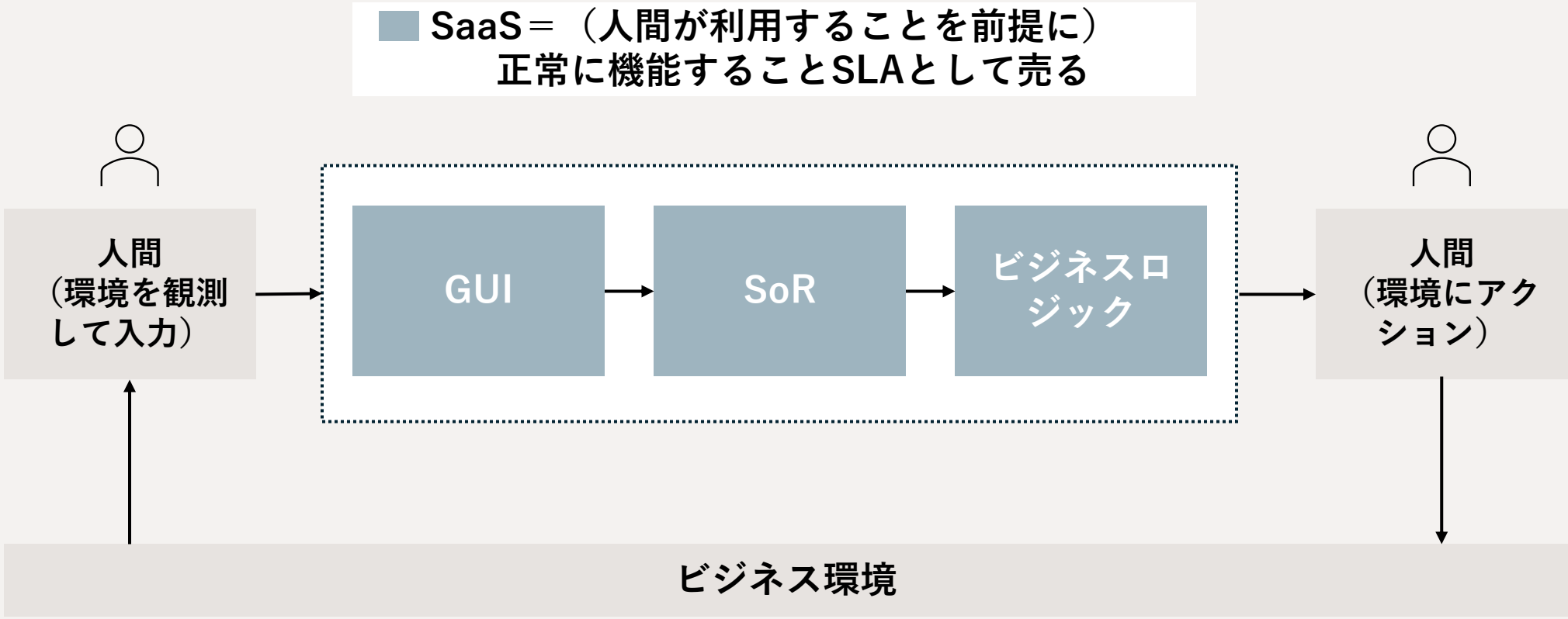
の企業向けAIパイロットは、測定可能な損益（P&L）へのインパクトを生んでいない。

原因

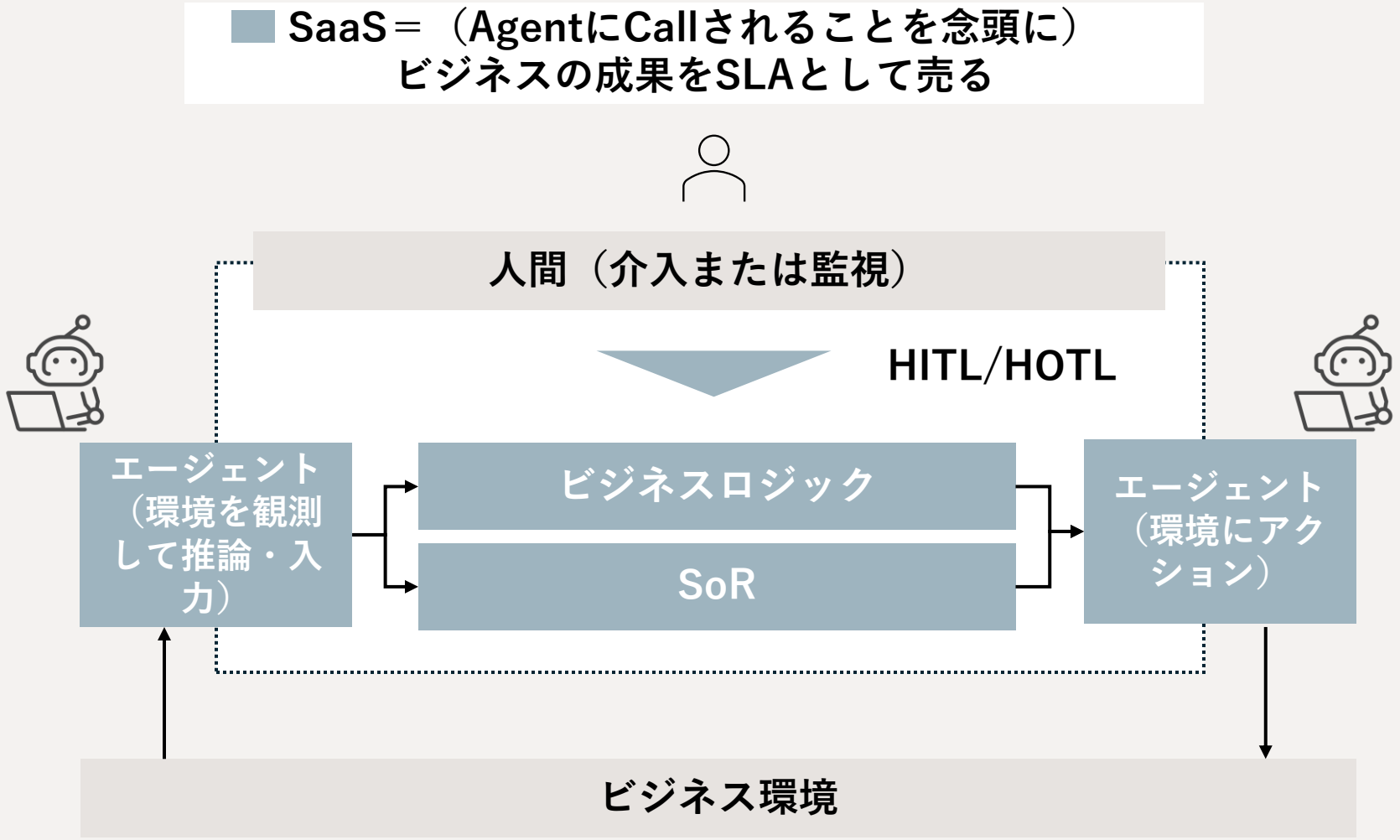
- **ボトルネックはモデルではない。**
- **AIエージェントが、各ドメイン固有の適切なビジネスコンテキストを見つけられず、理解できず、さらにそれに基づいて信頼して行動させることもできないから。**

1: MIT (2025) GenAI Divide

Agentic AI時代以前：「融通の利かないアルゴリズム」が駆動する人間依存のアプリケーションの世界



Agentic AI時代以後: 「融通の利くアルゴリズム」が駆動するエージェンティック（代理人的）なアプリケーションの世界

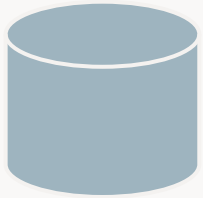


Agentic AI時代の本格到来でビジネスロジックとSoRの役割が変化

A: 概要		B: 役割の変化		C: 新技術トレンドの喚起
		人間中心	エージェント中心	
ビジネスロジック	業務ドメイン特有のオペレーション、意味解釈、アクションのルールを定める	オペレーションの固定化（FTS）を前提に、環境に干渉しない静的なアルゴリズムとして表現	オペレーションの変化を前提に、環境に干渉する動的なアルゴリズムとして表現	例: セマンティックレイヤー、オントロジー
SoR	ビジネス環境のステート（状態）を管理する	全社で「集約」してアルゴリズムに一様な形式で「入力」する	ドメインごとに「提供」してアルゴリズムが「参照」する	例: データメッシュ（データプロダクト）

21世紀以降のデータマネジメントにおける技術テーマの変遷

2004~06



Volume

大量のデータ

例

~2009



Variety

多様性のあるデータ

MongoDB. cassandra JSON

~2012



Velocity

非同期で流れるデータ

kafka Flink

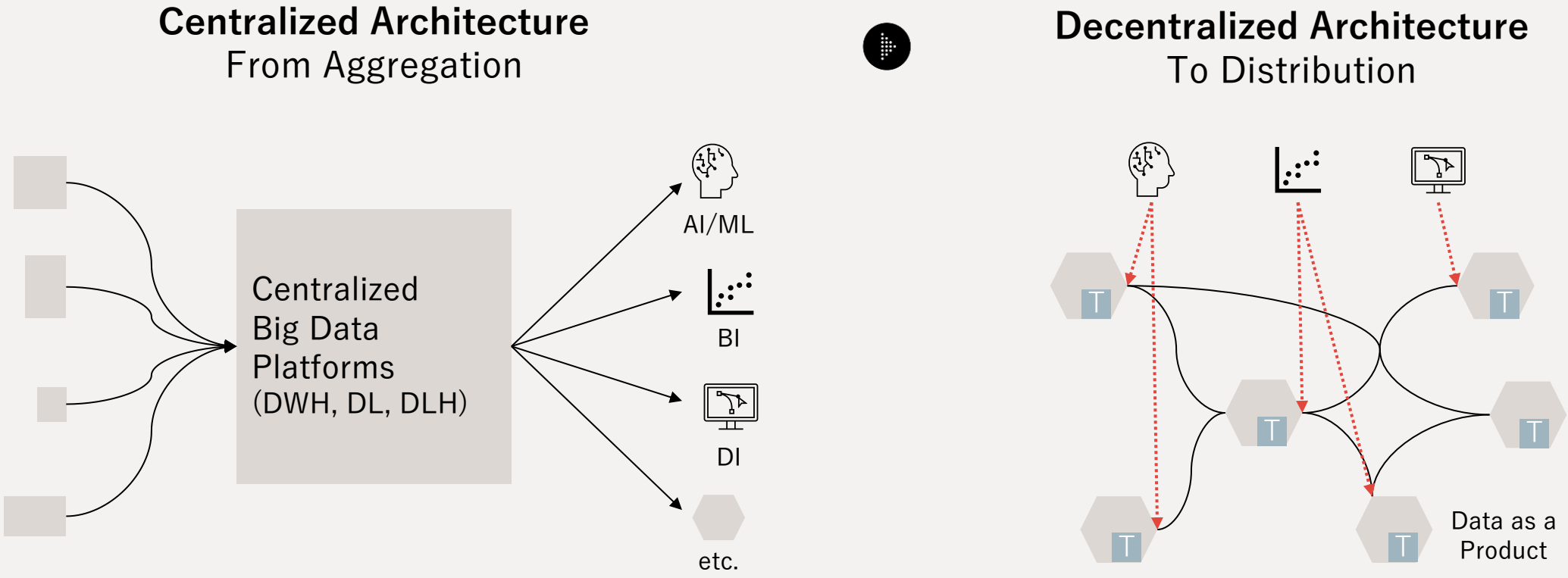
Big Data is Dead:
業務ドメイン特有の「コンテキスト」は中央集権のデータパイプラインで希薄化してしまう

中央集権型の3つの失敗モード

1. **中央チームがボトルネックになる**
   全社の要求が1チームに集中し、待ち行列ができる
2. **業務を知らない人がパイプラインを書く**
   変換の過程で、データのコンテキストが落ちていく
3. **提供側に品質責任の動機がない**
   集めるほど、誰も中身に責任を持たない状態になる

集めれば集めるほど、誰も中身に責任を持たないデータスワンプができる

分散データマネジメントアーキテクチャData Meshによるパラダイムシフト（2019年提唱）



しかし、2019年以降、データメッシュは一度頓挫。問題は組織だけではなく、ドメインごとに形式と道具が割れてしまう技術基盤の不在にもあった。

顕在的な理由：組織の壁

- 全社のドメインマップを描こうとして分析麻痺に陥る
- 既存の部署に「今日からあなたがオーナーです」と看板を付け替えるが、オーナーシップはない
- ガバナンスを、組織間の交渉として設計してしまう

・・・

18〜24ヶ月の文化的抵抗のあと、放置されたカタログだけが残る

潜在的な理由：技術の壁

各ドメインが自前でデータを持つと、2019年の技術で分散したのは責任ではなく、形式と道具のサイロだった。渡すたびにコピーが増え、「どれが正か」が分からなくなる。

マーケティング Airflow + Parquet	営業 スクリプト + CSV
財務 SQL Server	分析チーム ノートブック内で完結

第3の原則「セルフサーブ・データ基盤」が、当時は技術的に成立していなかった

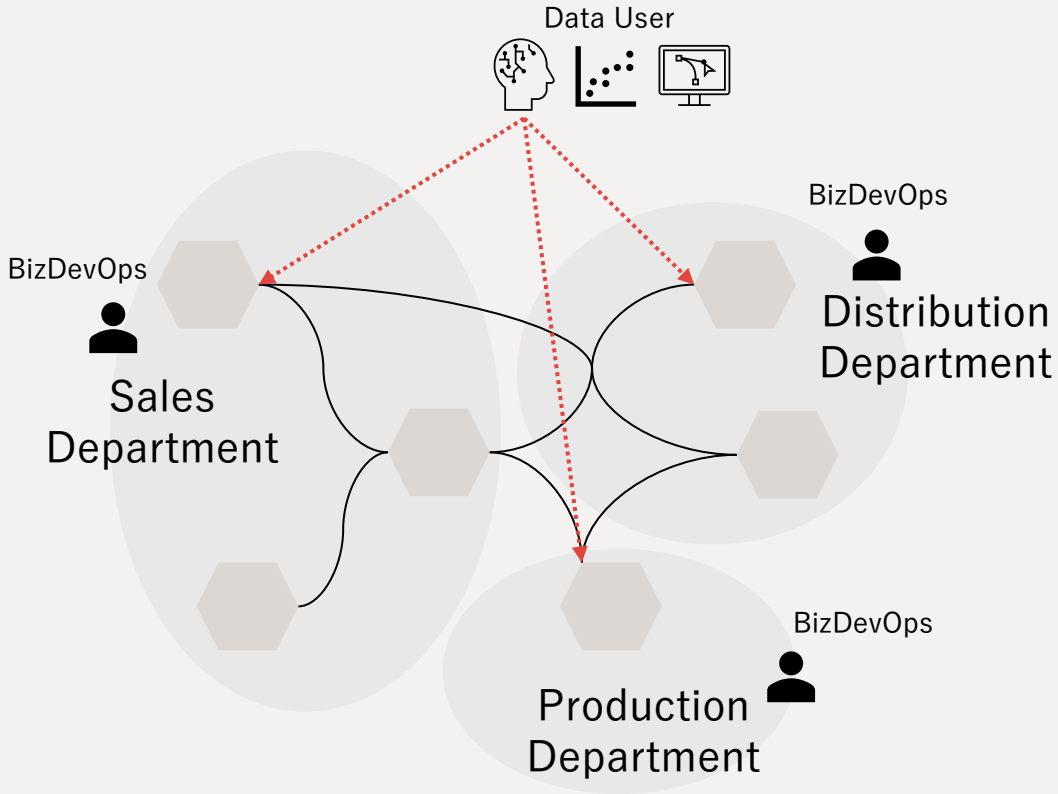
オープンテーブルフォーマットとRESTカタログにより技術要素が揃い、データを動かさずに提供する「データプロダクト」が現実的な手段になりつつある

対象	データレイクハウスの構成要素	技術課題	解決策
インフラ	**列志向のファイルフォーマット**  Apache Parquet（or Apache Avro） ＋ オブジェクトストレージ	データ構造の制約／スケーラビリティとコスト	コンピュートとストレージの分離
エンジン	**オープンテーブルフォーマット**  Apache Iceberg ／ Delta Lake (or Apache Hudi) +クエリエンジン（Apache Spark／Trino 等）	ストレージ形式の占有	データインスタンスとクエリエンジンの分離
テナント	**オープンRESTカタログ（＋Credential vending）**  Apache Polaris／Unity Catalog	メタデータとクレデンシャルの占有	データプレーンとコントロールプレーンの分離
ドメイン	**データ／メタデータのオープン共有プロトコル**  （INCUBATING） OpenSharing／Apache Ossie	データの複製／意味の乖離	データ提供と利用の分離

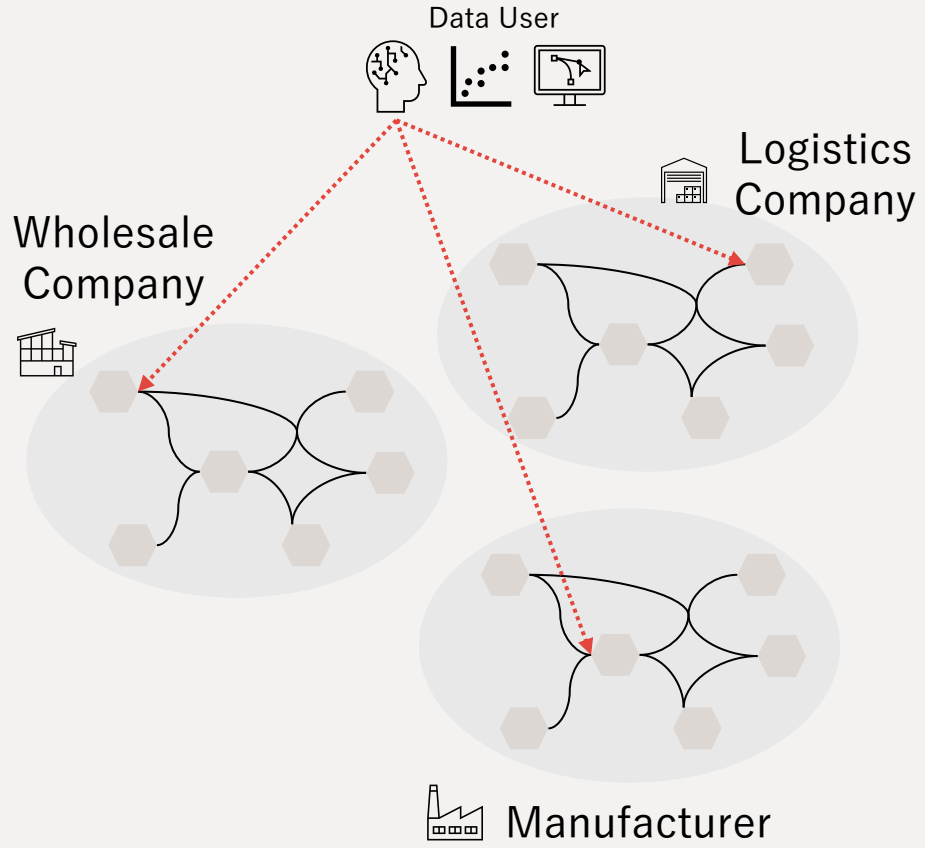
データを動かさずに、複数のエンジンから、そして複数のドメインから同じテーブルを読める。複製コストと「どれが正か」の問題が消える。

ドメイン駆動型のAgentic Industryの実現を目指す：Inter-Departmentから、Inter-Organizationへ

Data Mesh
From Inter-Department



?
To Inter-Organization



Open Data Spacesとは何か？
Agentic AI時代に必要な、ドメイン・産業を横断するコンテキストレイヤーの技術パラダイムを提供することが目的

概要

Open Data Spaces

国や組織ごとの多様性を尊重する、オープンでスケーラブルな**分散データマネジメント**の技術コンセプト

設計指針

- **Vendor-Agnostic**
- **Regulation-Agnostic**
- **Product-like & Service-Oriented**

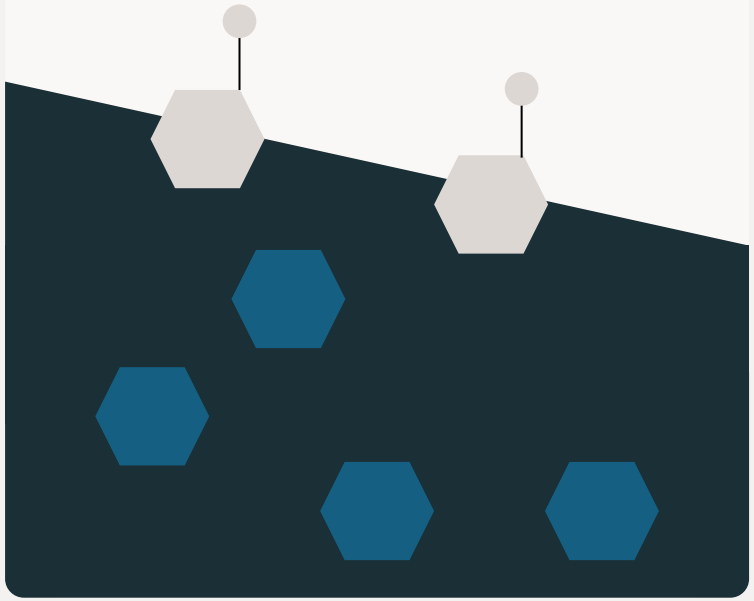
公開済のアーティファクト

設計思想とアーキテクチャパラダイム	ODS リファレンスアーキテクチャモデル (ODS-RAM)	ODS Protocols (OSP)	ODS Middleware and SDKs (OSS)
Ready for **New Paradigm**	Ready for **Onboarding**	Ready for **Interoperability**	Ready for **Go-To-Market**

Open Dataspacesは、横断的な分散型データマネジメントに秩序と規律を与えることを目的とする

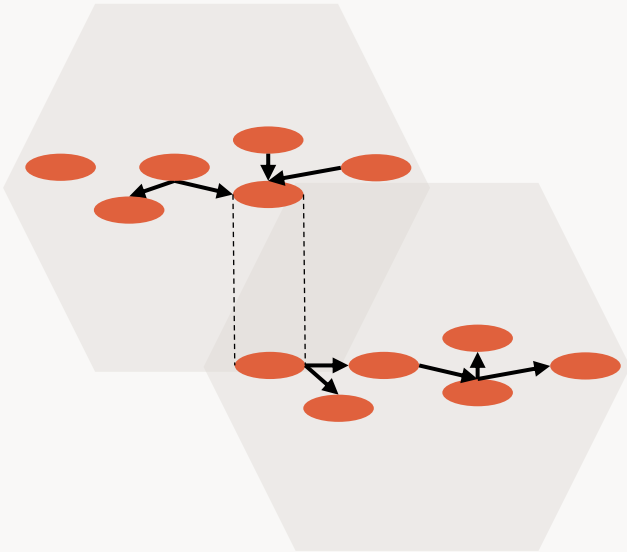
Where to Get:

Data Addressability and Discoverability（DAD）



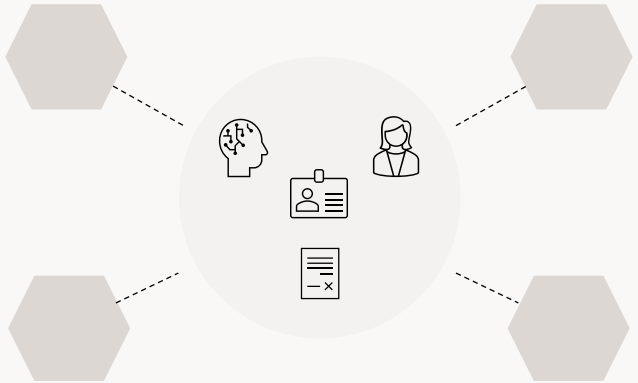
What to Mean:

Ontology and Semantic Interoperability（OSI）



Who and How to Use:

Identity and Usage Control（IUC）



分散型Data Managementにおける「ドメイン、企業横断の透過的なSSOT」と
「データ提供者が対価を得るメカニズム」を実現

高市政権 日本成長戦略（戦略17分野データプラットフォーム）、人工知能基本計画等にOpen Data Spacesが明示的に記載

日本成長戦略



日本成長戦略

2026年７月21日

（２）デジタル・サイバーセキュリティ

地方の社会経済を支える行政サービスやエッセンシャルサービスも含めたＡＸ／ＤＸの実現には、自律的なデジタル・エコシステムの実現が不可欠である。企業や公共領域等で管理されるデータを産業戦略上の焦点としてデータ基盤の高度化を推進するとともに、クラウド基盤の高度化を進めるほか、これらを支えるデジタル人材の育成と、新たな脅威に対応するセキュリティ産業基盤を強化する。

①データプラットフォーム

ⅰ）重要性

ＡＩの普及に伴い、データをＡＩで利用可能な状態にするデータ精製等を行うデータプラットフォームの重要性が増大している。産業競争力や経済安全保障に直結するデータを他国のプラットフォームに依存せずに安心して処理できる国内サービスの確保が急務である。

ⅱ）勝ち筋・方向性

製造業等で豊富なデータを有する強みをいかし、フィジカルＡＩの進展も見据え、データ精製技術や組織を超えたデータ連携技術（データスペース）の開発等を通じ、国内プラットフォームサービスの育成につなげる。

ⅲ）主な目標

・国内のデータプラットフォーム関連市場について、2035年までに市場規模５兆円を目指す。

ⅳ）主な施策

・ＡＩの利活用に不可欠なデータの精製（AI-Ready化）・連携（Open Data Spaces等）に関するミドルウェアの研究開発・実証を通じ、国内プラットフォームサービスを育成するとともに、その成果を幅広い産業に横展開する。

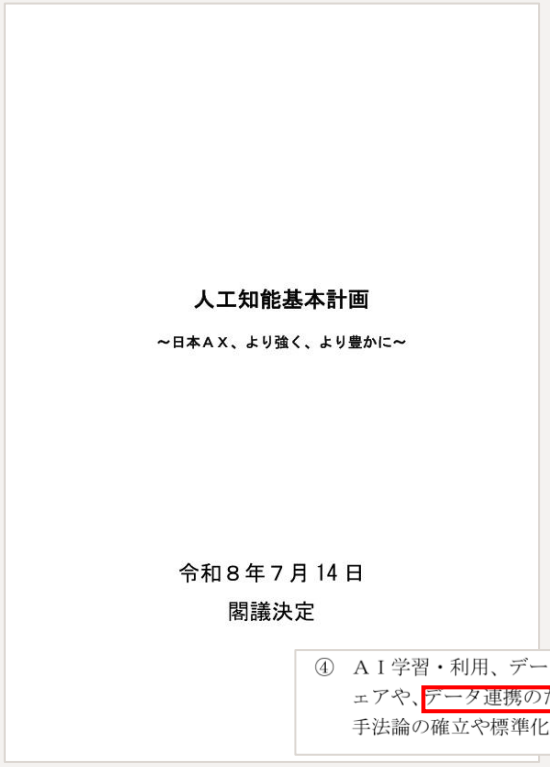
・産業界の実ニーズに基づいたデータ連携のユースケース創出を支援する。官民連携でＧ空間情報のデータ精製を進めるとともに、Ｇ空間データスペースの形成を通じ、多様な分野の課題をジオＡＩ[6]で解決するユースケースを創出する。

[4]「AI for Scienceの推進に向けた基本的な戦略方針」（2026年３月文部科学省決定）を踏まえ、研究プロセスのあらゆる段階においてＡＩを利活用することにより、研究の効率性・生産性を向上させ、科学研究の在り方を革新させる取組を推進。

[5] ＡＩセーフティ・インスティテュート。

[6] Ｇ空間情報とＡＩとの融合。

人工知能基本計画



人工知能基本計画

～日本ＡＸ、より強く、より豊かに～

令和８年７月14日

閣議決定

④ ＡＩ学習・利用、データ連携等のために不可欠なデータ精製技術に関するミドルウェアや、データ連携のためのデータスペース技術（Open Data Spaces等）について、手法論の確立や標準化に係る研究開発・実証を支援する。あわせて、データ精製・

Sovereignty is NOT isolationism

主権は、孤立主義ではない

THANK YOU!



Unauthorized reproduction, redistribution, or use of this material for AI training is strictly prohibited.
Contact: dadc-info@ipa.go.jp