DeepSeek V4 Pro の実力と 8 か月遅れ評価

ChatGPT-5.5

エグゼクティブサマリー

- CAISI/NIST の 2026 年 5 月 1 日評価がいう「DeepSeek V4 は最先端に約 8 か月遅れ」は、CAISI **が事前固定した評価スイート**を、5 **領域を等重み**で集約した 1PL-IRT **ベースの総合能力推定**に基づく表現である。CAISI は DeepSeek V4 を「これまで評価した中国モデルで最も高性能」と位置づけつつ、総合力では GPT-5 **相当**で、GPT-5.4 **や** Claude Opus 4.6 **相当ではない**と結論づけている。[1]

- この評価を押し下げている主因は、**抽象推論・サイバー・ソフトウェア工学**の一部、とくに ARC-AGI-2 semi-private、PortBench、CTF-Archive-Diamond での差である。逆に、**自然科学と数学はかなり接近**しており、OTIS-AIME-2025 では DeepSeek V4 が Claude Opus 4.6 を上回る。したがって「8 か月遅れ」は**全領域一様の遅れ**ではなく、CAISI の集約方法で強く重みが乗る弱点領域の影響が大きい。[2]

- DeepSeek 自身の公式資料はかなり異なる絵を描く。V4-Pro は**総** 1.6T/**有効** 49B、V4-Flash は**総** 284B/**有効** 13B、両者とも 1M **トークン文脈**、MIT **ライセンスのオープンウェイト**、32T **超の事前学習トークン**、CSA/HCA・mHC・Muon を中核とする設計で、推論・長文脈・エージェント用途を強く打ち出す。一方で、**データ内訳比率、総学習計算量、詳細な安全性評価、必要ハード要件**は公開資料で十分に特定されていない。[3]

- 第三者評価は、CAISI に完全には一致しないが、完全にも反しない。Artificial Analysis は V4 Pro を open-weight **で上位**、agentic 実務評価の GDPval-AA **で首位級**とみる一方、**「知らないときにも答えに行く」率が非常に高い**とする。LiveBench では DeepSeek V4 Pro は**公開・客観採点系ベンチで最前線に肉薄**するが、LM Arena の人間選好では**中上位止まり**で、閉

鎖型最前線ほどの支持は得ていない。要するに、**公開ベンチでは強く、held-out/実運用一般化ではなお差が残る**というのが、もっとも整合的な読みである。[4]

- 実務的には、DeepSeek V4 は**「安価・長文脈・オープンウェイト・強いコーディング/エージェント性能」**という大きな便益を持つが、**サイバー、未知推論、コード移植、** hallucination**、データガバナンス**では慎重な導入が必要である。開発者と導入企業は、DeepSeek の自己評価や単一リーダーボードだけでなく、**自社の** held-out **評価**を必須にすべきである。[5]

CAISI/NIST 評価の読み解き

CAISI の 2026 年 5 月 1 日評価は、2026 年 4 月に DeepSeek V4 Pro を評価したもので、結論は明快だ。DeepSeek V4 は CAISI **がこれまで評価した中国モデルの中で最も高性能**だが、**米国フロンティアの能力に約 8 か月遅れる**。この「8 か月」は、単一ベンチマークの勝敗ではなく、16 **ベンチマーク・35 モデル**を用いた時系列図の上で、IRT **に着想を得た能力推定**と**フロンティアモデルの最小二乗回帰**から読まれた差である。CAISI は同時に、「DeepSeek の自己報告では Opus 4.6 や GPT-5.4 に近いが、CAISI の評価では GPT-5 に近い」と明記している。[2]

方法論の肝は三つある。第一に、CAISI は**サイバー、ソフトウェア工学、自然科学、抽象推論、数学**の 5 領域を使い、**各領域内でベンチマークを等重み、さらに領域間も等重み**で集約している。第二に、評価には**公開ベンチだけでなく、**held-out/contamination-resistant **寄りのデータ**が含まれる。具体的には、ARC-AGI-2 **の** semi-private **セット**と、CAISI 内製の PortBench が中核である。第三に、実行条件も揃えており、DeepSeek V4 は**クラウド上の** H200/B200 で、**開発元推奨設定、内部推論の保持、最大思考**を維持しつつ、**重み付きトークン予算**と agent scaffolding を統制している。GPQA-Diamond では開発元自己報告の再現も行い、構成ミスではないことを確認している。[2]

ベンチマークごとの中身を見ると、「何が遅れているのか」がかなり具体的に見える。最も大きい差は**抽象推論**と held-out **ソフトウェア工学**だ。ARC-AGI-2 semi-

private では DeepSeek V4 は 46%で、Claude Opus 4.6 の 63%、GPT-5.5 の 79%に届かない。PortBench でも DeepSeek V4 は 44%で、Opus 4.6 の 60%、GPT-5.5 の 78%を大きく下回る。サイバーの CTF-Archive-Diamond でも DeepSeek V4 は 32%で、Opus 4.6 の 46%に劣る。一方、自然科学の FrontierScience は 74%で GPT-5.4 mini と並び、GPQA-Diamond も 90%で Opus 4.6 の 91%に近い。数学では OTIS-AIME-2025 が 97%、PUMaC 2024 が 96%、SMT2025 が 96%で、領域によっては米国勢にかなり迫っている。[2]

このため、CAISI の「約 8 か月遅れ」は、**数学や自然科学まで含めた“全体平均”で見てもなお、抽象推論・サイバー・held-out ソフトウェア工学の弱さが残る**、という意味に読むのが正確だ。CAISI の IRT-Elo では DeepSeek V4 は 800±28、GPT-5.4 mini は 749±46、Claude Opus 4.6 は 999±27、GPT-5.5 は 1260±28 で、CAISI 自身が「V4 は GPT-5 に近い」と整理する根拠もここにある。CAISI は 200 Elo **ごとに、ある課題を解けるオッズが 3 倍**になると説明しているため、V4 と Opus 4.6 の差は、CAISI の尺度ではかなり実質的である。[2]

この評価の限界も明示しておく必要がある。PortBench は**非公開**で、CAISI は今後詳細を公開予定としているが、2026 年 5 月 1 日時点では独立再現が難しい。ARC-AGI-2 についても、CAISI は**公式集約法と異なる平均スコア**を用いている。SWE-Bench Verified の点数は、CAISI 自身が system prompt、scaffolding、token budget **の違いで他評価者より低く出やすい**と注記している。CTF-Archive-Diamond の DeepSeek スコアは**サンプルの部分集合から IRT で補完**された値だ。さらに、CAISI はより詳細な方法論文書を**将来公開予定**としており、現時点では「8 か月」の完全再現には情報がまだ足りない。[2]

なお、この 5 月 1 日ノートは**能力とコスト**に焦点があり、2025 年の CAISI による DeepSeek 旧モデル評価で大きく扱われた**セキュリティ、脱獄耐性、検閲、採用率**のような軸は、V4 ノートでは新規に詳細評価されていない。したがって、V4 に関する CAISI の最新見解は、**「総合能力差の測定」と「類似能力モデルに対するコスト」**に主眼があると理解すべきだ。[6]

DeepSeek 公式情報の整理

DeepSeek の公式リリースでは、V4 系は V4-Pro と V4-Flash の 2 本立てで、いずれも 1M トークン文脈をサポートする。Hugging Face の技術報告では、V4-Pro は総 1.6T・有効 49B、V4-Flash は総 284B・有効 13B の Mixture-of-Experts で、公式 API は Thinking / Non-Thinking の両モードを持ち、JSON Output、Tool Calls、Chat Prefix Completion をサポートする。API は OpenAI 形式と Anthropic 形式の双方に対応し、最大出力 384K である。公式の料金表では、V4-Flash は入力$0.14 / 出力$0.28、V4-Pro は公称で入力$1.74 / 出力$3.48 だが、2026 年 4 月 26 日以降はキャッシュ価格が 1/10 に下がり、V4-Pro は 5 月末まで 75%割引として入力 $0.435 / 出力$0.87 が提示されている。CAISI のコスト比較はこの割引後価格ではなく、公称単価を用いている点に注意が必要だ。[7]

アーキテクチャ面では、DeepSeek は V4 を「超長文脈の効率化」に全力投球した世代として位置づけている。技術報告の中核は、Compressed Sparse Attention と Heavily Compressed Attention のハイブリッド、Manifold-Constrained Hyper-Connections、そして Muon optimizer だ。DeepSeekMoE と Multi-Token Prediction は V3 系列から継承しつつ、長文脈での計算量と KV キャッシュを大幅に削減し、1M 文脈で V3.2 比 27%の single-token FLOPs、10%の KV cache に抑えたとする。公式には、V4-Pro-Max は公開ベンチでオープンモデルの SOTA を再定義し、推論面では GPT-5.2 や Gemini 3.0 Pro を上回るが、GPT-5.4 と Gemini 3.1 Pro にはわずかに届かないと説明している。[8]

学習面では、DeepSeek は V4-Flash を 32T、V4-Pro を 33T トークンで事前学習した。データ構築については、V3 の土台に加え、より長い有効文脈、数学・コードの強化、mid-training での agentic data 導入、多言語の long-tail 知識、長文書データの重視を挙げている。モデルカードは、学習データの大分類を public data と licensed data とし、個人情報は意図的に収集せず、検出・除去を行うと述べる。最適化学習段階では、質問応答データの一部にユーザー入力が使われうるが、その場合は暗号化・de-identification・anonymization を行うとしている。ただし、

カテゴリ別の比率、ライセンスデータ比率、学習計算量、GPU/チップ構成、事後学習データ規模は公開資料で未指定だ。[9]

公開資料上のライセンスとガバナンスは、DeepSeek の価値提案とリスクの両方を形づくる。オープンソース配布の**重みとコードは** MIT、API 経由の利用は Open Platform Terms of Service に従う。利用規約は、違法・権利侵害行為の禁止、出力が**誤り・欠落・不快内容**を含みうること、**専門助言ではない**こと、AI 生成である旨の**明示的な免責表示**を付すことを定める。さらにプライバシーポリシーでは、入力やログ等の個人データを収集しうること、これらが**中国国内で処理・保管**されうることを明示している。重要なのは、V4 の公開資料が**モデルカード、技術報告、利用規約、プライバシーポリシー**中心であり、少なくとも 2026 年 5 月 1 日時点で、いくつかの米フロンティアラボが公表するような**詳細な** system card / risk report **相当の** V4 **文書**は確認しにくい点である。これは、生の能力とは別に、導入判断で効く非技術的要因だ。[10]

比較対象としては、OpenAI[11]の GPT-5/GPT-5.4、Anthropic[12]の Claude Opus 4.6、Google[13]の Gemini 3.1 Pro が主要な基準になる。以下のタイムラインは、DeepSeek の主要リリースと競合の節目を並べたものだ。日付は各社の公式リリースページに基づく。[14]

```
timeline
    title DeepSeek 主要リリースと競合モデルの時系列
    section DeepSeek
      2024-12-26 : DeepSeek-V3
      2025-01-20 : DeepSeek-R1
      2025-03-25 : DeepSeek-V3-0324
      2025-05-28 : DeepSeek-R1-0528
      2025-08-21 : DeepSeek-V3.1
      2025-09-22 : DeepSeek-V3.1-Terminus
      2025-09-29 : DeepSeek-V3.2-Exp
      2025-12-01 : DeepSeek-V3.2
      2026-04-24 : DeepSeek-V4 Preview
    section Frontier competitors
      2025-08-07 : GPT-5
      2026-02-05 : Claude Opus 4.6
      2026-02-19 : Gemini 3.1 Pro
      2026-03-05 : GPT-5.4
```

他の第三者評価との比較

2026 年 5 月 1 日までに確認できる独立検証は、**個別の査読済み V4 論文**よりも、**独立ラボの速報、継続更新型リーダーボード、主要メディアの技術報道**が中心だった。すなわち、V4 の評価は「学術的に完全に収束した定説」ではなく、**異なる測定手法が同時並行で走っている状態**にある。[15]

出典	日付	方法論	主結論	代表値
DeepSeek 技術報告 [16]	2026-04-24	公開ベンチと内部評価の**自己報告**	オープンモデル最上位。推論では最先端閉鎖モデルに**約 3–6 か月遅れ**と自己推定	LiveCodeBench 93.5、Codeforces 3206、HLE 37.7
CAISI/NIST [2]	2026-05-01	9 **ベンチ**/5 **領域**、held-out 含む、1PL-IRT **集約**、時系列回帰	中国モデル最強だが、総合力は**約 8 か月遅れ**	IRT Elo 800 ± 28
Artificial Analysis [17]	2026-04-24	独自の Intelligence Index、GDPval-AA、AA-Omniscience 等	open-weight で #2 **級**、agentic 実務で強いが hallucination **率が高い**	Intelligence 52、GDPval-AA 1554、hallucination 94%
LiveBench リーダーボード / 論文 [18]	2026-05-01 snapshot	毎月更新の contamination-limited 公開問題、**客観採点**	公開ベンチでは最前線クラスに肉薄	Overall 73.58
LM Arena Text / Code と方法論論	2026-05-01 snapshot	**盲検** pairwise human preference、Elo	会話選好と Web 開発選好では中上位だが最前線	Text 1463±9 / rank 25、Code 1455 / rank 16

出典	日付	方法論	主結論	代表値
文 [19]			ではない	

この表から見える一致点は、V4 **がオープンウェイト最上位群に復帰したこと**、そして**閉鎖型最前線を全面的には超えていないこと**だ。CAISI、Artificial Analysis、Reuters、Fortune はこの点で概ね一致する。Reuters は V4 を**長文・複雑テキストに強く、競合より安価だが、画像や動画のようなマルチモーダルには非対応**と整理し、Fortune は DeepSeek の技術報告を引きつつ、V4 が「GPT-5.4 / Gemini 3.1 Pro **にわずかに届かず、約 3–6 か月遅れ**」という自己評価だと伝えている。[20]

一方で、ズレの理由も明確だ。CAISI は held-out/非公開を含む評価で、しかも**領域等重み**なので、ARC-AGI-2 や PortBench の弱さが総合点に効く。LiveBench は contamination-limited ではあるが、あくまで**公開・客観採点系の継続更新ベンチ**であり、CAISI の PortBench や CTF のような非公開・運用寄りタスクは含まない。LM Arena はさらに別物で、ユーザーが**どちらの応答を好むか**を測る。つまり、LiveBench の高さと Arena のほどほど感は、CAISI と矛盾しているというより、**測っている能力の切り口が違う**と見るべきだ。[21]

8 か月遅れの定量解釈

「8 か月」を最も素直に読むなら、CAISI 自身が示すように、DeepSeek V4 の能力が GPT-5 **相当**であり、GPT-5 の公開が 2025 **年** 8 **月** 7 **日**、CAISI 評価の公開が 2026 **年** 5 **月** 1 **日**であることから、**約** 8.8 **か月前の米フロンティア水準**だ、という解釈になる。CAISI も図 3 の本文で「DeepSeek V4 は GPT-5 に近い」と書いており、これが“8 か月”の最も具体的なアンカーである。[1]

ただし、CAISI は月数を**単純な発売日差だけで定義した**とは書いていない。より厳密には、16 **ベンチ・**35 **モデル**の時系列データを 1PL-IRT で能力尺度に落とし、そこから U.S. frontier line と PRC frontier line の位置関係を見て、「U.S. capability frontier tends to lead the PRC frontier by roughly 8 months」としている。したがって「8 か月」は、**個別モデル A と B の単純比較**というより、**米中フロンティア系列の相対位置**であり、しかも CAISI の 5 領域重み付けの下での話だ。[2]

ここで DeepSeek 公式の**「約 3–6 か月遅れ」**と衝突して見えるが、実際には**測定器が違う**。DeepSeek 側は、公開ベンチを中心に V4-Pro-Max が GPT-5.2 **や** Gemini 3.0 Pro **を上回り、**GPT-5.4 **や** Gemini 3.1 Pro **にわずかに届かない**ことから、発展軌道として 3–6 か月遅れだと叙述する。しかし、その月数を導く**明示的な統計式は公開されていない**。つまり、DeepSeek の 3–6 か月は**技術報告上のナラティブ推定**、CAISI の 8 か月は**事前固定スイートの IRT 集約と時系列回帰**という違いがある。[22]

したがって、最も妥当な総括はこうなる。**「約 8 か月遅れ」は、CAISI の集約尺度では妥当だが、公開ベンチ・長文脈・一部コーディングではもっと差が小さい**。逆に言えば、V4 の能力差は 1 **本の数字で固定できるほど一様ではない**。ユーザーが重視するタスクが**数学・長文脈・公開コード問題**なら 3–6 か月寄りに見え、**未知推論・コード移植・サイバー演習**なら 8 か月、あるいはそれ以上に見える可能性がある。これは推論だが、公開されている両者の評価結果と整合的である。[23]

遅れの原因分析

以下は、公開資料からの**推論**を含む原因分析である。まず技術的には、DeepSeek V4 が明らかに**長文脈効率、**agentic coding**、価格性能**に開発資源を集中させていることが重要だ。技術報告は、CSA/HCA、mHC、Muon、FP4 QAT、sandboxed agent infrastructure、mid-training での agentic data 導入を前面に出している。これは V4 の差別化には成功しているが、CAISI が重視した PortBench、ARC-AGI-2、CTF のような**未知一般化や運用寄り能力**で、最終的な押し上げがまだ足りていない可能性がある。すなわち、V4 は「何でも少しずつ最高」ではなく、**長文脈とエージェントに最適化されたフロンティア追撃機**だという読みである。[24]

次に、**公開ベンチへの最適化**と held-out **一般化**の差である。CAISI 自身が、DeepSeek の自己報告では Opus 4.6 や GPT-5.4 に近く見える一方、CAISI の非公開ベンチでは GPT-5 相当に見えると明言している。これは、公開ベンチでの性能が悪いという意味ではない。むしろ逆で、LiveBench や DeepSeek 自身の技術報告が

示すように、V4 は公開・半公開系の課題では非常に強い。ただ、**それがそのまま“未露出タスク”に移らない**ことが、今回の最大の示唆である。[25]

三つ目は、計算資源と供給網だ。Reuters は、V4 が**華為技術向けに最適化**され、**一部の学習過程にも** Ascend **が使われた**と報じている。Reuters は同時に、Huawei がなお Nvidia に技術的に遅れ、DeepSeek の Pro 価格と可用性が**高性能計算資源の制約**で抑えられており、Ascend 950 supernode **が量産展開される下期には** Pro **価格が大きく下がる見込み**だと伝えている。これは、能力差だけでなく、**反復学習・実験回転数・提供コスト**そのものが供給制約の影響を受けていることを示唆する。[26]

さらに非技術面では、DeepSeek の V4 が MIT **のオープンウェイト**であること、API が OpenAI/Anthropic **互換**であること、しかも**大幅値引き**で攻勢をかけていることが、事業上の優先順位を物語る。DeepSeek は「最高性能の一点突破」だけでなく、**配布容易性、移植性、エコシステム拡張、採用の速さ**を強く狙っている。Reuters も、V4 が Hugging Face で急速にトレンド上位へ上がったことを伝えている。これはエコシステムにとっては利点だが、企業導入の観点では、**安全文書の薄さ、プライバシー/データ所在、**text-only **制約**が足かせになりうる。[27]

最後に、公開資料から見える安全・信頼性の弱さも無視できない。DeepSeek の利用規約は誤答や不快出力の可能性を明示し、Artificial Analysis は V4 Pro の hallucination rate 94%を報告している。もちろん、これは「常に虚偽を出す」という意味ではなく、**知らないときでも沈黙せず答えに行きやすい**という傾向だが、サイバー、法務、金融、医療のような高リスク用途では、能力差以上に運用上のリスクになる。[28]

リスク・便益評価と提言

便益は明確だ。DeepSeek V4 は、**オープンウェイト、MIT、1M 文脈、強い長文脈効率、強いコード**/agent **特性、低価格**という組み合わせで、2026 年 5 月 1 日時点でも極めて攻撃力の高いモデル群である。とくに、文書横断分析、巨大コードベース読解、コスト感度の高い開発、国や企業の技術主権の観点では、価値が大き

い。CAISI ですら、同等能力帯と見なした GPT-5.4 mini に対し、DeepSeek V4 が 7 **ベンチ中 5 ベンチでより安価**と報告している。[29]

リスクも同じくらい明確だ。第一に、held-out **な抽象推論、サイバー、コード移植**では、公開ベンチから受ける印象ほど強くない。第二に、text-only であり、マルチモーダル業務には不利だ。第三に、hallucination **傾向**と**安全文書の薄さ**が残る。第四に、サービス利用時の**データ処理・保管が中国にまたがる可能性**があり、規制産業や政府調達では追加精査が必要になる。この 4 点は、V4 を「使うべきでない」という意味ではなく、**強みが刺さる場面で選ぶべきモデル**だということを意味する。[30]

- **開発者向け**: V4 は、長文脈・コーディング・オープンウェイト・価格の観点で極めて魅力的だが、採用条件ははっきりしている。**自社の** held-out **課題**を用意し、少なくとも**未知推論、サイバー、コード移植、要約圧縮**で事前検証すべきだ。DeepSeek の自己報告だけでなく、CAISI 型の“知らされていない問題”でテストして初めて、導入の妥当性が見える。[31]
- **導入企業向け**: 規制・顧客データ・知財が絡む用途では、API **利用と重みの自前運用を分けて検討**するのがよい。MIT 重みを自前環境に載せれば、データ所在リスクを減らせる可能性がある一方、**安全フィルタや監査責任は自社側に移る**。API を使う場合は、**入力の取り扱い、保管場所、オプトアウト可否、ログ管理、人手レビューの境界**を契約・運用で明確化すべきだ。[32]
- **規制当局向け**: V4 事例の教訓は、「公開ベンチだけでは能力差を過小評価しうる」ことだ。今後の評価枠組みでは、pre-committed benchmark suite、held-out **課題、**scaffolding **統制、コスト比較、サービス版と** open-weight **版の差分**をセットで要求するのが望ましい。また、**モデル能力の月数表現**は直感的だが誤読されやすいため、Elo や IRT の**重み付け・推定不確実性・月数換算の前提**を必ず併記すべきだ。[33]

優先ソース

- NIST/CAISI **公式評価ページ** — 「約 8 か月遅れ」の主張、ベンチマーク定義、IRT 集約、コスト比較の起点。 [2]
- DeepSeek V4 **技術報告** — モデル仕様、アーキテクチャ、事前学習、自己報告ベンチ、3–6 か月推定の根拠。 [16]
- DeepSeek V4 **モデルカード** / API **リリース** / **料金表** — ライセンス、データ記述、利用条件、推論モード、API 機能、価格。 [34]
- Artificial Analysis — 独立ラボの能力・コスト・hallucination 評価。 [17]
- LiveBench — contamination-limited 公開ベンチのスナップショットと方法論。 [18]
- LM Arena — crowd preference ベースの最新スナップショットと方法論。 [35]

☒ navlist ☒最近の関連報道☒ turn27news29,turn27news27,turn27news28 ☒

---

[1] [2] [5] [6] [12] [21] [23] [25] [29] [31] [33] https://www.nist.gov/news-events/news/2026/05/caisi-evaluation-deepseek-v4-pro

https://www.nist.gov/news-events/news/2026/05/caisi-evaluation-deepseek-v4-pro

[3] [8] [9] [11] [16] [22] [24] https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro/resolve/main/DeepSeek_V4.pdf?download=true

https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro/resolve/main/DeepSeek_V4.pdf?download=true

[4] [15] [17] https://artificialanalysis.ai/articles/deepseek-is-back-among-the-leading-open-weights-models-with-v4-pro-and-v4-flash

https://artificialanalysis.ai/articles/deepseek-is-back-among-the-leading-open-weights-models-with-v4-pro-and-v4-flash

[7] https://api-docs.deepseek.com/news/news260424

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

[10] [13] [27] [32] [34] https://fe-static.deepseek.com/chat/transparency/deepseek-V4-model-card-EN.pdf

https://fe-static.deepseek.com/chat/transparency/deepseek-V4-model-card-EN.pdf

[14] https://api-docs.deepseek.com/news/news1226

https://api-docs.deepseek.com/news/news1226

[18] https://livebench.ai/

https://livebench.ai/

[19] https://arxiv.org/abs/2403.04132

https://arxiv.org/abs/2403.04132

[20] [30] https://www.reuters.com/technology/chinas-deepseek-returns-with-new-model-year-after-viral-rise-2026-04-24/

https://www.reuters.com/technology/chinas-deepseek-returns-with-new-model-year-after-viral-rise-2026-04-24/

[26] https://www.reuters.com/world/china/deepseek-v4-chinese-ai-model-adapted-huawei-chips-2026-04-24/

https://www.reuters.com/world/china/deepseek-v4-chinese-ai-model-adapted-huawei-chips-2026-04-24/

[28] https://cdn.deepseek.com/policies/en-US/deepseek-terms-of-use.html?locale=en_US

https://cdn.deepseek.com/policies/en-US/deepseek-terms-of-use.html?locale=en_US

[35] https://lmarena.ai/leaderboard/text

https://lmarena.ai/leaderboard/text