

MiMo-V2.6-Pro: 次世代オープンウェイトAIの衝撃と実像

圧倒的な性能とコスト効率



商用モデルの20～60倍のコスト効率

\$1.03 / 1M Tokens



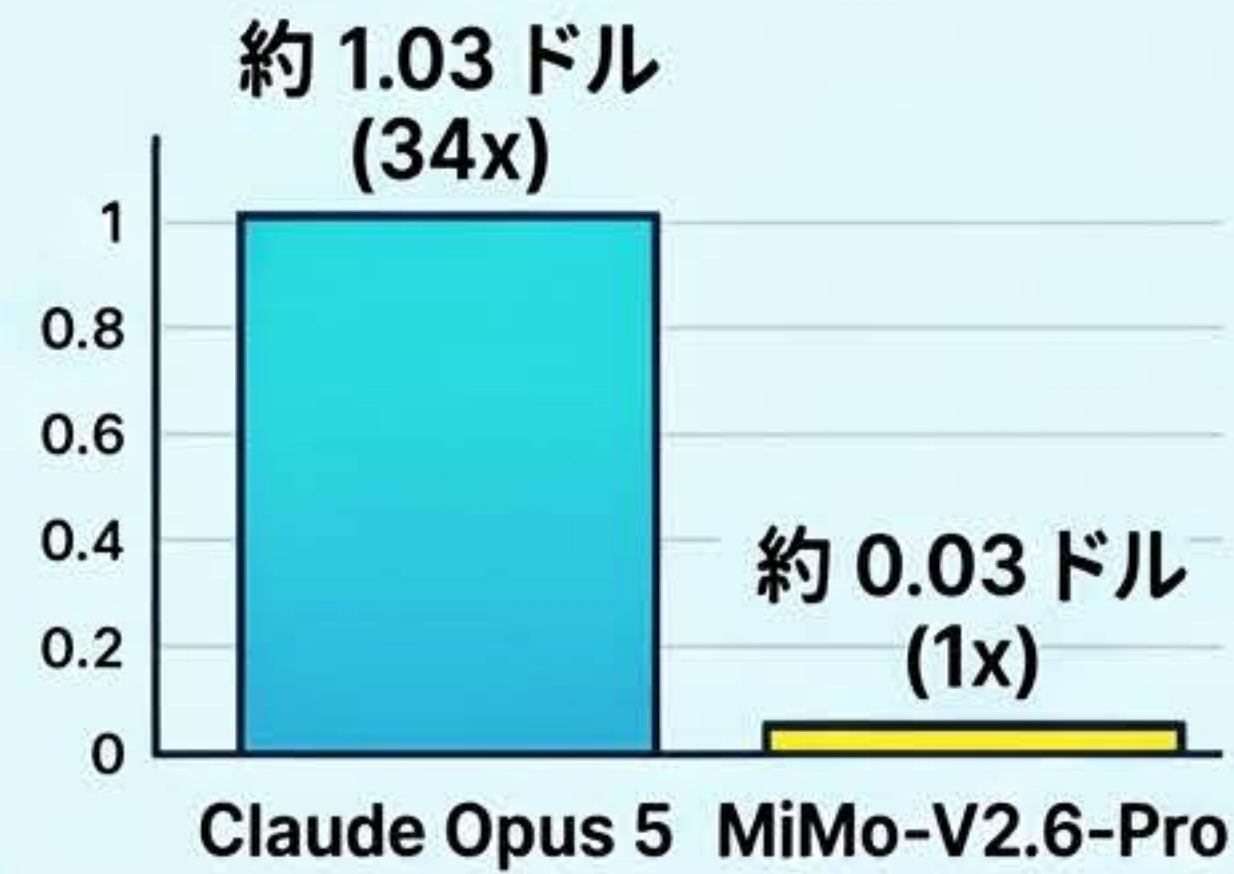
Claude Opus 5

\$0.435 / 1M Tokens



MiMo-V2.6-Pro

実際のタスク実行における
消費コスト比較



オープンウェイトモデルで
首位を獲得



46.32

知能指数

GLM-5.3

DeepSeek
V4系

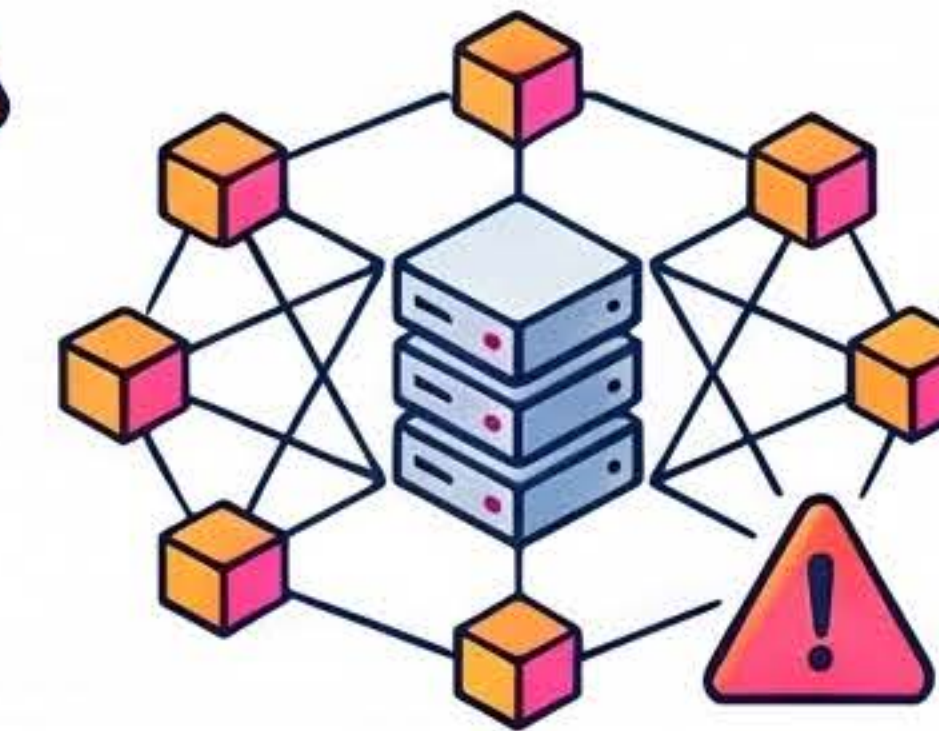
GLM-5.3やDeepSeek V4系を上回る最高性能

自己改善強化学習「GAR」の導入



導入・運用における技術的制約

1.02T全重みの保持とマルチノード環境



- 推論には1兆パラメータ分のメモリが必要
- 活性化は42B
- TP16/EP16構成が不可欠

実効コンテキスト長の精度検証



1Mトークン対応

強化学習時の軌跡: 110K~150Kトークン

謳うが、強化学習時の軌跡は110K~150Kトークンに留まる点に注意

API可用性と単一障害点 (SPOF)



- リリース初期は提供プロバイダが限定的
- 自動フェイルオーバーの構築が困難