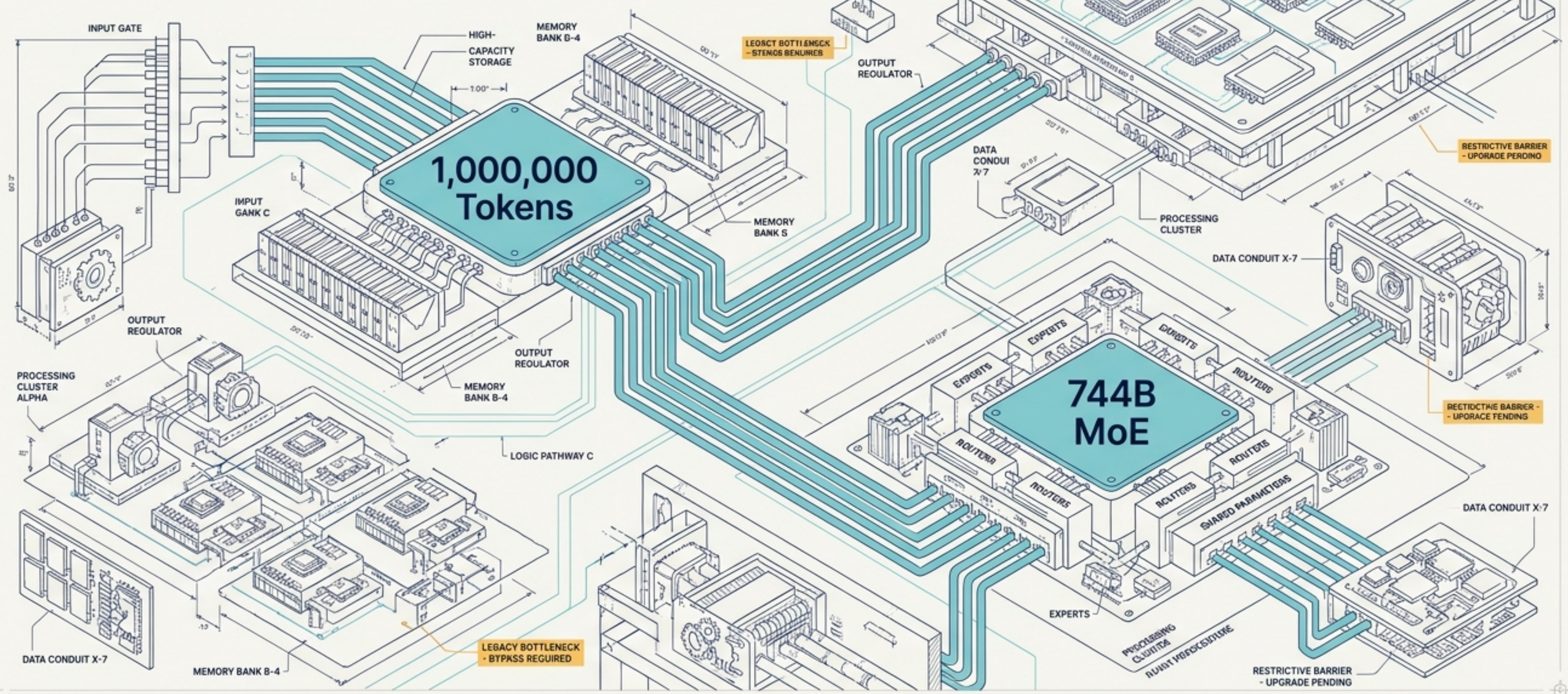


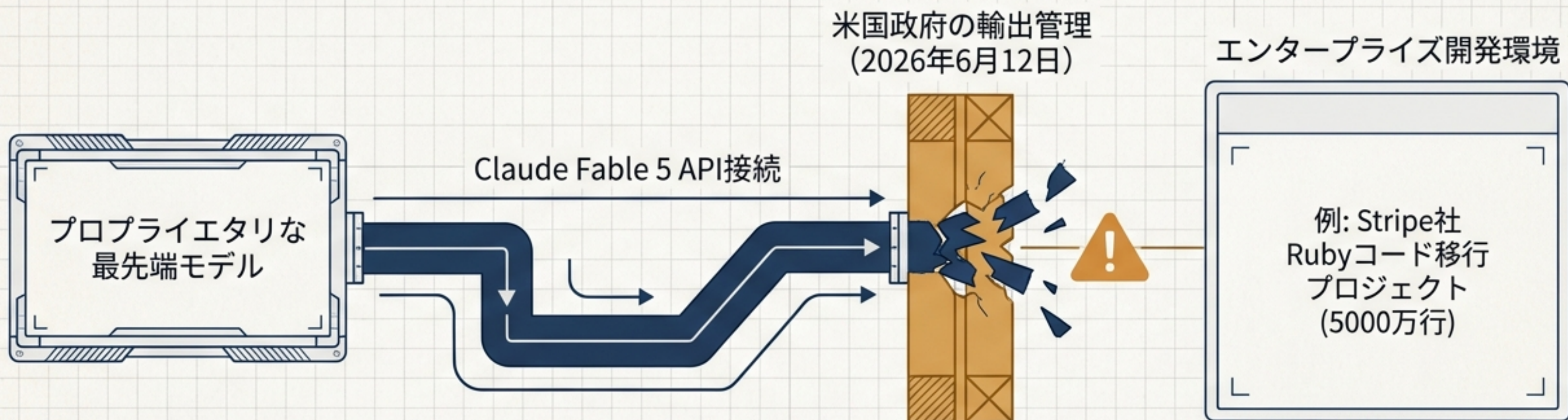
# Blueprint of a Paradigm Shift

GLM-5.2と「エージェント的エンジニアリング」の幕開け



# クローズドAPI依存がもたらす致命的リスク

特定プラットフォームへの依存は、事業継続における「サプライチェーンの脆弱性」を露呈させる



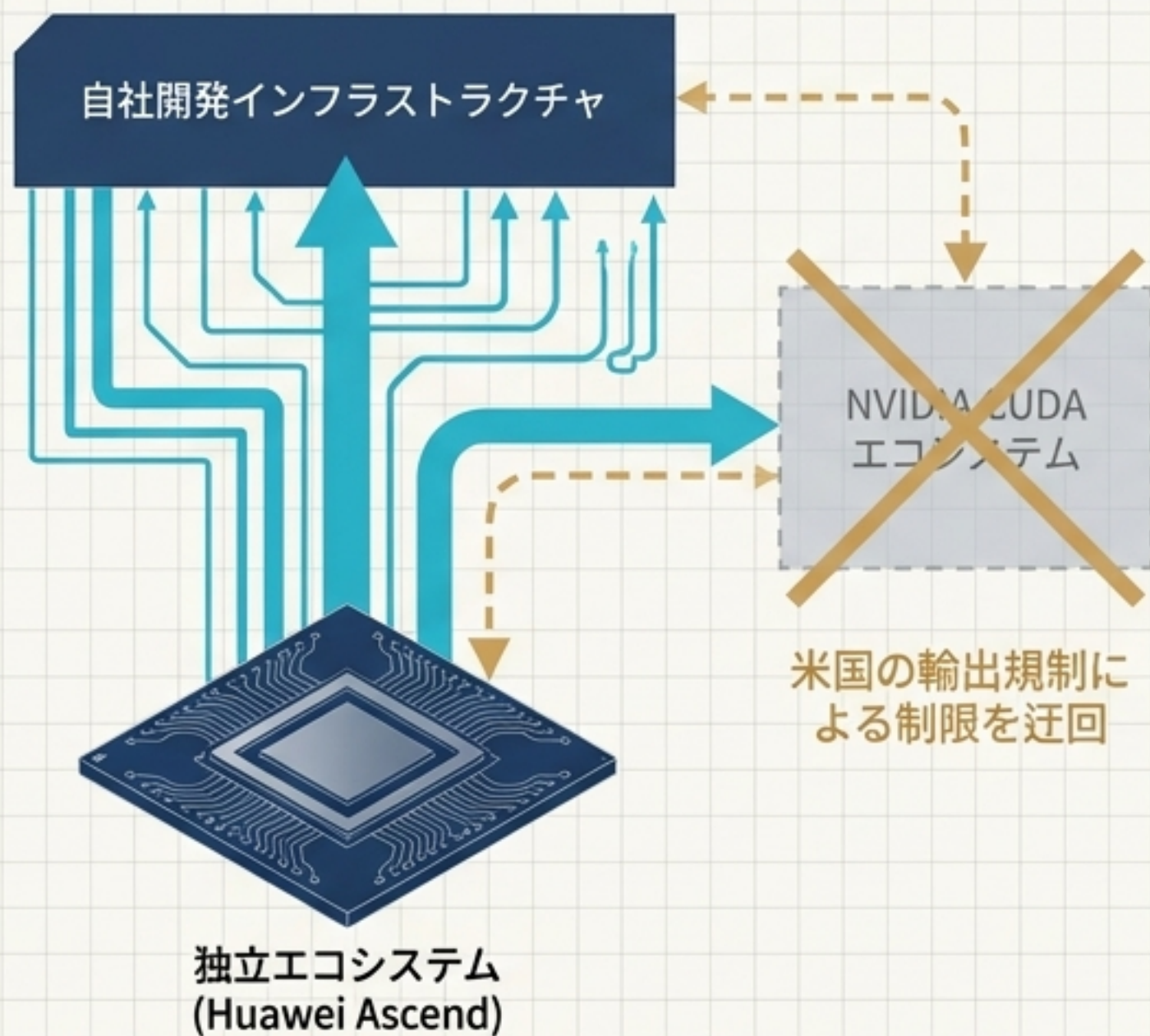
# 次世代フロンティアモデルの戦略的ポジショニング

評価軸	米国製プロプライエタリモデル (GPT-5.2 / Claude 4.5 Opus)	GLM-5.2
提供形態	特定APIプラットフォーム依存	MITライセンス・完全オンプレミス可
ハードウェア依存	NVIDIA / CUDAへの極端な依存	独立エコシステム (Ascend等) で完結
1Mコンテキストのコスト	高額な従量課金	破壊的低価格 (定額プラン等)
主要エージェント能力	トップクラス	トップクラス (SWE-bench等で互角以上)

# 制約を突破するMoEアーキテクチャと独立ハードウェア

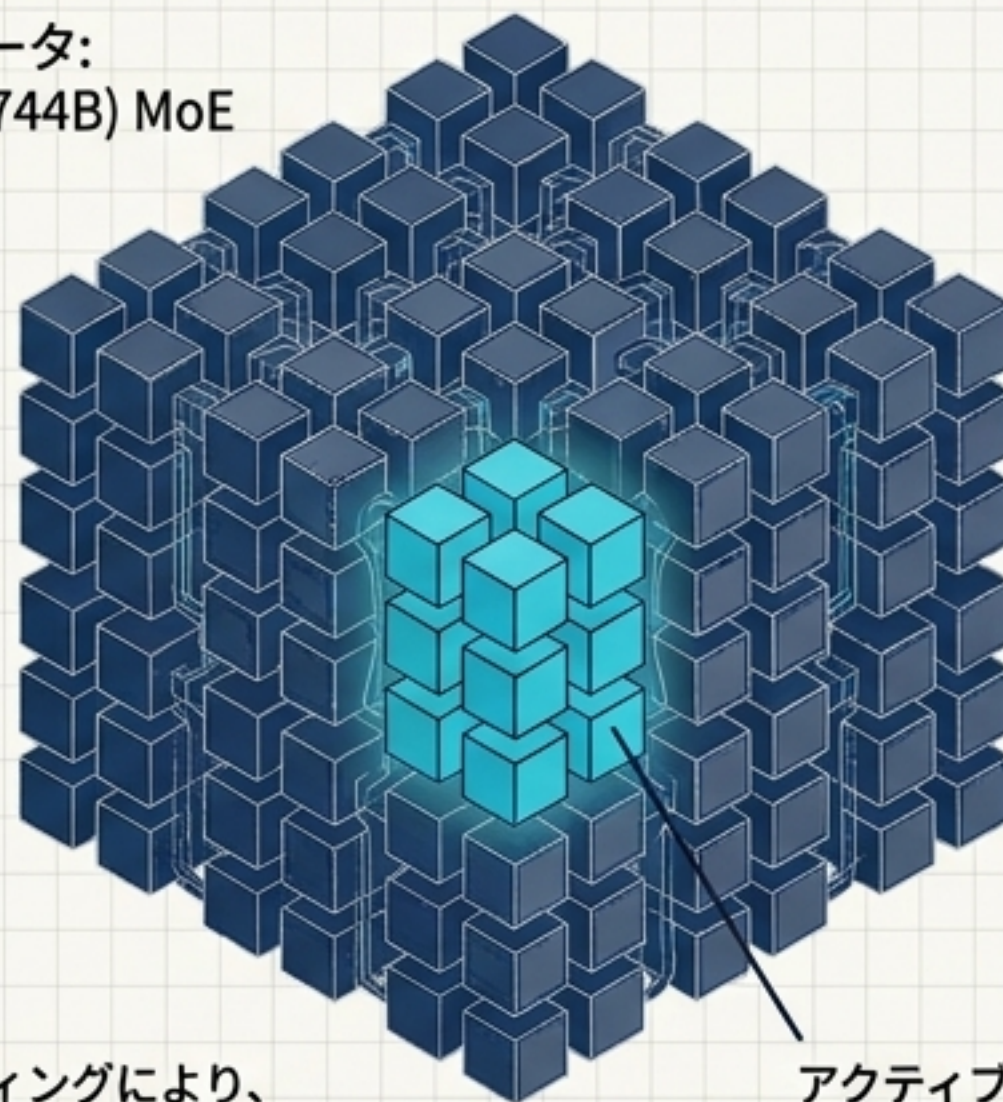
米国の輸出規制を迂回し、ゼロから独自インフラで世界トップ性能へ到達

## ハードウェア独立インシジョン



## 総パラメーター工率率を劇化

総パラメータ:  
7,440億 (744B) MoE



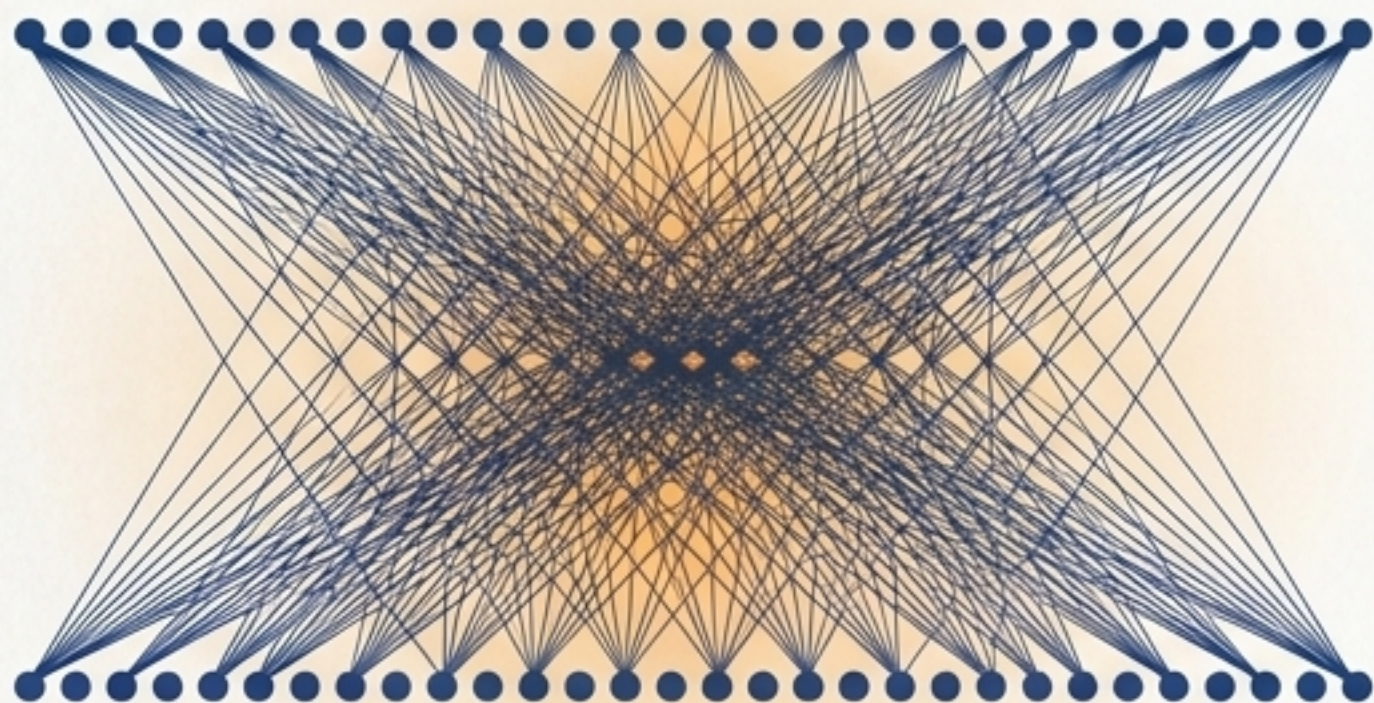
動的ルーティングにより、  
リクエストごとに必要な  
エキスパートのみを活性化

アクティブ・パラメータ:  
- 約400億 (40B)  
- 計算コストの極限最適化

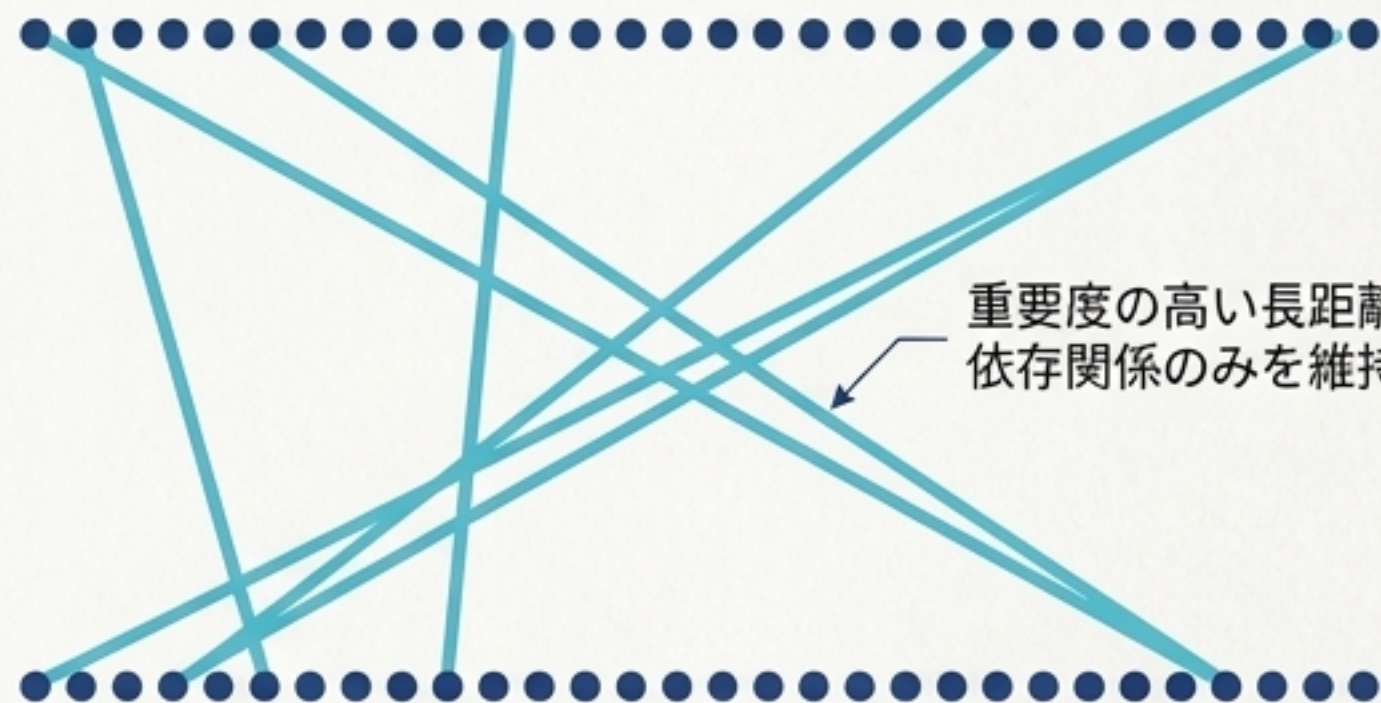
# DeepSeek Sparse Attention (DSA) : 1Mコンテキストの実用化

長距離の依存関係を維持したまま、計算量とメモリ消費の二次関数的増大を打破

従来型 Attention機構



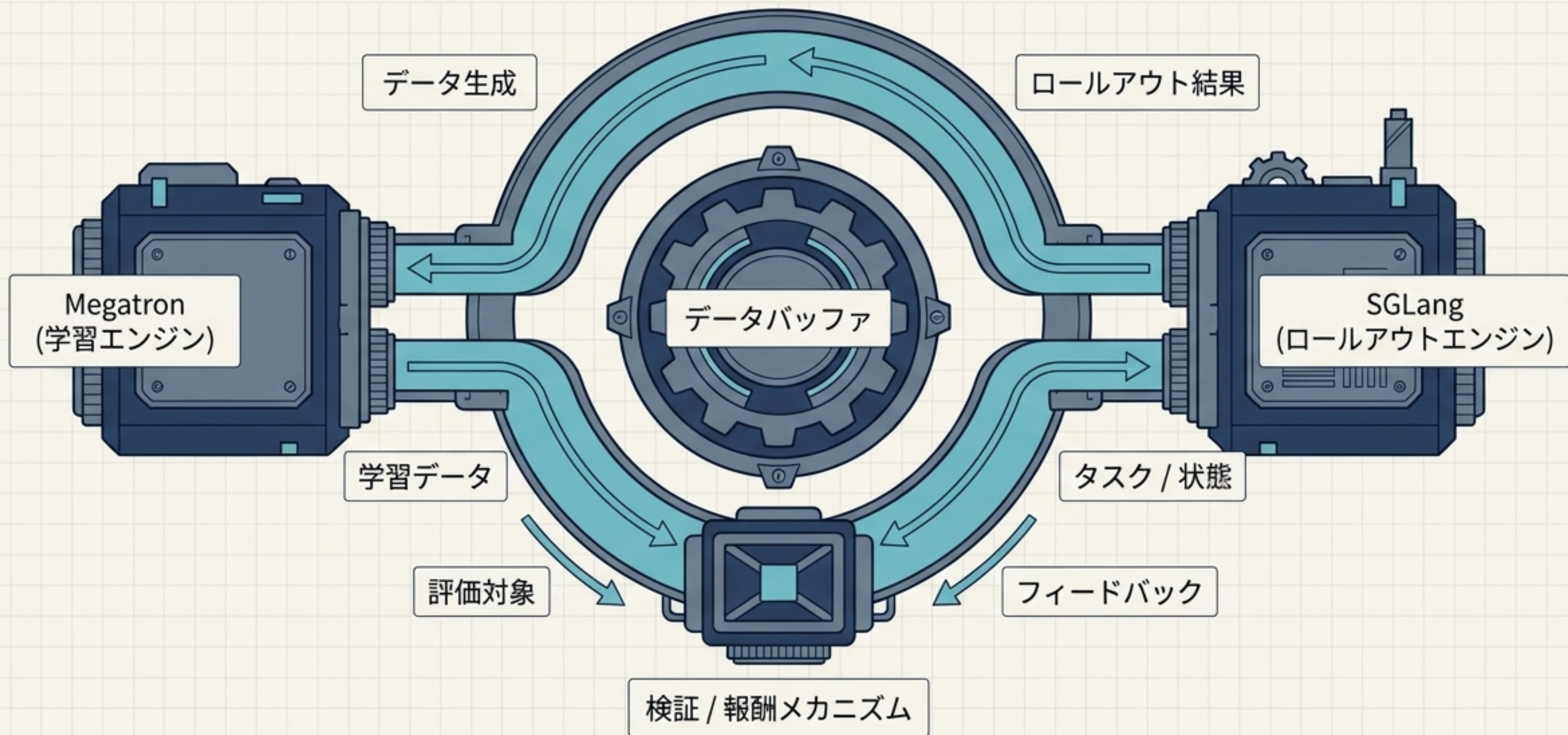
GLM-5.2 DSA (動的スパース化)



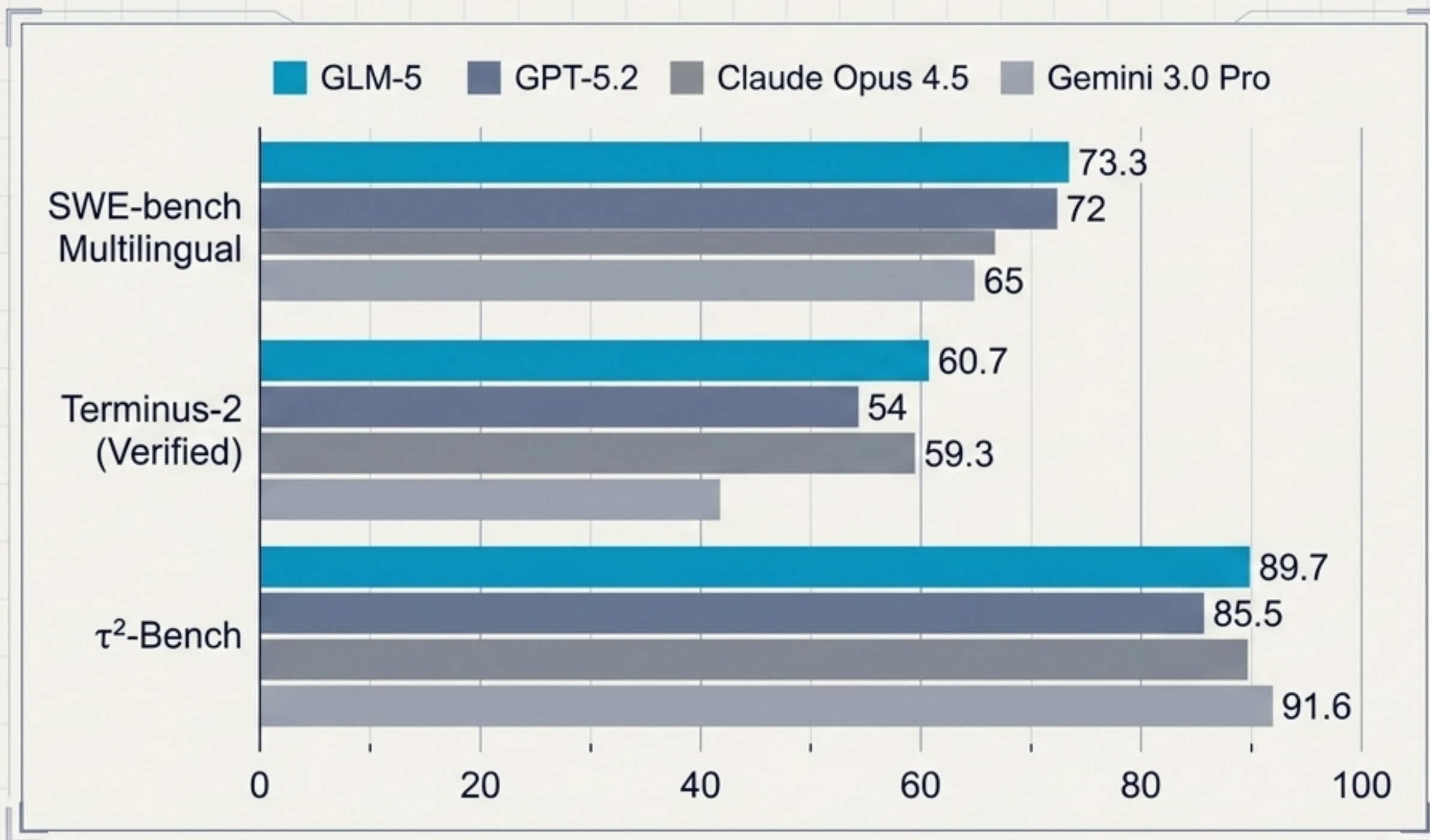
実用的な100万トークン処理の実現

# Slimeインフラストラクチャ：非同期強化学習のエンジン

推論と学習をシームレスに結合し、「思考の連鎖 (Chain-of-Thought)」をネイティブに獲得



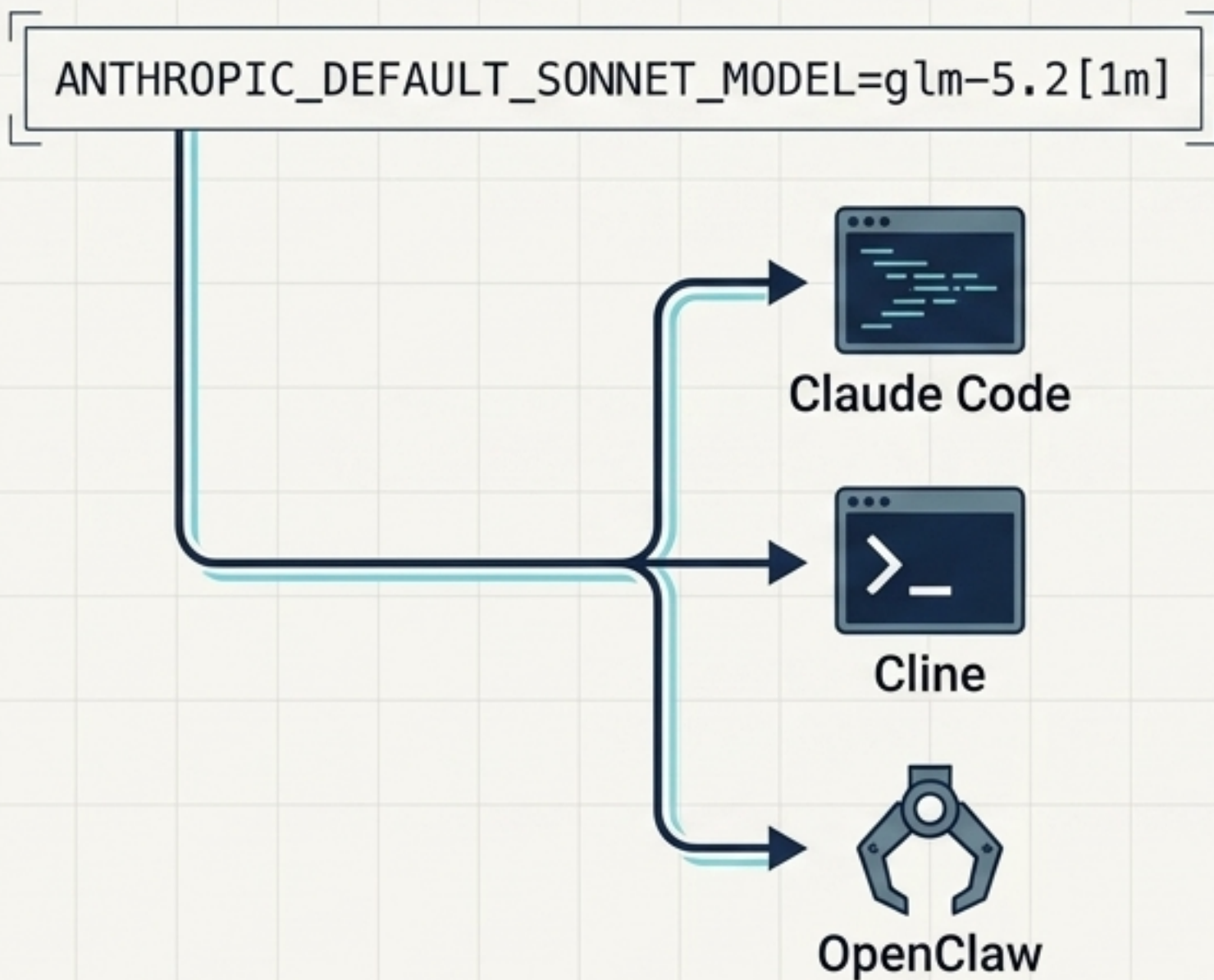
# ベンチマークの再解釈：Long-Horizonタスクでの優位性



「1ショットの一発書き」ではなく、状態の監視とエラーからの自己補正を伴う長期間タスク (Long-Horizon) においてプロプライエタリなトップモデルを凌駕

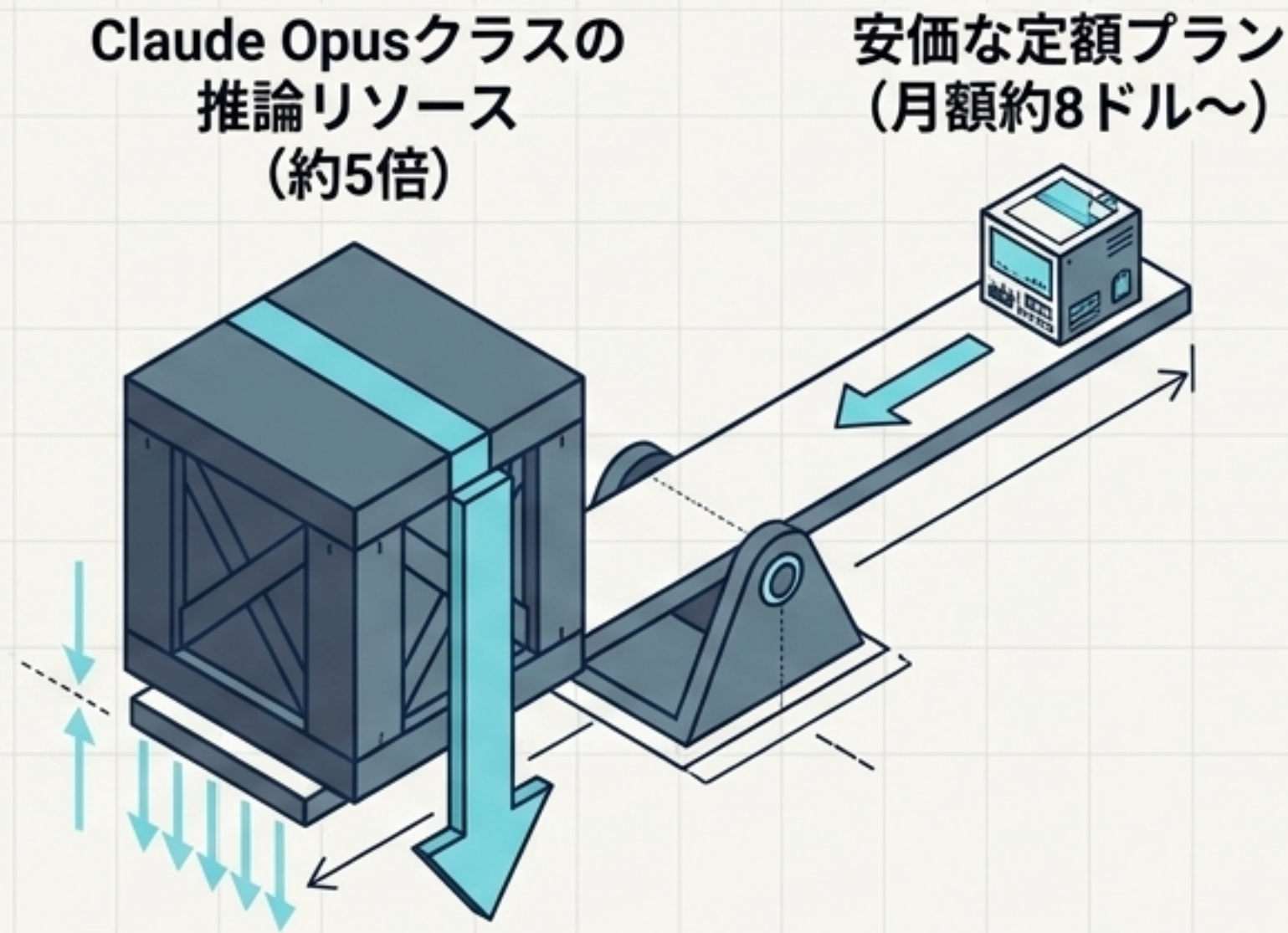
# 既存ワークフローへの統合と圧倒的なコスト競争力

## 開発インフラのシームレスな統合



エンドポイントの切り替えのみでネイティブ連携

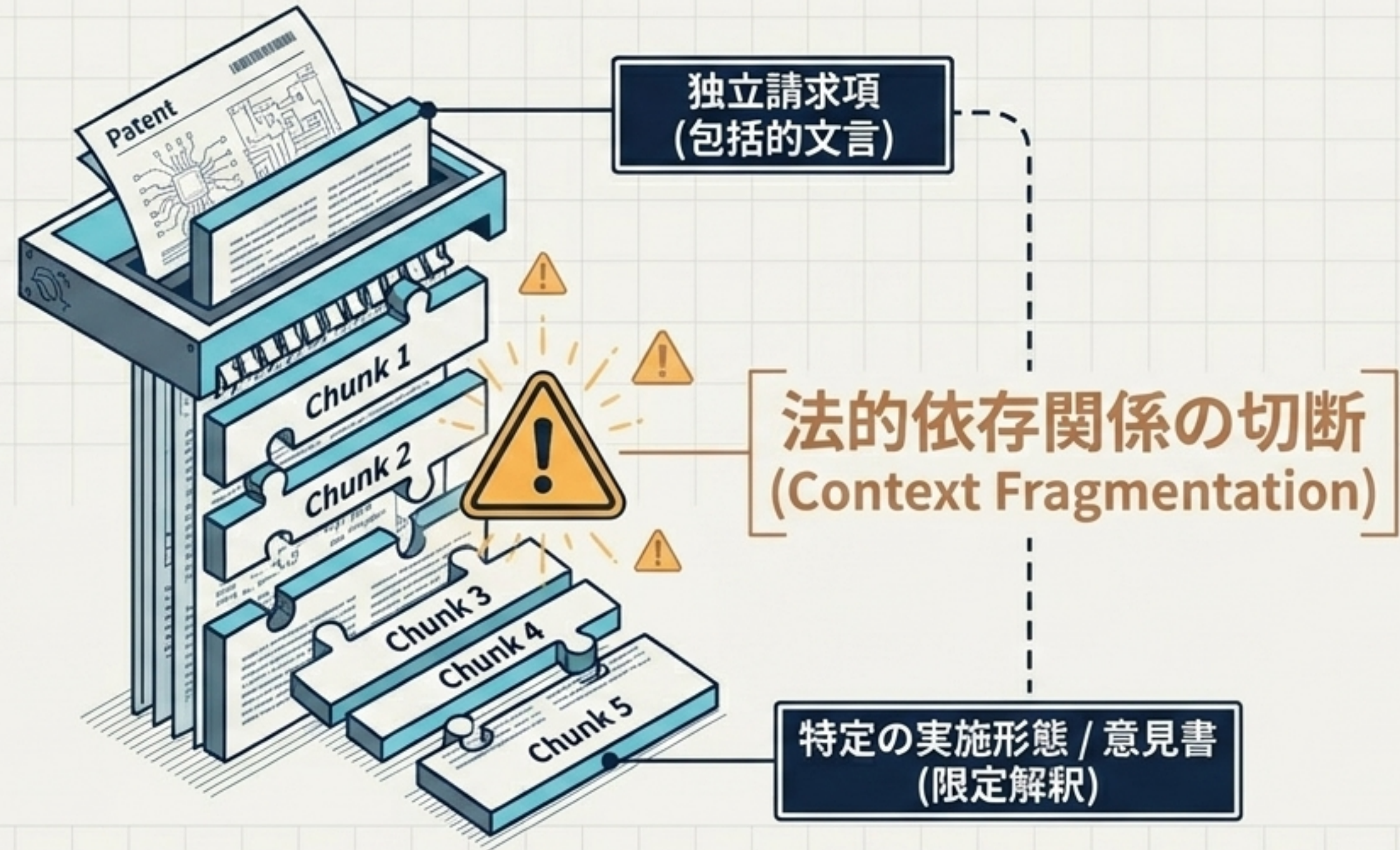
## 非対称なROIのレバレッジ



経済的破壊力による開発規模の拡張

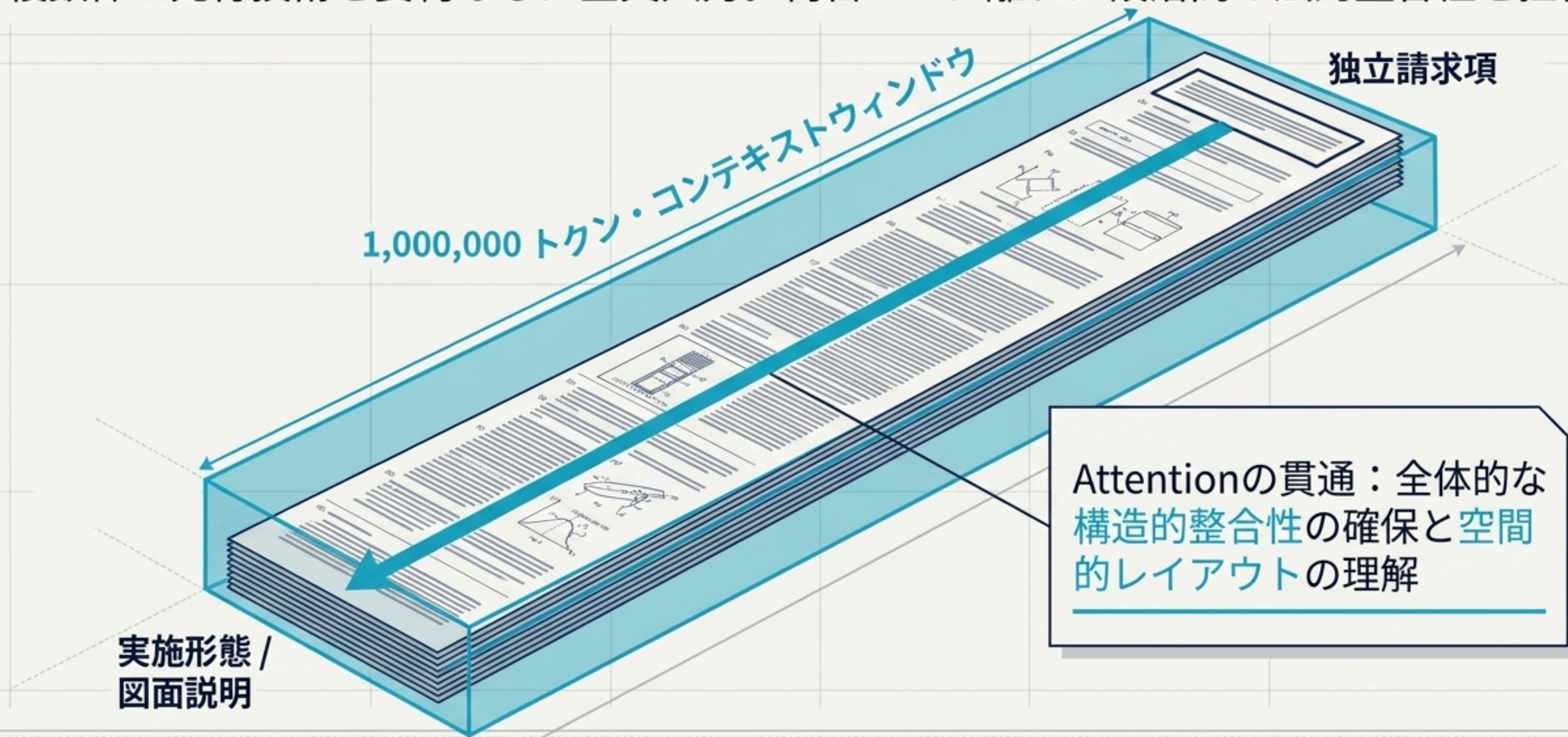
# 特許分析におけるボトルネック：「文脈の断片化」

従来のLLMによるチャンキングは、法的な権利範囲の限定解釈を切り離し致命的エラーを生む



# 1Mコンテキストによる「文書構造の完全な維持」

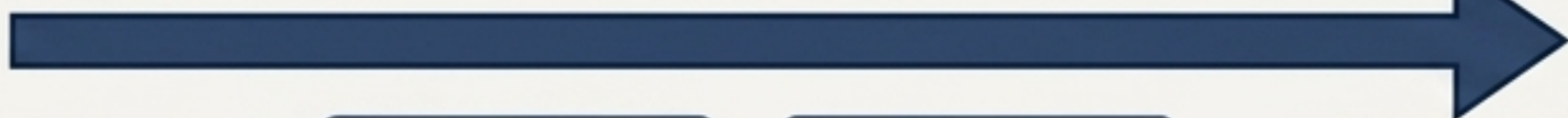
複数件の先行技術を要約なしに全文入力。何百ページ離れた段落間の法的整合性を担保



# 限界費用の崩壊と「反復的ストレステスト」の誕生

一発勝負の分析から、事実上無費用の多角的な仮説検証サイクルへ

人間のパートナー弁護士: 1回の高コストな分析

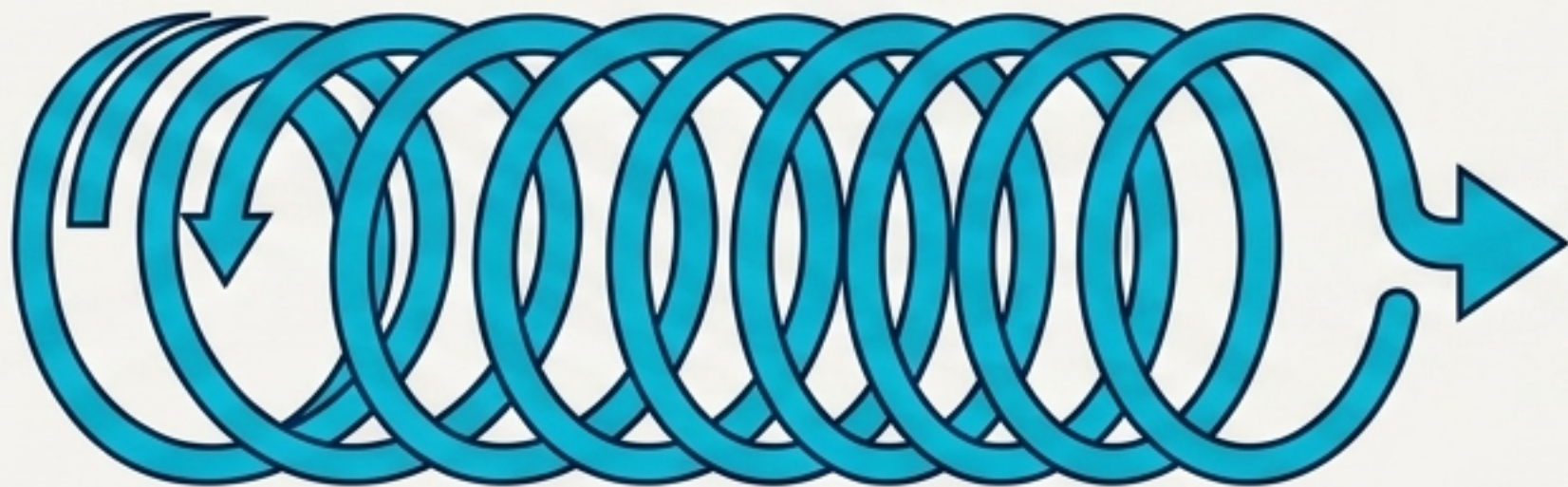


単一の結論

所要時間: 数時間

コスト: 数百ドル

GLM-5.2: 何十回ものマルチプル・パス検証



強固な防波堤  
(新規性)

所要時間: 数十秒

コスト: 約15~30セント

特許先行技術対比における分析コストの圧倒的格差



Data sources: IP Chimp Blog

# Context Distractionを防ぐハイブリッド・ファネル戦略

ハルシネーションを抑制し、厳密な証拠性 (Explainability) を確保する2段階のアプローチ

Step 1: 膨大な特許群の投入  
(1Mコンテキスト)



GLM-5.2による粗い  
スクリーニングの実行

Step 2: 該当する少数の段落  
(数百トークン) の抽出

厳密な法的マッピングの実行

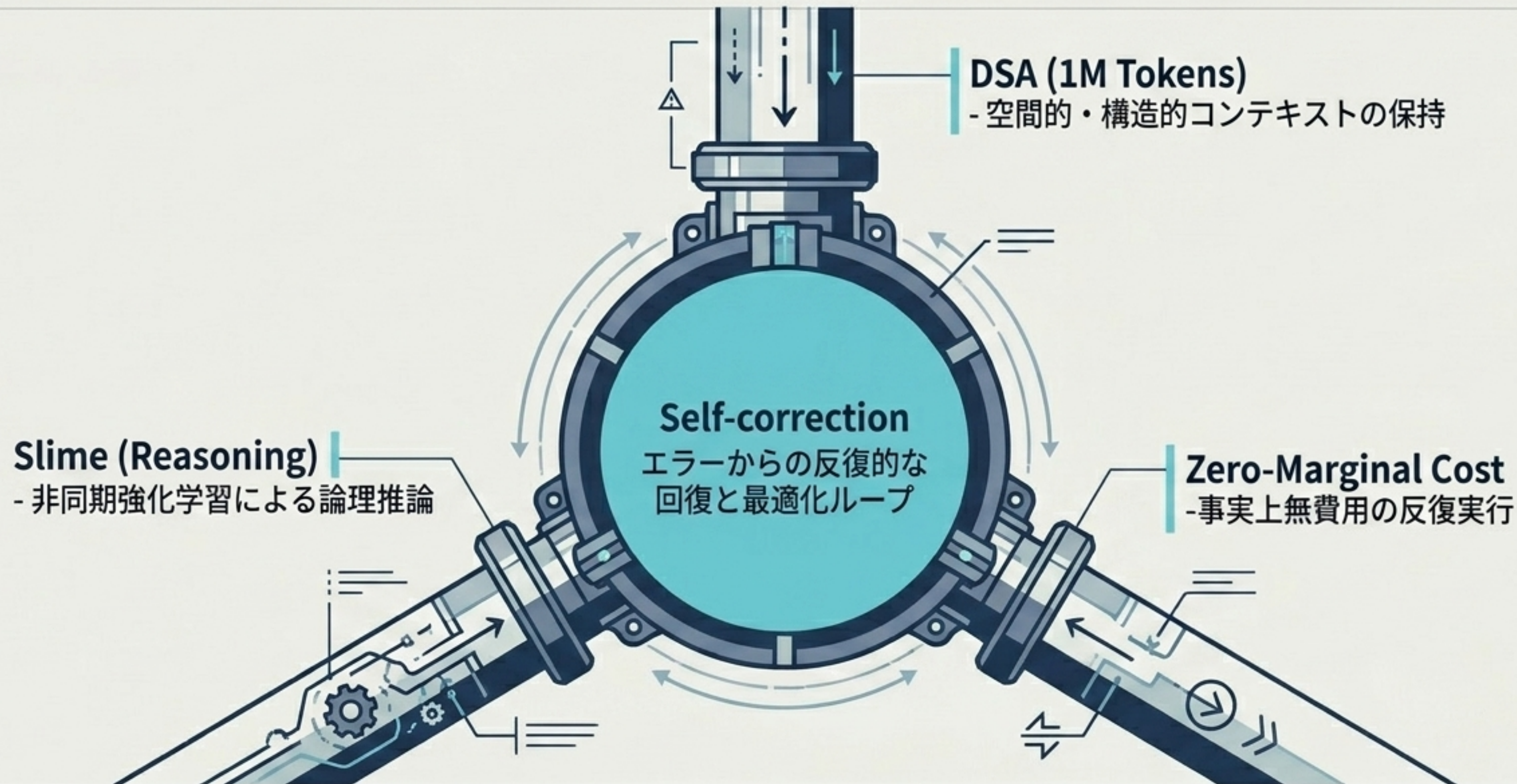
段落・行番号を特定し、  
専門家が検証可能な精緻な  
レポートを出力

# LLMを活用した特許分析アプローチの進化

	従来型LLM (チャンキング必須)	GLM-5.2単体 (1M全文入力)	GLM-5.2 + RAG (ハイブリッド戦略)
文脈の断片化リスク	● 高い (致命的)	なし	なし
ハルシネーション・ 混同リスク	中～高	中 (Context Distractionリスク)	極めて低い
証拠性 (Explainability)	低い	中	厳密 (段落・行番号の特定)
推奨ユースケース	実務適用不可	粗いスクリーニング・ 全体概要把握	最終的な法的クレーム・ マッピング

# The Synthesis: Agentic Engineeringへの到達

Vibe Coding（雰囲気コーディング）の終焉と、自律的システム設計へのパラダイムシフト



# 戦略的インフラとしてのオープンソースの未来

機密性を最優先する企業にとって、完全なMITライセンス下のローカル構築は「主権的インフラ」である

