

Kimi K3 徹底検証

中国 Moonshot AI の新型モデルは本当に OpenAI・Anthropic のフロンティアを「上回った」のか

調査日：2026 年 7 月 18 日（第 2 版・出典レビュー反映）

技術・ベンチマーク面と業界・地政学的インパクト面

Claude Fable 5

本レポートの確度区分（本文中に付記）

〔公式〕 開発元の公式発表・公式仕様 〔独立確認〕 第三者評価（Artificial Analysis／LMArena） 〔ベンダー公称〕

Moonshot 公称値・未再現 〔報道〕 報道ベース 〔推測〕 分析・推測

※「公式」は「企業がそう発表したことが確認済み」の意で、内容の真偽が独立検証済みという意味ではない。

要旨（TL;DR）

- 「Kimi K3」は実在する。Moonshot AI（月之暗面）は 2026 年 7 月 16 日に Kimi K3 を公式発表したと告知している。2.8 兆パラメータの MoE モデルで、「K2」やその派生（K2 Instruct／K2 Thinking）ではない。ただし「OpenAI・Anthropic のフロンティアを上回った」という主張は総合的には支持できない。公表値および一部の独立評価では、フロントエンド生成・長時間コーディング・検索・文書処理でフロンティア級を示す一方、総合評価や高難度の専門知識・推論では最上位モデルを一貫しては上回っていない。「特定評価での優位」と表現するのが妥当。〔独立確認〕^{2,4}
- 順位は数え方次第。Artificial Analysis Intelligence Index（v4.1、9 評価の合成）で K3 は 57.1。設定別ランキングでは 4 位（Sol xhigh にも次ぐ）、各モデルファミリーの最良設定のみで数えると 3 位相当で、AA 自身の見出しも「#3」。ただし Sol xhigh との差は約 0.54 点、指数全体の 95%信頼区間は約 1 点未満であり、決定的な能力差ではない。Fable 5（約 60）・GPT-5.6 Sol（約 59）に次ぎ、Opus 4.8（約 56）とほぼ同等。〔独立確認〕^{2,3,15}
- K3 が優位なのは特定領域。フロントエンド生成（LMArena Frontend Code Arena で 1679 Elo・暫定首位）、長時間エージェント型コーディング（SWE Marathon）、ディープ検索（BrowseComp）、文書理解（OmniDocBench）で首位級。一方、高難度評価 HLE では劣後し、知識評価 AA-Omniscience では正答率が上がる一方、誤りを認めず回答する傾向（幻覚）が K2.6 より強まったと報告されている。〔ベンダー公称〕 〔独立確認〕^{1,2}
- オープンウェイトは公開「予定」。K3 はオープンウェイト（重みのみ公開。学習データ・コード・完全レシピの公開は含意されない）として 7 月 27 日に公開予定。調査日時点で重みは未公開

のため、ライセンス本文・実ファイル構成・第三者環境での再現性・自前ホスティング要件は未確認。報道見出しでは「open-source」とも表記されるが、本文では「オープンウェイト」に統一する。〔公式〕^{1,14}

- **地政学的インパクトは大きいが因果は限定的。**発表は上海 WAIC 2026 と重なり、半導体株下落を説明する材料の一つとして報じられ「第 2 の DeepSeek ショック」との呼称も生じた。ただし下落は発表前から進行しており、地政学情勢・決算・設備投資懸念などとの寄与分解はできない。オープンウェイトの意義自体は本物で、米国ラボの価格支配力に構造的圧力を加える。〔報道〕^{7,8}

主要な発見

1. 事実確認：「K3」は実在、ただし用語の混同に注意

Moonshot AI は 2026 年 7 月 16 日に Kimi K3 を発表した。同社の系譜（K2〔2025 年 7 月〕→ K2.5〔2026 年 1 月〕→ K2.6〔4 月〕→ K2.7 Code〔6 月〕→ K3〔7 月〕）の最新フラッグシップである。2025 年 11 月の「Kimi K2 Thinking」は K2 世代の推論特化版で K3 とは別物。「最新は K2/K2 Thinking」という理解は 2026 年 7 月時点では誤りで、K3 が最新。混同が起きやすいのは、半年間で複数の派生版が高頻度に出たためである。〔報道〕^{10,22}

【競合モデル名についての注記】本文の比較対象名（GPT-5.6 Sol、GPT-5.5、Claude Fable 5、Claude Opus 4.8、Claude Sonnet 5 等）は、独立評価機関 Artificial Analysis や LMArena 等の集計上の表記に従う。OpenAI・Anthropic の公式モデルカードや API モデル ID は本調査では一次確認できていないため、とくに OpenAI 系の名称・提供開始日・評価設定は二次資料上の表記として扱う。正式モデルカード公開後の再確認を要する。〔推測〕

2. K3 の正体（公式発表済み事項）

- **〔報道〕開発元：**Moonshot AI（月之暗面、北京、2023 年 3 月設立、創業者は楊植麟〔Yang Zhilin〕ら清華大学の同窓生）。社名は Pink Floyd の『The Dark Side of the Moon』に由来。^{10,22}
- **〔報道〕発表日：**2026 年 7 月 16 日（上海 WAIC のタイミングに合わせた）。¹⁰
- **〔公式〕リリース形態：**発表時点で API と Kimi アプリ・Kimi Code で利用可。オープンウェイトの重みは 7 月 27 日に Hugging Face で公開予定（Modified MIT ライセンスの見込みだが本文未確認）。^{1,16}
- **〔公式〕アーキテクチャ：**2.8 兆総パラメータの MoE。公式ドキュメントによれば「896 エキスパートのうち 16 個をアクティベート」。MoE 部分の単純換算では約 500 億パラメータ相当が選択されるが、共有層・Attention・Embedding を含む正確なアクティブパラメータ数は未公開。1M トークン文脈、ネイティブ・マルチモーダル（画像入力）。新機構 Kimi Delta Attention（KDA、

100 万トークン文脈で最大 6.3 倍高速なデコーディングを主張)、Attention Residuals、Stable LatentMoE で K2 比約 2.5 倍のスケーリング効率を主張。量子化は SFT 段階からの QAT (MXFP4 重み/MXFP8 活性化)。^{1,5}

3. 「上回っている」の批判的検証

Moonshot 自身が「総合では Fable 5 と GPT-5.6 Sol に劣る」とローンチブログで明言している点は、モデル発表として異例に誠実。K3 が優位なのは主にコーディング/エージェント/マルチモーダルの特長評価で、高難度の専門知識・推論では劣後する。見出しの「フロンティア超え」はチェリーピーキングを含む。^{1,2,4}

詳細分析

技術・ベンチマーク面

〔独立確認〕 独立系評価（Artificial Analysis、2026 年 7 月 17 日）^{2,3}

- Intelligence Index v4.1（GDPval-AA v2／ τ^3 -Banking／Terminal-Bench v2.1／SciCode／HLE／GPQA Diamond／CritPt／AA-Omniscience／AA-LCR の 9 評価合成）：K3 は 57.1。設定別 4 位／ファミリー最良設定で 3 位相当。上位は Fable 5（約 60）、GPT-5.6 Sol max（約 59）、Sol xhigh（約 57.7）。Opus 4.8（約 56）とほぼ同等。GPT-5.6 ultra はマルチエージェント構成のため単一モデル比較から除外されている。^{2,15}
- エージェント系：GDPval-AA v2 で 1668 Elo（K2.6 の 1190 から大幅向上、GLM-5.2〔1514〕・GPT-5.5〔1494〕・Opus 4.8〔1600〕を上回るが Fable 5〔1760〕に及ばず）。AutomationBench-AA で約 53%・首位。AA-Briefcase で 1547 Elo・Fable 5 に次ぐ 2 位。^{2,3}
- 弱点：知識評価 AA-Omniscience 上では、未知・回答困難な問いに誤答を生成する傾向（幻覚）が K2.6 より強まったと報告される。ただしこの指標は当該ベンチ上の値であり、一般的な事実照会タスク全体の幻覚確率として解釈することはできない。出力は冗長（Index 完走に約 130M トークン、同種中央値 63M の倍）、生成速度は 62 tokens/秒と遅め。Moonshot 自身も「過度に能動的（excessively proactive）」で長時間タスク中に勝手な判断をしようと認めている。^{1,3}

〔ベンダー公称〕 公式ベンチマーク（K3 max vs 競合。ハーネス混在に注意）^{1,17}

「K3 max」の定義（API モデル名か／高推論予算設定か／複数試行の最大値か）は Moonshot 資料上で明示されていない。以下はいずれもベンダー公称値で、独立再現は限定的。

ベンチマーク	Kimi K3	Fable 5	GPT-5.6 Sol	Opus 4.8	評価（競合ごと）
HLE-Full（ツールなし）	43.5	53.3	44.5	49.8	Fable に大差／Sol には僅差
GPQA-Diamond	93.5	92.6	94.1	91.0	ほぼ首位級（Sol に僅差）
Terminal-Bench 2.1	88.3	84.6	88.8	—	僅差 2 位
FrontierSWE	81.2	86.6	71.3	—	Fable に劣後／Sol に優位
SWE Marathon	42.0	35.0	39.0	40.0	K3 構成が首位
Program Bench	77.8	76.8	77.6	—	K3 構成が首位
BrowseComp	91.2	88.0	90.4	—	K3 構成が首位

ベンチマーク	Kimi K3	Fable 5	GPT-5.6 Sol	Opus 4.8	評価（競合ごと）
OmniDocBench	91.1	89.8	—	—	K3 構成が首位

（注）評価はモデル単体ではなく Moonshot が採用した評価構成での比較。Terminal-Bench 2.1 は KimiCode ハーネスの値で、独立系（Vals）では 80.9 とより低い。HLE ツールありでは K3 56.0／Fable 5 63.0／Sol 58.0／Opus 4.8 57.9。BrowseComp は 300K のコンテキスト圧縮を使った値で、1M フルでは 90.4（Sol と事実上同値）。

- （独立確認）LMArena Frontend Code Arena（ブラインド投票）：K3 が 1679 Elo で暫定首位（Fable 5 1631、Sol 1618 を上回る）。7 ドメイン中 6 ドメインで首位。ただし投票数は集計中で「暫定」表示であり、Elo の絶対差（1679 対 1631）が統計的に有意かは併記情報がなく断定できない。またこの評価はフロントエンド成果物への人間の選好を測るもので、一般的なソフトウェア工学能力全体を意味しない。一般テキスト Arena では約 1486・#6～9 程度と平凡。^{4,14}

方法論上の重大な留保^{15,18}

Moonshot の表は K3 に KimiCode、競合に Claude Code／Codex／Terminus 2 など異なるエージェントハーネスを使っており、純粋なモデル比較ではない。Terminal-Bench・SWE Marathon・BrowseComp・Frontend Arena 等のエージェント系は、システムプロンプト、エージェントループ、ツール選択、コンテキスト圧縮、再試行回数、推論予算、タイムアウト、評価器モデルに大きく左右される。共通の mini-SWE-agent ハーネスを使う DeepSWE では K3 が 67.3（KimiCode の 67.5 とほぼ同じ）で感度チェックは良好だが、Sol の 73 には及ばない。全数値はベンダー公称（最大努力設定）であり、重みが 7 月 27 日に公開されて初めて真の検証が可能になる。¹⁸

コスト（指標の性質ごとに区別）^{2,3,16}

指標	K3	比較対象	出典・条件
公式 API 入力単価	\$3 /100 万	—	Moonshot 公式（キャッシュヒット \$0.30）
公式 API 出力単価	\$15 /100 万	K2.6 は\$4	Moonshot 公式（K2.6 比約 3 倍・Sonnet 5 同水準）
AA 評価タスク単価	\$0.94	Sol \$1.04／Opus 4.8 \$1.80	AA の指定評価条件下
対オープン勢	高い	GLM-5.2 \$0.32／DeepSeek V4 Pro \$0.04	AA 試算・オープンウェイト比では割高
実効コスト（推計）	条件次第で割高	—	二次資料・冗長性に起因・一般化不可

「冗長ゆえ GPT-5.6 より 5～7 割高くなりうる」という記述は、タスク・最大出力・キャッシュ利用・再試行回数に依存する二次推計であり、一般化できない。

- **〔報道〕** 学習：Moonshot は「潤沢な計算資源がなかったからこそ効率に集中した」と主張。公式ブログは H200 と別ベンダー GPGPU（Huawei Ascend と解釈）、輸出規制対応版 L20 でのカーネル最適化を示す。学習インフラの全容・正確なチップ構成は非公表。¹⁵

オープンウェイトの意義と現実¹⁶

2.8 兆パラメータの MXFP4 重みは理論上約 1.4TB（重み本体の最小値）。実運用では量子化スケール、メタデータ、KV キャッシュ、アクティベーション、通信バッファ、ルーティング、ランタイム予約領域が追加が必要。Moonshot は 64 以上のアクセラレータからなる「supernode」運用を推奨するが、各アクセラレータのメモリ容量・推論方式は明記されていないため再現性は限定的。事実上「自前運用できる」のは資金力のある組織のみで、多くのチームにとって「オープン」は推論市場での価格競争・継続性保険を意味する。¹⁶

業界・地政学的インパクト面

〔独立確認〕 〔報道〕 中国オープンウェイト勢の中での位置づけ ^{2,10}

2026 年の中国 AI は多極化：Zhipu/Z.ai（GLM-5.2、AA Index 約 51）、Alibaba（Qwen、最強の Max はクローズド）、DeepSeek（V4 Pro、最安値）、MiniMax（M3）、Moonshot（Kimi、エージェント/長時間コーディング特化）。重み公開後の K3 は AA 上で「ダウンロード可能な最強モデル」となる見込みで、GLM-5.2（51）・DeepSeek V4 Pro（44）を明確に上回るが、2.8 兆パラメータとオープン勢中で突出して大きい（GLM-5.2 は 753 億、DeepSeek V4 Pro は 1.6 兆）。^{2,10}

〔報道〕 米国フロンティアラボへの競争圧力 ²⁰

- OpenRouter 経由で米企業が使うトークンのうち中国モデル比率は 2026 年 2 月以降、毎週 30% 超（最高 46%）と報じられる。過去 12 か月平均 11%、2025 年上期は 4.5%。²⁰
- 中国オープンモデルは米大手より大幅に安いとされ、「安いオープンを既定にし難所だけフロンティア API を呼ぶ」ルーティング戦略が企業に拡大。ただし具体的な単価比較は出典・条件により変動する。^{16,20}

〔報道〕 半導体市場の反応 ^{7,8}

K3 発表は半導体株下落の説明材料の一つとして報じられ、フィラデルフィア半導体指数（SOX）の下落やベア相場入り、SoftBank・TSMC・香港上場中国 AI 株の下落と関連づけられた。ただし市場下落は発表前から進行しており、地政学情勢・決算・設備投資懸念など複合要因によるもので、K3 単独の因果寄与は特定できない。「3.3 兆ドルの時価総額喪失が K3 に起因した」といった強い因果づけは本文の範囲では確定できない。^{7,8}

〔報道〕 蒸留（distillation）論争 ^{11,12,13}

Anthropic は 2026 年 2 月 24 日の公式ブログ（本調査では二次報道経由で確認）で、DeepSeek・Moonshot・MiniMax が約 2.4 万の不正アカウントで 1600 万回超の Claude 対話を生成し能力を不正抽出したと申し立てた。うち Moonshot 分は 2 番目に多い約 340 万回で、エージェント推論・ツール使用・コーディング・データ分析・コンピュータビジョン等を標的にしたとされる。3 社は取材に応じていない。重要なのは区別で、知識蒸留そのものは一般的な機械学習手法である。争点は「他社 API を利用規約に反して大規模利用し能力を抽出したか」であり、Anthropic が行う合法的自社モデル蒸留とは別物である。専門家は違法/合法の境界が曖昧で、著作権より契約（利用規約）違反が主たる論点になると指摘する。^{11,12,13}

〔報道〕 輸出規制（一次資料は本調査で未取得） ¹⁰

報道によれば、米政府は 2026 年 7 月中旬に Nvidia H200 の対中輸出を条件付きで限定許可し始めた

され、政府への収益分配（約 25%とされる）やライセンス制、実出荷が「ごく僅か」との商務当局証言も伝えられる。ただし正式な規則名・対象製品・法的根拠・施行日・算定対象・米商務省/官報の一次資料は本調査では確認できていない。とくに「25%収益分配」は異例の内容であり、一次資料での再確認まで確定情報として扱わない。¹⁰

〔報道〕 政策論争と WAIC 2026^{9,10}

David Sacks (PCAST 共同議長) は X (7 月 17 日) で、中国モデルが初めて Frontend Code Arena で 1 位を取ったことを引き、規制がイノベーションを阻害すると批判したと報じられる。Gary Marcus は議会調査を、投資家 Gavin Baker は「クローズドな OpenAI・Anthropic には打撃だが、他レイヤー（電力・半導体・ハイパースケーラー・ソフト）にはプラス」と評したとされる。発表は上海 WAIC 2026 と同時で、習近平が初の基調講演を行い中国を「オープンソース AI の旗手」と位置づけたと報じられている。^{9,10}

今後の展望

〔報道〕 Moonshot の資金・企業戦略^{10,21}

- 2026 年 5 月、Meituan の VC 主導で約 20 億ドルを調達、評価額は約 200 億ドル超との報道（監査済み一次開示ではない）。²¹
- ARR は 2026 年 3 月の約 1 億ドルから 4 月に約 2 億ドルへ倍増との報道。さらに高評価額での次ラウンド模索、香港 IPO を初期協議中とされるが時期未定・見送りの可能性も。いずれも報道ベースで一次財務未確認。²¹

〔推測〕 中国 AI のフロンティアでの役割（分析・推測）²⁰

アナリストは中国トップ勢を米最高峰の「6~9 か月遅れ」と推定してきたが、K3 はこの差をさらに縮めた。「中国のオープンリリースが毎回予想より速くフロンティアとの距離を詰める」構図が定着しつつあり、断続的ショックではなく継続的な圧縮として計画すべき段階と考えられる。ただしコンテンツ規制、企業のデータガバナンス懸念、北京による最強モデルの海外アクセス制限検討という逆方向の動きもあり、普及には制約が残る。²⁰

〔推測〕 オープン vs クローズドの競争構図^{16,9}

K3 の含意は「オープンがクローズドに追いついた」こと自体より、「フロンティア級知能が恒久的寡占ではなく競争市場になりつつある」こと。OpenAI・Anthropic は次世代モデル（GPT-6、Claude Opus 5 等の名称が取り沙汰される）で再び距離を取ることが予想されるが、いずれも公式発表を確認できていない。米政府が安全保障を理由に一部 AI モデルの配布を制限する動きも報じられ、「取り消されないオープンモデル」がかえって安全な賭けに見えるという逆説的な状況が指摘される。^{16,9}

実務への提言

今すぐやるべきこと（開発者・技術意思決定者）

1. **限定パイロットを走らせる**：コーディングエージェント、ディープリサーチ、視覚的ドキュメント処理、長コンテキスト自動化のうち1つの高価値ワークフローに K3 API を投入し、K2.6/K2.7 Code と同一タスク・同一ハーネスでトークン数・\$・成功率を記録。ベンチ順位ではなく「完了品質」で採点する。
2. **正確性が要る用途は検証工程を必須化**：AA-Omniscience で幻覚傾向の悪化が報告されているため、事実系タスクでは引用確認・検索結果との照合・構造化された検証工程を導入し、K3 単独運用を避ける。ベンチの幻覚率をそのまま一般業務の幻覚確率と読み替えない。
3. **7月27日を待つ**：規制対象業界での自前ホストや戦略的オープンウェイト採用は、重み・ライセンス・モデルカード・技術報告の公開まで保留。Modified MIT の帰属条項を実リポジトリで確認してから判断。

判断を変える閾値（ベンチマーク）

- 重み公開後、独立ハーネス（共通の mini-SWE-agent 等）で SWE-bench 系・LiveCodeBench・HLE の再現が Moonshot 主張の±3 ポイント以内なら本格採用を検討。
- AA-Omniscience の幻覚傾向が後続版で改善するなら事実系ワークフローへの拡大を再検討。
- コミュニティが GLM-5.2/DeepSeek V4 より低い実効タスク単価を安定実現できたら、コスト最優先ワークロードの既定モデル切替を検討。

戦略的スタンス

- 「単一モデル」ではなくルーティング戦略：難所だけフロンティア（Sol/Fable 5 系）、量が出る「十分良い」タスクは安いオープン（GLM-5.2/DeepSeek/K3）、機密・レジリエンスにはローカル。
- データガバナンス上、可能なら重みを自前/西側クラウドでホストし、中国ホスト API へのプロンプト送信を避ける。中国のコンテンツ規制がモデル出力に及ぶ点も留意。

確度の総括（確定 vs 推測）

区分	本レポートでの主な該当
公式発表済み	K3 の存在・発表日（7/16）・2.8 兆パラメータ・1M 文脈・マルチモーダル・API 価格・重み 7/27 公開予定（＝Moonshot がそう発表したことが確認済み）

区分	本レポートでの主な該当
独立確認済み	AA Intelligence Index 57.1（設定別 4 位／ファミリー最良で 3 位相当）、AA 各種 Elo、LMarena Frontend Arena 暫定首位
ベンダー公称・未再現	Moonshot 公式ベンチ表全般（ハーネス混在・最大努力設定）。重み公開前は独立再現不可
報道ベース	資金調達・ARR・IPO 協議、市場下落との関連、蒸留申立ての数値、輸出規制（H200・25%等）、WAIC・要人発言
分析・推測	米中の技術差、価格支配力への影響、今後の競争構図、次世代モデルの投入予想

主な留意点：

- 「上回った」の評価は文脈依存。総合では上回っていない（AA で 3～4 位）。特定領域では首位級だが、順位の絶対差は小さく 95%信頼区間内に収まる場合がある。
- 競合モデル名・チップ構成・ライセンス・重み公開は未確認部分を含む。とくに OpenAI/Anthropic の公式一次資料、米商務省/BIS の規則、議会証言記録、Anthropic 蒸留ブログ原文、LMarena/AA の評価方法ページの直接参照は今後追加すべき。

参考文献

※ 本文中の上付き番号に対応。全 URL は 2026 年 7 月 18 日時点で参照。区分タグは出典の性質を示す。二次資料（補助）は主要な事実の唯一の根拠にはしていない。

A. 一次・独立評価・主要報道

- [1] **〔公式〕** Moonshot AI（月之暗面）「Kimi K3 Tech Blog: Open Frontier Intelligence（公式技術ブログ）」
<https://www.kimi.com/blog/kimi-k3>
- [2] **〔独立評価〕** Artificial Analysis 「Kimi K3 achieves #3 in the Artificial Analysis Intelligence Index, comparable to Opus 4.8 and GPT-5.5」 <https://artificialanalysis.ai/articles/kimi-k3-achieves-3-in-the-artificial-analysis-intelligence-index-comparable-to-opus-4-8-and-gpt-5-5>
- [3] **〔独立評価〕** Artificial Analysis 「Kimi K3 – Intelligence, Performance & Price Analysis（モデルページ）」
<https://artificialanalysis.ai/models/kimi-k3>
- [4] **〔報道〕** Decrypt 「China’s Kimi K3 Is Out—And Beats Claude Fable and GPT 5.6 Sol on Key Benchmarks」
<https://decrypt.co/373716/china-kimi-k3-largest-open-source-ai-model-ever-beats-claude-fable-gpt-5-6-sol>
- [5] **〔報道〕** Tom’s Hardware 「China’s 2.8-trillion-parameter Kimi K3 beats Claude Fable 5 in Frontend Code Arena benchmark」 <https://www.tomshardware.com/tech-industry/artificial-intelligence/moonshot-releases-2-8-trillion-parameter-kimi-k3>
- [6] **〔報道〕** Axios 「China’s open-weight Kimi model stuns AI world with frontier-level results」
<https://www.axios.com/2026/07/16/moonshot-kimi-ai-china-model-openai-anthropic>
- [7] **〔報道〕** Fortune 「Markets experience new DeepSeek shock after Moonshot AI releases Kimi K3」
<https://fortune.com/2026/07/17/china-moonshot-kimi-k3-markets-china-ai/>
- [8] **〔報道〕** The Next Web 「Kimi K3 spooked markets. The AI selloff was already loaded.」
<https://thenextweb.com/news/kimi-k3-china-ai-tech-rout-selloff>
- [9] **〔報道〕** The Next Web 「Trump’s AI advisor says Kimi K3 shows “how you lose the AI race.”」
<https://thenextweb.com/news/kimi-k3-reactions-sacks-khosla-marcus-ai-race>
- [10] **〔報道〕** Forbes 「Chinese AI Startup Moonshot Unveils Kimi K3 Model—Will It Challenge OpenAI And Anthropic?」 <https://www.forbes.com/sites/tylerroush/2026/07/17/chinese-ai-startup-moonshot-unveils-kimi-k3-model-will-it-challenge-openai-and-anthropic/>
- [11] **〔報道〕** CNN Business 「US AI giant Anthropic alleges China rivals DeepSeek, Minimax and Moonshot AI are cheating（2026 年 2 月 24 日）」 <https://www.cnn.com/2026/02/24/tech/anthropic-chinese-ai-distillation-intl-hnk>
- [12] **〔報道〕** VentureBeat 「Anthropic says DeepSeek, Moonshot and MiniMax used 24,000 fake accounts to distill Claude（2026 年 2 月）」 <https://venturebeat.com/technology/anthropic-says-deepseek-moonshot-and-minimax-used-24-000-fake-accounts-to>
- [13] **〔報道〕** CNBC 「Anthropic accuses DeepSeek, Moonshot and MiniMax of distillation attacks on Claude（2026 年 2 月 24 日）」 <https://www.cnbc.com/2026/02/24/anthropic-openai-china-firms-distillation-deepseek.html>
- [14] **〔報道〕** Winzheng 「Kimi K3 Tops Frontend Code Arena with 1679 Points; Open Weights to Be Released on July 27」 <https://www.winzheng.com/en/article/kimi-k3-tops-frontend-code-arena-moonshot-ai>

B. 補助・二次資料（唯一の根拠にはしない）

- [15] Kingy AI 「Kimi K3 Benchmarks: Ranking vs Frontier & Open Models（#3/#4 の集計方法を解説）」
<https://kingy.ai/blog/kimi-k3-benchmarks-specs-price-fable-5-gpt-5-6-sol/>
- [16] Trilogy AI (Substack) 「Kimi K3 Is Live: Pricing, Benchmarks, and the Wait for Public Weights」
<https://trilogyai.substack.com/p/kimi-k3-is-live-pricing-benchmarks>
- [17] OpenLM.ai 「Kimi K3（ベンチマーク集計ページ）」 <https://openlm.ai/kimi-k3/>
- [18] NxCode 「Kimi K3 Benchmarks Explained: A Coding-Agent Evaluation Guide」
<https://www.nxcode.io/resources/news/kimi-k3-benchmarks-coding-agent-evaluation-guide-2026>
- [19] Valletta Software 「Kimi K3 vs Claude Fable 5: The Full Benchmark Scorecard」
<https://vallettasoftware.com/blog/post/kimi-k3-vs-claude-fable-5>
- [20] Augmented Mind (Substack) 「Kimi K3: The Open-Source Model That Just Cracked the Frontier Moat」
<https://augmentedmind.substack.com/p/kimi-k3-the-open-source-model-that-cracked-the-frontier-moat>
- [21] The AI Rankings 「Moonshot AI in 2026: Kimi K3, Open Weights & Strategy」
<https://theairankings.com/moonshot/>
- [22] Wikipedia 「Moonshot AI / Kimi (chatbot)」 https://en.wikipedia.org/wiki/Moonshot_AI