Xiaomi MiMo-V2.6-Proのシステムアーキテクチャと本番運用の実現性

Gemini 3.8 Flash

Xiaomiが2026年9月21日（北京時間9月22日）にオープンウェイトとして公開した「MiMo-V2.6-Pro」は、総パラメータ数1.02兆（1.02T）、トークンあたりの活性化パラメータ数420億（42B）のスパースMixture-of-Experts（MoE）構造を備え、100万トークン（1M tokens）のコンテキストをネイティブ処理する全モダリティ（Omnimodal）基底モデルである[1]。第三者評価機関Artificial AnalysisのIntelligence Index v4.3.2においてスコア「46.32」を記録し、GLM-5.3やKimi K3、DeepSeek V4系を上回ってオープンウェイトモデルの首位を獲得した[3]。

本モデルの設計思想における中核概念は、「二値評価を脱却した自己改善強化学習（RL）のスケールアウト」にある[2]。従来の合否判定（Pass/Fail）のみに頼るRLが引き起こす冗長かつ防衛的なコード生成や報酬ハッキングを抑止するため、複数試行を同一空間で相対比較するGAR（Groupwise Advantage Redistribution）を実装した[2]。さらに、コーディング、汎用タスク、視覚設計、サイバーセキュリティの全領域を単一の実行バッチ内で混成学習させる「You Only RL Once」パラダイムを確立し、エージェントとしての実務遂行能力を大幅に引き上げた[2]。

本番導入における意思決定基準として、MiMo-V2.6-ProはClaude Opus 5と同等水準のエージェントタスク成功率を、入力100万トークンあたり0.435ドルという商用フロンティアモデルの20分の1から60分の1に達する劇的な低コストで実現する[1]。しかし、活性化が42Bであっても推論ホスティングには1.02Tの全重みとKVキャッシュをメモリに保持する必要があり、TP16/EP16によるマルチノード環境が不可欠となる[9]。加えて、リリース初期におけるAPIルーティング経路の単一依存（SPOF）や、実質110K〜150KトークンにとどまるRL実証軌跡長を踏まえた1Mコンテキストの実効推論精度の検証など、インフラエンジニアリング上の明確なトレードオフの精査が求められる[5]。

1.02TスパースMoEと全モダリティ統合アーキテクチャ

MiMo-V2.6-Proは、テキスト、画像、動画、音声を中間変換層を挟まずに同一シーケンス空間で直接処理する統合型全モダリティモデルとして設計されている[2]。長大なコンテキストを低レイテンシで処理するため、Sliding Window Attention（SWA）とGlobal Attention（GA）を混成させたハイブリッド注意機構を採用している[2]。

コンポーネント	仕様およびハイパーパラメータ
基本アーキテクチャ	スパース Mixture-of-Experts (MoE) + ハイブリッド SWA/GA[2]
パラメータ規模	総パラメータ数: 1.02T (1兆200億) / トークン活性化: 42B (420億)[2]
レイヤー構成	全70層 (Sliding Window Attention 60層 +

	Global Attention 10層)[2]
隠れ層次元 / 注意機構	Hidden Size: 6144 / Q Heads: 128, KV Heads: 8 (QK dim: 192, V dim: 128)[2]
スライディングウィンドウ長	128トークン (ローカルアテンションブロック)[2]
エキスパート構成	384 Routed Experts / トークンあたり Top-8 活性化 (共有エキスパートなし)[2]
コンテキスト長	入力最大: 1,048,576 トークン (1M) / 生成最大: 131,072 トークン[2]
視覚エンコーダ (MiMo ViT)	681M パラメータ (28層: 24 SWA + 4 GA), Patch 2x16x16, Merge 2x2[2]
音声エンコーダ群	AudioTokenizer (308M, 20 RVQ) + Patch Encoder (127M, 25Hz→6.25Hz)[2]
投機的デコーダ (MTP)	5層 SWA (DFlashスタイル), 1ステップあたり後続7トークンを並列予測[2]

言語バックボーンの第1層には高密度のDense FFNを備えたグローバル注意機構を配置し、以降の層では384個のエキスパートから動的に8個を選択するスパースFFNで構成される[2]。視覚エンコーダには28層（24層のSWAと4層のGA）からなる681Mパラメータの独自ViTを組み込み、時間軸2フレーム×空間16×16ピクセルを1パッチとしてトークン化を行う[2]。音声処理においては、20コードブック構成のResidual Vector Quantization（RVQ）を持つ308MのAudioTokenizerと、フレームレートを25Hzから6.25Hzに集約する127MのAudio Patch Encoderが連携し、低遅延なストリーミング入力を実現している[2]。

MoE構造に起因するデコード時のメモリ帯域ボトルネックを緩和するため、5層のSWAで構成されたMulti-Token Prediction（MTP）ドラフターが統合された[2]。メインモデルの1回のフォワードパスに対して後続7トークンを並列予測・検証する投機的デコーディングを実行することで、サービング時の実効スループットを大幅に引き上げている[2]。

自己改善強化学習のメカニズムと訓練安定化技術

従来のLLM強化学習はドメインごとに独立してファインチューニングされ、テストスイートの実行結果に基づく合否（0または1）を報酬としていた[2]。MiMo-V2.6-Proは、この手法では「長大で防衛的なハックコード」が正解扱いとなりモデルの思考が劣化するという課題に対し、RL計算量・環境多様性・グレーダー計算量を同時にスケールさせる新パラダイムを提示した[2]。

単一混合実行「You Only RL Once」と非同期GRPO

複数ドメインのRLを個別に反復するのではなく、コーディング（68%）、汎用タスク（12%）、視覚設計（13%）、サイバーセキュリティ（4%）、長文コンテキスト（3%）を同一バッチ内に混在させて単一のRL工

程を実行する[2]。異なるドメインのタスクとエージェントハーネスを共存させることで、コード生成で培われた論理検証能力がサイバーセキュリティや視覚推論へ汎化・転移する相互強化作用をもたらした[2]。

インフラ層では、Group Relative Policy Optimization（GRPO）を完全非同期で分散展開した[2]。各更新ステップにおいて、1,568個のプロンプトからそれぞれ16個のロールアウト（1ステップあたり計25,088試行）を生成し、1更新あたり2.7B〜3.7Bトークンを消費する[2]。実行時間のばらつきによるバッチ遅延を解消するため、タスクの実行完了率と許容スロットを適応的に管理する分散キャッシュ基盤が構築されている[6]。

相対評価型グレーディング（GAR / GRS）による報酬設計

二値報酬の限界を克服するため、同一プロンプトに対する16個の兄弟ロールアウトを同一ワークスペース内で相互比較・採点する「Groupwise Advantage Redistribution（GAR）」を導入した[2]。GARはテストを通過したパッチに対し、5つの評価軸に基づいて順位付けを行い、報酬アドバンテージを再配分する[6]。

評価軸	検証基準とアルゴリズムの挙動
問題適合性	提示された課題に対する解法アプローチが直接的かつ本質的であるかを評価する[6]。
実装の簡潔性	不要なフォールバック処理や過剰な防御的例外処理の排除度合いを測定する[6]。
変更の最小性	要求仕様を満たすパッチの差分（Diff）が最小限に抑えられているかを検証する[6]。
スコープ隔離性	課題のスコープ外にある既存コードやシステム設定への偶発的改変を検知する[6]。
コーディング規約	周囲の既存コードベースが採用している構文規則や設計作法との整合性を評価する[6]。

外部から解決策を不正スクレイピングするような明白な「報酬ハッキング」はスコアゼロとして厳格に排除される[6]。また、オフラインでタスク固有のルーブリックを作成し品質スコアを付与するGroupwise Reward Synthesis（GRS）を併用することで、トークン数の肥大化とターン数の発散を抑え込み、簡潔で堅牢な思考パスを自律選択するループを閉じた[2]。このGAR機構は、MiMo-V2.6-Proの全RL計算コスト（約262万ドル）の約12.7%を占めている[6]。

MoEルータ凍結による負荷偏位の抑制

超大規模RLの過程において、MoEのルーティング重みを更新対象に含めると、エキスパート間の負荷偏位（Expert Load Imbalance）が急激に悪化する現象が確認された[6]。実験データによると、わずか20ステップの訓練で負荷変動係数は0.78から2.0に急増し、ピーク負荷は平均の16倍、完全アイドル状態の休眠エキスパートは0.5%から22%に跳ね上がった[6]。

検証の結果、この崩壊はエキスパート内部の表現力低下ではなく、ルータ重みのドリフトに起因することが判明した[6]。XiaomiチームはRL訓練全体を通じて「MoEルータを完全に凍結（Freeze）」するアプローチを採用し、エキスパートの負荷分散の平滑性を維持しながら安全なバッチスケーリングを実現した[6]。

性能ベンチマークと主要競合モデルとの比較

MiMo-V2.6-Proは、多領域にわたるエージェント評価において商用フロンティアモデルに迫るスコアを達成している[1]。

評価カテゴリ / ベンチマーク	MiMo-V2.6 Pro	MiMo-V2.6 Flash	MiMo-V2.5 Pro	Claude Opus 5	GPT-5.6 Sol	GPT-6 Astra
総合知能指数						
AA Intelligence Index v4.3.2	**46.32** [cite: 3, 9]	-	26.0[3]	51.0[3]	47.0[3]	53.0[3]
コードエージェント						
DeepSWE v1.1	**71.9** [cite: 2, 3]	67.9[2]	19.0[2]	74.0[2]	73.0[2]	-
Program Bench	**26.5** [cite: 2, 3]	26.0[2]	12.5[2]	37.0[2]	25.0[2]	-
MiMo Code Bench (In-house)	**63.2** [cite: 2, 3]	61.2[2]	40.4[2]	68.6[2]	59.3[2]	-
汎用・デスクトップ操作						

AutomationBench v1.0.6	**53.1** [cite: 2, 3]	52.3[2]	16.0[2]	50.3[2]	45.8[2]	46.2[2]
Toolathlon-Verified	**76.9** [cite: 2, 3]	73.6[2]	49.1[2]	80.6[2]	74.9[2]	77.9[2]
GDPval-AA 2.1 (Elo)	**1673** [cite: 2, 3]	-	1107[2]	1708[2]	1588[2]	1595[2]
Agents' Last Exam	**31.6** [cite: 2, 3]	27.6[2]	13.2[2]	31.6[2]	30.8[2]	25.7[2]
Terminal Bench 4.0	**34.9** [cite: 2, 3]	28.8[2]	1.5[2]	49.0[2]	39.9[2]	59.6[3]
Terminal Bench 2.1	**89.9** [cite: 2, 3]	87.6[2]	65.2[2]	89.1[2]	88.8[2]	84.3[2]
OSWorld-Verified	**82.0** [cite: 2, 3]	80.8[2]	-	83.4[2]	83.0[2]	86.0[2]
JobBench	**62.0** [cite: 2, 3]	61.2[2]	25.0[2]	65.7[2]	45.4[2]	57.4[2]
セキュリティ・視覚						
CyberGym	**94.0** [cite: 2, 3]	95.1[2]	40.0[2]	-	-	-
MiMo Cyber Bench	**80.2** [cite: 2]	77.2[2]	0.0[2]	-	-	-
ExploitBench	**47.9** [cite: 2, 3]	25.3[2]	16.6[2]	70.0[2]	78.5[2]	100.0[3]

SEC Bench Pro	**66.3** [cite: 2, 3]	47.5[2]	17.7[2]	-	79.1[2]	-
MiMo VisualCoding	**72.3** [cite: 2, 3]	71.5[2]	-	70.0[2]	73.4[2]	69.1[2]

オープンウェイトモデルにおける競争力

Artificial Analysis Intelligence Indexにおいて、MiMo-V2.6-Proは46.32を記録し、競合するオープンウェイトモデルであるGLM-5.3（45）、Kimi K3（44）、DeepSeek V4.1 Flash（39）、DeepSeek V4 Pro（36）を抑えて首位に立った[3]。前世代MiMo-V2.5-Proのスコア26からの向上幅は20ポイントにおよび、特にTerminal Bench 4.0（1.5から34.9へ）やAutomationBench（16.0から53.1へ）に見られる通り、ツール呼び出しと環境フィードバックを伴うエージェント系タスクで大幅な性能向上を示している[2]。

クローズドモデルとの能力差

Claude Opus 5（スコア51）やGPT-6 Astra（スコア53）といった最上位クローズドモデルと比較すると、AutomationBench（53.1）やTerminal Bench 2.1（89.9）、MiMo VisualCoding（72.3）など一部のタスクで匹敵あるいは上回る結果を示している[2]。

一方で、高度なシステム脆弱性の探索と悪用を伴うExploitBenchにおいてGPT-6 Astraの100.0やClaude Opus 5の70.0に対して47.9にとどまり、Terminal Bench 4.0（34.9対49.0/59.6）でも明確な性能ギャップが存在する[2]。この差は、安全機構や長期推論探索の安定性における最適化の深度に起因するものと分析される。

提供形態・エコシステムおよび経済性

Xiaomiは、単なるウェイト公開にとどまらず、研究再現用のタスクセット、蒸留モデル、デスクトップ環境、多層的なAPIラインナップを同時に展開している[1]。

提供形態 / モデルエディション	目的と特徴	価格体系 **(Input / Output / Cache)**
MiMo-V2.6-Pro-RL (オープンウェイト)	フルスペックMoEモデル。Hugging FaceおよびModelScopeで配布[1]	MITライセンス (無料)[3]
MiMo-V2.6-Pro (Standard API)	クラウド推論エンドポイント。OpenAI/Anthropic互換API[1]	$0.435 / $0.87 (Cache: $0.0036) / 1M tokens[3]
MiMo-V2.6-Pro (Batch API)	非同期バルクジョブ用。50%の料金割引が適用[3]	$0.2175 / $0.435 / 1M tokens[3]

MiMo-V2.6-Pro-UltraSpeed	最大20倍の生成速度に最適化された低遅延ホスト層[1]	$4.35 / $8.70 (Cache: $0.036) / 1M tokens[3]
MiMo-V2.6-Flash-RL	309B総パラメータ / 15B活性化の軽量高効率版[3]	$0.14 / $0.28 / 1M tokens[3]
MiMo-V2.6-Distill-Qwen-9B	Qwen3.5-9Bベースの蒸留版。オープンRL研究用ベース[1]	MITライセンス (無料)[13]
MiMo Desktop	PC操作自動化・エージェント実行環境 (Win / Mac)[1]	サブスクリプション / 独自API キー設定可[1]

コスト対効果の実測検証

Standard APIにおけるMiMo-V2.6-Proの価格設定（入力0.435ドル / 出力0.87ドル / 100万トークン）は、先行するV2.5世代から据え置かれている[1]。プロンプトキャッシュヒット時には99%の割引（0.0036ドル/1M tokens）が適用され、キャッシュヒット7割、入力2割、出力1割の一般的なエージェント利用環境における加重平均コストは約0.18ドル/1M tokensまで圧縮される[5]。

Kingy.aiによる実践検証では、OpenCode内で23項目の自動テストを含む3つのコーディング・データ処理タスクを実行した際、MiMo-V2.6-Proは23項目すべてをクリアし、総消費コストは約0.03ドルであった[3]。同一テストで満点を記録したClaude Opus 5の消費コストが約1.03ドルであったことと比較して、実タスクにおけるコスト効率は30倍以上と報告されている[3]。Artificial Analysisのインデックスタスク換算でも、1タスクあたり約0.13ドル（全テスト実行費用206.66ドル）であり、Claude Opus 5の1タスクあたり5.86ドル（約45倍）に対して大きなコスト優位性を持つ[3]。

研究開発環境と派生ツールの公開

Xiaomiは研究コミュニティの再現性を高めるため、強化学習で使用した7,000件以上のタスク環境、学習フレームワーク、軽量実行基盤「mini-harnesses」をGitHub上のmimo-ossリポジトリ群を通じて公開した[1]。

さらに、Qwen3.5-9Bをベースに MiMo生成データで教師あり微調整（SFT）を行った「MiMo-V2.6-Distill-Qwen-9B」を同時に提供している[1]。研究者がこの9Bモデルに対して公開環境を用いてRLを追加適用したところ、SWE-bench Verifiedのスコアが61.1から66.2へ、内部サイバーセキュリティベンチマークが31.3から47.0へと伸長し、RLパイプライン自体の転移性と再現性が実証された[6]。

クライアント製品としては、画面認識とGUIツール操作を統合した「MiMo Desktop」がWindowsおよびApple Silicon Mac向けにリリースされた（法的・規制要件の精査により、韓国、英国、EU加盟国では提供保留）[1]。リアルタイム対話を要求するプロダクション向けには、推論速度を最大20倍に引き上げた「MiMo-V2.6-Pro UltraSpeed」サービング階層も提供されている[1]。

本番導入におけるインフラ制約と技術的課題

MiMo-V2.6-Proの成果はオープンウェイト分野を前進させた一方で、システム設計や運用の観点からは、複数の実務的制約や批判的見解が存在する[5]。

ホスティング要件と計算リソースの制約

公称スペックではトークン活性化パラメータ数が42Bと軽量に見えるが、これは推論計算時のルーティングパスに過ぎない[9]。モデルの重みとKVキャッシュをメモリ上に保持するためには、1.02T全パラメータのロードが不可欠である[9]。

公式の推奨デプロイ構成は、Tensor Parallelism 16（TP16）およびExpert Parallelism 16（EP16）を組み合わせた2ノード以上のクラスタ構成を要求する[9]。企業内の単一ノードやオンプレミスのハイエンドワークステーションでの自前運用は極めて非現実的であり、事実上、Xiaomiの公式エンドポイントまたは大規模クラウドインフラへの依存を前提としている[5]。

API可用性と単一障害点（SPOF）

OrcaRouterの分析が指摘するように、リリース当初においてMiMo-V2.6-Proを提供するAPIプロバイダは実質的にXiaomi公式プラットフォーム等ごく一部に限られていた[5]。複数プロバイダ間の自動フェイルオーバーやロードバランシングを組むことができず、上流のレートリミット（HTTP 429）やエンドポイントの障害が発生した場合、依存するアプリケーション全体が停止する可用性リスクを抱えている[5]。

100万トークンコンテキストの実効推論品質

モデルの設定ファイル上、最大位置埋め込み（max_position_embeddings）は1,048,576トークン（1M）に拡張されている[7]。しかし、AlphaLabによる強化学習プロセスの解析によれば、RL訓練中に生成・検証された典型的な軌跡長は11万～15万（110K～150K）トークンの範囲に留まっていた[9]。

1Mトークン近傍におけるNeedle In A Haystack（NIAH）の長文情報検索精度や、長大な依存関係にわたる複雑な因果推論の信頼性については、公開ベンチマークで十分に検証されておらず、現状では「受け入れ可能なバッファ容量」と「高精度な長文推論品質」を同一視すべきではないと評されている[9]。

自己改善フレームワークの位置付けとベンチマークバイアス

Xiaomiの技術レポート表題にある「Toward Self-Improvement」という表現について、厳密な再帰的自己改善（Recursive Self-Improvement: RSI）との混同に対する指摘がなされている[6]。MiMo-V2.6のループは、ポリシーが軌跡を生成し、MiMo自身を含むグレーダーがフィードバックを与えてRLで重みを更新する枠組みである[2]。しかし、タスク環境の選定、テストハーネスの設計、検証ルールの策定、アドバンテージ再配分（GAR）のアルゴリズム構築、学習スケジュール管理といった「外側のループ（Outer Loop）」はすべて人間の研究チームが制御している[9]。AI自身が学習アルゴリズムを自律改変し次世代モデルをゼロから設計する完全なRSIではなく、人間の設計枠組みを忠実にスケールさせた「人間主導の自己改善支援ループ」と位置付けるのが正確である[9]。

さらに、モデル評価の主軸とされたDeepSWE v1.1について、Epoch AIの第三者監査により、全113問の評価タスクのうち少なくとも23問においてテストハーネス側の設計不備に起因する偽陰性エラーが含まれていることが判明した[9]。また、30ステップのRL訓練を通じてDeepSWEのスコア（Proにおいて58.41から72.57へ上昇）が継続監視されていたことから、同ベンチマークがブラインドな汎化性能テストではなく、訓練プロセスにおける内部最適化メトリクスとして機能してしまった過剰適合の懸念も残されている[6]。

MiMo-V2.6-Proは、1.02T規模のMoEアーキテクチャにルータ凍結とGAR/GRSを組み込むことで、強化学習における報酬ハッキングと探索の頭打ちを打破し、オープンウェイトとして最高水準の推論・エージェント実行性能を達成した[2]。商用フロンティアモデル比で20分の1から60分の1に設定された

API料金は、高頻度なツール呼び出しを伴う自律エージェントの運用経済性を劇的に改善する[1]。しかし、その恩恵を享受する裏で、実質的なマルチノード専用インフラ要件、単一プロバイダによる可用性リスク、および極長文領域における実効精度の見極めというエンジニアリング上の明確な課題が存在しており、採用にあたってはタスク特性に応じた慎重なリスク検証が求められる[5]。

引用文献

1. Xiaomi、AIモデル「MiMo-V2.6」Pro／Flashをオープンウェイトで, https://gihyo.jp/article/2026/09/xiaomi-mimo-v2.6
2. XiaomiMiMo/MiMo-V2.6-Pro-RL - Hugging Face, https://huggingface.co/XiaomiMiMo/MiMo-V2.6-Pro-RL
3. MiMo-V2.6-Pro Benchmarks & Hands-On Test vs Claude Opus 5, https://kingy.ai/blog/mimo-v2-6-pro-benchmarks-specs-comparison/
4. MiMo-V2.6 | Xiaomi, https://mimo.xiaomi.com/mimo-v2-6
5. Xiaomi MiMo-V2.6-Pro：Open-Weights Indexで46 - OrcaRouter, https://www.orcarouter.ai/ja/blog/xiaomi-mimo-v2-6-pro-release
6. MiMo-V2.6: Scaling Reinforcement Learning Towards Self, https://www.alphaxiv.org/abs/2609.mimo-scaling-reinforcement-learning
7. MiMo-V2.6-Pro - API Pricing & Benchmarks | OpenRouter, https://openrouter.ai/xiaomi/mimo-v2.6-pro
8. MiMo-V2.6: Scaling Up Reinforcement Learning for Self-Improvement, https://mimo.mi.com/docs/news/latest/v2-6
9. MiMo-V2.6 深度解析：大規模RL，真的更接近模型自我改進？（2026）, https://www.alphalab.site/mimo-v2-6-rl-self-improvement
10. MiMo-V2.6-Pro-UltraSpeed, https://mimo.mi.com/models/en-US/mimo-v2.6-pro-ultraspeed
11. MiMo-V2.6-Pro-UltraSpeed - API Pricing & Providers | OpenRouter, https://openrouter.ai/xiaomi/mimo-v2.6-pro-ultraspeed
12. XiaomiMiMo/MiMo-V2.6-Flash-RL - Hugging Face, https://huggingface.co/XiaomiMiMo/MiMo-V2.6-Flash-RL
13. XiaomiMiMo/MiMo-V2.6-Distill-Qwen-9B - Hugging Face, https://huggingface.co/XiaomiMiMo/MiMo-V2.6-Distill-Qwen-9B