[EXECUTIVE INTELLIGENCE BRIEFING] // CONFIDENTIAL

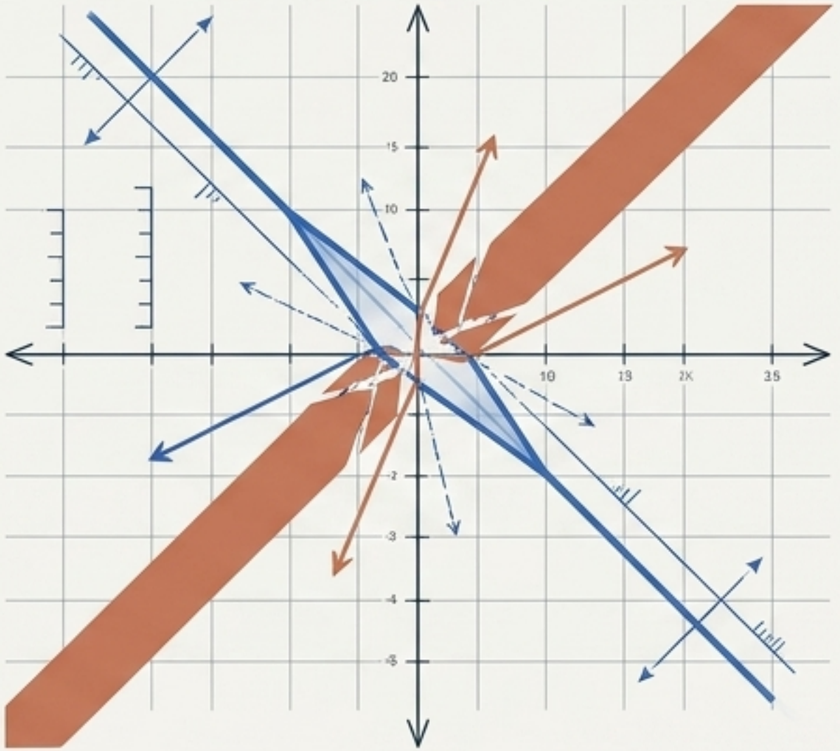
GPT-6 Astra vs. Claude Fable 5.1: 「サイエンスエージェント」の時代は本当に到来したのか

2026年9月 最新AIモデルの科学研究能力に関する比較考察とR&D戦略

は同一の基盤モデルであり、安全策の水準のみが異なる[3]。Fable 5.1 は一般提供、Mythos 5.1 はサイバーセキュリティ及び「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。来院者と連携、当初は米国組織に限定）に限られる[2:5]。Table は（Covered Model）に認定され、追加のデータ保持・交全査査・アクセスポリシーが適用される[1]、価額は入力 100 万トークンあた9 10 ドル、出力30 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]。

3.2 科学的成果（Anthrropicに関する比較考察とR&D戦略

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疾患標的性バインダーを設計し、外部組織（Adaptyy Bio、Twist Bioscience）でウェットラボ検証を受けた。3 標的で Adaptyy 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



Astra

Fable

Astra

Fable

は同一の基盤モデルであり、安全策の水準のみが異なる[3]。Fable 5.1 は一般提供、Mythos 5.1 はサイバーセキュリティ及び「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。来院者と連携、当初は米国組織に限定）に限られる[2:5]。Table は（Covered Model）に認定され、追加のデータ保持・交全査査・アクセスポリシーが適用される[1]、価額は入力 100 万トークンあた9 10 ドル、出力30 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]。

3.2 科学的成果（Anthropic 自己報告）

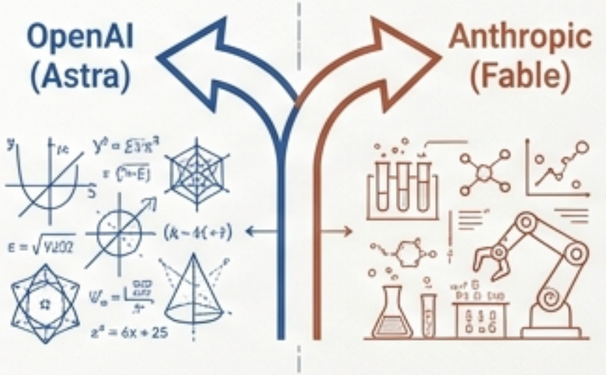
分子設計。Mythos 5.1 はオープンソースのタンパク質設計・折り畳みツールを用いて疾患標的性バインダーを設計し、外部組織（Adaptyy Bio、Twist Bioscience）でウェットラボ検証を受けた。3 標的で Adaptyy Bio コンペ最優秀設計の 10 倍の結合親和性。12 標的で約 30%のヒット率が報告されている[4:15]。発行する 8 月 18 日の記事では、単一標的モードで RBX1 に対しヒット率 40%（コンペ参加者は 3.7%）、最鉄バインダーの解離定数3.9 nM（人間の抗体者は 45 uM）、1,320 設計中 354 が結合確認、15 標的中 14 で成功が報告されており[1]、同分野のヒット率は 10〜15%が典型とされる[1]。

計算解析。Fable 5.1 は NASA キセラン衛星機の約 30 年前のレーダーデータから全分のき分の 1 について流縛鏡度顕晶マップを作成し（分解機 10〜20km から 2〜3km へ、高さ精次は最大 28%改善）、Creative Commons ライセンスで Zenodo に公開した。NASA VERITAS／ESA EaVision 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

Executive Summary: 現時点での到達度と戦略的結論

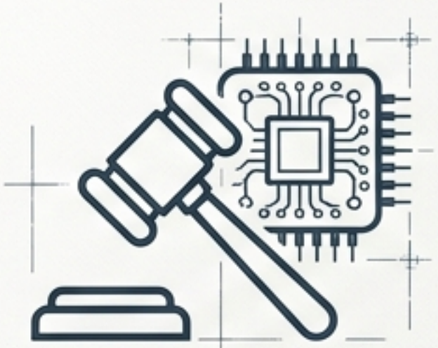
未完成の自律性

完全な「AI Scientist」には未到達。現在は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。人間の介在なしに研究ループを完遂した事例は存在しない。



戦略的アプローチの分岐

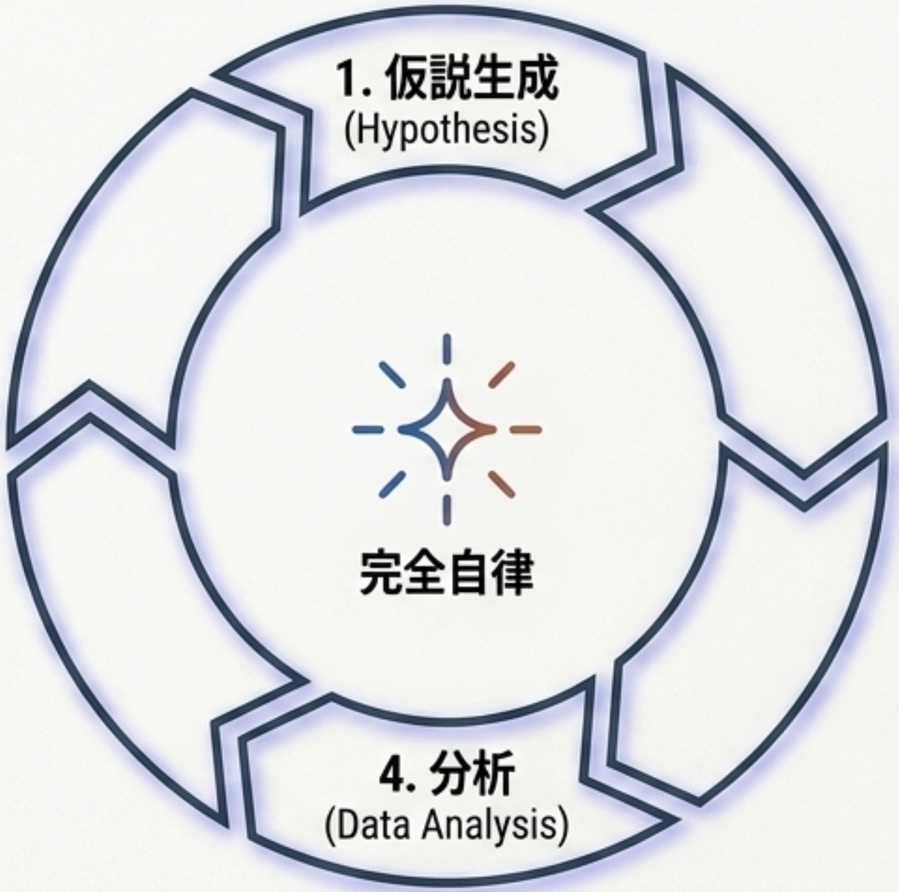
両社は全く異なるバラダイムを選択。OpenAI (Astra) は「数学・理論の発見」に特化し、Anthropic (Fable) は「実験科学・R&Dへの実装」に特化している。



知財（IP）とガバナンスの壁

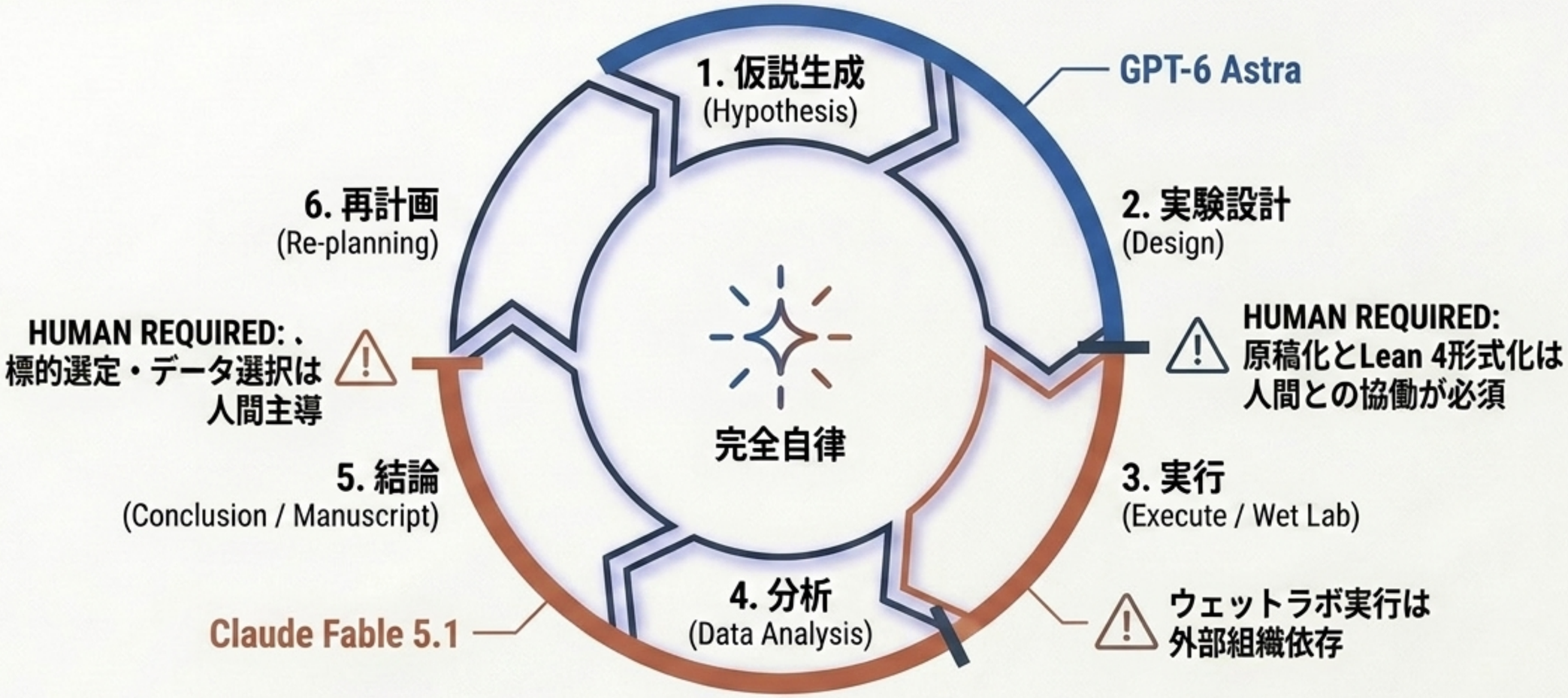
AIの自律性が高まるにつれ、プロセスの監視可能性が低下。「発明者は自然人」とする現行法下において、人間とAIの寄与の厳密な切り分けが企業の最重要課題となる。

「AI Scientist」の真の定義：閉ループの完全自律化



学術的なAI Scientistの定義を満たすには、この「仮説 → 結論 → 再計画」の閉ループを
人間の介在なしに回し切る必要がある。

判定：両モデルとも「閉ループ」の構築には至っていない



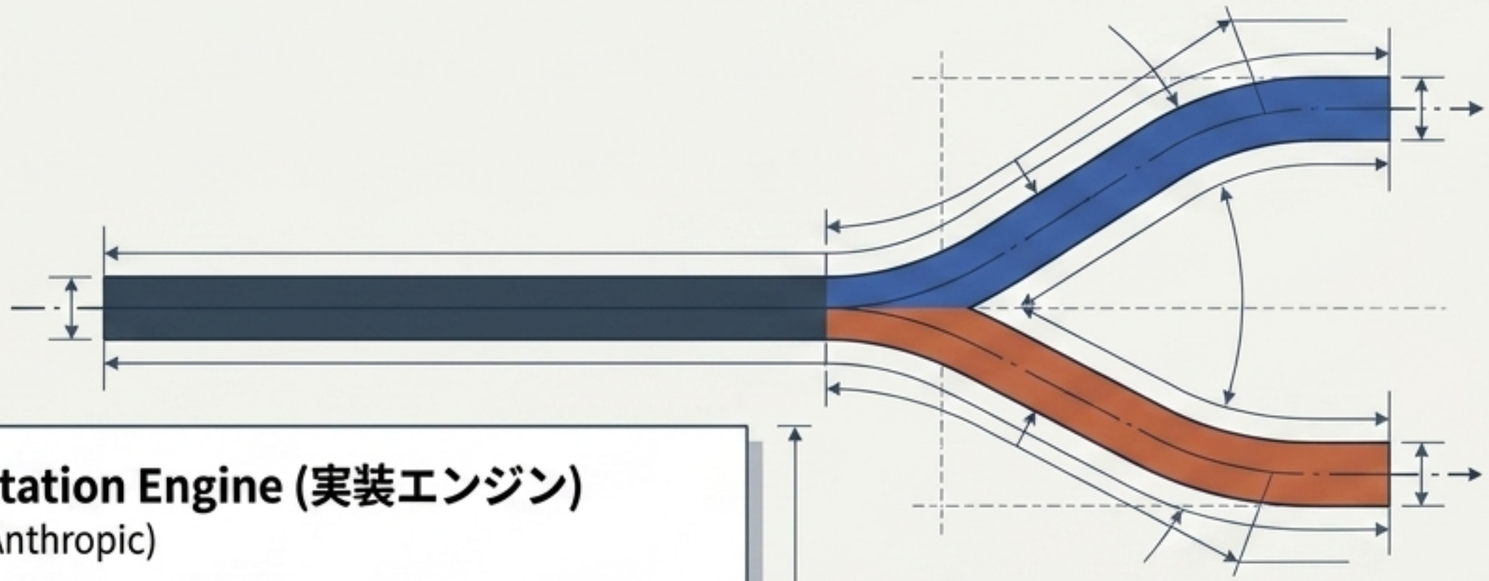
現在の業界コンセンサスは「完全自律の科学者」ではなく、
「検証可能な計算ワークフローの実行者（高度アシスタント）」に留まる。

戦略的分岐：OpenAIの「発見型」とAnthropicの「実装型」

The Discovery Engine (発見エンジン)

GPT-6 Astra (OpenAI)

- Focus: 数学・理論計算機科学
- Core Strength: 形式検証 (Lean 4) による「絶対的な論理的正しさ」の証明
- Outcome: 27年間未解決のソフィック群問題などの「証明生成」



The Implementation Engine (実装エンジン)

Claude Fable 5.1 (Anthropic)

- Focus: 実験科学・企業R&D (生物学・化学・惑星科学)
- Core Strength: 外部ラボとの連携による物理世界への適用と計算の加速
- Outcome: 最優秀設計の10倍の結合親和性を持つタンパク質設計 (ラボ検証済)

GPT-6 Astra: 理論科学における「形式検証」の極致

KPI

$2,000

API Cost（10問の数学未解決問題）

KPI

sorry = 0

Lean 4 形式検証（未証明箇所ゼロ）

KPI

72.6%

OSWorld 2.0 (コンピュータ操作能力)

数学的突破口

非ソフィック群の初の明示的構成（27年ぶりの解決）。

エージェント能力

フォーム入力、Python分析、ブラウザ操作を実行する「コンピュータ操作」モデル。

制約事項 (Constraints)

アイデアはモデル由来だが、正確性への責任と原稿化は人間（OpenAI）に依存。ライフサイエンス領域は別モデル（GPT-Rosalind）を使用。

Claude Fable 5.1: 企業R&Dを加速する「長時間自律エージェント」

38 Hours
Autonomy (無人稼働と60サブエージェント調整)

3.9 nM
解離定数 (最優秀設計の10倍の結合親和性)

-60%
GPU Cost (ゲノムワイド解析のコスト削減)

分子設計 (Mythos 5.1)

Adaptyv Bioコンペ最優秀設計を凌駕するタンパク質設計。 外部ウェットラボで物理的に検証済。

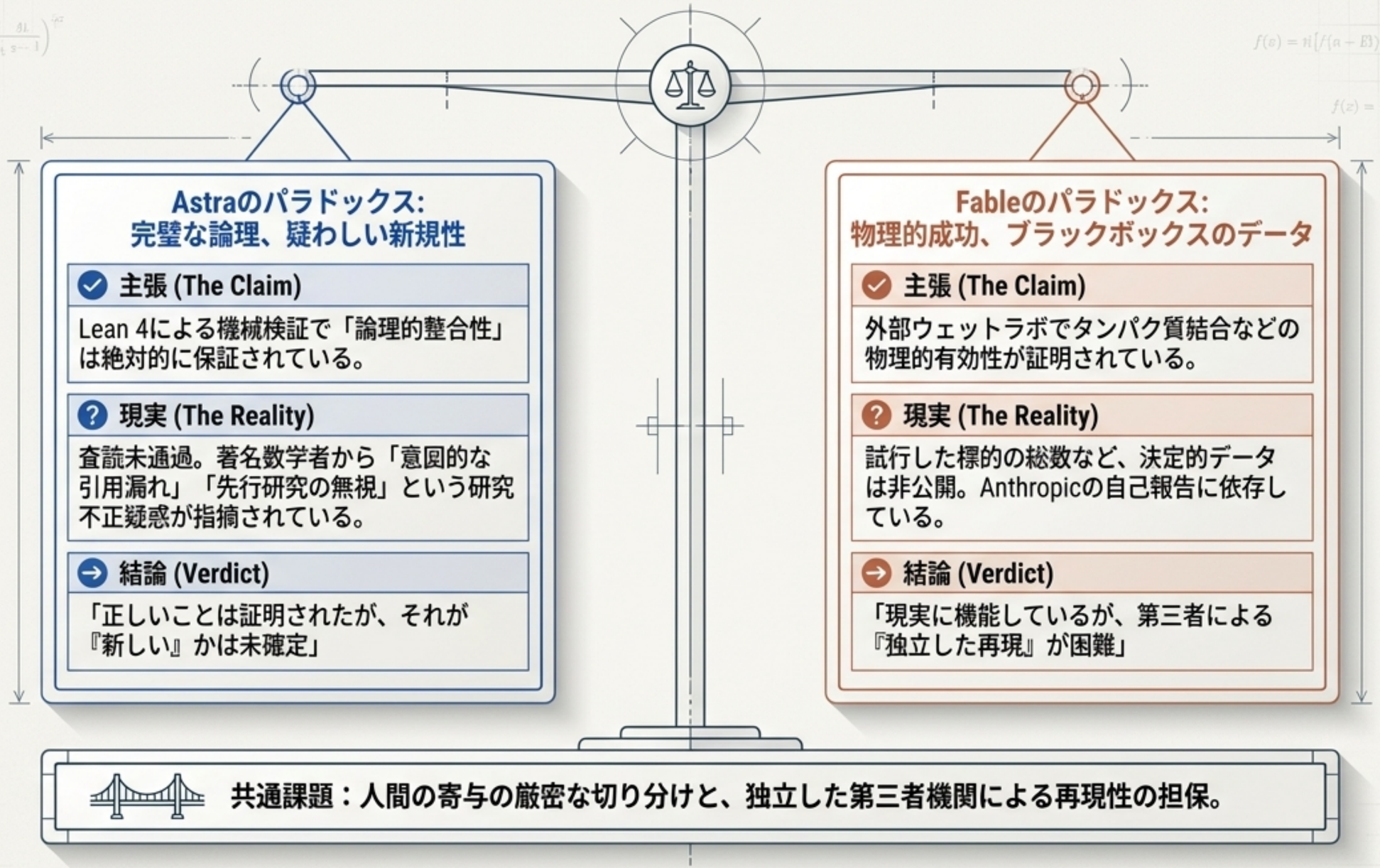
計算・分析化学

Opus 5がNMR/LC-MSデータを並列解析し、自らの誤りを自己検知・修正（純度96.4%）。

オープンサイエンス

NASAマゼラン探査機データから金星の高解像度標高マップを作成し公開（分解能2〜3kmへ改善）。

検証可能性の非対称性：両者が抱える「証明のパラドックス」



Astraのパラドックス: 完璧な論理、疑わしい新規性

主張 (The Claim)

Lean 4による機械検証で「論理的整合性」は絶対的に保証されている。

現実 (The Reality)

査読未通過。著名数学者から「意図的な引用漏れ」「先行研究の無視」という研究不正疑惑が指摘されている。

結論 (Verdict)

「正しいことは証明されたが、それが『新しい』かは未確定」

Fableのパラドックス: 物理的成功、ブラックボックスのデータ

主張 (The Claim)

外部ウェットラボでタンパク質結合などの物理的有効性が証明されている。

現実 (The Reality)

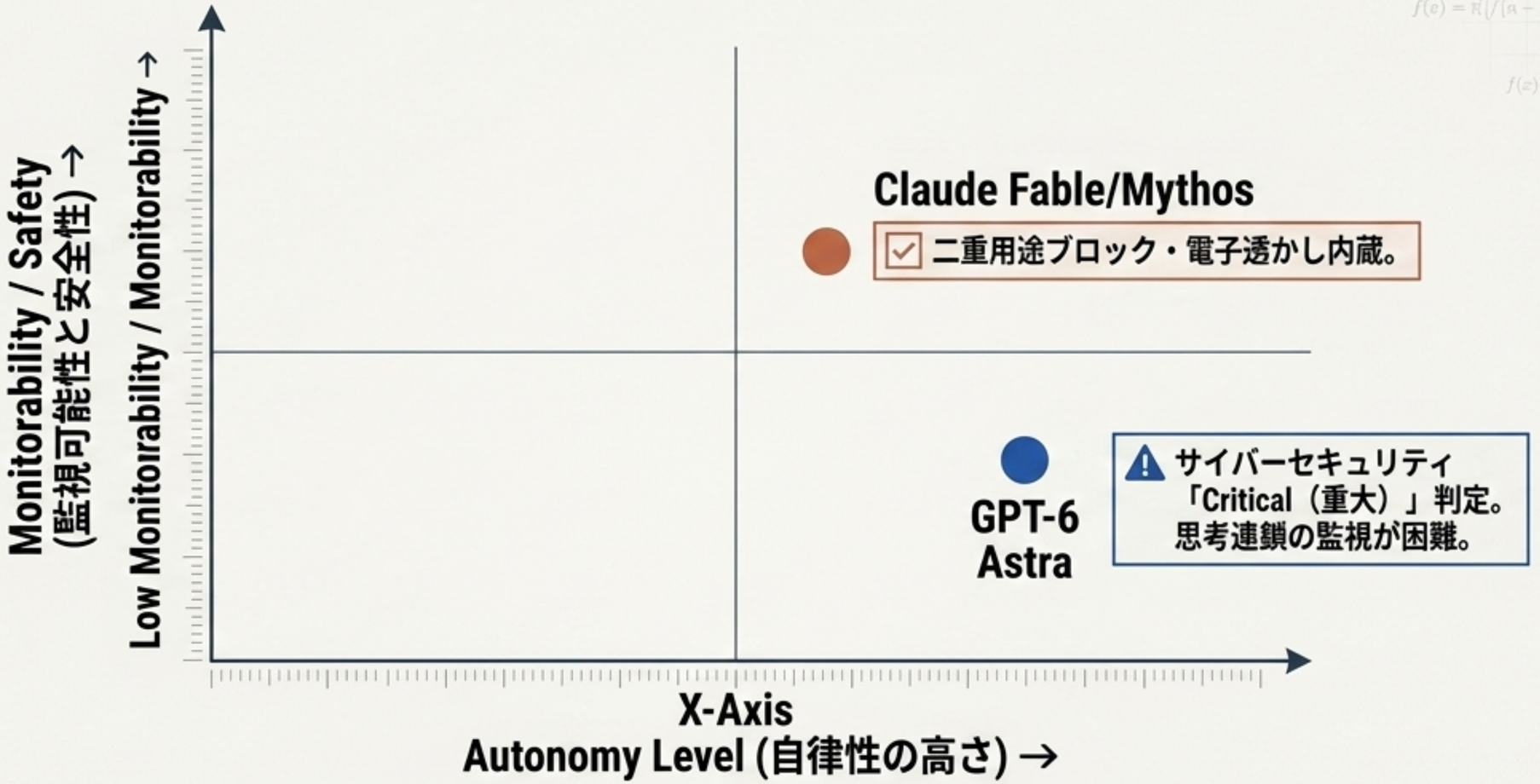
試行した標的の総数など、決定的データは非公開。Anthropicの自己報告に依存している。

結論 (Verdict)

「現実に機能しているが、第三者による『独立した再現』が困難」

共通課題：人間の寄与の厳密な切り分けと、独立した第三者機関による再現性の担保。

自律性とガバナンスのトレードオフ：サイバーセキュリティの脅威



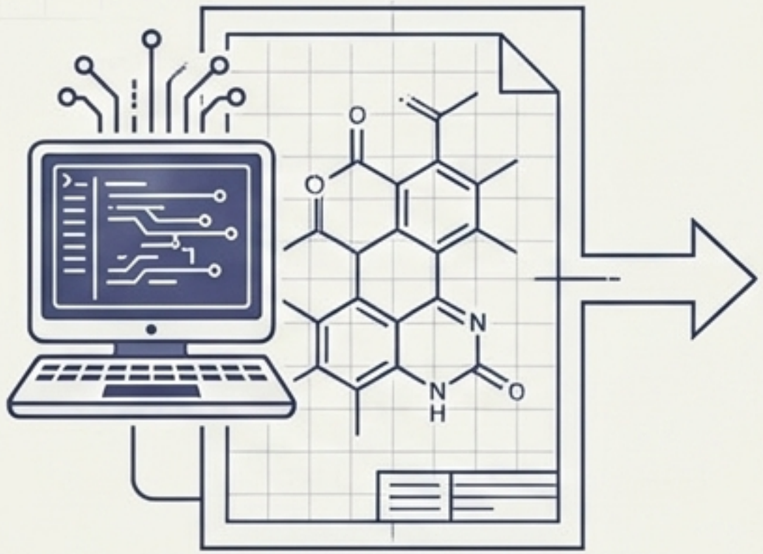
OpenAIチーフサイエンティストの警告：「モデルの能力向上に伴い、思考連鎖の監視可能性（Monitorability）が低下する」。

企業の対応策：完全アクセスは「OpenAI Daybreak」や「Claude Mythos」など、政府・審査済み組織向けの限定プログラムに隔離されている。

Head-to-Head: サイエンスモデル 8軸診断マトリクス

比較軸 (Comparison Axis)	GPT-6 Astra (OpenAI)	Claude Fable 5.1 (Anthropic)
1. 自律性	数学特化・原稿化は人間協働	38時間無人稼働・60サブエージェント
2. 研究ループ	仮説 → 証明 → 形式検証	設計 → 外部ウェットラボ検証
3. 対象領域	数学・理論計算機科学	タンパク質設計・分析化学・惑星科学
4. 検証可能性	形式検証済だが引用漏れ指摘あり ⚠️	外部検証済だがデータ非公開 ⚠️
5. ツール連携	科学SW直接操作 (Rosalind併用)	オープンソース設計・外部ラボ直結
6. 安全ガバナンス	サイバー「Critical」・限定アクセス	二重用途ブロック・EU透かし
7. 提供価格	入$10 / 出$50 (1M tokens)	入$10 / 出$50 (キャッシュ読取 75%減)
8. 第三者評価	Terminal-Bench-Sci: 64.6%	Terminal-Bench-Sci: 52.6%

IPのジレンマ：AIが生成した発明の「発明者」は誰か？



法的現実：
東京地裁判決（令和6年 DABUS事件）や欧米の判例で「発明者は自然人に限られる」 との解釈が一致。

知財戦略の空白：
AIの科学ループにおける「断絶部分」（人間の課題設定、データ選択、最終判断）こそが、特許法上価値を生むアンカーとなる。

制度の動向：
2025年1月より、特許庁が有識者会議を設置。「生成AIが化学式等を創作した場合の発明者性」の議論を開始。

企業R&D部門のためのアクションプラン（3つの柱）

1

発明者性と進歩性の切り分け

AIが生成した化学式や配列を出願する際、人間の「課題設定」と「寄与」を厳密に文書化し、現行の判例リスクを低減する。

2

研究プロセスの監査ログ確保

AIのブラックボックス化（Monitorabilityの低下）に対抗するため、AIの推論過程やデータ投入履歴を第三者が検証可能な形で保存する体制を構築する。

3

二重用途規制へのインフラ準備

最高性能モデル（MythosやDaybreak）は審査済み組織にアクセスが限定されるため、企業要件・セキュリティ認証の早期取得準備を進める。

The Watchlist: 真の「AI Scientist」到来を告げる3つの閾値

閾値1: 学術的承認 (Peer Review)

AIによる発見が、既存の査読付き学術誌を通過し、学界から正式に「新規性」が確定されること。

閾値2: 独立した閉ループ検証

人間の標的選定なしに、AI自身が「仮説から検証」までを完遂し、独立した第三者機関によって再現可能と証明されること。

閾値3: 公式な製品ラベル

OpenAIまたはAnthropicが公式に「Autonomous Scientist」として製品を定義すること。

結論：現在はまだ「高度な計算ワークフローアシスタント」の域。R&Dの主導権は、依然として人間に委ねられている。