米中フロンティア AI モデル比較（2026 年 8 月 15 日時点）

― 「中国は米国に半年遅れ」という言説は払拭されたのか ―

調査基準日：2026 年 8 月 15 日／作成：2026 年 8 月 16 日

Claude Opus 5

本文中の上付き番号は文末「参考文献」の番号に対応する。

---

1. 要旨

1.1 結論

ギャップの有無は「どの領域を見るか」で結論が逆転する。単純に払拭された／されていないと述べることはできない。

- クローズドの最高性能：米国が依然リード。ただし差は数点まで縮小した。Artificial Analysis 知能指数（v4.1.1）で Claude Opus 5 が 63.1 で首位、中国最上位の Kimi K3 は 60、DeepSeek-V4-Pro-0813 は 53 である [1,2,3]。
- オープンウェイト：中国が事実上の覇権を握った。上位のオープンモデルは Kimi K3・Qwen3.8-Max・DeepSeek-V4 系・GLM-5.3 が占め、米国側の対抗は Meta 系のみである [19,20,39]。
- コスト効率：中国が明確に優位。同等性能帯で数分の一から十数分の一の価格差がある [26,30,31]。
- 開発者採用：中国製が米国内でも過半を占める局面が生じている [36,37,38]。

したがって「半年遅れ」は、フロンティア最高性能という一点においてのみ部分的に妥当性を残し、それ以外の指標ではすでに実態を説明できていない、というのが本調査の評価である（分析）。

1.2 事実確認の結果（最優先事項）

ご提示いただいたモデル名・提供元・リリース日を一次情報および複数の報道で照合した結果、開発元表記に重要な誤りが 1 点、注意を要する点が 1 点確認された。

表 1　事実確認の結果

提示されたモデル	実際の開発元	リリース日	判定
GPT-5.6 Sol	OpenAI	2026 年 7 月 9 日 GA	実在を確認（提示に日付記載なし）
Claude Opus 5	Anthropic	2026 年 7 月 24 日	実在・提示情報は正確 [13]
Claude Fable 5	Anthropic	2026 年 6 月 9 日	実在・提示情報は正確 [14]
Gemini 3.7 Flash	Google	2026 年 8 月 13 日	実在・日付とも正確 [15,16]

提示されたモデル	実際の開発元	リリース日	判定
		（米国時間）	
Grok 4.6	SpaceXAI（旧 xAI）	2026 年 8 月 12 日	要注意：法人格の表記（下記参照）[6,7,17]
Muse Spark 1.2	Meta（Meta Superintelligence Labs）	2026 年 8 月 5 日	誤り：xAI 系ではなく Meta 製 [4,5]
Kimi K3	Moonshot AI	発表 7 月 16 日／重み 7 月 27 日	実在・提示情報は正確 [19,20]
Qwen3.8 Max	Alibaba	2026 年 8 月 3 日 GA	実在・日付とも正確 [23,24]
DeepSeek-V4-Pro-0813	DeepSeek	2026 年 8 月 12〜13 日 GA	実在・日付とも正確 [26]
GLM-5.3	Z.ai（智譜 AI）	2026 年 8 月 14 日	実在・日付とも正確 [27,28]

訂正が必要な 2 点

■ Muse Spark 1.2 は Meta 製である。 xAI 系の製品ではない。2026 年 8 月 5 日にコーディングエージェント「Muse Code」とともに公開され、8 月 10 日に Mark Zuckerberg 氏がオープンウェイト化を発表した。同日、24GB GPU で動作する軽量オープンモデル「Muse Glimmer」（300 億パラメータ）も公開されている [4,5]。

■ Grok 4.6 の開発主体は現在「SpaceXAI」である。 xAI は 2026 年 2 月に SpaceX に買収され、現行の公式表記は SpaceXAI（公式 X アカウントは@SpaceXAI）となっている。したがって「xAI が発表」は経緯としては誤りではないが、法人格としては SpaceXAI 表記が正確である [6,7]。

※本レポートの性能数値の多くはベンダー自己申告値である。第三者による独立検証が追いついていない指標が多い点に留意されたい（後述「留保事項」参照）。

2. 各モデルの基本情報

表 2　米国系フロンティアモデル

モデル	開発元	公開日	公開形態	主要諸元	API 価格（100 万トークン）
GPT-5.6 Sol	OpenAI	2026/7/9 GA	クローズド	推論特化。文脈約 105 万トークン、最大出力 12.8 万。Sol/Terra/Luna の 3 層構成	入力$5／出力$30
Claude Opus 5	Anthropic	2026/7/24	クローズド	effort ダイヤル（低〜高＋max）、Fast mode（約 2.5 倍速）	入力$5／出力$25[13]
Claude Fable 5	Anthropic	2026/6/9	クローズド	Mythos 級の一般公開版。	入力$10／出力$50[14]

モデル	開発元	公開日	公開形態	主要諸元	API 価格（100 万トークン）
			（セーフガード付）	6/12 に輸出規制で一時停止、7/1 復帰	
Gemini 3.7 Flash	Google	2026/8/13	クローズド	ネイティブマルチモーダル推論。コーディング・エージェント特化	入力$0.75／出力$3.75（2026 年末までの導入価格。2027 年 1 月から $1.50／$7.50）[15,16]
Grok 4.6	SpaceXAI	2026/8/12	クローズド	長期エージェント特化。文脈 50 万トークン、知識カットオフ 2026/2/1。パラメータ数非公表	入力$2／出力$6（fast 版は 2 倍）[7,17]
Muse Spark 1.2	Meta	2026/8/5（8/10 にオープン化）	オープンウェイト	コーディング特化。文脈 100 万トークン。テキスト・画像・音声・動画入力	入力$1.25／出力 $4.25[4,18]

表 3　中国系フロンティアモデル

モデル	開発元	公開日	公開形態	主要諸元	API 価格（100 万トークン）
Kimi K3	Moonshot AI	発表 2026/7/16、重み 7/27	オープンウェイト	2.8 兆パラメータ MoE（896 エキスパート中 16 活性化）、文脈 100 万トークン。Kimi Delta Attention	入力$3／出力$15[19,20]
Qwen3.8-Max	Alibaba	2026/8/3 GA	オープンウェイト化方針	2.4 兆パラメータ疎 MoE（95B 活性化）、ハイブリッドアテンション、文脈 100 万トークン、マルチモーダル	入力$2／出力$6（キャッシュ入力$0.25）[23,24]
DeepSeek-V4-Pro-0813	DeepSeek	2026/8/12〜13 GA（プレビュー4/24）	MIT 系オープンウェイト	1.6 兆パラメータ MoE（49B 活性化）、文脈 100 万トークン、最大出力 38.4 万。CSA+HCA 等の新アテンション	GPT-5.6 Sol 比で大幅に安価 [26,31]
GLM-5.3	Z.ai（智譜 AI）	2026/8/14	オープンウェイト（重み公開は約 2 週間後の予定）	ベースは GLM-5.2 のまま事後学習のみで性能 50%向上と主張。サイバーセキュリティ特化	当初は GLM コーディングプラン限定 [27,28]

※GLM-5.3・Qwen3.8-Max・DeepSeek-V4-Pro-0813 GA 版の重み公開状況は流動的であり、本稿執筆時点で一部未公開である。オープンウェイトとしての実効性は、ライセンス確定と重み公開後に再評価が必要である（留保）。

3. ベンチマーク横断比較

3.1 Artificial Analysis 知能指数（第三者測定）

ベンダー自己申告を離れた第三者合成指標として、Artificial Analysis 知能指数（v4.1.1）が最も横断比較に適する。

表 4　知能指数ランキング（2026 年 8 月時点）

モデル	開発元	国	知能指数	公開形態
Claude Opus 5 (max)	Anthropic	米	63.1	クローズド
Claude Fable 5	Anthropic	米	62.1	クローズド
GPT-5.6 Sol (max)	OpenAI	米	61	クローズド
Grok 4.6 (High)	SpaceXAI	米	60.9	クローズド
Kimi K3 (max)	Moonshot AI	中	60	オープンウェイト
Qwen3.8-Max	Alibaba	中	58	オープン化方針
Muse Spark 1.2 (xhigh)	Meta	米	57	オープンウェイト
DeepSeek-V4-Pro-0813 (max)	DeepSeek	中	53	オープンウェイト
GLM-5.3	Z.ai	中	未登録（前身 GLM-5.2 は 51〜53）	オープン化予定

※Kimi K3 は発表時点（2026 年 7 月 16 日）で AA 指数第 3 位だったが、7 月 24 日の Opus 5 登場後は概ね第 6 位程度に後退した。ただしオープンウェイト系では依然首位である[21]。Qwen3.8-Max のスコアはエンドポイント再測定により 53→56→58 と変動した経緯がある[25]。また effort（推論努力）設定により数点変動するため、本表は同一条件比較ではない（留保）。

3.2 コーディング（SWE-bench Verified）

最上位帯は Claude Opus 5 が 96〜97%、GPT-5.6 Sol が 96.2%（Vals AI による独立測定[44]）、Claude Fable 5 が 95%、Kimi K3 が 93.4%である。オープンウェイト系の首位は DeepSeek-V4-Pro-Max の 80.6%で、Qwen3.7 Max が 80.4%、Kimi K2.6 が 80.2%と続く[29,30,33]。

ただし SWE-bench Verified の数値は、各社が独自のスキャフォールド（実行環境・エージェント構成）で測定しているため、絶対値の単純比較には慎重を要する。集計サイトの llm-stats は独立検証値がゼロである旨を注記しており[43]、OpenAI は 2026 年 2 月に汚染を理由に同ベンチから撤退したとの報道もある。汚染耐性の高い SWE-bench Pro の方が比較指標としては信頼性が高い[29]（分析）。

3.3 エージェント／長期タスク

DeepSWE v1.1（長期エージェント型ソフトウェア工学）のスコアは以下のとおりである[32]。

表 5　DeepSWE v1.1 スコア

順位	モデル	スコア
1	Claude Opus 5（米）	74.0%

順位	モデル	スコア
2	GPT-5.6 Sol（米）	72.7%
3	Claude Fable 5（米）	69.7%
4	Grok 4.6（米）	67.0%
5	Gemini 3.7 Flash（米）	65.0%
6	DeepSeek-V4-Pro-0813（中）	63.0%
7	Qwen3.8-Max（中）	57.0%
8	Muse Spark 1.2（米）	55.0%

実務 44 職種を対象とした GDPval-AA v2 では、Claude Opus 5 が 1858 で首位、Kimi K3 が 1687 で第 3 位に入っている[21]。一方、OSWorld-Verified については Qwen3.8-Max が 86.1 を主張し、GPT-5.6 Sol Max（83.2）を上回ると自己申告しているが、これは独立検証されていない[24]（留保）。

3.4 数学・推論

GPQA Diamond では DeepSeek-V4-Pro が 90.1%、Kimi K3 が 93.5%を自己申告している[26,33]。数学領域は、米政府機関 CAISI ですら「DeepSeek V4 は米最上位にほぼ肉薄している」と認めた唯一の領域である[9,10]。

4. 「半年遅れ」言説の検証

4.1 言説の出所と各機関の評価

「中国は米国に半年遅れ」という言説には複数の出所があり、機関ごとに結論が大きく異なる。

表 6　米中ギャップ評価の比較

評価主体	ギャップ評価	根拠と特徴
Epoch AI	平均 7 か月（幅 4〜14 か月）	Epoch Capabilities Index（ECI）による。2023 年以降フロンティアは全て米国製。中国が初めて GPT-4 を超えたのは 2024 年 5 月（14 か月差）、o3（2025 年 4 月）は未超越[8]
NIST／CAISI	約 8 か月・差は拡大傾向	非公開ベンチによる評価。DeepSeek 自己申告（Opus 4.6／GPT-5.4 相当）に対し「GPT-5 相当＝8 か月前」と判定。ソフトウェア工学・サイバー・抽象推論で差が大きい[9,10]
Stanford HAI	2.7%・実質収束	LMArena Elo による（2026 年 3 月、Claude Opus 4.6=1503 対 ByteDance Dola-Seed-2.0-Preview=1464）。2023 年 5 月の 17.5〜31.6 ポイント差から急縮小[11]
AEI	数週間差	Kimi K3 を受けて、従来の 6〜8 か月説は数週間差にまで縮まったと主張[42]
Rice 大 MCNAIR	真の差は 3〜9 か月	CAISI は回帰直線で「差拡大」、Epoch はステップ関数で「横ばい」と読むため結論が逆転する。両者とも過信であると整理[12]

評価が割れる理由は、測定手法（合成指標か人間選好か、公開ベンチか非公開ベンチか）と、時系列の当てはめ方（回帰かステップ関数か）の違いにある。MCNAIR の「読み方次第で結論が逆転する」という整理が、現状として最も誠実な認識であると考えられる [12]（分析）。

※CAISI の評価は米国 AI Action Plan に由来する政府評価であり、政治的文脈を伴う点は割り引いて読む必要がある（留保）。

4.2 オープンウェイトとクローズドの区別

議論の混乱の主因は、この 2 領域を区別せずに論じている点にある（分析）。

- クローズド最高性能：米国が数点リード（Opus 5=63.1 に対し Kimi K3=60）。この差は縮小したが消えてはいない [1,2]。
- オープンウェイト：中国が事実上の覇権を握っている。Kimi K3 は世界最大級のオープンモデルであり、GLM-5.3・Qwen3.8・DeepSeek-V4 がオープン系上位を占有する。米国側の対抗は Meta のオープン系（Muse Spark 1.2、Muse Glimmer）のみである [19,20,39]。

Epoch AI は、2026 年 1 月以降、最も高性能なオープンウェイトモデルはフロンティアのクローズドモデルに平均 4 か月遅れ、ECI 差は平均 8 ポイントであると分析している。同時に、中国の主要モデルはほぼ全てオープンウェイトである一方、米国のフロンティアモデルはクローズドのままという非対称性を指摘している [8]。

4.3 コスト効率

この領域では中国が明確に優位である。DeepSeek V4 系は GPT-5.6 Sol の数分の一の価格帯にあり、Flash 系は入力$0.14／出力$0.28 という水準にある。CAISI ですら、DeepSeek V4 は同等性能帯の米国モデルより安価であり、7 ベンチマーク中 5 つで価格優位にあることを認めている [9,10,26,31]。

5. 構造的要因

5.1 投資額と計算資源

Stanford HAI の AI Index Report 2026 によれば、2025 年の民間 AI 投資額は米国 2859 億ドルに対し中国 124 億ドルで、23 倍超の開きがある [11]。この資金力の差は Nvidia／CUDA エコシステムへのアクセスと結びつき、米国側の構造的優位を形成している。

中国側は Huawei Ascend 910C（SMIC 7nm、Nvidia B200 比で約 3 分の 1 の BF16 スループット）による国産化を進めており、Bloomberg 報道によれば 2026 年に約 60 万個（前年比約 2 倍）、Ascend 系全体で最大 160 万ダイの流通が計画されている [34,43]。DeepSeek-V4-Pro の事後学習を Ascend 910C で

完了したとの報道もあるが、学習（プレトレーニング）は依然 Nvidia に依存しており、DeepSeek R2 は Ascend での学習に問題が生じ H800 に戻したと報じられている[34]（未確認情報を含む）。

なお、Epoch AI は 2025 年までに 29 万〜160 万 H100 相当の GPU が中国に密輸されたと推計している[8]。この推計値の幅の大きさ自体が、計算資源の実態把握の困難さを示している（留保）。

5.2 蒸留を巡る係争

Anthropic は公式に、DeepSeek・Moonshot・MiniMax が約 24,000 の不正アカウントを通じて Claude と 1600 万件超のやり取りを行い、うち Moonshot は約 340 万件（エージェント推論・コンピュータ操作が中心）を蒸留に使用したと指摘している[35]。Kimi K3 はこの指摘の対象であり、かつ米国最上位モデルと数点差の位置にある。米財務長官 Bessent 氏は中国 AI ラボへの制裁を示唆したとされる。

> ※本件は係争中であり、事実認定は確定していない。中国側の反論も含め、現時点で断定的な評価は避けるべきである（留保）。

5.3 戦略の非対称性

中国は低価格とオープンウェイトによって開発者エコシステムを獲得する戦略を、米国はクローズドの最高性能によって高収益を確保する戦略をそれぞれ採っている。両者は同じ土俵で競争しているわけではなく、「どちらが先行しているか」という問い自体が単一の答えを持たない構造になっている[28,39]（分析）。

6. 実運用・商用採用の状況

6.1 開発者トラフィック

OpenRouter における中国製オープンモデルのトークンシェアは、2024 年末には 1.2%未満だったが急増している。Bloomberg を引用した AI Weekly の報道によれば、米国企業が中国モデルを稼働させる比率は過去最高の 58%に達し、7 月第 1 週には一時 63%に触れたとされる[36]。モデル別では DeepSeek 単独で 17.6%、Qwen 系が 13.9%を占める[38]。米国製のシェアは 2025 年 6 月の約 70%から 2026 年 6 月には約 30%まで低下したとの集計もある[37,38]。

6.2 オープンモデルのダウンロード

Hugging Face のレポート（2026 年 8 月 14 日公開、Business Standard 報道経由）によれば、Qwen 系は 460 超のモデルで累計 30 億超のダウンロードと 30 万超の派生モデルを記録し、Google（4.18 億 DL）・Meta（2.27 億 DL）を大きく上回って世界首位に立っている[41]。

6.3 企業導入の天井

もっとも、OpenRouter の実運用比率は個人開発者では 70%超に達する一方、Fortune 500 企業では 30%未満にとどまる [38]。この差の背景には、データ主権とコンプライアンス上の懸念がある。具体的には、中国政府へのデータ共有義務、CAISI が指摘するエージェント乗っ取り耐性の低さ（米国モデルの約 12 倍脆弱）、および検閲挙動である [9,10]。

すなわち、ベンチマーク上のギャップが縮小しても、エンタープライズ採用のギャップは別の要因によって維持されている（分析）。

7. 実務的示唆

- ミッションクリティカル業務（複雑な多ファイルコーディング、長期自律エージェント、規制対応サイバー）：米国製（Claude Opus 5／Fable 5、GPT-5.6 Sol）を第一選択とすべきである。性能差は数点でも、独立検証・安全性・エコシステム成熟度で優位にある。判断を変える閾値としては、中国モデルが AA 知能指数で米国最上位と 1 点差以内となり、かつ独立検証機関がこれを追認した場合が挙げられる。
- 高 volume・コスト敏感・自己ホスティング要件：中国製オープンモデル（DeepSeek-V4 系、Qwen3.8、Kimi K3、GLM-5.3）が合理的である。4 倍から 100 倍に及ぶ価格差は無視できない。ただしエージェント乗っ取り耐性と検閲挙動を CAISI 報告に照らして事前検証すべきである [9,10]。
- 中間的選択肢：Meta のオープン系（Muse Spark 1.2、Muse Glimmer）は、西側製オープンウェイトとしてデータ主権要件とライセンス要件の両立を図る際の選択肢となる [4,5]。
- 機微データ処理：データ共有義務とセキュリティリスクを理由に、米国製または Meta 系オープンモデルを推奨する。中国モデルは非機微・実験的ワークロードでの段階導入が現実的である。

モニタリングすべき先行指標

以下の動きがあれば、上記の構造判断は更新を要する。

- GLM-5.3 の Artificial Analysis 知能指数の登録値
- DeepSeek-V4-Pro-0813 の独立測定による SWE-bench Pro 値
- Huawei Ascend 950（2026 年内予定）が学習用途で実用化するか
- Qwen3.8-Max および GLM-5.3 の重みが実際に公開されるか

8. 留保事項

■ 本レポートの性能数値の大半はベンダー自己申告値であり、独立検証（Artificial Analysis、Vals AI、Scale SEAL 等）が追いついていないものが多い。特に SWE-bench Verified は汚染とスキャフォールド差の問題が大きく、絶対値比較は慎重に扱うべきである[29,43,44]。

■ Artificial Analysis 知能指数はモデルの effort 設定により数点変動する。表 4 は主に max／xhigh 変種の値であり、厳密な同一条件比較ではない[1]。

■ 米中ギャップの定量値は測定機関により 3〜9 か月と大きくばらつき、政治的文脈も無視できない。単一の数字を断定的に用いるべきではない[9,10,11,12]。

■ GLM-5.3・Qwen3.8-Max・DeepSeek-V4-Pro-0813 GA 版の重み公開状況は流動的で、本稿執筆時点で一部未公開である。

■ 蒸留を巡る Anthropic 対 Moonshot の件、対中制裁、輸出規制は係争中・流動中であり、確定的評価は避けている[35]。

■ OpenRouter のトークンシェアは 1 プラットフォームの指標にすぎず、収益シェアや本番／実験の内訳は不明である[36,37,38]。

■ 参考文献のうち番号 11、35、41、42、43、44 については、本調査で一次ページを直接確認できておらず、二次情報経由の記述である。文献リストにおいて「URL 未確認」と明示した。

■ 「Muse Spark 1.2」「Grok 4.6」等については、公式ブログ・モデルカードの全文を直接確認できていない部分がある。開発元と公開日という重要事実は複数ソースで相互確認したが、細部の数値には二次情報依存が残る。

参考文献

番号は本文中の上付き数字に対応する。URL は 2026 年 8 月 16 日時点のもの。赤字イタリックで「URL 未確認」と記載した文献は、本調査において一次ページを直接確認できておらず、二次情報経由で参照した文献である。

[1] Artificial Analysis「Muse Spark 1.2 (xhigh) – Intelligence, Performance & Price Analysis」 https://artificialanalysis.ai/models/muse-spark-1-2（知能指数 v4.1.1 の横断比較データ）

[2] Fello AI「Best AI Models in August 2026: Updated Rankings and Comparisons」 https://felloai.com/best-ai-models/（2026 年 8 月時点の主要モデル横断ランキング（二次情報））

[3] The AI Rankings「Best AI Models in 2026: The Complete Ranking」 https://theairankings.com/best-ai-models/（ベンチマーク集計サイト（二次情報））

[4] Meta AI Developers Blog「Meet Muse Spark 1.2 and Muse Code: a coding model and the agent built to run it」 https://developer.meta.com/ai/resources/blog/build-with-muse-code/（Muse Spark 1.2 の開発元・仕様に関する一次情報）

[5] Meta AI Research「Introducing Muse Code and Muse Spark 1.2」 https://research.meta.ai/blog/introducing-muse-code-and-muse-spark-1-2（Meta 公式リサーチブログ（一次情報））

[6] 9to5Mac「SpaceXAI releases Grok 4.6, claiming GPT-5.6 Sol and Claude Fable 5-level intelligence」 https://9to5mac.com/2026/08/12/spacexai-releases-grok-4-6/（開発主体が SpaceXAI 表記である点の裏付け）

[7] MarkTechPost「SpaceXAI Releases Grok 4.6: A 500K-Context Frontier Model Tuned for Long-Running Agents, Coding, and Knowledge Work」 https://www.marktechpost.com/2026/08/12/spacexai-releases-grok-4-6/（Grok 4.6 の仕様・コンテキスト長）

[8] Epoch AI「The Geopolitics of AI: Data & Research」 https://epoch.ai/topics/geopolitics（ECI（Epoch Capabilities Index）による米中差の定量分析）

[9] NIST「CAISI Evaluation of DeepSeek V4 Pro（ニュースリリース）」 https://www.nist.gov/news-events/news/2026/05/caisi-evaluation-deepseek-v4-pro（米政府 CAISI による中国モデル評価（一次情報））

[10] NIST「CAISI Evaluation of DeepSeek AI Models（報告書本体）」 https://www.nist.gov/document/caisi-evaluation-deepseek-ai-models-report（同上の詳細報告書）

[11] Stanford HAI「AI Index Report 2026」 *URL 未確認（本調査で一次ページを直接確認できていない）*（LMArena Elo の米中差 2.7%、民間投資額比較の出典）

[12] Rice University, MCNAIR「The US-China AI gap is very uncertain.」 https://mcnair.center/china/（CAISI と Epoch の読み方の差異を整理した分析）

[13] Yahoo Tech「Anthropic releases Claude Opus 5 AI model on all platforms」 https://tech.yahoo.com/ai/claude/articles/anthropic-releases-claude-opus-5-062720389.html（Claude Opus 5 のリリース日・価格）

[14] Tech Times「Anthropic Launches Claude Fable 5: Most Powerful Public Model, Gated by Safeguards」 https://www.techtimes.com/articles/318082/20260609/anthropic-launches-claude-fable-5-most-powerful-public-model-gated-safeguards.htm（Claude Fable 5 のリリースとセーフガード構成）

[15] Google AI for Developers「Gemini 3.7 Flash（公式モデルドキュメント）」 https://ai.google.dev/gemini-api/docs/models/gemini-3.7-flash（Gemini 3.7 Flash の仕様（一次情報））

[16] VentureBeat「Google's Gemini 3.7 Flash targets coding and agents with a 50% introductory price cut」 https://venturebeat.com/technology/googles-gemini-3-7-flash-targets-coding-and-agents-with-a-50-introductory-price-cut（導入価格と位置づけ）

[17] TestingCatalog「ICYMI: xAI releases Grok 4.6 for long-running agent work」https://www.testingcatalog.com/icymi-xai-releases-grok-4-6-for-long-running-agent-work/（Grok 4.6 の発表内容）

[18] OpenRouter「Muse Spark 1.2 – API Pricing & Benchmarks」https://openrouter.ai/meta/muse-spark-1.2（API 価格・コンテキスト長）

[19] Northflank「Kimi K3: benchmarks, pricing, hardware requirements, and self-hosting」https://northflank.com/blog/what-is-kimi-k3-self-hosting（Kimi K3 のアーキテクチャ・価格）

[20] VentureBeat「China's Moonshot AI releases Kimi K3, the largest open-source model ever, rivaling top U.S. systems」https://venturebeat.com/technology/chinas-moonshot-ai-releases-kimi-k3-the-largest-open-source-model-ever-rivaling-top-u-s-systems（Kimi K3 の規模と位置づけ）

[21] Artificial Analysis「Kimi K3 achieves #3 in the Artificial Analysis Intelligence Index」https://artificialanalysis.ai/articles/kimi-k3-achieves-3-in-the-artificial-analysis-intelligence-index-comparable-to-opus-4-8-and-gpt-5-5（第三者測定による中国モデルの相対位置）

[22] Kili Technology「Kimi K3's Benchmarks and Hallucinations — What That Tells Us About AI Evaluation」https://kili-technology.com/blog/kimi-k3s-benchmarks-and-hallucinations----what-that-tells-us-about-ai-evaluation（ハルシネーション率に関する批判的検証）

[23] Technology Org「Alibaba Qwen3.8-Max: China's Biggest AI Model」https://www.technology.org/2026/08/03/alibaba-qwen38-max-24-trillion-parameters/（Qwen3.8-Max のパラメータ規模）

[24] DataNorth「Qwen3.8-Max: Alibaba's 2.4T open-weight AI model」https://datanorth.ai/news/alibaba-releases-qwen3-8-max（オープンウェイト化方針と仕様）

[25] OfficeChai「Qwen 3.8 Max Scores 56 On Artificial Analysis Intelligence Index」https://officechai.com/ai/qwen-3-8-max-scores-56-on-artificial-analysis-intelligence-index-ahead-of-all-us-companies-except-anthropic-and-openai/（AA 指数の再測定による変動の記録）

[26] Yotta Labs「DeepSeek V4: Release Date, Specs, and How to Access It (2026)」https://www.yottalabs.ai/post/deepseek-v4-release-date-specs-how-to-access-2026（DeepSeek-V4-Pro-0813 の仕様・公開時期）

[27] The Decoder「Zhipu AI releases GLM-5.3, claims it's the strongest open-weights coding model」https://the-decoder.com/zhipu-ai-releases-glm-5-3-claims-its-the-strongest-open-weights-coding-model/（GLM-5.3 の発表内容と自己申告値）

[28] Interconnects (Nathan Lambert)「GLM-5.3: How Chinese labs keep stride with the frontier」https://www.interconnects.ai/p/glm-53-how-chinese-labs-keep-stride（中国ラボの追随構造に関する分析）

[29] Local AI Master「SWE-bench Verified Leaderboard 2026: Top Models Ranked (+ SWE-bench Pro)」https://localaimaster.com/models/swe-bench-explained-ai-benchmarks（SWE-bench の汚染・スキャフォールド問題）

[30] Morph「Best AI Model for Coding (2026): 13 Models Ranked by Benchmarks and Cost per Task」https://www.morphllm.com/best-ai-model-for-coding（コーディング性能とタスク単価の比較）

[31] MACGPU Blog「DeepSeek V4 Full Release (July 2026): Pricing, Benchmarks & GPT-5.6 Comparison」https://macgpu.com/en/blog/2026-0720-deepseek-v4-full-release-pricing-benchmarks.html（価格比較（二次情報））

[32] CodingFleet「DeepSWE v1.1 Leaderboard: Claude Opus 5 Leads at 74% (2026)」https://codingfleet.com/blog/deepswe-v11-leaderboard-2026/（エージェント系ベンチの横断スコア）

[33] Codersera「Kimi K3 Benchmarks: How It Stacks Up vs Fable 5, GPT-5.6 Sol & Opus 4.8 (2026)」https://codersera.com/blog/kimi-k3-benchmarks-comparison-2026/（Kimi K3 の自己申告ベンチ値）

[34] Abhishek Gautam「Huawei Ascend 910C: China Plans 600,000 AI Chips in 2026」https://www.abhs.in/blog/huawei-ascend-910c-china-nvidia-alternative-2026（Ascend 910C の生産計画・性能（Bloomberg 報道の二次引用））

[35] Anthropic「Detecting and preventing distillation attacks」 *URL 未確認（本調査で一次ページを直接確認できてい*

ない）（蒸留を巡る中国ラボへの指摘）

[36] AI Weekly「Chinese AI Models Hit Record 58% of US OpenRouter Traffic」https://aiweekly.co/alerts/chinese-ai-models-hit-record-58-of-us-openrouter-traffic（米国開発者の実運用シェア）

[37] KuCoin「OpenRouter Data Shows 61% of Token Consumption by Chinese AI Models」https://www.kucoin.com/news/flash/openrouter-data-shows-61-of-token-consumption-by-chinese-ai-models?lang=en_US（トークンシェアの別集計値）

[38] MACGPU Blog「OpenRouter June 2026 Rankings Decoded: Chinese Models Own 61% of Developer Traffic」https://macgpu.com/en/blog/2026-0701-openrouter-june-rankings-chinese-models-61-percent.html（モデル別トークンシェアの内訳）

[39] Chris Zeoli (Data Gravity)「China's Open-Weight Takeover」https://www.datagravity.dev/p/chinas-open-weight-takeover（オープンウェイト領域の勢力図）

[40] Dasym「China Has Closed the Gap with the United States in Artificial Intelligence」https://www.dasym.com/en/the-weekly-worldview/china-has-closed-the-gap-with-the-united-states-in-artificial-intelligence（「追いついた」派の論拠）

[41] Hugging Face「State of Open Models Report（2026 年 8 月 14 日公開、Business Standard 報道経由）」*URL 未確認（原典を直接確認できていない）*（Qwen 系のダウンロード数・派生モデル数）

[42] AEI, Ryan Fedasiuk「米中 AI ギャップに関する論考（2026 年 7 月）」*URL 未確認（原典を直接確認できていない）*（「数週間差」とする最も楽観的な評価）

[43] llm-stats「SWE-bench トラッカー」*URL 未確認（集計サイト、原典を直接確認できていない）*（独立検証値の欠如に関する注記）

[44] Vals AI「独立ベンチマーク測定」*URL 未確認（原典を直接確認できていない）*（GPT-5.6 Sol の SWE-bench 独立測定値）