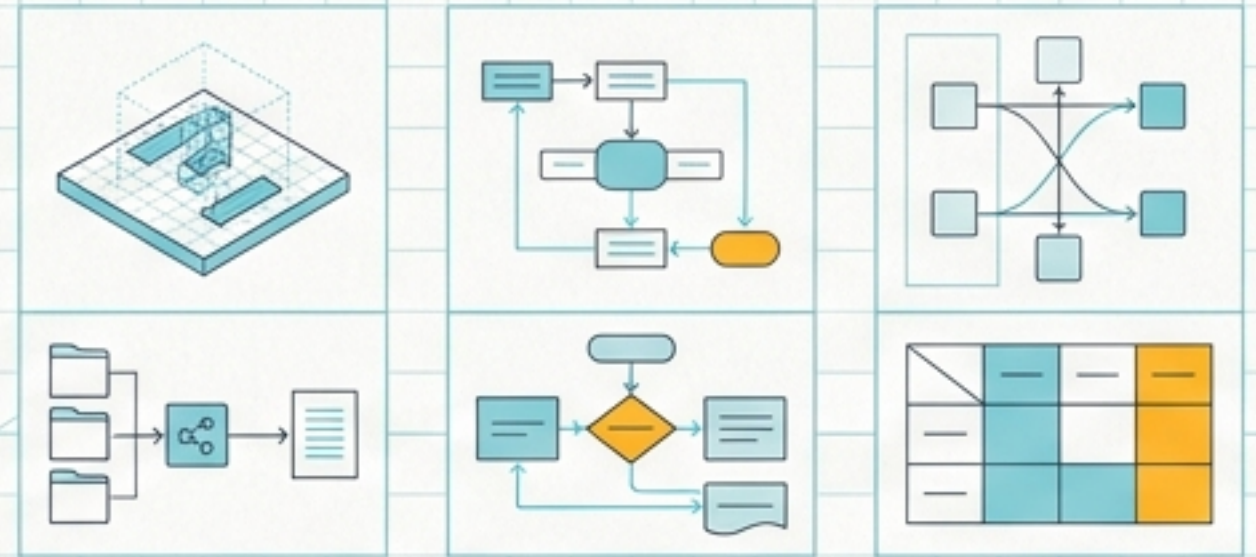




Aleph Alpha 「Kolibri」 の戦略的解剖

欧州発・主権AIモデルの技術仕様、規制対応と
日本企業向け導入プレイブック

対象読者: 企業CIO、AI戦略担当者、ITアーキテクト

エグゼクティブ・サマリー



基本仕様 (Technical Reality)

78.1Bパラメータ総量に対し、稼働は3.46B（約4.4%）。高並列・低計算コストを実現するMoEアーキテクチャ。



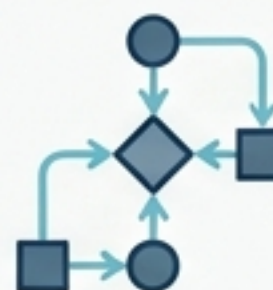
日本市場での現実 (The Japanese Reality)

「国産LLMの代替品」ではない。対象言語は英・独のみ。日本語業務への直接投入は非推奨。



真の戦略的価値 (The Real Value)

「主権AI」の青写真。EU AI Actを意識した透明性、学習データ管理、オンプレミスでの完全制御を実現。



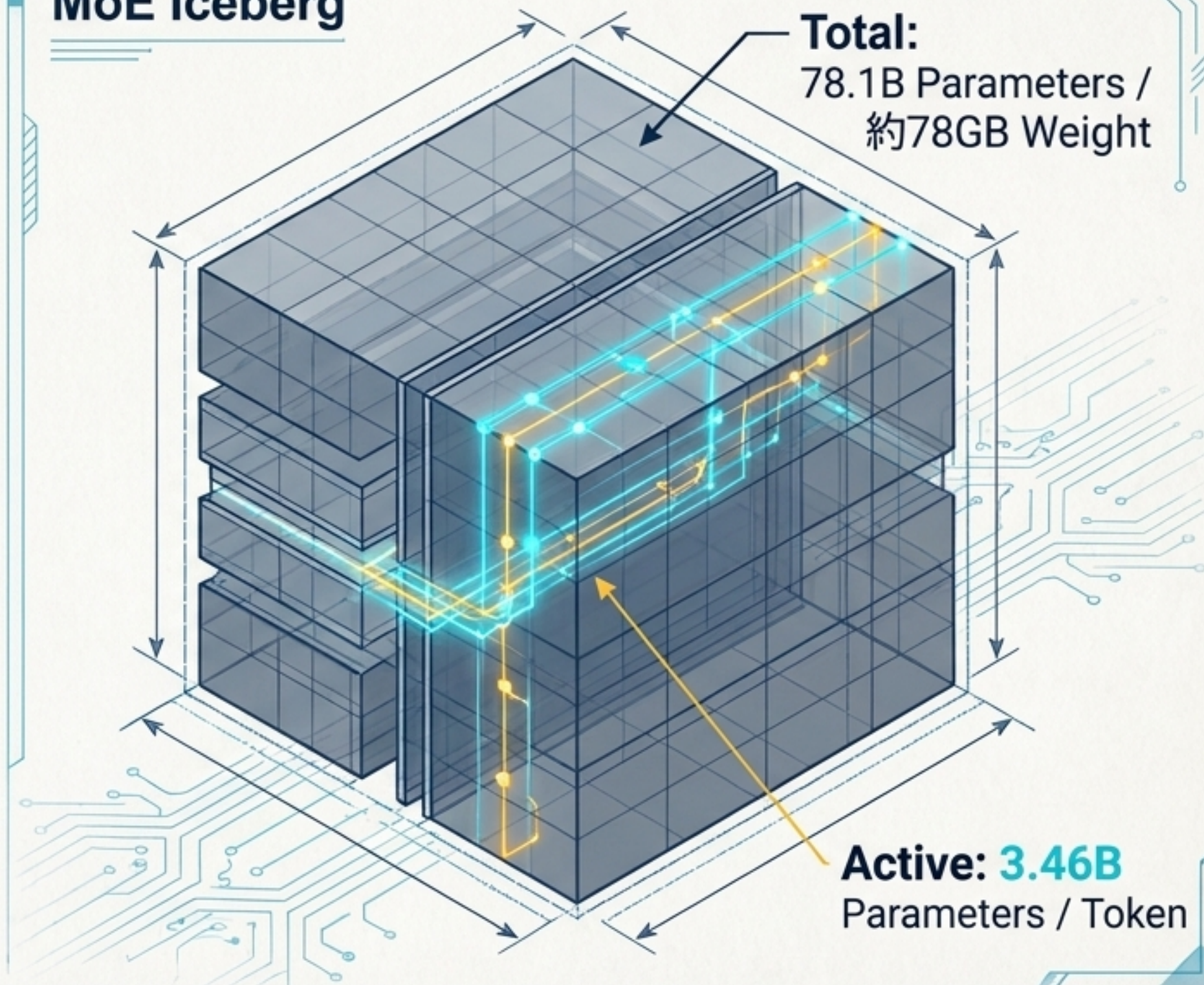
即時アクション (Immediate Action)

日系企業のEU拠点における限定的PoC、および「複数モデル・ルーティング」アーキテクチャへの移行準備。

技術仕様: Mixture-of-Experts

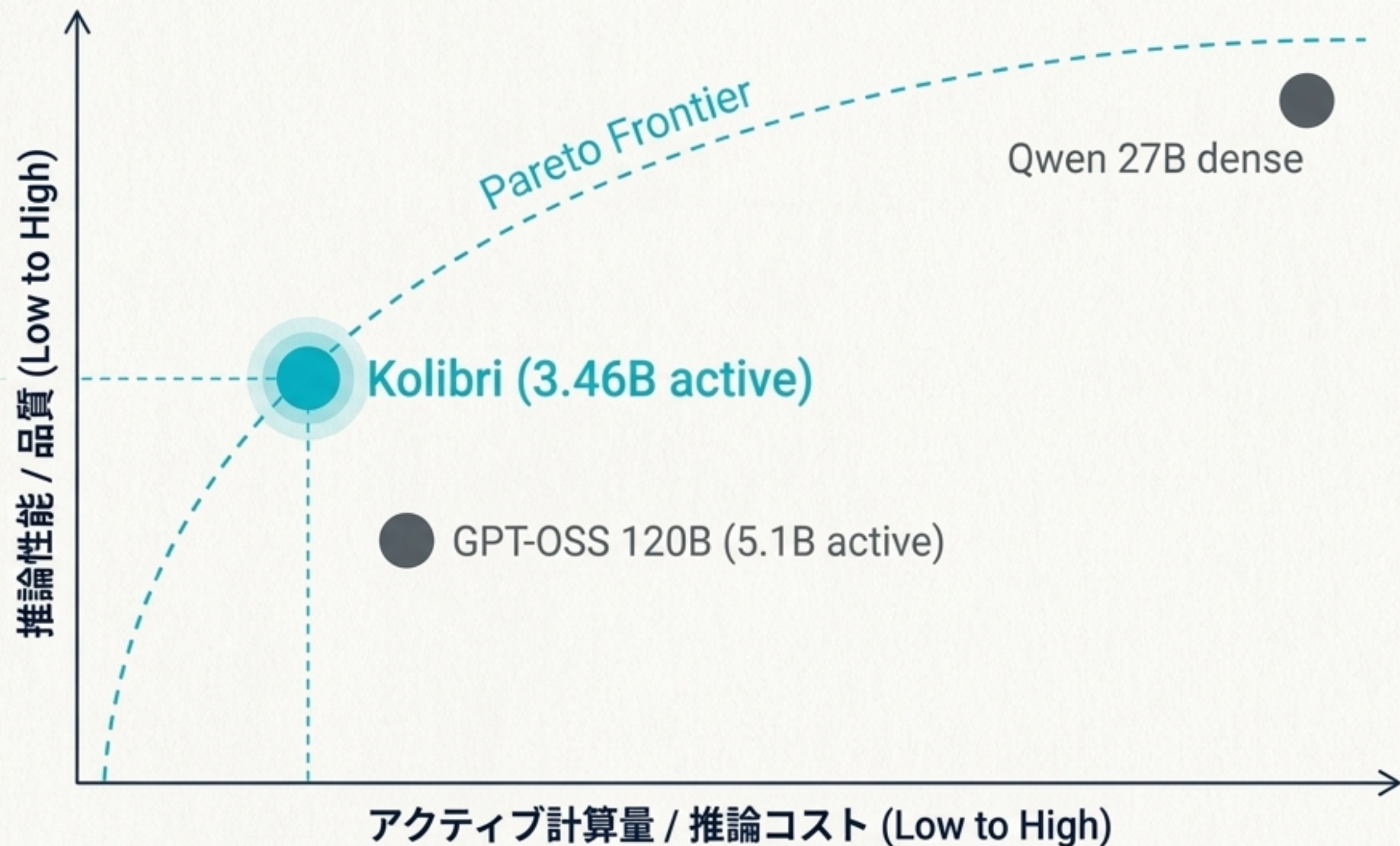
1 ● アーキテクチャ:	Decoder-only causal Transformer, sparse MoE
2 ● レイヤー構成:	50層 (40 SWA + 10 Full attention)
3 ● Expert分割:	● 384 routed experts ● + 1 shared expert (毎トークン上位6個を利用)
4 ● コンテキスト:	● 最大1,048,576 tokens (推奨 262,144)

MoE Iceberg



戦略的示唆: 推論コストは3Bクラスだが、運用には78GBのメモリ空間を要する。
単一ユーザーの応答速度ではなく、高並列サーバー効率 (throughput) に特化した設計。

AIME 2025 (英語): 96.9%
—
Industry RAG: 未見コーパスで比較MoE群を圧倒



戦略的示唆: Kolibriは『絶対的性能の世界一』を狙ったものではない。
許容可能なTCO内で最高水準の推論を引き出す最適解の提示である。

日本語能力の「未指定」リスク

公式対象言語は「ドイツ語・英語」のみ。128k語彙による日本語処理効率や、日本語学習データの厳密な比率は公式に「未指定」。

国内業務投入前の必須A/Bテスト項目

- ☐ 敬語・ビジネス日本語の適切性
- ☐ 日本語法務・契約用語の理解度
- ☐ 長文RAGにおける日本語特有の幻覚（Hallucination）率
- ☐ 日英独クロスリンガル検索の精度
- ☐ 処理1トークンあたりの日本語GPUコスト

結論: 偶然の日本語生成能力に依存すべきではない。国内業務への投入前には、TCOとエラー修正工数を含めた厳格なA/Bテストが不可欠。

エンタープライズ主権AI (Enterprise AI Sovereignty)

モデルウェイト

Apache 2.0による推論環境の自己保有



データ来歴

EU様式に準拠した学習データの透明性



インフラ制御

オンプレミス／エアギャップでのデプロイ



監査可能性

詳細なモデルカードとベンチマーク手法の開示



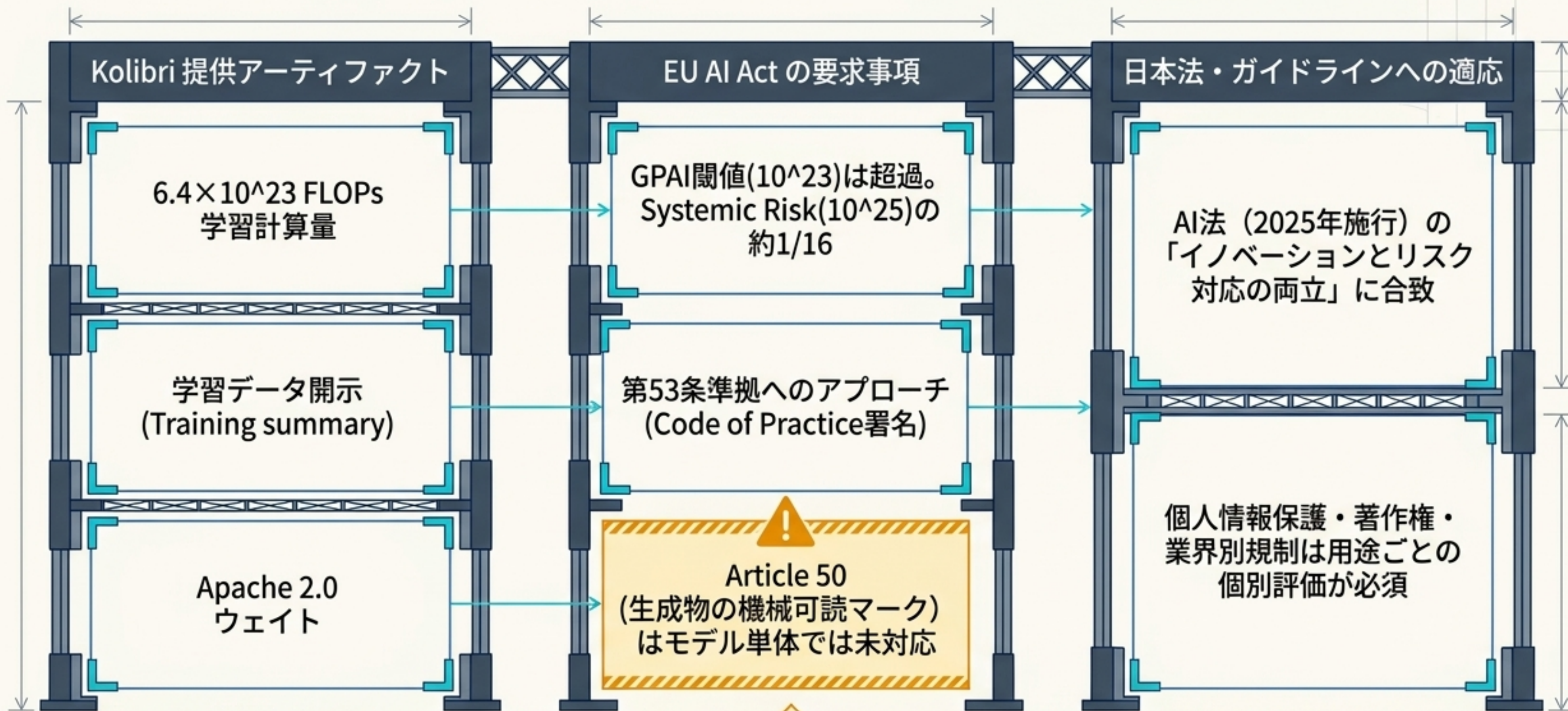
ガバナンス

バージョン・ログ・推論パラメータの自社管理



オープンソースを超える、企業向けAIの参照アーキテクチャ

規制アライメント: EU AI Actと日本



※Kolibriは「AI Act認証済み」ではなく、法令対応を前提に設計された「Compliant-Ready」モデルである。

エンタープライズ・セキュリティ要件

Kolibriの組み込み機能 (Built-in)



学習時におけるPII（個人情報）の
Redaction（匿名化）



特定の政治的バイアスに対する
専用フィルター



中央モデレーションに依存しない
ネットワーク隔離運用

企業側での必須実装 (Required Guardrails)



機密情報スキャナーおよびロール
ベースアクセス制御（RBAC）



Hallucination対策（必須のRAG
連携とソース検証）

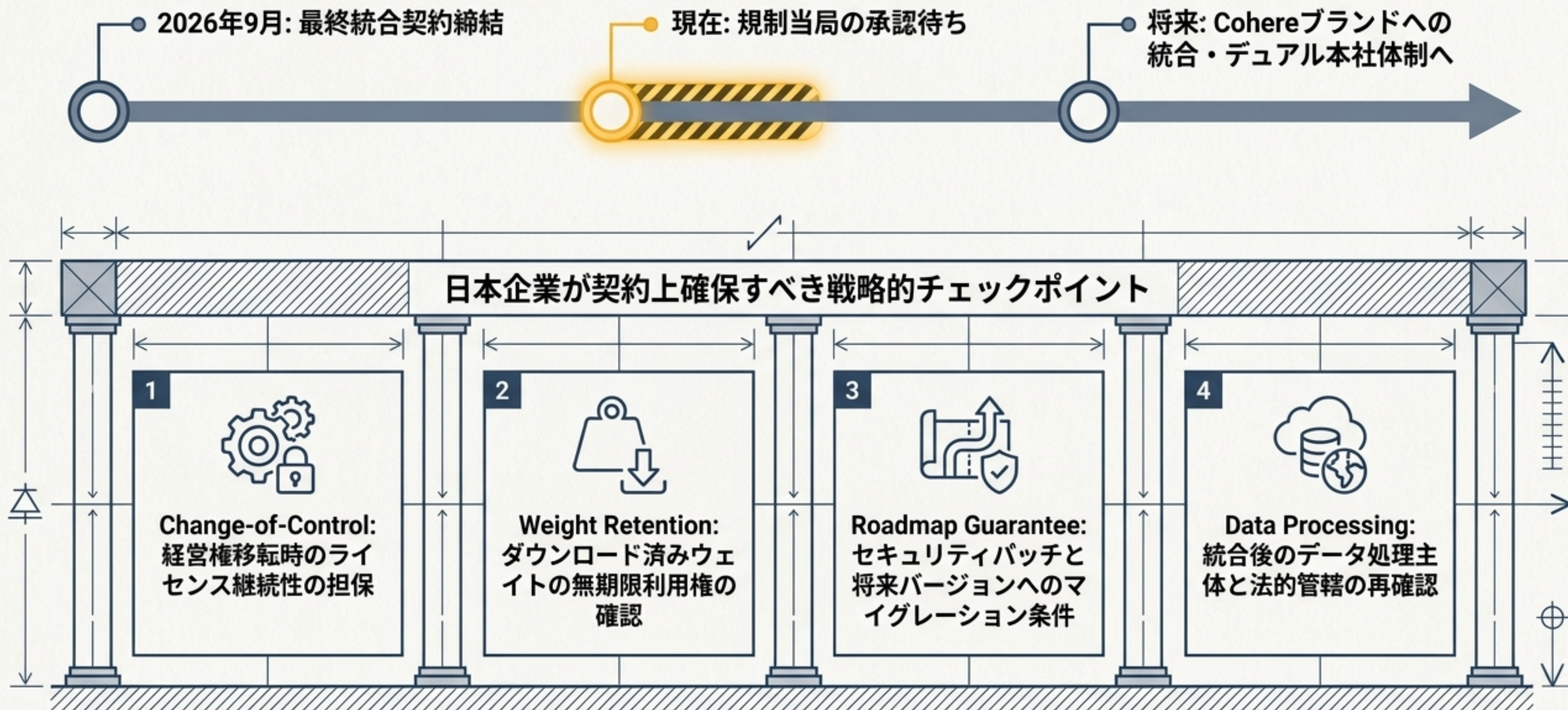


高リスク判断（医療・金融等）にお
けるHuman-in-the-loop（二重承認）

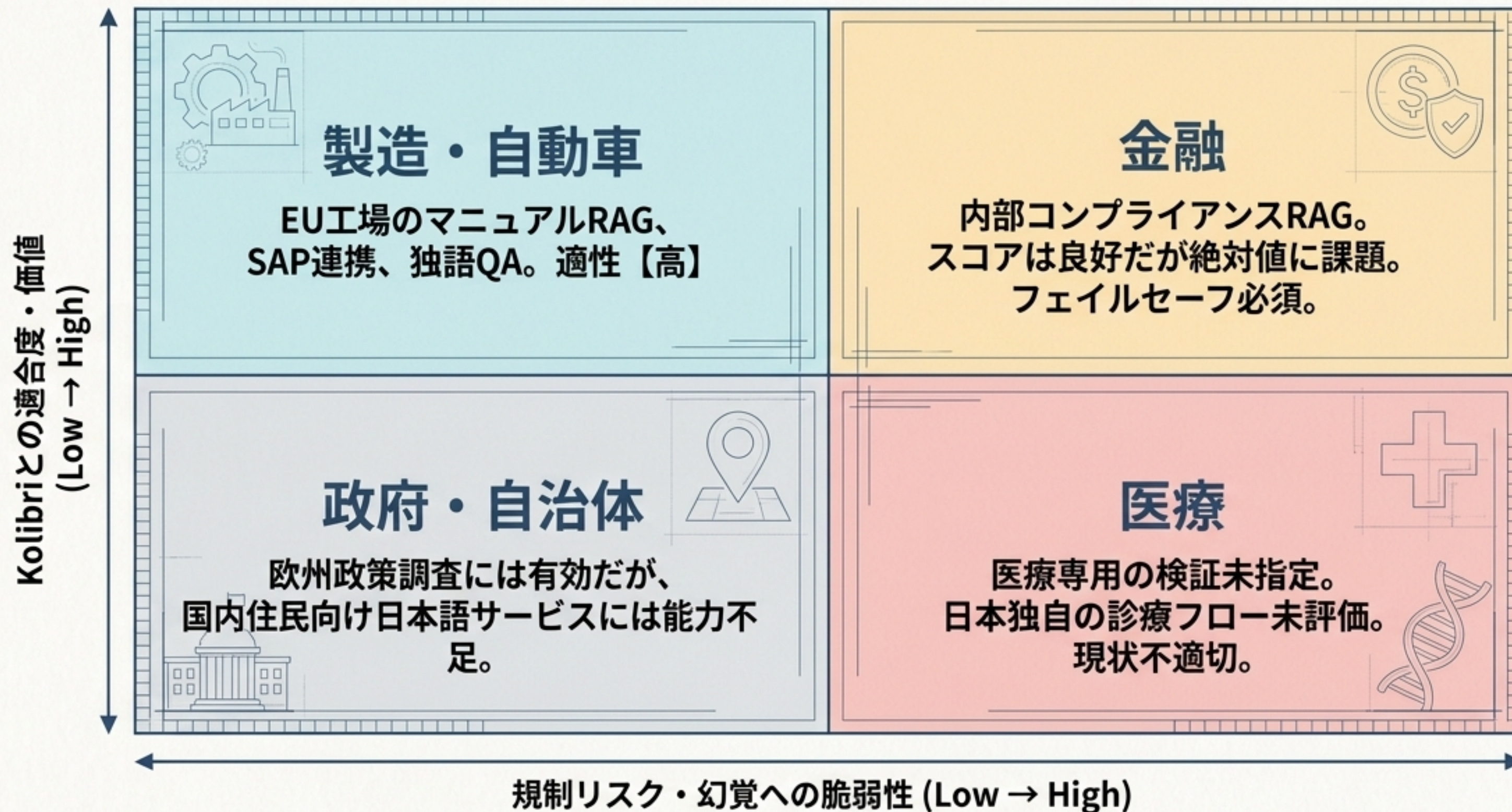


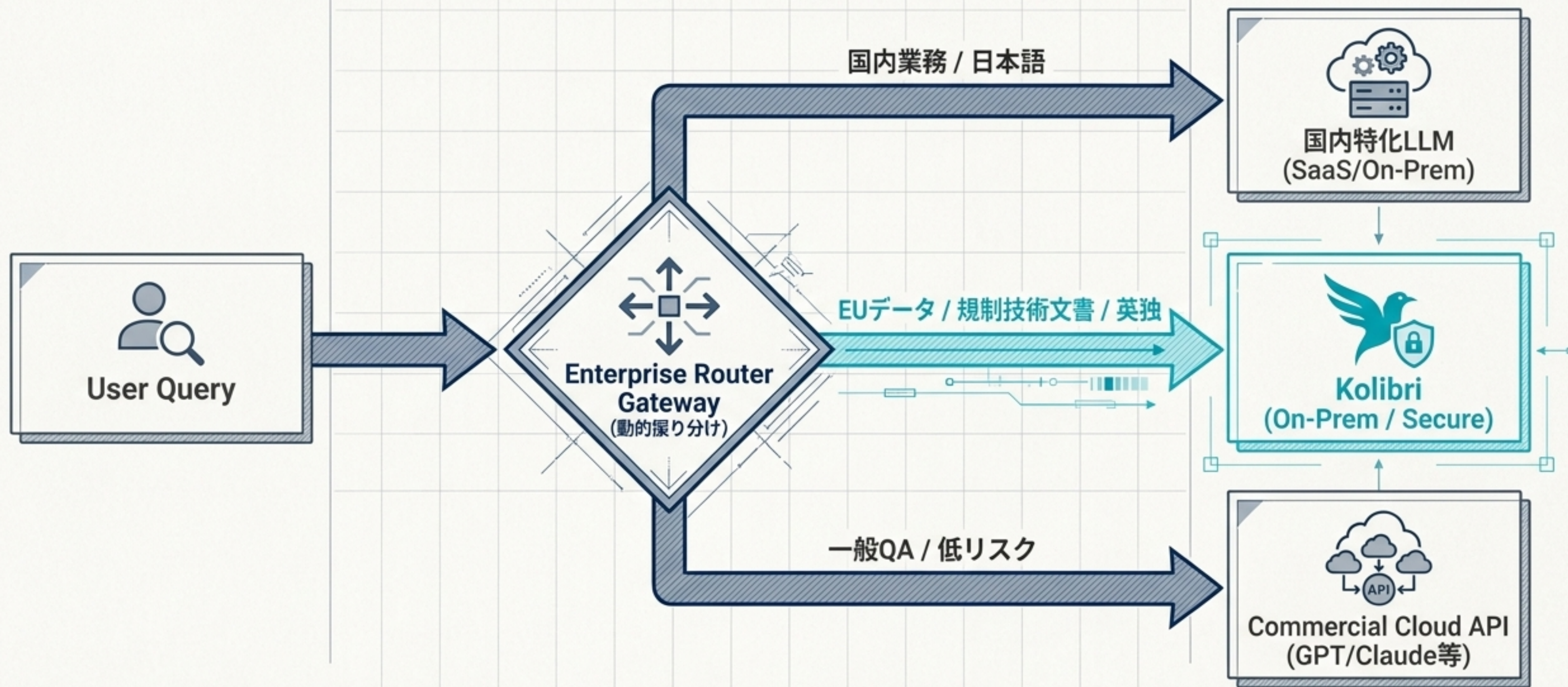
日本語および日本社会属性に基づ
く独自のバイアステスト

戦略的ベンダーリスク: Cohereとの統合



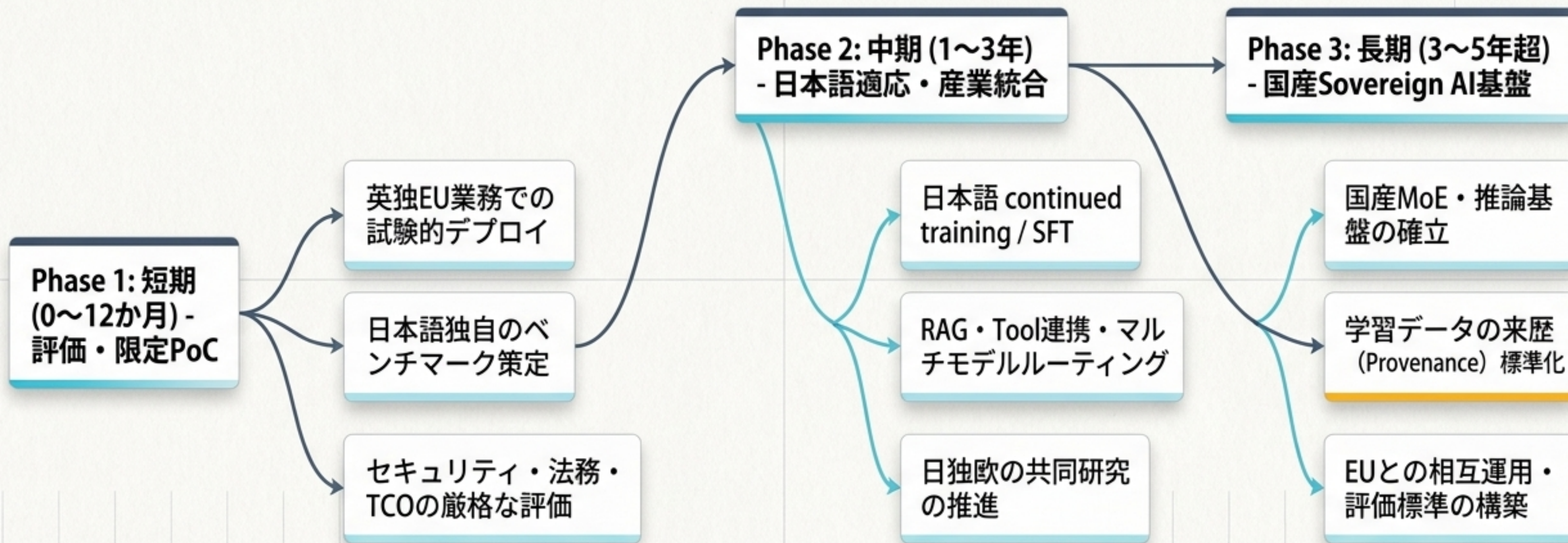
産業別インパクトと導入適合度





ターゲット・アーキテクチャ: 単一モデル主義からの脱却。言語・リスク・コスト・推論難易度に応じてクエリを動的に振り分けるマルチモデル・ルーティング。

日本企業向け導入プレイブック (戦略ロードマップ)



最終戦略提言 (Strategic Recommendations)

1

欧州拠点からの 段階的テスト

国内の日本語本番環境への即時投入は避けよ。まずは欧州拠点の英独RAGや内部ナレッジ管理に限定し、既存APIとのTCO（GPU稼働率・遅延等）を比較せよ。

2

モデル非依存インフラへの投資

特定のベンダーに依存せず、セキュアなRAG、データエンジニアリング、モデルルーティング基盤への投資を優先せよ。Cohere統合によるロードマップ変更への最高のリスクヘッジとなる。

3

次世代AI調達の「基準」化

Kolibriを単なる製品ではなく、日本が構築すべき主権AIの「参照アーキテクチャ」として扱え。透明性、データ来歴の開示レベルを、今後の公共調達・導入の必須要件にせよ。