

# 独自AI基盤モデル競争の「評価と挫折」

韓国「독파모」第2次評価の全貌と、日本（GENIAC）への戦略的示唆

|          |                    |
|----------|--------------------|
| DATE:    | 2026.09.01         |
| SUBJECT: | 国家プロジェクト設計と生存戦略    |
| TARGET:  | 政策立案者・AI経営層向けアナリシス |

# 01

## 評価設計の罫

技術トップの脱落：技術ベンチマーク1位の企業が総合最下位へ転落。実用性・受容性偏重の評価基準と、「透明性の圧力」が業界内の軋轢を露呈させた。

# 02

## 制度の限界と転換

資源集中への急旋回：「順次脱落・資源分散型」では米中フロンティアに対抗不可と韓国政府が認識。2027年以降、少数精鋭への「資源集中」へ抜本的再編を示唆。

# 03

## 日本への示唆

ドメイン特化へのピボット：「広く薄く」のGENIAC（日本）が学ぶべき教訓。最大45兆円規模のデジタル赤字を防ぐため、汎用スケール競争から「安全なインフラ・ドメイン特化」へ戦場を移す必要性。

## INCIDENT TIMELINE: 2026.08.18 - 08.27

8.18

MSITブリーフィング。3チーム通過・Motif脱落を発表。ステイグマ懸念から詳細スコアは非公開。

8.20

項目別1位とその点数のみを追加公開。しかし論争は収束せず。

8.26

脱落したMotifが異議申立。詳細スコアの完全公開と再審議を強く要求。

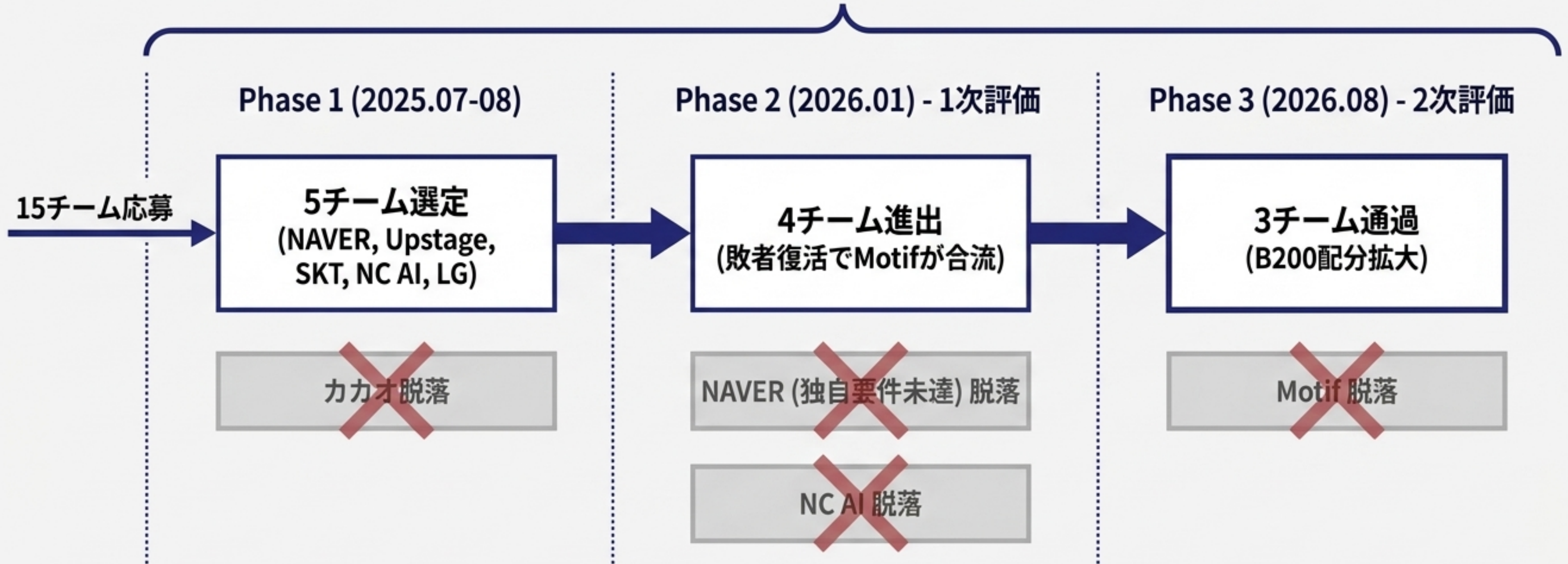
8.27

異例の措置。他チーム同意の上、全社・全項目のスコアを完全公開。



「透明性への圧力」が、結果として政府の評価の妥当性に対する根本的な問いを社会に突きつけた。

## MOVING TARGET: 6-Month Cycle (ムービングターゲット方式)



基準が半年ごとに変動する「ムービングターゲット方式」は、開発の柔軟性を担保する反面、参加企業から予見可能性を奪い去った。

## SCORE PARADOX: THE INVERSION MECHANISM

ベンチマークスコア  
(技術・Max 40)

ユーザー評価スコア  
(受容性・Max 25)

Motif  
(総合最下位 - 脱落)

24.6点  
(1位)

14.1点 (最下位)

SKT  
(総合1位 - 通過)

22.2点  
(3位)

19.1点  
(1位)

逆転のメカニズム: 上位3社の総合点差はわずか1.6点。「技術的性能」(4.0点差)よりも「実用性・受容性」(5.0点差)を重く見る設計が、ベンチマーク1位の脱落という致命的な逆転劇を生んだ。

## EVALUATION MATRIX ANATOMY (2ND ROUND)

| 評価軸                  | SKT<br>(総合70.6) | Upstage<br>(総合69.9) | LG<br>(総合69.0) | Motif<br>(総合65.8) |
|----------------------|-----------------|---------------------|----------------|-------------------|
| AAII<br>(Max 25)     | 8.8             | 9.4                 | 7.8            | 11.9              |
| NIA<br>(Max 15)      | 13.4            | 13.3                | 12.8           | 12.7              |
| 専門家<br>(Max 35)      | 29.3            | 29.1                | 29.5           | 27.1              |
| AI専門ユーザー<br>(Max 15) | 11.6            | 10.8                | 11.3           | 8.4               |
| 一般国民<br>(Max 10)     | 7.5             | 7.3                 | 7.6            | 5.7               |

AAII（国際指標）はエージェント分野等を重視。Motifが圧倒的1位だが、LGは最下位。

AAII最下位のLGが専門家評価で1位に。「誰がどの基準で評価したのか」と批判的的に。

一般国民評価は1次評価にはなかった新要素（実参加185名）。Motifのスコアが極端に沈む要因に。

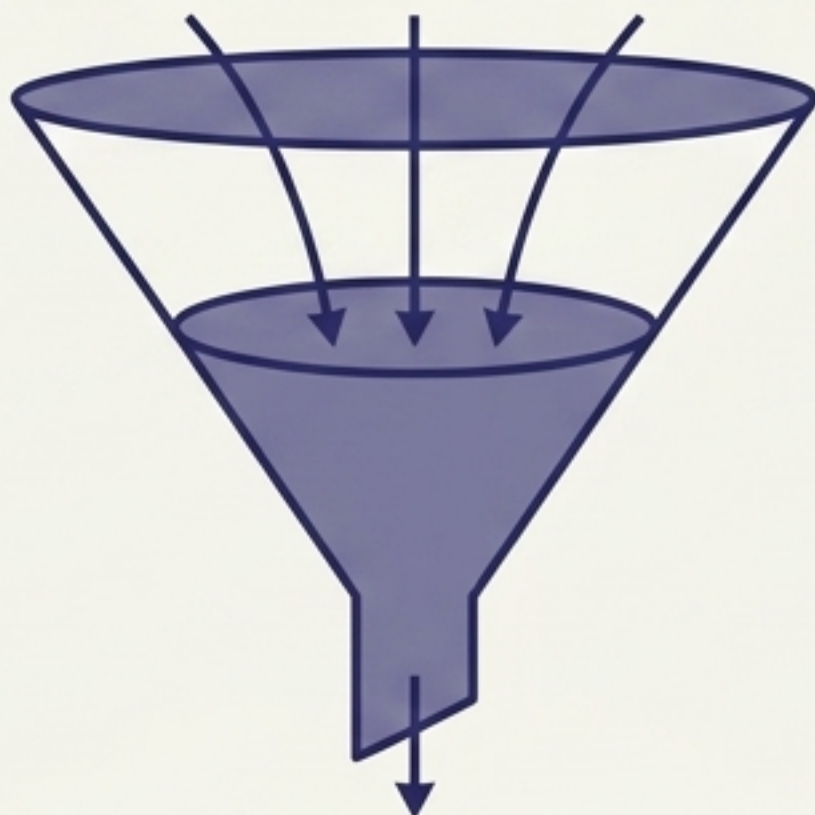
# MODEL ARCHITECTURE & SPECIFICATIONS

業界標準のシフト: 全4社が演算効率を高めるため「MoE (Mixture of Experts) 構造」を採用。

| SKT (A.X)                                                                                              | Upstage (Solar)                                                                                         | LG (EXAONE 2.0)                                                                                          | Motif (Motif 3)                                                                                          |
|--------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
|  688B / 33B (Active) |  250B / 15B (Active) |  750B / 37B (Active) |  314B / 13B (Active) |
|  希薄MoE+独自SGA        |  MoE                |  MoE                |  独自設計MoE            |
| ✓ 数学・韓国語に強み。<br>商用実績高評価。                                                                               | ✓ AIME 2026 97.1%、<br>ポータル連携・国産<br>NPU実証評価。                                                             | ✓ 最大100万トークン対<br>応。エージェント<br>ックAI能力高評価。                                                                  | ✓ AAILスコア首位。※<br>脱落后、政府支援に<br>頼らず独自開発を<br>表明。                                                            |

## 論争A: 「技術格差の圧縮」

国際指標(AAIL)の性能差: **47点** vs **31点** (圧倒的格差)



最終換算後: **わずか4点差**

**主張:** 定量的な技術格差が結果に正当に反映されない、評価アルゴリズムの構造的欠陥。

## 論争B: 「ベンチマクシング疑惑」

実際の能力  
(Real Capability)

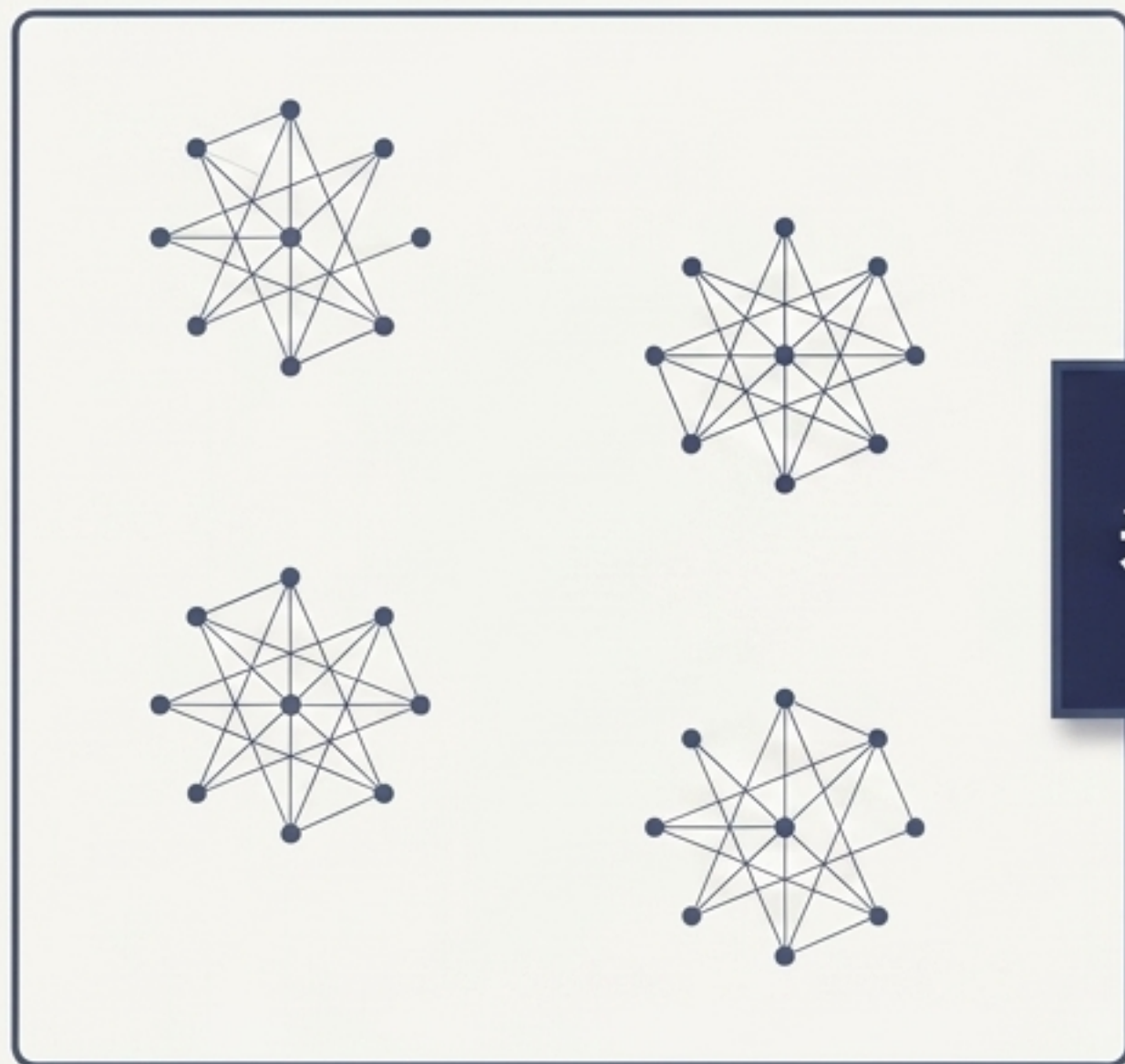
特定の指標への  
過度な最適化  
(Over-optimization)



**主張:** 参加企業がICML等でベンチマーク指標に過剰に最適化 (Benchmaxxing) した疑い。政府は公式に否定したが、評価指標への不信感は残存。

## POLICY PIVOT: RECOGNIZING THE LIMITS

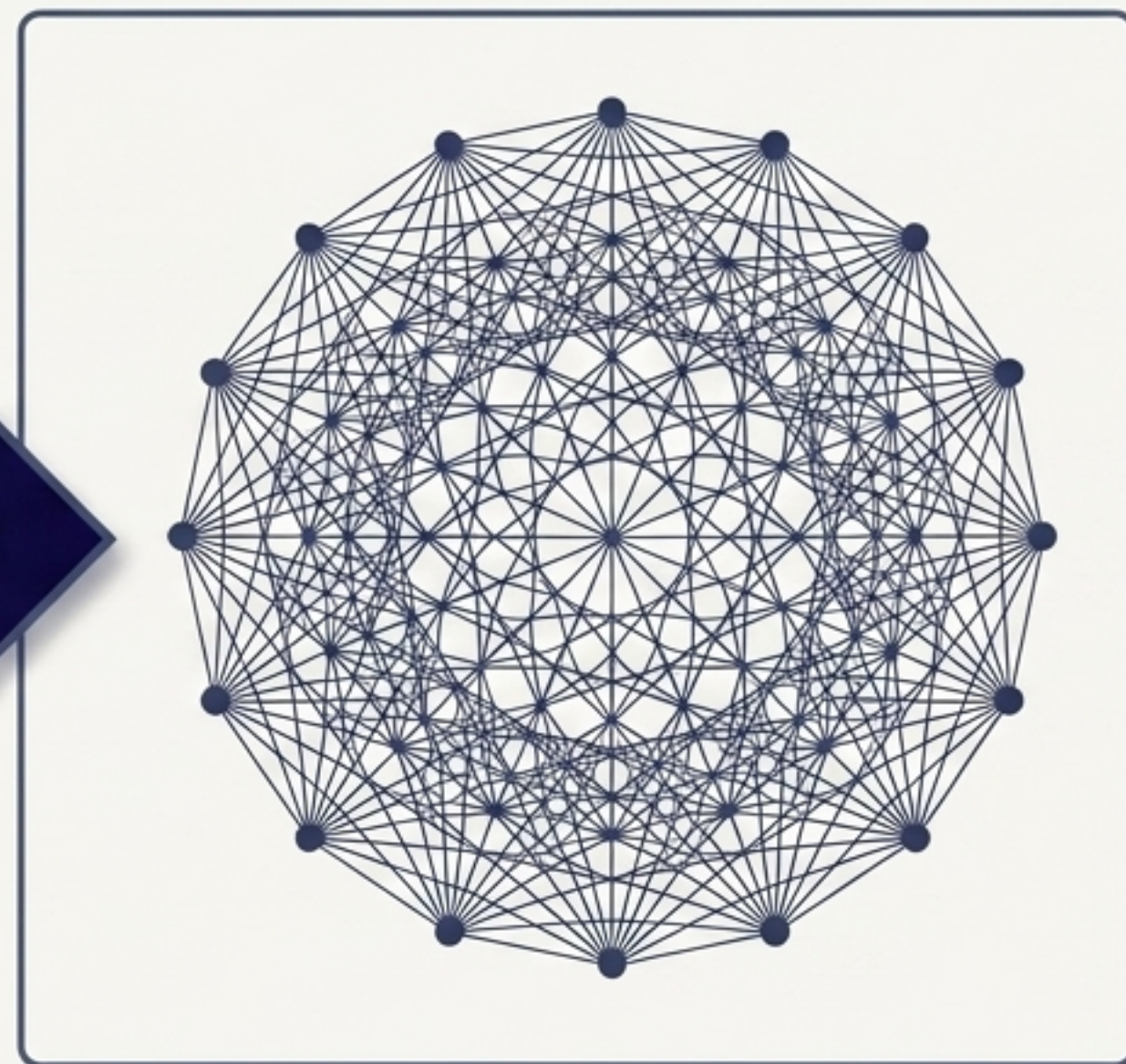
Before: 順次脱落・資源分散型



- ・ 少数を競わせ、GPU（数千枚規模）や予算を分散。
- ・ 限界: リュ・ジェミョン第2次官「資源分散型では米中フロンティアモデルの進化に到底追従できない」。

米中フロンティアへの焦り

After: 枠組みの抜本的再編・資源集中型



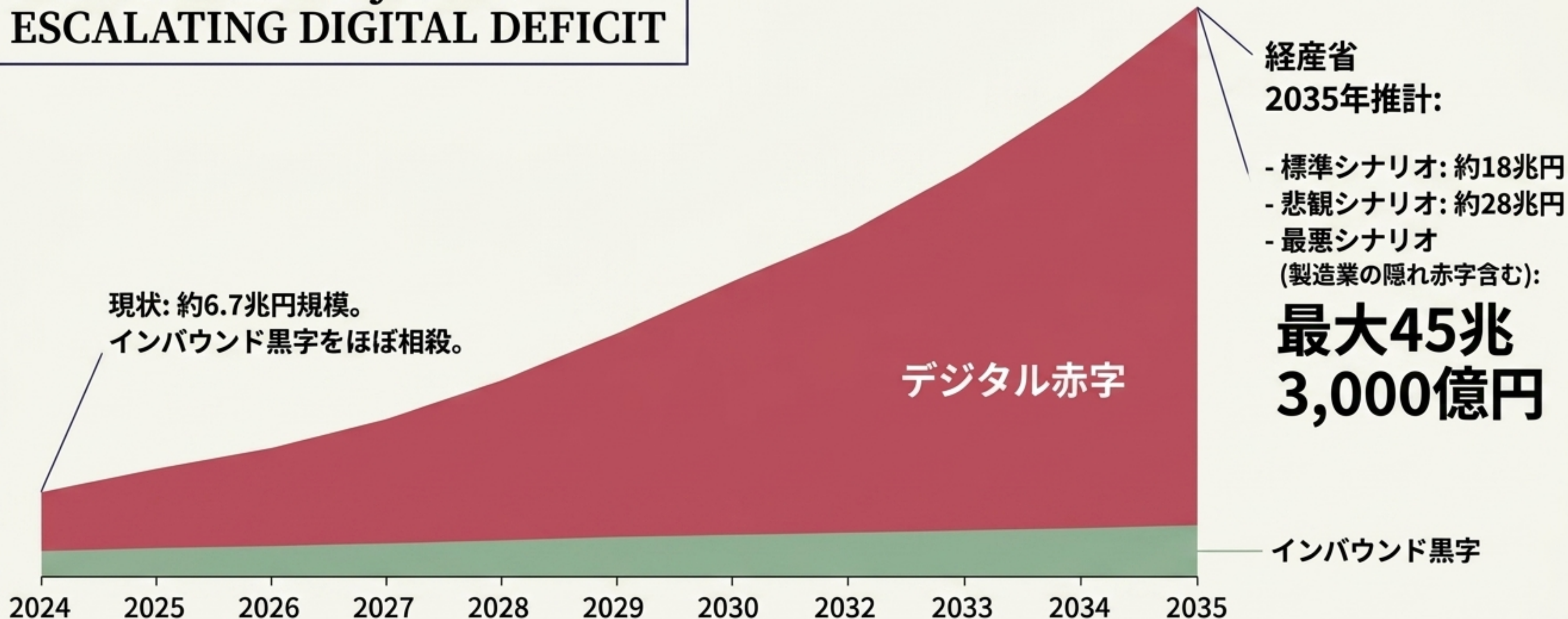
- ・ 2027年予算案: 独自AIモデル確保等に約5兆5,937億ウォンを計上。
- ・ 目標: 最先端GPU 5万枚の確保。現行の数倍規模の集中投資ヘシフト。

FRAMEWORK COMPARISON: KOREA VS JAPAN

| 評価軸     | 韓国（독파모）                                   | 日本（GENIAC 第1～4期）                |
|---------|-------------------------------------------|---------------------------------|
| 支援アプローチ | 少数を競わせ絞り込む<br>「競争・順次脱落型」                  | 広く薄く延べ30件を採択する<br>「非競争・インフラ提供型」 |
| 透明性と副作用 | 点数・順位を完全公開。<br>結果として紛争とスティグマ<br>（烙印）が発生 ⚠ | 順位非公開・コミュニティ重視。<br>健全だが競争原理は緩慢。 |
| 評価基準    | ムービングターゲット方式。<br>開発陣の予見性が著しく低下 ⚠          | 計算資源の提供と実証を重視。                  |

教訓: 評価基準や配点換算式を事前に固定・公開しない限り、「実用性」評価の導入は必ず紛争の火種となる。韓国の失敗は最大の反面教師である。

# THE THREAT TO JAPAN: ESCALATING DIGITAL DEFICIT



この莫大な富の流出を防ぐため、日本は巨大資本が支配する汎用モデルの  
「スケール競争」に正面から付き合い続けるべきか？

# JAPAN'S SURVIVAL PLAYBOOK: THE FLANK STRATEGY

独自性・優位性  
(National Advantage)

汎用スケール競争  
(米中・巨大資本)

**1. 脱・汎用スケール競争**  
汎用AIモデルで独自性を出す正面突破の困難さを直視（経産省方針と合致）。

Flank Strategy  
(側面攻撃)

## 2. 比較優位の最大化

DATA: インターネットデータではなく、現場データ（製造・自動車・ヘルスケア）の囲い込み。

INFRA: 低遅延インフラ（IOWN等）と厳格な法制を基盤とした「安全なAI活用基盤の提供者」へ。

日本の勝ち筋  
(安全なAIインフラ)

対象領域  
(Domain Focus)

## 3. 知財・開発アプローチの転換

「フルスクラッチ」への固執から脱却。「オープンウェイト活用＋事後学習＋ドメイン特化」におけるデータ権利処理とライセンス設計に注力する。

## STRATEGIC DIRECTIVES: 3 RULES FOR POLICYMAKERS

1

### 評価軸の事前固定化

「実用性」を評価に組み込む際は、配点換算式を含め事前公開を徹底せよ。事後の恣意性批判とスティグマを防ぐ唯一の防波堤である。

2

### 「ムービングターゲット」の排除

目標の柔軟な変更は、開発現場の透明性と予見可能性を著しく破壊する。国家プロジェクトのマイルストーンは不変であるべきだ。

3

### 強みへの資源集中

グローバルフロンティア比95%の達成は分散投資では不可能。日本の真の強み（エッジ・現場データ・インフラ）へ戦略的にリソースを集中投下せよ。