

韓国「独自AI基盤モデル」第2次評価の深層：ソブリンAI戦の略の転換点

グローバル首位の脱落とSKテレコム総合首位から読み解く、
国家AIインフラの新たな設計図



グローバル首位 モデルの脱落

ベンチマークで圧倒的1位（世界6位水準）のMotifが最下位で脱落。SKテレコムが総合70.6点で首位を獲得し、評価基準の妥当性が問われる事態に。



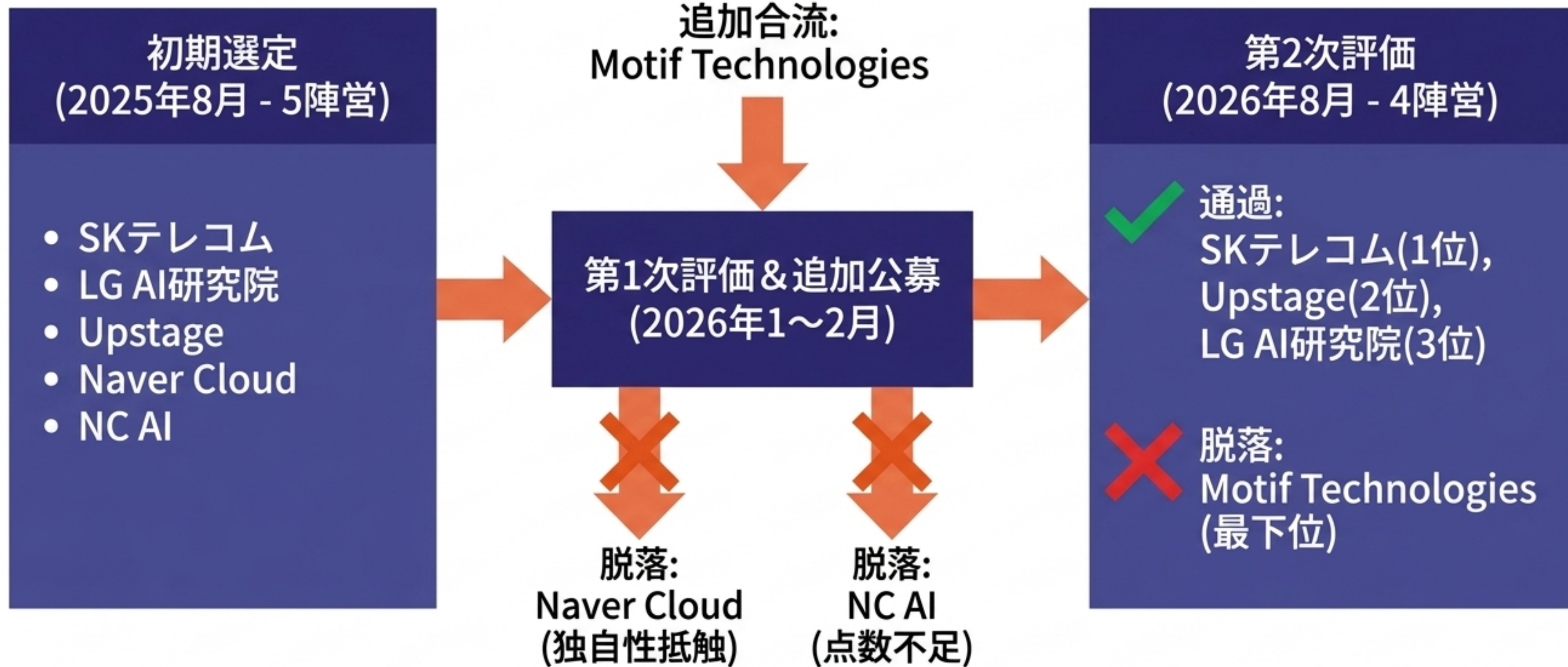
「純粹性能」 vs 「現場実働力」

基礎推論力だけでなく、通信・製造・防衛等の「実産業への即時統合力」と「軽量・高効率化」がソブリンAIの勝敗を分けた。



トーナメントから 「面的インフラ」へ

評価論争を契機に、政府は特定モデルの選抜から、約3.8兆ウォン規模のGPUインフラとデータ拡充を伴う「フロンティアAIエコシステム構築」へと政策を転換。



追加公募の救世主として合流したMotifが、
わずか半年後に波乱の主演となる。

World #6 Non-US/China #1

Last Place (65.8 pts)

なぜ、
グローバル
基準の勝者が
ローカル評価で
敗れたのか？

Motif 3は、グローバル指標
(AAII) で47点を獲得。オープン
ウェイトモデルとして世界6位、
米中以外の国家開発モデル
として世界トップの性能を証明。
Epoch AI 「注目すべきAI」 にも選出。

第2次評価において、SKT (70.6点)
などに敗れ総合最下位で脱落。

「世界最高水準のAI」が、「自国の
国家プロジェクト」から排除される
というパラドックスが発生。

AAII (グローバル) NIA (国内) 専門家評価 AIユーザー 一般国民

SKテレコム
(70.6点 - 1位)



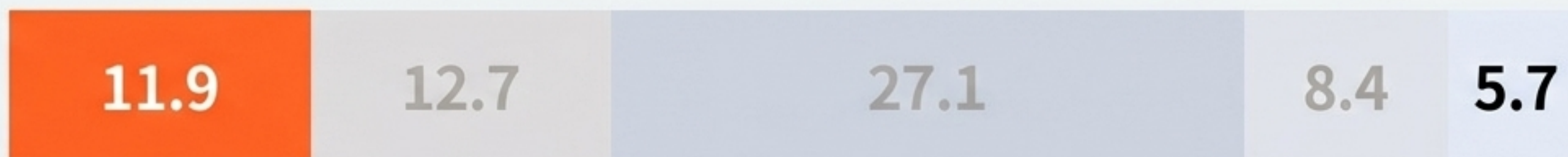
Upstage
(69.9点 - 2位)



LG AI研究院
(69.0点 - 3位)



Motif
(65.8点 - 脱落)



総合得点
70.6点
(1位)
現場実働型AI

純粋推論・論理性能 (Score: NIA 1位)

国際数学五輪 (AIME 2026) 基準で97.1点。
韓国語知識 (KMMLU-Pro) 80.5点、
文化理解 (CLiCk) 91.6点。

推論・コンテキストの高効率化

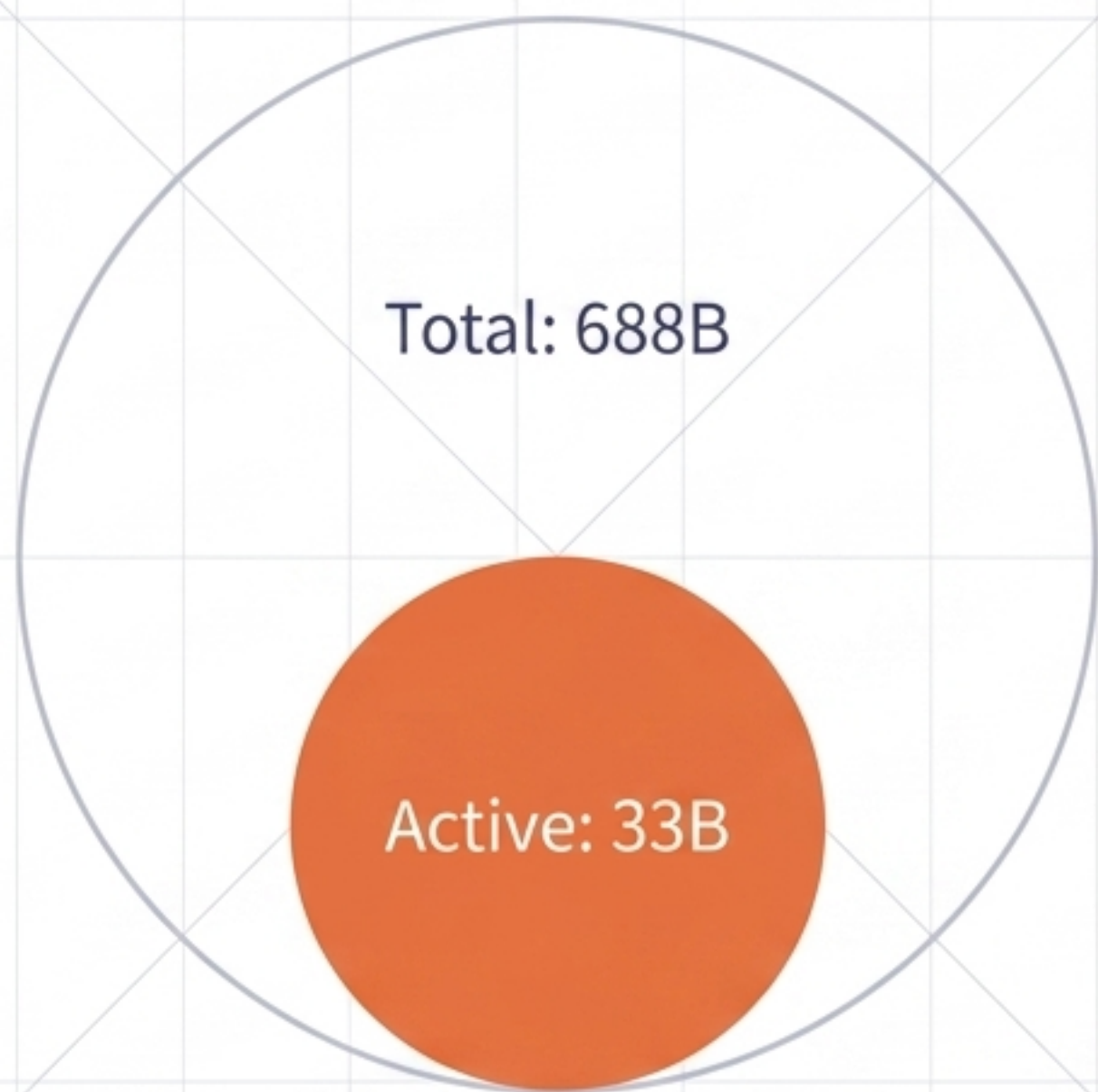
長文処理時の計算効率とスループットの維持。
通信・エージェント能力指標 (T-Bench) で
98.0点の圧倒的性能。

産業現場への即時実装力 (Score: AIユーザー 1位)

基礎研究を商用トラフィックへ即座に変換する
パイプラインの完成度。
製造、防衛、通信分野への垂直統合。

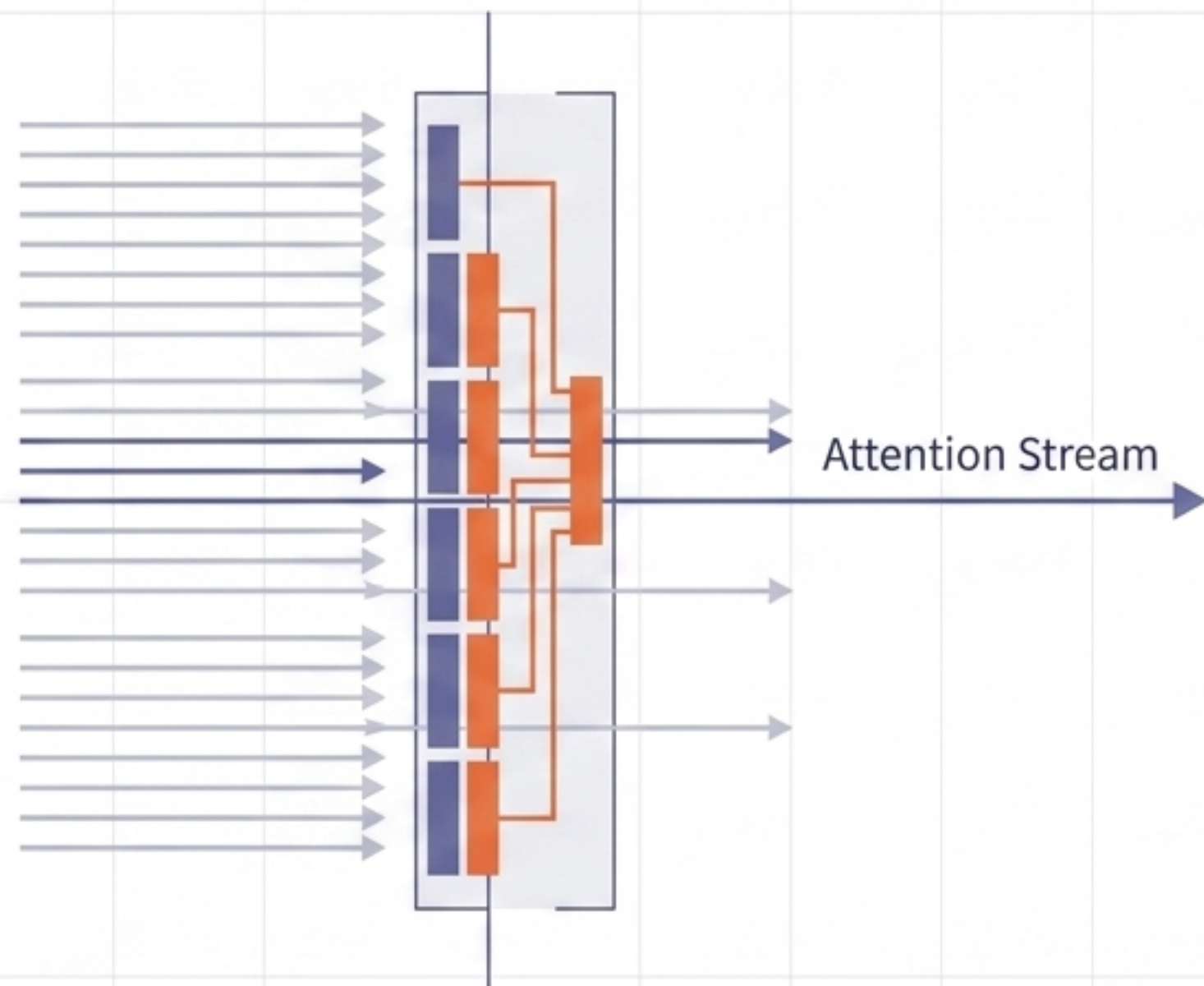
国家が求めたのは「テストの点数」ではなく、「社会基盤として稼働するシステム」であった。

MoE (Mixture of Experts) 構造



総パラメータ6,880億のうち、推論実行時はわずか330億のみを稼働。巨大モデルの性能と低遅延を両立。

SGA (Sparse Gated Attention)



SKT独自構造。大規模コンテキスト処理時のメモリ占有を抑制し、通信網でのリアルタイム処理（低遅延）を実現。

審査において「現場拡散実績」
として最高評価を獲得

製造業 (Manufacturing)

ターゲット：KGスチール、コーネック

不良事例データの原因分析と是正措置を自律提示する
「製造特化エージェント」の実証。

防衛・公共 (Defense / Public)

ターゲット：韓国国防部（閉鎖網環境）

データ主権を担保するオンプレミス環境向けの
軽量量子化モデル（FP8/FP4/INT4）の提供。

エンタープライズ&B2C (B2B / B2C)

ターゲット：法人向け「A.Biz」、個人向け「A. (エイドット)」

自社サービスへの即時組み込みと商用トラフィック化。

A.X K2
基盤モデル

Motifの主張 (The Challenger's Logic)

- スコア圧縮の不公平 : AAllの生点 (47点) の圧倒的優位が、線形換算によりわずかな差に圧縮された。

(Unfair Score Compression: AAll's overwhelming raw score (47 pts) was compressed into a marginal difference by linear conversion.)

- 定性評価の偏重 : 汎用推論力よりも、既存大手コンソーシアムに有利な「現場拡散実績」や「チャットボットの使用感」に審査が偏っている。

(Qualitative Bias: Evaluation was biased towards 'on-site deployment track record' and 'chatbot usability' favoring existing major consortiums, rather than general reasoning ability.)



政府の論理 (The Sovereign Logic)

- 事前の合意: 評価基準は事前に全陣営と合意済みであり、公正に執行された。

(Prior Agreement: Evaluation criteria were agreed upon beforehand by all parties and executed fairly.)

- 国家要件: ソブリンAIは公開ベンチマークのスコア稼ぎではなく、行政・公共への「適合性」と「安全性」が必須である。

(National Requirement: Sovereign AI is not about chasing open benchmark scores, but 'suitability' and 'safety' for administrative/public use are essential.)

【結果の現実】 Motifはベンチマーク以外の75点領域（専門家・ユーザー評価）すべてで最下位。挽回不可能な差が存在した。

Raw Data

Converted Score

Motif
生点: 47点

LG AI
生点: 31点

広大なギャップ
(16ポイント差)

25点満点への線形換算

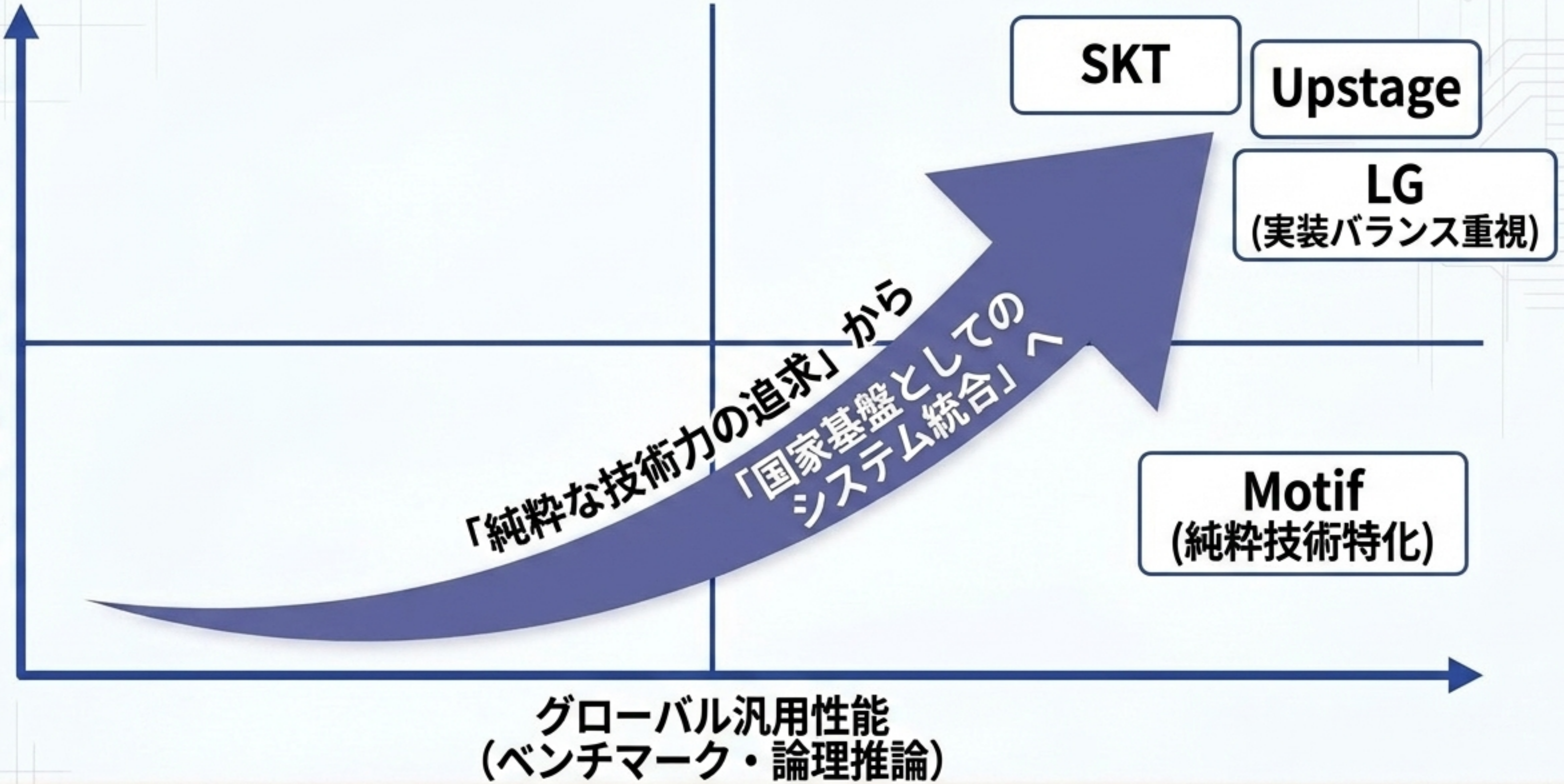
Motif
換算点: 11.9点

圧縮された差
(わずか4.1ポイント差)

LG AI
換算点: 7.8点

グローバル水準での圧倒的な性能差も、換算の枠組みを通すことで、定性評価で容易に覆る差へと縮小された。

ローカル適合・実働性（なら
産業実装・文化理解・安全性）

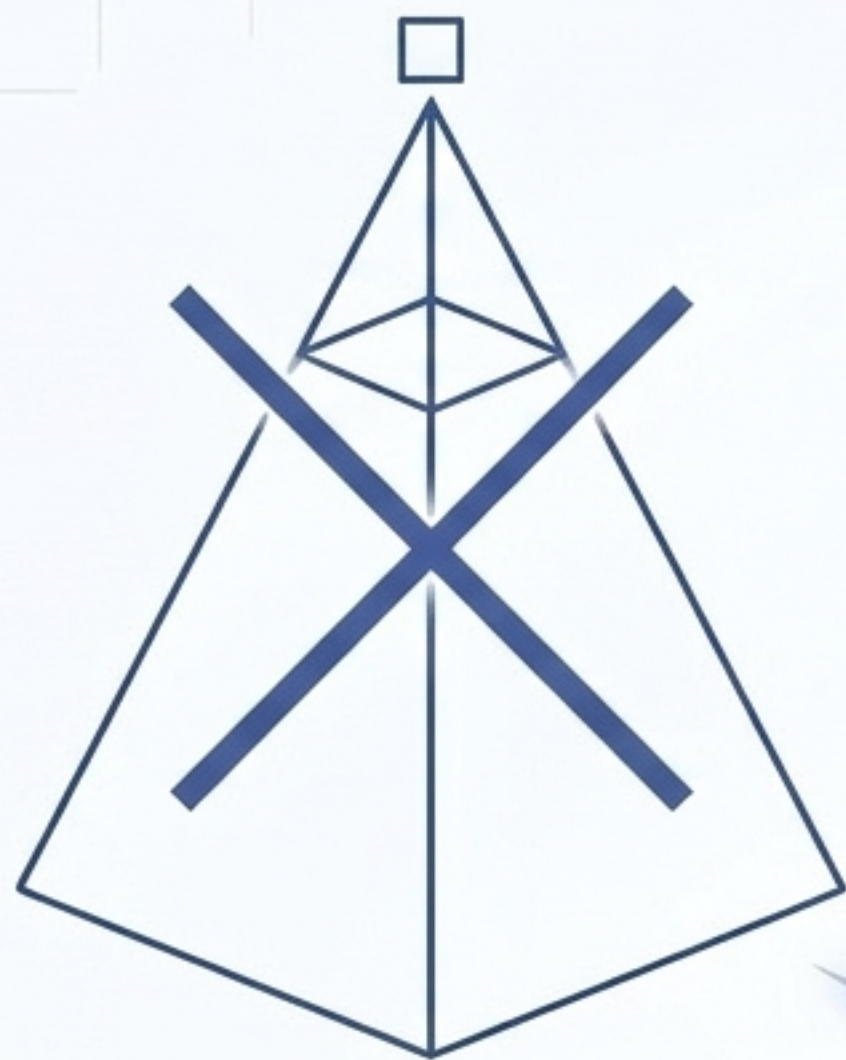


ソブリンAIの評価軸は、「どれだけ賢いか (Intelligence)」から
「社会システムにいかに組み込めるか (Integration)」へ完全に移行した。

グローバル水準での圧倒的な性能差も、換算の枠組みを通すことで、定性評価で容易に覆る差へと縮小された。

	SKテレコム (A.X K2)	Upstage (Solar Open 2)	LG AI研究院 (K-EXAONE 2.0)
規模 / 構造	688B (Active 33B) / MoE + SGA	250B規模 / 高効率MoE	750B (Active 37B) / 国内最大級MoE
戦略的重点	兆単位へのスケールアップ エージェントック 強化学習	圧倒的な コストパフォーマンス	オープンソース戦略 (Apache 2.0) の継続
展開計画	国産AI半導体 (Rebellions等) 活用による 推論コスト最適化	ポータル「Daum」連携、 アジア圏言語拡張による グローバル展開	10言語対応と産業特化 ナレッジによるB2B フロンティア市場攻略

※政府によるNVIDIA B200 GPU支援枠が、1チーム約1,000基（計1,200億ウォン規模）へ拡大。



**過去：特定モデルの選抜
(Tournament)**



**未来：面的インフラの供給
(Frontier AI Policy)**

トーナメント方式の制度的限界を認め、国家全体の計算資源とデータインフラを抜本的に強化する「フロンティアAI」政策へと支援体制を拡大。

政府予算、AI領域へ過去最大の9.4兆ウォン（全体技術予算の約1/3）を投入。

コンピュータ拡充
3.85兆ウォン

次世代アーキテクチャ
(NVIDIA Vera Rubin級)。
約1万基の導入。国家AIコン
ピューティングセンター構築。

データ競争力強化
8,000億ウォン

約1兆トークン規模の
産業特化・学術高品位コーパス
の構築と民間開放。

「国民のAI」事業
2,500億ウォン（新設）

SKT、KT、カカオを通じ、
予約・決済・医療支援を自律実行する
エージェントを全国民へ普及。

グローバルな指標を追う時代から、
国家の基盤産業を駆動する「**実働インフラ**」へ。

ソブリンAIの勝敗は、純粋な技術的優位性ではなく、
実用とエコシステムの構築力で決まる。