



成長を導くAIインフラ

次の10年の競争力を左右するインフラ戦略とソブリンAI

宝出 幸久

リサーチマネージャー

Infrastructure, Imaging and Devices,
IDC Japan

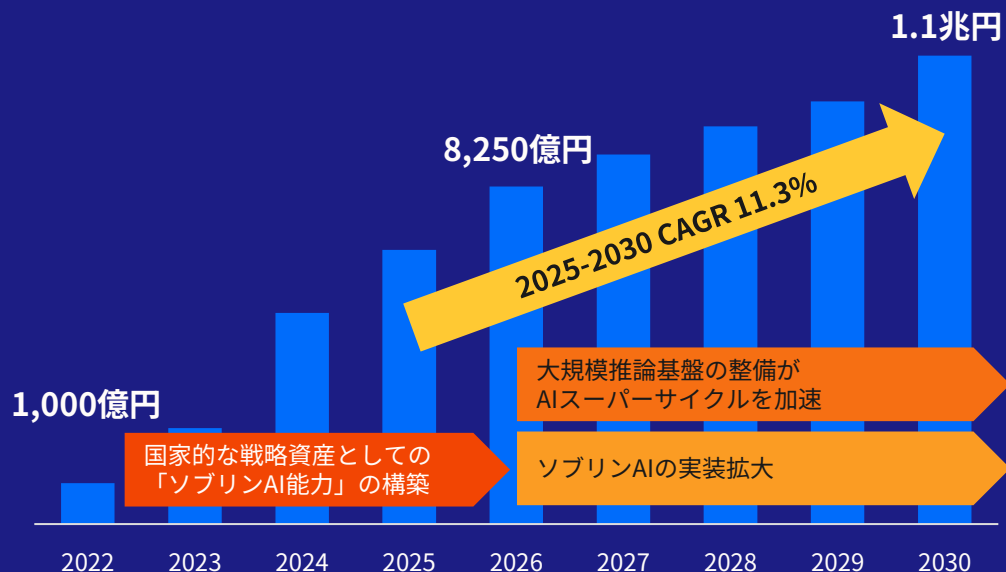
加藤 慎也

シニアリサーチマネージャー

AI and Automation, IDC Japan

学習から推論へ： AIスーパーサイクルの加速とソブリンAIの浸透

国内AIインフラ市場 支出額予測



- ➔ AIインフラは「価値創出の基盤」へ
- ➔ AIインフラの多様化が進む
- ➔ ソブリンAIの選択が拡大する

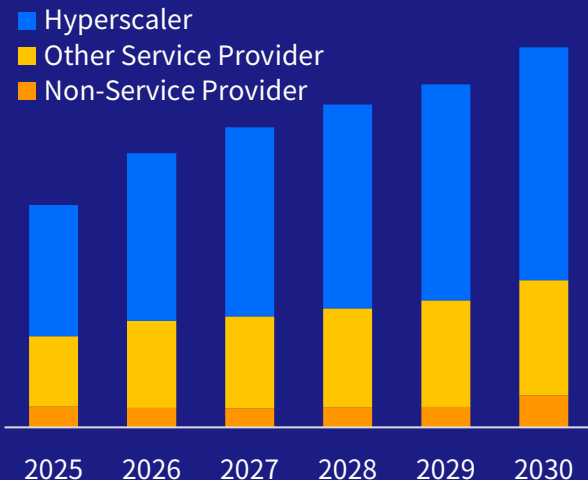
Note: AIサーバーとAIストレージを対象とする。CAGR : Compound Annual Growth Rate (年間平均成長率)



国内AIインフラ市場における3つの主要セグメント

国内AIインフラ市場

主要バイヤータイプ別 支出額予測



AIインフラ投資の実行主体としての3つの主要なセグメント



ハイパースケーラー

グローバル規模の調達力と広範な機能群で汎用的なAI基盤を供給



他のクラウドサービスプロバイダー

規制などへの対応と国内パートナーとの連携によって需要を獲得



エンタープライズ（ユーザー企業）

自社固有のワークロード・データに最適化した専有インフラを構築



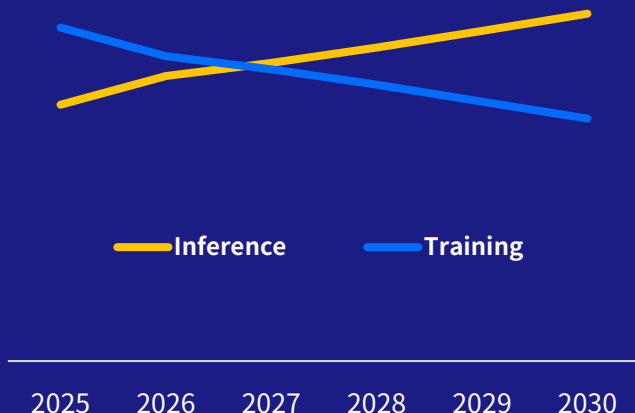
本格的な利用拡大が各セグメントを牽引

Note: Infrastructure Buyerとしてハイパースケーラー：Hyperscaler、他のクラウドサービスプロバイダー（Other Service Provider）：Cloud Service Provider, Communication Services, Digital Service Provider and Managed Services、エンタープライズ（ユーザー企業）：Non-Service Provider



国内AIインフラ市場をリードする推論

国内AIサーバー市場 学習／推論別
支出額構成比予測



◆ 推論向け支出が学習向け支出を上回る

- 2027年：推論向け支出＞学習向け支出
- 2030年：推論向け支出が約6割
- 2025年～2030年の5年CAGR 11.3%

◆ 学習「AIの開発」と推論「AIの利用」

- 「学習」の成長：生成AIモデルの開発基盤整備やサービスプロバイダーの投資による成長
- 「推論」の成長：生成AIチャットボット普及から業務アプリケーションへの実装へ
 - エージェントの展開で利用機会が増加

✓ 競争軸はGPUの規模から推論のデリバリーへ



推論へのシフトによって変わるAIインフラ需要

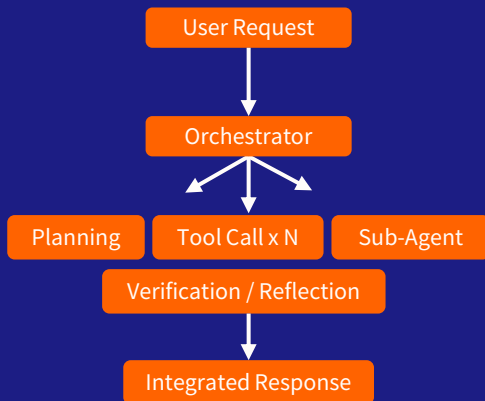
The Compute Multiplier Effect

Traditional AI



1 request = 1 inference call

Agentic AI



1 request = 10x – 100x inference calls

エージェンティックAIによる需要の爆発的増加



“2027年までに、Global 2000企業のエージェント利用は10倍に増加し、トークンやAPIコールの負荷は1,000倍に達する”

Source: IDC FutureScape Worldwide IT Industry 2026 Predictions: Japan Implications (Japan-P2) (IDC #JPJ53133025, December 2025)

トークン経済：モデル選択と経済性

- 2年で1/300：最適なモデルの選択と規模の経済の追求でトークンコストが劇的に下落
- どのようなモデルとインフラで推論インフラを構築するか

エージェンティックAI特化のインフラやモデルの最適解の追求



投資主体によるAIインフラ投資の違い

ハイパースケーラー



- 規模の優位が加速
 - カスタムシリコン採用とアーキテクチャ見直しにより高効率性
 - より大規模なAIインフラへの積極的な投資
- 国内DCだけではないソブリンAIの需要への対応



5年CAGR*: 12%

サービスプロバイダー



- 規模以外の優位性確立
 - 国内AIエコシステム構築による多様な要件対応
 - AI特化「ネオクラウド」による性能面の差別化
(IDC定義: Specialized Cloud Service Providers (SCSPs))
- 国内での設置・運用によるソブリンAI対応



5年CAGR*: 10%

エンタープライズ



- 保有する価値の整理
 - コスト最適化: 需要集約とコストの予測可能性
 - コントロール: データと推論プロセスの制御
- 多様な主権要件への対応によるソブリンAIの実現



5年CAGR*: 9%

Note: Infrastructure Buyerとして ハイパースケーラー: Hyperscaler、サービスプロバイダー: Cloud Service Provider, Communication Services, Digital Service Provider and Managed Services、エンタープライズ: Non-Service Provider

*5年CAGR: 2025年から2030年の5年における国内AIインフラ支出額予測(日本円)の年間平均成長率



AIインフラ導入の課題

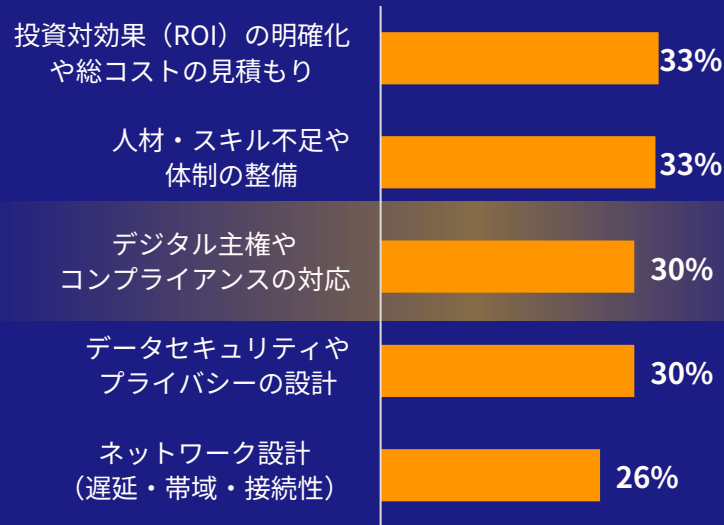


大企業では
**デジタル主権やコンプライアンスの対応は
第3位の課題（30%）**



AIインフラ導入において
ソブリンAIは優先的に検討すべき項目

AIインフラ導入の課題：大企業

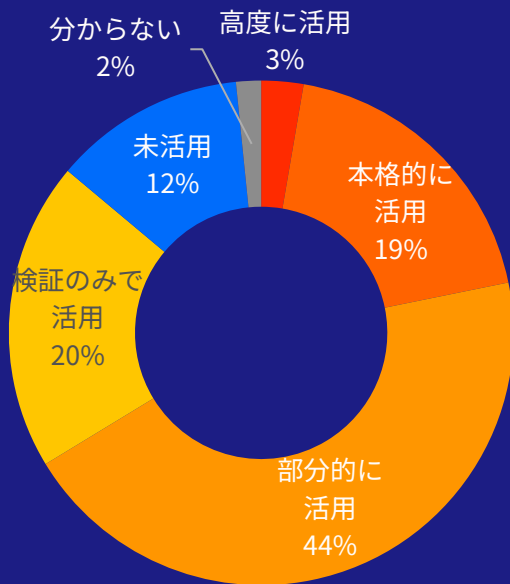


■ 大企業（n = 245）

Note: 複数回答、上位6項目、大企業（従業員数1,000人以上）

組織内データ活用の現実

AI向けの組織内データの活用状況



組織内データのAI本格活用は**22%**にとどまる

- データ基盤やアクセス制御を整備し、明確なルールや承認プロセスに基づいて機密情報や個人情報をAIに活用する企業は限定的

組織内データをAI活用する際の障壁

- データガバナンス 30%
- データ品質管理 30%
- データ構造 28%

データ活用の高度化に伴う課題

- AI／分析基盤との連携
- データアクセス制御やセキュリティ
- データのリアルタイム連携



AI活用の本格化に向けたデータガバナンスやデータ基盤の整備が不可欠



ソブリンAIとは

AIシステムおよびアプリケーション、ならびに
それらが依存する技術の設計、開発、導入、
利用可能性、運用、保守、ガバナンスに関して、
自由な選択とコントロールを行う能力



「すべてに当てはまる唯一の正解」
は存在しない

IDCのデジタル主権フレームワーク

国家、組織、個人によるデジタル自己決定能力

地政学的主権

国家・経済安全保障に立脚したデジタル政策

サプライチェーン主権

シリコンから電力に渡る供給安定性とスキル主権

アシュアランス主権

データの可用性・可搬性とレジリエンスの確保

運用主権

システムの可視性と運用者の特定

技術主権

外部からの干渉・監視に対する耐性

データ主権

データ管理の可視化・規制の遵守



ソブリンAIの浸透



国内企業の**46%**がソブリンAIを重要課題／検討課題としている



2027年までに、J1000企業の**60%**はAI主権の確保を追求し、非パブリックのホスティング、オープン技術、地域パートナーを組み合わせ、ミッションクリティカルなAI活用に対応する。



インフラ支出におけるソブリンAIの現実解

海外のソブリンAI事例

- 各国のインフラ戦略の違いで異なる取り組み
 - フルスタック型：中国は半導体からクラウドにいたる全ての内製化を狙う
 - 連合型：EUのAI大陸行動計画、クラウドやHPCで主権確保、仏政府のインフラベンダー国有化
 - 運用主権型：中東（UAE）は100%ソブリンクラウドを目指す
 - 主権確保における他のアプローチ：GPU前提（韓国）やモデル主権（インド）など

国内のソブリンAIへの取り組み

- 海外大手SPと国内SPの連携
 - 国内SP主権型（例：Oracle Alloy）
 - 構築運用ビジネス型（例：GCP）
 - インフラ資源提供型（例：Microsoft Azure）
- インフラハードウェアの国内提供
 - 半導体開発とHPC/AI国家プロジェクト
 - エアギャップ環境向けAIインフラ提供
 - AIサーバーの国内生産

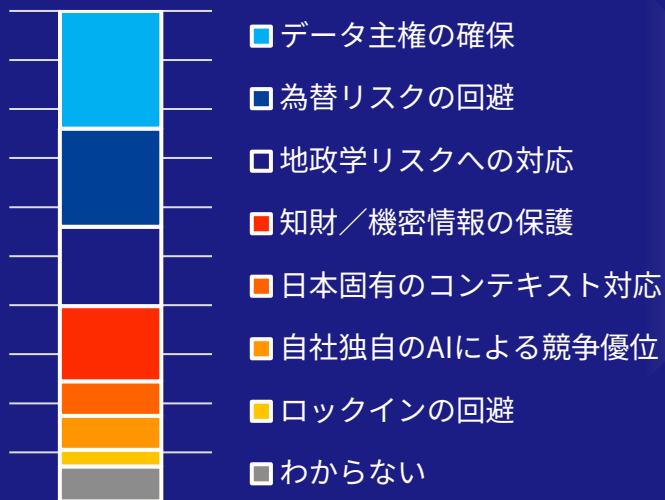
全てにおいて高いレベルのAI主権を確保していくことは非現実的

対象に応じた主権レベルの確保 “セミソブリンAI”



AIインフラのソブリンティ設計における3つの判断軸

Q. 「ソブリンAI」を検討課題や重要課題として
いる場合、「ソブリンAI」の推進目的として最も
重視しているものは何ですか？



Note: 「ソブリンAI」を検討課題や重要課題とした回答者（n = 580）を対象

セミソブリンAI、検討の判断軸

→ データの管轄権

- ・ どこに保存され、どの法律が適用されるか

→ ワークロードのリスク分類

- ・ 機微情報、規制対象、一般業務で統制レベルは全く異なる

→ 経済的持続性

- ・ 守るべき領域に主権を集中
- ・ 領域外はコスト効率を追求



主権の妥協ではなく、
主権をどこに集中させるかを設計する



本日のまとめ



推論の拡大

- ➔ 大規模推論基盤の拡大がAIスーパーサイクルを加速
 - AIインフラは「価値創出の基盤」へと進化
- ➔ 競争軸は「推論基盤のデリバリー」へシフト
- ➔ エージェントAI特化の技術選択や最適解の追求



AIインフラの多様化

- ➔ ハイパースケーラー
 - 規模の優位性と高効率性
 - ソブリンAI対応も進める
- ➔ サービスプロバイダー
 - 国内AIエコシステム
 - AI特化型基盤
 - 国内運用でのソブリンAI対応
- ➔ エンタープライズ
 - コスト最適化と予測可能性
 - 厳格な制御要件



ソブリンAIの浸透

- ➔ AIインフラ投資の優先的な検討項目
- ➔ セミソブリンAIが現実解
 - 対象に応じた主権レベルの確保
- ➔ ソブリンティ設計の判断軸
 - データの管轄権
 - ワークロードのリスク分類
 - 経済的持続性



IDCの提言



推論に最適化された包括的AI基盤オフリングを確立せよ
インフラアーキテクチャから基盤モデル、ライフサイクル支援までを
一体化したオフリングをパートナーと共に整備すべきである。

多様なソブリン要件に応えるAIインフラを構築せよ
可用性、セキュリティ、コストに加え、レイヤー別のソブリン要件に
対応したAIインフラを体系的に提供すべきである。

ハイブリッド戦略の新たな判断軸を捉えよ
AIワークロードの特性、ソブリン要件、供給制約、拡張性、コストと
いった複合的な判断軸に基づき、最適な基盤を提案すべきである。



 IDC
DIRECTIONS