中国製 LLM の現状と今後

—Kimi K3・GLM-5.2・Qwen3.7 Max・DeepSeek V4 Pro の実力評価と日本企業の活用戦略

—

2026 年 7 月 19 日

Claude Fable 5

1. エグゼクティブサマリー

2026 年前半、中国の AI 研究開発は「数か月遅れの追走者」から「事実上の並走者」へと局面を変えた。4 月の DeepSeek V4 Pro、5 月の Qwen3.7 Max、6 月の GLM-5.2、そして 7 月 16 日の Kimi K3 と、わずか 3 か月弱の間に 4 つのフロンティア級モデルが相次いで公開され、いずれも米国製最上位モデルとの直接比較が真剣に議論される水準に達している。特に Kimi K3 は、独立系評価機関の総合ランキングで全モデル中 4 位（Claude Fable 5、GPT-5.6 Sol に次ぎ、Claude Opus 4.8 を上回る位置）と評価された [2]。

本稿の結論を先に述べる。第一に、「ほぼ追いついた」という評価は、ベンチマーク性能とコスト効率の面では概ね妥当だが、最高難度タスクでの絶対性能、実運用での信頼性、安全性・データガバナンスの面では依然として留保が必要である。第二に、「追い越すか」については、単一の答えはなく、オープンウェイト分野では既に中国が世界のデファクトを握りつつある一方、クローズドの最先端では米国優位が続く可能性が高い。第三に、日本企業にとって中国製 LLM は「使うか使わないか」の二択ではなく、「どの形態で・どの用途に・どの統制の下で使うか」を設計する対象である。クラウド API 経由の利用と、重みをダウンロードして自社環境で運用するセルフホスト利用とではリスク構造が根本的に異なり、この区別を明確にした社内ポリシーの整備が急務である。

2. 主要 4 モデルの現状

2.1 DeepSeek V4 Pro（2026 年 4 月 24 日公開）

DeepSeek V4 Pro は、総パラメータ 1.6 兆・アクティブ 49B の MoE（Mixture-of-Experts）モデルで、100 万トークンのコンテキスト長を持ち、MIT ライセンスのオープンウェイトとして公開された [13,15]。CSA（Compressed Sparse Attention）と HCA（Heavily Compressed Attention）を組み合わせたハイブリッド注意機構により長文脈処理の効率を大幅に高めたことが技術的特徴である [13]。DeepSeek 自身の技術報告では、推論ベンチマークにおいて従来のオープンモデルを全面的に上回り、多くの指標でクローズドの最先端モデルに匹敵するとされる [13]。

ただし第三者評価はより慎重である。米国の CAISI（AI 標準・イノベーションセンター、NIST 傘下）は 2026 年 4 月の評価で、非公開ベンチマークを含む独自評価に基づき、DeepSeek V4 の実力は約 8 か月前の米国モデル（GPT-5 相当）にとどまると結論づけた。同時に、コスト効率では同等性能の米国モデルを多くのベンチマークで上回るとも認めている[14]。自己申告ベンチマークと独立評価の乖離は、中国製モデル評価全般に共通する留意点である。また、独立評価では「知らない問いにも回答してしまう」ハルシネーション傾向の高さが具体的に指摘されており、確信度の較正が重要な業務用途では考慮を要する[16]。API 価格は入力 100 万トークンあたり 0.4〜1.7 ドル程度と、米国フロンティアモデルの数分の一から十数分の一の水準にある[15]。

2.2 Qwen3.7 Max（2026 年 5 月 20 日公開）

アリババの Qwen3.7 Max は、同社が従来採ってきたオープンウェイト路線から転換し、API 限定のクローズドモデルとして提供される旗艦モデルである[10,11]。「エージェント時代の汎用基盤」を標榜し、100 万トークンコンテキスト、約 35 時間・1,158 回のツール呼び出しに及ぶ連続自律実行の実証が最大の訴求点となっている[10,11]。ベンチマークでは、GPQA Diamond 92.4（Claude Opus 4.6 Max の 91.3 を上回る）、Apex 数学推論 44.5（DeepSeek V4 Pro の 38.3 を上回る）など、コーディング・推論・長時間エージェントタスクで中国勢トップクラスの数値を公表している[11]。価格は入力 2.5 ドル／出力 7.5 ドル（100 万トークンあたり）で、Claude Opus 4.8 の概ね半分〜4 分の 1 の水準である[12]。

一方、独立集計では Artificial Analysis Intelligence Index で Opus 4.8（61.4）に対し 56.6 と差が残り、最難関のコーディングベンチマーク（SWE-Bench Pro 等）では米国最上位に届いていない[12]。Qwen シリーズの戦略的意味は旗艦モデル単体よりもエコシステムにある。Qwen 系オープンウェイトモデルは世界中の派生モデル開発の基盤として最も広く使われており、後述のとおり日本の「国産 LLM」の相当数も Qwen をベースにしている。

2.3 GLM-5.2（2026 年 6 月 13 日公開）

智譜 AI（Zhipu AI、現 Z.ai）の GLM-5.2 は、2026 年前半で最も市場インパクトの大きかった中国製モデルといえる。約 750B（報道により 743B〜753B）パラメータの MoE モデルで、100 万トークン級コンテキストを備え、MIT ライセンスで完全公開された[5,6]。長時間自律コーディングに特化した設計で、Terminal-Bench 2.1 で 81.0（Claude Opus 4.8 は 85.0）、FrontierSWE で Opus 4.8 に約 1 ポイント差まで迫り、GPT-5.5 を上回る項目も複数報告されている[6,8]。Artificial Analysis では総合 4 位・オープンウェイト 1 位となり、中国製モデルとして初めて主要グローバルベンチマークのトップ 3 圏に入った[9]。セキュリティ企業 Semgrep の独立検証では、脆弱性検出（IDOR）ベンチマーク

で Claude Code を上回る F1 スコアを記録し、「オープンウェイトモデルがもはや明白な劣位ではない」と評された[7]。

価格破壊も著しい。API 価格は入力 1.4 ドル／出力 4.4 ドル程度で Opus 4.8 の約 5 分の 1、コーディング向けサブスクリプションは米国競合の 10 分の 1 程度とされる[8,9]。シリコンバレーの著名実務家からも「日常使いに耐える初のオープンモデル」との評価が出ており、2025 年初頭の DeepSeek ショックの再来と形容されている[9]。なお同社は、GLM-5.2 が Anthropic の制限モデル Mythos とサイバーセキュリティ・脆弱性発見ベンチマークで同等と主張しているが、これは自社評価であり独立検証は部分的である点に注意が必要である[7,32]。

2.4 Kimi K3（2026 年 7 月 16 日公開）

Moonshot AI（月之暗面）の Kimi K3 は、総パラメータ 2.8 兆と、公開された重みを持つモデルとして史上最大であり、公開時点で中国製モデルの最高到達点を示した[1,2]。896 のエキスパートのうちトークンごとに 16 のみ（約 1.8%）を活性化させる高スパース MoE 構成、線形注意系の Kimi Delta Attention（KDA）と Attention Residuals という独自アーキテクチャ、100 万トークンコンテキスト、ネイティブなマルチモーダル（視覚）理解、常時オンの推論モードを特徴とする[1,3]。独立系評価では全フロンティアモデル中 4 位（Claude Fable 5、GPT-5.6 Sol に次ぐ）とされ、コーディングやエージェントタスクの一部では Claude Opus 4.8 や GPT-5.5 を上回ったと Moonshot は主張する[2,3,4]。

注目すべきは価格戦略の変化である。入力 3 ドル／出力 15 ドルと中国系ラボとして最高値を付けたが、それでも Opus 4.8 の 1 タスクあたりコストの概ね半分とされる[2]。「安さで追う」段階から「性能で並び、なお安い」段階への移行を象徴する価格設定といえる。重みの完全公開は 7 月 27 日予定とアナウンスされており[1,3]、本稿執筆時点（7 月 19 日）では未公開である。また推論トークン消費が非常に多いことが独立テスターから指摘されており、実効コストはタスク次第で変動が大きい[2]。

2.5 4 モデルの比較

項目	DeepSeek V4 Pro	Qwen3.7 Max	GLM-5.2	Kimi K3	（参考）米国勢
開発元	DeepSeek	アリババ	Z.ai（智譜 AI）	Moonshot AI	Anthropic/OpenAI
公開日	2026/4/24	2026/5/20	2026/6/13	2026/7/16	—
総パラメータ	1.6T（49B 活性）	非公開	約 750B（約 40B 活性）	2.8T（高スパース）	非公開
提供形態	オープン（MIT）	クローズド API	オープン（MIT）	オープン予定（7/27）	クローズド

コンテキスト	100 万	100 万	約 100 万	100 万	20 万〜100 万
API 価格(入/出, $/M)	約 0.4/0.9	2.5/7.5	約 1.4/4.4	3/15	Opus 4.8: 約 5/25
強み	コスト効率・長文脈	長時間エージェント	自律コーディング	総合力・規模	最高難度・信頼性

表 1　中国主要 4 モデルの概要（各種公表情報・報道に基づき筆者整理。価格・仕様は変動があり得る）

3. 「ほぼ追いついた」と評価してよいか

3.1 定量的な格差の推移（確認された事実）

まず、独立機関による定量評価を確認する。Epoch AI の分析によれば、2023 年以降、能力指数（ECI）で測ったフロンティアは一貫して米国モデルが占めてきたが、中国モデルの遅れは平均 7 か月（最小 4 か月〜最大 14 か月）で推移してきた [17]。スタンフォード大学 AI Index 2026（2026 年 4 月公表）は、Arena リーダーボード上の米中トップモデルの性能差が 2026 年 3 月時点で約 2.7%まで縮小したと報告している [18,19]。Google DeepMind のハサビス CEO も 2026 年 1 月に「中国トップモデルの遅れは数か月程度」と発言した [20]。さらに、GLM-5.2 公開後には米中の「クローズド最先端とオープン中国勢」の差を約 7 か月とする独立系論者の評価があり [9]、Kimi K3 は公開直後の独立集計で総合 4 位、すなわち米国最上位 2 モデルのみが上位という位置に達した [2]。

他方、慎重な評価も存在する。前述の CAISI 評価は DeepSeek V4 につき約 8 か月の遅れと結論しており [14]、AI Index 2026 も、性能差縮小の主因がオープンウェイト戦略にあること、クローズド最上位層では米国優位が残ることを併せて指摘している [18,19]。つまり「差は 2.7%」と「差は 7〜8 か月」は矛盾ではなく、前者はチャットボット的総合評価（Arena Elo）、後者は最高難度タスクを含む能力全体の時間換算という、測り方の違いを反映している。

3.2 評価：どの面で追いつき、どの面で差が残るか（分析）

以下は確認された事実に基づく筆者の分析である。「ほぼ追いついた」という命題は、次のように分解して評価すべきである。

- 追いついたと言える領域：①コーディング・エージェントタスクの実用性能（GLM-5.2、Kimi K3 が Opus 級と互角の項目多数）、②コスト効率（同等性能を数分の一の価格で提供）、③コンテキスト長等の仕様面（100 万トークンが標準化）、④オープンウェイト分野（世界の上位オープンモデルはほぼ中国製が独占）。
- 差が残る領域：①最高難度ベンチマーク（Humanity's Last Exam、SWE-Bench Pro 等では米国最上位が依然優位 [8,12]）、②独立評価と自己申告の乖離（CAISI 評価が示すとおり [14]）、③ハル

シネーション抑制・確信度較正などの信頼性[16]、④安全性評価・レッドチーミングの透明性、⑤最先端クローズドモデル（Fable 5 級）との絶対性能差。

総合すれば、「用途の 8〜9 割を占める実務的タスクではほぼ追いついた。ただし最先端の絶対性能と信頼性の層では、なお半年前後の差が残る」というのが、2026 年 7 月時点で最も証拠と整合的な評価と考える。数か月前まで「数か月〜1 年の遅れ」とされていたことを踏まえると、差の縮小速度自体が加速していることは明確である。

4. 追い越すと考えられるか

4.1 中国側の追い風

- 研究・人材基盤：AI Index 2026 によれば、中国は AI 論文数・特許出願数・人材供給で既に米国を上回り、投資額が米国の 23 分の 1 でありながら性能差を 2.7%まで詰めた[18,19]。
- アーキテクチャ革新の内製化：KDA、Attention Residuals（Kimi K3）[1,3]、IndexShare（GLM-5.2）[5]、CSA/HCA ハイブリッド注意（DeepSeek V4）[13] など、米国の後追いではない独自のアーキテクチャ改良が成果を上げ始めている。計算資源制約下でも事前学習スケーリングと構造革新の組合せで段階的性能向上が可能なことを K3 が実証したとの評価がある[33]。
- オープンウェイト戦略による世界普及：MIT ライセンス公開は、一度配布された重みを政府命令で停止できないという意味で「地政学的に止められない AI」であり、米国の Fable 5 輸出規制騒動（後述）はこの訴求力を皮肉にも強化した[6]。
- 収益基盤の確立：Moonshot AI の年間経常収益は 2 億ドル超に達し[2,4]、「安売り一辺倒」から持続可能な事業モデルへの移行が始まっている。

4.2 米国側の残存優位と中国側の制約

- 計算資源：先端 GPU・製造装置への輸出規制は 3 年間で中国のフロンティア到達を阻止できなかったが[2]、次世代の大規模学習では計算資源の量的格差が再び効いてくる可能性がある。
- クローズド最先端の厚み：AI Index 2026 は、性能向上の主戦場がクローズド上位層に移りつつあり、オープン戦略による追い上げが逓減局面に入る可能性を指摘する[18]。
- 蒸留規制の動き：米国では中国側による米国モデル出力の「蒸留」を制限する立法論が進んでおり[4]、実効性は不明だが、キャッチアップ経路の一つが狭まる可能性はある。

4.3 シナリオ評価（分析）

以上を踏まえると、「追い越すか」への答えは領域別に分かれる。第一に、オープンウェイト分野

では中国は既に事実上追い越しており、この構図が短期に覆る材料は見当たらない。第二に、コスト性能比（性能あたり価格）でも中国優位は確立している。第三に、絶対性能の最先端（クローズド最上位）では、今後 1〜2 年は米国優位の継続が相対的に確からしいが、確実とは言えない。予測市場では「年末時点で中国企業が AI 首位」の確率が年初の数%から 14%まで上昇しており [21]、市場参加者も「追い越し」をテールリスクではなく現実的シナリオとして織り込み始めている。なお、これらは本質的に不確実性の高い予測であり、次の大型モデル世代（GPT-6 級、Claude 次世代、DeepSeek V5 等）の出来如何で評価は数か月単位で変わり得ることを明記しておく。

5. 日本企業は中国製 LLM をどう使うべきか

5.1 前提となるリスク構造の整理

中国製 LLM のリスクを論じる際、最も重要なのは「クラウド API・アプリ利用」と「オープンウェイトのセルフホスト利用」を峻別することである。両者はリスク構造が根本的に異なる。

クラウド利用の場合、入力データが中国国内サーバーに保存され得る。中国の国家情報法（2017 年）第 7 条はすべての組織・個人に国家情報活動への協力義務を課し、データ安全法（2021 年）・サイバーセキュリティ法と併せ、当局のデータ提供要求を企業が法的に拒否し難い構造を作っている [22,23]。日本の個人情報保護委員会も、DeepSeek 等の利用に関し中国の法制度を考慮した注意喚起を行っている [24]。重要なのは、「実際にデータが渡った証拠があるか」ではなく「法的に要求可能な仕組みが存在すること自体」がリスクだという点である [22]。ジェイルブレイク耐性の脆弱性を指摘する検証も複数報告されている [23]。

これに対しセルフホスト利用では、MIT ライセンス等で公開された重みを自社のオンプレミスや国内クラウドの閉域環境にダウンロードして運用するため、データが中国側に送信される経路は原理的に存在しない [25]。重みは検証可能であり、ライセンス上、改変・商用利用も自由である。この場合の残存リスクは、①モデル出力の品質・バイアス（学習データに由来する価値観の偏り、中国当局にとって敏感な話題での応答傾向）、②サプライチェーンの継続性（後継モデルの公開が止まる可能性）、③将来の規制リスク（米国が同盟国に中国製モデル利用制限を求める可能性）などに限定される。

5.2 米国製モデル依存のリスクも顕在化した

2026 年 6 月の「Fable 5 事件」は、逆方向のリスクを世界の企業に突きつけた。米商務省は 6 月 12〜13 日、国家安全保障上の懸念（ジェイルブレイク報告）を理由に、輸出管理規則（EAR）に基づき Anthropic に対し Claude Fable 5 と Mythos 5 への外国籍者のアクセス全面停止を命令し、両モデ

ルは数時間で全世界のユーザーから利用不能となった[26,27]。規制は 6 月 30 日に解除されたが[28]、商用 AI モデルが政府命令により予告なく停止され得ることが史上初めて実証された。日本企業を含む非米国ユーザーにとって、最先端の米国製クローズドモデルへの依存は、それ自体が事業継続リスクを内包することが明らかになった。Z.ai が GLM-5.2 の完全公開をこの命令の翌日に発表したのは偶然ではなく、「ダウンロード済みの重みはいかなる政府も止められない」という対抗訴求である[6]。すなわち日本企業のモデル戦略は、「中国リスク対米国の安心」ではなく、「中国法域リスク対米国規制リスク」という二つの地政学リスクの管理問題として設計する必要がある。

5.3 日本企業の活用実例

実際、日本では中国製オープンウェイトモデルの活用が既に相当程度進んでいる。みずほフィナンシャルグループは、Qwen3-32B をベースに金融ドメインでファインチューニングした自社 LLM を内製し、オンプレミス運用で GPT-5.2 同等の精度を実現したと報じられている[29]。ソフトバンク系 SB Intuitions の Sarashina2 シリーズは Qwen2.5 をベースに日本語を強化したものであり[30]、国内の「国産 LLM」の相当数が Qwen 系を基盤としている。他方、国立情報学研究所（NII）の LLM-jp-4 のように、約 12 兆トークンの独自コーパスからフルスクラッチで学習し、一部ベンチマークで GPT-4o や Qwen3-8B を上回る性能を達成した純国産の取組みも進んでいる[31]。すなわち日本の実務では、「中国製モデルの重みを国内環境で使う」ことと「データ主権を確保する」ことは既に両立されており、問題は個社がこれを統制された形で行えるかに移っている。

5.4 提言：用途×形態のマトリクスで統制する

以上を踏まえ、日本企業への提言を示す。基本原則は「性能・コストの果実は取りに行く。ただし形態と用途を統制する」である。

- 原則 1：クラウド API・公式アプリ経由での中国製 LLM 利用は、機密情報・個人情報・営業秘密を扱う業務では原則禁止とする。個人情報保護委員会の注意喚起[24]および国家情報法等の法域リスク[22]を踏まえれば、これは最低限の統制である。公開情報のみを扱う調査・翻訳等の低リスク用途に限定する場合も、シャドーAI（従業員の無断利用）対策として利用ポリシーの明文化とログ監視を伴うべきである[23]。
- 原則 2：セルフホスト（オンプレミス・国内閉域クラウド）でのオープンウェイト利用は、積極的に検討に値する。GLM-5.2 や DeepSeek V4 級の性能をデータを一切外部に出さずに利用できることは、米国製クローズド API に対する明確な優位である。コーディング支援、社内文書 RAG、ドメイン特化ファインチューニングが有望な入口となる。ただし出力品質の独自評価（特にハルシネーションと敏感話題の応答傾向）を導入前に実施すること。

- 原則 3：マルチモデル体制を標準とする。Fable 5 事件が示すとおり [26,27]、単一ベンダー・単一法域への依存はそれ自体がリスクである。最高難度タスクは米国製最上位、大量・定型・コスト重視タスクは中国製オープンウェイトまたは国産モデル、という使い分けと、モデル切替えを前提としたアプリケーション設計（抽象化レイヤの導入）を推奨する。
- 原則 4：知財・コンプライアンス部門が関与するガバナンスを敷く。学習データの権利処理の不透明性は中国製モデルに限らないが、生成物の利用に伴う著作権・営業秘密リスクの評価、モデルライセンス（MIT、Apache 2.0 等）の条件確認、輸出管理（将来の対中・対米規制双方）のモニタリングは、法務・知財部門の継続的関与を要する。
- 原則 5：規制動向の定点観測を組み込む。米国の対中モデル規制・蒸留規制の立法動向 [4]、日本政府のガイドライン改定、中国側のデータ関連法運用は、いずれも半年単位で状況が変わる。モデル選定の前提条件として四半期ごとの見直しを推奨する。

6. おわりに

2026 年前半の 3 か月間に起きたことは、単なる「中国の追い上げ」ではなく、AI 市場の構造変化である。性能の頂点は依然米国にあるが、実務に必要な性能の大半は、無償でダウンロードできる中国製の重みで手に入るようになった。同時に、米国製モデルすら政府命令で止まり得ることが実証された。日本企業に求められるのは、いずれかの陣営への賭けではなく、両方の地政学リスクを直視した上で、オープンウェイトの果実を統制された形で取り込む実務体制の構築である。その巧拙が、今後 2〜3 年の AI 活用コスト競争力を大きく左右するだろう。

なお本稿の事実関係は 2026 年 7 月 19 日時点の公開情報に基づく。Kimi K3 の重み公開（7 月 27 日予定）や各社の次期モデルにより状況は短期間で変わり得る。ベンチマーク数値には開発元自己申告のものが含まれ、独立検証と乖離する例があることは本文で述べたとおりである。

参考文献

1. CodersEra「Kimi K3: Moonshot AI's 2.8T Open-Weight Model — Release, Specs & Pricing」2026 年 7 月 17 日. https://codersera.com/blog/kimi-k3-complete-guide-2026/
2. MLQ News「Moonshot AI Releases Kimi K3, a 2.8-Trillion-Parameter Open-Weight Model Rivaling Top U.S. Systems」2026 年 7 月. https://mlq.ai/news/moonshot-ai-releases-kimi-k3-a-28-trillion-parameter-open-weight-model-rivaling-top-us-systems/
3. VentureBeat「China's Moonshot AI releases Kimi K3, the largest open-source model ever, rivaling top U.S. systems」2026 年 7 月 16 日. https://venturebeat.com/technology/chinas-moonshot-ai-releases-kimi-k3-the-largest-open-source-model-ever-rivaling-top-us-systems
4. Fortune「Moonshot's Kimi K3 pushes Chinese AI into Fable-level territory」2026 年 7 月 16 日.

https://fortune.com/2026/07/16/moonshots-kimi-k3-pushes-chinese-ai-into-fable-level-territory/

5. VentureBeat「Z.ai's open-weights GLM-5.2 beats GPT-5.5 on multiple long-horizon coding benchmarks for 1/6th the cost」2026 年 6 月 16 日. https://venturebeat.com/technology/z-ais-open-weights-glm-5-2-beats-gpt-5-5-on-multiple-long-horizon-coding-benchmarks-for-1-6th-the-cost
6. Tech Times「GLM-5.2 Open Weights Live: Top Coding Benchmark, but API Use Carries China Data Risk」2026 年 6 月 17 日. https://www.techtimes.com/articles/318543/20260617/glm-52-open-weights-live-top-coding-benchmark-api-use-carries-china-data-risk.htm
7. Semgrep「We have Mythos at Home: GLM 5.2 beats Claude in our Cyber Benchmarks」2026 年 6 月. https://semgrep.dev/blog/2026/we-have-mythos-at-home-glm-52-beats-claude-in-our-cyber-benchmarks/
8. Technology.org「Zhipu's GLM 5.2 Rivals Opus 4.8 on Coding Benchmarks at a Fifth of the Cost」2026 年 7 月 2 日. https://www.technology.org/2026/07/02/zhipus-glm-5-2-rivals-opus-4-8-on-coding-benchmarks-at-a-fifth-of-the-cost/
9. Kie.ai「GLM-5.2 Benchmark Deep Dive: Open-Weight Frontier」2026 年 7 月. https://kie.ai/blog/glm-5-2-benchmark-deep-dive
10. VentureBeat「Alibaba's proprietary Qwen3.7-Max can run for 35 hours autonomously and supports external harnesses」2026 年 5 月 21 日. https://venturebeat.com/technology/alibabas-proprietary-qwen3-7-max-can-run-for-35-hours-autonomously-and-supports-external-harnesses-like-anthropics-claude-code
11. DataCamp「Qwen3.7-Max: Features, Benchmarks and Agent Capabilities」2026 年 5 月 22 日. https://www.datacamp.com/blog/qwen3-7-max
12. FrankX「Qwen3.7-Max: Alibaba's Agent Flagship Cracks the Global Top 5 — and Runs for 35 Hours」2026 年 6 月 5 日. https://www.frankx.ai/blog/qwen3-max-analysis-2026
13. DeepSeek-AI「DeepSeek-V4-Pro」（Hugging Face モデルカードおよび技術報告）2026 年 4 月. https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro
14. NIST/CAISI「CAISI Evaluation of DeepSeek V4 Pro」2026 年 5 月. https://www.nist.gov/news-events/news/2026/05/caisi-evaluation-deepseek-v4-pro
15. OpenRouter「DeepSeek V4 Pro - API Pricing & Benchmarks」2026 年 7 月時点. https://openrouter.ai/deepseek/deepseek-v4-pro
16. DeepInfra「DeepSeek V4 Pro: Model Overview, Features & Performance Guide」2026 年 4 月 30 日. https://deepinfra.com/blog/deepseek-v4-pro-model-overview
17. Epoch AI「Chinese AI models have lagged the US frontier by 7 months on average since 2023」2026 年 1 月 2 日. https://epoch.ai/data-insights/us-vs-china-eci
18. The Next Web「Stanford AI Index 2026: China narrows US lead to 2.7% while spending 23x less on AI investment」2026 年 5 月. https://thenextweb.com/news/stanford-ai-index-2026-china-us-performance-gap
19. China Biz Insider「Stanford 2026 AI Index: China Closes Gap With US Models」2026 年 6 月. https://chinabizinsider.com/stanfords-2026-ai-index-says-chinas-top-models-are-closing-the-gap-with-the-us/
20. VERTU「Is China Catching the US in AI? Hassabis Says Gap is "Months"」2026 年 1 月 19 日. https://vertu.com/lifestyle/the-global-ai-race-why-china-is-just-months-behind-the-us-according-to-deepminds-ceo/
21. ExplainX「Why Did the US Gov Ban Fable 5? The Full Anthropic Story」2026 年 6 月（6 月 27 日更新）.

https://www.explainx.ai/blog/us-government-bans-fable-5-mythos-5-anthropic-export-control-2026

22. Crystal Method「DeepSeek 危険性の構造と対策｜エンジニアが判断すべきリスク 4 層」2026 年 6 月 13 日. https://crystal-method.com/blog/deepseek-risk/
23. LANSCOPE「【重要】DeepSeek の安全性とセキュリティリスクを徹底解説」2026 年 3 月更新. https://www.lanscope.jp/blogs/it_asset_management_emcloud_blog/20250328_26112/
24. 個人情報保護委員会「DeepSeek 等の生成 AI の利用に関する注意喚起」（二次資料経由で確認。原典 URL は要確認）. https://genaiabc.com/deepseek-suspicious/
25. 株式会社アタリ「"DeepSeek は危険"という誤解 〜LLM をローカルで使う選択肢〜」2025 年 3 月. https://note.com/atali/n/n6018a70060b2
26. Cloud Security Alliance Lab Space「Fable 5 Suspension: Enterprise AI Under Export Controls」2026 年 6 月 14 日. https://labs.cloudsecurityalliance.org/research/csa-research-note-ai-model-export-controls-enterprise-govern/
27. Anthropic「Statement on the US government directive to suspend access to Fable 5 and Mythos 5」2026 年 6 月 12 日. https://www.anthropic.com/news/fable-mythos-access
28. CNBC「Anthropic says Trump admin has lifted export controls on Claude Fable 5 and Mythos 5」2026 年 6 月 30 日. https://www.cnbc.com/2026/06/30/anthropic-says-trump-admin-has-lifted-export-controls-on-claude-fable-5-and-mythos-5.html
29. AI/DX Media「みずほ FG が自社 LLM を Qwen3-32B で内製化 - GPT-5.2 同等精度でオンプレ運用を実現」2026 年 3 月 6 日. https://media.image-pit.com/articles/japan/2026-03-06-mizuho-fg-llm-qwen3-32b-gpt52-accuracy-on-premise
30. Crystal Method「国産 LLM 比較｜2026 年版ガイド」2026 年 7 月. https://crystal-method.com/blog/llm-comparison-domestic/
31. 国立情報学研究所「約 12 兆トークンの良質なコーパスで学習した新たな国産 LLM『LLM-jp-4』をオープンソースライセンスで公開」2026 年 4 月 3 日. https://www.nii.ac.jp/news/release/2026/0403.html
32. Creati.ai「China's Zhipu Z.AI Claims GLM-5.2 Matches Anthropic Mythos on Cybersecurity Benchmarks」2026 年 6 月 29 日. https://creati.ai/ai-news/2026-06-29/chinas-zhipu-z-ai-glm-52-matches-anthropic-mythos-cybersecurity-benchmarks/
33. CNBC「China's Moonshot AI unveils Kimi K3 that rivals OpenAI, Anthropic」2026 年 7 月 17 日. https://www.cnbc.com/2026/07/17/moonshot-ai-kimi-k3-model-openai-anthropic-china.html