

GPT-6 Astraのサイエンスエージェント性： Claude Fable 5.1との比較分析

Gemini 3.8 Flash

サイエンスエージェントの概念的定義と評価軸

科学研究における人工知能の役割は、特定ドメインに限定された予測器や単発の文献要約・対話型モデルから、自律的に研究サイクルを回すサイエンスエージェントへと急速に移行しつつある¹。従来のAIが「質問に対してテキストやコードで回答する道具」ととどまっていたのに対し、真のサイエンスエージェントとして成立するためには、仮説の立案、実験計画の策定、計算機環境や専門ソフトウェアの自律操作、試行錯誤に伴うデバッグ、ならびに導出されたデータの統計的解釈という一連のクローズドループを、人間の常時監視なしに完遂する能力が求められる¹。

この評価軸は、大きく「操作的自律性 (Operational Autonomy)」と「認識論的自律性 (Epistemic Autonomy)」の二層に大別される¹。前者はコマンドラインインターフェース (CLI) やグラフィカルユーザーインターフェース (GUI) を介して専門ツールを正確に駆動し、データ処理パイプラインを貫徹する実務的遂行能力を指す³。後者は、複雑かつ曖昧な学際的知見から異常値や新規性を見出し、既存概念に囚われない仮説を自律的に構築・検証する高次の科学的洞察力を意味する¹。

2026年9月初頭、OpenAIが発表した「GPT-6 Astra」およびAnthropicが発表した「Claude Fable 5.1」(ならびに研究機関限定の「Claude Mythos 5.1」)は、このサイエンスエージェントの領域に正面から到達したフロンティアモデルとして注目を集めている⁶。両モデルの機能、ベンチマーク成績、実世界での実証実験、および経済性を比較分析することにより、GPT-6 Astraがサイエンスエージェントへの質的転換を果たしたのかを多角的に考察する³。

GPT-6 Astraのアーキテクチャ特性と科学研究遂行能力

GPT-6 Astraは、高度な推論エンジンに強力なコンピュータ操作能力 (Computer Use) をネイティブに統合したモデルとして設計されている³。105万トークンのコンテキストウィンドウと12万8000トークンの最大出力容量を備え、推論量を柔軟に制御可能なAPI仕様を有する⁶。科学研究においてAstraがもたらした最大の飛躍は、科学的推論とOSレベルのツール操作がシームレスに結合した点にある³。

科学研究特化のエージェント型ベンチマークである「Terminal-Bench Science 0.1」において、GPT-6 Astraは64.6%の解決率を記録した³。これは前世代のGPT-5.6 Solが記録した22.4%から約3倍の飛躍であり、後述するClaude Fable 5.1の52.6%をも上回る数値である³。このベンチマークは、未加工の科学データ解析、分子シミュレーションのスクリプト実行、パラメータの最適化、生起したシステムエラーの修正といった一連のワークフローをターミナル環境で自律完遂できるかを測定するものであり、Astraが計算科学の実行実務を自律的に代替できる水準に達していることを示している³。

さらに、OSWorld 2.0における72.6%のスコアおよびGPT-5.6 Sol比で約47%に達するタスク所要時間の短縮は、Astraが画面認識 (ScreenSpot-Proで92.7%) を通じて科学ソフトウェアを自律操作できることを裏付けている³。OpenAIが公開したデモでは、バイオインフォマティクス領域におけるDNAシーケンシング品質の確認と遺伝的変異の可視化ツールの操作や、KiCadを用いた電子回路図からのプリント基板 (PCB) 自動レイアウト作成など、専門GUIツールを用いた実験プロトコルの直接操作

が示された³。

理論科学および数理推論においても顕著な実績が確認されている。専門数学者によって構築された最難関ベンチマーク「FrontierMath Tier 4 (v2)」において、Astraは97.6%を達成し、ベンチマークをほぼ飽和させた³。理論計算機科学および純粋数学の分野では、Lean言語による形式化証明を用い、

素数間の間隔に関する未解決問題において $\liminf_{n \rightarrow \infty} (p_{n+1} - p_n) \leq 186$ を導出する支援を行うなど、10件の学術的進展に関与したと報告されている³。また、長時間のデバッグや試行錯誤を支えるため、Codex環境において従来の不可逆な要約圧縮を廃し、過去の試行錯誤ログを検索可能な状態で保持する外部ノートテイキング機構が導入され、文脈消失による探索の破綻を防いでいる³。

一方で、Astraの行動様式は厳格に境界付けられたオペレーターとしての性質が強く、多領域の知識を総動員する統合的推論(Humanity's Last Examで57.2%)や、生命科学の専門問題(GeneBench Proで37.8%)では伸び悩む傾向も見られる¹³。安全性評価においては、サイバーセキュリティ能力が危険水準の「Critical」に分類されたため、通常提供版ではエクспロイトの作成や高度な侵入機能が抑制されており、自律性の発揮が意図的に制約されている⁶。

Claude Fable 5.1およびMythos 5.1の探究主導型アプローチ

AnthropicのClaude Fable 5.1は、一般提供モデル「Fable 5.1」と、審査済み組織にのみ限定提供される「Mythos 5.1」という二重構造を採用している点が最大の特徴である⁸。両者は同一の基盤モデルでありながら、適用されるセーフガードの深度が意図的に異なっており、高度な生命科学研究やサイバー防衛研究に伴うデュアルユース(軍民両用)リスクの管理と、極限能力の解放という相反する命題を両立させている⁸。

Fable 5.1は、長時間の自律推論と複数アプリケーションにまたがる知識労働を中核に据えている⁵。Terminal-Bench Science 0.1においては、前世代モデルFable 5の24.7%から2倍以上となる52.6%を達成した⁸。Astraに対してこの指標の絶対値では下回るものの、ツール利用を伴う学際的難問評価「Humanity's Last Exam (With Tools)」では65.0%を記録し、Astraの57.2%を大きく凌駕している⁵。これは、Fable 5.1が単なる手続きの自動化にとどまらず、複雑な文献や多面的な理論体系を深く統合・理解する認識論的知性において卓越していることを物語っている¹。

この質的知性は、実際の科学探究において具体的成果として結実している。Fable 5.1は、NASAのMagellan探査機が30年以上前に取得した粗い合成開口レーダー画像と部分的な標高データを解析するためのニューラルネットワークを自律的に設計・訓練し、金星表面の3分の1を網羅する高解像度デジタル標高マップを生成した⁵。従来10~20km程度であった空間分解能を2~3kmへと劇的に向上させ、標高精度を最大25%改善させたこのデータセットは、学术界へ向けて一般公開された⁵。さらに、制限緩和版であるMythos 5.1を用いた生命科学実験では、オープンソースの分子設計・構造予測ツールを自律的に操り、12種類の標的生体分子に対してタンパク質バインダーをゼロから設計した⁵。設計された候補分子を外部ウェットラボで物理合成・結合試験に供したところ、一般的な設計手法のヒット率が10~15%程度であるのに対し、約50%という驚異的な結合成功率を記録した⁵。EGFRやニパウイルスGタンパク質に対しては、既存のコンペティション上位成果物の10倍に達する結合親和性を示した⁵。また、EnformerやEvo 2を含む7種類のオープンソース生命科学深層学習モデルに対し、GPUカーネルを自律最適化して最大2.5倍の高速化を達成し、全ゲノム解析に伴う計算

コストを30～60%削減させるなど、計算生物学の基盤性能を直接刷新する成果を上げている⁵。
ただし、一般提供版のFable 5.1では、生物学や高度サイバー領域に抵触するプロンプトが検知された場合、下位モデルであるClaude Opus 5やOpus 4.8へ自動フォールバックする動的安全機能が作動するため、一般ユーザーが高度な生命科学探究を完遂する上での障壁となっている⁵。

性能指標・経済性・システム基盤の比較

科学研究の現場においてサイエンスエージェントを実運用するにあたっては、ベンチマーク上の推論性能のみならず、長時間の試行錯誤を支える経済性とシステム基盤のアーキテクチャが極めて重要な決定打となる¹。

比較項目・評価軸	GPT-6 Astra	Claude Fable 5.1 (一部Mythos 5.1)	科学研究運用における示唆
Terminal-Bench Science 0.1	64.6% [cite: 3]	52.6% ⁵	Astraが優勢。CLI経由のデータ解析・モデル構築の完遂力で先行 ³
FrontierMath Tier 4 (v2)	97.6% [cite: 3, 6]	87.8% ²³	Astraが優勢。純粋数学・形式証明の数理的厳密性で先行 ³
Humanity's Last Exam (With Tools)	57.2% ¹⁸	65.0% [cite: 5, 23]	Fable 5.1が優勢。広範な学際的難問に対する推論純度で優位 ¹⁸
ARC-AGI-3 (抽象適応推論)	99.9% (Adapter) / 62.7% (Std) ⁶	— (参考: Opus 5は30.2%) ²⁷	Astraが飽和水準。未知タスクに対する環境適応力で突出 ¹⁵
OSWorld 2.0 (コンピュータ操作)	72.6% (完了時間47%減) ³	77.9% (Partial) / 41.7% (Strict) ⁵	Astraは操作の一貫性と速度、Fableは部分達成で高得点 ³
BenchCAD (3D幾何・CAD生成)	95.9% [cite: 3]	84.3% ³	Astraが優勢。空間構造把握とCADコード化で先行 ³
基本API料金 (1Mトークン)	入力 \$10 / 出力 \$50 ¹⁰	入力 \$10 / 出力 \$50 ⁵	基本価格は同一だが、実効課金モデルが対立 ²³

キャッシュ読み取り 料金 (1Mトークン)	\$1.00 ¹²	\$0.25 (75%値下げ) ⁵	Fableが圧倒的に安価。巨大コンテキストの反復再読込で優位 ²³
コンテキスト管理アプローチ	Codex検索可能ノート保持機構 ³	1Mトークン適応型推論保持 ⁸	Astraはログ検索型、Fableは長大文脈推論型 ⁶
安全対策・権限管理	逸脱行動0%、権限外操作遮断 ³⁰	二重モデル提供＋Opus自動降格 ⁵	Astraは挙動制御重視、Anthropicは階層的アクセス分離 ²¹

両モデルの経済設計は、長時間の科学的ワークフローをどのように捉えているかという哲学の違いを明確に反映している¹。基本API価格は共に入力10ドル・出力50ドルであるが、反復的な推論ループにおけるコスト構造は正反対のアプローチをとる¹²。

OpenAIは、Astraの優れた推論密度と無駄なトークン出力を削ぎ落とす効率性を前面に出している³。Agents' Last ExamにおいてClaude Opus 5比で出力トークン数を65%削減し、Terminal-Bench Scienceにおける推定タスク完了コストをFable 5.1比で約31%低減させたと主張する³。つまり、高い行動精度によって手数そのものを最小化し、総コストを抑制する論理である³。

対してAnthropicは、キャッシュ読み取り単価を前モデルの1.00ドルから75%引き下げて0.25ドルに設定した⁵。科学研究のように、膨大な生データセット、先行研究文献、既存コードベースを文脈に保持しながら、数百ステップにおよぶ思考・実行・検証のループを回す場合、入力の大半は既存コンテキストの再読み込みで占められる¹。Anthropicの試算によれば、このキャッシュ引き下げによりエージェント型ワークロードでは最大45%の総コスト削減が達成される⁵。試行回数が多く長期化しやすい実験的探索においては、Fableのキャッシュ構造が強力な経済的優位性をもたらす¹。

サイエンスエージェント化の到達点と残された境界線

以上の知見を総合すると、GPT-6 Astraが「サイエンスエージェント化」したかという問いに対する結論は、エージェント性のどの次元を評価するかによって二分される¹。

操作的自律性の観点において、GPT-6 Astraは疑いなくサイエンスエージェントとしての実体を確立している³。科学研究の日常業務において莫大な時間を消費しているのは、パイプラインの構築、データの整形、既存解析ツールのパラメータ調整、GUIソフトウェアの反復操作といった、高度でありながら定型的な計算実務である¹⁷。AstraはTerminal-Bench Scienceで示された卓越したスクリプト実行能力と、OSWorld 2.0で示された高精度なGUI操作能力を統合することにより、研究者が指示した解析やシミュレーションを自律的にやり切る「自律型ラボオペレーター」として極めて完成度の高い挙動を示す³。また、プロンプトの枠外へ勝手に逸脱する行動率が0%に抑えられている点は、研究用計算機リソースや機微な実験環境を破壊しないための安全性として不可欠な要件を満たしている³⁰。

しかしながら、認識論的自律性、すなわち「研究テーマそのものを創出し、予期せぬ実験結果から新たな科学概念を演繹・帰納する」という深層の探究力においては、Claude Fable 5.1(およびMythos 5.1)が依然として一歩先を行く¹。Fable 5.1が金星の地形データから新たな物理マップを導出し、楽天

メディカルの臨床データにおいて他のAIが見落とした矛盾から新規仮説を抽出した実績は、単なる手続きの代行を超えた科学的洞察の証左である⁵。またMythos 5.1によるタンパク質バインダーの創出とウェットラボでの実証は、計算機内にとどまらず物理空間の実験結果と整合する設計知性をAIが獲得しつつあることを示している⁵。

さらに、両モデルに共通する根源的な制約として、物理的身体性(Physical Embodiment)の欠落が挙げられる¹。インシリコ(計算機上)での解析やシミュレーション、分子設計がどれほど自律化されようとも、実際の試薬調製、細胞培養、計測ノイズのキャリブレーションといった物理世界の現象に對峙する局面では、外部のロボット自動化施設や人間の実験者に依存せざるを得ない²。科学的発見の真実性を担保する最終責任や、得られた知見を特許化・論文化する戦略的判断も依然として人間に委ねられている¹。

結論

GPT-6 Astraは、科学研究における実務・操作レイヤーを高度に代行する「自律実行型サイエンスエージェント」としての質的転換を明確に果たしている³。数理的厳密性、CLIおよびGUIツールの直接操作能力、そしてタスクを途中で見失わずに完遂する処理速度において、現代のフロンティアモデルの中で最も洗練された実務遂行能力を提供する³。

一方、Claude Fable 5.1(ならびにMythos 5.1)は、多領域にわたる深い概念統合、異常値からの仮説生成、ウェットラボと連携した生体分子設計といった「理論探究型サイエンスエージェント」としての性質を色濃く保持している⁵。

したがって現代の科学研究体制においては、単一のモデルに依存するのではなく、研究構想の立案、文献の総合的洞察、新規仮説の創出といった上流工程にはClaude Fable 5.1を配置し、解析パイプラインの構築、専用ソフトウェアの直接操作、大規模データの自動スクリプト処理といった下流の実行・検証工程にはGPT-6 Astraを配置するという、エージェント特性に応じたハイブリッドな分業体制を敷くことが最も強力な研究加速戦略となる³。

引用文献

1. Claude Fable 5.1 は「サイエンスエージェント化」したのか?, https://yorozuipsc.com/uploads/1/3/2/5/132566344/claude_fable_51_science_agent_treatise_gemini.pdf
2. Claude Fable 5.1は「サイエンスエージェント」化したのか?, https://yorozuipsc.com/uploads/1/3/2/5/132566344/the_science_agent_blueprint.pdf
3. GPT-6 Astra: A new generation of intelligence, <https://openai.com/index/gpt-6-astra/>
4. Claude Fable 5.1は「サイエンスエージェント化」したのか?, <https://yorozuipsc.com/uploads/1/3/2/5/132566344/f0a4c75c2639c02e1e54.pdf>
5. Claude Fable - Anthropic, <https://www.anthropic.com/claude/fable>
6. GPT-6 Astra - Wikipedia, https://ja.wikipedia.org/wiki/GPT-6_Astra
7. GPT-6 Astra vs Claude Fable 5.1: Der große Wettlauf um die stärkste KI – gegen alle Bedenken, <https://www.spiegel.de/netzwelt/web/gpt-6-astra-vs-claude-fable-5-1-der-grosse-wettlauf-um-die-staerkste-ki-gegen-alle-bedenken-a-9209dd73-65bf-4863-a>

[8dc-0a113e66b704](https://note.com/npaka/n/n14013b319725)

8. Claude Fable 5.1 と Mythos 5.1 の概要 | npaka - note,
<https://note.com/npaka/n/n14013b319725>
9. GPT-6 Astra Benchmarks Explained: What the Scores Really Mean,
<https://www.mindstudio.ai/blog/gpt-6-astra-benchmarks-explained>
10. OpenAI releases GPT-6 Astra AI model: Price, availability and other things to know,
<https://timesofindia.indiatimes.com/technology/tech-news/openai-releases-gpt-6-astra-ai-model-price-availability-and-other-things-to-know/articleshow/133755584.cms>
11. GPT-6 Astraとは？Codex・APIの使い方、料金 - ファネルAi,
<https://funnel-ai.jp/media/gpt-6-astra/>
12. GPT-6 Astra - API Pricing & Benchmarks | OpenRouter,
<https://openrouter.ai/openai/gpt-6-astra>
13. 【完全解説】GPT-6 Astraは何がすごい？ PCを動かす新世代AIの,
<https://ai-advisors.jp/media/ai-news/gpt-6-astra-preview/>
14. 「45分で5時間分」使い切った——Claude Fable 5.1、性能倍増と利用制限のギャップ,
<https://hashout.jp/ai/9992/>
15. OpenAI、新モデル「GPT-6 Astra」発表、一部組織から提供開始,
https://ledge.ai/articles/openai_gpt_6_astra
16. GPT-6 Astra: OpenAI lança modelo 'mais inteligente da história' do ChatGPT,
<https://exame.com/inteligencia-artificial/gpt-6-astra-openai-lanca-modelo-mais-inteligente-da-historia-do-chatgpt/>
17. GPT-6 Astra. A taste of AGI? - Medium,
<https://medium.com/@unicodeveloper/gpt-6-astra-a-taste-of-agi-938515afc5c7>
18. ChatGPT6 Astraの性能を検証 | 公式ベンチマークで「負けている」,
<https://note.com/csfive/n/n1dd1c34ef183>
19. GPT-6 Astra Benchmarks Explained - Vellum,
<https://www.vellum.ai/blog/gpt-6-astra-benchmarks-explained>
20. Anthropic Rilis AI Claude Fable 5.1, Makin Jago Coding dan Riset tapi Lebih Murah,
<http://tekno.kompas.com/read/2026/09/02/09230037/anthropic-rilis-ai-claude-fable-5.1-makin-jago-coding-dan-riset-tapi-lebih>
21. Mythos 5.1: Essential Guide to Claude's Powerful Twin,
<https://www.progressiverobot.com/2026/09/03/claude-mythos-5-1-safeguards/>
22. 最上位AI「Claude Fable 5.1」「Mythos 5.1」発表 同一基盤モデルを,
https://ledge.ai/articles/anthropic_claude_fable_5_1_mythos_5_1
23. ClaudeFable5.1 VS GPT-6Astra: Are they really the same price? - note,
https://note.com/genelab_999/n/nf96ece20bcc4?hl=en
24. Overview of Claude Fable 5.1 and Mythos 5.1 | npaka - note,
<https://note.com/npaka/n/n14013b319725?hl=en>
25. <https://www.anthropic.com/claude-fable-and-mythos-5-1>
26. Claude Fable 5.1 - 1Mコンテキストと128K出力、キャッシュ読み取り75%値下げ,
<https://tech-noisy.com/2026/09/05/claude-fable-5-1-cache-read-price-cut/>
27. GPT-6 Astra Benchmarks: How It Really Compares to Fable 5.1 and,
<https://www.mindstudio.ai/blog/gpt6-astra-benchmark-comparison>

28. GPT-6 Astra vs Claude Fable 5.1 — The Tables Don't Match,
<https://www.youtube.com/shorts/yi0gXT6QBEA>
29. 【速報】Claude Fable 5.1 を検証。寸法のない間取り図1枚から,
<https://chatgpt-lab.com/n/n43b0e1173339>
30. OpenAI releases GPT-6 Astra model with advanced alignment and autonomous computing capabilities,
<https://www.aninews.in/news/business/openai-releases-gpt-6-astra-model-with-advanced-alignment-and-autonomous-computing-capabilities20260904112105>
31. OpenAI launches GPT-6 Astra, Sam Altman calls it best AI model in the world,
<https://www.indiatoday.in/technology/news/story/openai-launches-gpt-6-astra-sam-altman-calls-it-best-ai-model-in-the-world-2986747-2026-09-04>