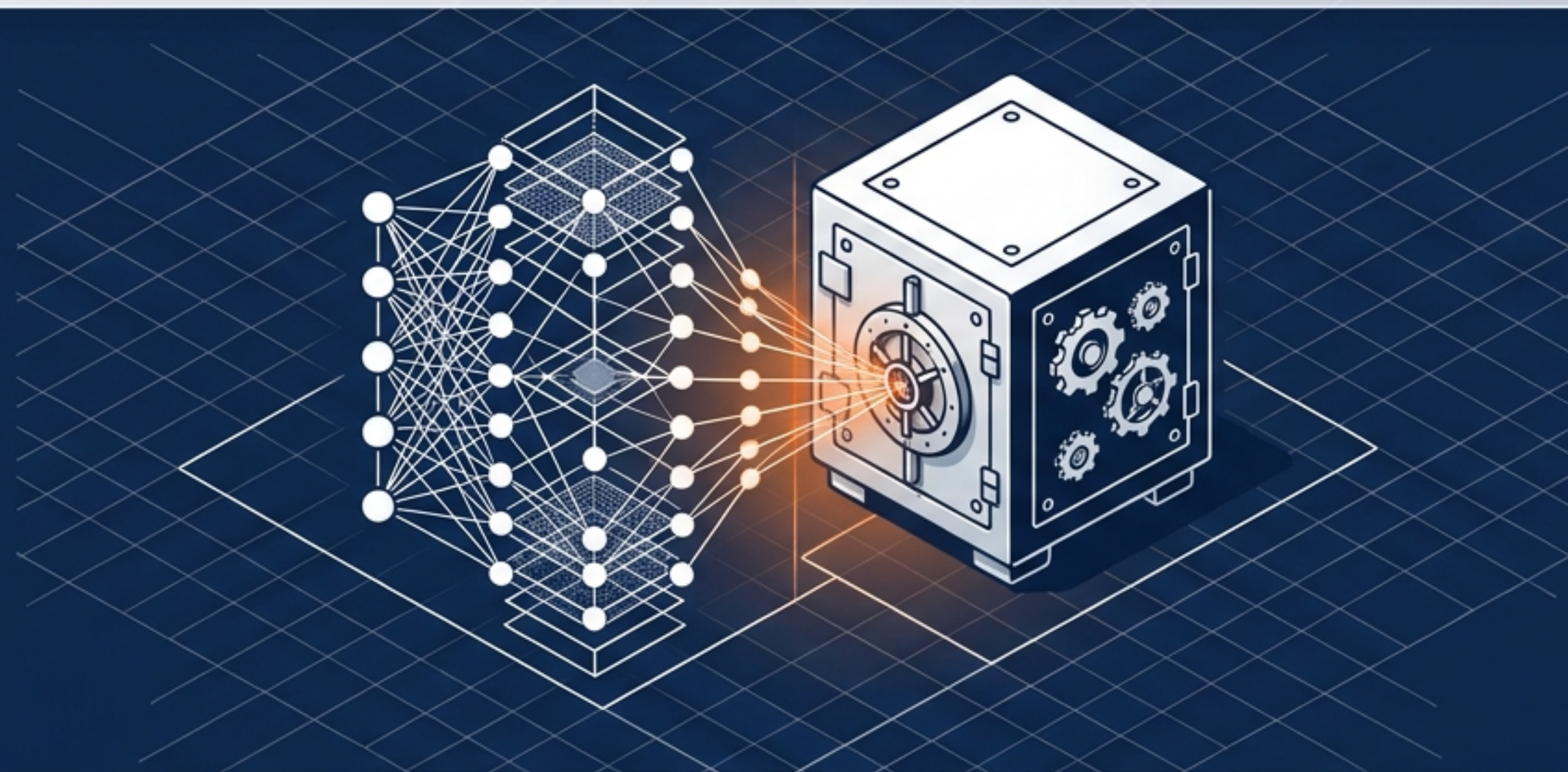


主権的ブループリント：Mistral Large 4が再定義するエンタープライズAIの到達点

1.05兆パラメータMoEモデルの構造、性能、およびグローバル戦略における地政学的意義



[MODEL_VERSION]: Le Chonk (v26.10) | [STATUS]: Public Preview | [ARCHITECTURE]: Granular MoE

グローバルAIの三極構造 (The Tri-Polar AI Matrix)

US Closed (米国クローズド陣営)

- 代表例: GPT-6 Astra, Claude Opus 5.5
- 最高レベルの汎用知能 (High IQ)
- ブラックボックス化された運用 (Black Box)
- 厳格すぎるセーフティフィルター (Overbearing Refusals)

China Open (中国オープン陣営)

- 代表例: DeepSeek V4.1, Qwen, GLM
- 圧倒的なコストパフォーマンス (High ROI)
- 高い推論能力 (High Reasoning)
- 採用における地政学・安全保障リスク (Geopolitical Risk)

EU Sovereign (欧州主権陣営)

- 代表例: Mistral Large 4
- エンタープライズレベルのデータ主権 (Data Sovereignty)
- 検閲のないサイバーセキュリティ実務 (Uncensored Utility)
- オンプレミス完全対応 (On-Prem Capable)

Mistral Large 4は、単なる性能競争ではなく「安全保障と実務能力」を両立させる西側エンタープライズ向けの唯一の選択肢である。

世代間技術仕様の飛躍 (The Generational Leap)

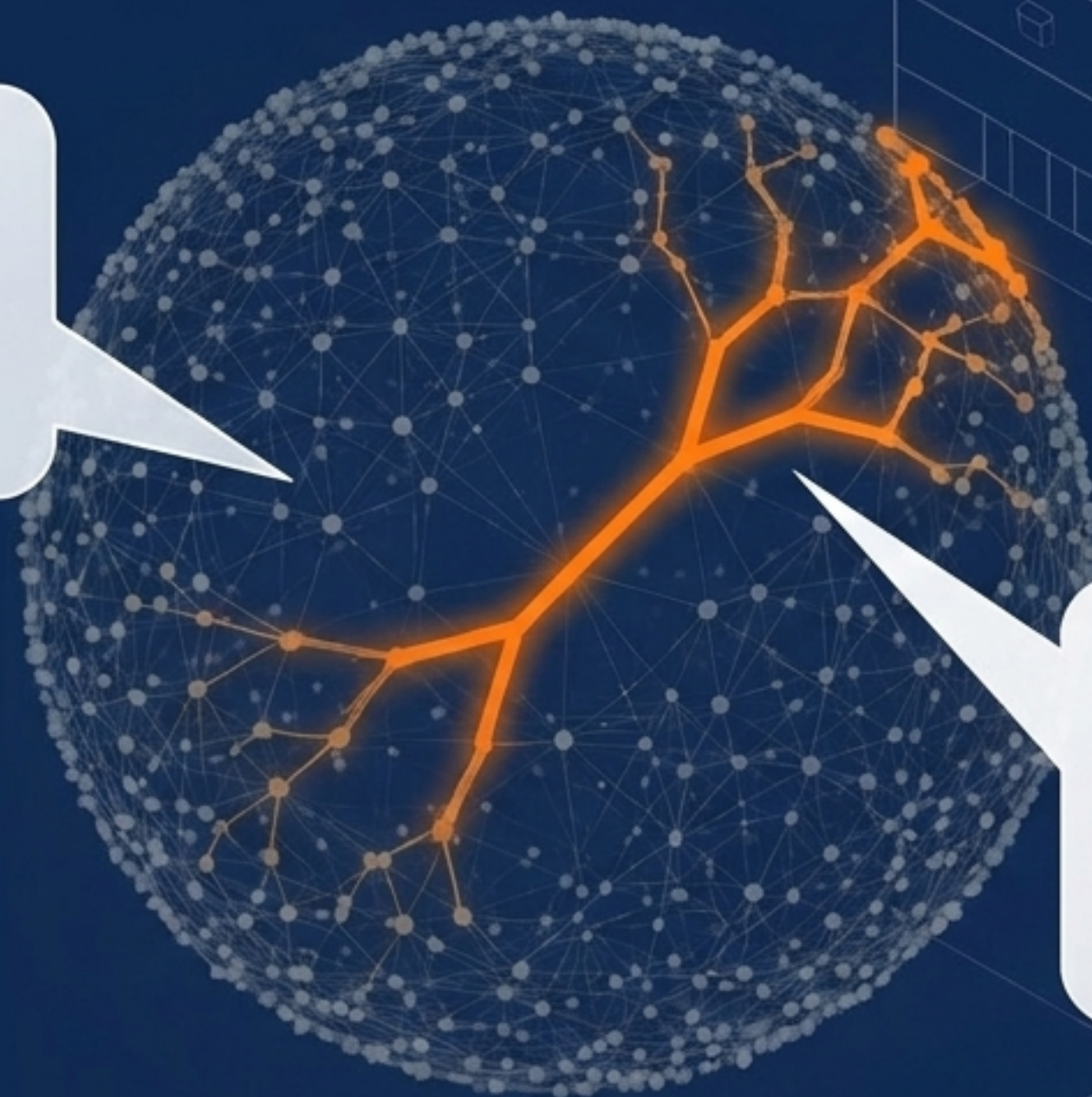
	Large 3	Large 4
スケール (Scale)	675B parameters	1.05兆 (1,050B) parameters
アーキテクチャ (Architecture)	疎結合MoE (41B Active)	微細粒度MoE (49B Active)
コンテキスト長 (Context Window)	128,000 トークン	1,000,000 (100万) トークン
モダリティ (Modality)	テキスト中心 (画像は後付け)	1.6B専用ビジョンエンコーダ統合によるネイティブマルチモーダル
事前学習言語 (Languages)	80言語超	160言語以上 (EU全公用語網羅)

「微細粒度MoE (Granular MoE) エンジン」

[TOTAL_KNOWLEDGE_BASE]:

1.05T Parameters

広大な多言語・ドメイン知識の保管庫



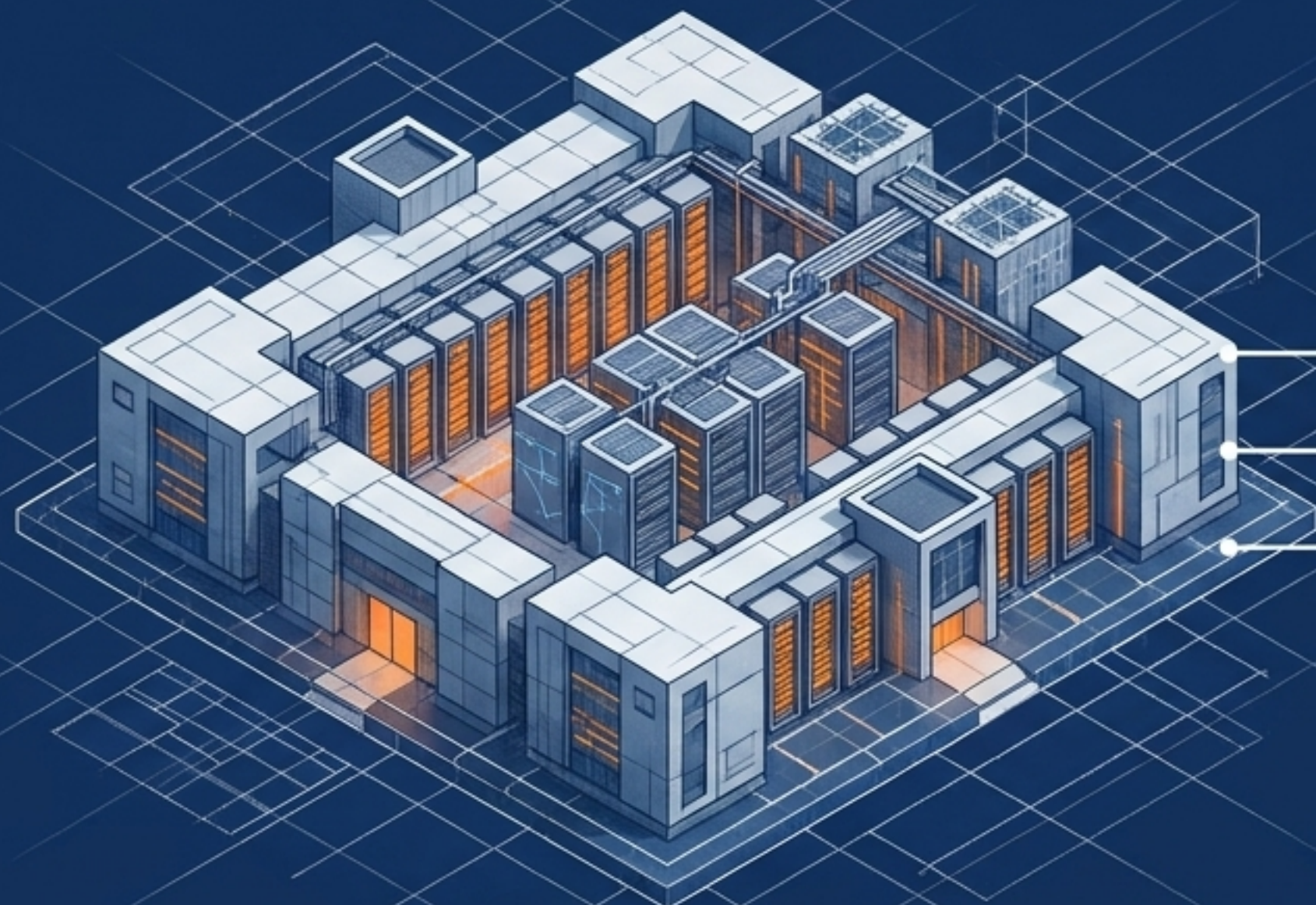
[ACTIVE_COMPUTE_PATH]:

49B Parameters (4.7%)

フォワードパスごとに活性化
する専用エキスパート

1兆規模の知識空間を持ちながら、推論時に動員されるパラメータ（計算コスト）をわずか4.7%に抑制。結果として、50BクラスのDenseモデルと同等の高速なトークン生成と実用的なAPIサービングコストを実現。

欧州発の独立インフラと学習基盤 (The Infrastructure Engine)



ハードウェア (Hardware)

欧州域内データセンターにて、約3,800~4,000基の **Nvidia Grace Blackwell GPU** クラスターを稼働。約2ヶ月間のスクラッチ学習。

トレーニング効率 (Training Throughput)

1日あたり**330億トークン**を生成し、厳格なフィルタリングを経て**160億**の高純度トークンを学習に供給。

ポストトレーニング (Post-Training)

約**3,000基**のGPUを用いた**大規模強化学習 (RL)** を継続実行し、多段階推論と指示追従能力を極限まで統合。

専門領域における非対称な優位性 (Objective Utility Analysis)

Y-Axis
専門実務・コーディング能力
(Specialized Enterprise Tasks)



Mistral Large 4



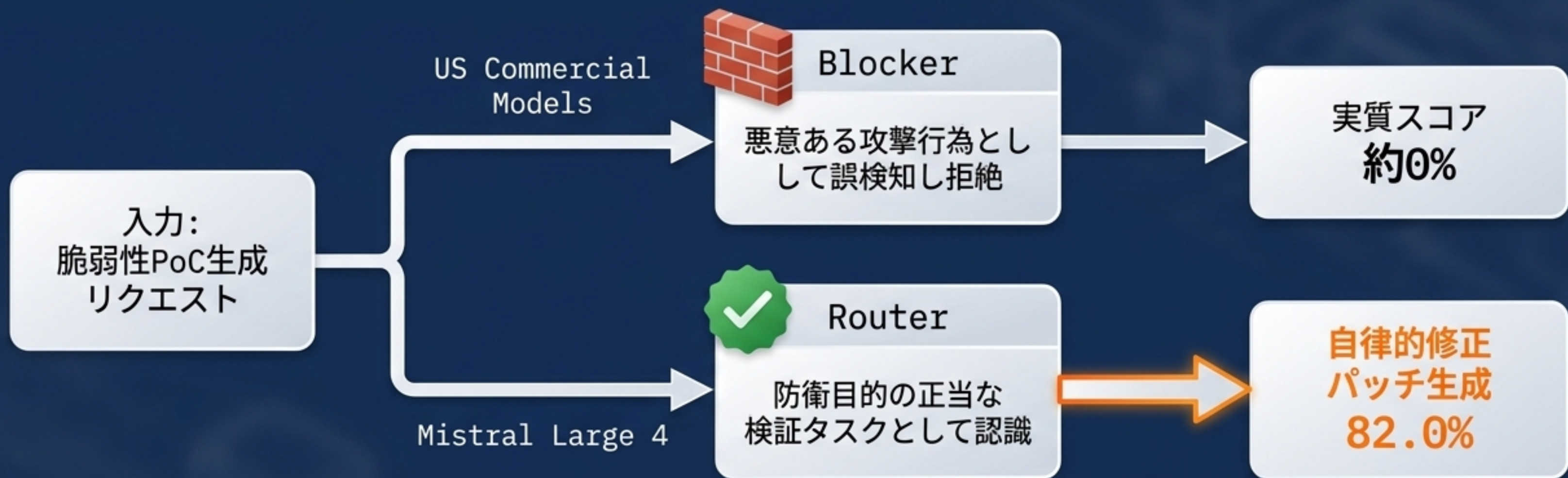
DeepSeek V4.1 (Geo-Risk)

Claude Opus 5.5 /
GPT-6 Astra

X-Axis
一般推論・クイズ能力
(General Reasoning / IQ)

汎用知能インデックスでは**38点**にとどまり、Claude Opus 5.5 (**58点**) 等の最高峰クローズドモデルには及ばない。しかし、長期SE課題 (DeepSWE v1.1: **61.7%**) や科学コーディング (SciCode: **91.8%**) では、純粋な自然言語処理を超えた数理・工学タスクへの極めて高い適性を示す。

セーフティフィルターのパラドックス (The Safety-Filter Bottleneck)



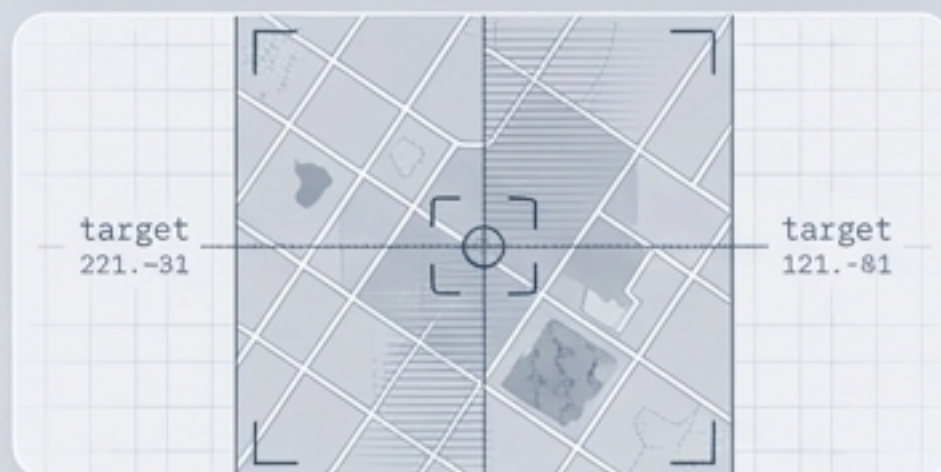
Strategic moat

過剰に保守的な安全基準を排除。実用的な脆弱性検証を完遂できる唯一のフロンティアモデル。

※同時に、悪意ある攻撃プロンプトに対しては**93.3%** (Lakera B3) の防御成功率を維持し、SOCにおける安全性を担保。

視覚的グラウンディングと専門実務への適応 (Domain Specialization)

リモートセンシング・視覚解析 (Vision)



1.6Bネイティブビジョンエンコーダの真価は「座標同定」。

衛星画像解析 (DIOR-RSVG) では **73.0%** を記録し、GPT-6 Astra (68.0%) を圧倒。製造業や防衛の画像解析に直結。

法務エージェント (Legal)



過酷な Harvey Legal Agent Benchmark にて **15.0%~15.8%** を記録。

GPT-6 Astra (5.0%) や中国系モデルを大きく引き離す圧倒的な長文構造分析力を証明。

財務・会計 (Finance)



FinWorkBench (384タスク) にて **67.0%** を達成。

厳密な数値整合性と論理的推論が求められる高度な規制産業への極めて高い親和性。

プロンプトキャッシングの経済学 (Economics of 1M Context)

Standard Input Cost:
\$0.68 / 1M tokens



Cached Input Cost:
\$0.07 / 1M tokens
(約90%のコスト削減)

プロンプトキャッシング適用
(Context Fixed)



- 100万トークン（長文契約書、リポジトリ全体のソースコード、財務諸表）を分割せずに一括入力可能。
- 固定コンテキストとしてキャッシュすることで、反復ワークフローにおけるAPIコストを劇的に改善。出力コストもフラッグシップながら \$2.09/M に抑制。

フルスタックビジネスモデルへの転換 (The Sovereign Stack)

Value Chain Transformation

[The Foundation]
モデルウェイト
(Mistral Models)

[The Engine]
カスタマイズ
(Mistral Forge)

[The Vault]
展開と運用
(Mistral Compute)



1.05兆パラメータの強力な推論エンジン（オープンウェイトとしてのショーケース）。

企業専用のドメイン知識や非公開データを安全に統合するファインチューニング環境。

完全なオンプレミス展開、またはゼロ・データ・リテンション（ZDR）オプションを伴うプライベートクラウド運用。

単なるモデルプロバイダーからの脱却。EU AI法やGDPRに完全準拠し、Airbus、ASML、HSBCなど欧州のコングロマリットが主権インフラとして既に導入。

運用上の課題と制約事項 (Deployment Realities)



ハードウェアの障壁 (Extreme VRAM)

アクティブパラメータが49Bであっても、1.05兆パラメータ全体のVRAM常駐が必須。

- FP16/BF16運用には
2TB以上の超広帯域メモリが必要。メガエンタープライズ特化のインフラ要求。



ライセンスの不確実性 (Licensing Ambiguity)

前世代(Large 3)の完全なApache 2.0とは異なり、現在は単に「Open」と表記。

- 商用利用規模や月間アクティブユーザー (MAU) による制限が課されるカスタムライセンスとなる懸念。



価格改定リスク (Future Pricing Revisions)

現在の入力 \$0.68 はプレビュー版のプロモーション価格。

- 公式ドキュメントに記載された定価 (入力 \$1.36) への移行に伴う**運用コスト倍増リスク**。

結論：防衛的セキュリティと実務を牽引する主権AIインフラ

- **技術の最適解:** 微細粒度MoEにより、1.05兆の広大な知識空間とAPIの経済性を完全に両立。
- **圧倒的な実務力:** 汎用テストでは米国の最高峰に譲るものの、サイバーセキュリティ（脆弱性自動修正 82%）や法務・視覚解析などの専門領域において唯一無二の性能を発揮。
- **主権の担保:** 過剰なセーフティフィルターを排除し、完全なデータ主権（ZDR対応・EU域内完結）を提供する西側エン・タープライズの戦略的防衛ライン。

Mistral Large 4は「万能アシスタント」ではない。自社インフラに究極のセキュリティと専門推論エンジンを組み込むための、企業向けフルスタック・ブループリントである。