韓国製LLM「Motif 3」の技術的革新とグローバルAI評価における日韓エコシステムの構造比較

Gemini 3.7 Flash

韓国の人工知能スタートアップであるMotif Technologiesが開発した「Motif 3」は、AIモデルの第三者評価機関であるArtificial Analysisの総合性能評価指標「Artificial Analysis Intelligence Index（AAII）」において「47点」のハイスコアを記録し、グローバルなフロンティアLLMの領域に突如として躍り出た[1]。この記録は、1.6兆パラメータ規模を誇る中国DeepSeekの最新鋭モデル「DeepSeek V4 Pro」に並ぶものであり、米中以外の地域で開発されたモデルとしては世界最高水準、オープンウェイトモデル全体の中でもトップクラスの位置を占める[1]。
本研究レポートでは、Motif 3の技術的アーキテクチャの全貌、開発背景、および各種ベンチマークにおける実証結果を精緻に解読する。さらに、わずか30名規模の韓国新興企業が巨大プラットフォーム企業に対抗し得るモデルを創出し得たエコシステム的要因と、対照的に日本のLLM開発勢がグローバルな主要リーダーボードにおいて露出を拡大できていない構造的理由について、技術政策・産業戦略の観点から深掘りした比較分析を提示する。

1. Motif 3の全貌とアーキテクチャの画期的突破

開発経緯とプロジェクト的背景

Motif Technologiesは、AI半導体およびシステムソフトウェアの開発企業であるMoreh（モレ）から2025年2月にスピンオフして設立された新興企業である[3]。同社は設立間もない段階で、韓国政府（科学技術情報通信部）が主導するナショナルプロジェクト「独自AIファウンデーションモデル（通称：Dok-Pa-Mo / 独覇模）プロジェクト」の第2期開発主体として追加選定を受けた[4]。
政府から提供された約700枚のGPU基盤と、Moreh時代から培われた並列処理最適化技術を活用し、Motif Technologiesはプロジェクト選定からわずか5ヶ月という異例の短期間でMotif 3の事前学習および事後学習を完了させた[1]。同モデルは、既存のオープンソースモデル（LlamaやQwenなど）のパラメータを微修正した派生モデルではなく、アーキテクチャの基幹構造から完全に自社で新設計（Built from scratch）された高効率なデコーダー専用Mixture-of-Experts（MoE）モデルである[3]。

設計仕様と主要スペック

Motif 3は、全パラメータ規模に対してトークン生成時に実際に計算へ投入されるアクティブパラメータの割合を極小化する「細粒度疎性（Fine-grained Sparsity）」を徹底追求している[2]。

項目	仕様および技術数値
全パラメータ数（**Total Parameters**）	約3,140億（314B）[2]

アクティブパラメータ数（**Active Parameters**）	約132億（13.2B / 1トークンあたり約4.2%）[4]
レイヤー構造（**Layer Configuration**）	全53層（高密度Dense層 2 ＋ MoE層 51）[9]
MoE構造（**Expert Routing**）	384 Routed Experts（Top-8 Routing）＋ 1 Shared Expert[4]
コンテキスト長（**Context Window**）	256K（262,144トークン）[4]
事前学習データ量（**Pre-training Volume**）	約12.5兆トークン（WEB、STEM、コード、数学、法務・金融等）[4]
ライセンス形態（**License**）	MIT License（Hugging Faceにてオープン公開）[3]

独自アーキテクチャのメカニズム解析

Motif 3の競合優位性は、従来のTransformer構造におけるボトルネックを解消するために導入された一連の自社特許級ビルディングブロックにある[3]。

Grouped Differential Latent Attention (GDLA)

従来のアテンション機構は、入力文脈の長大化に伴ってKey-Value（KV）キャッシュのメモリ占有量が急増する点と、文脈に含まれる余分なノイズ情報にアテンション重みが分散する点が課題とされていた[3]。Motif 3では、アテンションマップ差分によってノイズを削除するDifferential Attentionのアイデアと、KVキャッシュを幾何学的に圧縮するMulti-head Latent Attention（MLA）を統合した「GDLA」を考案した[3]。

GDLAでは、信号経路（Signal Path）とノイズ評価経路（Noise Path）のアテンションヘッドを非対称に割り当て、トークンごとに動的に算出される減衰係数 $\boldsymbol{\lambda}_t = \sigma(\mathbf{x}_t \mathbf{W}^{\lambda}) \in (0,1)^{n_S}$ を用いてノイズ成分を減算補正する[8]。これにより、長文脈処理時におけるKVキャッシュのメモリフットプリントを大幅に削減しながら、高い文脈保持精度を維持することに成功した[3]。

Expert-Specific PolyNorm

MoE構造を大スケールで安定して学習させる際、特定のエキスパートネットワークにおいてアクティベーションの突発的な異常値（Activation Outliers）が発生し、全体の学習が不安定化する問題が存在する[3]。Motif 3では、一般的なSiLUアクティベーションと非線性ゲートの組み合わせを排し、各エキスパートごとに独立して学習される多項式正規化層「Expert-Specific PolyNorm」を導入した[3]。これが外れ値の発生を自律的に抑圧し、384個に及ぶ細粒度エキスパートの高度な機能特化を安定して促進させる[3]。

Modified Manifold-constrained Hyper-Connections (mHC)

深層ネットワークにおける勾配消失・爆発を防ぐため、従来の単純な残差接続（Residual Connection）を拡張し、4つの並列残差ストリームによる二重確率混合構造「mHC」を組み込んでい

る[3]。事前学習時には動的なストリーム間干渉の抑制を行い、多層のMoE層を通過する際のアクティベーション値の累積的な膨張を防いでいる[3]。

Multi-Token Prediction (MTP) と Multi-Teacher On-Policy Distillation (MOPD)

予測効率の向上を目的として、1層のMTPヘッドを搭載し、事前学習時の学習効率を高めるとともに、推論時には追加モデルなしでセルフスペキュレイティブ・デコーディングによる高速化を実現している[3]。さらに事後学習（Post-training）では、コード生成、数学的推論、法務・金融実務などに特化した6つの強化学習済み専門教師モデルから、オンポリシー蒸留（MOPD）を用いて単一のUnifiedモデルへ能力を集約させている[9]。

2. ベンチマーク実績とフロンティアモデルにおける競合優位性

AAII指標および専門タスク評価

Artificial Analysis Intelligence Index（AAII）は、単一の静的テストセットでの単語選択にとどまらず、エージェント的タスク遂行能力、長文脈理解、高度な科学的推論など9つの多角的なベンチマークを統合した総合指標である[6]。Motif 3は、主要なオープンウェイトモデルおよび米国企業の商用APIモデルと比較して極めて強力なベンチマーク結果を示している[1]。

モデル名	開発組織 / 国	パラメータ規模 (総 / 活性)	AAII スコア	Terminal-Bench 2.1	SWE-Bench Verified	GPQA Diamond
Motif 3 (Beta)	Motif Tech (韓国)	314B / 13.2B	**47**	**74.9**	**76.2**	**83.4**
DeepSeek V4 Pro	DeepSeek (中国)	1.6T / 非公表	44 - 47	64.0	77.4	-
GLM-5.1 / 5.2	Zhipu AI (中国)	744B / 非公表	51 (GLM-5.2)	61.8	76.4	86.8
Kimi K2.6	Moonshot AI (中国)	1T / 非公表	44	-	76.2	91.1
Solar Open 2	Upstage (韓国)	非公表	37	-	-	-
A.X-K2	SK Telecom	688B / 非公表	35	-	-	-

	(韓国)					
K-EXAONE 2.0	LG AI Research (韓国)	非公表	31	-	-	-

実務型エージェント能力と回答校正

Motif 3のベンチマークにおいて最も顕著な特徴は、CLI（コマンドライン）環境における複雑な操作能力を計測する「Terminal-Bench 2.1」で記録した「74.9点」というスコアである[1]。これはGLM-5.1の61.8点やDeepSeek V4 Proの64.0点を10ポイント以上上回る数値であり、実際のシステム構築やコード修正を伴う自律型エージェント環境での能力の高さを提示している[1]。

また、金融機関の実務業務シナリオでの遂行能力を評価する「τ^3 -Banking」においては35.3点を記録し、DeepSeek V4 Proの30.1点を凌駕した[1]。さらに、根拠のない虚偽応答を抑止するハルシネーション評価「OmniScience Non-Hallucination」において71.6点を獲得しており、自身が確証を持てない高度な設問に対して適切に応答を辞退（Abstention）する高い確率校正精度を備えている[3]。

3. 韓国AIエコシステムが高い世界評価を獲得する構造的背景

韓国のLLM開発勢がグローバル市場で極めて高い評価を獲得している背景には、政府による国家的なソブリンAI推進策と、市場の要請に対応した開発企業の高効率な開発戦略が相互に機能しているエコシステムが存在する[4]。

ソブリンAI構想「Dok-Pa-Mo」とスタートアップ主導型開発

韓国政府は、米中の巨大ITプラットフォームに対する技術的従属を回避し、自国の文化的・言語的アイデンティティとデータ主権を維持するために「ソブリンAI（Sovereign AI）」戦略を強力に推進している[4]。その中核である「Dok-Pa-Mo（独覇模）」プロジェクトは、単に大型助成金を給付するだけでなく、実効的な計算インフラであるGPUリソースを直接提供し、短期間で測定可能なベンチマーク成果を上げることを参加企業の至上命令として課している[1]。

この枠組みにおいて特筆すべきは、大企業（LG AI Research、SK Telecom、Naverなど）と新興スタートアップ（Upstage、Motif Technologies）を同一の土俵で競わせるコンペティショナルな構造である[1]。これにより、大企業特有の社内政治や意思決定の遅延を排し、Motif Technologiesのような僅か30名規模の技術者集団が最先端のアルゴリズム実装に特化して突出した成果を叩き出す環境が醸成された[2]。

資本力の劣勢を跳ね返すアルゴリズム優先戦略

米中の巨大企業が数万台規模のGPUクラスタを投入して数兆パラメータのモデルを力任せに学習させるアプローチに対し、韓国のベンチャー勢は「パラメータ効率と推論コストの極小化」に焦点を当ててきた[2]。Motif 3の設計に見られるように、全体で314Bものパラメータ容量（知識蓄積量）を保持しな

がら、1トークンの生成時に稼働するのはわずか13.2B（約4%）に過ぎない[2]。この超疎性 MoE構造と独自アテンション（GDLA）の組み合わせにより、限られた計算リソースでありながら、巨大モデルと互角以上の高度な推論・エージェント能力の抽出に成功している[2]。

性能評価を巡る課題：ベンチマーク最適化論争

その一方で、韓国LLM業界の急速なスコア上昇に対しては、過剰なベンチマークハックやデータ汚染に関する議論も発生している[2]。2026年8月、Dok-Pa-Moプロジェクトの最終選定を控え、一部の参加企業がベンチマークのスコア向上を請け負う海外のコンサルティング企業（米AfterQueryなど）と接触していたのではないかという疑念が現地報道で取り上げられた[2]。特定ベンチマークのテストセットに対する「Benchmaxxing（スコア最適化）」が過剰に行われた場合、ベンチマーク上の数値が高くとも未知の実務問題に対する汎用性能が伴わないという「過学習リスク」が指摘されている[2]。
これに対し、主務官庁である韓国科学技術情報通信部は、公開された自動評価ベンチマークだけでなく、非公開データによるブラインドテスト、領域エキスパートによる直接審査、および一般市民による利便性評価（Usability Assessment）を複合的に組み合わせた多角的な性能実証を進めており、評価の公正性と実用性の担保に注力している[2]。

4. 日本のLLMがグローバル主要リーダーボードで露出しない要因分析

日本国内においても、経済産業省とNEDOが主導する「GENIAC」プロジェクトや、理化学研究所・東京工業大学等による「富岳LLM」、Preferred Networks（PFN）による「PLaMo」シリーズなど、活発なモデル開発活動が行われている[12]。しかしながら、Artificial Analysisのようなグローバル主要リーダーボードにおいて日本のLLMモデルがトップランカーとして表面化することは極めて稀である[3]。この不整合は、単なる技術力の落差ではなく、日韓における開発戦略、評価ターゲット、および企業市場の需要構造の違いに起因している[12]。

評価目標と言語域のミスマッチ

Artificial Analysis Intelligence Index（AAII）をはじめとするグローバル評価基盤は、主に英語をベースとした高度な数学、プログラミング、論理的思考、および複合的ツール利用の能力を測定するよう設計されている[10]。
これに対し、日本の国産LLM開発者たちの主要な開発生成物は、「高度な日本語理解」「日本語独特の敬語・文脈・ビジネス慣習への追従」「日本独自の法制・文化知識の補完」に特化されている[12]。国内モデルの性能測定には、Jaster-GENIAC、JMT-Bench、ELYZA-tasks-100、あるいは日本語指示追従性能を可視化するJFBenchなどが標準的に用いられている[13]。日本語ベンチマークにおいて海外のフロンティアモデルを上回る成果を出していても、それらの成果指標がグローバルな英語中心リーダーボード（AAII等）へ直接接続されないため、世界的ランキング上では「日本の存在が認識されない」という非対称性が生じている[13]。

日本産業界における「Dual-Stack」導入戦略の定着

日本のエンタープライズ市場における生成AI活用は、海外フロンティアモデルと国産ローカルモデルを役割分担させる「Dual-Stack（二重化）」アプローチへと急速に収れんしている[19]。
高度な汎用推論、複雑なコード生成、およびグローバル展開業務には、OpenAIのGPT-5、

AnthropicのClaude、GoogleのGeminiといった最高峰の海外SaaS型モデルを採用する[15]。一方で、オンプレミス環境や高度な機密保持が求められる国内業務、官公庁・金融・医療・製造などの特定ドメイン、およびコスト効率を重視した大量の定型処理には、国産の小型・中型モデルやドメイン特化型モデル（PLaMo、TSUZUMI、創薬特化AI等）を適用する戦略が一般的である[12]。
この構造下では、国産モデルに「米中フロンティアモデルと汎用知能のトップラインで競り合うこと」は投資対効果の観点から必ずしも要求されず、「特定領域での高い安全性、低遅延、および低コスト」が最重視される[12]。結果として、世界ランキングの上位を狙うような汎用巨大モデル開発への開発資材集中が起きにくい環境が形成されている[12]。

アーキテクチャ創出支援とエコシステム公開戦略の差異

日韓の開発プロジェクトを比較した際、基礎的ビルディングブロックの創出に対するアプローチの違いも明確である[3]。Motif 3の事例に見られるように、韓国の戦略はアテンションや正規化層そのものをゼロから自社設計し、汎用ベンチマークのトップを獲得することで自国の技術威信（Technological Prestige）を世界に誇示するブランディング戦略をとっている[3]。
対照的に、日本のGENIAC等の施策では、既存の信頼性の高いアーキテクチャ（Llama等）をベースにした日本語継続事前学習（Mid-training）や、スーパーコンピュータ「富岳」等の国内固有インフラにおける動作最適化、あるいは複数モデルの連携（Sakana AIに見られるオーケストレーション技術等）にリソースが多く充当されてきた[12]。また、開発されたモデルのオープンウェイト公開や、Artificial Analysis等のグローバル第三者評価ハーネスへの積極的・能動的なエントリー・API提出に対するインセンティブが弱かった点も、グローバル露出の乏しさに拍車をかけている[3]。

5. 結論と日本AI産業への提言

韓国の「Motif 3」が提示した成果は、巨額の資本力と巨大な計算資源を持つ米国・中国の巨大IT企業に対し、徹底的なアーキテクチャの効率化とスマートな設計論（細粒度MoE、GDLA、PolyNorm等）を駆使すれば、僅か30名規模の技術者チームであっても世界の最前線に立ち得るという事実を実証した[2]。これは、基盤モデル開発における競争の軸が、単なるパラメータ規模の拡大（Brute-force Scaling）から、アルゴリズムと推論効率の質的転換へとシフトしていることを明示している[3]。
日本のAI産業および政策主導プロジェクトが、国内市場での実務的有用性を維持しつつ、グローバルな技術的発言力を確保するためには、以下の戦略的アプローチが不可欠である。
第一に、既存モデルの継続学習にとどまらず、アテンション機構やメモリ圧縮、計算疎性の創出といったアーキテクチャの根幹部分に対する基礎研究開発へ戦略的投資を傾注することである[3]。
第二に、国内専用ベンチマークでの成果にとどまることなく、Artificial AnalysisやSWE-benchといった世界標準の独立リーダーボードへ能動的に参入し、自国の技術的到達度を客観的かつ定量的なデータで世界へ向けて発信・証明する姿勢を強めることである[1]。
第三に、過度なベンチマークハック（Benchmaxxing）に陥ることを警戒しつつも、Motif 3に見られるようなターミナル操作や金融・法務実務などの「高度な自律型エージェント環境（Agentic Framework）」で機能する実用型推論能力の獲得へ開発の焦点（Focus）を合致させることである[1]。国内市場に閉じた最適化から一歩を踏み出し、極限の推論効率と客観的性能評価を両立させる開発エコシステムの構築が、今まさに日本のAI戦略に求められている[3]。

引用文献

1. Motif Technologies Releases Open-Source 'Motif 3' — Tops South Korean Field with AAII Score of 47 - BigGo Finance, https://finance.biggo.com/news/57ab7685-e1d6-4f97-9a48-8bf1aabcc3d6
2. Benchmark Score Manipulation Allegations Surface Ahead of South Korea's National AI Model Second-Round Evaluation - BigGo Finance, https://finance.biggo.com/news/463fa495-706e-4d6e-ac28-e3279aef08b4
3. South Korean AI Attacks with 314B MoE—Motif-3 and the Asian LLM War｜みさき - note, https://note.com/bright_jacana710/n/naf15992560a0?hl=en
4. Korean AI Attacks with 314B MoE—Motif-3 and the Asian LLM War｜みさき - note, https://note.com/bright_jacana710/n/n53b459fda38f?hl=en
5. Motif-3 Leads Outside U.S./China, Ranks Third in Open-Source AI, https://www.chosun.com/english/industry-en/2026/07/21/C66HLBGLXFCDRORTJ44H3EROBA/
6. 自社設計LLM「Motif-3」、AAII 44点…DeepSeek V4 Pro（max）と同スコア - KORIT, https://www.korit.jp/news/ai/platum-motif-technologies-llm-260725/
7. Korean AI Strikes with 314B MoE—Motif-3 and the Asian LLM War｜みさき - note, https://note.com/bright_jacana710/n/nfa1a1da6bff2?hl=en
8. Motif 3: Technical Report - arXiv, https://arxiv.org/html/2608.09119v1
9. Motif-Technologies/Motif-3 - Hugging Face, https://huggingface.co/Motif-Technologies/Motif-3
10. Artificial Analysis: AI Model & API Providers Analysis, https://artificialanalysis.ai/
11. The Sovereign AI Debate and Prospects of “Korean” AI - Korea Economic Institute, https://keia.org/analysis/the-sovereign-ai-debate-and-prospects-of-korean-ai/
12. GENIAC - T&Mコーポレーション株式会社, https://tm-co.co.jp/glossary/geniac/
13. GENIAC事業 第1期 性能評価結果公開 - 経済産業省, https://www.meti.go.jp/policy/mono_info_service/geniac/selection_1/result_1/index.html
14. 生成AI・LLM必須リンク集2026 - Qiita, https://qiita.com/aokikenichi/items/fca5a1898cc2a9c961f6
15. 【2026年最新】LLM比較表・性能ランキング！LLM比較サイト一覧 | Data Driven Knowledgebase, https://since2020.jp/media/llm-ranking/
16. Sovereign AI in Asia: Why Every Country Is Building Its Own LLM, https://digitalinasia.com/sovereign-ai-asia-every-country-building-own-llm/
17. [2025 Latest] 4 AI Tool Ranking Sites | (2) Benchmarking LLM ... - note, https://note.com/onemorevision/n/n8f19e37c116e?hl=en
18. LLMs4All: A Review of Large Language Models Across Academic Disciplines - arXiv, https://arxiv.org/html/2509.19580v5
19. Where is LLM Access Available Across Asia in 2026? ChatGPT, Claude, Gemini, and Chinese Models Tracker - Digital in Asia, https://digitalinasia.com/which-llms-work-asia-accessibility-tracker/
20. Intelligenz-Benchmarking | Artificial Analysis, https://artificialanalysis.ai/de/methodology/intelligence-benchmarking

21. PLaMoベースの強い日本語報酬モデルの構築 - Preferred Networks Tech Blog, https://tech.preferred.jp/ja/feed
22. Riding a New Wave: Is ByteDance's Waver 1.0 the Future of AI Video? - Skywork, https://skywork.ai/blog/riding-a-new-wave-is-bytedances-waver-1-0-the-future-of-ai-video/
23. Kimi K3で上がったのは天井ではなく床｜AIの下限が変わった日, https://trendmap.jp/blog/kimi-k3-open-weights-2026
24. 生成AI・デジタルニュース（26/08/13）｜Kick@AnyGrab｜スキルアップ・人材育成 - note, https://note.com/anygrab_kick/n/n20a35f55b37f
25. Changelog · Directories by Enterprise DNA, https://enterprisedna.co/directories/changelog/