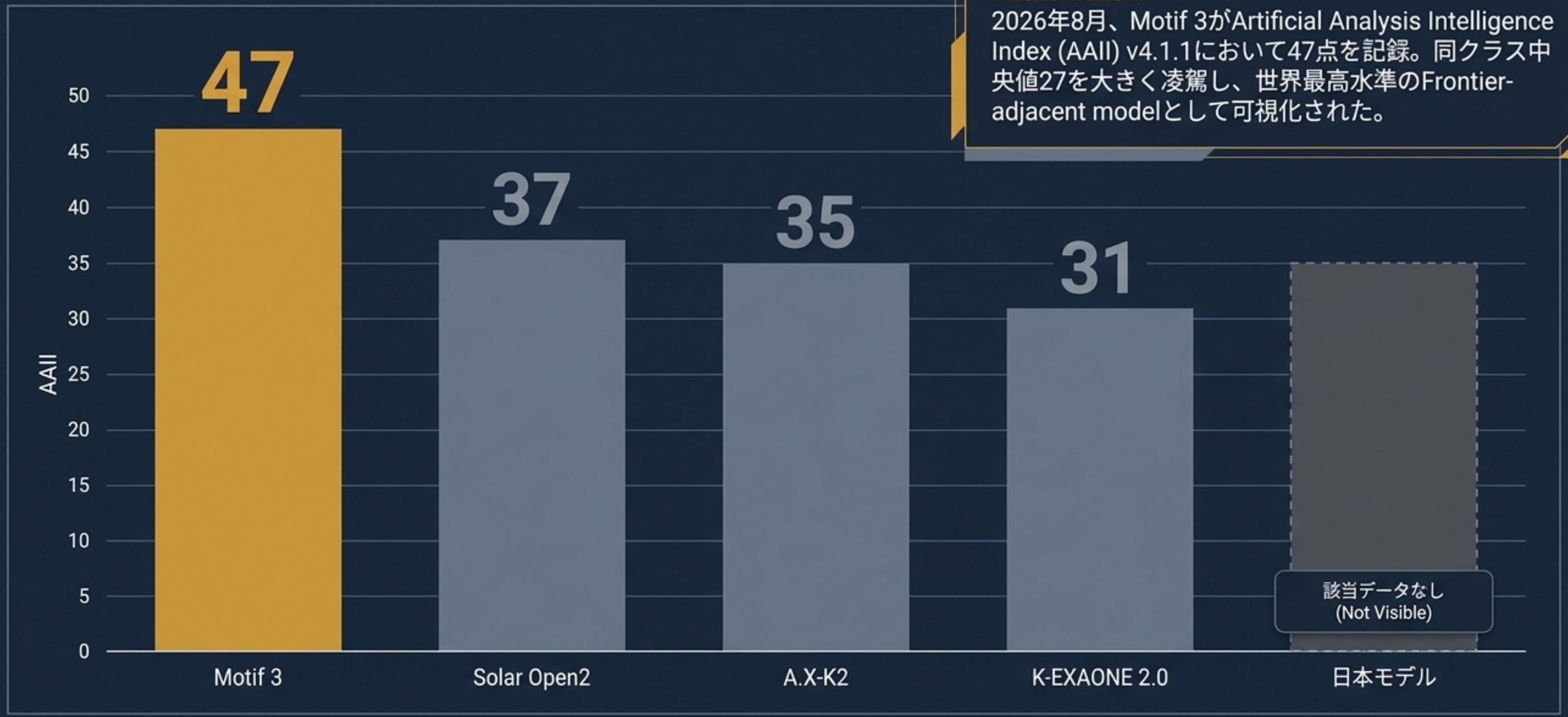
[CONFIDENTIAL / FOR STRATEGIC REVIEW]

錯覚と現実：Motif 3の高評価が暴く日韓LLMエコシステムの構造的真実

AAII 47点の技術的解剖と、日本が取るべき「二層戦略」への転換

Strategic Blueprint

Phenomenon:
2026年8月、Motif 3がArtificial Analysis Intelligence Index (AAII) v4.1.1において47点を記録。同クラス中央値27を大きく凌駕し、世界最高水準のFrontier-adjacent modelとして可視化された。

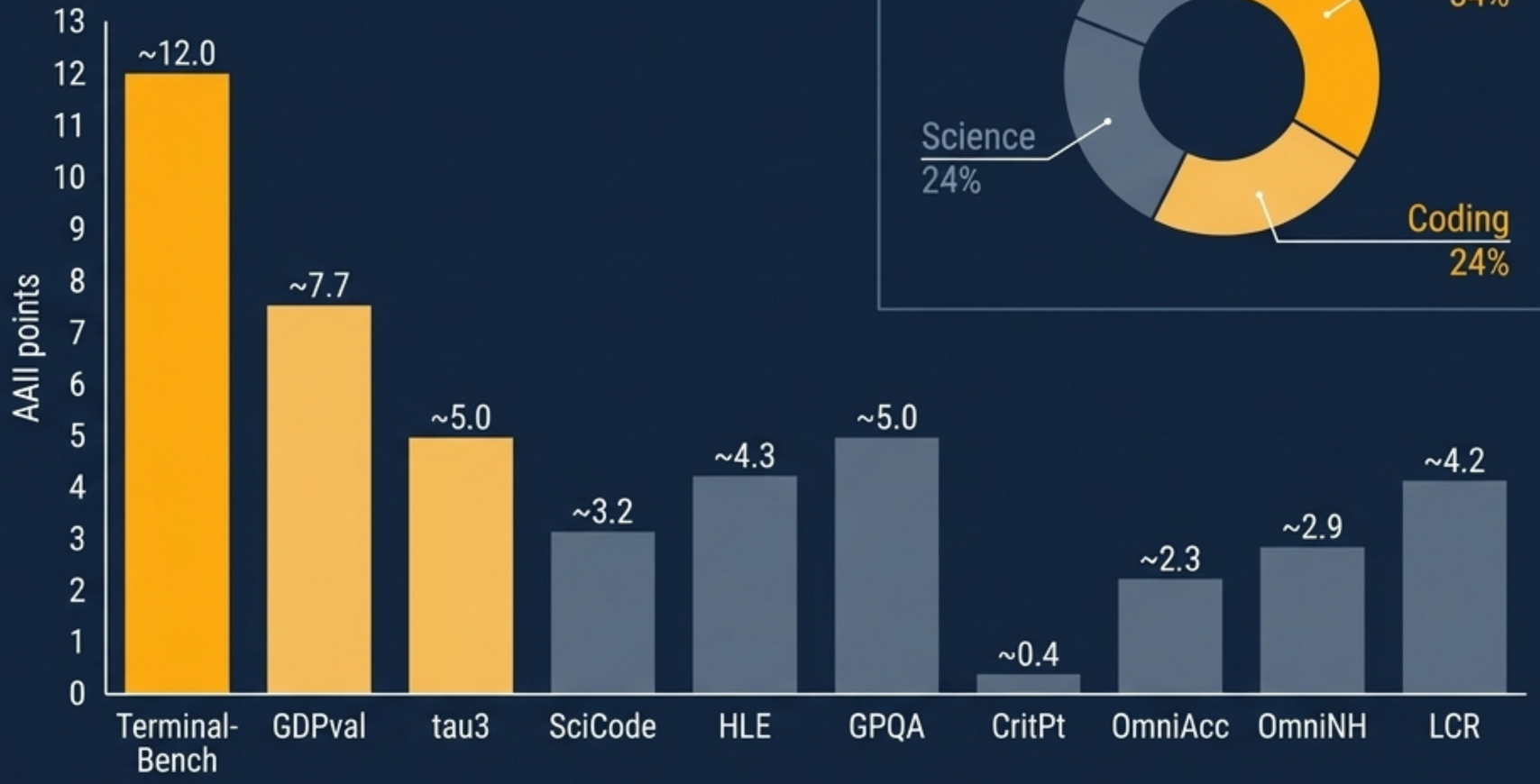


CORE INSIGHT: この結果は単なる「韓国語能力の高さ」でも、「純粋な知能の差」でもない。評価指標の性質と国家KPIが完全に一致した結果である。

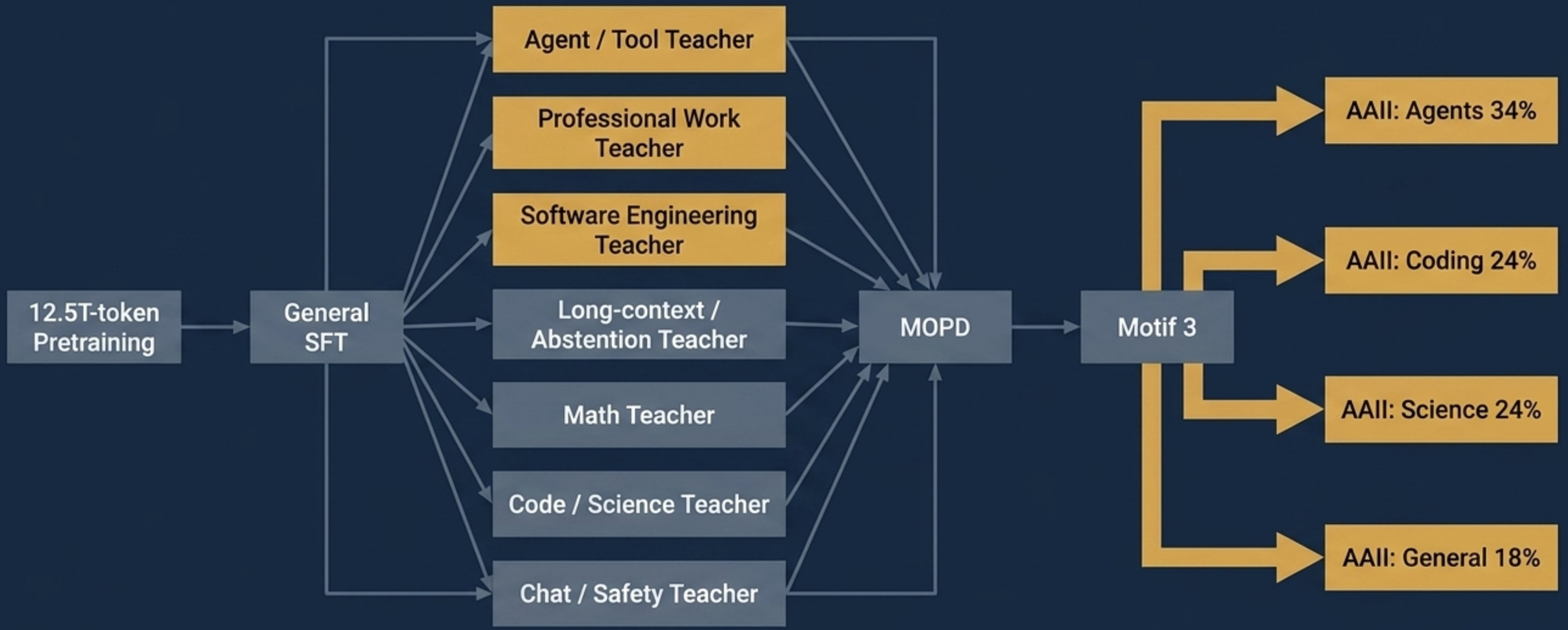
Terminal-Benchの圧倒的寄与:
12.0 points
(全体の25%)

Agentic/Professional実行力:
GDPval + tau3
= 12.7 points

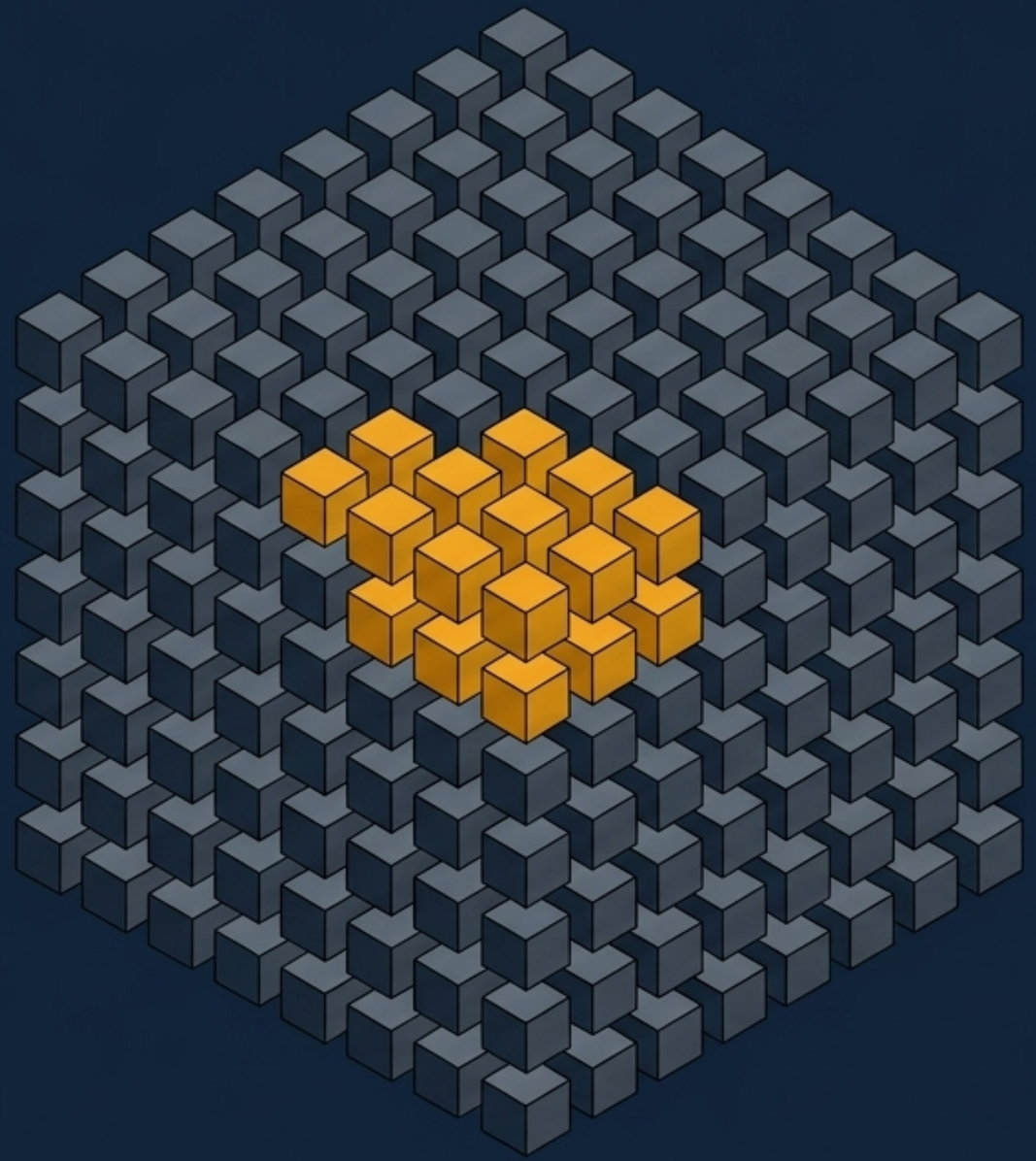
Motif 3のAAII 47.36点への寄与



47点は百科事典的知識ではなく、「端末操作」「コード」「ツール利用」といった
Agentic/Codingタスクを完遂する能力（AAIIの58%を占める）に極端に最適化された結果である。



従来の「巨大な知識を詰め込む」アプローチではなく、目的のベンチマークに直結する専門教師（Specialist Teachers）を作り、MOPDで統合する「2025-2026年型Agentic Model設計」。



Total Parameters:

314B

Active Parameters:

13.2B (4.2%)

Context Window:

262k

Architecture:

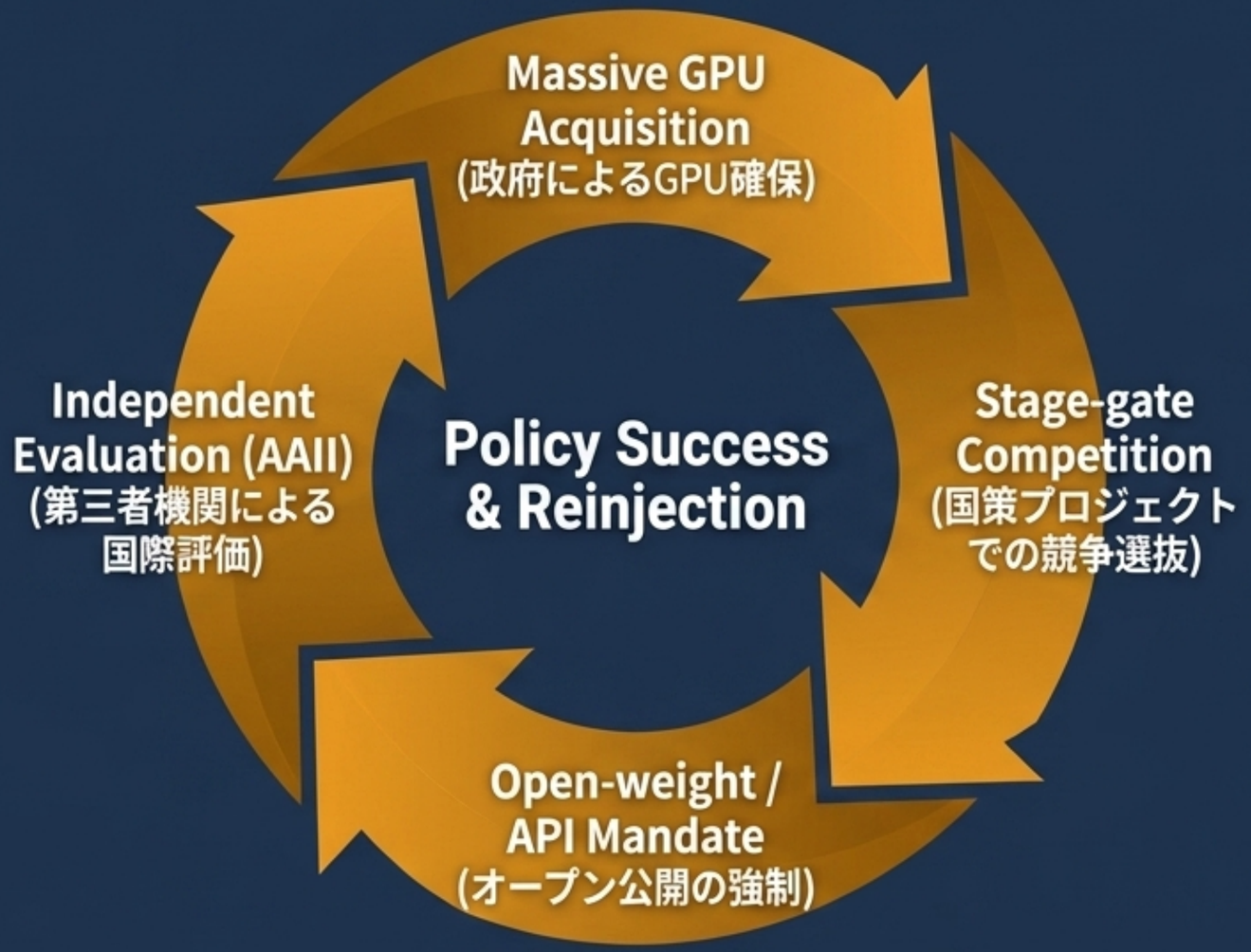
384 Routed Experts (Top-8)
+ 1 Shared Expert, GDLA

巨大なパラメータ数による推論能力を持ちながら、MoEアーキテクチャとGDLAの導入により、推論時の計算負荷を劇的に抑制。単純なパラメータ競争ではなく、「実用的な効率性」を兼ね備えている。

	韓国 (Korea)	日本 (Japan)
政策的KPI	Sovereign AI / グローバル・ベンチマークでの競争力	国内産業実装 / 物理AI・特化型AI
計算資源配分	1.46兆ウォンで1万GPU確保、 少数チームへ集中投下	GENIAC等による多数案件への分散支援
商用戦略	API / オープンウェイト / 国際リーダーボード掲載	国内企業向け / オンプレミス / 機密データ保護
主要プレイヤー	Motif, SKT, LG, Upstage	NTT, PFN, NII/LLM-jp, NEC等

差の正体は「能力」ではなく「ゲームのルールの違い」。韓国はグローバル・ベンチマークを国家KPIに設定しているのに対し、日本は国内の産業的価値とオンプレミスに最適化している。

可視化を設計する政策的フィードバックループ



計算されたリリースループ

- 2026-08-10: Technical Report公開
- 2026-08-12: AAIIへ評価掲載
- 2026-08-13: 国内首位として大々的に報道

韓国政府は単なる研究助成ではなく、「開発→公開→独立評価→成果アピール」という完全なFeedback Loopを政策としてデザインしている。

Objective Mismatch（目的の不一致）

日本モデルは「日本語の精緻さ」「軽量化」「企業内データの秘匿性」を重視。Agentic/Coding重視の英語AAIIとは交わらない。

Access Friction（アクセス形態）

tsuzumi等の主力はオンプレミス/プライベートクラウド。AAIIのような第三者が評価可能なAPI/Weight公開とは逆行する。

Release Timing & Scale（投入スケール）

韓国勢が数百Bクラスを同時投入したのに対し、日本は段階的リリース（LLM-jp-4の8B/32Bなど）を採用。

Policy Incentives（政策インセンティブ）

GENIACは素晴らしい支援だが、対象が分散的。単一のGeneral Leaderboardへの集中を促すインセンティブがない。

日本モデルの「不在」は、技術力の欠如ではなく、評価機関が観測できないエコシステム（Closed/Domestic）を選択している結果である。

The Myth (神話)

~~データ不足~~

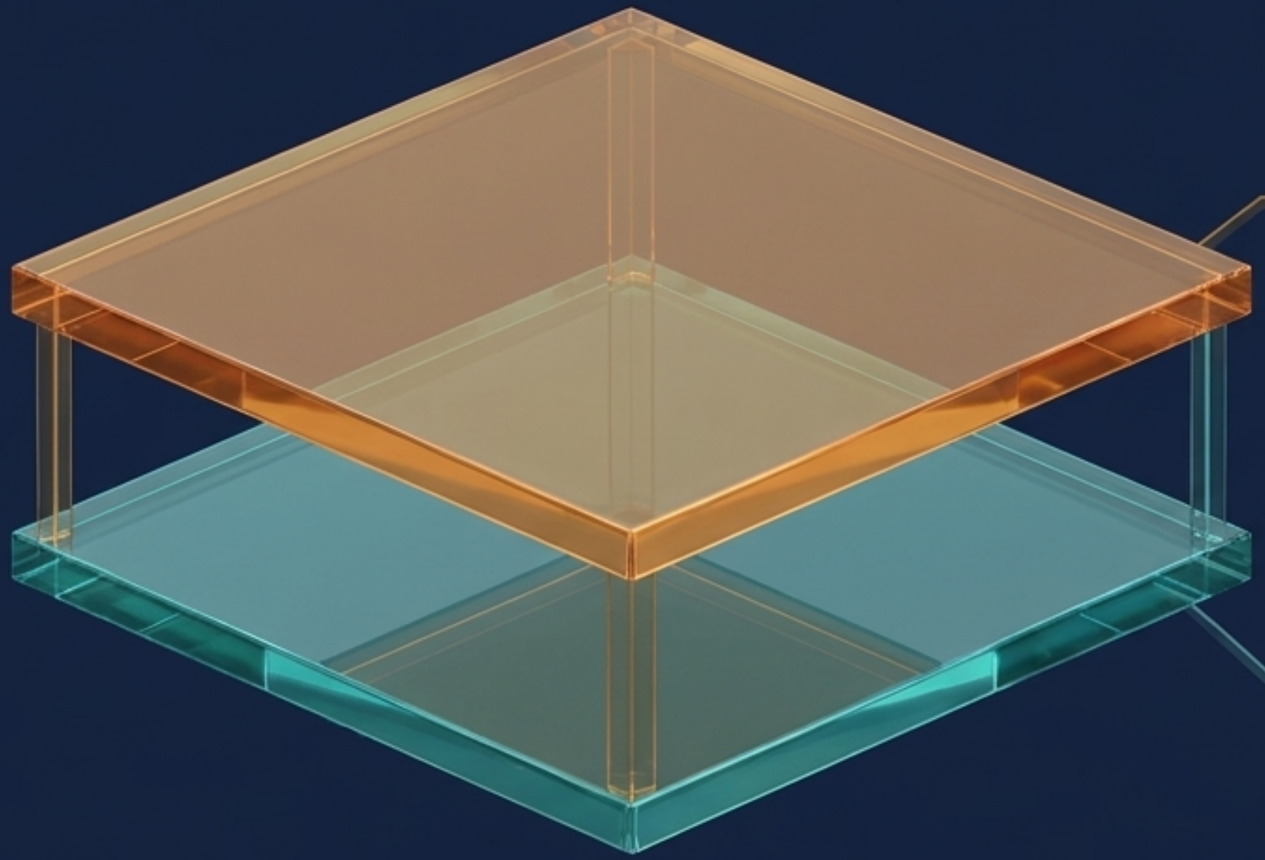
The Reality (現実)

LLM-jp-4 (2026年4月): 約12兆トークンの巨大コーパス（政府文書、公開Web、合成データ含む）をゼロから構築。

Pretraining capability: LLM-jp-3における172Bモデルのフルスクラッチ開発実績。

2026年現在、日本にデータや基盤構築能力がないという説明は成立しない。真の課題は「そのデータを使って、何Bのモデルに、どれだけの計算資源を集中させ、エージェント的強化学習（RL）を施すか」という資源の集中と後学習（Post-training）の設計にある。

Two-Tier Strategy (二層戦略)



Layer 1: Frontier General Model Track

目的: グローバルでの可視性・比較可能性の確保

アプローチ: 少数のモデルへの計算資源の極端な集中、AAII等への最適化、API/Open-weightでの標準公開。

Layer 2: Sovereign & Industrial Models

目的: 日本特有の価値（機密性、軽量化、物理AI）の維持・拡大

アプローチ: オンプレミス、業界特化型データ、日本語最高精度、セキュリティ。

ベンチマークハックに踊らされて既存の強み（Layer 2）を捨てるべきではない。
しかし、無視を続ければグローバルでの存在感を失う。

両者を明確に分離した「二層戦略」が必要である。

ACTION PLAN: DEVELOPERS & ENTERPRISES

Action 1: Release Standard Portfolios (評価ポートフォリオの標準化)

日本語特化ベンチマーク（JGLUE等）だけでなく、AAII、SWE-Bench等のグローバル標準スコアをリリース時に併記し、「比較可能な可用性」を自ら作る。

Action 2: Pivot to Post-Training (事後学習への投資シフト)

Motif 3の教訓に従い、Agent/Tool/SWEなどの「専門教師モデル」とOn-policy RLの構築にプレトレーニングと同等の予算と戦略的重要性を置く。

Action 3: Leverage Open-Science (透明性を武器にする)

LLM-jpのようなデータレベルからの完全な透明性（Decontamination reportの開示等）は、韓国勢が十分に行えていない領域であり、日本の強力な差別化要因となる。

Intervention 1

GENIAC内に「Frontier General Model Track」を分離し、少数のチームにGPUを長期集中投下する。

Intervention 2

補助金の成果条件として「リリース後30日以内の海外独立リーダーボードでの評価可能化（API/Weight公開）」をKPI化する。

Intervention 3

単純なGPU枚数ではなく、「1チームあたりのUsable Training FLOPs」と「Post-training Budget」を追跡指標とする。

Motif 3の47点は、純粋な技術的勝利以上に、
「研究目標」「公開方法」「政策インセンティブ」を
一つのループに繋ぎ合わせた**韓国の戦略的勝利**である。
日本に必要なのは韓国の模倣ではなく、
自らの強みを活かした上での、
グローバルな「評価のゲーム」への戦略的参戦である。

END OF REPORT