

# 2026年最新：中国主要LLM 4モデル比較ガイド — 性能・運用・透明性の徹底検証

2026年4月末時点の中国主要4モデル (DeepSeek, Kimi, Qwen, MiMo) の性能評価。総合スコアではKimiとMiMoがリードし、DeepSeekが僅差で続く。Qwenは実務特化のポジションを確立。選定は「絶対性能」と「運用現実性」のバランスが鍵。100万トークン文脈はDeepSeek, エージェントはKimi、自前運用はQwen、効率的APIはMiMoが最適。透明性とライセンスの確認も不可欠。

## DeepSeek-V4-Pro 長文読解と推論のスペシャリスト



MMLU-Pro: 87.5

LiveCodeBench: 93.5

GPQA: 90.1

100万トークンの文脈と圧倒的な長文処理効率。新しい注意機構 (CSA/HCA) でKVキャッシュを10%削減し、長大なコード解析やRAGに最適。



企業向けの大規模な計算資源が必要。1.6兆パラメータのMoEモデルで、通用には大規模なマルチGPUサーバーやクラスター環境が必須。

## Kimi K2.6 エージェント型コーディングの最強機輔



自稔的な開発エージェントに特化した設計。24時間稼働のバックグラウンドエージェントや、マルチモーダルなコーディング支援において、最も高い適性。

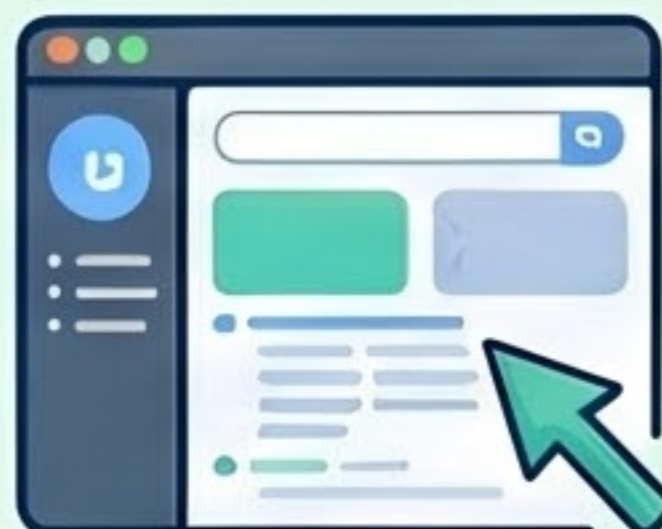


HLE w/tools: 54.0

DeepSearchQA: F1 92.5

ツール利用と検索性能で圧倒的スコア

外部ツールの操作や連発な検索タスクに非常に強い。



UI生成やマルチモーダルタスクに最適

画像・動画入力をサポートし、デザインからコーディングまでを一貫して行うエージェント基盤。

## Qwen3.6-27B 運用の現実性と汎用性のバランス



「小さくても強い」現場主義の強モデル

27BのDenseモデルでありながら、Trillion級のMoEに迫る性能。Apple Siliconや6 GPU構成での自前運用が可能。



Apache-2.0ライセンスによる高い商用利用性

商用利用の制約が少なく、社内での調整 (SFT/DPO/GRPO) や独自プロダクトへの組み込みが最も容易。



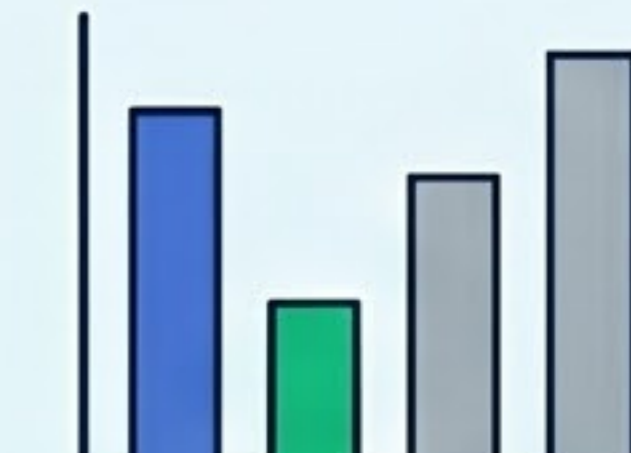
日本語・多言語対応への期待値が最大

201以上の言語・方言への対応を明言。日本語専用スコアは水公園だが、多言語基盤としての怪類性は最高。

## MiMo-V2.5-Pro 長時間エージェントとトークン効率



APIファーストの長時間自律タスク向け。1,000回を組めるツール呼び出しを伴う長時間タスクでも一貫性を維持。プロプライエタリなAPIモデルとして完成度が高い。



競合他社を凌ぐトークン効率

同様のタスクをGPT-5.4やClaude 3.1 Proより40~60%少ないトークン量で完



最大の懸念は「情報の透明性不足」

学習データ、ライセンス、詳細な安全設定が非公開の「ブラックボックス」。企業導入迷、コスト効率が非常に高い。阿の監査コストリスクが高い。

## モデル選定のクイックガイド

| 項目     | DeepSeek-V4-Pro | Kimi K2.6           | Qwen3.6-27B           | MiMo-V2.5-Pro |
|--------|-----------------|---------------------|-----------------------|---------------|
| 主な強み   | 1M文脈・長文効率       | エージェント・開発           | 運用性・ライセンス             | 長時間タスク・効率     |
| ライセンス  | MIT (重み公開)      | Modified MIT (重み公開) | Apache-2.0 (重み公開)     | 非公開 (APIのみ)   |
| ハードウェア | 巨大マルチGPU        | 高級GPU群              | 8 GPU / Apple Silicon | 不要 (API利用)    |
| 透明性    | 中 (一部未開示)       | 中 (系列情報は豊富)         | 高 (運用情報が充実)           | 低 (ブラックボックス)  |
| 日本語用途  | 長大コード/文書解析      | 高性能エージェント           | 最も推奨 (微調整可)           | 自社ベンチでの検証必須   |