# 主要 3AI モデルの評価比較考察

## —— Artificial Analysis 個別評価に基づく GPT-6.2 Astra / Claude Fable 5.2 / Gemini 4 Carbon ——

*作成日：2026 年 11 月 18 日　情報基準日：2026 年 11 月 17 日（Artificial Analysis 公表値）*

**Perplexity**

Artificial Analysis（以下、AA）の最新評価（Intelligence Index v5.0.1、2026 年 11 月 17 日公表）では、Claude Fable 5.2（65 点）＞ GPT-6.2 Astra（62 点）＞ Gemini 4 Carbon（60 点）の順位となり、GPT-6.2 Astra は旧 GPT-6 Astra（57 点）から 5 点伸ばして第 2 位に浮上した[1][5]。

【前提とするデータの範囲】AA の比較ページで公開されている個別評価の詳細数値は、Fable 5.2 対 Carbon、および Fable／Carbon 対旧 GPT-6 Astra のものが中心であり、GPT-6.2 Astra の項目別数値は AA のリリース記事ならびに第三者レポートで部分的に開示されているにとどまる[1][4]。以下の対比表は判明している公表値に基づく。

### 1. 個別評価の対比（上位設定）

| 評価項目 | Fable 5.2 (max) | GPT-6.2 Astra (max) | Carbon (high) |
|---|---|---|---|
| Intelligence Index | 65 | 62 | 60 |
| Terminal-Bench-Science（科学研究ワークフロー） | 73.3% | 78.1%（1 位） | 64.8% |
| Terminal-Bench 4.0（端末操作） | 59.6% | 64.1% | 65.2%（1 位） |
| AA-Briefcase（実務成果物, Elo） | 1785（1 位） | 1644 | 1611 |
| HLE-Diamond（知識・推論） | 80.2%（1 位） | 2 位 | 69.1% |
| CritPt v2（物理推論, hard） | 76.5%（1 位） | 2 位 | 69.0% |
| AA-Omniscience（事実知識正解率） | 67（1 位） | 2 位（ハルシネーション率 59%） | 54 |
| AA-3D（3D 体験生成, Elo） | 1271 | 1318（1 位） | 未公表 |

| | | | |
|---|---|---|---|
| AA-Robotics（現実世界の理解・計画） | 75 | 79（1 位） | 未公表 |
| Coding Agent Index | 58（1 位, Claude Code） | 53（Codex、旧世代から横ばい） | — |
| FrontierCode-AA 2.0（AI コーディング） | 49.1%（1 位） | 42.4%（旧 Astra 値） | 40.6% |
| AutomationBench-AA（業務自動化） | 76.4% | 68.5%（旧 Astra 値） | 77.8%（1 位） |
| GDP.pdf（視覚的文書の知識労働） | 24.2% | 31.0%（旧 Astra 値） | 28.4% |
| タスクあたり費用（max/high） | $26.54 | $4.00 | $5.48 |

*出典：[1][2][3][4]。「旧 Astra 値」は GPT-6.2 向けの項目別公表値が未提供のため、旧 GPT-6 Astra の測定値で代用したもの。*

## 2. GPT-6.2 Astra

### 2.1 すぐれている分野

- 科学・エージェント型研究ワークフロー：Terminal-Bench-Science で 78.1%と他 2 者を大きく引き離し首位。CritPt（物理推論）、AA-Omniscience（事実知識）、HLE-Diamond でも Fable に次ぐ 2 位で、AA は「科学と知識の強み」と総括している[1][4]
- 3D・ロボティクス：AA-3D（1318 Elo）と AA-Robotics（79）の両方で首位。テキストから操作可能な 3D 体験の生成や、現実世界の計画・行動に強い[1]
- コスト効率：最高設定でもタスクあたり約$4 で、Fable（$26.54）の約 1/6.6。AA の Cost per Task パレートフロンティアの大半を OpenAI モデルが占有する[1][5]

### 2.2 今一つの分野

- コーディング：Coding Agent Index は旧 Astra から横ばいの 53 で、Fable の 58 に及ばず。独立評価の FrontierCode 系でも Fable に後れを取る[1][6]
- 実務成果物の品質：AA-Briefcase は Elo 1644 で、Fable（1785）に大きく水をあけられている[4]
- 信頼性：最大設定のハルシネーション率 59%は旧 Astra（50%）より悪化しており、確答の絞り込みが甘い側面がある[4]

## 3. Claude Fable 5.2

### 3.1 すぐれている分野

- 総合知能と実務品質：Intelligence Index 首位（65）に加え、AA-Briefcase、HLE-Diamond、CritPt、AA-Omniscience と個別評価でもほぼ首位[1][2]
- コーディング・長時間の自律開発：Coding Agent Index 58 で首位。独立評価の FrontierCode でも 49.1%（Astra 42.8%、Carbon 41.1%）とリード[4][6]
- 専門領域別指数：財務・会計 65、法務 63、エンジニアリング 73、経済学 67 など、AA の能力指数のほぼ全領域でトップ[2]

### 3.2 今一つの分野

- コスト：最大設定でタスクあたり$26.54 と Astra の 6 倍以上。低コスト帯の設定（Low で$3.10）を持つものの、最高知能の利用は高額[2]
- 業務自動化の一部：AutomationBench-AA では Carbon（77.8%）に 0.4 ポイント差で譲り、Terminal-Bench 4.0（59.6%）も Carbon（65.2%）に及ばない[2]
- 科学研究フロー：TB-Science 73.3%は Astra の 78.1%に次ぐ 2 位にとどまる[4]

## 4. Gemini 4 Carbon

### 4.1 すぐれている分野

- 定型業務フローの自動化：AutomationBench-AA 77.8%で 3 者首位。Google 公表値でも 53.9%（Fable 51.4%、Astra 41.4%）と一貫して強い[2][6]
- 端末操作：Terminal-Bench 4.0 で 65.2%と首位[2]
- 入力モダリティ：3 者中唯一、テキスト・画像に加え音声・動画入力に対応[2]
- 価格と性能のバランス：入力$2／出力$10（100 万トークン）と Fable（$8/$40）の 1/4 で、Intelligence Index 60 は Astra（62）に肉薄[2]

### 4.2 今一つの分野

- 知識・推論の深度：AA-Omniscience 54（Fable 67）、HLE-Diamond 69.1%（Fable 80.2%）など、深い知識を問う評価で明確に 3 位[2]
- 科学・物理：CritPt 69.0%、TB-Science 64.8%と両者に大きく離されている[2][4]
- AI コーディング：FrontierCode-AA 2.0 で 40.6%と最下位[2]
- 速度：出力 64 tok/s・タスクあたり約 32 分と遅く、Fable（112 tok/s）に大きく劣る[2]

## 5. 実務への含意

3 モデルは「総合知能の Fable」「科学・3D・コスト効率の Astra」「定型自動化・マルチモーダルの Carbon」と棲み分けが鮮明である。特許調査・知財実務の文脈では、クレーム分析や無効資料調査のような深い推論と正確な知識が要る作業は Fable、大量の明細書スクリーニングや要約のようにコスパ重視で回す作業は Astra／Carbon、という使い分けが数値と整合する。ただし Astra はハルシネーション率 59%と高めであるため、引用・番号照合のような正確性が命の工程では検証ステップを必須とすべきである[4]。

【留意点】GPT-6.2 Astra の項目別の一部（AutomationBench、GDP.pdf 等）は、現時点では旧 GPT-6 Astra 値での比較しか公開されていない。AA の比較ページが 6.2 対応に更新され次第、対比表の該当箇所を再確認する必要がある[1][3]。

## 参考文献


[1] Artificial Analysis, "OpenAI has the #2 model in the Intelligence Index with the release of GPT-6.2 Astra and GPT-6.2 Sol", November 17, 2026.https://artificialanalysis.ai/articles/openai-has-the-2-model-in-the-intelligence-index-with-the-release-of-gpt-6-2-astra-and-gpt-6-2-sol （最終アクセス：2026 年 11 月 18 日）

[2] Artificial Analysis, "Claude Fable 5.2 vs. Gemini 4 Carbon：リリースの知能・性能・料金の比較".https://artificialanalysis.ai/ja/models/comparisons/claude-fable-5-2-vs-gemini-4-carbon （最終アクセス：2026 年 11 月 18 日）

[3] Artificial Analysis, "Gemini 4 Carbon vs. GPT-6 Astra：リリースの知能・性能・料金の比較".https://artificialanalysis.ai/ja/models/comparisons/gemini-4-carbon-vs-gpt-6-astra （最終アクセス：2026 年 11 月 18 日）

[4] HandyAI, "Model Drop: GPT-6.2 Astra & GPT-6.2 Sol", Substack, November 17, 2026.https://handyai.substack.com/p/model-drop-gpt-62-astra-sol （最終アクセス：2026 年 11 月 18 日）

[5] R&D World Online, "GPT-6.2 Astra trails Fable 5.2 on AI index at lower cost", November 17, 2026.https://www.rdworldonline.com/openais-gpt-6-2-astra-trails-claude-fable-5-2-by-three-points-on-an-independent-ai-index-at-a-fraction-of-the-cost-per-task/ （最終アクセス：2026 年 11 月 18 日）

[6] けん（.ai Futures）, "最強 AI3 選の実務別比較!Fable5.2・Astra・Carbon", note, 2026 年 11 月 14 日.https://note.com/ken0001/n/n372591e7b784 （最終アクセス：2026 年 11 月 18 日）