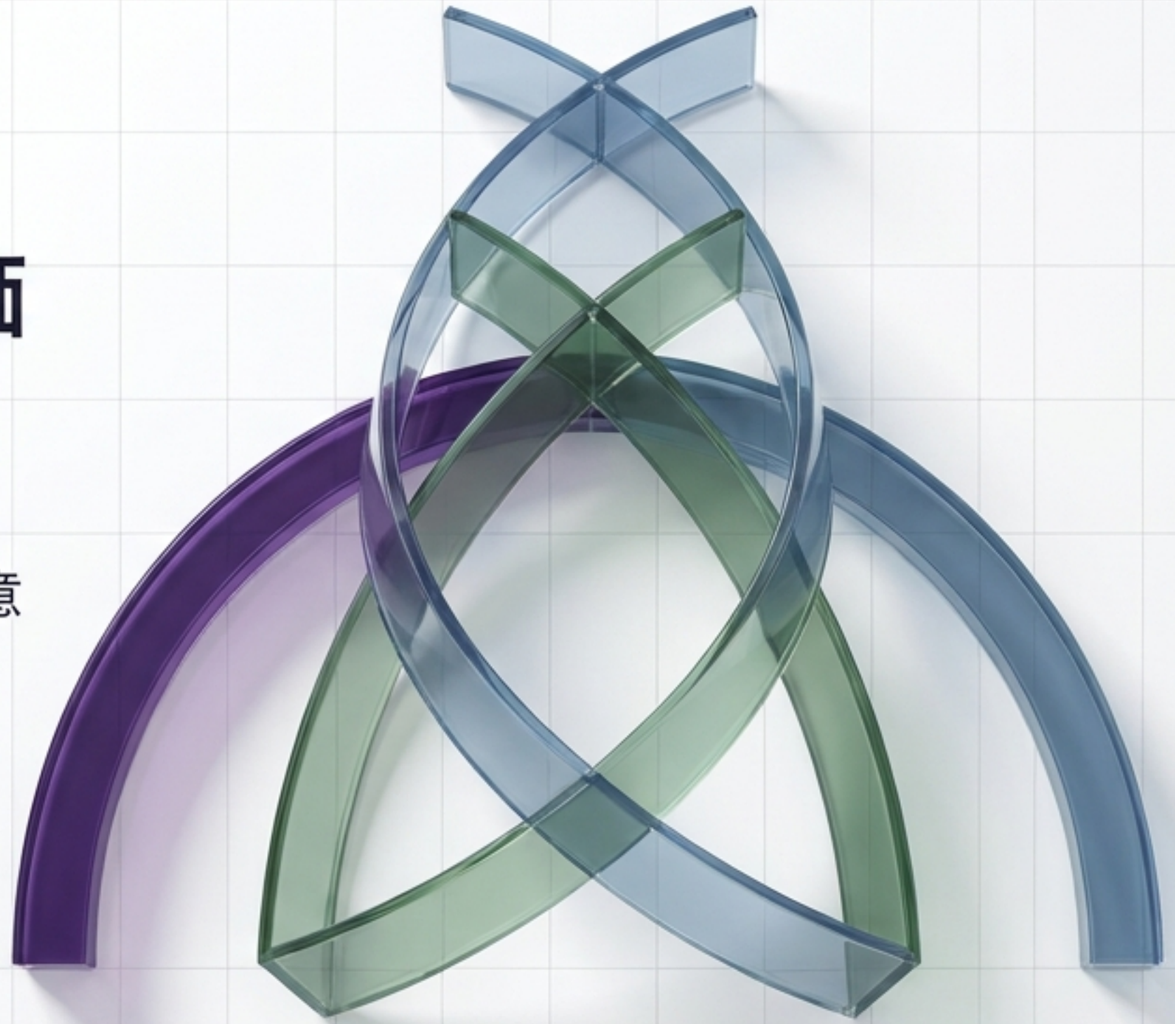
2026年 次世代AIモデル評価値比較と実務導入戦略

Claude Fable 5.2 / GPT-6.2 Astra / Gemini 4 Carbonの徹底解剖と知財・リサーチ業務への含意

情報基準日: 2026年11月17日
データソース: Artificial Analysis Intelligence Index v5.0.1
ステータス: CONFIDENTIAL / STRATEGIC OVERVIEW



Gemini Notebook

「単一の最強モデル」から、
「適材適所のマルチモデル・ポートフォリオ」へ。



Claude Fable 5.2

妥協を許さない最高品質の推論・自律開発エンジン

圧倒的な総合知能を持つが、コストは他モデルの6倍以上。品質が絶対視されるタスク専用。



GPT-6.2 Astra

科学領域と圧倒的コスパを制する、破壊的チャレンジャー

Fableに肉薄する知能を1/6のコストで実現。ただし最大設定時のハルシネーションに注意。



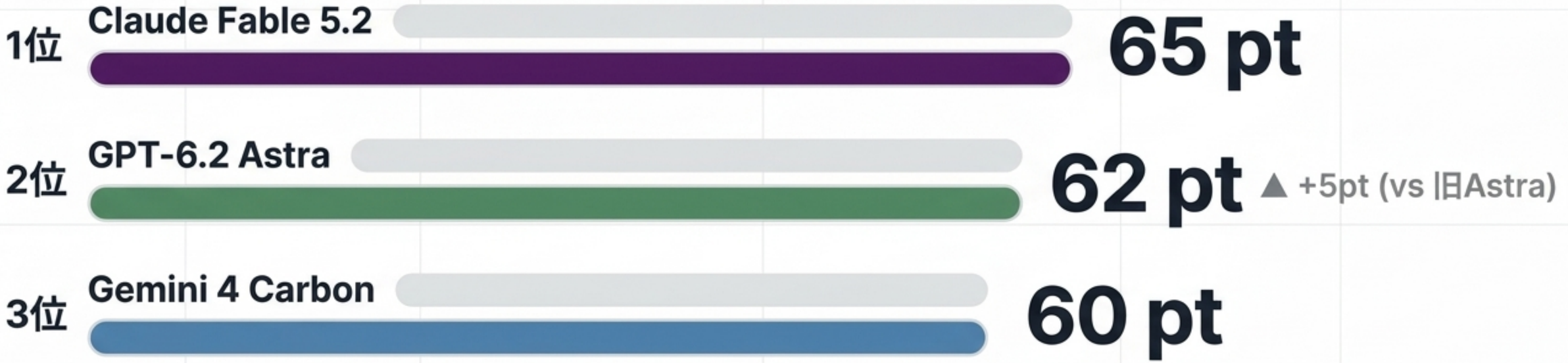
Gemini 4 Carbon

定型業務フローと多次元入力を支配する、自動化のスペシャリスト

端末操作と音声・動画を含む真のマルチモーダル対応。大量の定型処理に最適。

Gemini Notebook

Intelligence Indexの地殻変動：僅差の三つ巴へ



GPT-6.2 Astraのリリースにより、上位陣のスコアがかつてないほど肉薄。トップモデル間の能力差が縮まったことで、『性能至上主義』から『コストと用途の最適化』へ戦略の重心が移行している。

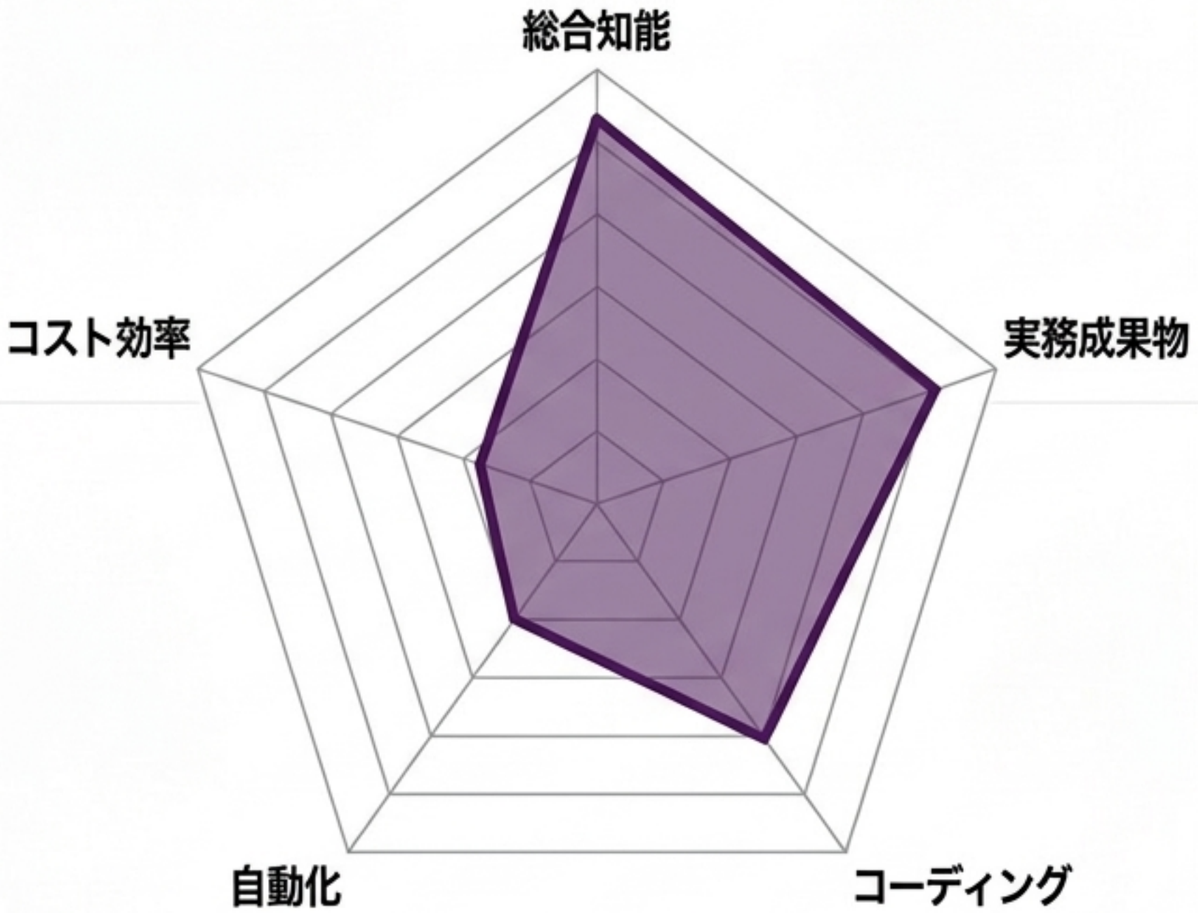
Gemini Notebook

性能評価ヒートマップ：14の指標が示す能力の偏在

Metric	Fable	Astra	Carbon
【総合・推論】			
実務成果物(Briefcase Elo)	**1785**	1644	1611
知識・推論(HLE-Diamond)	**80.2%**	2位	69.1%
物理推論 (CritPt v2)	**76.5%**	2位	69.0%
事実知識 (Omniscience)	**67**	2位	54
【特定タスク・空間】			
科学研究フロー (TB-Science)	73.3%	**78.1%**	64.8%
3D体験生成 (AA-3D Elo)	1271	**1318**	未公表
現実世界理解 (Robotics)	75	**79**	未公表
【実務・自動化・コード】			
業務自動化 (AutomationBench)	76.4%	68.5%	**77.8%**
端末操作 (Terminal-Bench 4.0)	59.6%	64.1%	**65.2%**
AIコーディング (FrontierCode)	**49.1%**	42.4%	40.6%
【コスト】			
タスクあたり費用	$26.54	**$4.00**	$5.48

Gemini Notebook

Claude Fable 5.2 | The Reasoning Engine



High Points (強み)

最高峰の知能と品質: Briefcase 1785 / HLE-Diamond 80.2%

専門領域の支配: 法務(63)、財務(65)、エンジニアリング(73)の全域でトップ

開発力: Coding Agent Index 58で首位

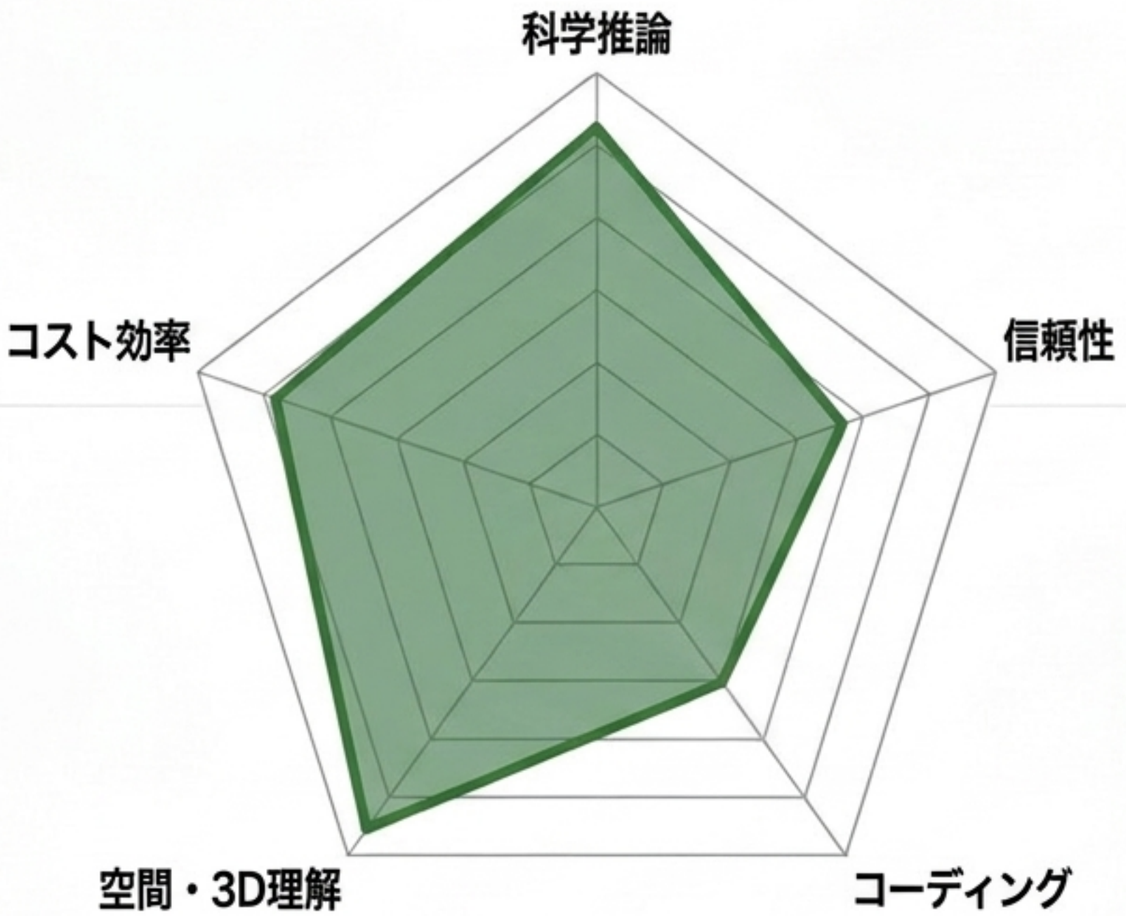
Frictions (摩擦・弱点)

極限のコスト: 最大設定でタスクあたり$26.54（Astraの6倍超）

高コストを許容してでも、法務・財務・知財における致命的なミスを排除し、最高品質の推論を求める『最終防衛ライン』として運用すべきエンジン。

Gemini Notebook

GPT-6.2 Astra | The Scientific & Cost Leader



High Points (強み)

科学研究の突破力: TB-Science 78.1%で他を圧倒

物理・空間把握: AA-3D(1318) / Robotics(79)で首位

破壊的コスト効率: 最高設定でもタスクあたり$4.00

Frictions (摩擦・弱点)

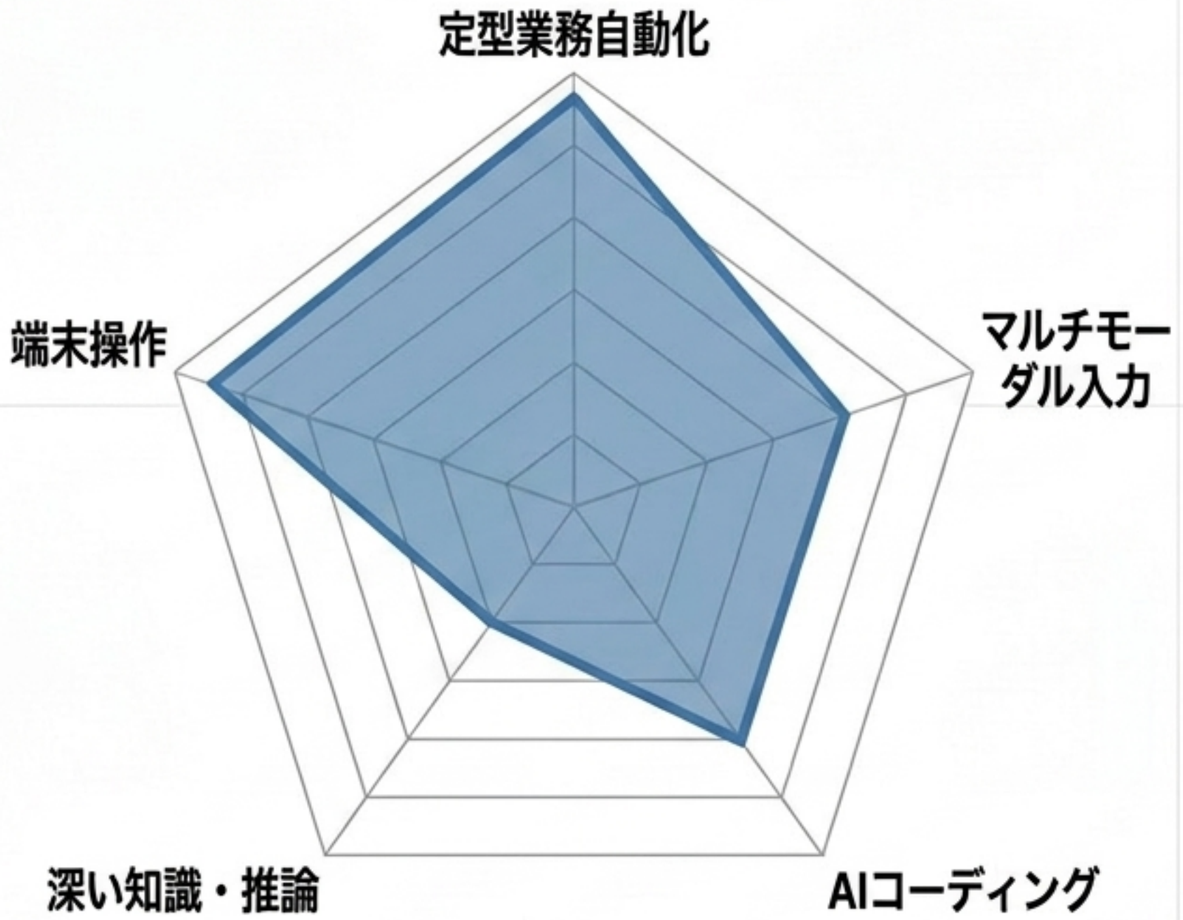
ハルシネーション: 最大設定での発生率が59%（旧モデルの50%から悪化）

停滞するコード生成: 指数53（旧世代から横ばい）

圧倒的なコスパで大量の科学技術文献を処理できるが、事実関係の確答を要するタスクでは『幻覚』のリスクが高まる。検証プロセスとの併用が前提。

Gemini Notebook

Gemini 4 Carbon | The Automation Specialist



High Points (強み)

自動化の覇者: AutomationBench 77.8% / Terminal-Bench 65.2%（共に首位）

多次元モダリティ: テキスト・画像に加え、唯一「音声・動画」の直接入力に対応

バランスの取れた価格: 入力$2 / 出力$10（100万トークンあたり）

Frictions (摩擦・弱点)

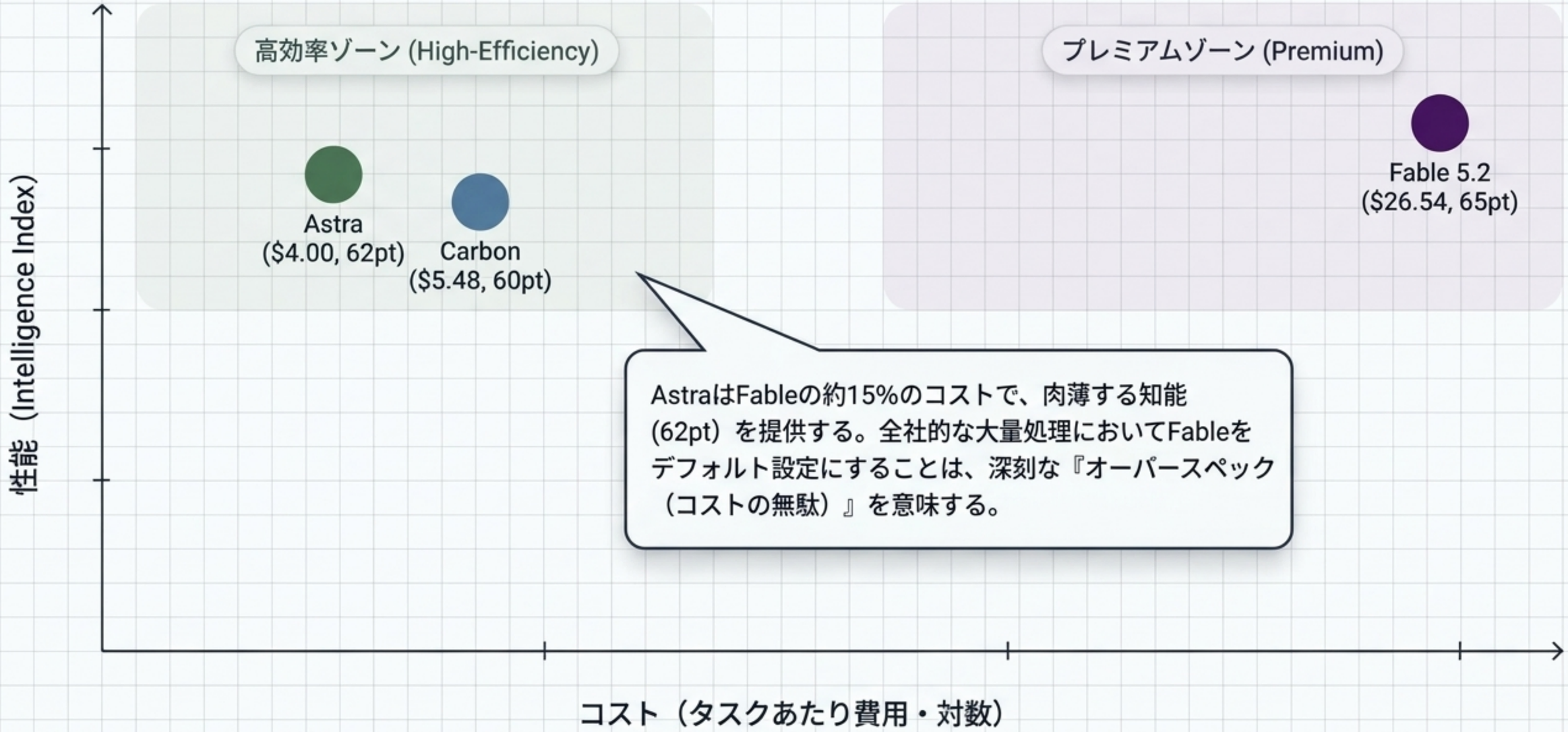
浅い推論深度: HLE-Diamond 69.1% / Omniscience 54（事実知識で明確に最下位）

コード生成能力: FrontierCode 40.6%で3者中最下位

深い専門知識を問う思考作業には不向きだが、GUI操作や動画解析を含む『ルーチンワークの完全自動化』においては比類なきパフォーマンスを発揮する。

Gemini Notebook

コスト vs パフォーマンス：知能の経済学



Gemini Notebook

リスク分析：Astraの「ハルシネーション」パラドックス

能力

科学推論スコア
78.1%
(首位)





リスク

最大設定時の
ハルシネーション率
59%
(旧モデルから悪化)



旧Astra (50%) ➡ GPT-6.2 Astra (59%)

高度な科学的文脈を理解する力はトップクラスだが、出力された『事実』の絞り込みが甘い。引用文献のページ数、特許番号の照合など、１文字のミスが致命傷となる業務においては、Astraの出力をそのまま信頼することは極めて危険である。

Gemini Notebook

戦略的デプロイメント：知財・リサーチ業務の最適化フロー

【ユースケース A】

クレーム分析・無効資料調査・侵害鑑定

要件: 深い論理推論、絶対的な正確性

→ **Claude Fable 5.2**

（高コストだがヒューマンエラーを防ぐ防波堤として機能）

【ユースケース B】

大量の明細書スクリーニング・技術動向の要約

要件: コスト効率、大量処理、科学技術理解

→ **GPT-6.2 Astra / Carbon**

（一次スクリーニングを爆発的なスピードと低コストで処理）

【⚠️ 必須検証ステップ (Human-in-the-Loop)】

Astraで処理したデータの「引用文献・特許番号の抽出」工程には、人間の専門家によるチェック、またはFableによるクロスチェック（交差検証）APIを必ず噛ませるアーキテクチャを構築する。

Gemini Notebook

Mission Checklist：留意事項と今後のアクション

【データの透明性に関する留意点】 (Caution)

現在、AA比較ページにおけるGPT-6.2対応は進行中。本資料内の「AutomationBench」「GDP.pdf」等の指標は、暫定的に「旧GPT-6 Astra値」を用いて比較している。（出典: Artificial Analysis公表値・透明性基準に基づく）

【ネクストアクション】 (Action Items)

- [ ] **APIルーティングの再設計**
  タスクの「推論深度」と「処理量」に応じ、FableとAstra/Carbonを自動で振り分ける社内プロンプトルーティングを構築する。
- [ ] **ハルシネーション検証フェーズの導入**
  Astraの知財業務適用に向け、事実照合専用の「クロスチェック・プロンプト」を標準実装する。
- [ ] **次回アップデートのスケジューリング**
  GPT-6.2の全公式データ（AutomationBench等）が更新され次第、本評価マトリクスの再検証を実施する。

Gemini Notebook