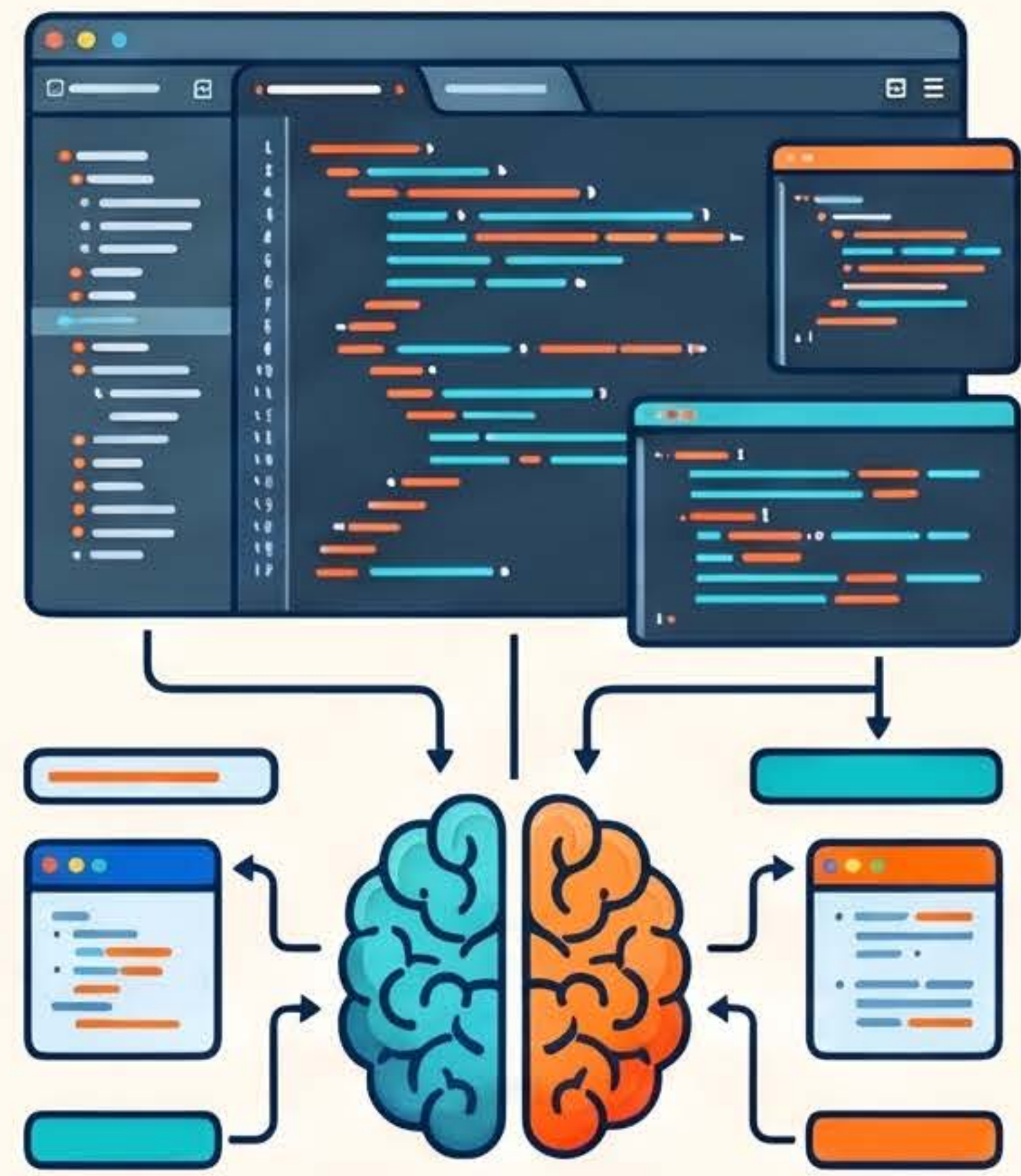


Kimi K3 徹底検証：中国発 2.8兆パラメータモデルは「フロンティア」を超えたのか？

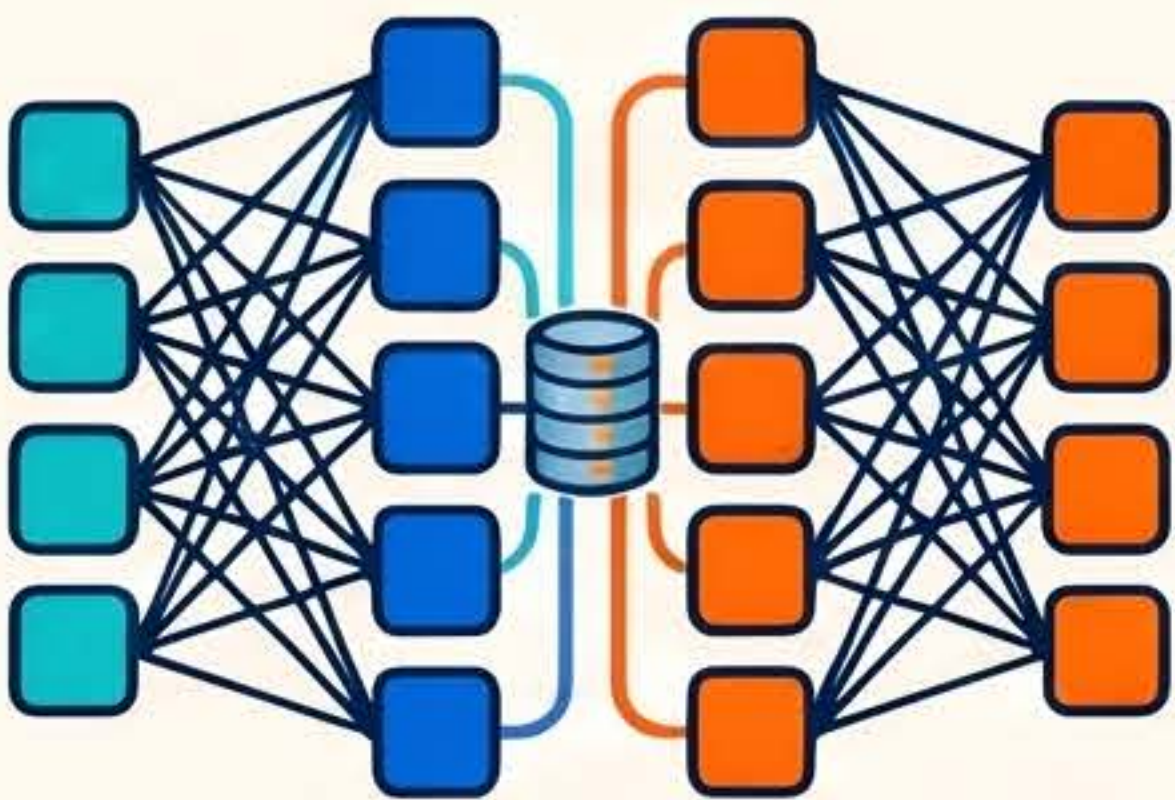
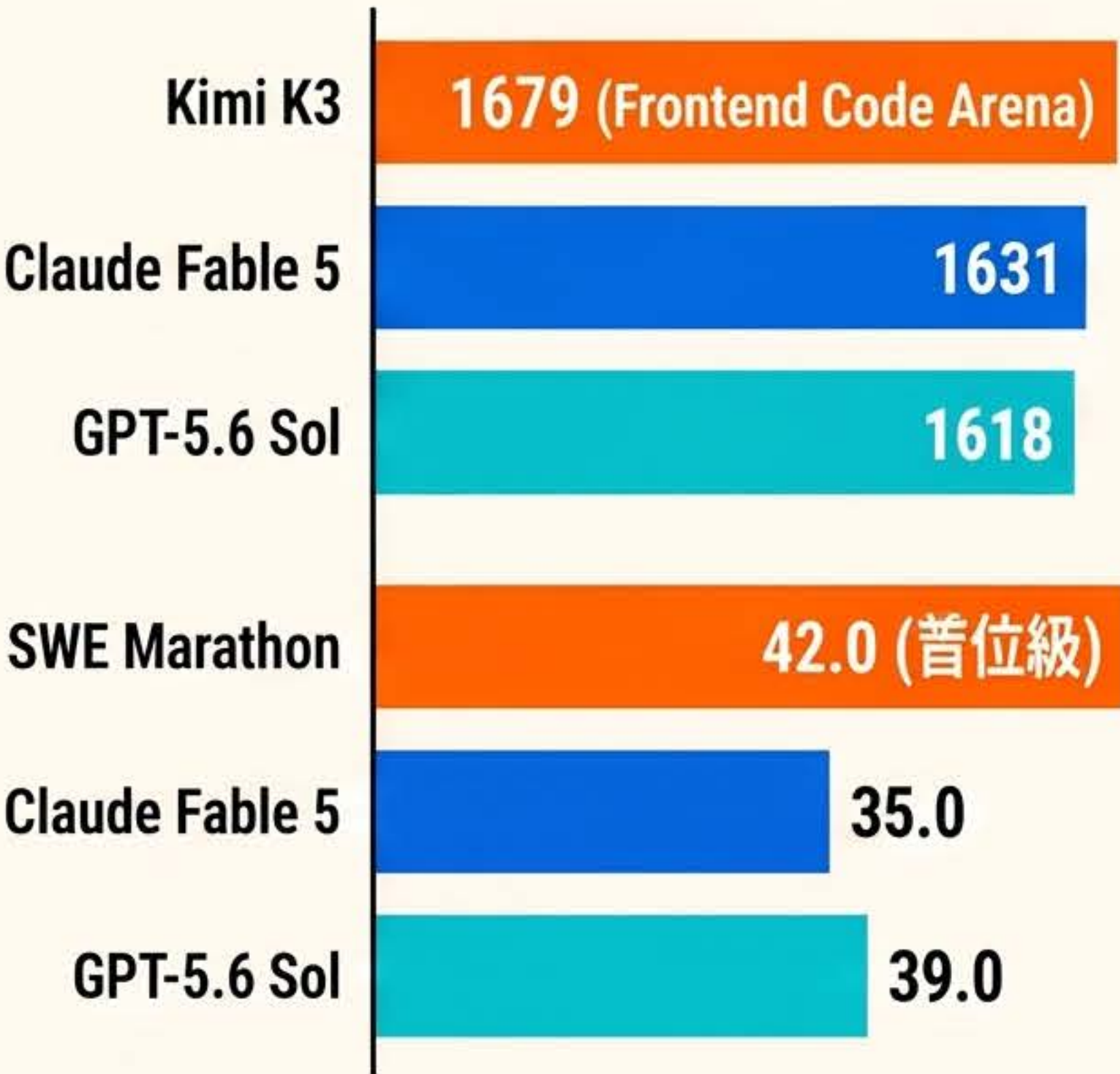
特定領域での圧倒的優位

1679 Elo
Frontend Code Arena
世界暫定首位



コーディングと長時間エージェント能力
SWE MarathonやBrowseComp等のエージェント系タスクで首位級のスコアを獲得。

主要なベンチマーク比較（特定領域）



2.8兆パラメータの巨大MoE構造
896個のエキスパートを備え、100万トークンの広大なコンテキスト窓をサポート。

総合評価と実運用の課題

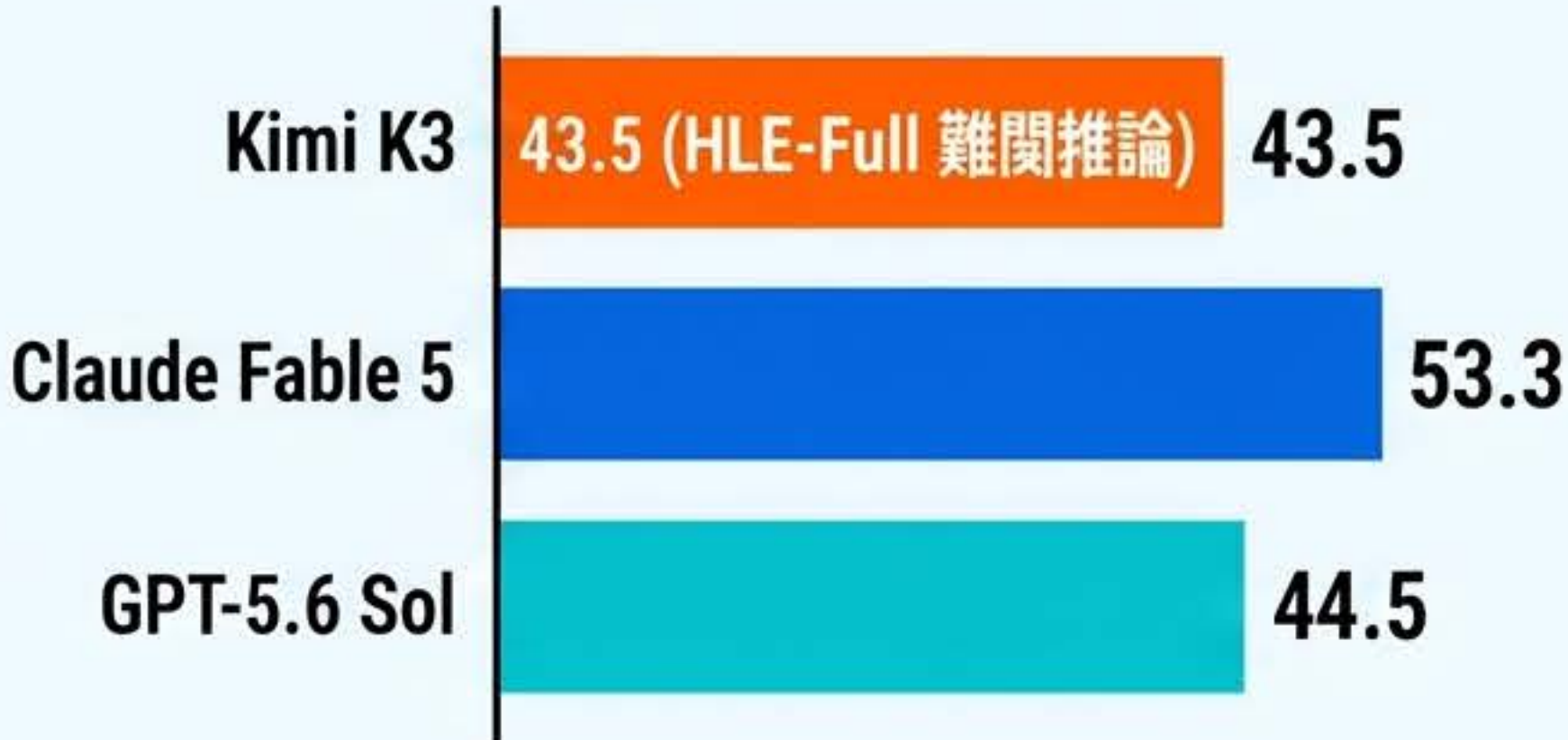


総合指数は世界3位～4位相当
Artificial Analysis Indexでは57.1。
Fable 5やGPT-5.6 Solに次ぐ位置。

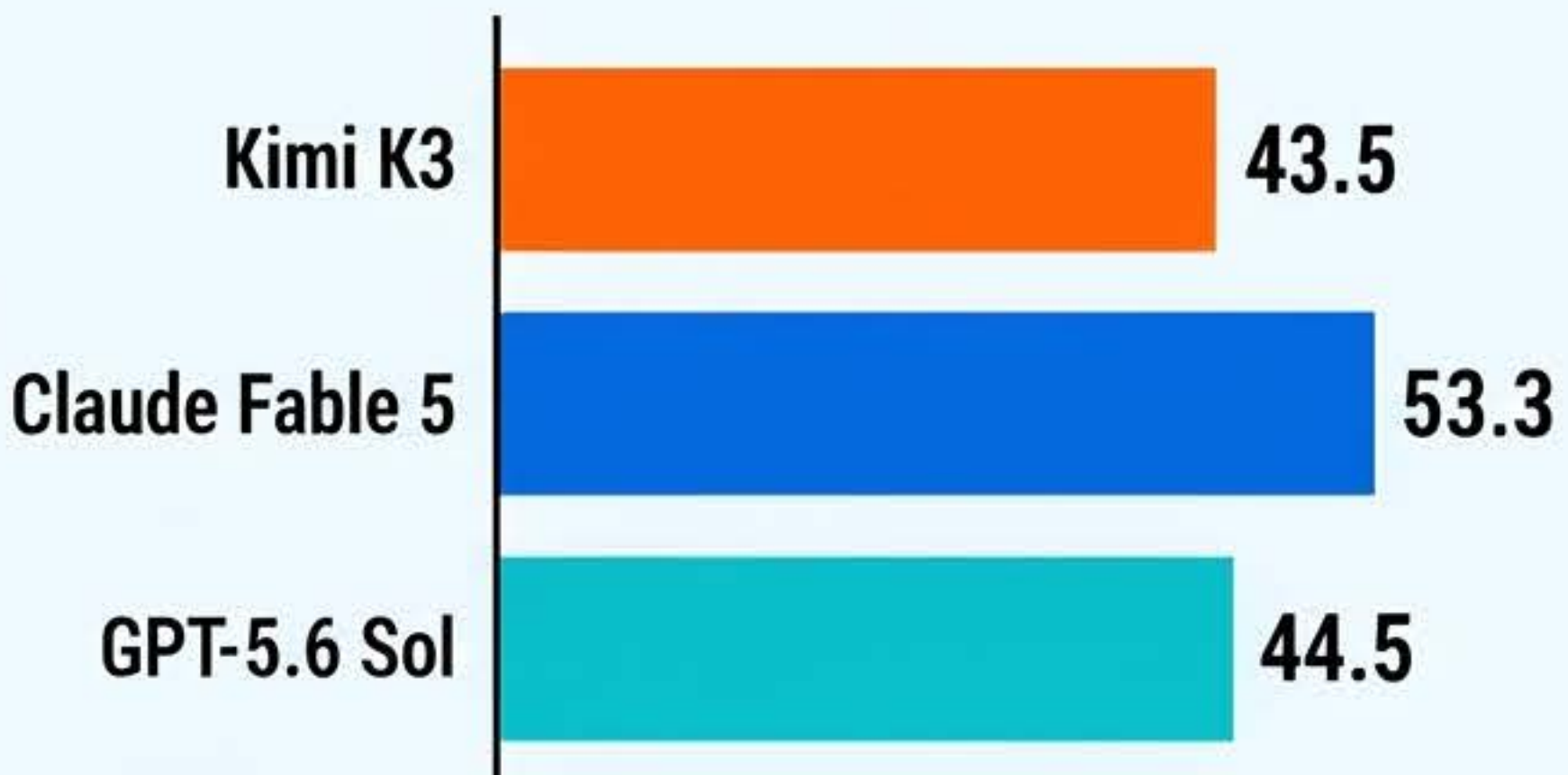


「幻覚」傾向の増大に注意
知識評価（AA-Omniscience）において、
前モデル（K2.6）より誤答を諍断する傾向が
強まったとの報告。

主要なベンチマーク比較（総合/難関）



主要なベンチマーク比較（総合/難関）



7月27日にオープンウェイト公開予定
重みの公開により、コスト効率の高い自前運用
の選択肢（戦略的バックアップ）となる可能性。