Claude Opus 5.5

特徴と評判に関する調査報告

調査基準日 2026 年 9 月 23 日

GPT-6 Astra

Claude Opus 5.5 は、長時間のコーディングや複数資料を使う知識労働を強化し、料金と処理効率も改善したモデルである。Anthropic は 2026 年 9 月 22 日（米国時間）、Claude 5.5 シリーズの最初のモデルとして公開した。第三者評価でも高性能が確認されており、実務で試す価値の大きい更新と評価できる。一方、最大推論設定での費用、作業を最後まで完成させる確実性、専門分野でのモデル切替には留意が必要である。[1][2]

本稿は、公式資料、第三者ベンチマーク、先行利用者の実測・レビューを照合し、特徴、性能、費用、評判、導入上の条件を整理する。知財実務への適用可能性は、公開評価に基づく考察として示す。一般公開の翌日時点の調査であり、長期運用後の評価はまだ出そろっていない。本稿独自の同一条件によるモデル比較実験は行っていない。

基本仕様は、長い文脈とツール利用を伴う業務を想定している。

項目	確認できた内容
API モデル名	claude-opus-5-5
主な用途	長時間のエージェント型コーディング、知識労働
コンテキスト長	100 万トークン
最大出力	通常 12.8 万トークン。Batch API では 30 万トークンのベータ機能あり
入出力	テキスト・画像を入力し、テキストを出力
推論方式	Adaptive thinking。無効化できない
推論強度	low／medium／high／xhigh／max。標準は medium
知識の基準時点	2026 年 6 月
提供先	Claude API、Amazon Bedrock、Google Cloud、Microsoft Foundry など

出典：Anthropic のモデル仕様。[3]

モデルの重みは非公開で、パラメータ数も公表されていない。ローカル PC にモデルをダウンロードして動かす方式には対応していない。[4]

性能面では、資料やツールを使いながら複数段階の仕事を進める能力が評価されている。

GitHub は、自社の先行テストで、Opus 5 と同程度に課題を解決しながら、手順とトークンを大幅に減らし、途中のエラーからも速く立て直したと報告している。単発のコード生成より、既存コードの調査、修正、検証を繰り返す作業で効果が期待できる。GitHub Copilot への提供も発表されている。[5]

知識労働では、Box がスプレッドシート、PDF、プレゼンテーション、画像を横断して成果物を作る課題で評価した。Box によると、Opus 5 に対し、作業完了が 30%速く、使用トークンは約 3 分の 1、成果物の長さは 42%短くなり、評価対象で精度の低下はなかった。文書分析から報告資料の作成までを任せたい利用者にとって、有用な実測報告である。ただし、Box 自身の業務課題と評価環境での結果である。[6]

図表・画像についても、公式のプロンプトガイドは、グラフの数値、図中の接続関係、技術図面などの読み取り改善を説明している。ただし、細密な図では高解像度画像や切り出し・拡大ツールが依然有効である。図面を入力できることと、細部を取り違えずに解釈できることは、別々に検証する必要がある。[7]

第三者ベンチマークでは、Artificial Analysis の総合指標で首位となった。

調査時点の Artificial Analysis Intelligence Index では、最大推論設定・フォールバック有効の Opus 5.5 が 58 を記録している。この値は総合指標のスコアであり、正答率 58%という意味ではない。現行ページの指標は v4.3.2 で、過去の異なるバージョンの数値とはそのまま比較できない。[4][8]

Artificial Analysis の評価では、10 項目のうち 6 項目で首位となり、知識労働の成果物を評価する AA-Briefcase では 1,822 Elo、GDPval-AA では 1,846 Elo を記録した。一方、CritPt、AA-LCR、GDP.pdf では首位ではなく、用途による得意・不得意は残っている。[2]

Anthropic が発表した主要な比較値は次のとおりである。メーカーの発表表に掲載された値であり、全モデルを完全に同一条件で再測定した結果ではない。[1]

評価項目	Opus 5.5	Fable 5.1	Opus 5	GPT-6 Astra
Terminal-Bench 4.0	66.4%	55.8%	52.3%	57.9%
FrontierCode v1.1	54.4%	50.3%	48.0%	53.3%
AutomationBench	40.0%	31.4%	26.9%	41.4%
Terminal-Bench-Science	58.7%	52.6%	29.0%	64.6%

出典：Anthropic 公式発表。科学研究評価の版は Terminal-Bench-Science 0.1。[1]

Terminal-Bench 4.0 の Opus 5.5 は、Anthropic 発表では 66.4%、Artificial Analysis の測定では 59.6%となる。後者では GPT-6 Astra と同水準とされている。実行環境、推論強度、ツール構成などが違うため、数値を混ぜて他社モデルを一律に大差で上回ると解釈することは適切ではない。[1][2]

公式表には誤差や安全機構の介入条件も付されている。たとえば、Terminal-Bench 4.0 では Opus 5.5 は xhigh、GPT-6 Astra は OpenAI 公表の high 設定の値が使われている。ベンチマーク上の小差だけで、実際の業務での優劣を確定することはできない。[1]

料金は下がったが、40%安いという説明は条件付きで理解する必要がある。

通常の API 料金は次のとおりである。単位は 100 万トークン当たりの米ドルで、税や個別の追加料金は含めていない。[9]

課金項目	Opus 5	Opus 5.5	値下げ率
通常入力	5 ドル	4 ドル	20%
出力	25 ドル	20 ドル	20%
キャッシュ読み込み	0.50 ドル	0.20 ドル	60%
キャッシュ書き込み 5 分	6.25 ドル	5 ドル	20%

出典：Anthropic 料金資料。値下げ率は表の単価から算出。[9]

Anthropic のいう実行コスト 40%削減は、標準設定の典型的な仕事で、単価低下と処理効率改善を合わせた同社の評価である。すべてのリクエストが 40%安くなるという料金制度ではない。[1]

長い資料や共通の指示を繰り返し使う仕事では、キャッシュ読み込みの値下げが効く。一方、ツール利用には追加費用が生じる場合があり、クラウド事業者や処理地域によっても料金が異なる。Batch API は通常の入出力単価の半額、Fast mode は入力 8 ドル・出力 40 ドルで、利用できる環境も限定されている。[9]

実際の費用を左右するのは、推論強度の設定である。

Artificial Analysis が測定した、同じ Opus 5.5 の設定別結果は次のとおりである。評価時にはフォールバックが有効になっている。[10]

推論強度	Intelligence Index	評価タスク当たりの費用
low	42	0.55 ドル
medium 標準	51	1.34 ドル
high	54	1.82 ドル
xhigh	56	3.46 ドル
max	58	5.98 ドル

出典：Artificial Analysis の設定別比較。費用は評価課題の加重平均。[10]

この費用はベンチマークを構成する課題の加重平均であり、通常のチャット 1 往復の料金ではない。表の数値から計算すると、max は medium の約 4.5 倍である。[10]

さらに、Artificial Analysis では、max 設定の出力トークンが 1 課題当たり約 11.9 万となり、Opus 5 の約 7.3 万から増加した。その結果、max 同士のタスク当たり費用は Opus 5 と同程度だったと報告している。安くなったから常に max を使うという選び方は、合理的とは限らない。[2]

公式ガイドも medium から始め、実際の課題で品質差が確認できる場合に推論強度を上げるよう勧めている。利用者に見える回答が簡潔になっても、内部の推論を含む課金トークンが少なくなるとは限らない。[7]

評判は好意的だが、実測レビューには具体的な弱点もある。

先行利用を 7 日間行った Every のレビューは、肯定面と否定面の両方を報告している。同誌は Anthropic から先行アクセスを受けているが、記事への同社の関与はなかったと明示している。[11]

コーディングやデザインでは、日常の開発で Fable 5.1 から乗り換えた評価者がいる。対話ではフィードバックに応じやすくなり、文章の読みやすさも改善した。一方、文章の要点を後ろに置く癖が残り、編集作業では他モデルを好む評価者もいる。[11]

同レビューの具体的な失敗例として、研修の進行表を頼んだのに周辺教材を作り込み、期限までに進行表を完成できなかったケース、スライドでロゴや色を取り違えて根拠のない記述を加えたケース、自動チェックを通過しても実際には主要画面でエラーとなったケースが挙げられている。[11]

これらは限定されたテストであり、失敗率を示すものではない。ただし、見栄えの良さ、会話の自然さ、テストに通ったという自己報告だけでは、完成品の品質を判断できないという示唆は実務上重要である。

一般利用者の Reddit 投稿でも、速度、UI の不具合発見、指示の理解を評価する声が確認できる。ただし、投稿者自身が初日の高揚感の可能性に触れており、こうした投稿を利用者全体の評価として扱うことはできない。[12]

安全機構は、実際に利用できる能力にも影響する。

公式ヘルプによると、Opus 5.5 では内容に応じて次のような切替・制限が生じる。[13]

分野	主な扱い
通常の安全なコード開発	ソースコードの脆弱性確認や修正などに利用可能
高リスクと判定されたサイバー関連依頼	Opus 4.8 へ切替となる場合がある
一部の生物学・分子設計関連依頼	Opus 5 へ切替となる場合がある
一部の最先端 LLM 開発関連依頼	Opus 5 へ切替となる場合がある
内部推論の抽出を狙う依頼	別モデルへ切り替えず、直接拒否

出典：Anthropic 公式ヘルプ。切替は各分野の全依頼に一律適用されるものではない。[13]

判定対象には、ユーザーが書いた依頼だけでなく、ファイル、検索結果、接続先の情報なども含まれる。Claude のアプリでは自動切替が標準で有効だが、API ではフォールバックを利用者側で設定する必要がある。アプリでは、切替後もその会話のモデル選択が切替先に残ることがある点も、比較試験では見落とせない。[13]

生物学研究向けには、審査を経た組織が利用する Life Sciences Verification Program がある。同プログラムには監視のための 30 日間のデータ保持という条件が明記されている。創薬やバイオ分野の未公開技術情報を扱う場合、この制度の条件を通常利用と同一視しないことが必要である。[14]

METR の外部評価は、研究開発の完全自動化には至っていないと判断している。

METR は 10 営業日のアクセス期間に 5 つの課題などを使って、AI 研究開発を加速する能力を評価した。その結論は、Fable 5.1 からの改善はあるが、AI 研究開発を完全に自動化できる可能性は低いというものである。専門家なら通常は示さない長期課題での弱点が残るとしている。[15]

また、METR 自身が、これは包括的なアライメント特性や Anthropic の特定の安全基準への適合を認証する評価ではないと説明している。外部機関がテストしたことを、あらゆる用途で安全性が保証されたと読み替えることはできない。本稿の安全性に関する分析は、Anthropic の公開説明と METR 自身の公開評価報告を根拠とする。[1][15]

API を組み込んだ既存システムでは、モデル名だけの置換では済まない場合がある。

公式の移行資料では、主に次の変更が示されている。[16]

・推論を無効化する指定や、手動の推論予算指定が受け付けられない。

・特定ツールの利用を強制する従来の指定がエラーになる。

・推論ブロックがモデルと会話履歴に結びつき、履歴や指示の変更でエラーになる場合がある。

・Claude API と Google Cloud では、旧 Computer Use ツールからの移行が必要となる。

・ツール実行間の進捗説明の返却形式が変わり、従来の画面実装では途中経過が表示されなくなる場合がある。

さらに公式ガイドには、長い自律作業で、途中報告をしてターンを終了するケースへの対処も記されている。モデルの能力だけでなく、未完了の仕事を管理して継続させる実行システムの設計が、成果に影響する。[7]

文章出力では、電子透かしの扱いも確認しておく必要がある。

Opus 5.5 の文章出力は透かしの対象となっている。Anthropic の説明では翻訳文も対象だが、透かしは個人や組織を識別するものではなく、Claude が生成・編集に関与した可能性を判定する仕組みである。短い文章や軽微な校正では検出に限界があり、著作権の帰属や著作者性を判定するものでもない。[17][18]

知財実務では、資料の統合と成果物作成の改善が期待できる。

以下は、公開評価からの考察であり、日本語特許実務における性能を実測した結果ではない。

業務	期待できる改善	導入時に確かめたい点
IP ランドスケープ・技術調査	複数資料の統合、分析、経営向け報告資料の作成	根拠資料との対応、数値・引用の正確性
特許明細書・意見書の検討	長い資料間の整合確認、論点整理、文章修正	構成要件、限定事項、引用箇所の取り違え
FTO・クレーム対比	候補特許の整理、対比表の下書き、検討作業の連続実行	重要文献・構成要件の見落とし
特許図面・化学分野	図面の説明、技術情報の関連づけ	細部の読み取りと専門分野のモデル切替
知財部内の業務自動化	文献整理から表・報告書作成までの一連の処理	完了条件、実行時間、修正工数を含む総費用

上表は本稿の考察。基礎となる能力・制約は参考文献[2][6][7][11][13]を参照。

今回確認した公開資料では、日本語特許文書のクレーム解釈、FTO の見落とし率、日本の拒絶理由対応について、Opus 5.5 の優位性を確立する比較検証は確認できていない。一般的な知識労働での高得点を、そのまま特許実務の精度保証に置き換えることはできない。

導入判断では、既に正解や重要論点を把握している過去案件を使い、まず medium と high を比較する方法が適している。評価項目は、回答の印象よりも、重要論点の捕捉率、誤引用・見落とし、完成までの時間、人による修正時間、総費用に置くことを推奨する。Anthropic 自身も、多くの仕事では Opus 5.5 から始め、難しい推論や長期作業で不足する場合に Fable 5.1 を検討する方針を示している。[19]

参考文献

全資料の閲覧日：2026 年 9 月 23 日。随時更新される資料の数値・仕様は、この閲覧時点の内容に基づく。本文の番号は以下の文献番号に対応する。

[1] **Anthropic.** Introducing Claude Opus 5.5. 2026 年 9 月 22 日.

https://www.anthropic.com/claude-opus-5-5

公式発表。機能、比較ベンチマーク、費用、安全機構、提供状況。 閲覧日：2026 年 9 月 23 日。

[2] **Artificial Analysis.** Claude Opus 5.5 takes the top spot on the Artificial Analysis Intelligence Index, along with a 20% price cut and larger cache hit discount. 2026 年 9 月 22 日.

https://artificialanalysis.ai/articles/claude-opus-5-5

第三者測定。総合指標、知識労働、推論設定と出力トークン量。 閲覧日：2026 年 9 月 23 日。

[3] **Anthropic.** Claude Opus 5.5 — Claude Platform Docs. 随時更新.

https://platform.claude.com/docs/en/models/opus-5-5/overview

モデル仕様、入出力、コンテキスト長、推論設定、提供先。 閲覧日：2026 年 9 月 23 日。

[4] **Artificial Analysis.** Claude Opus 5.5 (Adaptive Reasoning, Max Effort, Default Fallback) Intelligence, Performance & Price Analysis. 随時更新.

https://artificialanalysis.ai/models/claude-opus-5-5

調査時点の総合評価、モデル公開形態、仕様。 閲覧日：2026 年 9 月 23 日。

[5] **GitHub.** Claude Opus 5.5 is now available in GitHub Copilot. 2026 年 9 月 22 日.

https://github.blog/changelog/2026-09-22-claude-opus-5-5-is-now-available-in-github-copilot/

GitHub による初期テストと Copilot への提供案内。 閲覧日：2026 年 9 月 23 日。

[6] **Aditi Tuli／Box.** Claude Opus 5.5: Faster and leaner for high quality work. 2026 年 9 月 22 日.

https://blog.box.com/claude-opus-55-faster-and-leaner-high-quality-work

複数文書を使う業務課題での Box 自身の評価。 閲覧日：2026 年 9 月 23 日。

[7] **Anthropic.** Prompting Claude Opus 5.5 — Claude Platform Docs. 随時更新.

https://platform.claude.com/docs/en/build-with-claude/prompt-engineering/prompting-claude-opus-5-5

推論強度、視覚入力、長時間作業、途中停止などの公式運用ガイド。 閲覧日：2026 年 9 月 23 日。

[8] **Artificial Analysis.** Announcing the Artificial Analysis Intelligence Index v4.3. 2026 年 9 月 7 日.

https://artificialanalysis.ai/articles/artificial-analysis-intelligence-index-v4-3

指標改訂と評価方法。閲覧時点の表示には v4.3.2 を含む。 閲覧日：2026 年 9 月 23 日。

[9] **Anthropic.** Pricing — Claude Platform Docs. 随時更新.

https://platform.claude.com/docs/en/about-claude/pricing

通常、キャッシュ、Batch API、Fast mode などの料金。 閲覧日：2026 年 9 月 23 日。

[10] Artificial Analysis. Claude Opus 5.5 Models — Intelligence, Performance & Price Comparison. 随時更新.

https://artificialanalysis.ai/models/releases/claude-opus-5-5

5 段階の推論設定別の総合指標と評価タスク当たり費用。 閲覧日：2026 年 9 月 23 日。

[11] Katie Parrott／Every. Vibe Check: Opus 5.5 Is Pulling Our Codex Converts Back to Claude. 2026 年 9 月 22 日.

https://every.to/vibe-check/vibe-check-opus-5-5-is-pulling-our-codex-converts-back-to-claude

7 日間の先行利用レビュー。先行アクセスの提供と編集への不関与を開示。 閲覧日：2026 年 9 月 23 日。

[12] Reddit r/ClaudeAI. Opus 5.5 in Claude Code is crazy fast, especially at spotting UI bugs. 利用者投稿.

https://www.reddit.com/r/ClaudeAI/comments/1wnil7n/opus_55_in_claude_code_is_crazy_fast_especially/

公開直後の体験談。統計的な利用者調査ではない。 閲覧日：2026 年 9 月 23 日。

[13] Anthropic. Why Claude switched models in your conversation with Opus 5 or Opus 5.5 — Claude Help Center. 随時更新.

https://support.claude.com/en/articles/16049681-why-claude-switched-models-in-your-conversation-with-opus-5-or-opus-5-5

分野別フォールバック、アプリと API の違い。 閲覧日：2026 年 9 月 23 日。

[14] Anthropic. Introducing the Life Sciences Verification Program. 2026 年 9 月 17 日.

https://www.anthropic.com/news/life-sciences-verification-program

研究組織向け制度、利用条件、30 日間のデータ保持。 閲覧日：2026 年 9 月 23 日。

[15] METR. Summary of METR's predeployment evaluation of Claude Opus 5.5. 2026 年 9 月 22 日.

https://metr.org/blog/2026-09-22-claude-opus-5-5/

AI 研究開発能力に関する公開前評価と、その対象・限界。 閲覧日：2026 年 9 月 23 日。

[16] Anthropic. What's new in Claude Opus 5.5 — Claude Platform Docs. 随時更新.

https://platform.claude.com/docs/en/models/opus-5-5/whats-new-opus-5-5

API の非互換変更、対応機能、応答形式の変更。 閲覧日：2026 年 9 月 23 日。

[17] Anthropic. How Claude marks AI-generated content — Claude Help Center. 随時更新.

https://support.claude.com/en/articles/16266773-how-claude-marks-ai-generated-content

Opus 5.5 を含むモデル別の電子透かし対応。 閲覧日：2026 年 9 月 23 日。

[18] Anthropic. How Claude's text watermarking works. 2026 年 9 月 1 日更新.

https://www.anthropic.com/news/claude-text-watermark

透かしの仕組み、翻訳・校正への適用、識別・権利に関する説明。 閲覧日：2026 年 9 月 23 日。

[19] Anthropic. Choosing the right model — Claude Platform Docs. 随時更新.

https://platform.claude.com/docs/en/about-claude/models/choosing-a-model

Opus 5.5 を起点とするモデル選択と Fable 5.1 への切替方針。 閲覧日：2026 年 9 月 23 日。