

Motif 3 徹底調査: 韓国LLMの飛躍と 日本LLMの「見えない」実力

Artificial Analysis Intelligence Index
47点の意味、技術構成、そして
二層型国家AI戦略への提言

Target: Japanese AI Strategists & Policymakers | Focus: Sovereign AI & Global Benchmarking

47点は「韓国語能力」 の評価ではない

科学的推論 (24%):
HLE, GPQA
Diamond, CritPt

24%

エージェント (34%):
GDPval-AA v2,
 τ^3 -Banking (実務
成果物、ツール利用、
銀行業務フロー)

34%

WEIGHT
DISTRIBUTION
MAP

18%

一般能力 (18%):
AA-LCR,
AA-Omniscience

24%

コーディング (24%):
Terminal-Bench v2.1,
SciCode (端末操作、
ソフトウェア修正)

Motif 3が高得点を得た直接の理由は、韓国語の優位性ではなく、**英語圏の「環境を操作して仕事を完了する能力（エージェント・コード）」にモデル設計を正面から適合させた結果**である。

細粒度MoE：巨大な容量と軽量な実行の分離

区分	量	主含評価
一般能力	18%	AA-LCR, AA-Omniscience

出所：Artificial Analysis Intelligence Benchmarking Methodology. [2]

エージェント型タスクのベンチマークは、従来型の問題や学問問題だけでなく、外部環境を操作し、複雑なタスクを遂行する能力が要求される。2026年6月v4.1 設計仕様、よりエージェント型の業務負荷を反映している。[2.3]

Motif 3 の学術分野は、エージェント型ツール利用、専門業務、数学、コード・科学、対話の七つの専門機能を統合する。この構成は、AAII の評価に適合している。[4]

3.3 AAII を過大解決してはいけない理由

- 日本語・韓国語の自然さ、物語、専門用語、文化的文脈は AAII 47%
- Motif 3 はテキスト専用であり、画像・音声・動画を扱うマルチモーダル能力は低い
- 日本語 (Tanniwaprio) と英語 (P21) の両方、文化の Bias: AAII 1.43%
- Motif 3 はテキスト専用であり、画像・音声・動画を扱うマルチモーダル能力は低い
- 出力トークン数が多いため、性能と専用対換率・応答時間は別個で評価される
- ベンチマークへの適合が実用価値と重なる部分は大きいですが、特定実用価値は保証しない。

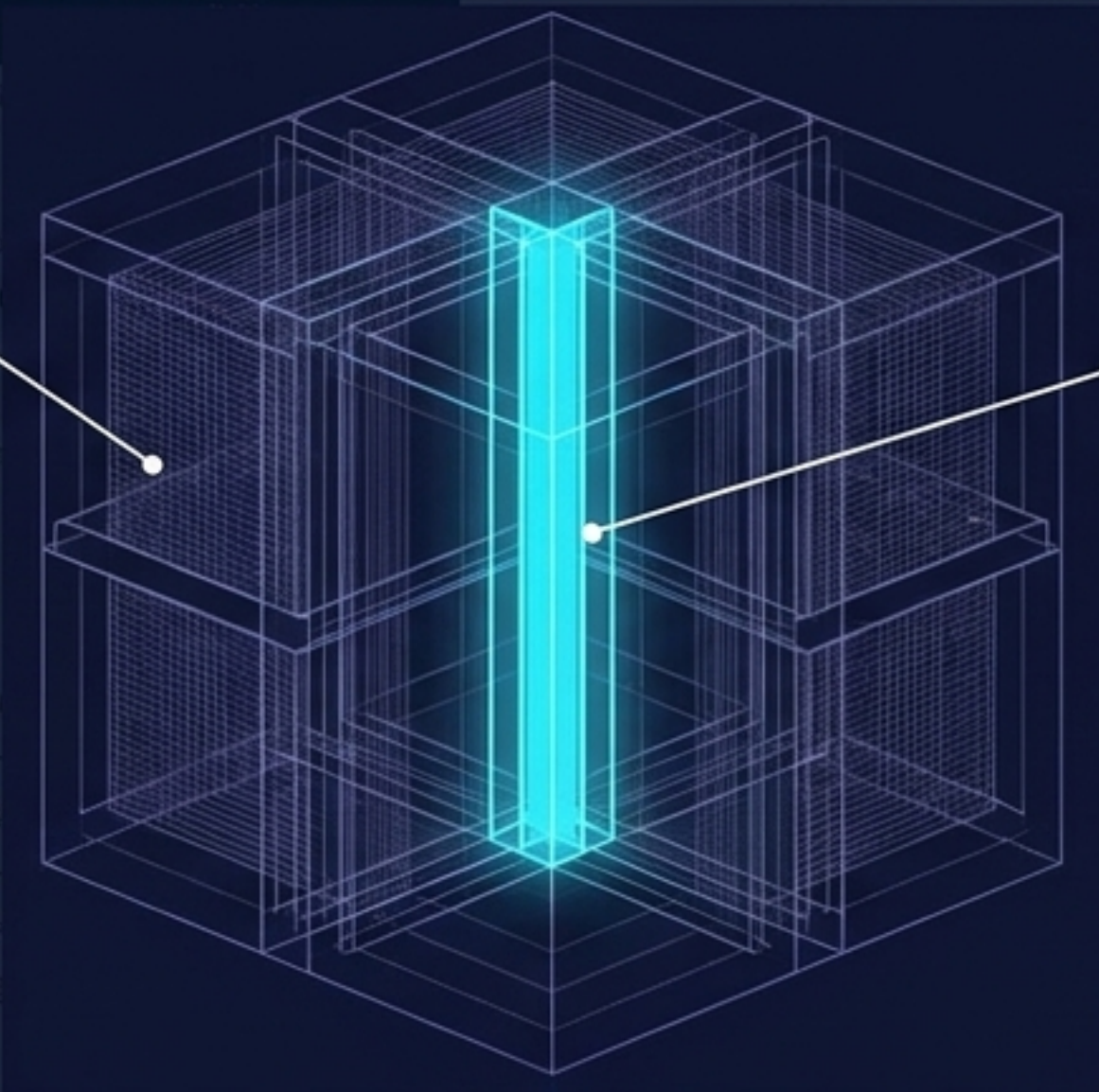
4 Motif 3 の技術的構成

Context Engine:

GDLA (Grouped Differential Latent Attention) により、ノイズを抑制しながら262,144トークン (256K) の長文推論負担を軽減。

Motif 3 はデコーク専用の自己回帰型 MoE で、33 層 (2 dense + 31 MoE) からなる。各層は MoE の 384 のルーティング専門家と 1 つの共有専門家があり、各トークンについて 8 専門家が選択される。大量の専門家を保持しながら、各トークンの計算負担は「13Bクラスの小型モデル」と同等に抑えるアーキテクチャ。ただし、全専門家を保持するため複数GPU (H200/B200等) の分散基盤は必須である。

出所：Artificial Analysis Intelligence Benchmarking Methodology. [2]



24%を合計すると 38%になる。従来型の問題や学問問題だけでなく、外部環境を操作する能力が要求される。2026年6月v4.1 設計仕様、よりエージェント型の業務負荷を反映している。[2.3]

エージェント型ツール利用、専門業務、ソフトウェア長文と変換判断、数学、コードを統合する。この構成は、AAII の評価に適合している。[4]

アクティブパラメータ 約13.2B / トークン。毎回の推論で 8専門家のみを選択 (top-8)。

Motif 3
約 314B / 約 13.2B
53 層 (2 dense + 51 MoE)
384 routed, top-8, shared 1
Grouped Differential Latent Attention (GDLA)
修正版 manifold-constrained hyper-connections (mHC)
Expert Specific PolyNrom
1 層の Multi-Token Prediction head
262,144 トークン

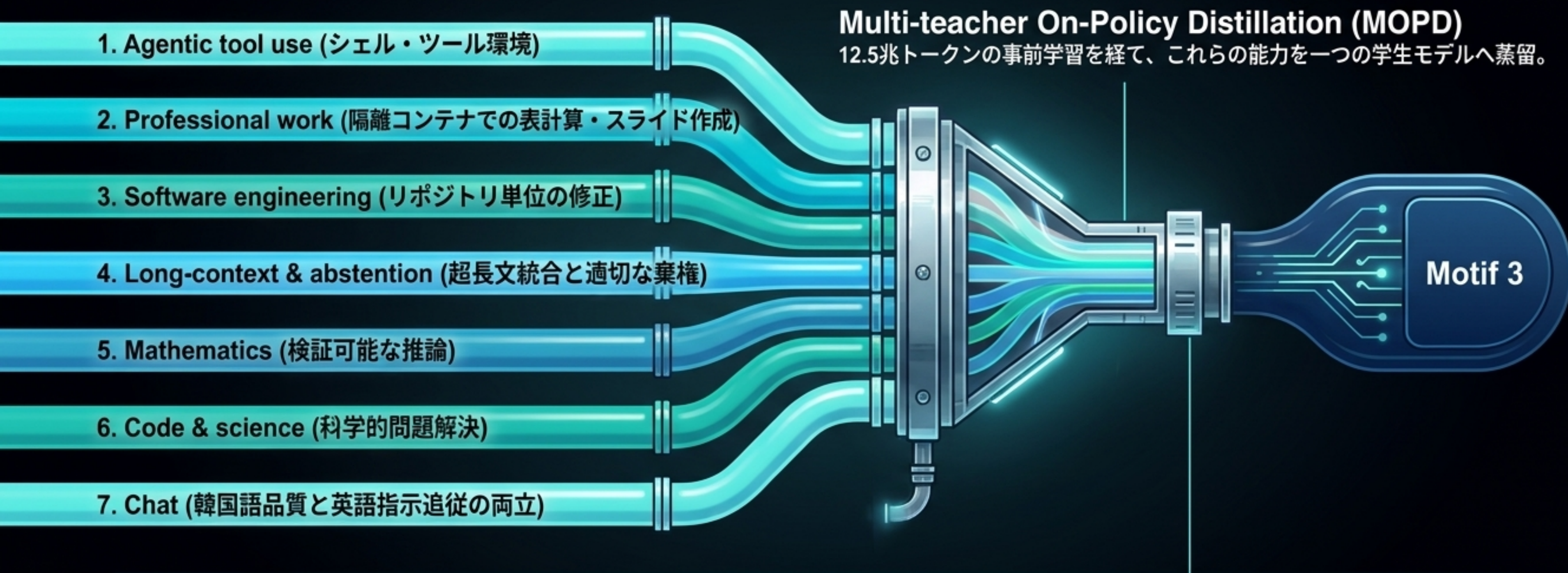
出所：Motif 3 Technical Report, Table 1. [4]

4.2 GDLA：長文推論と KV キャッシュ効率

Multi-head Latent Attention を組み合わせる。固有値を多く、ノイズ抑制効果を少なく低次元でノイズ成分を抑制する設計である。長文推論時の KV キャッシュ負担を抑え、選択性を高める。

巨大な知識容量を持ちながら、各トークンの計算負担は「13Bクラスの小型モデル」と同等に抑えるアーキテクチャ。ただし、全専門家を保持するため複数GPU (H200/B200等) の分散基盤は必須である。

事後学習の戦略的アライメント：7つの専門教師



単一の報酬でモデルを混乱させるのではなく、各能力を別々の教師モデルに専門化させてから統合。
この構成が、AAILの評価配分と構造的に完全に一致している。

韓国の可視化を生んだ「ソブリンAI」政策エンジン

Step 1: ゼロからの設計 (Mandate)

海外モデルの派生を禁じ、
国内設計・独自事前学習を義務化。

Step 2: 大規模クラスタ集中 (Compute)

13,000基の先端GPU (B200/H200)
を確保し、少数チームに数百～数千
GPU単位で提供。

Step 3: 勝ち抜き型評価 (Pressure)

15チームから段階的に脱落させる競争。
評価基準として「13の国際ベンチマーク」
と「世界SOTAとの比較」を要求。

Step 4: 製品としての公開 (Execution)

MITライセンスでのウェイト公開、
英語技術報告、vLLM/SGLang実行
手順の標準化。

Policy-to-
Performance
Flywheel

Motif 3の成功は偶発的ではない。
開発者の目標を「グローバルベンチマークでの勝利」へ強制的に
向けさせる、極めて集中度の高い政策設計の成果である。

日本のLLMは「不在」ではなく「異なるゲーム」をプレイしている

モデル	主たる強み・焦点	なぜAAIの公開ランキングで見えないのか
PLaMo 3.0 Prime (Preferred Networks)	日本語性能とコストの両立、256K、API/オンプレミス提供中心。	商用・API提供が主体であり、独立評価パイプラインに接続しにくい。
tsuzumi 2 (NTT)	単一GPU推論、機密情報処理、日本語ビジネス文書への特化。	企業向け・小型省計算特化であり、大型・長期推論を競うAAIの目的と完全に異なる。
LLM-jp (国立情報学研究所)	モデル・コーパス・ツールの公開、組織横断の研究共同基盤の構築。	研究エコシステムとしての価値は、単一のベンチマークスコアには表れない。

日本勢の「見えなさ」は能力不足ではない。英語・エージェント中心の評価軸との不一致と、国内企業向けに最適化された提供形態（オンプレ・API）による実質的なズレである。

日韓の国家AIエンジン比較：集中突破 vs 多様性ポートフォリオ

【韓国】勝ち抜き型集中投資

- 支援方式：少数精鋭・段階脱落型
- 計算資源：特定チームへの大規模クラスター集中
- 評価基準：13の国際ベンチマーク、世界SOTA比較
- 結果：グローバルリーダーボードでの早期可視化

【日本】GENIAC 多様性支援

- 支援方式：多様テーマ採択のポートフォリオ型（第3期24社、第4期16件）
- 計算資源：複数事業者への分散提供
- 評価基準：社会実装、産業特化（製造、医療、ロボット等）
- 結果：失敗リスクの分散と産業基盤の底上げ、しかし、フラッグシップは不在

GENIACの戦略は誤りではない。しかし、資源を広く薄く分散させる設計上、「世界標準の汎用フラッグシップ」が生まれにくいという数学的必然を抱えている。

日本への提言：「二層型」国家AI戦略への移行



最適解は二者択一ではない。国内特化の明確な経済価値を守りつつ、世界の技術潮流に影響を与える「公開フラッグシップ」を分離して育成する必要がある。

グローバルフラッグシップ獲得へのロードマップ

Phase 1: 0～6か月

- 独立機関による客観評価への接続（API/ウェイトの規格化）。
- 日本独自の強み（特許、製造、医療）を「多言語エージェント評価ベンチマーク」として世界へ公開。

Phase 2: 6～18か月

- 少数チームへのGPU長期・大規模クラスタ配分。
- 事後学習基盤（RL環境、専門教師モデル）の共同資産化による重複開発の排除。

Phase 3: 18～36か月

- 300B超級MoE（または同等能力）の公開モデル完成と国際リーダーボード継続参加。
- 高性能フラッグシップから国内向け小型・オンプレミス版への「蒸留技術」による技術還流。

日本は「国産LLMを作る」段階から、「世界の公開フロンティアで測らせ、勝たせる」段階へシフトしなければならない。これは技術の問題ではなく、システム設計の問題である。