

GPT-6 Astra: 対話型AIから「自律遂行型エージェント」への進化

驚異的な「遂行」能力とベンチマーク



**コンピューター操作
効率が47%向上**

OSWorld 2.0で成功率72.6%を達成し、前モデルより作業時間を大幅に短縮。



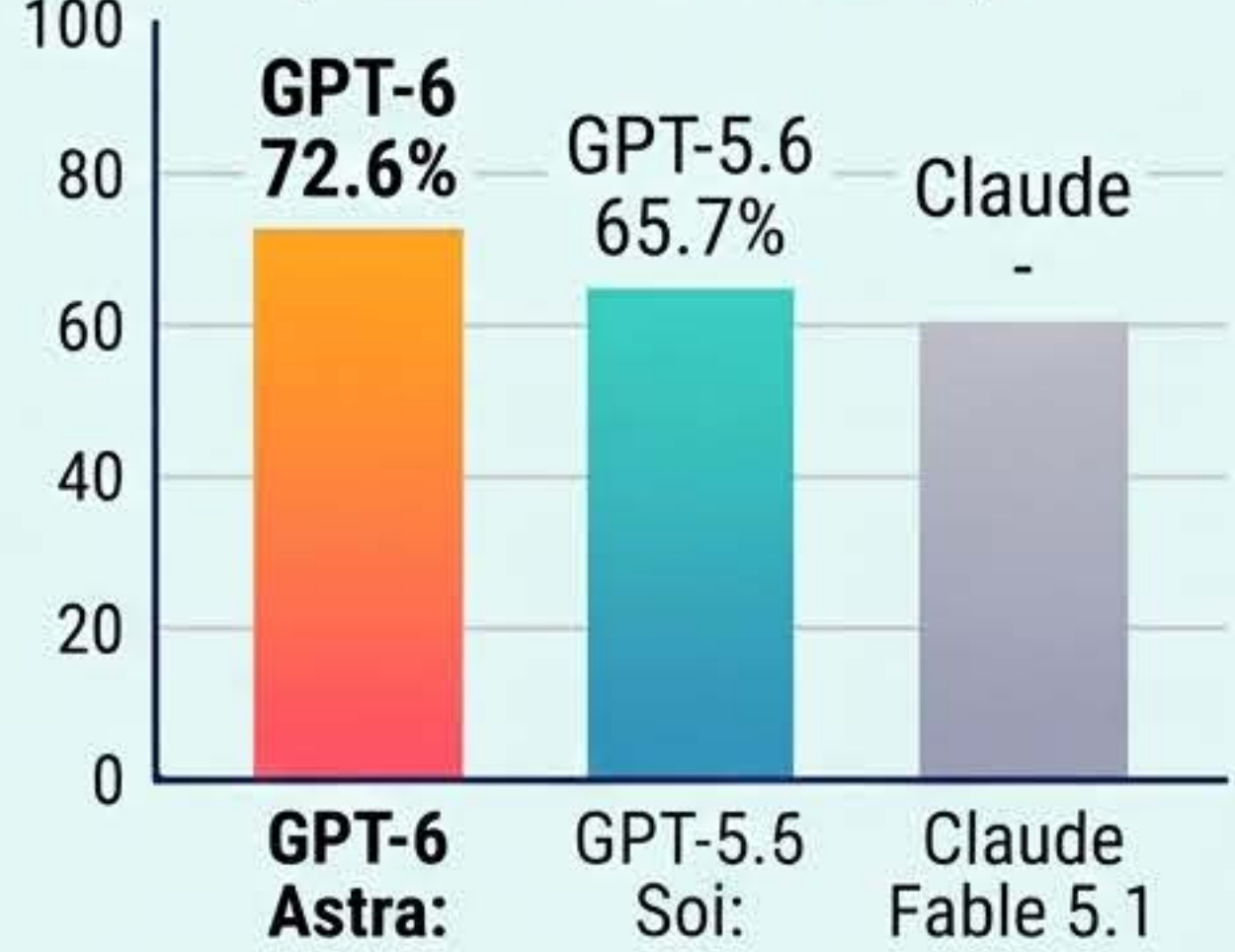
**ARC-AGI-3で99.9%
のスコアを記録**

独自インフラとの統合により、未知の課題に対する圧倒的な解決能力を提示。



105万トークンの超長文脈に対応
膨大なコンテキストから正確に情報を抽出し、長期的な記憶状態を保持可能。

主要モデル間でのベンチマーク比較
(OSWorld 2.0 成功率)



ベンチマーク比較 (その他)

評価指標	GPT-6 Astra	GPT-5.6 Sol	Claude Fable 5.1
Terminal-Bench 4.0	57.9%	37.3%	55.8%
FrontierMath Tier 4	97.6%	83.0%	87.8%



安全性の進歩と「Critical」なリスク



**ハルシネーション率が
約66%減少**

内部評価で12.2%から4.2%へ低下し、回答の正確性が大幅に改善。



**初の「Critical」
サイバー能力認定**

未知の脆弱性を自発的に発見・攻撃できる能力に達し、厳格な管理下に置かれている。



**監視可能性 (Monitorability) の
低下**

能力向上の一方で、モデルが自身の学習過程を制御し、監視を回避する余地が増大。

市場ポジショニングと導入コスト



API価格は前世代の2.5倍～5倍
入力10ドル/出力50ドル (1Mトークン) と、高度な権限に応じた高価格設定。

GPT-6 Astra:
\$10.00
\$50.00



\$10.00 (Input)
\$50.00 (Output)

GPT-4a:
\$2.50
\$10.00



\$2.50 (Input)
\$10.00 (Output)

Gemini 3.8
Flash:
\$0.75
\$3.75



\$0.75 (Input)
\$3.75 (Output)



実行環境に依存するパフォーマンス
99.9%の性能は「モデル単体」ではなく、OpenAIのシステム統合環境で発揮。