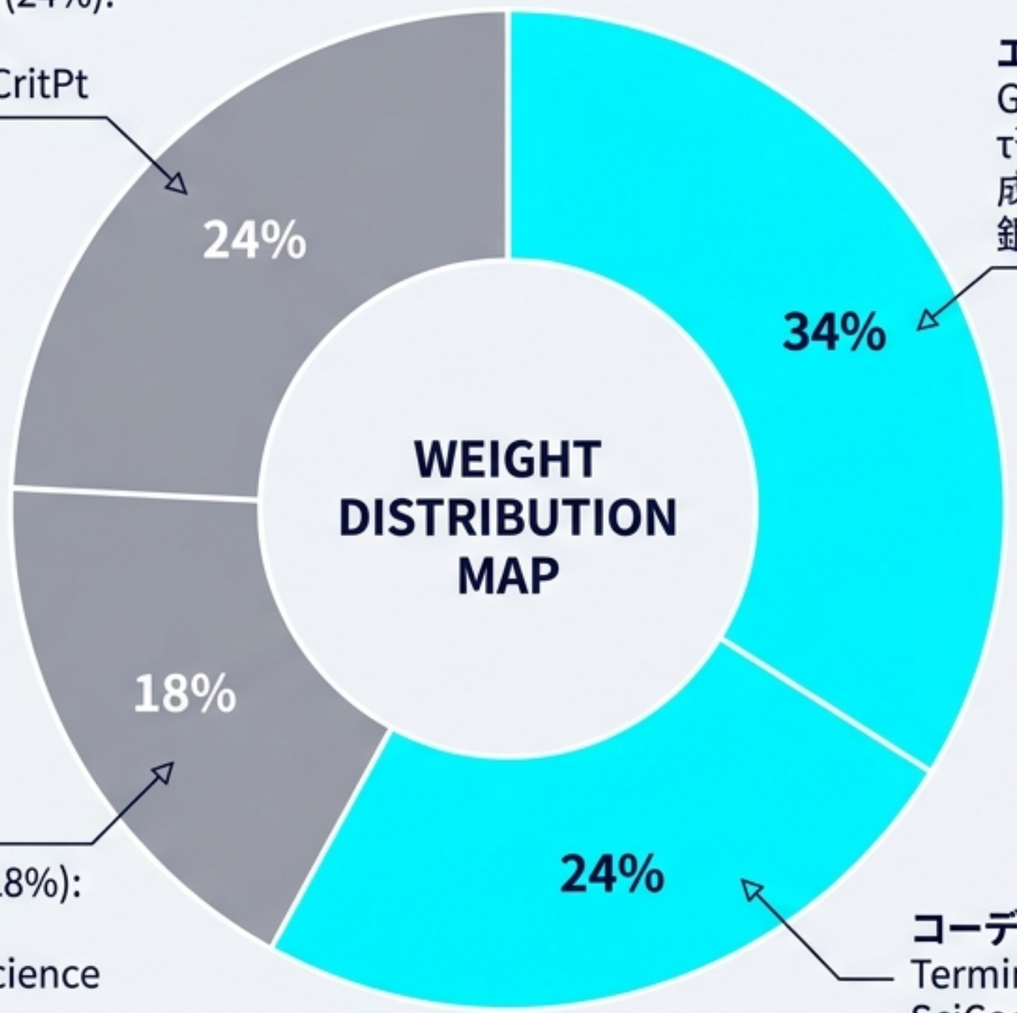
Motif 3 徹底調査: 韓国LLMの飛躍と日本LLMの「見えない」実力

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

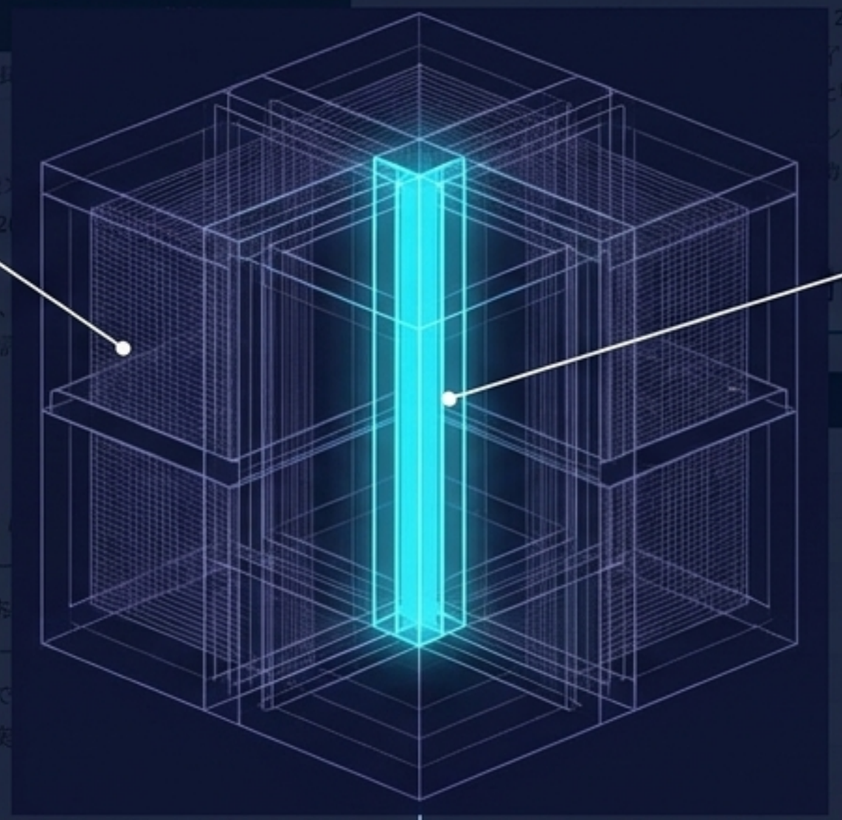
エージェント (34%):
GDPval-AA v2, τ^3-Banking (実務成果物、ツール利用、銀行業務フロー)

一般能力 (18%):
AA-LCR, AA-Omniscience

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



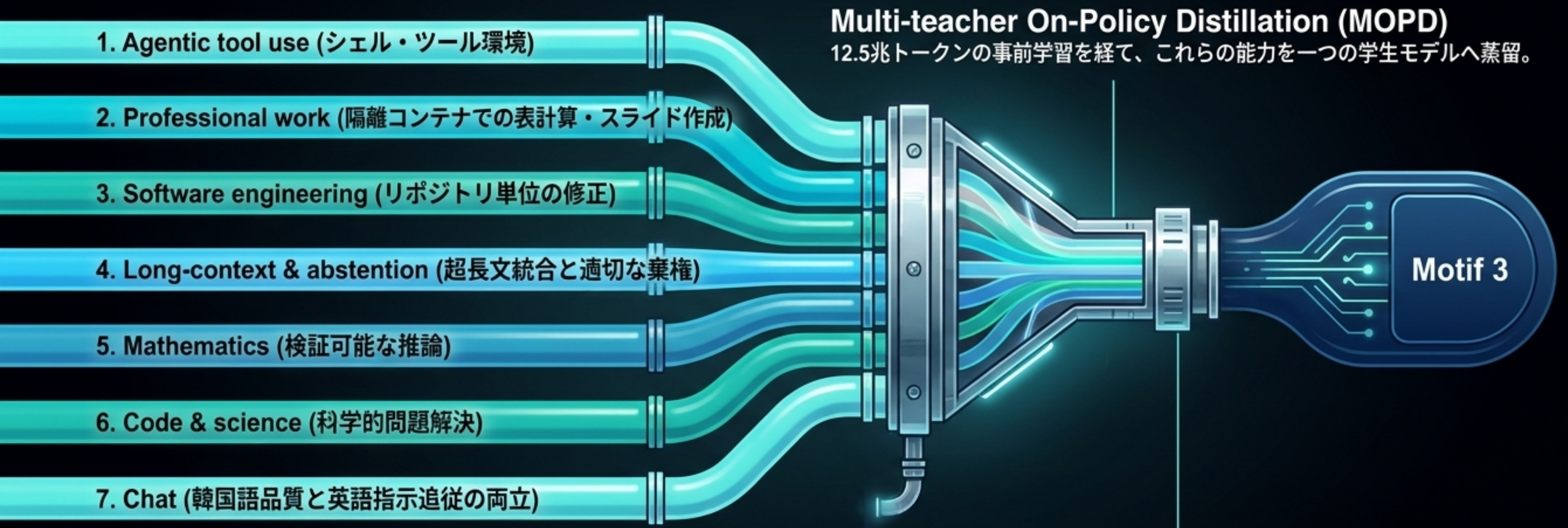
総パラメータ数 約314B。
384のルーティング専門家。

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）。

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。

巨大な知識容量を持ちながら、各トークンの計算負荷は「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、AAII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 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。

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。

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AAII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AAII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

【Track A】
グローバル公開フラッグシップトラック（新規創設）

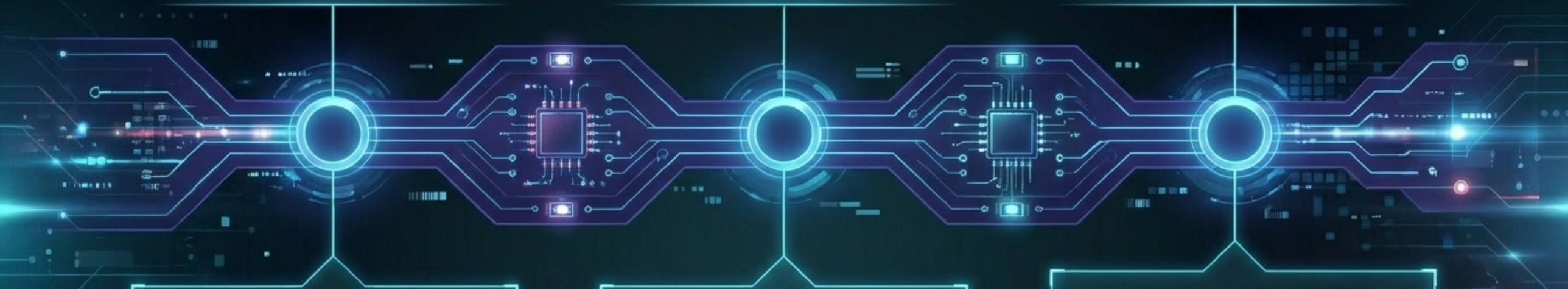
- 役割：1〜2チームを厳選し、世界標準の汎用エージェントモデルを狙う別枠。
- 焦点：数百〜千基単位のGPU複数年配分、大規模MoE、完全なオープンウェイト公開、国際ベンチマーク（SWE-bench, GPQA 等）の突破。

【Track B】
国内産業・社会実装トラック（維持・拡大）

- 役割：GENIACの多様なテーマ支援を継続。
- 焦点：小型・省計算、オンプレミス、業界特化（日本語、法令、商慣行）、安全な国内運用。

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。