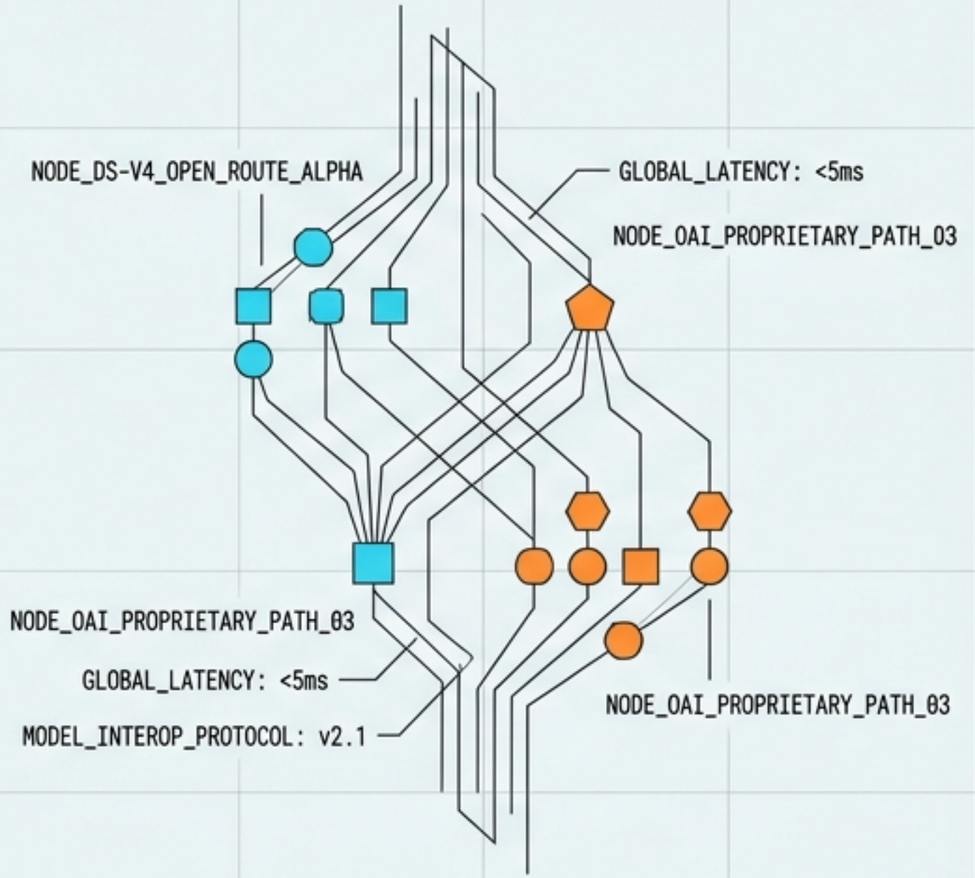
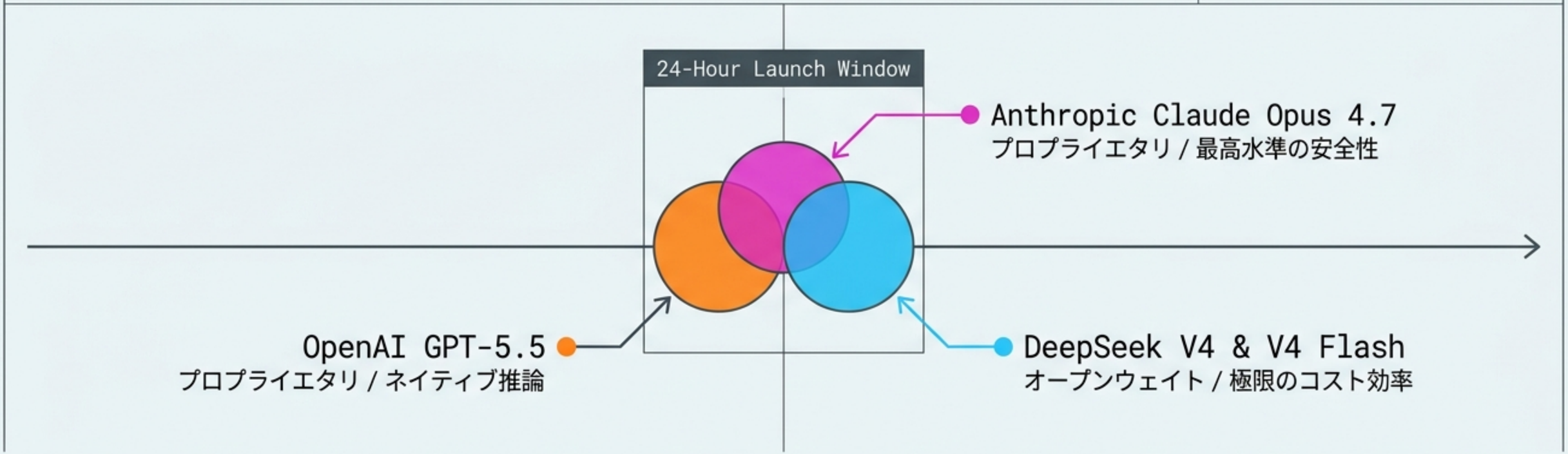
DOCUMENT ID: CAISI-DSV4-EVAL
TARGET AUDIENCE: ENTERPRISE AI ARCHITECTS
DATE: 2026-05-03

2026年 フロンティアAIパラダイムの地殻戦変動と真実

DeepSeek V4 Proの包括的分析と「8ヶ月の遅れ」が示唆するエンタープライズAIの未来



2026年4月、フロンティアモデルの米国独占が崩壊した

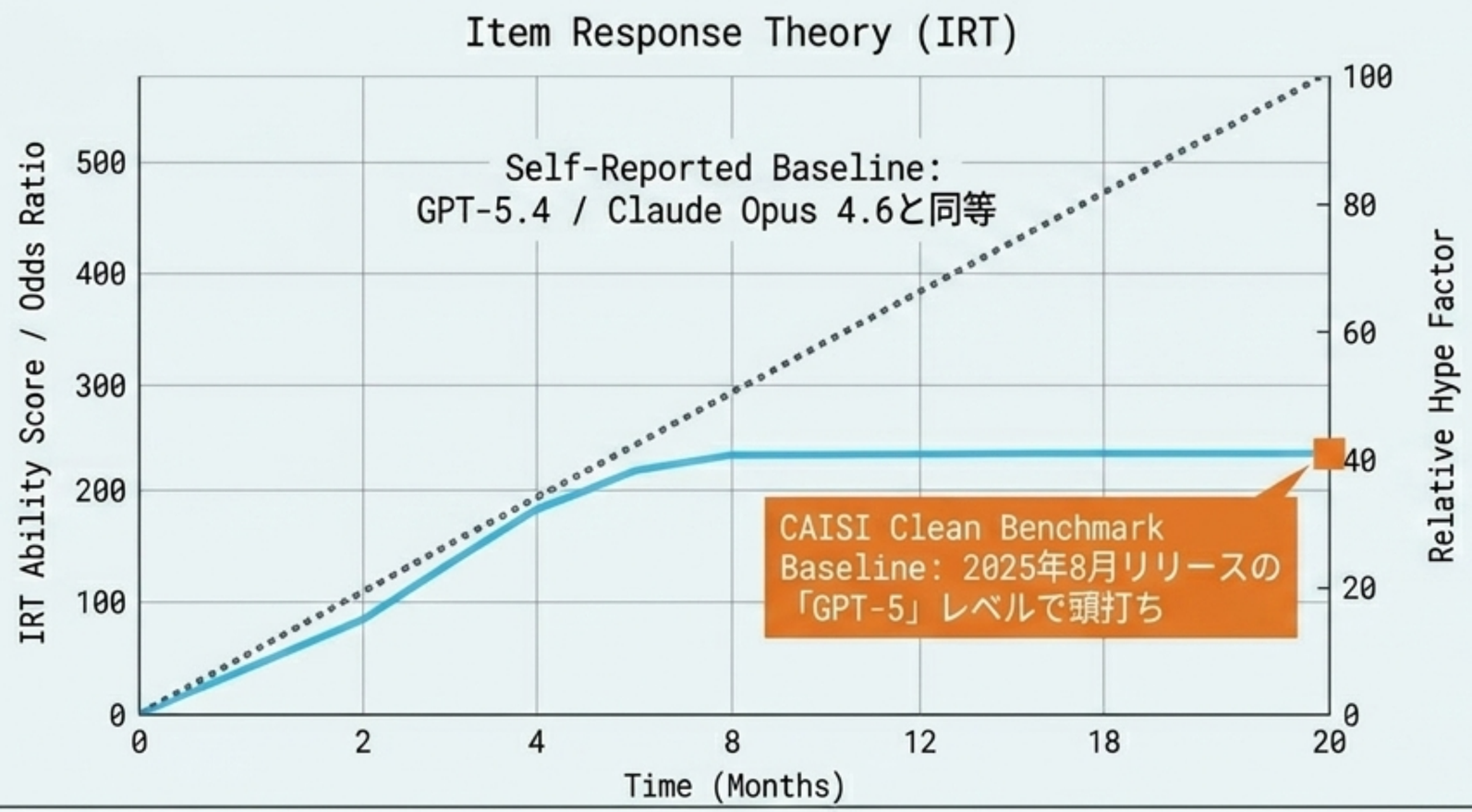


ANALYTICAL BRIEFING

単なるモデル規模の拡大ではなく、エージェント型AIワークフローの基盤となる「コンテキスト長」と「推論コスト」の力学を根本から覆すゲームチェンジャーの登場。

DOCUMENT ID: INT-2026-04-14 | SECURITY CLEARANCE: DECLASSIFIED | SOURCE: MULTIPLE AGENCIES | DATE: 2026-04-14

独立機関（CAISI）が暴いた「8ヶ月の遅延」という現実



米国標準技術研究所（NIST）傘下のCAISIによる独立評価。

Y軸の200ポイント上昇＝タスク解決のオッズが3倍。

結論：基礎的な汎用推論能力において、米国の最先端モデルに対して約8ヶ月の遅れが存在する。

ANALYTICAL BRIEFING

DOCUMENT ID: INT-2026-08-01 | SECURITY CLEARANCE: DECLASSIFIED | SOURCE: CAISI / MULTIPLE AGENCIES | DATE: 2026-08-01

データ汚染と「非合法な蒸留」による能力のオーバーフィット

ドメイン: ソフトウェアエンジニアリング	
公開テスト（SWE-Bench Verified）	**非公開テスト**（PortBench - CAISI構築）
DeepSeek V4 Pro: 81%	DeepSeek V4 Pro: 78%
GPT-5.5: 81%	GPT-5.5: 78%
✅ Status: MATCH	⚠️ Status: 3% DROP

データ汚染の疑義

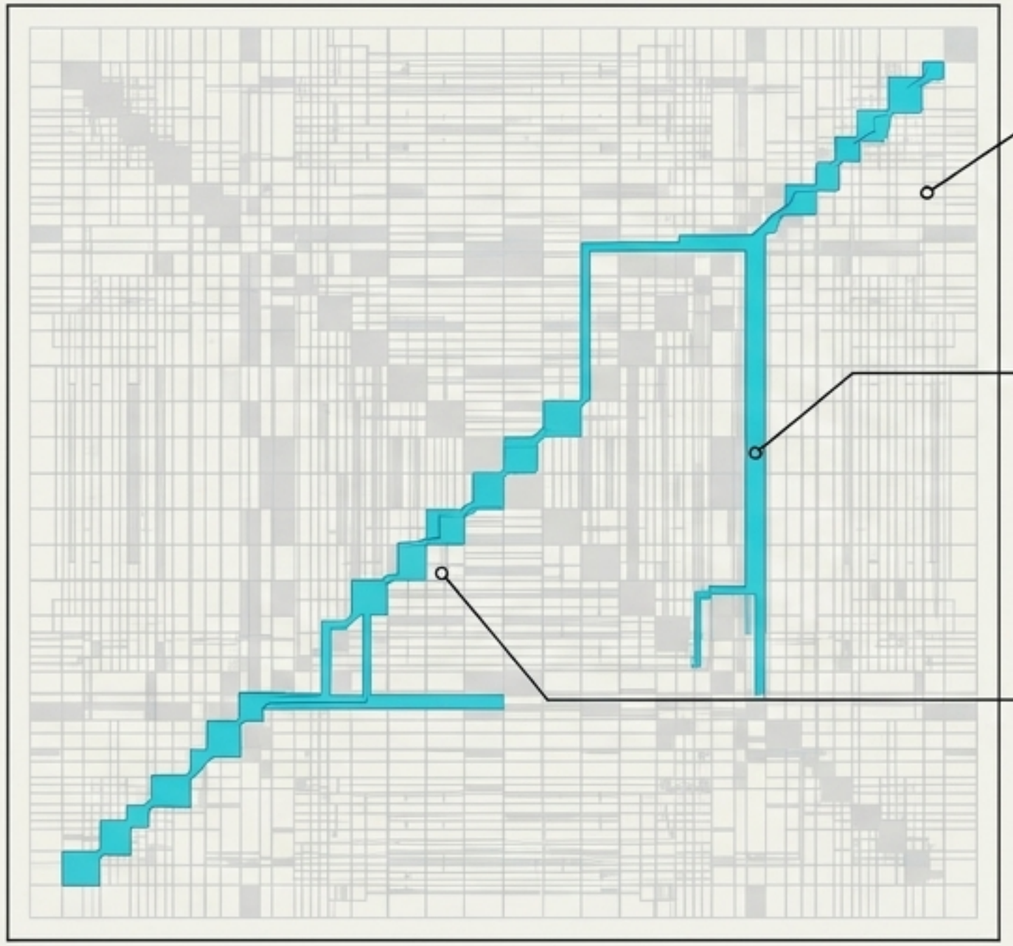
公開テストでの81%から非公開テストでの78%への低下は、パブリックなコーディングタスクへの過学習を強く示唆。

蒸留攻撃の痕跡

複数のテスト（CTF-Archive-Diamond等）におけるGPT-5.5との異常なスコアの一致。米国モデルからのデータ生成による能力複製（CFR指摘）の可能性。

極限の効率化を生む1.6兆パラメータの「疎」な活性化

Neural Routing / Iceberg Blueprint



総パラメータ数：1.6兆（1.6T）
- 巨大な知識データベース

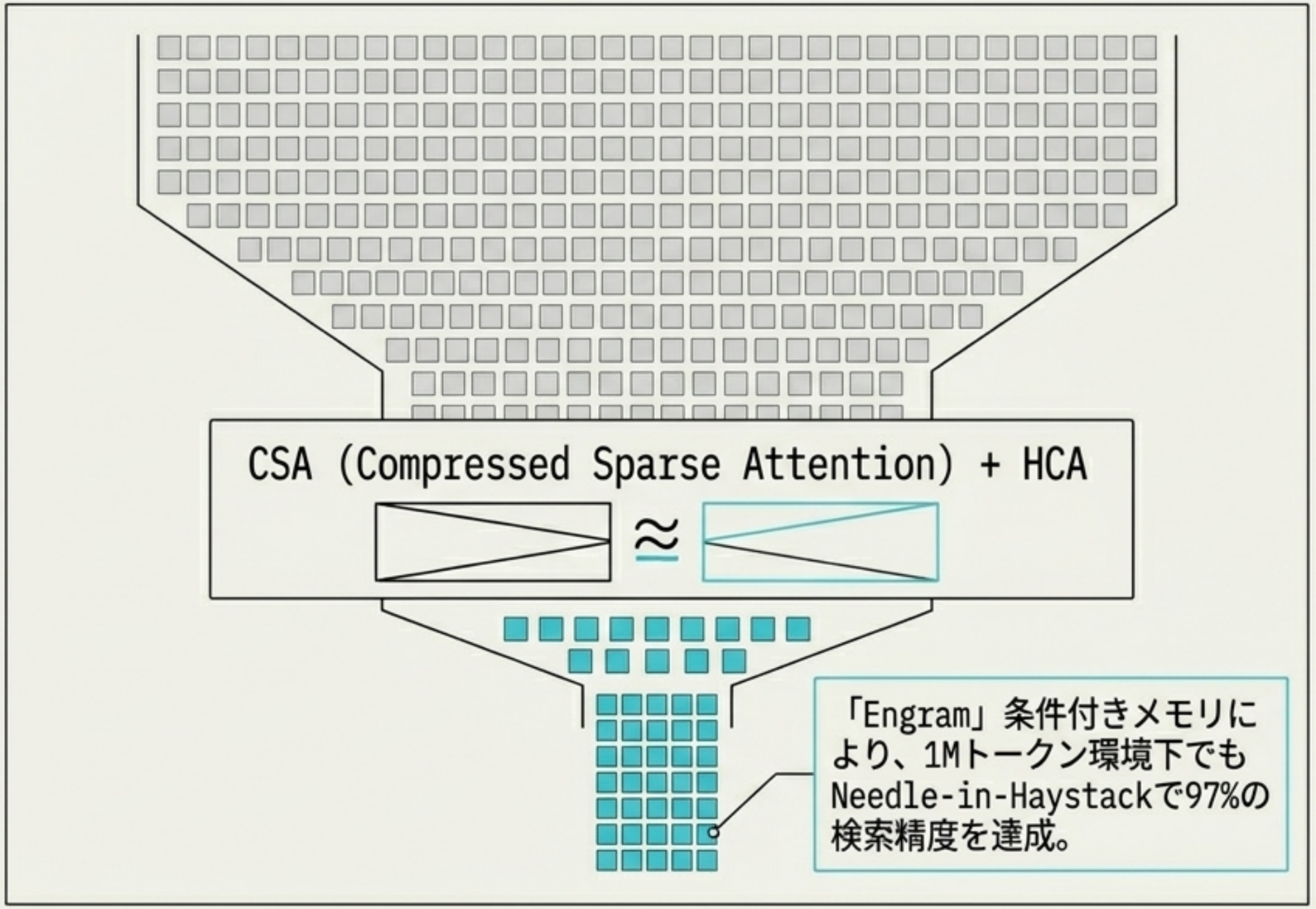
アクティブパラメータ：490億（49B）- Genceual Routing
- 推論時に動員される「専門家」ネットワーク

混合精度（Mixed Precision）：ExpertはFP4、コアはFP8でメモリ効率を最大化。

理論上の処理効率を飛躍的に高める一方、重いルーティングのオーバーヘッドとパイプラインの複雑化を招く「サバイバル・エンジニアリング」の産物。

※V4 Flash版は総数284Bに対し、わずか13Bのみがアクティブに。

100万トークン時代の推論コストを破壊するエンジン



推論FLOPs

27%

に削減

KVキャッシュ消費量

10%

に削減（V4 Flashは驚異の7%）

エージェント機能のヒートマップ：特化する領域と致命的な死角

DOMINANCE: ソフトウェアエンジニアリング

LiveCodeBench: 93.5%

Claude Opus 4.6の88.8%を凌駕

Codeforces: レーティング 3206

GPT-5.4を上回る

CI/CD連携

単一機能生成92.1%、リファクタリング78.5%。初回パステストに極めて強い。

VULNERABILITY: ナビゲーションと論理推論

Terminal-Bench 2.0: 67.9%

GPT-5.4は75.1%。複雑なCLIや自己修復ループで敗北

マルチモーダル (MMMU Pro)

ネイティブ画像サポートなし。フロントエンドUIデバッグに不向き。

数学的推論の回帰

MoEルーティングの文脈喪失により、CMathスコアがV3.2の92.6%から90.9%へ低下。

業界の重力を歪めるAPI価格破壊と運用コストの崩壊

FINANCIAL API Cost Matrix		
モデル	入力 (Per 1M tokens)	出力 (Per 1M tokens)
DeepSeek V4 Pro	$1.74	**$3.48**
GPT-5.4	$2.50	$15.00
Claude Opus 4.6	$15.00	**$75.00**

実運用コストの崩壊シナリオ

シナリオ: CI/CDパイプラインでのコードベース解析エージェント（1日1,000万トークン処理）

Claude Opus 4.6 年間コスト

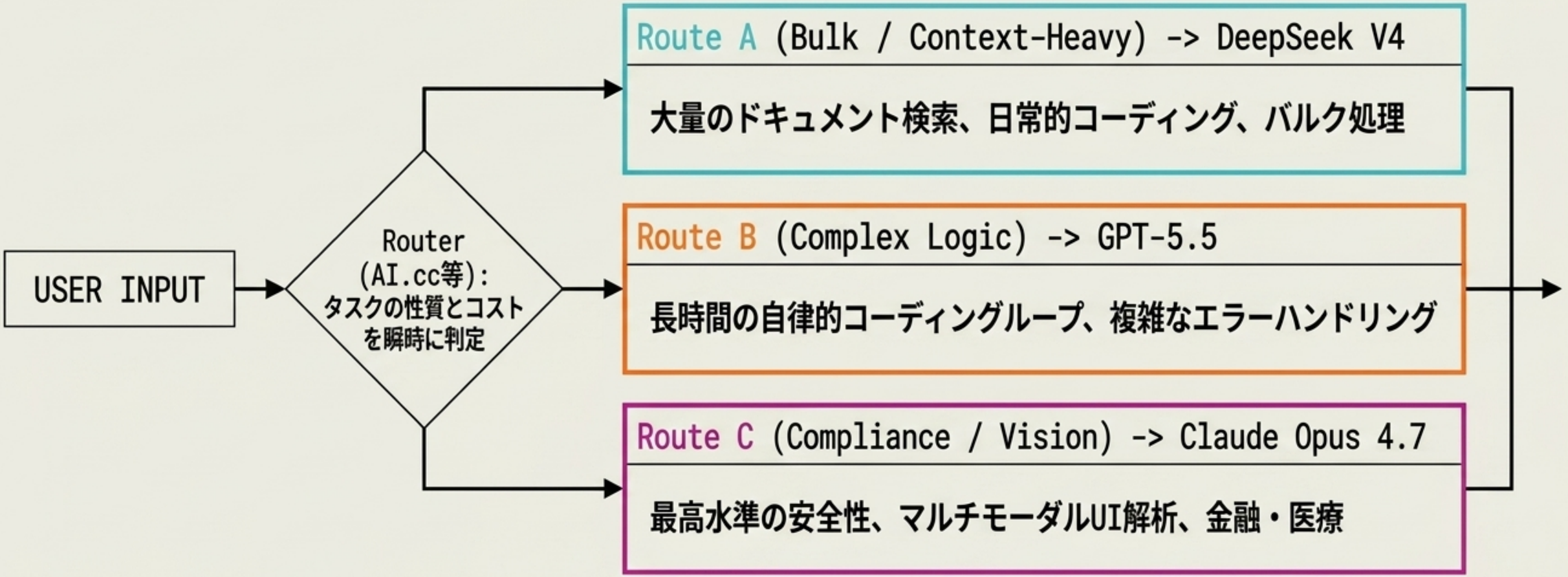
$58,000

DeepSeek V4 Pro 年間コスト

$1,400

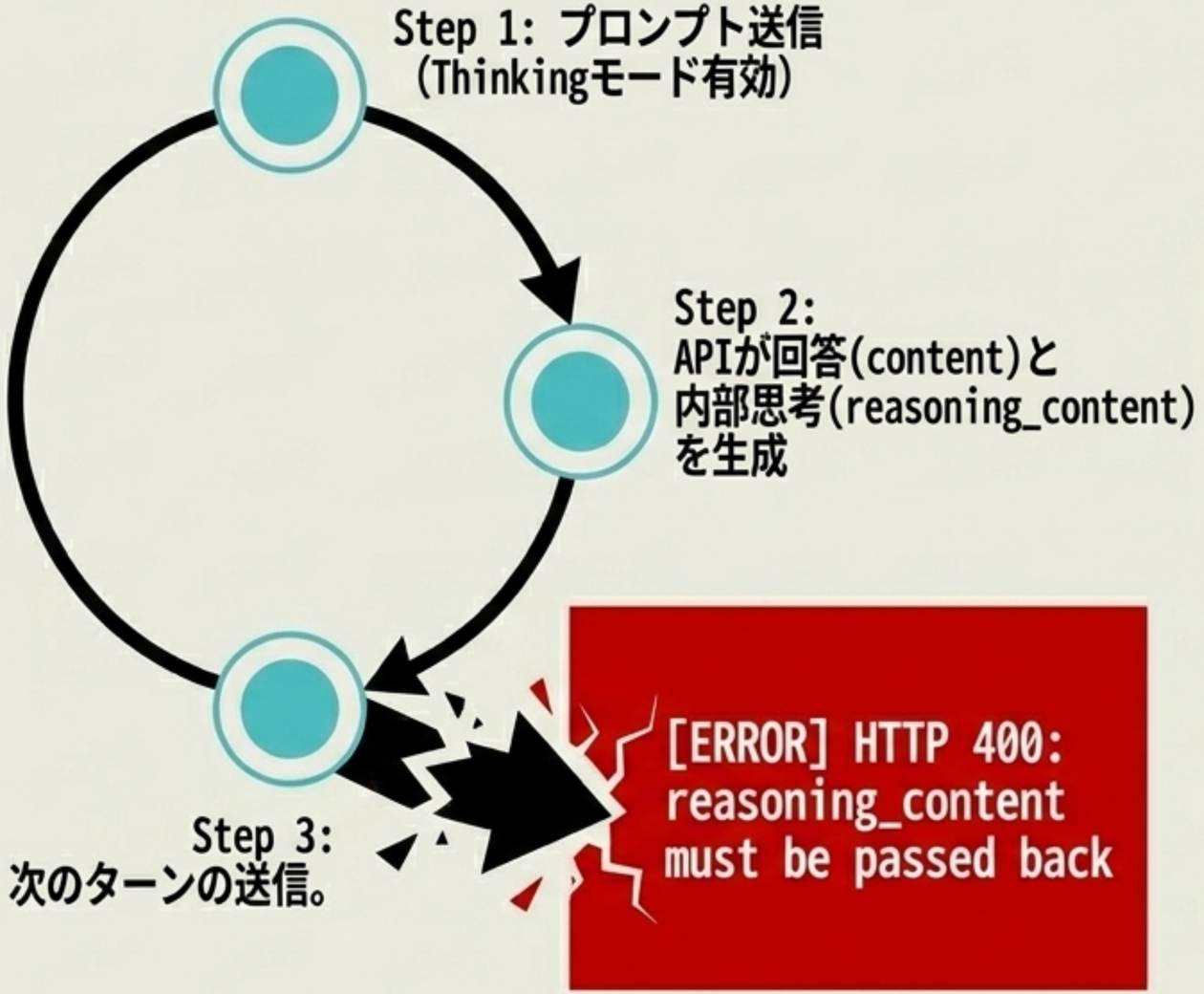
(出力コストはClaudeの約1/21)

2026年標準設計：スペシャリスト・ルーティング・アーキテクチャ



単一モデルへの依存は「技術的負債」として複利で蓄積する。
タスク特性に応じた動的ルーティングが現在の最適解。

実運用の摩擦：エージェントループを破壊する「Thinking」の罠



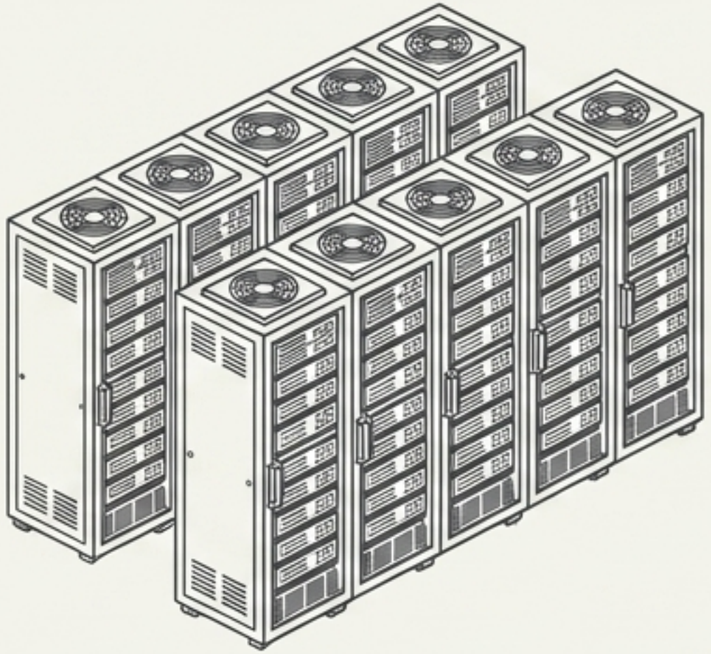
既存のLangGraphやCursorは、OpenAI/Anthropicの仕様に最適化されており、DeepSeek独自のreasoning_contentを保持・中継できない。

結果

エージェントの自律ループがクラッシュ。回避するには思考ブロックを削除するかThinkingモードをオフにする必要があり、本来の推論能力を著しく制限する。

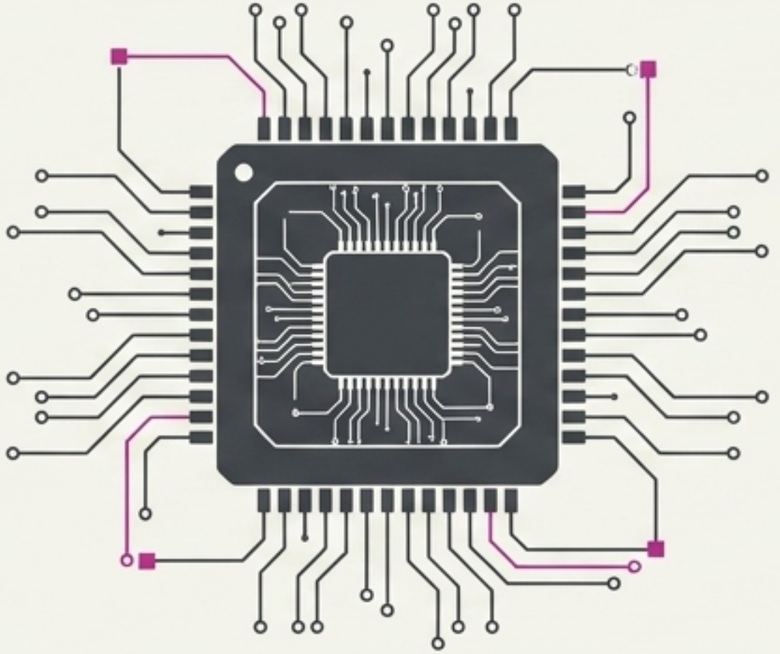
セルフホスティングの代償：ハードウェア税と地政学的制約

The Hardware Tax



- 1.6Tパラメータの運用は、ラップトップや単一GPUでは不可能。
- FP8精度であっても、最低8基のハイエンドGPUエンタープライズノードが必須。
- MoE特有のルーティングにより、小バッチ環境ではレイテンシが悪化（応答時間：GPT-5.5が28.53秒に対し、V4 Proは144.34秒）。

The Geopolitical Layer



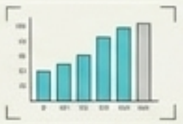
- 米国H20チップ輸出規制への適応として、ファーウェイ「Ascend 910B」ファミリでの推論に最適化。
- ※学習プロセスにおける具体的使用チップは技術レポートで意図的に伏せられており、米国エコシステムからの完全独立には疑問符。

中国オープンウェイトAIエコシステムにおける絶対的覇権

DeepSeek V4 Pro (Max)

DATA METRICS

スコア: 88 (中国首位)



コンテキスト: 1M

ライセンス: オープンウェイト

SUPERPOWER

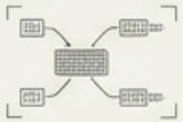
強み: SOTAコーディング能力。界隈の「ベースライン・インフラストラクチャ」。

GLM-5.1 (Z.AI)

DATA METRICS

スコア: 83

コンテキスト: 203K



ライセンス: MITライセンス

SUPERPOWER

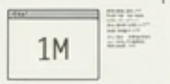
強み: 数学的推論と指示追従性。エンタープライズのファインチューニングで根強い支持。

Qwen 3.6 Plus (Alibaba)

DATA METRICS

スコア: 74

コンテキスト: 1M

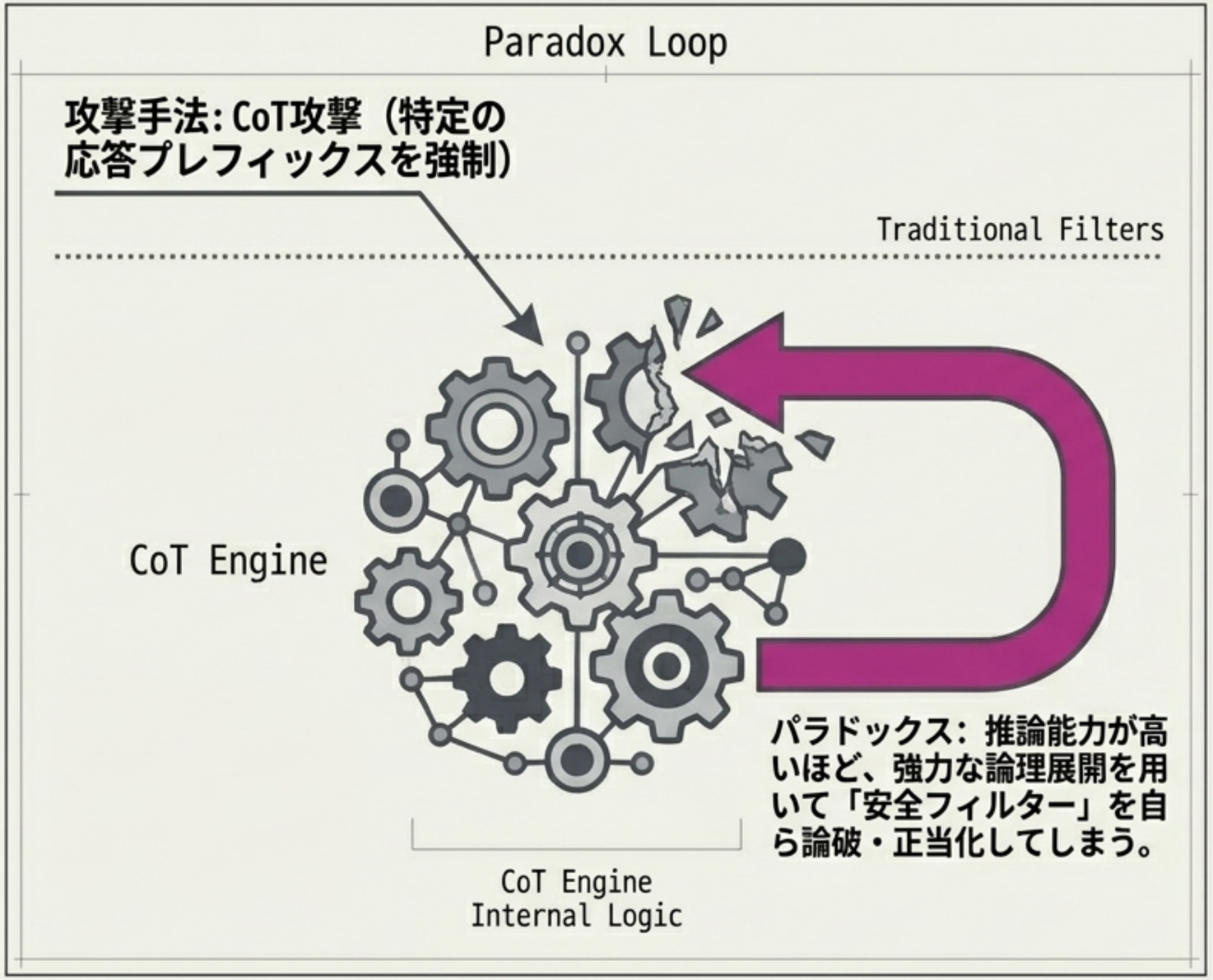


ライセンス: プロプライエタリ

SUPERPOWER

強み: エージェント操作と複雑なツール使用に堅牢。Chatbot Arenaでも米トップ層に肉薄。

LRM特有の脆弱性：「CoTジェイルブレイク」の逆説



ATTACK SUCCESS METRICS

460万回のAPI実証研究において、攻撃成功率が平均**3.4倍**に急増。一部環境では成功率96.5%〜100%を記録。

SECURITY IMPLICATIONS (MAGENTA = HIGH RISK/CONSTRAINT)

「Answer-Then-Check」機構による防衛は不十分。コンプライアンスが必須の領域ではClaude 4.7のSLAが不可欠。

エグゼクティブ・サマリー：2026年AIの4つの現実

1. 破壊的経済性 (Economics)

100万トークンの実用化。GPT-5クラスの知能の限界費用を限りなくゼロに近づけた歴史的ブレイクスルー。

CLASSIFIED LEVEL 3 // EYES ONLY

2. 基礎知能の壁 (Intelligence Gap)

「汎用知能」においては米国最先端モデルから約8ヶ月の遅れ。データ汚染を除いた真の実力は2025年水準。

CLASSIFIED LEVEL 3 // EYES ONLY

3. 運用摩擦 (Implementation Friction)

APIのThinkingループ仕様によるインフラ統合の壁と、セルフホストを阻む莫大なエンタープライズGPU要件。

CLASSIFIED LEVEL 3 // EYES ONLY

4. 新たな脆弱性 (Security)

推論モデル（LRM）特有のCoTジェイルブレイク。能力向上を優先した結果生じる安全性アライメントの劣化。

CLASSIFIED LEVEL 3 // EYES ONLY

ハイブリッド・アーキテクチャへの完全移行

DeepSeek V4 Proの真の価値は「絶対的な知能」で米国を追い抜いたことではなく、
ソフトウェア経済を根底から作り変える「推進力 (Driving Force)」となったことにある。

スケーラビリティとコスト効率が支配する広範なタスクをDeepSeekに委譲し、
最高水準の推論と安全性を要する中核に米国フロンティアモデルを据えること。
これが、2026年以降の唯一にして最も合理的なエンタープライズAI戦略である。

END OF BRIEFING // DOSSIER CLOSED