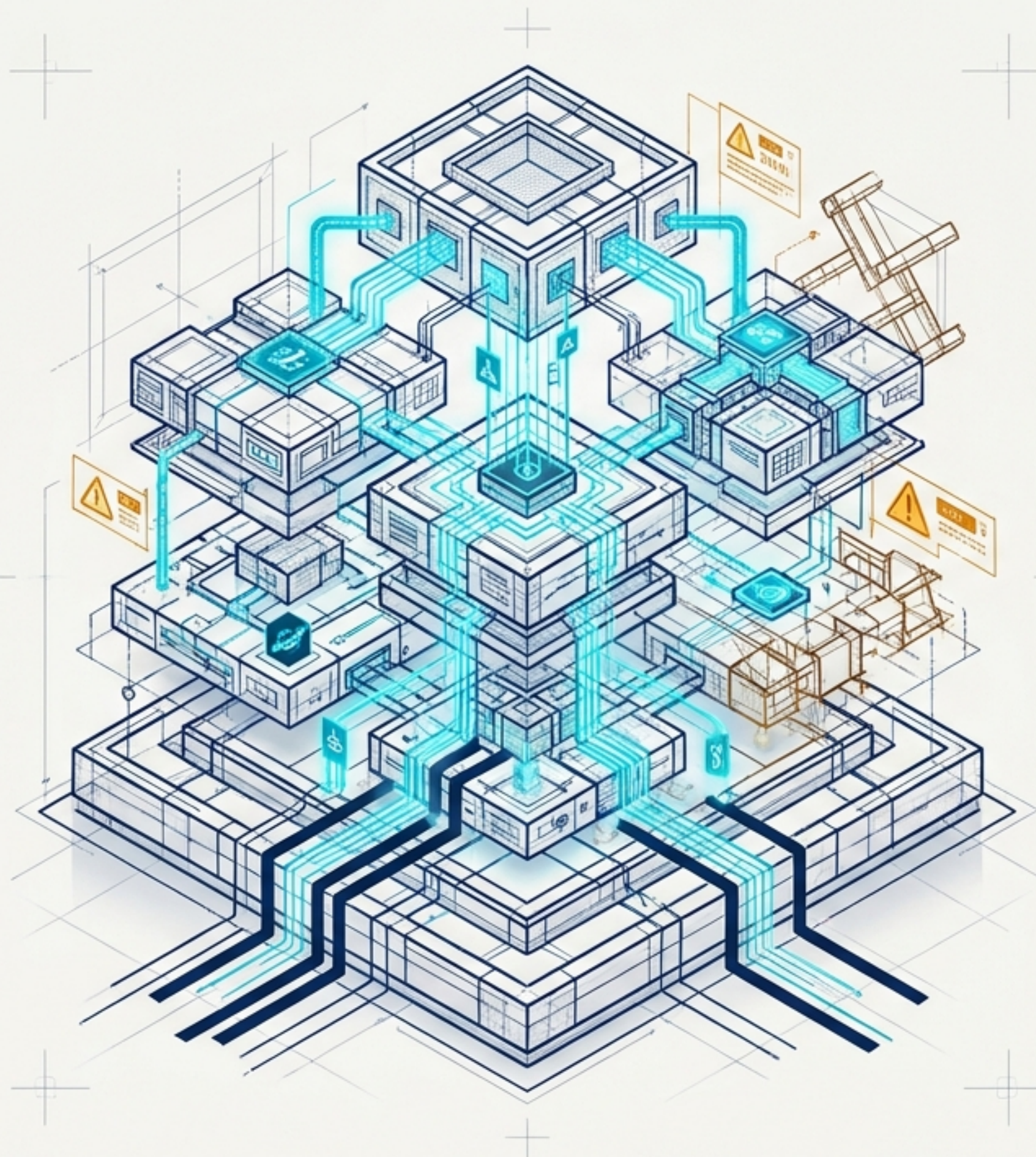




A BLUEPRINT FOR JAPAN'S AI SOVEREIGNTY

巨大AI競争の 終焉と 「主権スタック」 の設計図

Mistral Large 4が日本へ突きつける
真の教訓：規模の競争から、
代替可能性とエージェント実装へ

「1兆パラメータの国産モデル」は、もはやゴールではない

2026年公開のMistral Large 4 (ML4) は、AI開発の競争軸が「パラメータの絶対数」から完全に移行したことを証明した。日本が目指すべきは、単一の巨大国産モデルの開発ではない。

◆ 古い認識 (Old Paradigm) ◆

MONOLITHIC
BLACK BOX MODEL
(1 Trillion Parameters)

WARNING:
HIGH RESOURCE
CONSUMPTION

WARNING: HIGH RESOURCE
CONSUMPTION

OBSOLETE
PARADIGM
DETECTED

RISK: VENDOR LOCK-IN

- 競争軸: 総パラメータ規模による力技
- 目的: 「国産100%」の巨大モデル保有
- 用途: チャットボットと補助業務

◆ 新たな現実 (New Paradigm) ◆

MODULAR MoE BLOCKS

SPARSE EXPERTS

ROUTING LAYER

OPEN WEIGHTS

OPTIMIZED
EFFICIENCY

MODULAR
INTEROPERABILITY

EFFICIENCY > SCALE

EFFICIENCY E NODE

REPLACEABLE NODE

AUTONOMOUS
AGENT

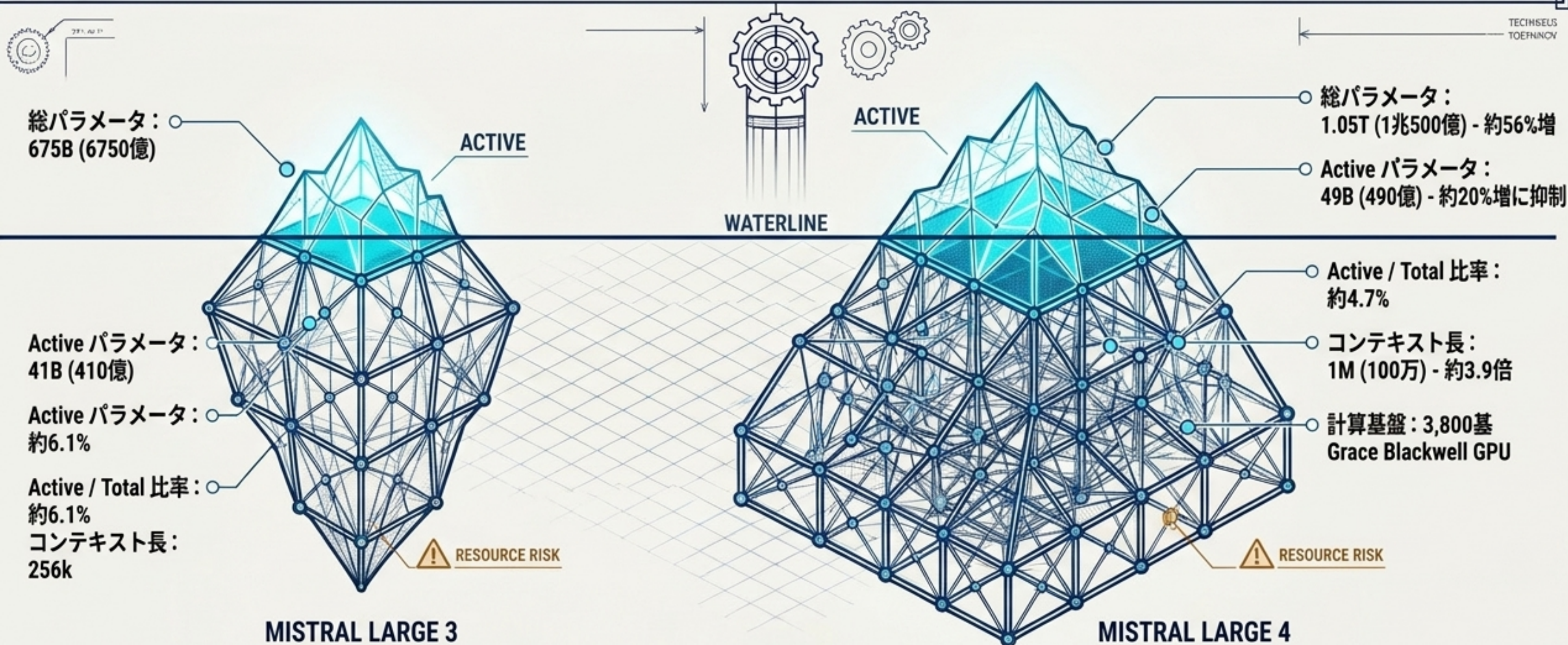
FUTURE-PROOF
ARCHITECTURE

FUTURE-PROOF
ARCHITECTURE

- 競争軸: 疎な計算効率 (MoE) とオープンウェイト
- 目的: モデルの「退出可能性 (代替可能性)」の確保
- 用途: 複数ステップを自律実行する業務エージェント

氷山の水面下：アーキテクチャの本質は「490億のActive」

フロンティア競争は「パラメータ数」から
「巨大な知識容量を、いかに少ない計算量とメモリで駆動させるか」へ移行した。



グローバル・フロンティアの三極構造と日本の現在地

	アーキテクチャと効率	公開形態と主権	リスクと課題
米国 (GPT-6 Astra)	クローズド。極めて高い総合性能。	プロプライエタリAPI（提供者による中央管理）。	プロバイダのポリシー変更・停止リスクへの依存。
中国 (DeepSeek V4 Pro)	1.6T/49B。アルゴリズム効率（Attention/KV Cache削減）の追求。	高効率オープンウェイト。	地政学的・データガバナンス要件の精査が必要。
欧州 (Mistral Large 4)	1.05T/49B。疎なMoE（Mixture-of-Experts）。1.6B Vision Encoder。	欧州データセンターでのSovereign運用＋オープンウェイト。	ライセンス実運用要件や日本語評価の未確定。



レジリエンスの最適解:

いずれか一つに依存するのではなく「国内モデル+欧州/中国オープン+米国API」を常時切り替え可能なマルチモデル・アーキテクチャこそが日本の目指す姿である。

AI主権の再定義：7層の「モジュール化された主権スタック」

AI主権とは「国籍」ではない。「いつでも他国モデルから退出・移行できる統制能力 (Exit-ability)」である。

Layer 6: 安全保障・法制度
(個人情報・輸出管理・デプロイヤー責任)

Layer 5: データ・競逐検証
(日本・電力・度致した規格・を確認)

Layer 3: モデル・レイヤー
(国産モデル開発 + 池端等の国内実行)

Layer 7: 実装・オーケストレーション
(モデルを自由に切替可能な安全な産業・行政実装)

Layer 5: 評価・検証
(日本独自の評価基盤・AISI/JAI-Trust)

Layer 4: データ・チューニング
(日本語・産業データによるPost-training / RAG)

Layer 3: モデル・レイヤー
(国産モデル開発 + 海外Open-Weightの国内実行)

Layer 2: 計算基盤
(国内Compute・AIスパコン)

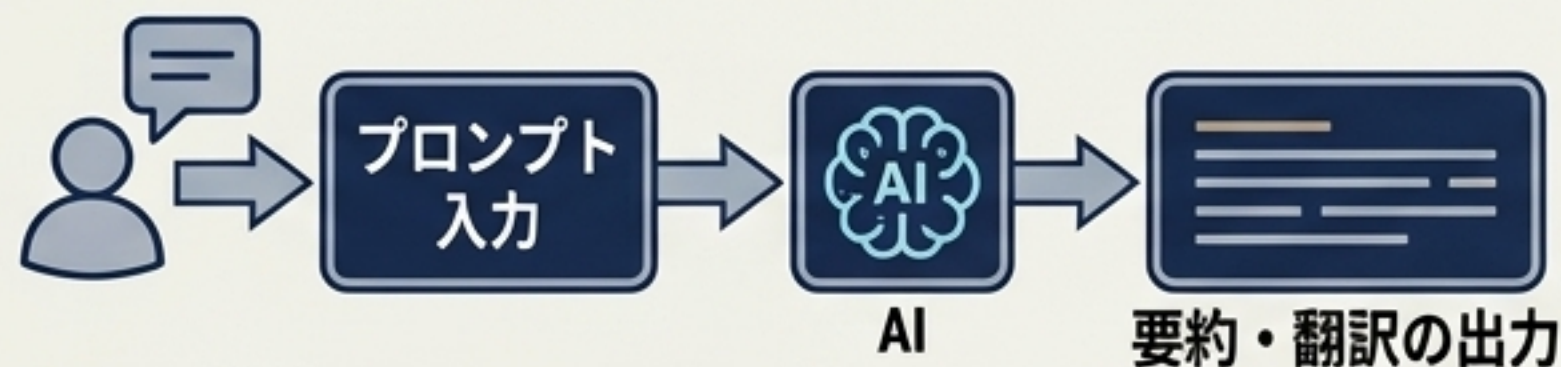
Layer 1: インフラ・資源
(半導体・電力・通信網のTCO確保)



産業実装のシフト：「AI利用率」から「業務変革率」へ

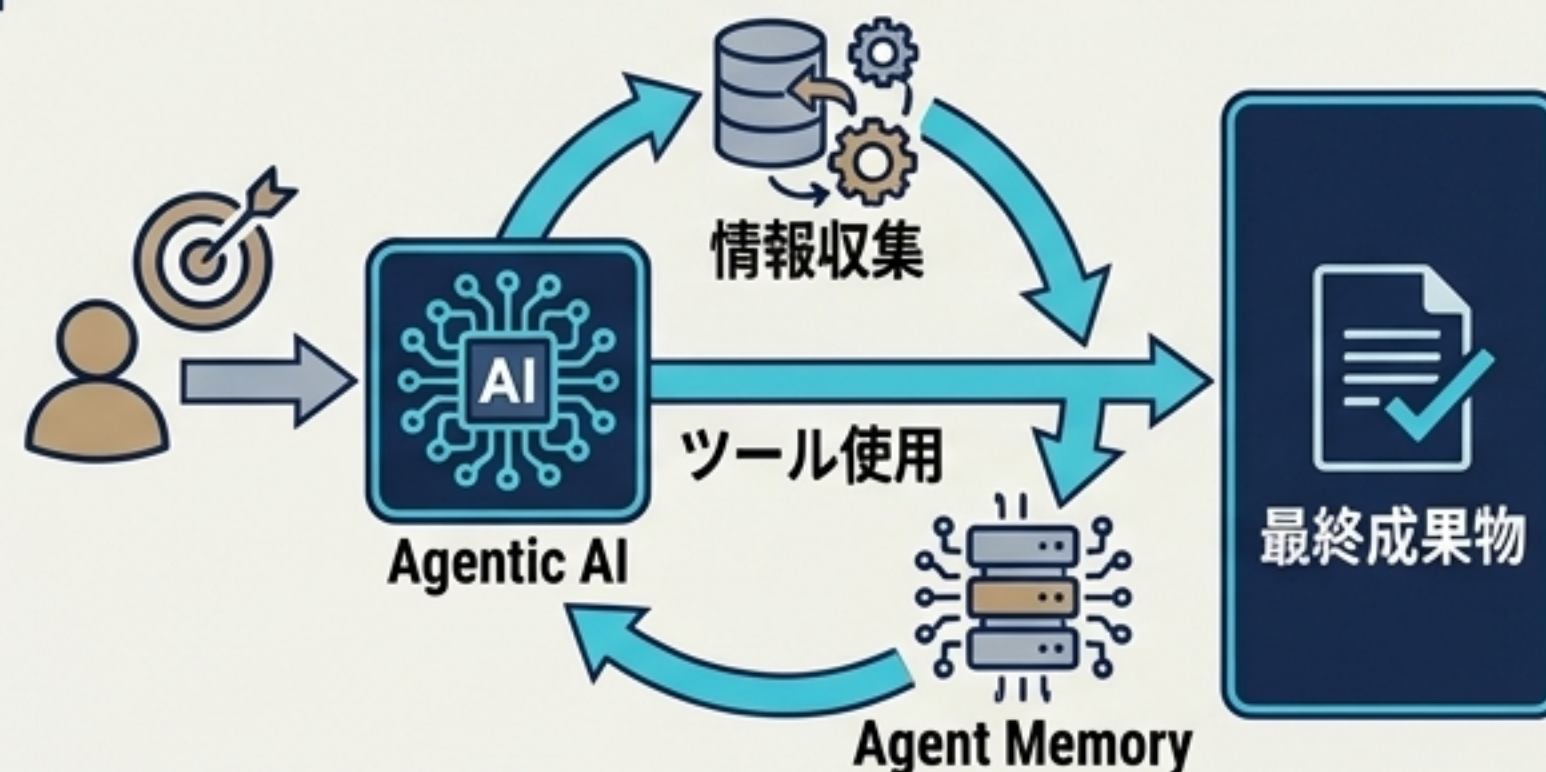
- 企業でのAI導入・試行は58.0%に達するが、その用途の80%以上が要約・翻訳・文書作成などの「補助業務」に偏っている。（IPA「DX動向2026」より）

現状のチャットボット利用



- 競争軸：「チャットの賢さ」
- フロー：単発のプロンプト入力 → 要約・翻訳の出力
- 影響：個人の生産性向上（微増）

自律型業務エージェント



- 競争軸：「一連の仕事を完遂する能力」（AutomationBench重視）
- フロー：ツール使用 → 情報収集 → 複数ステップの自律実行（Agent Memory） → 最終成果物
- 影響：受発注、設計審査、コンプライアンス等、プロセス全体の再設計

製造・産業資本との融合：日本最大のオポチュニティ

一般Webコーパスでは代替できない日本の産業資産（設計図、品質管理記録）と、1Mコンテキスト・視覚処理能力の結合。



フロンティア・オープンウェイト（例: ML4）のコア能力

-  **1M Context:** 長大なマニュアル、複数資料の横断理解
-  **1.6B Vision Encoder:** 技術図面、複雑なPDF、地理空間画像、部品画像の理解
-  **Self-Hosting (自己ホスト):** 機密データ（CADや製造ノウハウ）を国外APIに送信せず、国内サーバー/オンプレミスで運用可能

アクション:

「汎用AIをそのまま使う」ことから、国内データでファインチューニングした「製造・金融・法務特化のVertical AI」へ投資を集中させる。

Open Weightsのジレンマ：安全保障におけるデュアルユース性

資産 (Asset) としてのOpen Weight

自己ホストと防衛力:
機密情報を外部に出さず、
防御側が独自にマルウェア
解析や検知ルール
作成 (Fine-tune) を自
己運用可能。
プロバイダー依存からの
脱却。

ML4のサイバー能力:
Cybench 93% / 脆弱性再現・Patch試験 82%
(Mistral公表最高値)



脅威 (Threat) としてのOpen Weight

高度な攻撃能力の拡散:
APIのようなアクセス制限
(監視・停止) が不可能。
セーフガードを外した
独自の派生モデル (Air-
gapped deployment) に
よるサイバー攻撃や脆
弱性悪用のリスク。

「公開か非公開か」の二元論は終わった。能力レベルに応じた「重み (Weight)」
のガバナンス、レッドチーム基準、輸出管理指針のアップデートが急務。

信頼性の再定義：「ハルシネーション・ゼロ」から「検証可能性」へ

AIエージェントが自律的に検索から生成までを行う時代において、真に重要なのは文章の流暢さではなく、情報の出所を遡れる「Provenance（来歴）」である。



JAI-Trustの戦略的価値

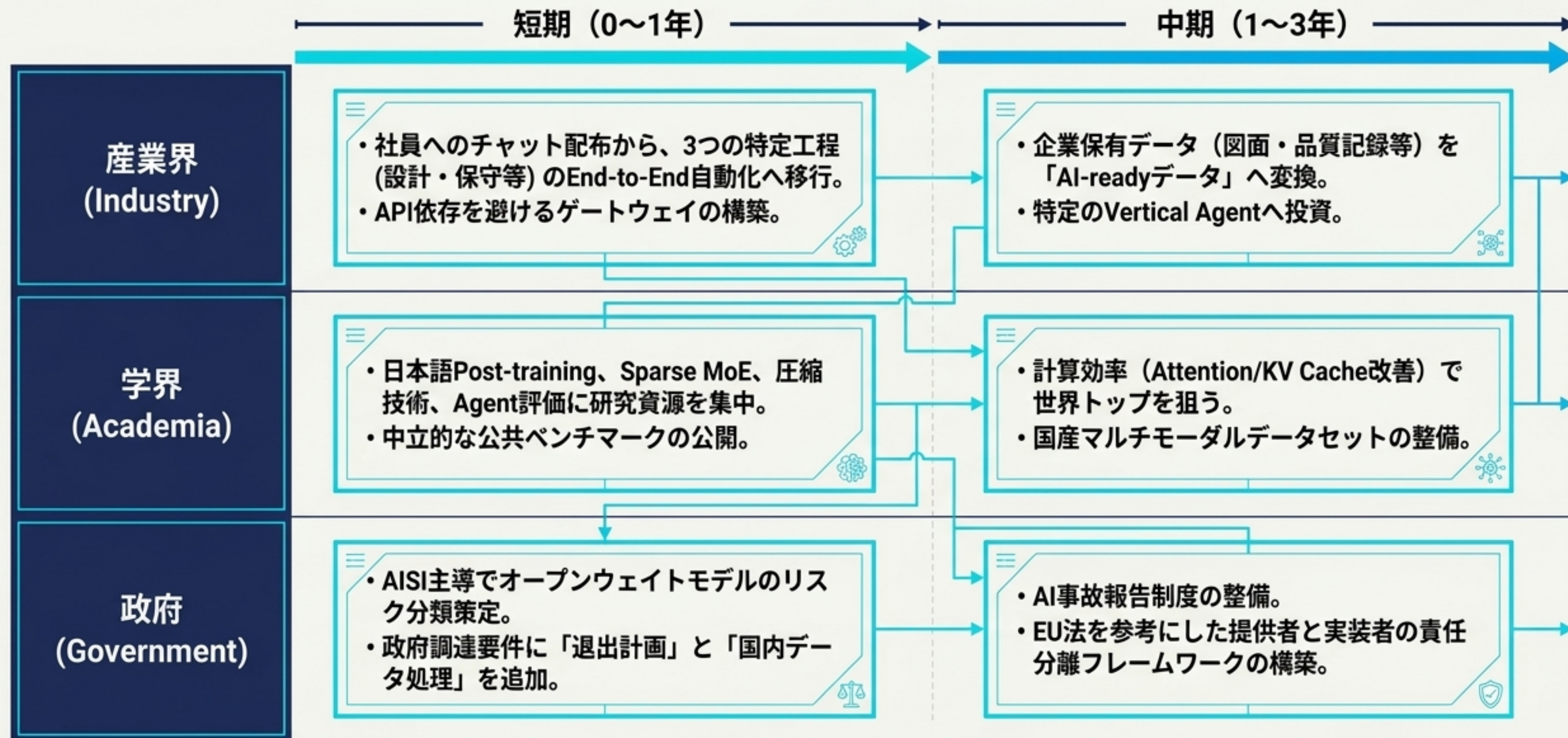
英語のベンチマークで安全でも、日本社会で安全とは限らない。バイアス、偽情報、情報漏洩を独立して定量評価する「日本独自の評価基盤（JAI-Trust）」が、多言語オープンモデル時代には不可欠。

ガバナンスの構造：モデル提供者と実装者の責任分離

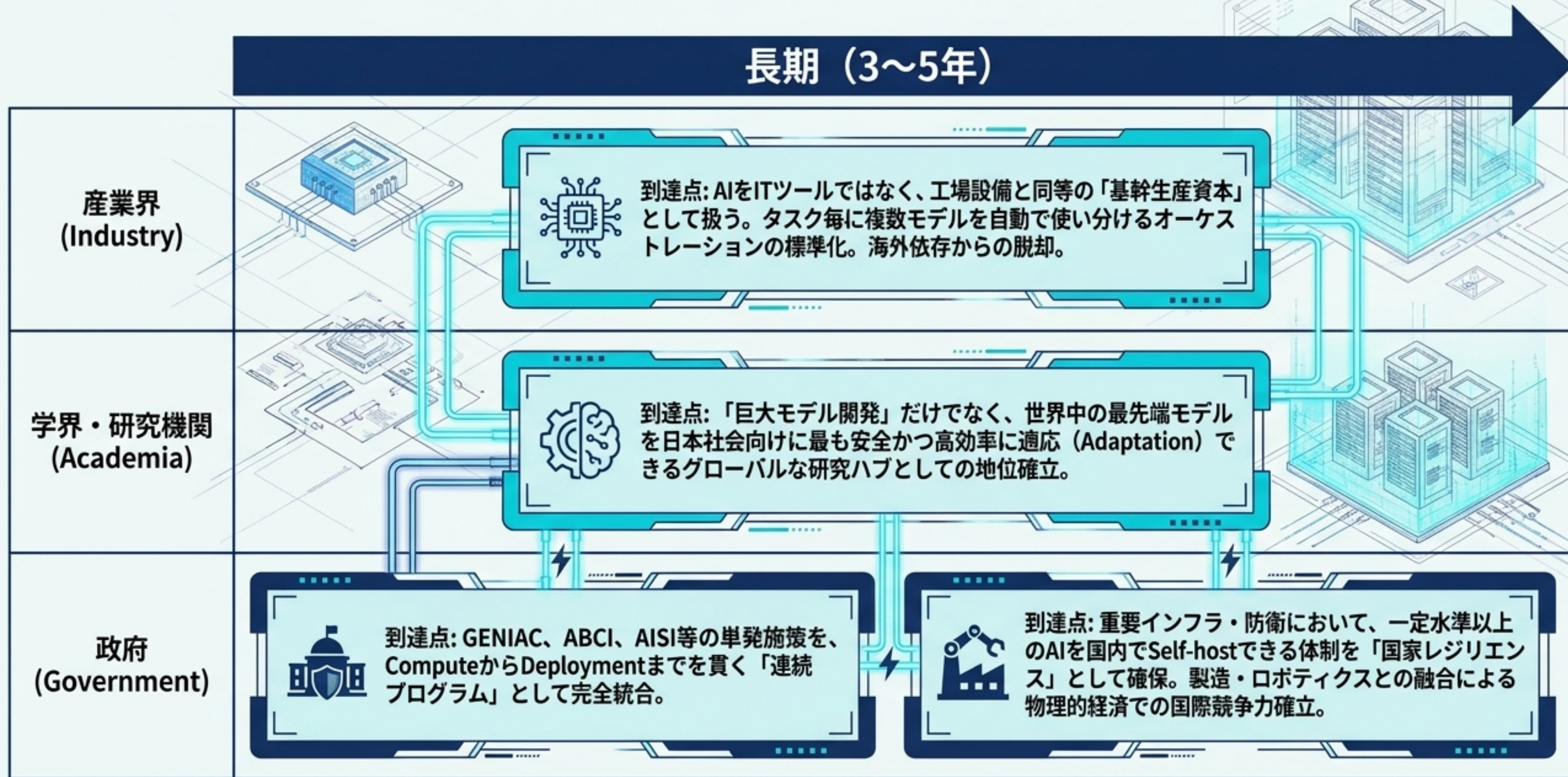
EU AI Act（2025-2026年適用開始）の最大の教訓は、基盤技術を作る者と、それを社会の特定用途（医療・金融など）に組み込む者の責任を明確に分離したことにある。



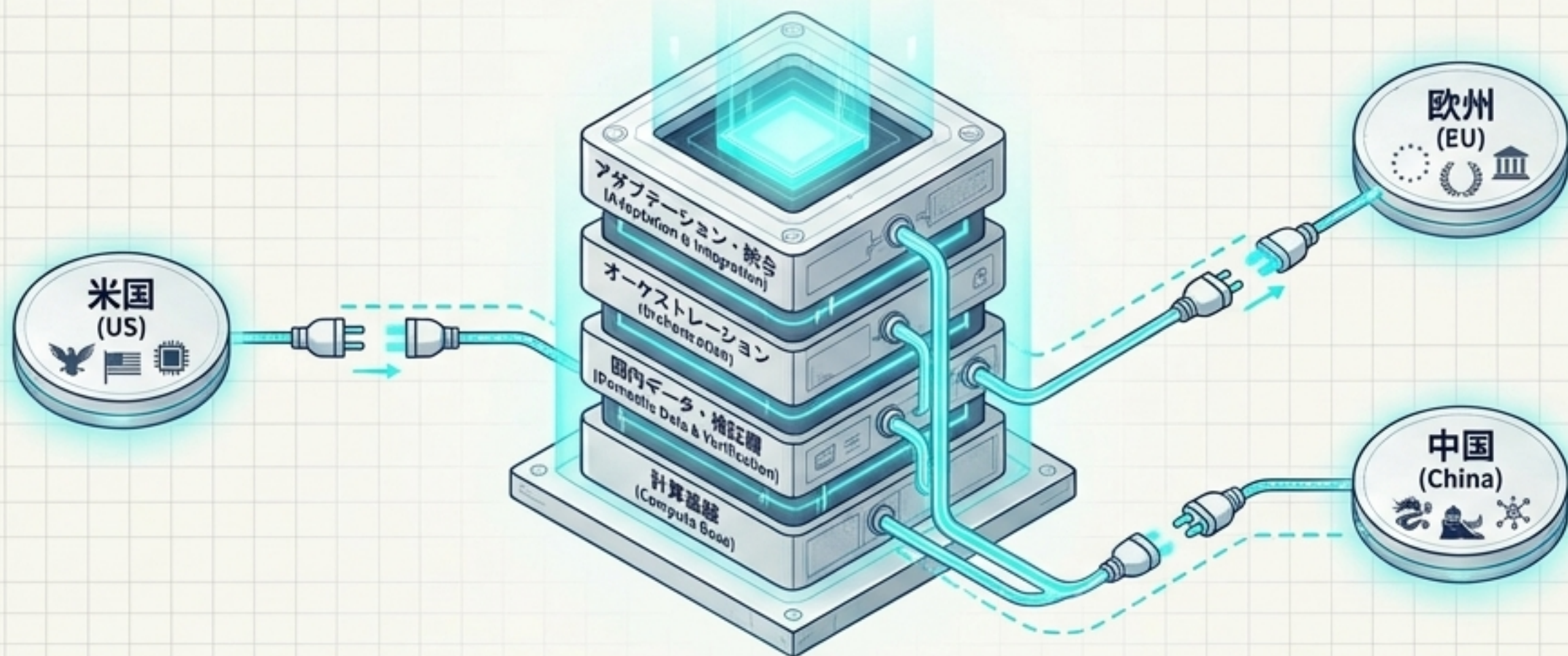
Sovereign Roadmap Part 1: 実装と評価の立ち上げ（0～3年）



Sovereign Roadmap Part 2: 戦略的自律とインフラ統合 (3~5年)



結論：真の「AI主権」とは退出する力である



Mistral Large 4が証明したのは、計算資源のアルゴリズム的節約（疎なスケーリング）とオープンウェイトの戦略的組み合わせにより、米中と同じ資本規模がなくとも独自の競争空間を作り出せるという事実である。

日本が目指すべきは「世界最大のモデルを持つ国」ではない。
「世界のどのモデルが登場しても、それを自国で検証・改変し、国内データで価値を加え、必要であればいつでも別のモデルへ乗り換えられるレジリエンスを備えた国」
それこそが、新しいAIアーキテクチャにおける日本の最終的な青写真である。