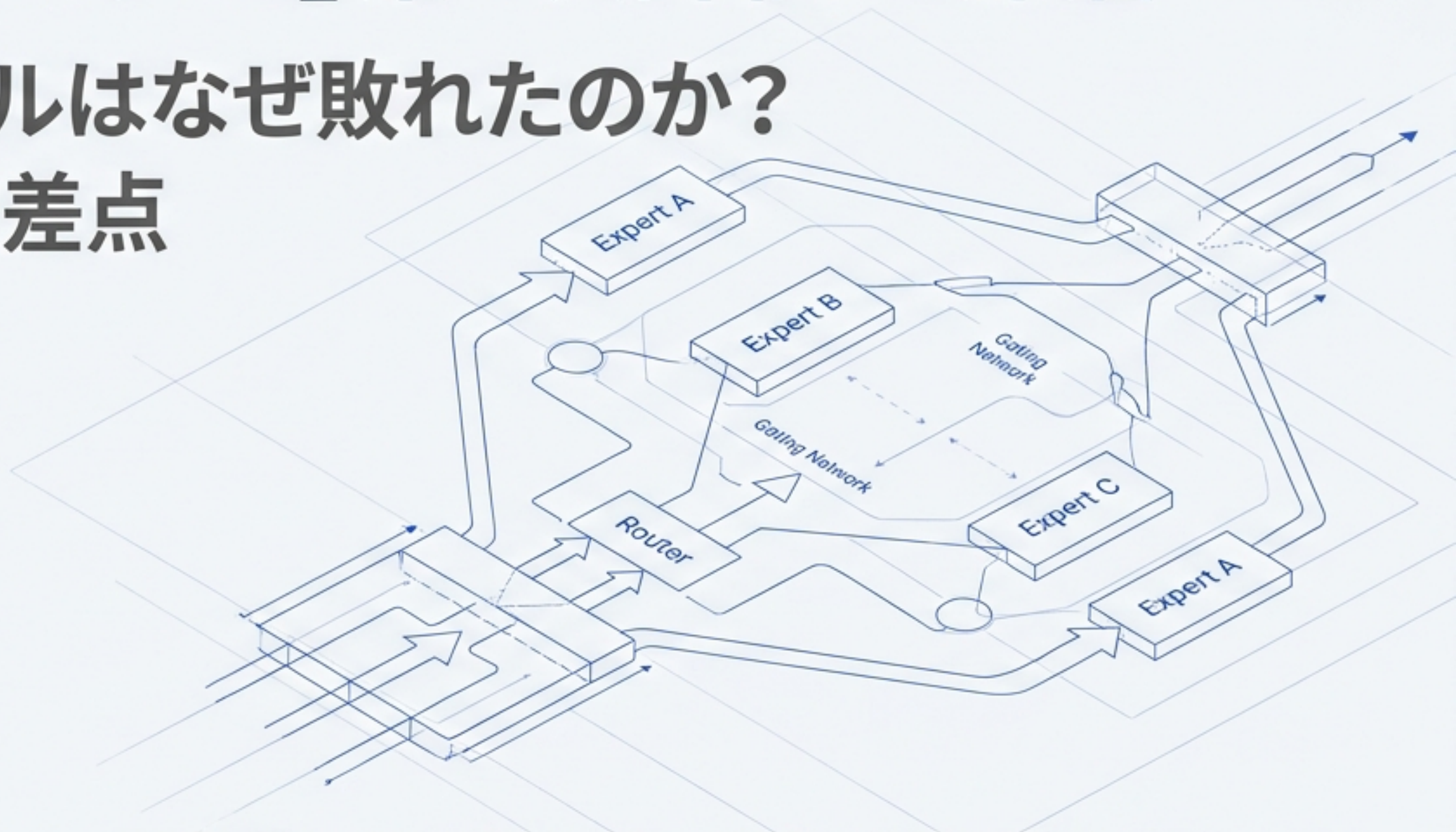


# 韓国「独自AI基盤モデル」第2次評価の深層

ベンチマーク最強モデルはなぜ敗れたのか？

技術力と国家政策の交差点



SKT A.X K2

Upstage Solar Open 2

LG K-EXAONE 2.0

Motif 3

# 報道の見出しと、隠された「4つの勝者」



**総合1位はSKテレコム  
(70.6点)**

**わずか0.7点差での辛勝**

**国際技術力 (グローバル)**

**Motif** (AAll: 11.9点 / 原点 47.4)

国際一般能力で他を圧倒。

**国内知識 (ローカル)**

**SKT** (NIA: 13.4点)

韓国語・国内文脈で僅差のトップ。

**戦略・安全性 (専門家)**

**LG** (専門家: 29.5点 / 一般国民: 7.6点)

戦略、波及効果、一般利用性で首位。

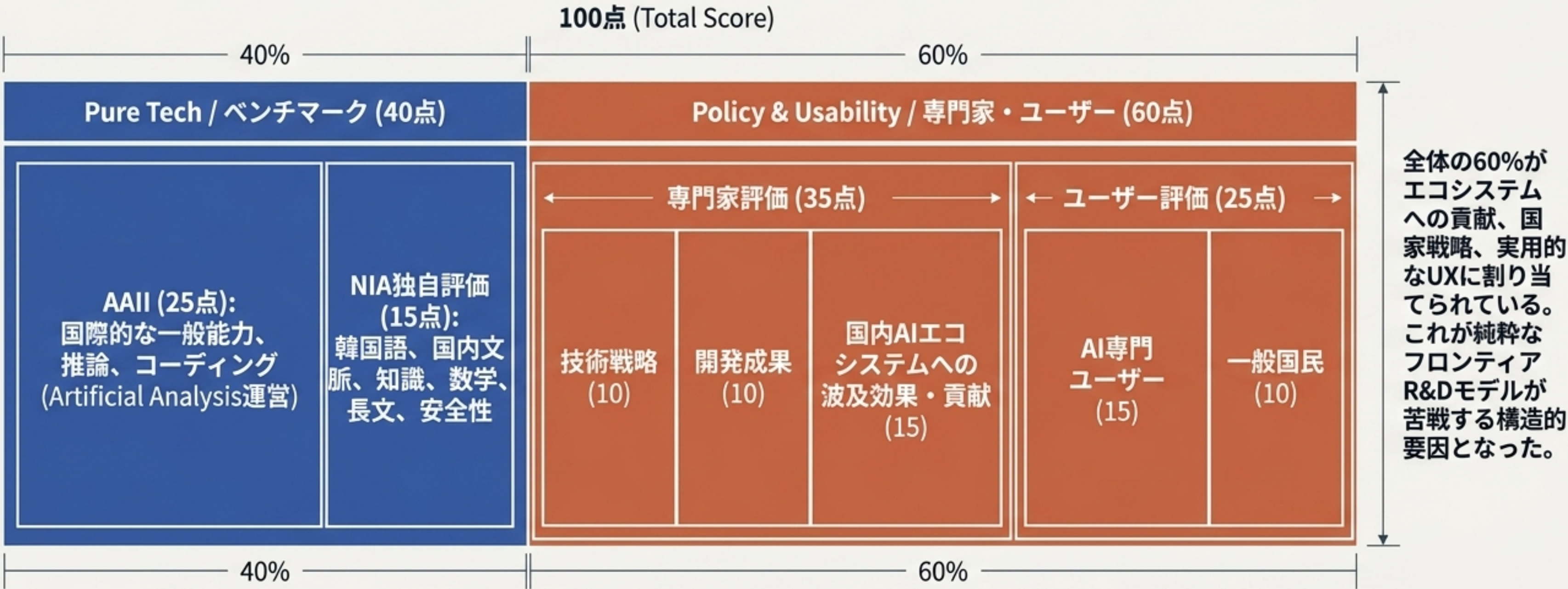
**実用性 (プロユーザー)**

**SKT** (AI専門ユーザー: 11.6点)

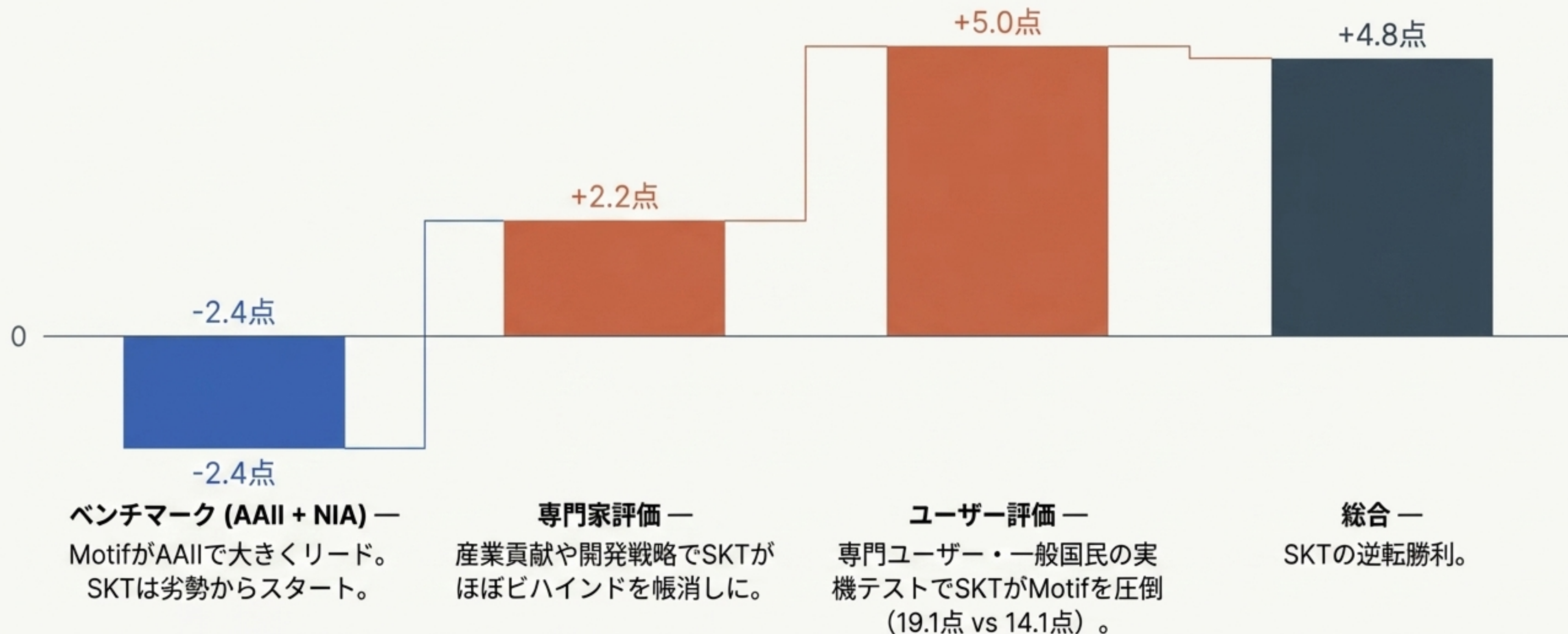
実利用での圧倒的な使い勝手。

**結論：「全性能でSKTが1位」のではなく、  
評価軸ごとに勝者が完全に分散した異例の結果。**

# 「純粋な技術」対「政策・実用性」の採点アーキテクチャ

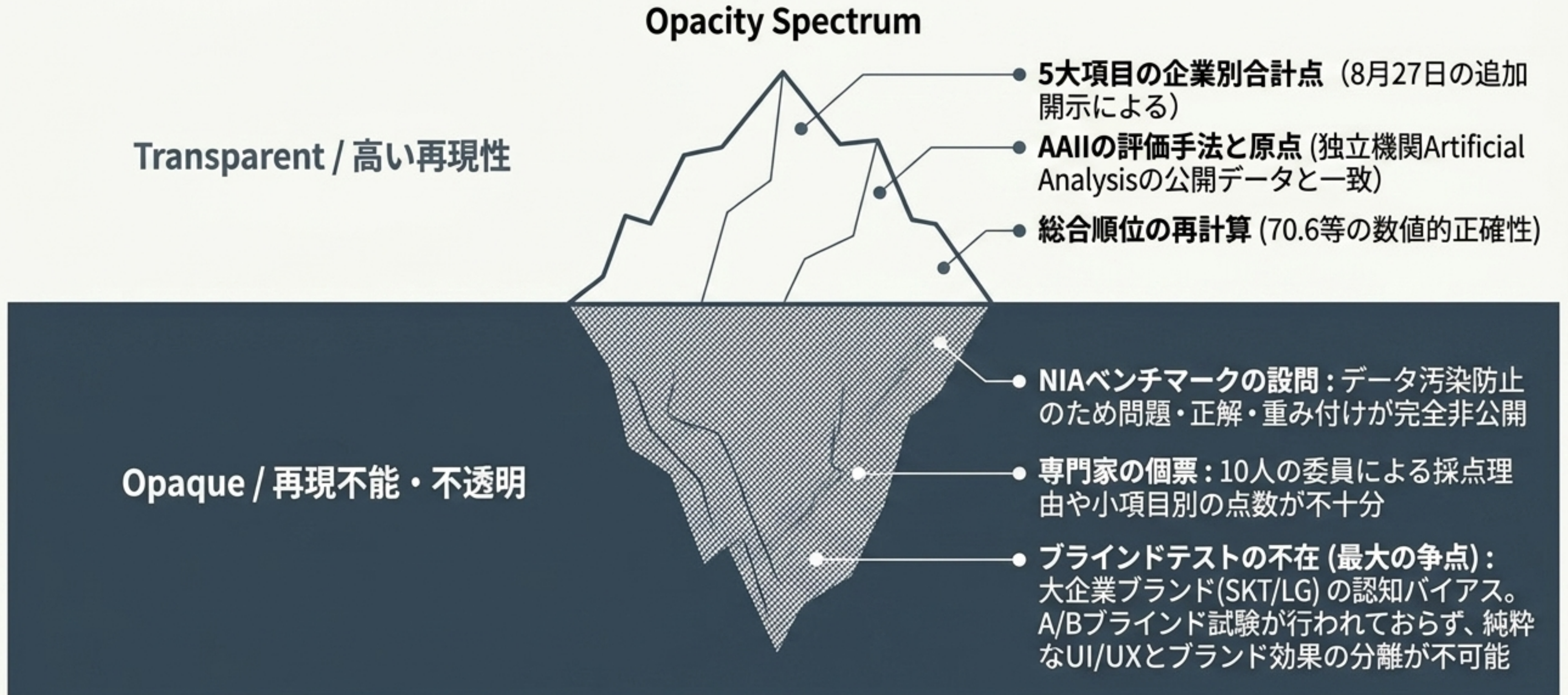


# 逆転の構造：SKTはいかにしてMotifを4.8点差で覆したか



勝負を分けたのは「AIモデルの基礎知能」ではなく、「Webインターフェースを含めた実サービスの安定性と使いやすさ (UX)」であった。

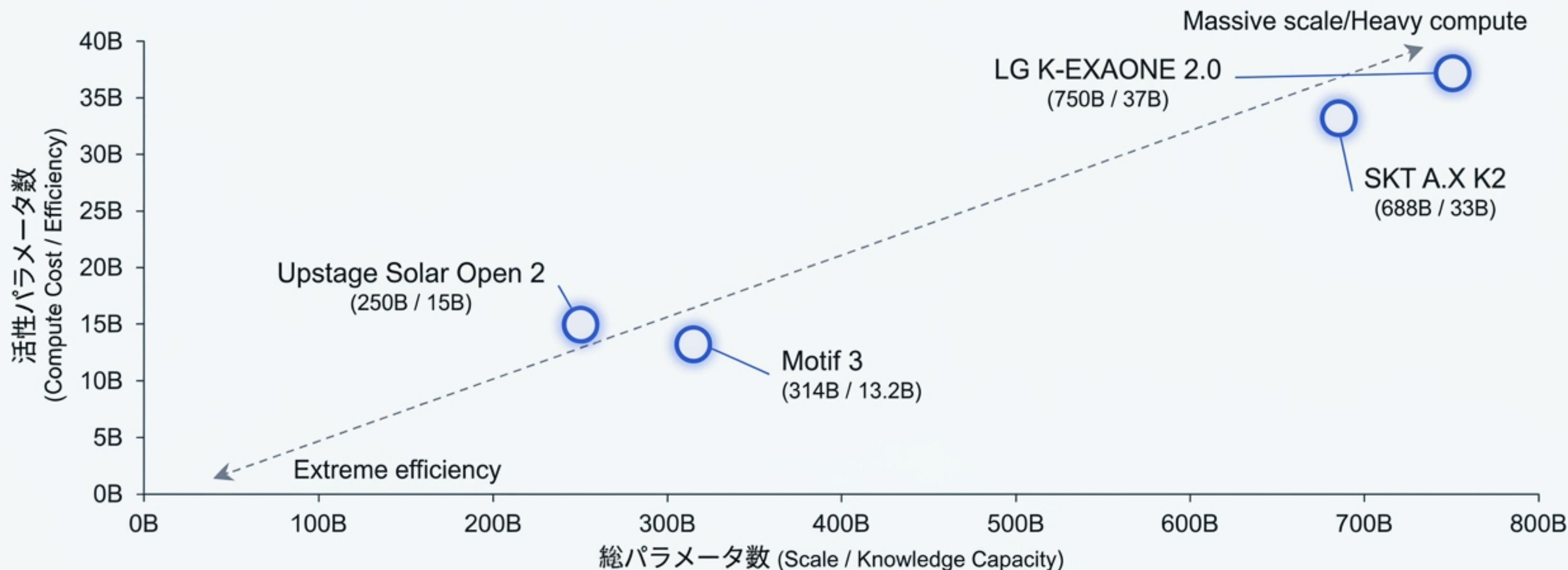
# 評価の透明性と残された「ブラックボックス」



# コンテNDER・マトリックス：上位4モデルの技術診断表

|                      | SKT A.X K2                      | Upstage Solar Open 2                 | LG K-EXAONE 2.0     | Motif 3                     |
|----------------------|---------------------------------|--------------------------------------|---------------------|-----------------------------|
| 総パラメータ数              | 688B                            | 250.3B                               | 750B                | 約314B                       |
| 活性パラメータ<br>- 1 Token | 33B                             | 15B                                  | 37B                 | 約13.2B                      |
| 最大コンテキスト             | 256K                            | 1M                                   | 256K                | 256K                        |
| アーキテクチャ<br>の特徴       | Sparse Gated<br>Attention + FP8 | Hybrid-Attention<br>(Softmax+Linear) | 既存モデルの<br>upcycling | 独自設計<br>(MTP, modified mHC) |
| ライセンス                | Apache 2.0                      | 独自<br>(Solar License)                | Apache 2.0          | 非商用中心<br>(ベータ版)             |
| 最大の強み                | 利用者UXと<br>実用バランス                | 圧倒的な推論効率                             | 安全性体系と<br>専門家評価     | 国際ベンチマーク<br>(AAII 47.4)     |

# MoEアーキテクチャと推論効率の視覚化



4社すべてが**推論コストを下げるMoE (Mixture of Experts)**を採用。特にUpstageとMotifは**活性パラメータを15B以下に抑え**、SKT/LGの**半分以下の計算負荷**でフロンティア級の性能を狙う設計思想が明確に表れている。

# 実用主義の二極：バランスのSKT、効率のUpstage

## SKT A.X K2（総合1位） 「FP8ネイティブとUXの最適解」

### Key Specs:

- 8.2兆トークンをゼロから事前学習。
- 長文時の関連トークンをtop-kで選ぶSGA（Sparse Gated Attention）を採用。

### Analysis:

基礎研究の頂点ではないが、韓国語データ、商用サービス運用（Apache 2.0）、低精度推論設計の総合力で実利用評価を勝ち取った。

## Upstage Solar Open 2（総合2位 - 0.7点差） 「1Mコンテキストの超効率ハイブリッド」

### Key Specs:

- 12兆トークン学習。
- Softmax層を限定しLinear attentionを組み合わせることで1Mトークンの長文処理と省メモリを両立。

### Analysis:

活性15Bという軽量さでAAll2位、NIA原点1位を記録。しかし独自ライセンスと専門ユーザー評価の伸び悩みが首位奪取を阻んだ。

# 両極端の哲学：安全性のLG、純粋技術のMotif

LG K-EXAONE 2.0 (総合3位)

## 「アップサイクリングと堅牢な安全性」

### Key Specs:

- 既存236Bモデルから750Bへ拡張し、追加8兆トークンを学習。
- 100+70領域に及ぶ「Korea-Augmented Universal Taxonomy」構築。

### Analysis:

単なるベンチマーク最適化ではなく、企業・国家導入への最適化を提示。安全性と専門家評価で首位を獲得した。

Motif 3 (総合4位)

## 「国際ベンチマークを粉砕する異端児」

### Key Specs:

- オープンソースの流用ではなく完全な自社設計（Grouped Differential Latent Attention等）。
- AAIで47.4という驚異的スコア。

### Analysis:

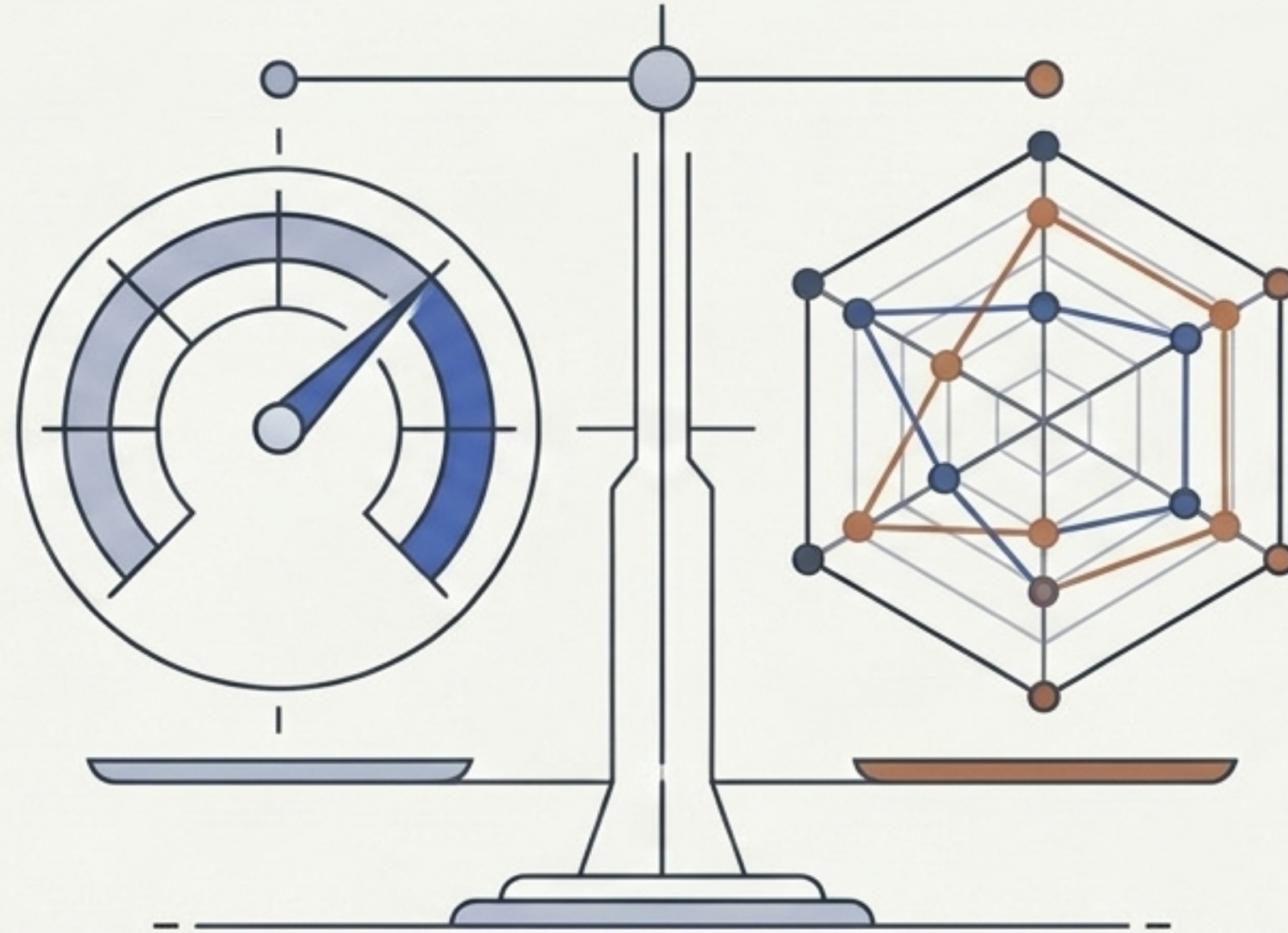
グローバル技術力は圧倒的だが、UX（ユーザー評価最低）とライセンス（非商用中心のBeta）がネックとなり、政府の「実用化」基準に合致せず敗退。

# 評価哲学の衝突：「IQテスト」対「十種競技」

## Motifの哲学： 「IQテスト」

国家の基盤モデル支援は  
「フロンティア級の純粋  
な知能（Foundation  
Modelの性能差）」に報  
いるべき。

AAIIスコア至上主義。



## MSITの哲学： 「十種競技」

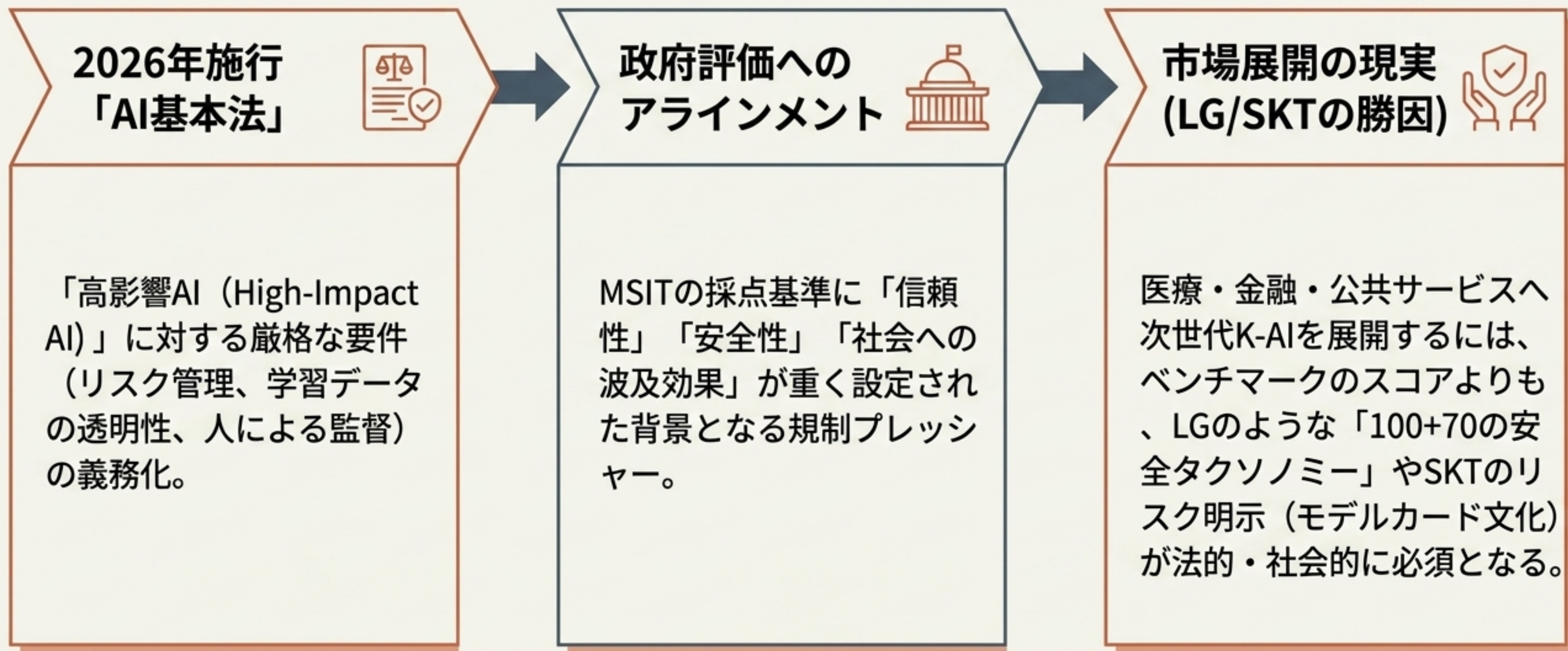
国家資金とGPUを投入する  
以上、「産業政策」として  
国内データ、独自のアー  
キテクチャ、実用的な  
UX、エコシステム貢献を  
総合評価すべき。

UX、安全性、ローカライゼ  
ーションを含む複合評価。

## Synthesis:

この論争に「正解」はない。評価関数（目的）が根本的に異なっていたことが、ベンチマーク王者の敗退というパラドックスを生んだ。

# 迫る規制の影：韓国「AI基本法」と安全性の義務



# オープンウェイトの現実：商用化を左右するライセンス

## ライセンス許諾のスペクトラム



「モデルの重みが公開されている（Open Weights）」ことと「商用利用の自由度」はイコールではない。

# ソブリンAI・マチュリティモデル：韓国の真の国家目標

## Level 4

**産業エコシステムの完成 (MSITの最終目標):**  
法制対応、安全性体系、UX、商用ライセンス、  
API化を通じた国家社会への実装

## Level 3

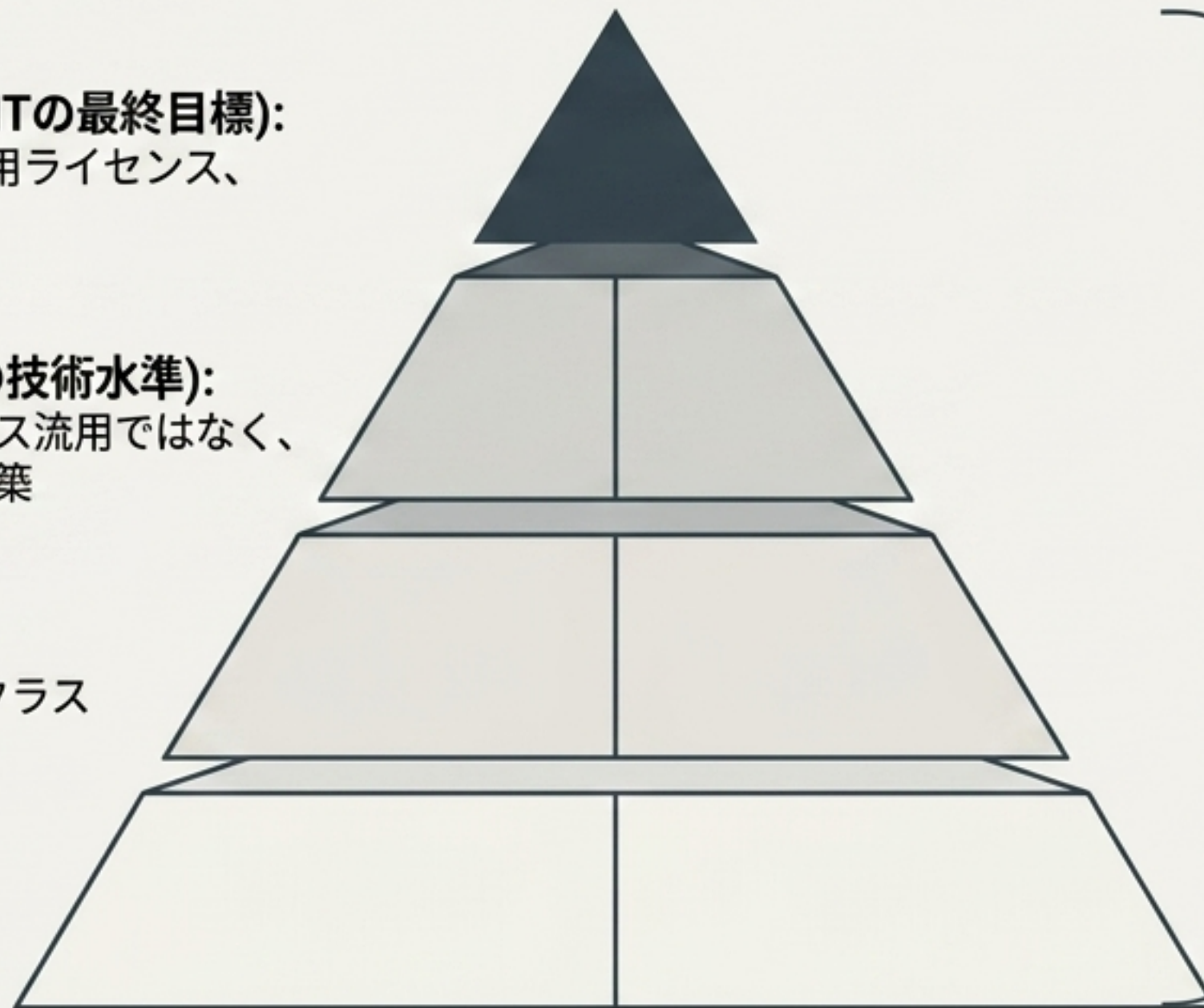
**アーキテクチャの独立 (現在の技術水準):**  
数百B規模のMoEをオープンソース流用ではなく、  
独自設計や高度なUpcyclingで構築

## Level 2

**インフラの独立 (基盤):**  
国内でのデータ構築、自社GPUクラス  
タの運用

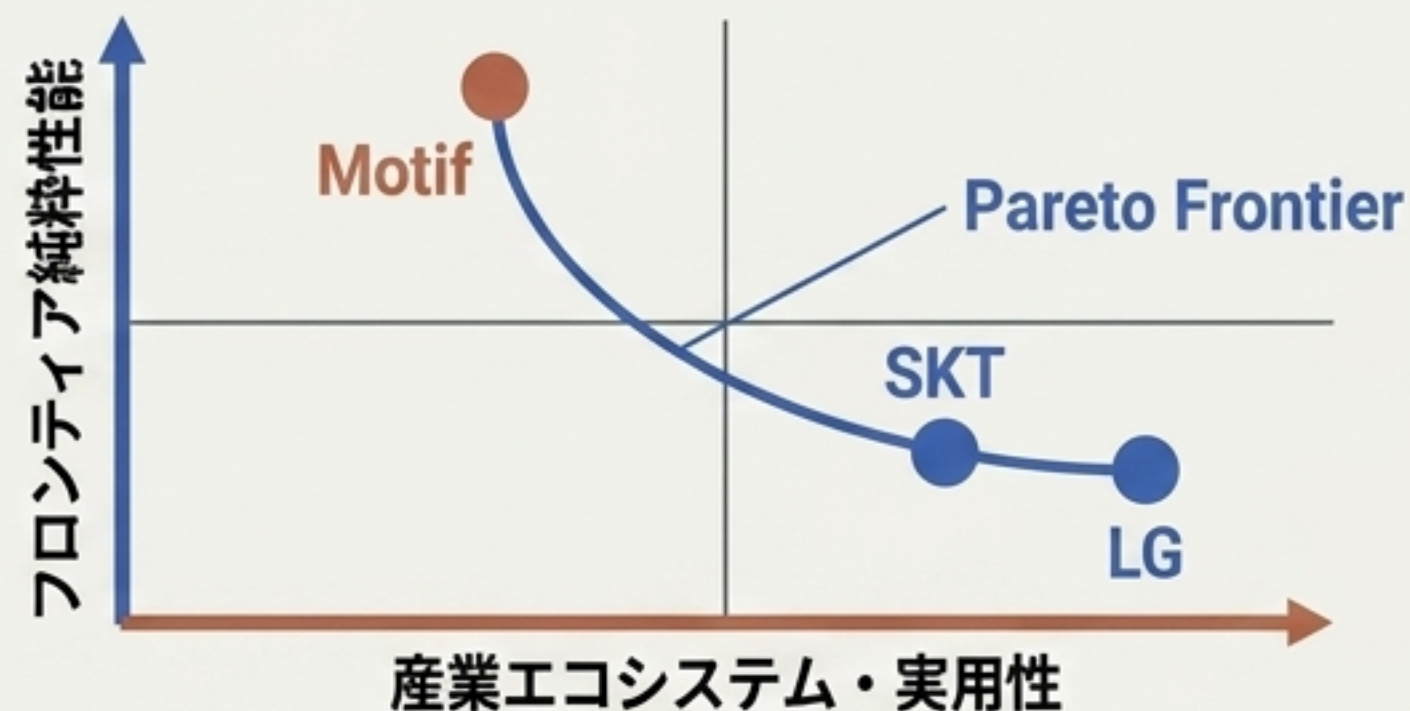
## Level 1

**言語の独立 (過去):**  
単なる韓国語LLMの構築



今回の評価は、韓国が単なる  
「API利用国」から、フロン  
ティア規模のモデルを設計・  
訓練・評価し、Level 4の産業  
実装までを自国で完結できる  
技術層を獲得した証拠である。

# 次期評価に向けた提言：「パレート・フロンティア」アプローチ

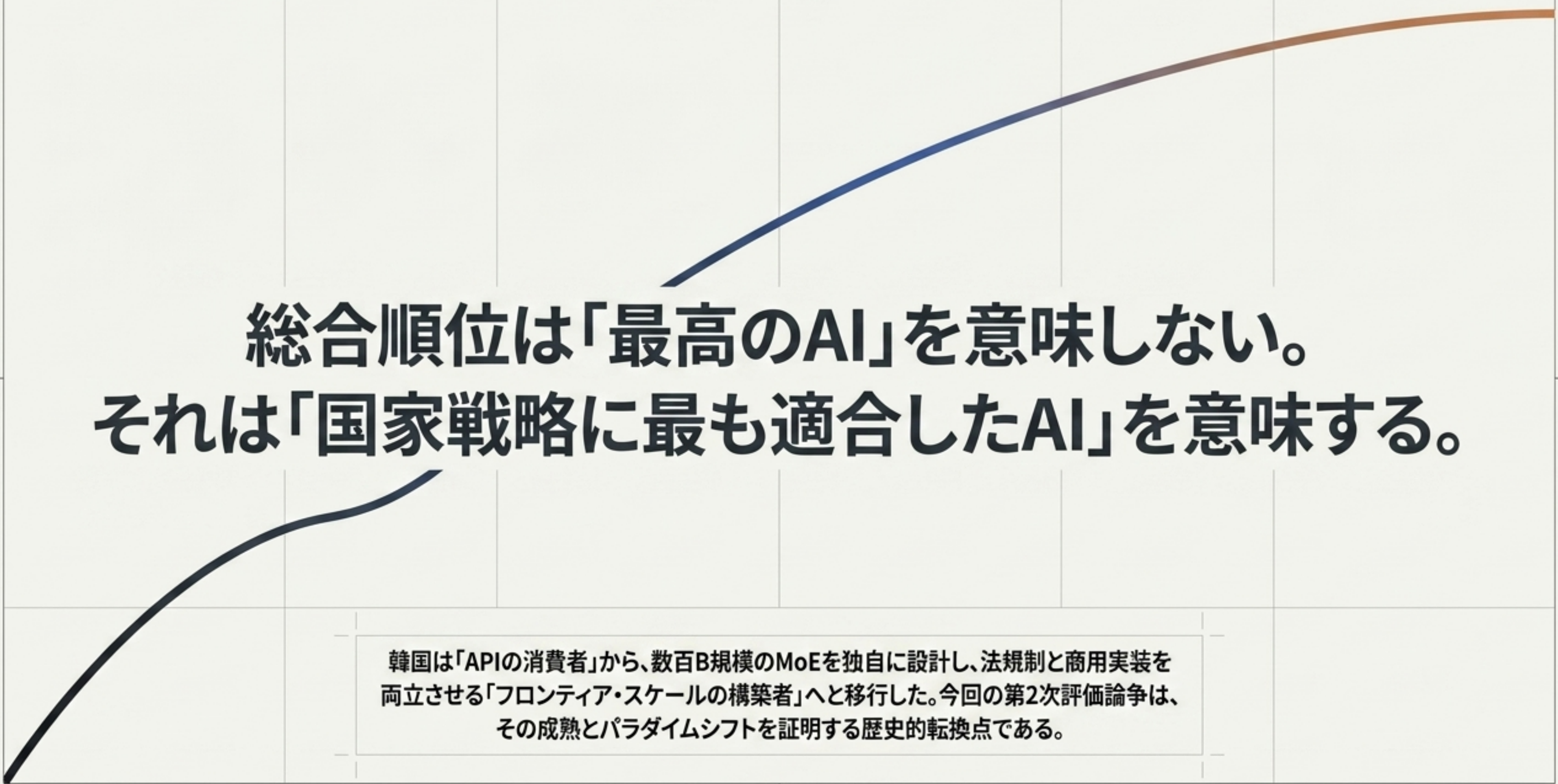


## トラックの分離 (Dual-Track Funding)

「世界最高性能を追求するR&D支援 (Motif型)」と「国内に広く展開する実用・エコシステム支援 (SKT/LG型)」を単一のスカラー得点に押し込めず、パレート・フロンティアとして複数モデルを同時支援すべき。

## 透明性の確保 (Evaluation Transparency)

- ・ユーザー評価: 企業名・UIを統一した「二重ブラインド (A/Bテスト) の導入 (UI品質とモデル品質の分離)。
- ・NIA/専門家: 汚染を防ぎつつ、領域別重みや匿名化された個票・評価理由の公開。



総合順位は「最高のAI」を意味しない。  
それは「国家戦略に最も適合したAI」を意味する。

韓国は「APIの消費者」から、数百B規模のMoEを独自に設計し、法規制と商用実装を両立させる「フロンティア・スケールの構築者」へと移行した。今回の第2次評価論争は、その成熟とパラダイムシフトを証明する歴史的転換点である。