Claude Fable 5.2 の内容と評価

公式発表と第三者評価から見る性能 費用 評判 知財実務への示唆

調査基準日 2026 年 11 月 10 日 日本時間

GPT-6 Astra

Claude Fable 5.2 は、高難度の研究・開発や長時間の自律作業で大きな進歩を示している。独立評価でも総合首位を獲得しているが、最高性能の設定では費用と待ち時間が増え、法律実務などでは評価が分かれる。[1][5][8]

公式発表日は米国時間 2026 年 11 月 9 日、日本時間では 11 月 10 日である。本稿は公式資料、第三者ベンチマーク、先行利用者のレビューを整理したものであり、筆者自身によるモデルの実測試験ではない。発売直後の初期評価であるため、長期利用での安定性に関する判断は今後の検証を待つ。[1]

1 製品の位置づけと主な仕様

項目	内容
位置づけ	複雑な推論、研究開発、長時間のエージェント作業向けの最上位モデル
同時発表	Claude Mythos 5.2。基盤モデルは共通だが、サイバー・生物分野の利用制限とアクセス条件が異なる
コンテキスト	100 万トークン
最大出力	12 万 8,000 トークン
入出力	テキスト・画像を入力し、テキストを出力
通常 API 料金	100 万トークン当たり入力 8 米ドル、出力 40 米ドル。5.1 から 20%値下げ
高速モード	入力 16 米ドル、出力 80 米ドル。トークン生成を最大 2.5 倍に高速化
通常時の生成速度	前世代より出力生成が最大 30%高速化したと Anthropic は説明

出典 [1]〜[3]。料金は米ドル建て。高速化はトークン生成速度に関する説明であり、思考量を含めた作業全体が常に短時間になるという意味ではない。

Anthropic が強調しているのは、大きな課題を分解し、多数のエージェントを統括しながら、検証と修正を続ける能力である。数日単位の開発や複数のリポジトリにまたがる作業を想定している。一方、日常的なコーディングや知識業務については、同社自身がまず Opus 5.5 から始めることを勧めている。[4]

2 第三者ベンチマークによる評価

Artificial Analysis の総合指標では、Fable 5.2 の Max 設定が首位に立った。以下は調査時点の表示値であり、推論設定も併記した。[5]

モデルと推論設定	総合指標	評価課題当たり費用
Fable 5.2 Max	65	26.54 米ドル

モデルと推論設定	総合指標	評価課題当たり費用
Fable 5.2 High	62	9.08 米ドル
Fable 5.2 Medium	56	3.80 米ドル
Gemini 4 Carbon High	60	7.76 米ドル
Opus 5.5 Max	59	10.04 米ドル
GPT-6 Astra Max	57	6.24 米ドル

出典 [5][6]。総合指標は正答率ではない。費用は評価課題を重み付けした平均であり、通常のチャット 1 回の料金ではない。

最高性能と費用対効果は分けて評価する必要がある。Fable 5.2 は High でも強い結果を示し、Max にすると指標が 62 から 65 へ上がる一方、費用は約 2.9 倍になる。同機関が出力トークン量と速度から算出する処理時間指標も、Max では約 46 分／課題である。これも通常の質問への応答時間を示すものではない。[5][6]

個別分野では、次のように得意・不得意がある。すべて同一機関の比較ページの値であり、他機関や Anthropic 自身の測定値とは区別した。[7]

評価項目	Fable 5.2 Max	Gemini 4 Carbon High	GPT-6 Astra Max
FrontierCode 2.0 コーディング	48.8%	40.6%	42.4%
Terminal-Bench-Science 0.1 科学研究	73.3%	64.8%	63.3%
Terminal-Bench 4.0 ターミナルを使う課題	59.6%	65.2%	59.1%
GDP.pdf 長文 PDF 関連の評価	24.2%	28.4%	31.0%

総合首位を、あらゆる作業での首位と解釈することはできない。特に、図表や文書を扱う実務では、総合指標だけでなく該当する個別項目を確認する必要がある。[7]

別の独立評価機関 Vals AI でも、High 設定で総合指標 71.31%、36 モデル中 1 位であった。ただし、Sonnet 5.5 の 70.85%、Gemini 4 Carbon の 70.45%との差は小さく、この評価では圧倒的な差とまではいえない。科学分野の総合指標では GPT-6 Astra に次ぐ 2 位である。[8]

3 注目された実例と実験条件

電子機器の設計

Anthropic は、スピーカーフォン、音声アシスタント端末、携帯型バーチャルペットの設計例を公表した。Fable 5.2 が回路、基板、筐体、ファームウェアを扱い、人間が製造した試作品は回路修正なしで動作したという。設計から動作する試作品へ進んだ点が重要である。ただし、長期耐久性や商用認証への適合は未確認である。[9]

フェルマーの最終定理の形式化

約 1,000 のエージェントが、約 29 時間で Lean による証明の形式化を再現した。60 の中間命題、既存の 47 補題などが与えられた実験であり、既に知られた定理の証明をコンピュータで検証可能にする成果である。大規模な分業と形式検証の能力を示すが、人間から何の材料も与えられずに新しい定理を発見したという意味ではない。[10]

理論物理の研究

Anthropic は、ブラックホール関連の未解決問題 2 件を解いたと発表し、関連プレプリントを案内している。本稿では同社の研究成果報告として扱い、当該計算の独立再現や学術的評価まで確認できたとはしていない。[1]

4 先行利用者の評価と評判

初期レビューでは、能力への驚きと、運用上の負担への指摘が併存している。以下はいずれも実際に試した人の報告であるが、課題や設定が異なるため、利用者全体を代表する調査としては扱わない。

Ethan Mollick による評価

Mollick は、事業構想、研究、教材、3D 世界の制作など、幅広いデジタル課題で専門家に近い能力を示すと評価し、「Jagged AGI」という表現を用いた。これは、能力の凹凸を残しながら汎用的な知的作業に近づいたという本人の見解であり、AGI 達成が客観的に確定したことを意味しない。[11]

同時に、冗長化、発想の反復、過剰なルール作りを自分で問題と認識できず、人間の介入が必要だったと報告している。成果物を作る力に比べ、その成果物が長すぎる、退屈、運用困難といった問題を自分で見抜く力が弱いという指摘は、実務上重要である。[11]

Simon Willison と Spring Prompt の検証

Willison は SVG 制作を推論設定別に試し、Xhigh で良好な結果を得た一方、Max では 12 万 8,000 トークンを思考に使い切り、成果物を返さなかった事例を報告した。推論を最大にすれば常に良い結果が得られるとは限らない。[12]

Spring Prompt は、自社試験でアンケート分析と広告予算運用が首位だったと評価した。一方、高コストと遅さを指摘し、29 件の創作コーディング課題のうち 8 件は、当初の設定で思考上限に達した後、推論を弱めて完了したと報告している。単発の制作課題の結果を一般的な失敗率とみなすことはできない。[13]

Claire Vo と開発者コミュニティの反応

Claire Vo は動画制作、家具の組み立て支援、ゲーム制作などを評価したが、Web デザインは単独で任せるには不満が残ると述べている。[14]

Hacker News では、Max 設定のトークン消費と費用への批判に対し、Medium や High で比較すべきだという反論が見られる。評判を読む際には、どの推論設定で、どの作業を試した感想なのかを確認する必要がある。[15]

5 利用条件と実務上の注意点

Pro と Team 標準席では、米国太平洋時間 2026 年 11 月 16 日 23 時 59 分、日本時間 11 月 17 日 16 時 59 分まで、週の利用上限の 50%まで Fable を追加料金なしで利用できる。その後は利用クレジットによる課金に戻る。利用枠のリセット特典には、契約開始時期などの条件がある。[16]

Claude アプリと Claude Code の推論初期設定は Medium、API は High である。公式 API 資料も、Max は時間やトークンを重視しない限られた課題向けとしている。モデル名だけで比較せず、推論設定も記録することが望ましい。[1][3]

API には原則 30 日間のデータ保持要件があり、ゼロデータ保持は Anthropic が明示的に認める場合に限られる。未公開発明や顧客資料を利用する際は、既存モデルと同じ契約条件だと考えず、適用される保存条件を確認する必要がある。[2][3]

6 知財実務への示唆

知財実務への適用では、総合ランキング以上に、法律分野の評価の違いが重要になる。調査時点で、次の差が確認できた。[8][17][18]

法律分野の評価	Fable 5.2 の結果
Artificial Analysis Legal Index	Max 設定で 63 点 掲載モデル中首位
Vals Legal Index	High 設定で 69.17% 37 モデル中 20 位

両者は対象課題、重み付け、推論設定が異なる。Artificial Analysis は法律知識、推論、文書処理などを組み合わせ、Vals は米国法律事務所の業務別収入を基に、調査や文書作成等を重み付けしている。この違いは、日本語の特許実務での優劣を総合スコアから直接判断できないことを示している。[17][18]

公開された一般能力から考えると、知財分野でまず試す価値があるのは、複数文献を突き合わせる技術分析、複雑なクレーム対比、実験データの解析、長時間の調査工程の統括である。ただし、これは本稿の考察であり、日本の特許業務で改善が実証されたという意味ではない。

導入判断には、公開済みの実案件を使い、Opus 5.5 などと同じ資料と指示で比較することが有効である。評価軸を、引用箇所の正確さ、重要事項の見落とし、技術的な読み違い、修正に必要な人間の時間、総費用に置くことで、Fable 5.2 を使う価値がある工程を見極めやすくなる。

参考文献

本文中の番号は以下の文献に対応する。閲覧日はすべて 2026 年 11 月 10 日。動的なランキング、料金、利用条件は同日時点の表示内容に基づく。個別レビューは著者の観察・評価として引用した。比較ページ[7]は Fable 5.2 Max、Gemini 4 Carbon High、GPT-6 Astra Max の表示を参照した。

[1] Anthropic. Introducing Claude Fable 5.2 and Claude Mythos 5.2. Anthropic. 2026 年 11 月 9 日.
https://www.anthropic.com/claude-fable-and-mythos-5-2

[2] Anthropic. Claude Fable 5.2. Claude Platform Docs. 公開日記載なし 随時更新.
https://platform.claude.com/docs/en/models/fable-5-2/overview

[3] Anthropic. What's new in Claude Fable 5.2. Claude Platform Docs. 公開日記載なし 随時更新.
https://platform.claude.com/docs/en/models/fable-5-2/whats-new-fable-5-2

[4] Anthropic. Claude Fable 5.2. Anthropic 製品紹介. 公開日記載なし 随時更新.
https://www.anthropic.com/claude/fable

[5] Artificial Analysis. LLM Leaderboard — Comparison of AI models. Artificial Analysis. 公開日記載なし 随時更新.
https://artificialanalysis.ai/leaderboards/models

[6] Artificial Analysis. Claude Fable 5.2 (Max, Default Fallback) — Intelligence, Performance & Price Analysis. Artificial Analysis. 公開日記載なし 随時更新.
https://artificialanalysis.ai/models/claude-fable-5-2

[7] Artificial Analysis. Model Comparison Tool. Artificial Analysis. 公開日記載なし 随時更新.
https://artificialanalysis.ai/models/comparisons

[8] Vals AI. Claude Fable 5.2 Benchmarks, Cost and Capabilities. Vals AI. 2026 年 11 月 9 日の評価告知を掲載 随時更新.
https://www.vals.ai/models/anthropic_claude-fable-5-2

[9] Anthropic Science Team. Testing Claude on real-world electronics design. Anthropic Research. 2026 年 11 月 9 日.
https://www.anthropic.com/research/claude-electronics-design

[10] Barry Zhang, Tianyi Peng, Elle Najt, Keller Scholl and Lance Martin. Safely scaling multi-agent teams with Claude Fable 5.2. claude.dev Blog. 2026 年 11 月 9 日.
https://claude.dev/blog/safely-scaling-multi-agent-teams-with-fable-5-2/

[11] Ethan Mollick. Fable 5.2 and Jagged AGI. One Useful Thing. 2026 年 11 月 9 日.
https://www.oneusefulthing.org/p/fable-52-and-jagged-agi

[12] Simon Willison. Release: llm-anthropic 0.31. Simon Willison's Weblog. 2026 年 11 月 9 日.
https://simonwillison.net/2026/Nov/9/llm-anthropic/

[13] Ellis Crosby. Claude Fable 5.2 review: first on our survey and ad budget tests, and one of the dearest models to run. Spring Prompt. 2026 年 11 月 10 日.
https://springprompt.com/blog/claude-fable-5-2-review

[14] Claire Vo. Claude Fable 5.2 review: fast, creative, and finally not annoying. Lenny's Newsletter / How I AI. 2026 年 11 月 9 日.
https://www.lennysnewsletter.com/p/claude-fable-52-review-fast-creative

[15] Hacker News 投稿者各氏. Claude Fable 5.2 scores 65 on Artificial Analysis, highest so far. Hacker News 討論スレッド item 50431144. 2026 年 11 月 9 日から 10 日の投稿を参照.
https://news.ycombinator.com/item?id=50431144

[16] Anthropic. Claude Fable 5.2 launch offers. Claude Help Center. 公開日記載なし 2026 年 11 月 10 日閲覧時更新.
https://support.claude.com/en/articles/17506494-claude-fable-5-2-launch-offers

[17] Artificial Analysis. Legal Index. Artificial Analysis. 公開日記載なし 随時更新.
https://artificialanalysis.ai/indexes/legal

[18] Vals AI. Vals Legal Index Leaderboard and Methodology. Vals AI. 公開日記載なし 随時更新.
https://www.vals.ai/benchmarks/legal_index