

Kimi K3 (Moonshot AI) オープンウェイト公開

— 事実関係・技術仕様・第三者評価・ライセンス及び政策的論点の調査 —

2026 年 8 月 1 日

Claude Opus 5

1. 要旨

- ・ 公開は事実であるが、公開日は修正を要する。Kimi K3 はまず 2026 年 7 月 16 日に API・製品として公開され^{8,9}、完全なオープンウェイトは Moonshot が「2026 年 7 月 27 日までに」と予告したうえで、7 月 26 日から 27 日にかけて Hugging Face 上で公開された^{1,9}。ご指摘の「7 月 28 日公開」は不正確であり、7 月 28 日はアリババ・ファーウェイのサーバー対応発表や米中当局の応酬が続いた日付に相当する¹⁴。
- ・ 性能は「オープンウェイトとしてはフロンティア級」だが総合首位ではない。総パラメータ 2.8 兆・活性 104B の MoE 構成^{1,3}。第三者機関 Artificial Analysis の Intelligence Index v4.1 で 57 点・全モデル中 3~4 位^{5,6}、LMArena の Frontend Code Arena では 1 位 (1,679 Elo)⁷。Moonshot 自身も最上位プロプライエタリモデルに総合では及ばないと明記している¹。
- ・ 論点は三つに集約される。すなわち、①MIT ではない独自「Kimi K3 ライセンス」による商用しきい値条項^{1,21}、②米ホワイトハウスによる蒸留および輸出規制違反の疑惑（証拠未提示、Moonshot は否定）^{10,11,14}、③総容量約 1.56TB という実運用上のハードルの高さ^{24,27}である。

2. 公開の事実関係

開発元は Moonshot AI (月之暗面)。北京拠点、アリババが出資し、創業者は楊植麟 (Yang Zhilin) である³³。

2.1 公開の時系列 (一次情報で確認)

- ・ 2026 年 7 月 16 日 : Kimi K3 を Kimi.com、Kimi アプリ、Kimi Code、API で公開。上海の WAIC 開催に合わせたタイミングであった^{8,9}。
- ・ Moonshot は「2026 年 7 月 27 日までに」完全ウェイトを公開すると予告し、実際に 7 月 26 日から 27 日にかけて Hugging Face で公開した^{1,9,27}。技術報告書 (arXiv:2607.24653、7 月 27 日付) およびインフラ技術 (MoonEP、FlashKDA、AgentEnv) も同時に公開されている^{2,3}。

- ・「7月26日夜（米東部時間）に1日前倒しで公開された」との記述は二次情報である。Moonshot 自身のページには「by July 27, 2026」としか記載がなく、前倒し公開の具体的時刻は確認できていない^{1,9}。

2.2 公開場所・公開形態

項目	内容
Hugging Face	huggingface.co/moonshotai/Kimi-K3 (like 9,300 超、月間ダウンロード約 49 万件) ¹
GitHub	github.com/MoonshotAI/Kimi-K3 (技術報告書 k3_tech_report.pdf を収録) ^{2,4}
ModelScope	modelscope.cn/organization/moonshotai (公式にリンク) ¹
技術ブログ	kimi.com/blog/kimi-k3 ¹
公開バリエーション	単一 SKU (kimi-k3)。常時推論 (thinking) モデルで、reasoning_effort (low/high/max、既定は max) により調整。旧世代のような Base/Instruct/非 thinking の分割はない ¹
派生版	Unsloth 等のコミュニティが GGUF 量子化版を派生公開 ²⁴
ライセンス	Hugging Face 上の表記は独自の「kimi-k3」ライセンス (MIT でも Modified MIT でもない。第 7 章参照) ^{1,21}

3. 技術仕様

以下はいずれも公式モデルカードおよび技術報告書に基づく一次情報である^{1,3}。

項目	値
アーキテクチャ	Mixture-of-Experts (MoE)
総パラメータ	2.8 兆 (技術報告書では 2.78T)
活性パラメータ	104B (1 トークンあたり)
層数	93 層 (うち Dense 1 層)
アテンション構成	69 KDA + 24 Gated MLA (ブロックごとに KDA:MLA=3:1、最終層は Gated MLA)
専門家数	896 (ルーティング) / トークンあたり 16 活性 + 共有専門家 2
アテンション隠れ次元	7,168 (ヘッド数 96)
語彙サイズ	160K
コンテキスト長	1,048,576 トークン (1M)
活性化関数	SiTU-GLU
ビジョンエンコーダ	MoonViT-V2 (401M、スクラッチ学習、対照学習なし)
量子化	MXFP4 重み/MXFP8 活性 (量子化認識学習 QAT を SFT 段階から適用)

項目	値
モダリティ	テキスト・画像（動画理解も可）

3.1 技術的新規性

- ・ Kimi Delta Attention (KDA)：ハイブリッド線形アテンション。1M 文脈で最大 6.3 倍の高速デコード、KV キャッシュ最大 75%削減を主張³。
- ・ Attention Residuals：残差接続の置き換えにより学習効率が約 25%向上³。
- ・ Stable LatentMoE：極疎ルーティングの安定化 (Quantile Balancing + SiTU-GLU)³。
- ・ Per-Head Muon 最適化³。総合して K2 比で約 2.5 倍のスケーリング効率を主張している^{3,19}。

3.2 非開示事項

学習トークン数および学習計算量は非開示であり、技術報告書には絶対値の記載がなく、「K2 比 2.5 倍効率」という相対値のみが示されている³。学習データは領域 (Web／コード／数学／知識／ビジョン) でのみ記述され、データセット名、ライセンス、知識カットオフはいずれも非開示である³。学習データおよび学習コードも未公開であり、実行・改変は可能であるが再現学習はできない^{1,3}。

3.3 前世代からの変化

K2 系は 1 兆パラメータ・テキスト中心・256K 文脈であった。K3 は総パラメータが約 3 倍となり、ネイティブなマルチモーダル対応と 4 倍の文脈長 (1M) を得た^{18,30}。2025 年後半の K2 Thinking は INT4 量子化認識学習・32B 活性・256K 文脈・逐次 200～300 回のツール呼び出しを特徴としており、K3 はこれを MXFP4、マルチモーダル、KDA へと発展させた系譜と位置づけられる¹⁸。

4. ベンチマーク性能

4.1 第三者による独立評価

Artificial Analysis の Intelligence Index v4.1 (2026 年 7 月 23 日発表) において、Kimi K3 (max) は 57 点を獲得し、全 190 モデル中 3～4 位となった。同社は K3 の知能水準を Opus 4.8 および GPT-5.5 と同等とし、Fable 5 と GPT-5.6 Sol には及ばないと評価している^{5,6}。集計上、K3 (57) を上回るのは Claude Opus 5 (61)、Claude Fable 5 (60)、GPT-5.6 Sol (59) の 3 モデルのみであり、GLM-5.2 (51) や DeepSeek V4 Pro (44) を上回ってオープンウェイト最上位となる。ただし後発の Claude Opus 5 が首位となったため、「最新の総合最強」ではない^{5,6}。

コスト面では、タスクあたり 0.94 ドルと GPT-5.6 Sol (1.04 ドル) と同水準、Opus 4.8 (1.80 ドル) の約半額であるが、オープンウェイト勢の中では高い⁶。一方で出力が非常に冗長かつ低速で、出力トークンは中央値の約 2 倍、AA-Briefcase の 1 タスク平均時間は約 56 分 (Fable 5 の約 2.5 倍) に達する⁶。GDPval-AA v2 では Elo 1,668 を記録し、K2.6 の 1,190 から大幅に改善した⁶。幻覚については、AA-Omniscience で精度が 33% から 46% へ改善する一方、幻覚率は 39% から 51% へ悪化しており、指標が精度改善をより重く評価する構造のため総合スコアは向上している²⁰。

LMarena の Frontend Code Arena では Elo 1,679 で 1 位 (1,757 票時点)。K2.6 の 18 位から 17 ランク上昇し、7 ドメイン中 6 分野で首位、ゲーム分野のみ Fable 5 に次ぐ 2 位であった。ペアワイズ勝率は 76% (Fable 5 は 63%、GPT-5.6 Sol は 58%) である⁷。

4.2 Moonshot 公式評価表

以下は Moonshot による自己申告値であり、モデルごとに異なるハーネスで測定されている点に注意を要する¹。

ベンチマーク	K3	Fable 5	GPT-5.6 Sol	Opus 4.8	GPT-5.5	GLM-5.2
GPQA Diamond	93.5	92.6	94.1	91.0	93.5	91.2
HLE-Full (無ツール)	43.5	53.3	44.5	49.8	41.4	—
HLE-Full (有ツール)	56.0	63.0	58.0	57.9	52.2	—
AA-LCR (長文脈推論)	74.7	70.0	73.7	67.7	74.3	71.3
DeepSWE	67.5	70.0	73.0	59.0	67.0	46.2
ProgramBench	77.8	76.8	77.6	71.9	70.8	63.7
Terminal-Bench 2.1	88.3	88.0	88.8	84.6	83.4	82.7
FrontierSWE	81.2	86.6	71.3	66.7	64.9	67.3
SWE-Marathon	42.0	35.0	39.0	40.0	14.0	13.0
BrowseComp	91.2	88.0	90.4	84.3	84.4	—
τ^3 -Banking	33.4	26.8	33.0	27.6	31.3	26.8
MMMU-Pro (無ツール)	81.6	81.2	83.0	78.9	81.2	—
MMMU-Pro (有ツール)	83.4	86.5	84.6	82.7	83.2	—

注：DeepSWE は mini-SWE-agent 使用時 67.3。BrowseComp の 91.2 は 300K 圧縮時、1M 時は 90.4¹。

4.3 評価上の重要な注意

- ・ Moonshot の公式表には「SWE-bench Verified」「SWE-bench Multilingual」「AIME」「LiveCodeBench」「MMLU-Pro」の独立行は存在しない。一部の二次メディアが報じる「SWE-

bench Verified 76.8%」は公式表に見当たらず、出典不明のため未確認として扱うべきである¹。コーディング系は DeepSWE、ProgramBench、Terminal-Bench、FrontierSWE、SWE-Marathon 等の独自スイート、視覚系は MMMU-Pro（テキストの MMLU-Pro ではない）である¹。

- ・ハーネス混在の問題。K3 は Kimi Code ハーネス、競合は Claude Code、Codex、Terminus 2 等、モデルごとに異なるハーネスで測定されている。ハーネス差だけでスコアが 10~26 点動くため、モデル単体の比較には使用できない旨が第三者から繰り返し指摘されている^{17,31}。
- ・エージェント能力の実証として、48 時間の連続自律動作によりオープン EDA ツールと Nangate 45nm ライブラリを用いた専用 AI チップ（面積 4mm²、146 万標準セル、INT4 MAC アレイ、100MHz でタイミング収束）の設計に成功したとする概念実証が示されているが、これは自己申告であり第三者検証は確認できていない^{2,3}。

5. 反響と評価

5.1 メディア報道

VentureBeat、Bloomberg、Fortune、South China Morning Post、日本経済新聞、Forbes JAPAN、Caixin 等が大きく報じた。VentureBeat は、K3 を同社が「世界最大のオープンソース AI モデル」と称する 2.8 兆パラメータのモデルであり、ベンチマーク上は Anthropic および OpenAI の最有力プロプライエタリシステムと互角に競っていると紹介する一方、ライセンス上の注意喚起を強調している⁸。日本では日経が「アリババ・ファーウェイが 7 月 28 日に Kimi K3 対応を発表」と報じ、Forbes JAPAN 等が企業導入の観点から解説している¹⁴。

5.2 開発者コミュニティの反応

r/LocalLLaMA、Hacker News、X 上では Frontend Code Arena 首位が特に話題となった^{7,34}。Unsloth が 7 月 29 日にダイナミック GGUF（1bit 版で約 594~620GB）を公開し²⁴、one-shot で three.js シーンや Minecraft クローンを生成できたとの報告が多数寄せられている^{17,32}。

5.3 批判・懐疑論

- ・実運用ギャップ：独立したハンズオン評価では、単純タスクでは Opus 4.8 とほぼ互角である一方、罫を仕込んだ複雑タスクでは失敗率 36%（Opus は 8%）との報告がある³²。
- ・ベンチマーク汚染懸念：LiveBench 等の非汚染評価による検証を求める声が上がっている。ProgramBench 側からは指標選択が部分点を過大評価するとの指摘もある^{20,31}。

- ・冗長さとコスト：常時推論のため出力トークンを多く消費し、実効コストが見かけより高いとの指摘がある^{6,17}。
- ・セキュリティ用途：Sengrep は、偽陽性率の観点から精度が低く、エンタープライズ実務では見かけほど強くないと評価している³²。

6. 産業・政策的インパクト

6.1 市場反応

K3 の発表を機に中国 AI 競合各社の株価が下落したと報じられている（Z.ai が最大 30%安、MiniMax が最大 16%安、アリババが 4%安）¹⁶。半導体株への影響について Tech Times は「3.3 兆ドルが消失した」と報じているが、この巨額の数値は単一媒体の見出しであり裏付けが不十分である¹⁶。Moonshot の日次売上は K3 公開後に 6 倍以上に伸び、ARR は 6 月時点で 3 億ドル、香港 IPO および 300～500 億ドル規模の評価での調達を模索していると報じられている¹⁶。

6.2 米中の応酬

2026 年 7 月 22 日、ホワイトハウス OSTP 局長 Michael Kratsios が X 上で、Moonshot が Anthropic の Fable を蒸留して K3 を開発したとの情報があること、検知を回避するため複数のアクセス手段を切り替える大規模蒸留用の内部プラットフォームを構築したこと、さらにタイ経由で Nvidia GB300 サーバーにアクセスしたことを主張した^{11,14}。財務長官 Scott Bessent は知財窃盗を巡り Moonshot への制裁権限があると述べ、制裁や Entity List 入りが選択肢となると警告した¹⁴。米商務省 BIS が正式調査中であるが、7 月 28 日時点で正式措置は確認されていない¹⁴。

前史として、Anthropic は 2026 年 2 月の申立てにおいて、Moonshot が DeepSeek、MiniMax とともに産業規模の蒸留攻撃を行い、Moonshot 単独で不正アカウント経由の 340 万件超のやり取りを生成したと主張したと報じられている（報道ベース）¹¹。中国大使館はこの蒸留主張を全くの中傷であると反論した¹¹。

反証として、TechCrunch および SCMP は、Fable 5 が 7 月 1 日によりやく一般公開されたばかりであることから、K3 が主として同モデルからの蒸留によって開発され得たという見方に疑問を呈する専門家が存在すると報じている。Fable 5 の一般公開（7 月 1 日）から K3 公開（7 月 16 日）まで 15 日しかない点が根拠である¹⁰。Moonshot の従業員 Randy Xian は「15 日で新たなフロンティアモデルを訓練したのならギネス級だ」と皮肉り¹⁰、Moonshot 自身も中国「毎日経済新聞」に対し、アーキテクチャの独自改良によるものと説明して疑惑を否定している¹⁰。中国商務省は 7 月 27 日に「AI 覇権主義」と反発し、必要なあらゆる措置を取ると警告した^{11,14}。

業界の反応としては、7月24日に Nvidia CEO の Jensen Huang 主導でオープンソースモデルの重要性を訴える公開書簡が出され、当初 25 社から後に 132 社（Meta、Microsoft、Amazon、Google、OpenAI 等）へ拡大した¹¹。

6.3 エコシステムと採用

Fireworks AI および Together AI が Day-0 ホスティングを実施した²⁶。7月30日から31日にかけては Microsoft Foundry 上で Fireworks AI 経由の Kimi K3 が利用可能となり、価格は入力 100 万トークンあたり 3.30 ドル、出力 16.50 ドルである^{12,13}。ファーウェイ昇騰（Ascend）CANN が Day-0 で適配し、アリババクラウドも対応した¹⁴。一方、AWS Bedrock や Google Vertex 等の大手ハイパースケーラーは、政策リスクもあって当初対応を見送り、専業推論プロバイダが受け皿となった²⁸。

7. ライセンスと知的財産上の論点

7.1 ライセンスの構造

ライセンス名は独自の「Kimi K3 License」（Hugging Face 上の表記は「kimi-k3」）である。K2 系の「Modified MIT」から変更された点が実務上重要であり、MIT 由来ではあるが二つの商用しきい値条項を持つ^{1,21,22}。なお一部の二次媒体は「Modified MIT」と報じているが、モデルカードの表記と一致せず、誤りの疑いがある³²。

条項	内容
第2条（MaaS条項）	ライセンサー又はその関連会社が「Model-as-a-Service」事業を運営し、グループ合算売上が連続12か月で2,000万ドルを超える場合、商用利用前に Moonshot との別途契約が必要。クラウド・推論プロバイダ（AWS、GCP、Fireworks、Together、CoreWeave 等）を主たる対象とする。2,000万ドルは低いしきい値であり、実質的に多くの事業者が該当するとの指摘がある ^{21,22}
第3条（帰属表示条項）	K3 又はその派生を用いた商用製品が月間1億 MAU 超、又は月間売上2,000万ドル超の場合、UI 上に「Kimi K3」を明示的に表示する義務を負う ^{21,22}
第4条（免除）	内部利用（第三者に成果・能力を提供しない場合）、及び Moonshot 公式製品・認定推論パートナー経由の利用は第2条・第3条の適用外 ^{21,22}

商用利用については、使用、改変、デプロイ、ファインチューニング、再配布、サブライセンス、販売の各権利が明示的に付与されており（表示条件および大規模事業者の特則を除く）、派生モデルの作成・再配布も明示的に許可されている^{1,21}。

7.2 知財・訴訟リスクに関する論点

学習データの出所およびライセンスが非開示であるため、学習データに由来する知的財産上の論点は

現時点で評価不能である³。米政府の蒸留疑惑については証拠が提示されておらず、また蒸留自体が直ちに違法となるとは限らない（モデル出力に著作権は及ばないとする専門家見解がある）。より法的に問題化し得るのは輸出規制（GB300 のタイ経由アクセス）の側であると整理されている^{10,11}。いずれも係争中の未確定論点であり、断定的な評価は避けるべきである。

8. 分析と示唆

Kimi K3 は「オープンウェイトがフロンティアに追いついた」ことを示す象徴的事例である。Nathan Lambert（Interconnects、2026 年 7 月）は、K3 の登場によりオープン対クローズド、あるいは米国対中国のモデル性能差が、従来議論されてきた 6~9 か月から 3~5 か月程度へ縮小したと分析している¹⁸。

他方で、「史上最大のオープンモデル」という見出しはパラメータ数、すなわち能力の容量指標にすぎず、出力品質の保証ではない。品質指標では 3 位である点に留意が必要である^{5,6}。

技術的に見ると、K3 の本質はパラメータ数の積み増しではなく「スケーリング効率 2.5 倍」にある。KDA（線形アテンション）と Gated MLA の 3 対 1 の混在、LatentMoE による極疎ルーティング（896 専門家中 16 活性=1.8%）、MXFP4 量子化認識学習という組み合わせが、1M 文脈と 2.8 兆パラメータを実運用可能なコストに落とし込む設計思想を示している^{3,19,23}。

ただし「オープン=ローカル実行可能」ではない。公式リポジトリは約 1.56TB に達し、Moonshot は 64 アクセラレータ以上のスーパーノード構成を推奨、現実的な最低ラインは 8×H100 とされる^{24,25,27}。1bit GGUF（約 594~620GB）であっても 650GB 超の RAM と VRAM が必要であり、単一のコンシューマ機では動作しない^{24,29}。したがって、このオープン化は実質的にクラウドおよび推論プロバイダのためのものであるというのが実態評価となる²⁸。

9. 実務上の推奨

以下は本調査に基づく筆者の分析・提案であり、事実の記述ではない。

- ・まず API で評価し、自社ホスティングは慎重に判断する。短期評価は Moonshot API（入力 3 ドル／出力 15 ドル、キャッシュヒット 0.30 ドル）または Fireworks、Together、Microsoft Foundry 経由が現実的である^{12,13}。2.8 兆パラメータの自社運用は GPU コストと運用スキルの観点から大半のチームには推奨できない。継続的な高トラフィック（キャッシュヒット率 90%超のエージェント／RAG 用途）かつデータ主権要件がある場合に限り検討する^{23,27}。

- ・用途を絞って導入する。フロントエンドコード生成、ドキュメント理解、長文脈処理、自動化タスクでは K3 が強い⁷。一方、罣のある複雑なデバッグ、単発の深い推論、幻覚許容度の低い業務（セキュリティ精査等）では Fable 5 や Opus 系を優先すべきである^{20,32}。いずれにせよベンチマークではなく自社コードベースでのブラインド比較（A/B テスト）を必ず実施されたい。
- ・ライセンスを法務レビューに付す。MaaS としての再販や、月間 1 億 MAU・月間売上 2,000 万ドル級の製品に組み込む場合は、第 2 条（別途契約）および第 3 条（帰属表示）を事前に確認し、license@moonshot.ai に照会する。内部利用は原則免除であるが、「ログインの背後にあるから内部利用である」とは限らない点に注意を要する^{21,22}。
- ・データ主権と政策リスクを織り込む。中国の国家情報法、および BIS 調査・Entity List 入りの可能性を踏まえ、機微データについては API 経由（中国サーバー）を避け、セルフホストまたは認定パートナー経由を検討する^{14,15}。米政府が中国製モデルのホスティング規制に動けば供給が変わり得るため、継続的な監視が必要である²⁸。
- ・再評価のトリガーを定めておく。独立ハーネス（同一条件）での SWE 系再現結果、および LiveBench 等の非汚染評価スコアが公表された時点で、採否を再評価することが望ましい³¹。

10. 留保事項

- ・公開日：ご指摘の「7 月 28 日公開」は不正確である。API 公開は 7 月 16 日、ウェイト公開は 7 月 26 日から 27 日である^{1,8,9}。7 月 28 日はアリババ・ファーウェイのサーバー対応発表日等に相当する¹⁴。
- ・競合モデル名：本書中の「Claude Fable 5」「Claude Opus 4.8/5」「GPT-5.6 Sol」「GPT-5.5」「GLM-5.2」「DeepSeek V4 Pro」等は、いずれも出典中に現れる 2026 年時点の世代呼称である。実在性および正式名称については各社の一次情報で別途確認されたい。
- ・自己申告と第三者検証の区別：Moonshot 公式の評価表はハーネス混在の自己申告である。独立検証は Artificial Analysis（Intelligence Index 57 点、3～4 位）および LMArena（Frontend 1 位）が中心となる^{1,5,6,7}。48 時間でのチップ設計デモは自己申告であり未検証である^{2,3}。
- ・未確認・裏付け不足の情報：「SWE-bench Verified 76.8%」は公式表に存在せず出典不明。「半導体株 3.3 兆ドル消失」は単一媒体の見出しであり裏付けが不十分。「7 月 26 日前倒し公開」の具体的時刻は二次情報。株価下落率（Z.ai 30%等）は報道ベース。「340 万件の不正アカウント経由のやり取り」は Anthropic の主張（報道ベース）である^{11,16}。
- ・蒸留・輸出規制疑惑：米政府の主張は証拠が提示されておらず係争中である。法的にも技術的にも未確定であり、本書は事実認定を行わない^{10,11,14}。

- ・ 学習の詳細：学習トークン数、計算量、データセット名、知識カットオフはいずれも非開示であり、学習データおよびコードも未公開であるため再現学習はできない³。
- ・ URL の検証状況：参考文献に掲げた URL のうち、一次情報（1～4）および第三者検証（5～7）は内容を確認済みである。二次情報として分類したものは記事の存在を確認しているが、記載内容の一次情報との突合までは行っていない。

参考文献

末尾の角括弧は情報源の種別を示す。一次情報＝開発元による公式公表、第三者検証＝独立した評価機関・リーダーボード、報道＝報道機関、専門家ブログ＝著名実務家による解説、二次情報＝上記以外の解説記事・まとめ記事。

1. Moonshot AI 「moonshotai/Kimi-K3」 Hugging Face モデルカード。
<https://huggingface.co/moonshotai/Kimi-K3> [一次情報]
2. MoonshotAI/Kimi-K3 (GitHub リポジトリ) . <https://github.com/MoonshotAI/Kimi-K3> [一次情報]
3. Kimi K3 Technical Report, arXiv:2607.24653 (2026 年 7 月 27 日) . <https://arxiv.org/pdf/2607.24653> [一次情報]
4. Kimi-K3/k3_tech_report.pdf (GitHub 収録の技術報告書) . https://github.com/MoonshotAI/Kimi-K3/blob/main/k3_tech_report.pdf [一次情報]
5. 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> [第三者検証]
6. Artificial Analysis 「Kimi K3 (max) – Intelligence, Performance & Price Analysis」 .
<https://artificialanalysis.ai/models/kimi-k3> [第三者検証]
7. LMArena 公式 X 投稿 (Frontend Code Arena 順位) . <https://x.com/arena/status/2077893862778183737>
[第三者検証]
8. 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> [報道]
9. Quartz 「Moonshot AI releases Kimi K3 open-weight model for download」 . <https://qz.com/moonshot-ai-kimi-k3-open-weights-download-072726> [報道]
10. South China Morning Post 「Global AI experts push back on US "distillation" claims against Moonshot's Kimi K3 model」 . <https://www.scmp.com/tech/tech-war/article/3361625/global-ai-experts-push-back-us-distillation-claims-against-moonshots-kimi-k3-model> [報道]
11. Calcalist Tech 「The AI cold war escalates as China and U.S. clash over Moonshot's Kimi model」 .
<https://www.calcalistech.com/ctechnews/article/hjbji0hsze> [報道]
12. TechNode 「Moonshot AI's Kimi K3 reaches enterprise users through Fireworks on Microsoft Foundry」
(2026 年 7 月 30 日) . <https://technode.com/2026/07/30/moonshot-ais-kimi-k3-reaches-enterprise-users-through-fireworks-on-microsoft-foundry/> [報道]
13. Neowin 「Kimi K3 model is now deployable on Microsoft Foundry through Fireworks AI」 .
<https://www.neowin.net/news/kimi-k3-model-is-now-deployable-on-microsoft-foundry-through-fireworks-ai/>
[報道]
14. Tech Times 「China Fires Back as US Targets Moonshot AI Over Kimi K3 Anthropic Fable Theft Claim」
(2026 年 7 月 28 日) . <https://www.techtimes.com/articles/321757/20260728/china-fires-back-us-targets-moonshot-ai-over-kimi-k3-anthropic-fable-theft-claim.htm> [報道]

15. Tech Times 「Kimi K3 Open Weights Arrive Sunday: Self-Hosting Cuts China Data Risk the API Never Can」 (2026 年 7 月 25 日) . <https://www.techtimes.com/articles/321551/20260725/kimi-k3-open-weights-arrive-sunday-self-hosting-cuts-china-data-risk-api-never-can.htm> [報道]
16. Tech Times 「Kimi K3 Wipes \$3.3T From Chip Stocks: Moonshot Moves Toward Hong Kong IPO」 (2026 年 7 月 20 日) . <https://www.techtimes.com/articles/321066/20260720/kimi-k3-wipes-33t-chip-stocks-moonshot-moves-toward-hong-kong-ipo.htm> [報道 (数値裏付け不十分)]
17. Simon Willison 「moonshotai/Kimi-K3」 (2026 年 7 月 27 日) . <https://simonwillison.net/2026/Jul/27/kimi-k3/> [専門家ブログ]
18. Geopolitechs 「Moonshot released Kimi K3 model weights and technical report」 . <https://www.geopolitechs.org/p/moonshot-released-kimi-k3-model-weights> [二次情報]
19. MoClaw 「Kimi K3 Technical Report: What It Reveals」 . <https://moclai.ai/blog/kimi-k3-technical-report> [二次情報]
20. Kili Technology 「Kimi K3's Benchmarks and Hallucinations」 . <https://kili-technology.com/blog/kimi-k3s-benchmarks-and-hallucinations---what-that-tells-us-about-ai-evaluation> [二次情報]
21. Digital Applied 「Kimi K3 Open Weights Shipped: What the Licence Says」 . <https://www.digitalapplied.com/blog/kimi-k3-open-weights-shipped-license-restrictions-2026> [二次情報]
22. Roo Newsletter 「Kimi K3 Open Weights Are Live: 2.8T Parameters, 1M Context, and a License Nobody Actually Read」 . <https://roo.beehiiv.com/p/kimi-k3-open-weights-license-benchmarks> [二次情報]
23. Runpod 「Kimi K3: KDA, MXFP4, and the self-host breakeven math」 . <https://www.runpod.io/articles/guides/kimi-k3-technical-faq> [二次情報]
24. ModemGuides 「Run Kimi K3 Locally: 1-Bit GGUF Size and RAM Requirements」 . <https://www.modemguides.com/blogs/ai-news/run-kimi-k3-locally-hardware-reality-check> [二次情報]
25. explainx.ai 「Run Kimi K3 Locally: Confirmed Hardware + vLLM (2026)」 . <https://explainx.ai/blog/kimi-k3-run-locally-open-weights-desktop-july-2026> [二次情報]
26. explainx.ai 「Kimi K3 Open Weights: 2.8T Params, Day-0 Hosting (2026)」 . <https://www.explainx.ai/blog/kimi-k3-open-weights-2-8-trillion-parameters-july-2026> [二次情報]
27. TECHi 「Kimi K3's open weights arrive July 27. The catch is 1.4TB」 . <https://www.techi.com/kimi-k3-open-weights-inference-economics/> [二次情報]
28. Memeburn 「Kimi K3 Cloud Freeze Shows the AI Distribution Wall」 . <https://memeburn.com/kimi-k3-cloud-freeze-shows-the-ai-distribution-wall/> [二次情報]
29. Kingy AI 「Kimi K3 Open Weights: Download and Hardware Guide」 . <https://kingy.ai/blog/kimi-k3-open-weights-download-hardware/> [二次情報]
30. Build Fast with AI 「Kimi K3 Review: Benchmarks, Pricing, and K2 Comparison」 . <https://www.buildfastwithai.com/blogs/kimi-k3-review> [二次情報]
31. BenchLM 「Kimi K3: The Open Model Closing the Gap」 . <https://benchlm.ai/blog/posts/kimi-3-release-data-coming-soon> [二次情報]
32. AIToolsRecap 「Kimi K3 Weights Are Live: Download From HuggingFace, Modified MIT License Confirmed」 . <https://aitoolsrecap.com/Blog/kimi-k3-weights-live-download-huggingface-july-27-2026> [二次情報 (ライセンス名の記載に誤りの疑い)]
33. Wikipedia 「Kimi (chatbot)」 . [https://en.wikipedia.org/wiki/Kimi_\(chatbot\)](https://en.wikipedia.org/wiki/Kimi_(chatbot)) [二次情報]

34. X (旧 Twitter) トレンド「Moonshot AI Releases Massive 2.8 Trillion Parameter Kimi K3 Model」 .
<https://x.com/i/trending/2081575186881540491> [二次情報]