

中国LLM四モデル比較レポート

更新基準日: 2026-04-26 JST

エグゼクティブサマリ

- 独立評価の総合感では、Kimi K2.6 と MiMo-V2.5-Pro がともに 54 点、DeepSeek-V4-Pro が 52 点、Qwen3.6-27B が 46 点で、今回の四者の中では **Kimi と MiMo が最上位、DeepSeek が僅差、Qwen は「小ささの割に非常に強い」実務型** という並びでした。ただし、**透明性は性能と一致せず**、最も情報開示が薄いのは MiMo-V2.5-Pro です。 ¹
- DeepSeek-V4 Preview は単一モデルではなく **Pro / Flash のシリーズ公開** です。比較の主対象にすべきは上位版の **DeepSeek-V4-Pro** で、1.6T 総パラメータ・49B active・1M context・CSA/HCA ベースの新注意機構を持ち、**長文効率で最も尖っている** のが特徴です。 ²
- Kimi K2.6 は **オープンウェイト陣営の中で、エージェント型コーディングと長時間タスク遂行が最も強い選択肢** です。ただし、K2.6 固有の学習データや安全対策の開示は薄く、公開情報の多くは K2 / K2.5 系列の系譜を参照しないと埋まりません。 ³
- Qwen3.6-27B は **自前運用・追加学習・配布互換性** の観点で今回もっとも扱いやすいモデルです。27B dense、Apache-2.0、公式の 8 GPU 起動例、豊富な推論フレームワーク対応が揃っており、“**性能の絶対値**”ではなく“**運用しやすい強モデル**”という位置づけが非常に明確です。 ⁴
- MiMo-V2.5-Pro は **API-first の長時間エージェント用途** ではかなり有望ですが、今回の四者で最も“**ブラックボックス寄り**”です。公式ページは実例・効率・ケーススタディに強い一方、**重み公開、ライセンス、詳細ベンチ、安全報告、学習データの実態** はまだ十分には見えません。 ⁵

比較表

オープンウェイトの三者は、主に Hugging Face ⁶、ModelScope ⁷、GitHub ⁸ を通じて配布・実装情報が公開されています。独立比較では Artificial Analysis ⁹ の指標も参考になります。 ¹⁰

モデル	概要	訓練データ	性能ベンチの要点	推論速度・推奨ハードウェア	調整方法	ライセンス・商用	安全既知
DeepSeek-V4 Preview	2026-04-24 公開。シリーズ構成は Pro / Flash。主比較対象の Pro は 1.6T total / 49B active / 1M context の MoE。CSA+HCA、mHC、Muon を採用。	Pro 33T / Flash 32T の事前学習トークン量は開示。だが言語分布、公開 vs 独自データ比率、収集ポリシーの詳細は未公開。	MMLU-Pro 87.5、GPQA 90.1、LiveCodeBench 93.5、SWE Verified 80.6、Terminal-Bench 67.9、SWE Multilingual 76.2、MRCR 1M 83.5。 長文系が非常に強い。	DeepSeek 公式 API で 36.9 t/s、TTFT 1.92s 。公式は Think Max で 少なくとも 384K context を推奨。最小 GPU 構成は未公開で、実質は大規模 multi-GPU / cluster 向け。	SFT + RL(GRPO) による領域別 expert 育成後、 on-policy distillation で統合。	MIT。重み公開あり。商用利用は MIT 条件準拠で可能。	有害策のは未公開ら tool が平 JSO する A 報告 力制限 挙する ニテがあ
Kimi K2.6	2026-04-20 公開。 1T total / 32B active / 256K context の MoE。MLA、MoonViT 400M、thinking/instant、画像・動画入力をサポート。	K2.6 固有の新規学習 トークン数は未公開。 系列背景として、K2 は 15.5T text 、K2.5 はその上に 約15T mixed visual+text を継続学習。K2.5 論文では web/code/math/knowledge と豊富な multimodal data が説明される。	HLE w/tools 54.0、BrowseComp 83.2、DeepSearchQA F1 92.5、Terminal-Bench 66.7、SWE Pro 58.6、SWE Verified 80.2、AIME 2026 96.4、GPQA 90.5。 agentic/coding が最強クラス。	公式プロバイダ Kimi で 106.3 t/s 、ただし reasoning workload では first answer token 43.12s と重い。推論エンジンは vLLM / SGLang / KTransformers 推奨。巨大モデルなので self-host は高級 GPU 群向け。	K2.6 自体の詳細 recipe は未公開。系列として zero-vision SFT + joint text/vision RL と大規模 agentic post-training が公開済み。	Modified MIT 。重み公開あり。商用可否は原則ライセンス本文準拠。	専用 safe system card でき ミュ では Anth と名 過剰 創造 下・ 加な 期報 る。

モデル	概要	訓練データ	性能ベンチの要点	推論速度・推奨ハードウェア	調整方法	ライセンス・商用	安全 既知
Qwen3.6-27B	2026-04-22 公開。27B dense、Vision Encoder 付き causal LM。64 層、Gated DeltaNet + Gated Attention、MTP。262K native / 1.01M まで拡張可能。	3.6-27B 固有の corpus 規模は未公開。ただし系列として Qwen3 は約36T / 119 言語、Qwen3.5 は 201 言語・方言対応と early-fusion multimodal 訓練を公表。	SWE Verified 77.2、SWE Pro 53.5、SWE Multilingual 71.3、Terminal-Bench 59.3、SkillsBench 48.2、GPQA 87.8、AIME26 94.1、MMMU-Pro 75.8、VideoMME 87.7。小型 dense として非常に強い。	Alibaba Cloud で 62.9 t/s、TTFT 3.88s。公式の SGLang / vLLM 起動例は TP=8 GPU。llama.cpp、MLX、Transformers も対応。	公開チェックポイントは pre-training + post-training 済み。追加学習は SFT / DPO / GRPO を Swift / UnSloth / Llama-Factory 等で実施可能。	Apache-2.0。重み公開あり。商用利用しやすい。	安全 詳細 team system card 得。 認し 公開 は、 運用 比べ safe disc が薄

MiMo-V2.5-Pro	2026-04-22 に public beta。長時間 agentic / complex SWE / long-horizon coherence を売りにする。V2.5-Pro 自体の正確なパラメータ数は未公開。前世代 V2-Pro は 1T 超 / 42B active / 1M context / hybrid attention 7:1 + MTP。	規模・言語分布・公開/独自比率・収集方針は未公開。V2-Pro では agent scaffold 上で SFT + RL を明示。V2.5 系では post-training pipeline 最適化が語られるが、Pro の具体は非公開。	公式公開値は主に ClawEval 64% Pass^3 / 約70K tokens per trajectory、SysY compiler hidden tests 233/233、長時間自律コーディング事例など。独立評価では Intelligence Index 54。公的な math / multilingual / Japanese 指標は乏しい。	Xiaomi API で 62.0 t/s、TTFT 2.68s。現時点では API / AI Studio 中心で、self-host 用ハードウェア要件は未公開。価格も「V2-Pro から据え置き」で、V2-Pro は 256K まで \$1/\$3、256K-1M で \$2/\$6。	V2-Pro 系列では SFT + RL、V2.5 系では post-training 強化が示唆されるが、Pro 版の詳細は未公開。	重み未公開。ライセンス未公開。商用利用条件は API 規約依存で、公開資料だけでは断定不可。	safe system card 認。 リス 能よ 明性 ベー prop なの 達・ 再現 明負 い。
---------------	--	--	--	--	---	--	--

DeepSeek-V4 Preview

DeepSeek-V4 Preview は 2026 年 4 月 24 日に公開されたシリーズで、実体は DeepSeek-V4-Pro と DeepSeek-V4-Flash の 2 モデルです。比較の中心となる Pro は 1.6T total / 49B active / 1M context の巨大

MoE で、注意機構には **Compressed Sparse Attention** と **Heavily Compressed Attention** のハイブリッド、残差経路には **mHC**、最適化には **Muon** を導入しています。技術報告書では、1M 文脈で DeepSeek-V3.2 比 **single-token FLOPs 27%、KV cache 10%** まで削減したことを強調しており、ここが V4 の最大の差別化点です。 ²

訓練データについては、V4 系全体で **32T 超** の “diverse and high-quality tokens” を使ったと公式に述べられており、技術報告書の本文では **Flash 32T、Pro 33T** と書かれています。ただし、**言語分布、公開データと独自データの比率、クローリング方針、権利処理方針** は今回の公開資料からは確認できません。ポストトレーニングは、領域特化 expert を **SFT** と **RL (GRPO)** で個別に鍛え、その後 **on-policy distillation** で単一モデルへ統合する二段構えです。つまり、量は比較的多く開示されていますが、**データの中身の透明性はまだ足りない** と評価すべきです。 ¹⁵

性能は非常に強いです。公開モデルカードの Pro-Max 比較では **MMLU-Pro 87.5、GPQA-Diamond 90.1、LiveCodeBench 93.5、Codeforces 3206、HMMT 2026 Feb 95.2、IMOAnswerBench 89.8** と、推論・数学・コードの広い範囲でトップ級です。エージェント系でも **Terminal-Bench 2.0 67.9、SWE Verified 80.6、SWE Pro 55.4、SWE Multilingual 76.2、BrowseComp 83.4** を記録しています。日本語専用ベンチは公開されていませんが、**MMMLU 90.3、MGSM 84.4、MultiLoKo 51.1** といった多言語 proxy は悪くありません。ただし、**MMLU-ja** や **Rakuda** 系の日本語専用スコアは**未公開** です。 ¹⁶

運用面では、公式 API は **OpenAI / Anthropic 互換** で、公式のニュースページは OpenClaw、OpenCode、Claude Code との統合を前面に出しています。独立の API ベンチでは、DeepSeek 公式プロバイダで **36.9 t/s、TTFT 1.92s** が観測されています。とはいえ、このモデルは **1.6T 級の open weights** なので、self-host の実像は “ローカル PC で試せる” ではなく **企業用 multi-GPU サーバ / cluster 向け** です。公式はローカル実行用の `inference` フォルダを置いていますが、**最小 GPU 台数や最低 VRAM** は明記していません。さらに、チャットテンプレートは Jinja を同梱せず、独自の `encoding` フォルダ方式なので、エコシステム統合では一手間あります。 ¹⁷

実務上の長所は、**1M context、長文推論効率、コード・数学の高さ、MIT での重み公開** です。長文要約、長大コードベース解析、RAG を超える long-context 研究、社内オンプレの高性能基盤には非常に向きます。短所は、**安全性開示の薄さ** と **運用資源の重さ** です。公開直後の GitHub ⁸ issue では、tool calling 時に `tool_calls` フィールドではなく本文に平文 JSON が出る不具合が報告されており、さらに一部ユーザーは出力制御や政治的話題での挙動の揺れも報告しています。現時点では、「**能力は高いが、API と安全まわりは早すぎる公開の粗さが残る**」という評価が妥当です。 ¹⁸

Kimi K2.6

Kimi K2.6 は 2026 年 4 月 20 日に公開された **native multimodal agentic model** で、公式モデルカードでは **1T total / 32B active / 256K context、61 層、384 experts、token ごとに 8 experts + 1 shared expert、MLA、MoonViT 400M** と整理されています。製品面では、**long-horizon coding、coding-driven design、agent swarm、24/7 background agents** が前面に出ており、典型的な “会話モデル” ではなく、**自律実行する開発エージェントの中核モデル** として設計されていることがはっきりしています。 ¹⁹

訓練データの透明性は、今回の四者の中では **中程度** です。K2.6 のモデルカード自身は新しい学習トークン数を開示せず、むしろ「**K2.6 は K2.5 と同じアーキテクチャ**」と述べ、デプロイ方法も K2.5 を再利用できるとしています。そこで系列文書を見ると、K2 base は **15.5T の text token** で事前学習され、K2.5 はその上に **約 15T の mixed visual + text token** を継続学習しています。さらに K2.5 論文では、text corpus が **web text / code / mathematics / knowledge** の 4 領域からなり、multimodal corpus には **caption / interleaving / OCR / knowledge / perception / video / agent data** が含まれ、**multilingual OCR** や image-code paired data も使うと説明されています。ポストトレーニング手法としては **zero-vision SFT** と

joint text-vision RL が公開されています。したがって、K2.6 は **非常に強い基盤と豊富な系列情報** を持つ一方、**2.6 固有に増えたデータや RL の詳細は未公開** です。 ²⁰

ベンチマークは、今回の四者の中で最も “agentic coding に振り切れている” と言えます。モデルカードの公開比較では **HLE-Full w/ tools 54.0、BrowseComp 83.2、BrowseComp (Agent Swarm) 86.3、DeepSearchQA F1 92.5、Claw Eval pass^3 62.3、OSWorld-Verified 73.1** と、ツール利用・検索・コンピュータ操作系で極めて強い数値を出しています。コード面でも **Terminal-Bench 2.0 66.7、SWE-Bench Pro 58.6、SWE-Bench Multilingual 76.7、SWE-Bench Verified 80.2、LiveCodeBench v6 89.6** です。数学も **AIME 2026 96.4、GPQA-Diamond 90.5、HMMT 2026 92.7** と高水準です。一方、**日本語専用ベンチは未公開** で、多言語性を直接保証するのは主に SWE-Bench Multilingual と、K2.5 論文で明示された multilingual OCR データの存在です。 ²¹

推論速度は、モデルそのものより **プロバイダ差** が大きいです。独立ベンチでは 9 つの API プロバイダが追跡されており、**Clarifai 162.0 t/s、Kimi 106.3 t/s、Together FP4 102.2 t/s** と高速な一方、reasoning モデルゆえに **first answer token は数十秒級** になり得ます。公式は vLLM / SGLang / KTransformers を推奨し、K2.6 は **native INT4 quantization** を採用します。とはいえ、**1T 級 open weights** なので、dense 27B の Qwen のような “気軽な自前運用” ではありません。公式 API は OpenAI/Anthropic 互換で、**thinking / instant、image / video、preserve_thinking、multi-step tool call** をサポートし、Kimi Code CLI と強く結びついています。ライセンスは **Modified MIT** です。 ²²

実用上の長所は、**エージェント型コーディング、長時間自律実行、UI/フロント生成、multimodal coding** での圧倒的な強さです。反対に短所は、**巨大すぎる self-host 難易度** と **K2.6 固有の透明性不足** です。コミュニティの初期報告では、Hugging Face discussion に **韓国語で Anthropic を名乗る** という指摘があり、また kimi-cli の issue では **思考が重く、創造性や安定性が下がった** というユーザー報告もあります。これらは公式再現結果ではなく **アネクドット** ですが、公開直後の品質リスクとしては無視できません。総じて、**外部 API を許容できるなら、今回の四者の中で agentic coding の最有力候補** です。 ²³

Qwen3.6-27B

Qwen3.6-27B は 2026 年 4 月 22 日に公開された **Qwen3.6 系の最初の open-weight variant** で、今回の比較対象では唯一の **27B dense** モデルです。モデルカードでは、**Causal LM with Vision Encoder、64 layers、Gated DeltaNet と Gated Attention を組み合わせた 16×ブロック構造、MTP、262,144 native context、最大 1,010,000 token まで拡張** とされています。重要なのは、このモデルが “小さい” のに “弱くない” ことです。巨大 MoE を使わずとも、agentic coding でかなり高いところまで来ています。 ²⁴

訓練データの開示は **系列レベルでは比較的厚いが、3.6-27B 個別では薄い** という状態です。公開モデルカードと GitHub README は、Qwen3.6 が **Qwen3.5 の基礎的ブレイクスルーの上で、安定性と real-world utility を優先した更新** と説明する一方、**3.6-27B 専用の新規 pretraining corpus 規模** は出していません。系列の背景としては、Qwen3.5 が **201 言語・方言対応、early-fusion 型の multimodal foundation**、強化学習の大規模化を公表しており、さらに Qwen3 は **約 36T token・119 言語** を公式ブログで公表しています。つまり、**多言語基盤は最も説明がある** 一方で、**Qwen3.6-27B そのものの corpus の追加量やデータソース比率は未公開** です。 ²⁵

ベンチマークでは、サイズを考えると非常に優秀です。言語・コードでは **SWE-bench Verified 77.2、SWE-bench Pro 53.5、SWE-bench Multilingual 71.3、Terminal-Bench 2.0 59.3、SkillsBench Avg5 48.2、NL2Repo 36.2、Claw-Eval Pass^3 60.6**。知識・推論では **MMLU-Pro 86.2、GPQA Diamond 87.8、AIME26 94.1、LiveCodeBench v6 83.9**。視覚・動画でも **MMMU-Pro 75.8、CharXiv RQ 78.4、VideoMME 87.7、AndroidWorld 70.3** を出しています。今回の四者で唯一、“日本語に直接効きそうな公開材料” が比較的多いのも Qwen です。理由は、日本語専用スコアがあるからではなく、**201 言語・方言対応**

を明言していること、そして dense 27B で fine-tuning しやすいからです。逆に言うと、**日本語専用 benchmark はやはり未公開**です。²⁶

運用面は今回最も実務的です。公式の SGLang / vLLM 例は **TP=8 GPU** で 262K context を想定しており、さらに **text-only モード** で vision encoder を外してメモリを節約できます。加えて、**llama.cpp**、**MLX**、**Transformers**、**Qwen-Agent**、**Qwen Code** に対応しており、Alibaba Cloud ²⁷ の Model Studio API も OpenAI/Anthropic 互換です。独立 API ベンチでは、Alibaba Cloud で **62.9 t/s**、**TTFT 3.88s** が観測されています。ライセンスは **Apache-2.0** で、公式 README は **SFT / DPO / GRPO** による追加学習のツール群も案内しています。⁴

実務上の長所は、**性能・サイズ・互換性・ライセンスのバランス**です。オンプレ日本語アシスタント、社内コパイロット、コード支援、追加学習ベースの業務特化、Apple Silicon やローカル推論の実験まで、用途が広いです。短所は、**生の agentic ceiling は Kimi や DeepSeek V4 Pro に一步譲る**こと、そして**安全対策や 3.6 個別データの説明が十分ではない**ことです。それでも、「**最終性能を 5 点落としても、運用コストと制御性を取りたい**」なら今回の最有力です。²⁸

MiMo-V2.5-Pro

MiMo-V2.5-Pro は 2026 年 4 月 22 日に public beta 入りした Xiaomi の最上位モデルで、リリースページは **general agentic**、**complex software engineering**、**long-horizon tasks** の改善を強く打ち出しています。特徴は、**1,000 回超の tool calls をまたぐ長時間タスク**を維持できること、文脈中に埋まった subtle requirements に従いやすいこと、そして AI Studio / API 上に即時展開されたことです。ただし、重要な点として **V2.5-Pro そのもののパラメータ数は公式ページで再掲されていません**。参考として前世代 V2-Pro は **1T 超 / 42B active / 1M context / hybrid attention ratio 7:1 / MTP** を明示していました。よって、V2.5-Pro のサイズ欄は厳密には **未公開**と書くべきです。²⁹

訓練データの透明性は、四者中で最も低いです。V2.5-Pro 公開ページは、データ量・言語分布・公開データと独自データの比率を出していません。系列文書を見ると、V2-Pro では **complex, diverse agent scaffolds** 上で **SFT** と **RL** を行ったことが書かれており、MiMo-V2.5 (無印) は **perception / reasoning / tool use** を **joint** に整列する **optimized post-training pipeline** を説明しています。しかし、**V2.5-Pro 固有の pretraining corpus** や **post-training compute** の詳細は**公開されていません**。したがって、性能は見えても、**その性能が何に依存しているかは見えにくい**モデルです。³⁰

ベンチマーク開示の性格も独特です。公式は、一般的な公開ベンチの長い表よりも、**ケーススタディと token efficiency** を前面に出しています。代表例は、Peking University ³¹ の課題由来の **SysY compiler in Rust** を **4.3 時間・672 tool calls・hidden tests 233/233** で通した事例、**8,192 行の video editor** を **1,868 tool calls / 11.5 時間** で構築した事例、ngspice を回しながら analog EDA 最適化を行った事例です。定量では **ClawEval 64% Pass^3 / 約 70K tokens per trajectory** を出し、同等能力帯の Claude Opus 4.6、Gemini 3.1 Pro、GPT-5.4 より **40-60% 少ない token** で到達すると主張します。独立の Artificial Analysis ⁹ でも Intelligence Index **54** と強い数字ですが、**公開された標準 math / multilingual / Japanese** ベンチの厚みは他の三者に劣ります。³²

運用面では、現時点では **API-first** と考えるべきです。公式ページは、V2.5-Pro が API Platform / AI Studio にロールアウト済みで、価格は **据え置き**と記します。前世代 V2-Pro の API 価格表では **256K まで input \$1 / output \$3**、**256K-1M で input \$2 / output \$6** とされており、V2.5-Pro も少なくとも当面はこの枠組みを継承すると読むのが自然です。独立計測では Xiaomi API が **62.0 t/s**、**TTFT 2.68s**。エコシステム面では Claude Code、OpenCode、Kilo などの scaffold との統合を歓迎しており、“**Coming Open Source**”と予告されているものの、**重み・ライセンス・self-host 推奨 HW はまだ出ていません**。³³

実務上の長所は、**長時間エージェント性能と token 効率** です。API で使う前提なら、かなり魅力的です。一方で短所は明確で、**監査性・再現性・法務確認・安全性確認に必要な情報が最も少ない**。公開資料からは dedicated safety/system card を確認できず、ベータ段階でもあります。したがって、“**いま一番強そうな proprietary Chinese agent API を短納期で試したい**” 場面には向くが、**調達審査や再現性説明が厳しい大企業導入にはまだ早い**、というのが公平な結論です。 ³⁴

結論と推奨

結論を一行で言うと、**能力の絶対値で選ぶなら Kimi K2.6 と DeepSeek-V4-Pro、運用現実性で選ぶなら Qwen3.6-27B、API-only の長時間エージェント賭け枠が MiMo-V2.5-Pro** です。Kimi K2.6 は agentic coding と multimodal agent の総合力で最も魅力があり、DeepSeek-V4-Pro は 1M context と長文効率で別格です。Qwen3.6-27B は dense 27B / Apache-2.0 / 8GPU 公式構成という“現場で回る強モデル”で、MiMo-V2.5-Pro は実力が高そうだが documentation debt が大きい、という整理になります。 ³⁵

日本語用途に絞って推奨を出すなら、**最も無難なのは Qwen3.6-27B** です。理由は、日本語専用スコアがあるからではなく、**公開資料の多言語説明が最も厚く、追加学習しやすく、ライセンスが扱いやすく、ハードウェア要件も trillion 級より現実的** だからです。反対に、**日本語での raw capability の天井を追求** なら Kimi K2.6、**長大文書や長大コードベースを 1M context で処理したい** なら DeepSeek-V4-Pro が有力です。MiMo-V2.5-Pro は、日本語性能を含めて **実力評価を自社ベンチで必ず補うべき候補** です。四者とも **日本語専用 benchmark を十分に公開していない** ため、本番採用前には社内の Japanese eval suite を必須にすべきです。 ³⁶

最終的な推奨は次の通りです。**外部 API を許容し、最先端の agentic coding を最優先するなら Kimi K2.6。長文効率と open-weight 研究基盤を重視するなら DeepSeek-V4-Pro。自前運用・微調整・多言語業務導入まで見据えるなら Qwen3.6-27B。API-only で高難度長時間エージェントを先行検証するなら MiMo-V2.5-Pro** です。能力競争そのものは確実に前進していますが、**透明性と安全性の説明責任はモデルごとに大きく差があり、そこでは Qwen、次いで DeepSeek、Kimi、MiMo の順で安心感がある** と私は評価します。 ³⁷

¹ ⁶ ³⁵ <https://artificialanalysis.ai/models/kimi-k2-6>

<https://artificialanalysis.ai/models/kimi-k2-6>

² ⁹ ¹¹ ¹⁷ <https://api-docs.deepseek.com/news/news260424>

<https://api-docs.deepseek.com/news/news260424>

³ ²⁰ ²¹ ²² <https://huggingface.co/moonshotai/Kimi-K2.6>

<https://huggingface.co/moonshotai/Kimi-K2.6>

⁴ ²⁴ ²⁵ ²⁶ <https://huggingface.co/Qwen/Qwen3.6-27B>

<https://huggingface.co/Qwen/Qwen3.6-27B>

⁵ ⁸ ¹⁴ ²⁹ ³¹ ³² ³³ ³⁴ <https://mimo.xiaomi.com/mimo-v2-5-pro/>

<https://mimo.xiaomi.com/mimo-v2-5-pro/>

⁷ ²³ ²⁷ <https://huggingface.co/moonshotai>

<https://huggingface.co/moonshotai>

¹⁰ ¹⁵ ¹⁶ <https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro>

<https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro>

¹² ¹⁹ <https://www.kimi.com/blog/>

<https://www.kimi.com/blog/>

13 28 37 <https://github.com/QwenLM/Qwen3.6>

<https://github.com/QwenLM/Qwen3.6>

18 <https://github.com/deepseek-ai/DeepSeek-V3/issues/1244>

<https://github.com/deepseek-ai/DeepSeek-V3/issues/1244>

30 <https://mimo.xiaomi.com/mimo-v2-pro>

<https://mimo.xiaomi.com/mimo-v2-pro>

36 <https://huggingface.co/Qwen/Qwen3.5-27B>

<https://huggingface.co/Qwen/Qwen3.5-27B>