

AIは「科学の知識人」から「研究の実践者」へ：GPT-6 Astra vs. Claude Fable 5.1

科学遂行能力の直接対決

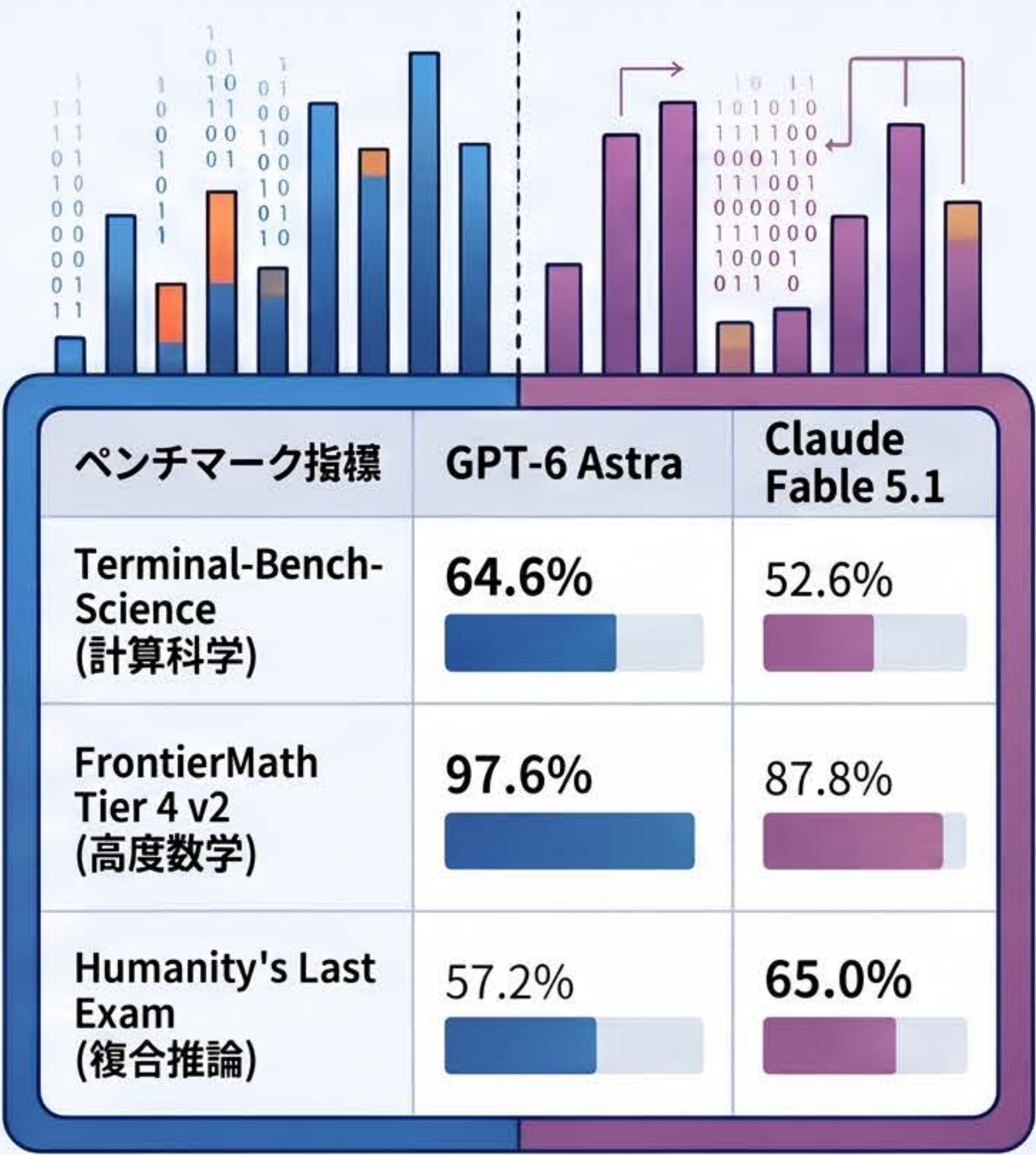


GPT-6 Astra

Astraは検証可能な計算科学で11.9ポイントのリード
Terminal-Bench-ScienceでAstraは64.6%を記録し、Fableを上回りました。

実務遂行のAstra
数学 ソフト操作
Astraは数学・ソフト操作に強く、長時間のエージェント作業で優位です。

両者ともに「知識」を超えた研究成果物を生成
Astraは素数研究の証明、Fableは金星の標高モデルを公開しました。



主要な科学ベンチマークにおける両モデルのスコア比較

Claude Fable 5.1

Fableは長時間のエージェント作業で優位です
Astraは数学・ソフト操作に強く、Fableは長時間のエージェント作業で優位です。

長期探索のFable
長時間 エージェント作業
オーケストレーション
長時間のエージェント作業で優位です。

両者ともに「知識」を超えた研究成果物を生成
Astraは素数研究の証明、Fableは金星の標高モデルを公開しました。



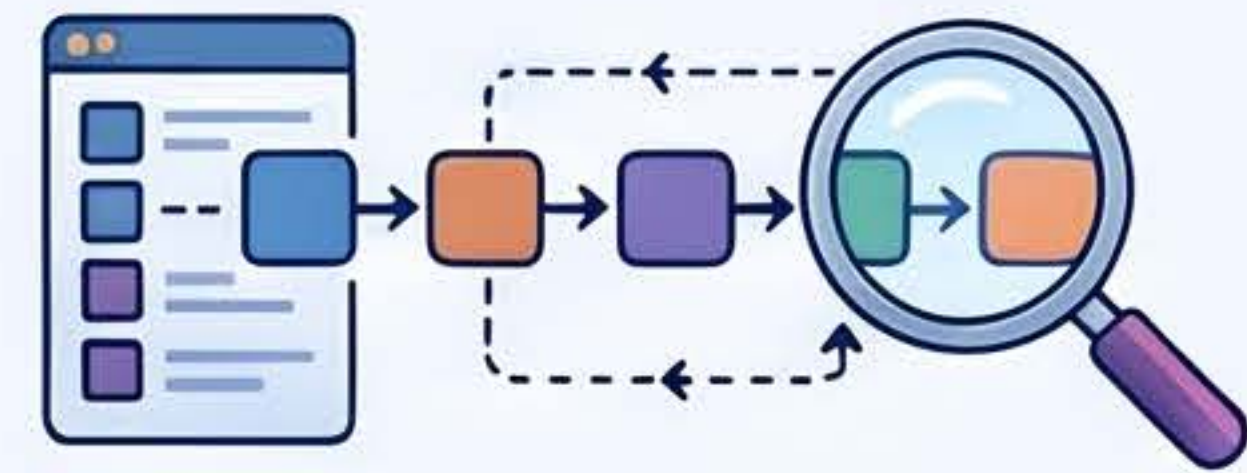
サイエンスエージェントの現在地と限界



「計算科学型エージェント」への条件付き移行
検索・解析・検証のサイクルを回す「研究実務者」として定着されます。



完全自律には未達：物理実験と倫理の壁
ウェットラボ操作や独創的な仮説更新には、依然として人間の介入が必要です。



監査の焦点は「思考」から「行動ログ」へ
内部思考の監視は困難なため、実行したツールの履歴による検証が重要です。