Qwen3.8-Max 徹底調査レポート

米中AIギャップの縮小と日本企業への戦略的含意

調査基準日：2026年8月4日 ｜ 対象読者：経営層・CIO・法務コンプライアンス担当

Executive Summary: Qwen3.8-Maxがもたらす「破壊的恩恵」と「構造的リスク」

光: 破壊的恩恵

第2集団への肉薄

2.4兆パラメータ。
2.4兆パラメータ。コーディング・マルチモーダル領域で米国フロンティア勢（Claude Fable 5, GPT-5.6 Sol）に匹敵。

圧倒的な価格破壊

100万トークンあたり
入力$2 / 出力$6。
米国勢の約3〜4分の1の低コスト。

影: 構造的リスク

名ばかりのオープン化

2.4兆パラメータは
コーディング・集局アクリスと、
2.4兆パラメータは巨大すぎて
自己ホスト実質不可。
ライセンス条文も未公開。

レッドライン（調達の壁）

中国国家情報法（第7条・14条）のリスクに加え、米国防総省1260Hリスト（2026年6月更新）による調達制限対象。

Model Architecture: 2.4兆パラメータの「スパースMoE」構造解剖

総パラメータ数

2.4兆

アクティブパラメータ

約950億/トークン

推論モード

thinking常時有効
(デフォルトxhigh)

コンテキスト長

100万トークン

アーキテクチャ基盤

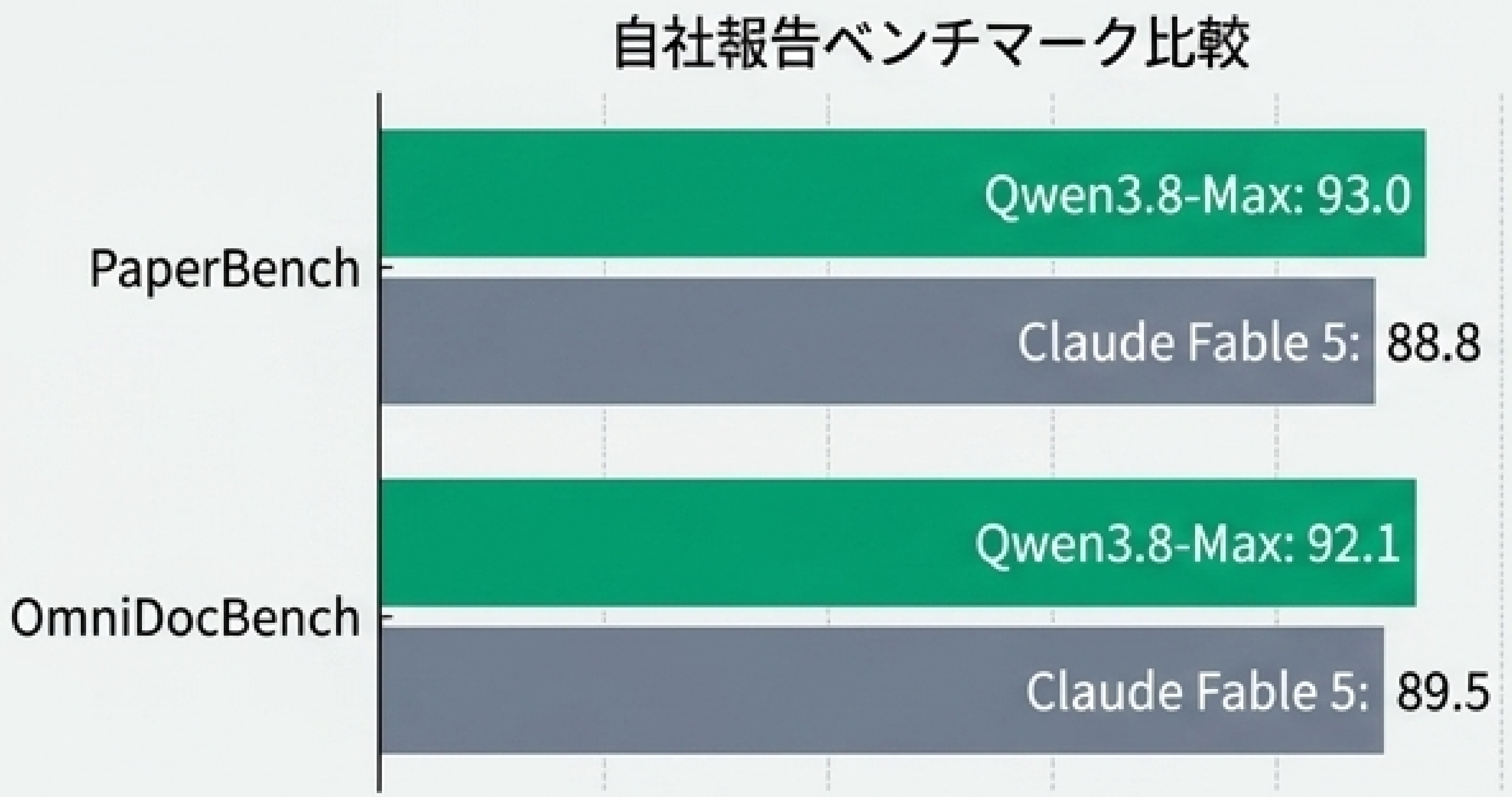
Gated DeltaNet +
スパースMoE

マルチモーダル

テキスト・画像・動画・
文書をネイティブ入力

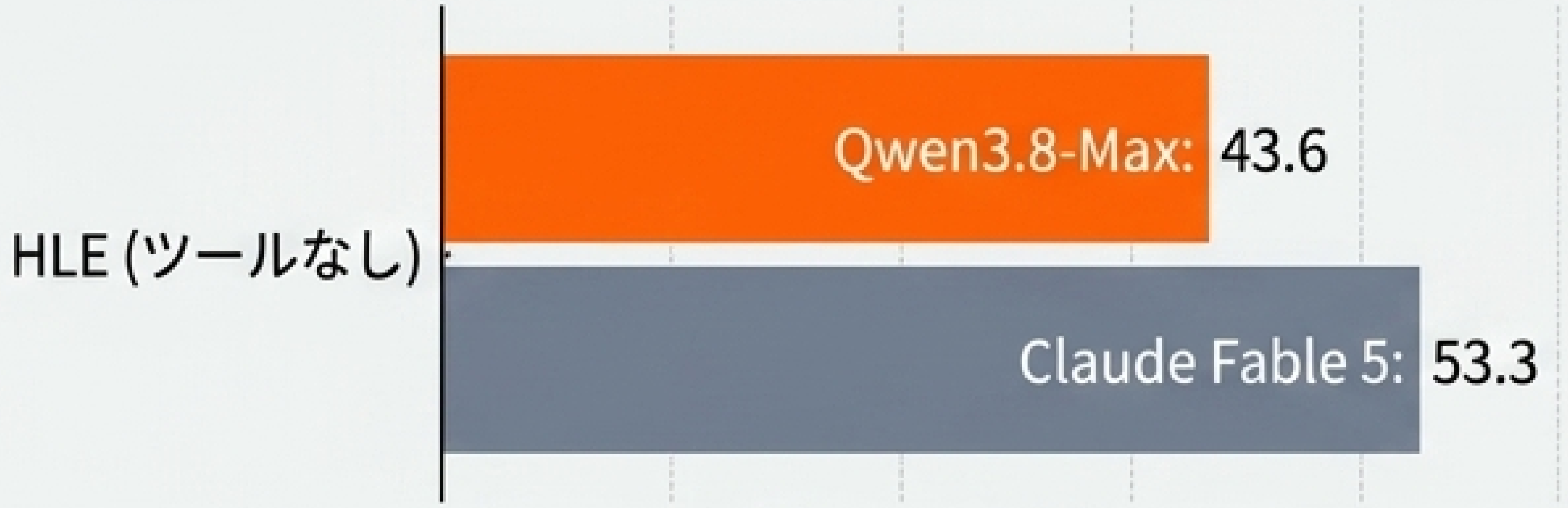
Benchmark Reality: コーディングの躍進と、残存する「推論・知識」の壁

Expectation (Alibaba 自己申告値)



コーディングとマルチモーダル領域で米国フロンティア勢を凌駕する自己申告スコア。

Reality (独立評価・推論の弱点)



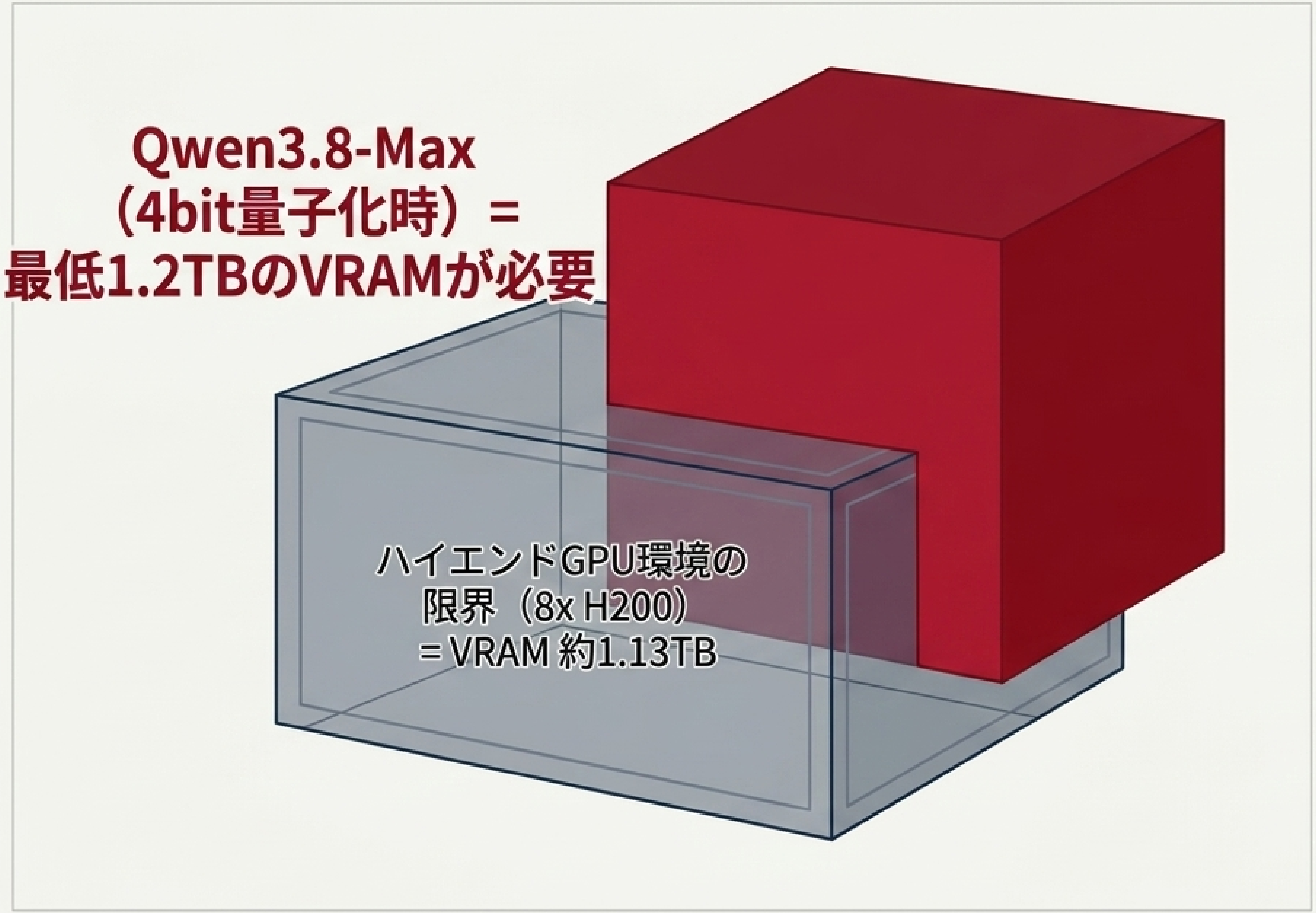
- 第三者による全表再現は未実施
- 独立評価はArena 5位、Artificial Analysis Index 53点のみ
- 前世代であった幻覚率22.9%の懸念が未払拭

Competitive Matrix: フロンティアモデル群における価格対性能のポジショニング

比較項目	Qwen3.8-Max	Claude Opus 5 (Fable 5)	GPT-5.6 Sol	Kimi K3
パラメータ規模	2.4兆	非公開	非公開	2.8兆
得意領域	コーディング・マルチモーダル	総合・汎用	総合・汎用	長文コンテキスト
APIコスト (入力/出力 per 1M)	**$2.00 / $6.00**	$5.00 / $25.00	Qwenの約4倍	$3.00 / $15.00
制約・リスク	知財未確定・中国3法	コスト高	コスト高	独自制限ライセンス・幻覚率

インサイト: 米国フロンティア勢の3〜4分の1という圧倒的なAPI価格。ブレンド単価約$1.18（キャッシュ考慮）は、エンタープライズの大量バッチ処理の経済性を根本から変える。

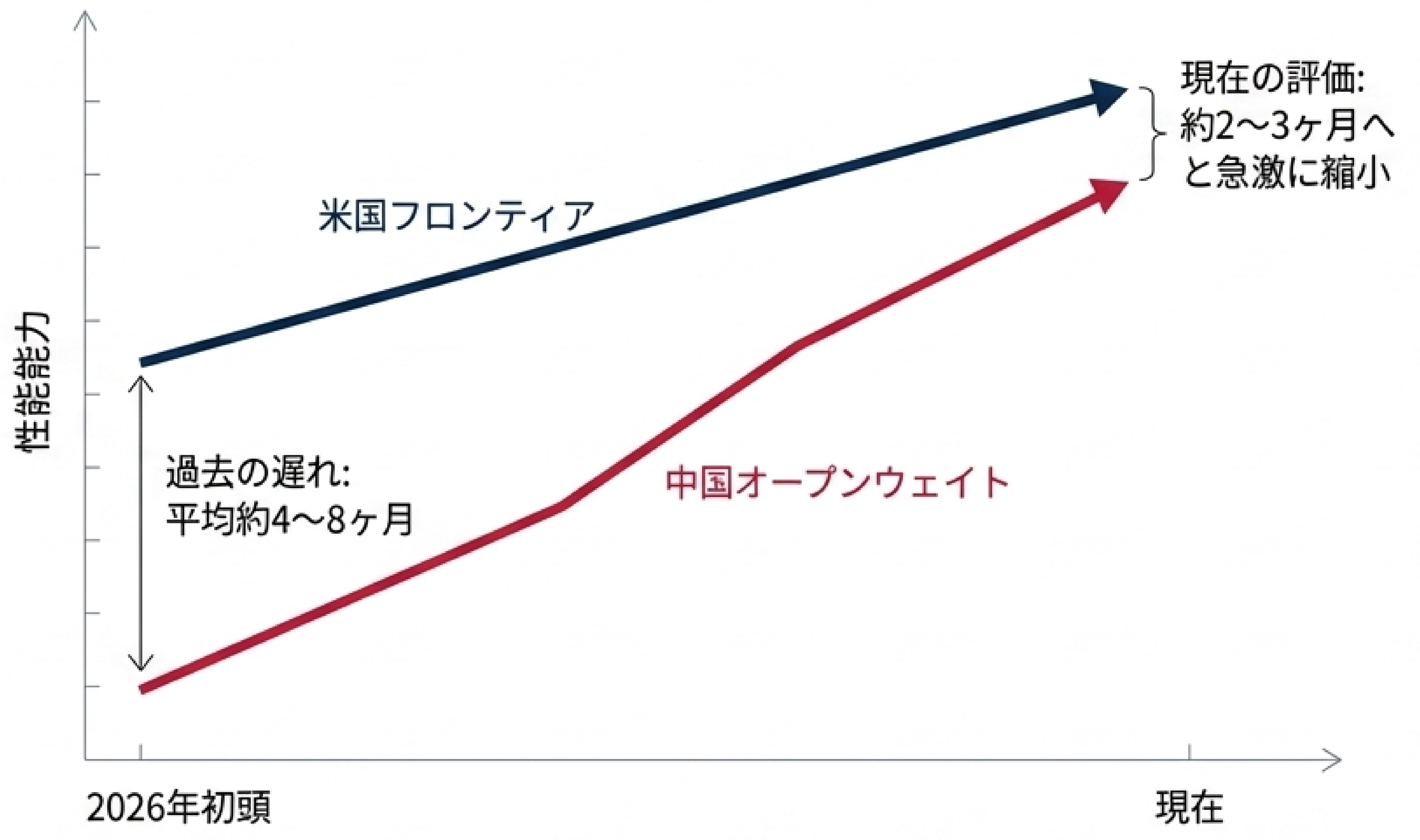
The "Open-Weight" Illusion: 2.4兆パラメータの自己ホストは事実上不可能



データセンター級のインフラがなければ稼働しない

現実的なオンプレミス（自己ホスト）経路は、同時に公公開される小型版「Qwen3.8-27B」のみに限定される。

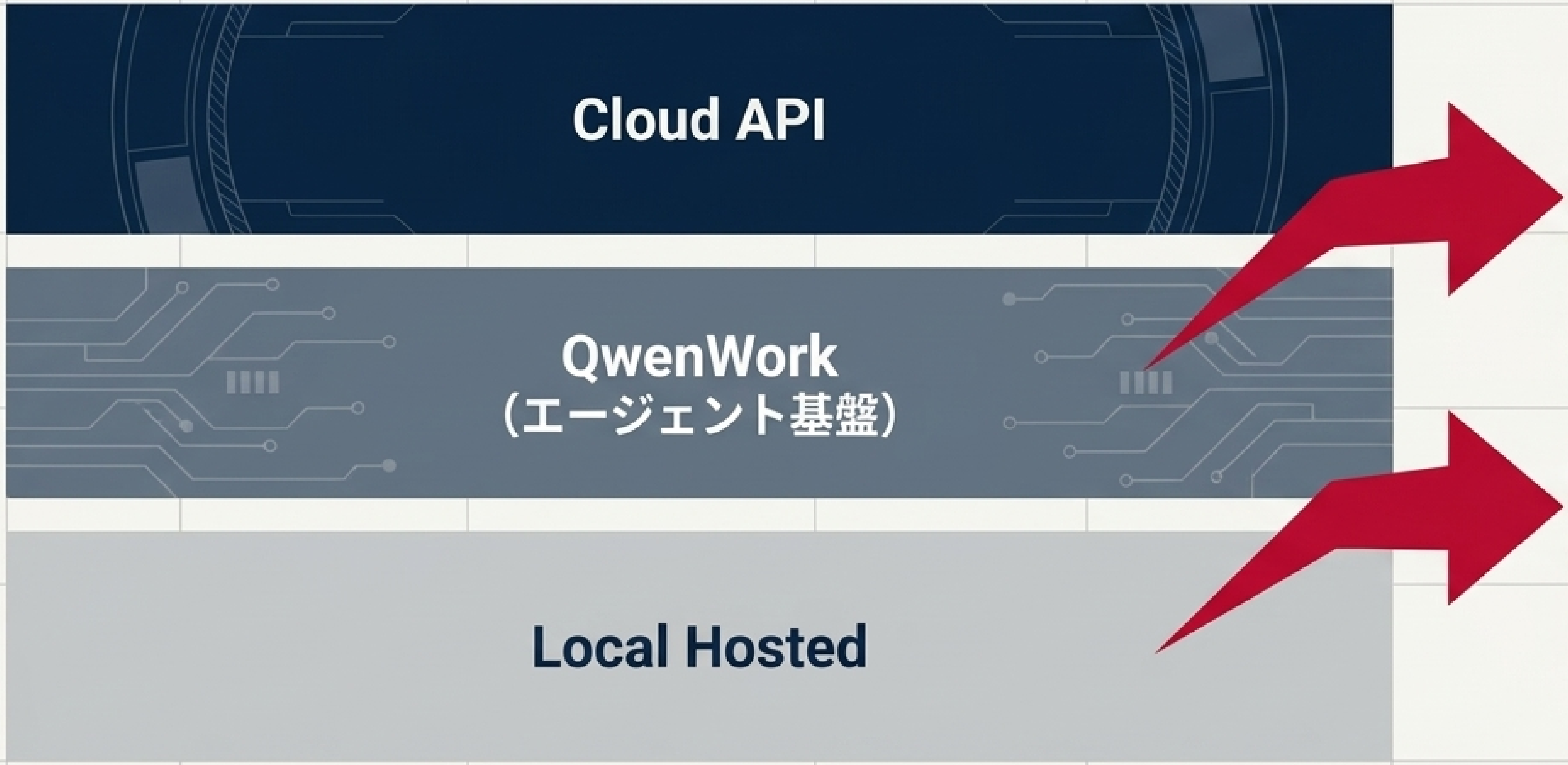
Geopolitical AI Landscape: 縮小する米中フロンティアの性能ギャップ



Qwen3.8-Maxはこのトレンドを完全に裏付けるマイルストーン。ただし推論系（HLE等）には依然として10ポイント近い差が存在。

⚠ ※学習チップの謎: AlibabaのPPU「Zhenwu M890」の存在。学習ハードウェアが未公表であることの不気味さ。

Governance & Security Risks: 導入を阻む「中国3法」と「米国防総省リスト」の壁



中国3法のリスク

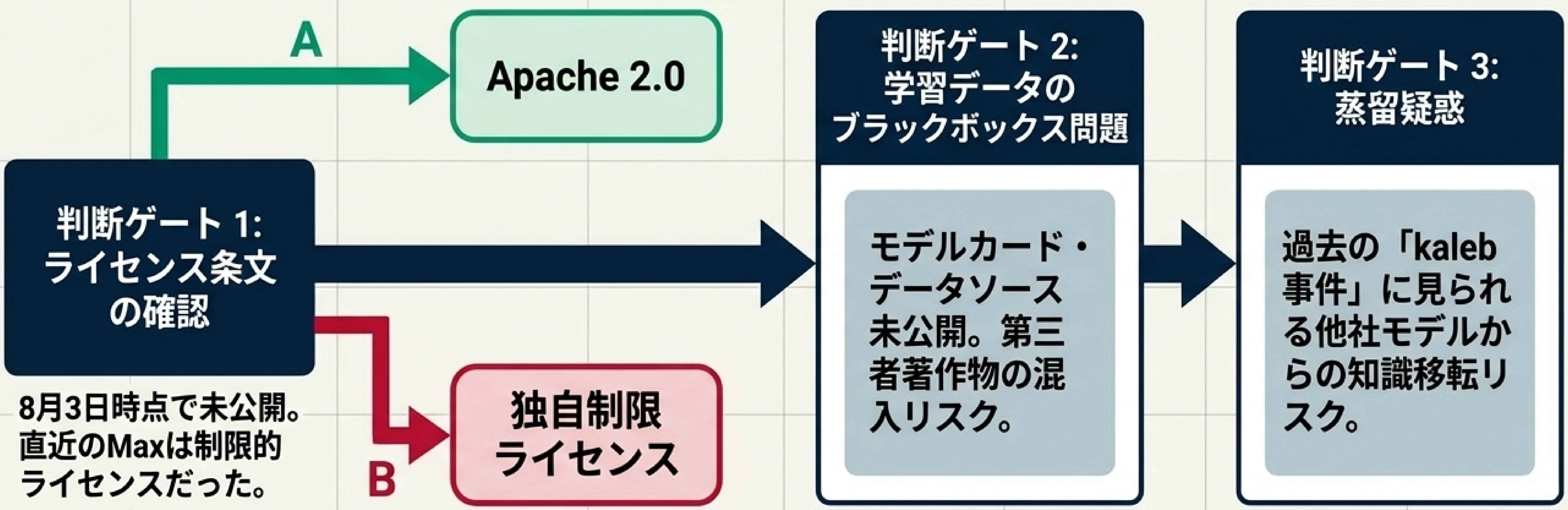
国家情報法（第7条・14条）、サイバーセキュリティ法、データセキュリティ法。サーバ所在地に関わらず適用される懸念。

米国防総省1260Hリスト

2026年6月更新で追加。直接契約禁止および間接調達禁止のリスク。

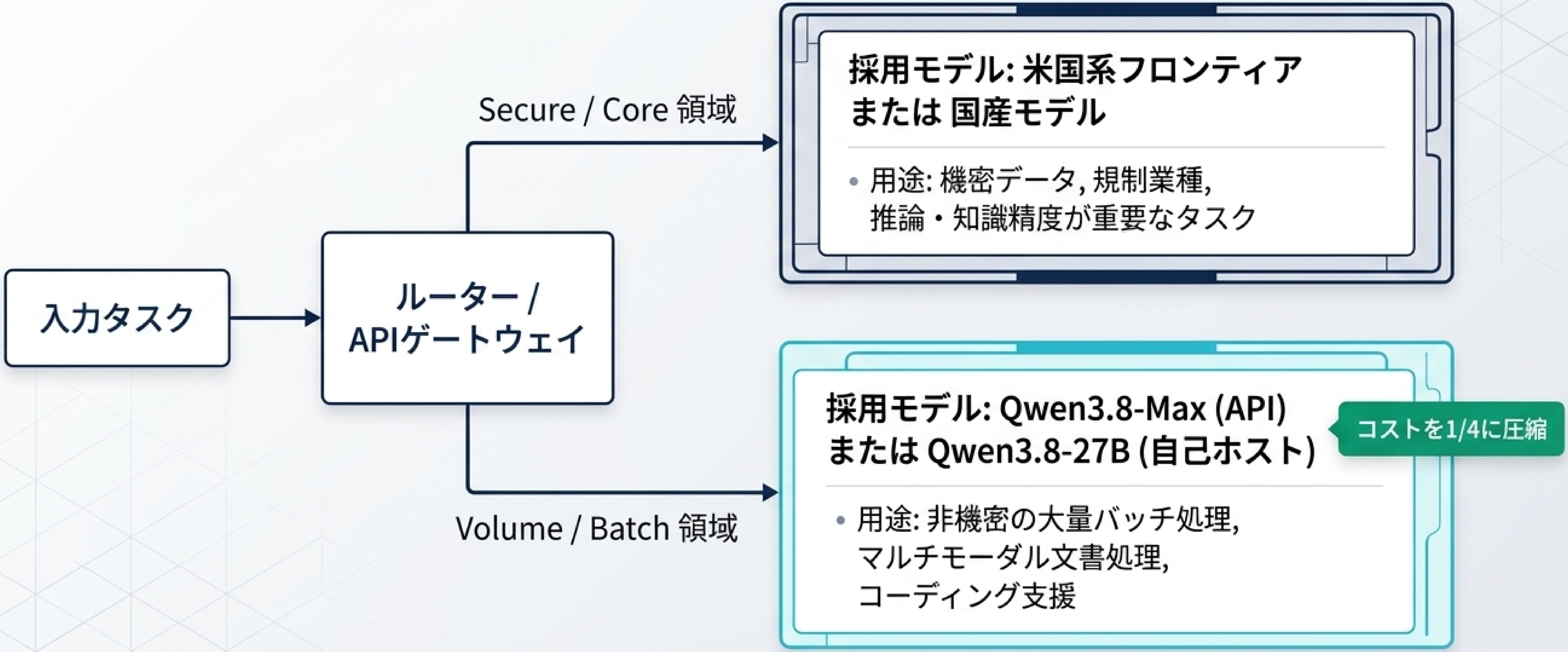
結論: エージェント基盤（QwenWork）による全社文書の取り込みは極めてハイイリスク。

IP & Copyright Gates: 知財実務上の「未確定リスク」と確認フロー



提言: ライセンス条文（Hugging Face）の法務レビュー完了まで、製品組込みは完全凍結。

Recommended Architecture : 適材適所の「デュアルスタック」運用戦略



Action Plan & Triggers: 段階的検証アプローチと導入判断の閾値

Roadmap

1 即時（評価フェーズ）

東京リージョン環境で、自社の実プロンプト(20-30件) による A/B テストを実施。本番トラフィックは移行しない。

2 近日中（知財ゲート）

オープン ウェイト公開後、Apache 2.0であることを確認。

3 中期的（ガバナンス評価）

米国関連取引の有無、機密データ 要件と照らし合わせ、クラウド利用可否とオンプレ限定運用を策定。

判断を変える閾値 (Triggers)

TRIGGER A

If: 独立系評価でコーディング性能がAlibaba申告値の±3pt以内と証明された場合

↓

Then: コーディング用途での全社採用を格上げ。

TRIGGER B

If: 知識タスクの独立幻覚率が前世代(22.9%)から大幅改善した場合

↓

Then: 知識検索用途への適用を再評価。