GPT-6 Astra は「サイエンスエージェント化」したのか？

― Claude Fable 5.1 との比較考察 ―

Claude Fable 5.1

2026 年 9 月 6 日

要旨

本稿は、OpenAI が 2026 年 9 月 3 日に公開した「GPT-6 Astra」が「サイエンスエージェント（AI Scientist）」と呼びうる段階に達したかを、Anthropic が同年 9 月 1 日に公開した「Claude Fable 5.1」と対比しつつ検討する。結論は三点である。

第一に、両モデルは実在し一次ソースで確認できるが、両社とも自社モデルを「AI Scientist」とは呼んでいない[1,2]。「サイエンスエージェント化」は、筆者およびメディア・アナリスト側の解釈枠組みである。第二に、両モデルは「アシスタント→コパイロット→自律エージェント」の連続体上で、限定的な自律性を備えた「高度コパイロット～準自律エージェント」の段階にある。「仮説→実験→検証→再計画」の閉ループを人間の介在なしに回した事例は、いずれにも確認できない。第三に、両者の差は「対象領域」と「検証の様式」にある。Astra は数学・理論計算機科学における形式検証済みの「発見」を前面に出し、Fable 5.1 は実験科学・企業研究への「実装」（外部ウェットラボ検証、分析化学、計算生物学）を前面に出す。この「発見型」と「実装型」の分岐が、本稿の中心的な観察である。

なお本稿では、一次ソースで確認できた事項を【確認済み事実】、二次報道に依拠する事項を【報道ベース】、筆者の分析・推論を【分析】として区別する。未確認事項は末尾の付記に集約した。

1. 前提の事実確認 ― 「GPT-6 Astra」とは何か

【確認済み事実】「GPT-6 Astra」は実在する。OpenAI は 2026 年 9 月 3 日にシステムカード（Safety overview）を公開し [1]、同日のブリーフィングを経て 9 月 4 日から段階的に提供を開始した [3,4]。API 上のモデル ID は gpt-6-astra、コンテキストウィンドウは 100 万トークン、標準価格は入力 100 万トークンあたり 10 ドル、出力 50 ドルである [3,4]。ChatGPT（Plus／Pro／Business／Enterprise）、OpenAI API、Microsoft Azure、AWS Bedrock で提供され、上位プラ

ン向けに「GPT-6 Astra Pro」も用意された[3,4]。

「Astra」の名称は、2026 年 8 月 1 日の数学成果発表の時点では未公開の「次期主要モデル（next major model）」を指す呼称として登場し[7]、その後「GPT-6 Astra」として製品化された。すなわち 8 月 1 日の「未解決問題 10 問解決」は内部版 Astra の成果であり、公開版との厳密な性能差分は一次ソースに明示されていない。

「サイエンスエージェント化」という語について。OpenAI は 8 月 1 日の発表で Astra を「研究協働者（research collaborator）」と位置づけ[7,8]、研究者 Noam Brown 氏は X で「科学的推論にとって大きな一歩」と述べた[8,9]。他方、「AI 研究者」「自律的研究者」は OpenAI の中期ロードマップ上の目標（2026 年 9 月までにインターン級の研究助手、2028 年までに本格的な AI 研究者）として語られたものであり[27]、Astra 自体に付されたラベルではない。Anthropic も Fable 5.1 を「コーディングと知的作業のための最先端モデル」と説明しており[2]、「AI Scientist」とは呼んでいない。

2. GPT-6 Astra の科学研究能力の実態

2.1 数学・理論計算機科学 — Astra の主戦場

【確認済み事実】2026 年 8 月 1 日、OpenAI は内部版 Astra が数学・理論計算機科学の未解決問題 10 件を解決したと発表し、249 ページの原稿と Lean 4 による形式証明を GitHub リポジトリ（openai/ten-proofs、Apache 2.0）で公開した[6,7,8]。Lean 証明における未証明箇所（sorry）はゼロ、すなわち全ステップが機械検証済みである[8,10]。OpenAI は、10 件の解を見つけるのに要したトークン数が Sol API 料金換算で約 2,000 ドルに相当すると明記している（訓練コスト、失敗した試行、人間の労力は含まれない）[9]。

目玉は非ソフィック群（non-sofic group）の初の明示的構成である。ソフィシティの概念は Gromov が 1999 年に導入し[24]、以後 27 年間未解決であった[8]。ほかに Connes の剛性予想の反証、Ehrhart の体積予想の証明、多色三角形ラムゼー数の超指数的下界（Erdős 問題 183 の解決）などが含まれる[8,11]。発表時点の性能指標として、FrontierMath Tier 4（v2）で 97.6%、GPQA Diamond で 96.0%が示され、素数間隔に関する 2 つの新結果（短い間隔での定数の 240 から 186 への改善、80 年以上更新のなかった長い間隔に関する項の改善）も公表された[9,11]。

2.2 実験科学・ライフサイエンス

Astra 本体は、シーケンシングの品質検査や細胞追跡ワークフローなど、科学ソフトウェア上で

データを直接検査・可視化できるとされる [3,11]。ただしこれは研究実務の支援であり、自律的なウェットラボ連携ではない。ライフサイエンス領域については、OpenAI は 2026 年 4 月に専用モデル「GPT-Rosalind」を別途提供している。同モデルは証拠統合・仮説生成・実験計画・多段ワークフローを支援し、Codex 向け Life Sciences 研究プラグインを通じて 50 超の科学ツール・データソースに接続する [14]。もっとも OpenAI のライフサイエンス研究リード Joy Jiao 氏は、AI が単独で新しい疾患治療を生み出せる段階にはないと述べている [14]。

2.3 エージェント能力と自律性

【確認済み事実】Astra は「コンピュータ操作」モデルとして位置づけられ、フォーム入力、CRM 更新、スプレッドシート操作、Python による分析、ブラウザ操作などを実行する。OSWorld 2.0（オフラインサブセット）で 72.6%（GPT-5.6 Sol は 65.7%）、Terminal-Bench Science 0.1 で 64.6%（Fable 5.1 は 52.6%）と報告されている [3,11]。

数学成果における「人間とモデルの寄与の切り分け」については、8 月 1 日の発表本文自体が説明している。すなわち「数学的論証そのものは我々のシステムが生成した」うえで、「論証はその後、同じモデルとともに人間によって原稿にまとめられ、その後モデルが各論証を Lean 証明書に形式化した」[6,9]。要するに、数学的アイデアはモデル由来、原稿化と形式化は人間とモデルの協働、正しさへの責任は OpenAI が負う、という構造である。

なお「ルートエージェントが問題を分割し、サブエージェントが数時間から数日並列稼働する」というマルチエージェント構成の説明は二次報道・アナリスト記事に見られる枠組みであり [11]、8 月 1 日の公式発表本文には明記されていない（付記参照）。

2.4 第三者評価と論争

【報道ベース】Lean 検証は「論理的整合性」を保証するが、「問題設定の忠実さ・新規性・重要性」までは保証しない、との指摘が複数の専門家・媒体からなされている [10,12]。

より深刻なのは、引用漏れをめぐる研究不正疑惑である。Scientific American（Joseph Howlett 記者、2026 年 8 月 6 日掲載、8 月 7 日・17 日に編集注記で改訂）は、数学者 Stephen Miller 氏（Yeshiva 大学）と Francesco Fournier-Facio 氏（ケンブリッジ大学）の指摘を報じた [12]。Miller 氏は、球充填の証明が自身の 2016 年論文の議論を出所を明示せず用いたとし、先行研究を意図的に無視する組織的な行為であって研究不正を指し示すと批判した [12,13]。Fournier-Facio 氏は、非ソフィック群の結果が 2016 年と 2019 年の 2 論文のアイデアを結合したものだと指摘しつつ、詐欺ではなく誇張の問題であるとし、正確性を軽視する巨大な PR 装置の存在を問題視した [12]。

2019 年論文の共著者 Andreas Thom 氏（TU Dresden）は MathOverflow で結果を「独創的でありながら初等的」と評した [12]。OpenAI は「結果の正しさに責任を負い、人間の数学者に一般に期待される基準を満たしている」との立場を示し、原稿を 8 月 6 日に更新（オリジナル版へのリンクを併記）、「10 年以上進展がなかった」という当初の表現も修正した [12,13]。

査読については、2026 年 9 月初旬時点で 10 件のいずれも従来の学術査読を通過していない [10,13]。肯定的評価としては、erdosproblems.com を運営する Thomas Bloom 氏（マンチェスター大学。2025 年 10 月に OpenAI の誤った主張を論駁した人物）が結果を「大きなニュース」と評し、単位距離予想の反例より重要だと述べている [10]。ただし 2026 年 5 月の Erdős 単位距離予想の反例が約 9 名の外部数学者の検証を経ていたのに対し、8 月の 10 問セットは同種の外部レビューを伴っていない点も指摘されている [10,13]。

2.5 安全ガバナンスと監視可能性

【確認済み事実】OpenAI の Preparedness Framework において、Astra はサイバーセキュリティで初めて「Critical」判定を受けた。人間の逐次指示なしに堅牢なシステムの未知の脆弱性を発見・悪用できる水準であり、評価中に 2 件のゼロデイ脆弱性を発見・開示した [1,5]。公開版は高度なサイバータスク（概念実証コードの生成など）を拒否するよう訓練され、より緩い安全策は「OpenAI Daybreak」プログラムを通じた審査済み組織に限定される [1,5]。ExploitBench で 100%、SRE-Bench で 88.0%（初回試行）が報告されている [5]。生物・化学は「High」に留まり、システムカードは「Critical として扱う必要はないが High の安全策を維持すべき」と結論している [1]。

チーフサイエンティスト Jakub Pachocki 氏は、モデルの能力向上に伴い思考連鎖の監視可能性（monitorability）が低下し、より高性能なモデルはより少ない言語トークン、あるいはトークンなしで難しいタスクを遂行できると述べており、Astra の監視可能性が GPT-5.6 Sol より低下したことを OpenAI 自身が認めている [1]。自律性の向上とガバナンスのトレードオフを示す重要な論点である。社長 Greg Brockman 氏は公開時に「AGI 時代へようこそ」と述べ、個人的にはすでに到達していると考えると発言したが、同時に AGI の定義は曖昧だとも認めている [3]。これは検証済み事実ではなく、個人的見解・マーケティングとして扱うべきである。

3. Claude Fable 5.1 の科学研究能力の実態

3.1 位置づけと提供形態

【確認済み事実】Claude Fable 5.1 と Claude Mythos 5.1 は 2026 年 9 月 1 日に公開された。両者

は同一の基盤モデルであり、安全策の水準のみが異なる[2]。Fable 5.1 は一般提供、Mythos 5.1 はサイバーセキュリティおよびライフサイエンス向けの信頼済みアクセスプログラム（Cyber Verification Program／Life Sciences Verification Program。米政府と連携、当初は米国組織に限定）に限られる[2,15]。Fable は「Covered Model」に指定され、追加のデータ保持・安全審査・アクセスポリシーが適用される[15]。価格は入力100万トークンあたり10ドル、出力50ドルとAstraと同一で、キャッシュ読み取りは75%値下げされた（100万トークンあたり0.25ドル）[2,16]。

Fable 5.1 は二重用途（サイバー・ライフサイエンス）のブロッキング分類器を内蔵し、拒否時には HTTP 200 で stop_reason: "refusal" を返す。生物学分野の安全策は良性の初等生物学・医療質問での発火が 85%減少し、サイバー分野の誤検知は約 60%減少したとされる[15]。ペネトレーションテスト、エクスプロイト生成、バイナリ脆弱性スキャンなどは Opus モデルへ転送される[15]。また EU AI Act 透明性行動規範（2026 年 7 月署名）に基づき、2026 年 8 月 2 日以降のモデル出力に不可視の電子透かしを付与し、検出 API を限定プレビュー提供している[15]。

3.2 科学的成果（Anthropic 自己報告）

分子設計。Mythos 5.1 はオープンソースのタンパク質設計・折り畳みツールを用いて高親和性バインダーを設計し、外部組織（Adaptyv Bio、Twist Bioscience）でウェットラボ検証を受けた。3 標的で Adaptyv Bio コンペ最優秀設計の 10 倍の結合親和性、12 標的で約 50%のヒット率が報告されている[2,15]。先行する 8 月 18 日の記事では、単一標的モードで RBX1 に対しヒット率 40%（コンペ参加者は3.7%）、最良バインダーの解離定数3.9 nM（人間の優勝者は45 nM）、1,320 設計中 354 が結合確認、15 標的中 14 で成功が報告されており[21]、同分野のヒット率は 10〜15%が典型とされる[17]。

計算解析。Fable 5.1 は NASA マゼラン探査機の約 30 年前のレーダーデータから金星の 3 分の 1 について高解像度標高マップを作成し（分解能 10〜20km から 2〜3km へ、高さ精度は最大 25%改善）、Creative Commons ライセンスで Zenodo に公開した。NASA VERITAS／ESA EnVision ミッションを見据えたものである[2,18]。計算生物学では、Mythos 5.1 がカスタム GPU カーネルにより 7 つのオープンソース深層学習モデルを最大 2.5 倍高速化し、ゲノムワイド解析の GPU コストを 30〜60%削減した[2,15]。

分析化学。8 月 18 日の記事では、Opus 5 が NMR／LC-MS データを 23 分・19 分で並列解析し、契約ラボの結果（純度 96.4%対 96.33%）と一致、その過程で自らの誤りを検出・修正したと報告されている[21]。

数学。2026 年 8 月 10 日の記事によれば、未公開の研究版 Claude が、リーマン予想に関連する

下限（臨界線上に存在するゼータ関数の非自明零点の割合）を 41.6%から 67.2%に改善した。指示者は非数学者のスタッフ（Jarred Sumner 氏）であり、Claude は当初 650 のアイデアで失敗した後、約 36 時間で約 60 のサブエージェント、2,400 のシェルコマンド、数百の Python スクリプト、3,100 万出力トークンを調整し、54 本の arXiv 論文を参照した。人間による詳細な数学的指導はなく、Anthropic の数学者 Levent Alpöge 氏と Ralph Furman 氏が検証し、外部専門家 Brian Conrey 氏と Dan Goldston 氏が精査、Lean 4 証明も作成された [22]。

3.3 第三者ベンチマークとパートナー証言

Terminal-Bench-Science 0.1 は、Stanford 大学主導・Terminal-Bench チーム（Laude Institute）が構築した、実際の科学研究計算ワークフロー70 タスク（生命・物理・地球・数学・工学）からなるベンチマークである [19]。OpenAI 公表値では Astra 64.6%、Fable 5.1 52.6%[11]、Snorkel AI のリーダーボードでは Astra 65.4%、従来トップは Claude Opus 5 の 30.0%である [20]。Fable 5.1 の 52.6%は Anthropic 自己報告値であり [16]、公開リーダーボード上の最高値は本稿執筆時点でも Opus 5 の 30.0%である [18,20]。同ベンチマークは「客観的に検証可能なタスクに限定し、仮説生成や文献レビューのようなオープンエンドのタスクは対象外」とし、「科学者を創造的・知的作業で置き換えるのではなく、信頼できる科学アシスタントへの進歩を促す」ことを設計思想として明示している [19]。

パートナー企業の証言としては、Ramp の「38 時間の無人ランでラベルアーティファクトを診断し、6 つの並列実験を起動、翌朝に結果と次のステップを提示」、Rakuten Medical の「他の 3 つのフロンティアモデルが承認した臨床研究プロジェクトで見逃されていたギャップを発見し新仮説を提案、切り捨てていたデータセットを新たな研究方向に転換」、iGent の「18 か月取り組んできたグランドチャレンジ級の問題で実質的進展」などが公開されている [2]。ただしこれらは Anthropic が選定・公開した証言であり、独立検証ではない。

4. 比較考察

4.1 8 軸による比較

以下の表は、前節までの事実を 8 つの比較軸で整理したものである（【分析】）。

比較軸	GPT-6 Astra（OpenAI）	Claude Fable 5.1（Anthropic）
(a) 自律性	数学：論証はモデルが生成、原稿化・Lean 形式化は人間との協働。コンピュータ操作	数学：約 60 サブエージェント・約 36 時間を調整、人間の詳細な数学的指導なし。ML

比較軸	GPT-6 Astra（OpenAI）	Claude Fable 5.1（Anthropic）
	型エージェント（OSWorld 2.0 で 72.6%）	実験で 38 時間の無人ラン（Ramp 証言）
(b) 研究ループ	仮説→証明→形式検証。実験工程を含まない	設計→外部ウェットラボ検証。標的選定・データ選択は Anthropic 主導
(c) 対象領域	数学・理論計算機科学が中心。ライフサイエンスは別モデル（GPT-Rosalind）	実験科学・企業 R&D への実装を重視（タンパク質設計、分析化学、惑星科学、計算生物学）
(d) 検証可能性	Lean 4 で機械検証可（sorry ゼロ）。査読未通過、引用漏れの指摘あり	外部ウェットラボ検証あり。決定的データは非公開、外部数学者による精査あり
(e) ツール・実験インフラ連携	科学ソフトウェアの直接操作。GPT-Rosalind＋Codex Life Sciences プラグイン（50 超のツール接続）	オープンソース設計ツール、外部ラボ（Adaptyv Bio・Twist Bioscience）、成果の Zenodo 公開
(f) 安全ガバナンス	Preparedness Framework：サイバーで初の Critical、生物・化学は High。Daybreak による審査済み組織への限定緩和	Fable に二重用途ブロッキング分類器、Mythos は検証プログラム限定。EU 行動規範に基づく出力透かし
(g) 提供形態	ChatGPT／API／Azure／Bedrock で段階展開。Pro 版あり。入力 10 ドル・出力 50 ドル（100 万トークン）	全プラットフォームで即日提供。Mythos は信頼済みアクセス限定。価格は Astra と同一、キャッシュ読み取り 75%値下げ
(h) 第三者評価	数学者の実名評価（Bloom 氏の高評価と研究不正疑惑が併存）。Terminal-Bench-Science 64.6%（自己報告）／65.4%（Snorkel）	外部ラボ・外部数学者の検証。企業パートナー証言（Anthropic 選定）。Terminal-Bench-Science 52.6%（自己報告）

4.2 戦略的分岐 ― 「発見型」と「実装型」

【分析】両社のローンチ資料と科学関連記事の内容配分を比較すると、OpenAI は数学・理論計算機科学における「発見」（形式検証で正しさを証明できる成果）を旗印にし、Anthropic は実験科学・企業 R&D への「実装」（ウェットラボ検証、分析化学、計算生物学、企業パートナー証言）を前面に出す傾向が明確である。両社ともリーマン予想関連の数学に取り組んでいる点は共通するが、OpenAI は「証明の生成」を、Anthropic は「長時間エージェント調整による既存下界の改善」を、それぞれ見せ方の中心に据えている。

4.3 検証可能性の非対称性

【分析】Astra の数学的主張は Lean 4 で機械検証でき（sorry ゼロ）、その意味で「正しさ」の検証可能性は極めて高い。しかし査読は未通過であり、著名数学者から引用漏れと新規性の誇

張が指摘された。Fable／Mythos の生物学的成果は外部ウェットラボ検証を受けたが、決定的データ（試行した標的の総数など）は非公開で、Anthropic の自己報告に依存する。すなわち、Astra は「正しいが新しいかは未確定」、Fable は「有効そうだが再現材料が不足」という、異なる種類の検証課題を抱えている。両者に共通するのは、「人間の寄与の切り分け」と「独立した再現」が未解決な点である。

5. 判定 — 両モデルは「サイエンスエージェント化」したのか

5.1 「AI Scientist」の定義と判定基準

学術的な「AI Scientist」「AI co-scientist」の定義を参照すると、Sakana AI の「AI Scientist」は着想・実装・実験・論文執筆・査読までの研究ライフサイクル全体の自動化を標榜し[23]、Google DeepMind の「AI co-scientist」は多エージェントによる仮説生成・実験設計の支援を掲げる[28]。ここから導かれる判定基準は、「仮説生成→実験設計→実行→分析→結論→再計画」の閉ループを、人間の介在なしにどこまで回せるか、である。

なお Sakana AI Scientist に対する独立評価（Beel, Kan & Baumgart）は、文献レビューにおける新規性判定が貧弱で確立された概念を新規と誤分類することが多く、実験実行では 42%がコーディングエラーで失敗したと報告しており[23]、「フルライフサイクル自律」の主張には慎重な検証が必要であることを示す先例である。

5.2 両モデルの判定

【分析】GPT-6 Astra は、数学領域で「発見の一部（論証生成）」を担うが、原稿化・形式化は人間との協働であり、実験科学の閉ループは未達である。位置づけは「特定領域における準自律の発見エンジン＋汎用の高度コパイロット」となる。

Claude Fable 5.1 は、長時間無人ラン（38 時間）や約 60 エージェントの調整、外部ウェットラボ検証にまで到達しているが、実験の実行・検証は Anthropic や外部組織の人間とインフラが担い、標的選定・データ選択にも人間が関与する。位置づけは「実装寄りの長時間自律エージェント＋実験科学コパイロット」となる。

いずれも完全な「AI Scientist」ではなく、「サイエンスエージェント化」は進行中だが未完成、というのが証拠に基づく判定である。Terminal-Bench-Science 自体が「仮説生成・文献レビューは対象外、科学者の置き換えではなく信頼できるアシスタントへの進歩を測る」と定義していること[19]は、現時点の業界コンセンサスが「エージェント＝完全自律の科学者」ではなく「検

証可能な計算ワークフローの実行者」にあることを示唆している。

6. 知財・企業研究への含意

AI 生成発明の発明者性。日本では東京地裁判決（令和 6 年 5 月 16 日、いわゆる DABUS 事件）を含め、日英米豪の裁判所は「発明者は自然人に限られる」との解釈でおおむね一致している [25]。特許庁は 2025 年 1 月に有識者会議を設け、人が課題設定・アイデアを出し生成 AI が化学式等を創作した場合の発明者性や、発明に用いた AI の開発者を共同発明者とみなしうるかといった論点の検討を開始したと報じられている [26]。立法論は流動的であり、2026 年時点で確定した改正内容は本調査では確認できていない。

R&D 部門への示唆。Astra／Fable の成果は「早期発見・計算集約部分の加速」に強みがあり、実験の実行・検証・意思決定は依然として人間が担う。日本企業は、(1)AI が生成した発明の発明者性・進歩性の切り分け、(2)研究プロセスの監査可能性（ログ）の確保、(3)二重用途規制（Mythos／Daybreak 型の限定アクセス）への対応、を先行して整備する必要がある。とくに AI が生成した化学式・タンパク質配列・数学的構成を含む成果を出願する際は、人間の課題設定と寄与を文書化しておくことが、現行の「発明者は自然人」という判例下でのリスク低減につながる。

7. 結論と今後の注視点

本稿の問い「GPT-6 Astra はサイエンスエージェント化したか」への答えは、「数学・理論計算機科学という特定領域で発見の一部を担う準自律エンジンへと進化したが、完全な自律サイエンスエージェントではない」である。Fable 5.1 との対比では、Fable を「実装型・長時間エージェント」、Astra を「発見型・形式検証」と類型化することで、両社の戦略の違いが最も明瞭に見える。

判定を上方修正すべき閾値は次の三つである。(1)10 件の数学結果のいずれかが査読付き学術誌に掲載され、新規性が確定した場合。(2)Astra／Fable が人間の標的選定・データ選択なしに「仮説→実験→検証→再計画」の閉ループを、独立再現可能な形で回した事例が第三者により検証された場合。(3)OpenAI／Anthropic が「autonomous scientist／AI researcher」を公式に製品として位置づけた場合。とくに(3)については OpenAI が 2026 年 9 月と 2028 年をマイルストーンに掲げているため [27]、今後 12〜24 か月が注視期間となる。

付記：本稿で確認できていない事項

- 内部版 Astra と公開版 GPT-6 Astra の厳密な性能差分。
- 数学ランにおけるマルチエージェント稼働（数時間〜数日）の詳細は二次報道の枠組みであり、8 月 1 日の公式発表本文には明記されていない。
- OpenAI が 8 月 6 日の原稿更新で具体的にどの引用を追加したかを個別に列挙した報道は確認できなかった。
- Fable／Mythos のタンパク質設計成果の決定的データ（試行標的の総数など）は非公開。
- ベンチマーク値は提供者の自己報告を含み、ハーネス・設定の差により単純比較には注意を要する。
- 2026 年 8 月 18 日の Anthropic 記事は Mythos Preview／Opus 5 による成果であり、9 月 1 日の Fable／Mythos 5.1 発表はこれを更新・拡張したものである。
- 日本の特許法改正は検討段階であり、確定内容は未確認。

参考文献

※「URL 未確認」と付したものは、本調査において URL を直接取得・確認できなかった資料である。

[1] OpenAI「Safety overview: GPT-6 Astra」（GPT-6 Astra システムカード）、2026 年 9 月 3 日。https://openai.com/index/safety-overview-gpt-6-astra/

[2] Anthropic「Introducing Claude Fable 5.1 and Claude Mythos 5.1」、2026 年 9 月 1 日。https://www.anthropic.com/claude-fable-and-mythos-5-1

[3] AI Weekly「OpenAI launches GPT-6 Astra, Brockman invokes 'AGI era'」、2026 年 9 月。https://aiweekly.co/alerts/openai-launches-gpt-6-astra-brockman-invokes-agi-era

[4] Notebookcheck「OpenAI releases GPT-6 Astra AI: The 'Terminator' version」、2026 年 9 月。https://www.notebookcheck.net/OpenAI-releases-GPT-6-Astra-AI-The-Terminator-version.1389198.0.html

[5] Codersera「Why GPT-6 Astra Was Held Back: The Critical Cyber Finding」、2026 年 9 月。https://codersera.com/blog/gpt-6-astra-safety-cyber-capabilities-2026/

[6] OpenAI「Ten advances in mathematics and theoretical computer science」、2026 年 8 月 1 日（8 月 6 日更新）。原稿（249 頁）および Lean 4 証明は GitHub「openai/ten-proofs」で公開。※URL 未確認（内容は文献 7〜11 経由で確認）

[7] The Decoder「OpenAI announces its "next major model" Astra by dropping ten previously unsolved math solutions」、2026 年 8 月。https://the-decoder.com/openai-announces-its-next-major-model-astra-by-dropping-ten-previously-unsolved-math-solutions/

[8] Pondero「OpenAI reveals Astra through ten machine-verified open-problem proofs, headlined by first non-sofic group construction in 27 years」、2026 年 8 月 2 日。https://pondero.ai/news/2026-08-02-openai-astra-math-proofs/

[9] SiliconANGLE「OpenAI's Astra solves 10 long-open math problems and publishes the proofs」、2026 年 8 月 2 日。https://siliconangle.com/2026/08/02/openais-astra-solves-10-long-open-math-problems-publishes-proofs/

[10] Tech Times「OpenAI's Astra Solves Ten Decade-Old Math Problems With Machine-Checkable Lean Proofs」、2026 年 8 月 2 日。https://www.techtimes.com/articles/322710/20260802/openais-astra-solves-ten-decade-old-math-problems-machine-checkable-lean-proofs.htm

[11] Tech Insider「OpenAI Astra Solves 10 Open Math Problems 2026」、2026 年。https://tech-insider.org/openai-astra-solves-10-open-math-problems-2026/

[12] Howlett, J.「OpenAI's latest math breakthroughs commit research misconduct, experts say」Scientific American、2026 年 8 月 6 日（8 月 7 日・17 日に編集注記で改訂）。Yahoo! News 転載版：https://www.yahoo.com/news/science/articles/openai-latest-math-breakthroughs-commit-184300525.html（Scientific American 原記事の URL は未確認）

[13] AI Weekly「OpenAI's Astra math proofs draw research misconduct claims」、2026 年 8 月。https://aiweekly.co/alerts/openais-astra-math-proofs-draw-research-misconduct-claims

[14] Drug Patent Watch「GPT-Rosalind: What OpenAI's Life Sciences Model Actually Does to Drug Development」、2026 年。https://www.drugpatentwatch.com/blog/gpt-rosalind-what-openais-life-sciences-model-actually-does-to-drug-development/

[15] Atomic Agent「Claude Fable 5.1」、2026 年 9 月。https://atomicagent.io/blog/claude-fable-5-1/

[16] MarkTechPost「Anthropic Releases Claude Fable 5.1 and Claude Mythos 5.1: 52.6% on Terminal-Bench-Science and 75% Cheaper Cache Reads」、2026 年 9 月 1 日。https://www.marktechpost.com/2026/09/01/anthropic-releases-claude-fable-5-1-and-claude-mythos-5-1-52-6-on-terminal-bench-science-and-75-cheaper-cache-reads/

[17] McDonald, L.「Technology: Anthropic's Claude Designs Protein Binders Validated in Lab Tests」、2026 年。https://www.laynemcdonald.com/post/technology-anthropic-s-claude-designs-protein-binders-validated-in-lab-tests

[18] Kingy AI「Can Claude Fable 5.1 and Mythos 5.1 Do Real Science?」、2026 年 9 月。https://kingy.ai/blog/claude-fable-mythos-5-1-scientific-research/

[19] Terminal-Bench「Terminal-Bench-Science 0.1」（Stanford 大学・Laude Institute）。https://www.tbench.ai/news/terminal-bench-science-0-1

[20] Snorkel AI「Terminal-Bench-Science Leaderboard」。https://snorkel.ai/leaderboard/terminal-bench-science/

[21] Anthropic「How Claude is accelerating protein design and analytical chemistry」、2026 年 8 月 18 日。※URL 未確認（既往調査で英語版・日本語版 PDF に基づき内容確認済み）

[22] Anthropic「Learning more about Claude's mathematical capabilities」、2026 年 8 月 10 日。※URL 未確認

[23] Beel, J., Kan, M.-Y., Baumgart, M.「Evaluating Sakana's AI Scientist for Autonomous Research: Wishful Thinking or an Emerging Reality Towards 'Artificial Research Intelligence' (ARI)?」arXiv:2502.14297v3、2025 年 10 月改訂。https://arxiv.org/abs/2502.14297 ※本調査で直接取得しておらず、題名・URL は要確認

[24] Gromov, M.「Endomorphisms of symbolic algebraic varieties」Journal of the European Mathematical Society, 1(2), 109–197, 1999.

[25] 東京地方裁判所判決 令和 6 年 5 月 16 日（令和 5 年（行ウ）第 5001 号、特許出願却下処分取消請求事件、いわゆる DABUS 事件）。裁判所ウェブサイト掲載。※URL 未確認

[26] 日本経済新聞報道「特許庁、AI 発明の発明者性について有識者会議で議論開始」（趣旨）、2025 年 1 月 17 日前後。※見出し・URL 未確認

[27] OpenAI ライブ配信（Sam Altman・Jakub Pachocki）「AGI 研究ロードマップ」、2025 年 10 月 28 日。インターン級研究助手（2026 年 9 月）・本格的 AI 研究者（2028 年）の目標を提示。※URL 未確認

[28] Google Research／DeepMind「Accelerating scientific breakthroughs with an AI co-scientist」、2025 年 2 月。https://research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist/ ※本調査で直接取得しておらず、URL は要確認