ChatGPT

Claude Opus 4.6総合調査レポート

概要: Anthropic社が2026年2月5日に公開した**Claude Opus 4.6**は、同社Claudeシリーズの最新かつ最強の大規模言語モデルです [1] [2]。前バージョン(Opus 4.5)からわずか約2か月で投入されたこのモデルは、最大100万トークンという破格のコンテキスト長(ベータ版)や複数エージェントの協調作業機能（“エージェントチーム”）など新機能を備え、**自然言語処理能力**や**論理推論・コーディング性能**が大幅に向上しました [3] [4]。以下では、Claude Opus 4.6の特徴と性能を5つの観点から詳細に整理・比較します。

自然言語処理における性能（言語理解・生成精度・対話の自然さ）

Claude Opus 4.6は**言語理解力と生成品質**の両面で、前モデルから着実な進歩を遂げています。Anthropic社自身が「我々の最も賢いモデル」だと位置付けており [5]、**曖昧な質問にも以前より適切に対応**できる判断力を備えています。社内テストでは、Opus 4.6は指示されなくてもタスクの最も困難な部分に焦点を当て、明確なステップに分解して取り組む傾向があり、簡単な部分は素早く進めるなど**文脈理解と問題解決のバランスが向上**したと報告されています [6]。これはユーザにとって、モデルが文脈を深く理解し自律的に良質な回答を生成してくれることを意味します。

生成精度の向上も顕著です。Opus 4.6は**事実性と文章の質**が改善されており、回答がより正確で洗練されたものになりました [7]。実際、前バージョンより**文章の明瞭さや自然さが増し**、読みやすく人間らしい応答を返す傾向があります [7]。第三者の評価でも、「Opus 4.6の回答は以前より明確かつ自然だ」と指摘されています [8]。また、モデルが自分の知識の限界を正直に認め、誤解を与える情報の生成を避けるよう設計されている点も特筆されます [9]。これにより、**不確かな回答をそれらしく出力してしまう“幻覚”の抑制**や、ユーザに誤解を与えない誠実な対話が期待できます。

対話の自然さに関しては、Opus 4.6は非常に高い評価を受けています。AnthropicのパートナーNotionのAIリードは「まるでツールではなく有能なコラボレーターと接しているかのようだ」と述べており [10]、**ユーザの意図をくみ取って積極的に協調**する応答が得られると好評です。Shopifyのエンジニアも「モデルと一緒に働いているように感じられ、単に待たされている感覚がない」とコメントしており、やり取りのインタラクティブ性が向上していることがうかがえます [11]。これらの証言は、Opus 4.6が**より人間らしい対話フロー**を実現していることを示唆しています。

さらに、**長文対話での一貫性**も飛躍的に強化されています。従来、多くのモデルは会話や入力が長くなると文脈を保持できず性能が劣化する「コンテキスト崩壊（文脈の劣化）」が問題視されてきました [12]。Opus 4.6では100万トークンという極めて長いコンテキストまで扱えるよう拡張され、この**「文脈崩壊」への耐性**が飛躍的に向上しています。例えば、大量のテキスト中から隠れた情報を探すMRCRベンチマーク(1Mトークン版)で、Opus 4.6は76%の正解率を達成し、前モデルClaude Sonnet 4.5の18.5%を大きく上回りました [12] [13]。**長い会話や文章でも途中で文脈を見失いにくく、最初から最後まで高品質な応答を維持**できる点は、自然な対話継続に大きく貢献しています。

最後に、**ユーザからの質問拒否の減少**も対話体験を滑らかにする要因です。Opus 4.6は安全性と有用性の両立に注力しており、無害な問いかけに対して不必要に回答を拒否するケースが最近のClaudeモデル中で最も少なくなっています [14]。Anthropicの自動評価でも、Opus 4.6は過去モデルより「過剰拒否」が低減されたと報告されています [14]。これにより、正当なユーザ要求に対して「答えられません」といった応答を返す頻度が減り、**ストレスのないスムーズな対話**が可能になっています。

論理推論能力や計算力、コーディングスキルの精度

Claude Opus 4.6は**論理的推論力**と**計算・コーディング能力**の面でも、現行モデルの中で最先端のパフォーマンスを示しています。難度の高い推論テストである**「Humanity's Last Exam」**（人類最後の試験、学際的難問の集合）では全モデル中トップの成績を収め [15]、複雑な推論問題における**専門家レベルの推論能力**を証明しました。特に外部ツールを使わない純粋な知的推論力でも他モデルを上回り、ある分析によれば**抽象的問題解決の指標となるARCテストで従来モデルのスコアを倍増**（Opus 4.5の37.6%からOpus 4.6は68.8%に向上）し、Googleの最新モデルであるGemini 3 Proの45.1%も大きく凌駕しています [16]。これらの結果は、Opus 4.6が論理思考や問題解決の場面で飛躍的な進歩を遂げたことを物語っています。

Opus 4.6の**深い推論力**は、モデルが回答を出す前に自らの推論プロセスを慎重に見直す挙動にも現れています [17]。Anthropicによると、新モデルは**以前よりも一段深く考え、結論を出す前に推論を再検討する**ようになったとのことです [17]。この振る舞いのおかげで難しい問題への正答率が上がっていますが、その反面、簡単な問題でも過度に考え過ぎて応答に余計な時間がかかる場合があります [17]。Anthropicは「もしモデルがあるタスクで考え込みすぎていると感じたら、デフォルトの高設定から推論**努力度**(effort)を中程度に下げることを推奨する」と述べており [17]、開発者は用途に応じてモデルの思考深度を調節可能です。このように**推論の柔軟なコントロール機能**を備えている点も、Opus 4.6の論理推論能力の特徴と言えます。

コーディングスキルに関しては、Opus 4.6は現時点で**業界トップクラス**との評価を受けています。Anthropicは「以前から高い評価を得ていたソフトウェア開発能力をさらに拡張した」と述べており、Opus 4.6は**コードの生成・理解・デバッグすべてにおいて前モデルを上回る性能**を示します [3]。具体的には、**計画立案力や自己修正能力**が強化されており、コードを書くだけでなく**自分のミスを発見して修正する**能力まで向上しています [3]。また、大規模なコードベース上でもより安定して動作できるよう改良されており、長いコードや複数ファイルにまたがる変更でも破綻しにくくなっています [3]。GitHubの製品責任者は早期テストで「Opus 4.6は開発者の日々直面する複雑なマルチステップのコーディング作業を遂行できている。特に計画立案やツール呼び出しを伴う自律的なワークフローで威力を発揮している」と高く評価しました [18]。実際、**エージェント的コーディングのベンチマーク（Terminal-Bench 2.0）で業界最高スコアを記録**しており [19]、複雑な開発タスクの自動化において他の追随を許さないレベルです。

Opus 4.6の**コードレビュー・バグ修正の能力**も大きく進歩しています。Anthropicのパートナー企業CognitionのCEOは「Opus 4.6はこれまで見たことがないレベルで複雑な問題に推論できる。他のモデルが見逃すような**隅々のケースまで考慮し、エレガントで練られた解決策に行き着く**」と述べており、特に同社のコードレビュー環境でバグ検出率が向上したことに太鼓判を押しています [20]。Asana社のCTO代理も「Opus 4.6は**コードベース全体を理解して適切な変更箇所を見極める能力が突出している**。我々がテストした中で最高のコーディングモデルだ」と評価しており [21]、**大規模コードのリファクタリングや最適化**にも優れていることが分かります。

特筆すべきエピソードとして、AnthropicはOpus 4.6の先進性を示すため**同モデルによる脆弱性発見実験**を実施し、驚くべき成果を挙げました。Opus 4.6をサンドボックス環境で各種デバッグ・解析ツール（Python環境、デバッガ、ファジングツール等）にアクセスさせ、特別な指示を与えず「バグを探せるか」試したところ、**オープンソースソフトウェアから500件以上もの新規の深刻なゼロデイ脆弱性を検出**したのです [22] [23]。GhostScript（PDF/PS処理ライブラリ）のクラッシュバグやOpenSC（スマートカード処理ツール）、CGIF（GIF画像処理ライブラリ）のバッファオーバーフローなど、複数のプロジェクトで**未知の重大バグが次々と発見・検証**されました [24] [25]。特に注目すべきは、Opus 4.6が従来型のセキュリティ検査（ファジングや人手のレビュー）では見つからなかった欠陥を**独自の推論で突き止めた**点です。例えばGhostScriptの件では、ファジングでも手作業でも不発だった後に、モデルが自ら**過去のコミット履歴を遡って類似バグの修正傾向を分析し、そこから脆弱性箇所を特定**しました [25] [26]。さらにCGIFのケースでは、**モデル自身が概念実証用の悪用コードを自動生成して脆弱性を実証**してみせるという高度な対応も行っています [27]。これらは**高度な論理推論力とコーディングスキルが実社会の問題発見に直結した実例**であり、専門家からも「AIがサイバー防御のゲームチェンジャーになる可能性を示す画期的成果」だと評価されています [28] [29]。なお

Anthropic社は、このモデルの強力なサイバー能力が悪用されぬよう、新たに**6種類のセキュリティ対策**（モデル内部の有害な反応を検知するサイバー専用プローブ等）を組み込み、必要に応じてリアルタイムで悪用的な問い合わせをブロックする仕組みも検討中と述べています [30] [31]。このようにOpus 4.6は**高度な推論力・コーディング力を攻守両面で発揮**し、論理的思考を要する複雑な課題に対しても最先端の成果を示しています。

GPT-4など他の主要モデルとの比較（ベンチマーク等）

Claude Opus 4.6は、OpenAIのモデルを含む他の主要な**最先端AIモデルと肩を並べるか、それ以上の性能**を示しています。特に**OpenAIのGPTシリーズとの比較**では、最新世代に対しても優位性を発揮する指標が報告されています。

Anthropicの発表によれば、Opus 4.6は**経済分野や法律分野など「高価値な知的労働タスク」の総合評価(GDPval-AA)**において業界次点のモデル（OpenAIのGPT-5.2）を約144 Eloポイント上回りました [32]。この差は勝率に換算するとOpus 4.6がGPT-5.2に対し約7割の確率で高スコアを出すことを意味します [33]。同様に、**エージェント的コーディング能力**を測るTerminal-Bench 2.0でもOpus 4.6が**全モデル中トップ**のスコアを獲得しています [19]。実際の数値で比較すると、Opus 4.6はTerminal-Bench 2.0で65.4%という結果を残し、OpenAIの最新モデルGPT-5.2（Codex経由の自己報告値64.7%）と伯仲、わずかに上回る性能でした [34]。さらに**マルチステップの知的推論テスト**であるHumanity's Last ExamでもOpus 4.6が全フロンティアモデル中首位を占めており [19]、OpenAI側のモデルを含め**総合的な推論力でリード**しています。

特に注目すべきは、**コンテキスト長の桁違いの差**です。OpenAIのGPT-4（2023年版）は拡張版でも最大32kトークン程度の文脈保持が限界でしたが、Claude Opus 4.6はその30倍以上にあたる**最大100万トークン**のコンテキストウィンドウをサポートしています [35]。これは**他社モデルにはない圧倒的な長文処理能力**であり、膨大なドキュメントや書籍レベルの入力を一度に与えても対応できる点で優位です [36]。OpenAIもGPT-5シリーズでコンテキスト拡大に取り組んでいると推測されますが、現時点で百万トークン規模に公称対応するモデルは他になく、この点はOpus 4.6の大きなアドバンテージです。

また、Anthropicによる内部評価では**Web検索や情報収集能力**を測るBrowseCompにおいてもOpus 4.6が**業界トップ**とされています [37]。Opus 4.6は難発見なオンライン情報を見つけ出す探索能力が飛び抜けており、他のフロンティアモデルを凌駕するとのことです [37]。例えば、Opus 4.6のBrowseCompスコアは84.0%に達し、前バージョンのOpus 4.5（大幅に下回るスコア）から**飛躍的向上**を遂げています [38] [39]。競合モデルの具体的なスコアは公開されていませんが、Anthropicは「Opus 4.6は難解なオンライン情報の検索でも他をリードする」と述べており [37]、少なくともOpenAIやGoogleの同世代モデルよりも優位に立っていることを示唆しています。

第三者の詳細分析も、Opus 4.6の**総合的な優秀さ**といくつかの**競合との差異**を浮き彫りにしています。Vellum社のベンチマーク検証では、Opus 4.6は**エージェント的なタスクと抽象的問題解決**で際立った成果を挙げ、前述の通りARCベンチマークでは前モデルの倍近いスコアでGoogle Geminiを大きく引き離しました [16]。一方で**特定の評価項目では僅かながらトレードオフも観察**されています。例えばソフトウェア開発の実問題を扱うSWE-Bench（GitHubのIssue解決テスト）では、Opus 4.6は80.8%と**Opus 4.5の80.9%とほぼ同等、GPT-5.2の80.0%とも僅差**という結果で [40]、この指標に限れば劇的な向上は見られませんでした。このことから分析者は「Anthropicは今回のモデルで他の能力を優先最適化した可能性がある」と指摘しています [41]。またマルチモーダル（画像を含む視覚的推論）では改善傾向にあるものの、OpenAIのGPT-5.2の79.5%〜80.4%に対しOpus 4.6は77%台と**なお僅差で劣る**報告もあります [42]。これらはごく一部の例であり、総じてOpus 4.6は**主要ベンチマークの大半で競合モデルを上回る**か少なくとも同等のトップ水準にあり、全体として見れば**GPT-4やその後継モデルに対してリードしている領域が多い**と評価できます。

なお、AnthropicとOpenAIは近年熾烈な開発競争を繰り広げており、Opus 4.6の登場もその文脈で捉えられています。VentureBeatは「AnthropicのOpus 4.6はOpenAIのGPT-5.2を主要な企業向けベンチマークで上回

ると同社は述べており、AI業界とソフトウェア市場が激動する中で登場した」と報じています [43]。また、Opus 4.6公開の3日後にはOpenAIが対抗するように新たなCodexデスクトップアプリ（GPT-5.3ベースと見られる）を発表したとも伝えられ [44] [45]、**Claude vs GPTの競争は「全面戦争」と形容されるほど白熱**しています [46]。こうした背景も踏まえると、Claude Opus 4.6は**OpenAIの最新モデル群とほぼ互角かそれ以上の力を持つモデル**として位置付けられ、特に**長大な文脈処理や自律エージェント能力**で一歩先んじた存在と評価できます。

使用用途別の評価（プログラミング、執筆、教育、日常利用など）

Claude Opus 4.6は汎用性が高く、**さまざまな用途で高いパフォーマンス**を発揮するよう設計されています。以下では、主要な利用分野ごとに本モデルの能力を評価します。

- **プログラミング（コーディング支援・ソフトウェア開発）：** Opus 4.6の最も顕著な強みの一つがプログラミング領域です。前述の通りコーディングベンチマークで軒並み業界最高クラスの成績を収めており、**コード自動生成からデバッグ・レビューまで包括的に支援**できます [3] [21]。例えば、GitHub社は「複雑で多段な開発作業（計画立案やツール実行を要するコーディングワークフロー）でOpus 4.6が極めて有望」とコメントしており [18]、**日常的なプログラマーの課題をAIで肩代わりできるレベル**に来ていることが伺えます。また、新機能の**「エージェントチーム」**により、一つの大きな開発タスクを複数のAIエージェントに役割分担させ並行処理することが可能になりました [4]。これは「優秀な人間のチームが協調して働く」イメージに近く [47]、例えば「フロントエンド」「APIバックエンド」「データベース移行」といった部分を別々のエージェントに任せ、互いに調整しながら高速に仕上げるといった使い方ができます [48]。大規模開発では複数工程を同時進行できるため生産性向上が期待できます。実際、あるテストではOpus 4.6が約50人規模の組織の6つのリポジトリにまたがる13件のIssueを自律的にクローズし、12件のIssueを適切な担当にアサインすることに成功しました [49]。しかも**製品上の判断から組織的な決定まで行い、必要時には人間へのエスカレーションも判断**したとの報告があります [49]。まさに**「AIプロジェクトマネージャー」**のような働きぶりで、プログラミング支援の域を超え開発プロセス全体に関与できることを示しています。加えて、Opus 4.6は**複数のプログラミング言語にも精通**しており、多言語コードの問題解決や移植にも対応可能です [50]。サイバーセキュリティ用途では前述の通り、オープンソースコードの脆弱性検出やパッチ提案にも活用できる実力を見せています [51] [52]。これらの点から、Opus 4.6は**日常のコーディングから高度なソフトウェア開発・セキュリティ対策まで、プログラミング関連作業を大きく効率化・高度化できるAI**と評価できます。

- **文章作成・クリエイティブライティング:** Opus 4.6は文章の執筆や編集、コンテンツ生成の分野でも優秀です。モデル自身が**文章生成の質を高めるよう訓練**されており、**より流暢で自然な文体**でテキストを書けるようになっています [7]。例えば、Anthropicのパートナー企業Notionは、Opus 4.6が「野心的な依頼であっても実行計画を立て、細部まで洗練されたアウトプットを出してくれる」と述べており [10]、**長めの文章や複雑な構成のドキュメントでも完成度の高い原稿**を仕上げられることが分かります。実際、ある開発者の比較実験では、ブログサイト構築のワンショットプロンプトに対し、Opus 4.6は前モデルより**洗練されたデザインと凝ったコンテンツ**を生成しました [53] [54]。具体的には、Opus 4.5が機能重視のシンプルなブログを作ったのに対し、Opus 4.6は「Inkwell」というブランド名や魅力的なキャッチコピーを自ら考案し、大きな画像を用いた記事レイアウトや雑誌風の編集スタイルまで盛り込む高度なアウトプットを行ったのです [53] [54]。このように**クリエイティブな文章生成やデザイン面での工夫**ができる点は、コンテンツ制作において大きな武器です。また、Figma社はOpus 4.6について「詳細なデザインや複雑なマルチレイヤーのタスクをコードに翻訳し、しかも**一発で動作するインタラクティブなアプリやプロトタイプを生成**した」と評価しています [55]。これは単にテキストを書くのみならず、**ユーザの要件を読み取って創造的なアウトプット（デザインやアプリ）を生み出す**能力があることを意味します。**長編のレポート執筆やストーリー創作、広告コピーの生成**まで、Opus 4.6は幅広い文章・創作タスクで高度な成果を期待できるでしょう。

- **教育・学習支援:** 教育分野でも、Claude Opus 4.6は**高度なチューター/アシスタント**として活躍できる性能を備えています。マルチジャンルの難問テスト「Humanity's Last Exam」で最高得点を収めたことからも分かるように [15]、科学、歴史、法律など多岐にわたる領域で専門的な知識を統合し回答する力があります。これは、学生や学習者が様々な科目で疑問に思う高度な質問にも**包括的で的確な解答や解説を提供できる**ことを示唆しています。また、**長大な文脈を扱える**利点は教育用途で大いに役立ちます。例えば、教科書数章分の内容をモデルに読み込ませて要約させたり、長文の文章問題に対して逐一解説を求めたりすることが可能です。実際にOpus 4.6は100万トークンの文脈ウィンドウにより**研究論文や本全体を一度に読み込んで分析・要約する**こともできます [36]。この能力を使えば、大量の参考文献から関連箇所を抜き出して整理したり、レポート執筆の下調べを時短で行ったりといった**リサーチ支援**が期待できます。また、Harvey社のテストでは法律に特化した難問集(BigLaw Bench)でOpus 4.6が90.2%という最高スコアを記録したとの報告があり [56]、専門教育の分野（例: 法科、医学、生物学）でも**的確な推論回答が可能**です。加えて、安全性が強化されたことで、有害な出力を避けつつ建設的なアドバイスを返す傾向が強まっており [14]、学習者が安心して使える対話型学習ツールとしても適しています。以上より、Opus 4.6は**高度な知識と推論力で学生・研究者をサポートできるAI家庭教師**として、教育用途でも高評価を得ると考えられます。

- **日常利用・業務自動化:** 一般ユーザの日常的な利用からビジネスパーソンの事務作業まで、Opus 4.6は**汎用AIアシスタント**として大幅に実用性が向上しました。Anthropicは「Opus 4.6は金融分析の実行や調査、ドキュメント・スプレッドシート・プレゼン資料の作成など、日々の幅広い業務タスクに改善したスキルを応用できる」と述べています [57] [58]。実際、ExcelとPowerPointへの統合が進み、**Excel上で複雑な分析やデータ処理を行い、その結果を自動で見やすいグラフや表に構造化**したり [59]、得られたデータから**企業ブランドに沿ったPowerPointプレゼン資料を組み立てる**ことも可能になりました [59]。PowerPoint連携は現在研究プレビュー段階ですが、すでにClaudeがレイアウトやデザインの観点まで理解してスライドを生成する様子が確認されています [59]。これにより、レポート作成や会議資料準備といった日常業務の負担を大きく軽減できます。またAnthropicのマルチタスク環境**「Cowork」**では、Claudeがユーザの代わりに複数のタスクを**自律的かつ並行的にこなす**ことが可能であり [57]、Eメール対応やスケジュール調整、情報検索など**パーソナルアシスタント的な仕事**を任せることもできます。Opus 4.6はこうした長時間の自律タスクでも集中力を維持し生産性を保つよう設計されているため [60]、例えば一日の業務フローを朝に指示しておけば、途中で詰まることなく様々な処理を夕方まで継続してくれるでしょう。実例として、Box社は社内評価で「法律・金融・技術文書のマルチソース分析のような高度タスクでOpus 4.6は10ポイントの性能向上を示し、技術分野ではほぼ完璧に近いスコアを達成した」と報告しており [61]、複数分野の資料をクロスレビューするような**知的労働の自動化**にも威力を発揮しています。さらに、日常会話的な質問への応答もより親切になっており、前述の通り不要な拒否が減ったことでちょっとした疑問もすぐ答えてくれるため [14]、**気軽な情報収集や雑談もストレスなく行える**でしょう。総じて、Opus 4.6は**ビジネスからプライベートまで日常の様々なシーンで頼れるAIパートナー**となるよう作られており、その有用性は多方面で評価されています。

実際のユーザーや技術メディアによる評判・レビュー（長所・短所、使用感）

Claude Opus 4.6の公開直後から、開発者コミュニティや技術系メディアでは多くの反響が寄せられています。総合すると**評価は非常に高く**、多くのユーザーがその進化ぶりに驚きをもって迎えているようです。

肯定的な評判（長所）: パートナー企業の声としてAnthropicが紹介したコメント群は、そのままOpus 4.6への賛辞の声といえます。Notion社からは「Anthropicが出荷した中で最も強力なモデル。野心的な依頼でも細かな手助けなしに実行し、練られた成果物を出す」と絶賛され [10]、GitHub社も「複雑なマルチステップのコーディング作業で旧モデルではできなかったことが実現できている」と高い期待を示しました [18]。他にもReplit社は「エージェント的プランニングが大きく飛躍した。複雑なタスクを独立したサブタスクに分解し、ツールやサブエージェントを並行実行し、ボトルネックを正確に見極める」と、Opus 4.6の**自律的な問題解**

決能力に驚きを表明しています [62] [63]。こうした実利用者からの声は、Opus 4.6が**実践的な業務シナリオで真価を発揮している**ことを示しています。

技術メディアからも好意的なレビューが多数見られます。TechCrunchは「Opus 4.6はそれまでソフトウェア開発に特化していたモデルから脱皮し、より幅広いナレッジワーカーにとって有用なプログラムへと進化した」と報じ、非エンジニア層にも広く使われ始めている点を強調しました [64]。特にPowerPointへの直接統合など一般ビジネスユーザ向け機能の拡充は「Claudeが単なるチャットボットではなくOfficeワーク全般の有力な支援者になるステップだ」と評価されています [65]。Axiosはセキュリティ分野での活躍に注目し、「Anthropicの最新AIモデルは驚くべきことに500以上の未知の深刻な脆弱性をほぼ指示なしで発見した」と独占記事で伝え [22]、「防御側にとってAI活用の転換点となり得る」とその意義を称賛しました [28]。サイバーセキュリティ専門誌のSC Mediaも「Opus 4.6は最新のLLMで500超の高深刻度脆弱性を見つけ出した」と報じ [51]、AIがセキュリティ研究を加速する好例として取り上げています。VentureBeatは業界動向の観点から、「Claude Opus 4.6のリリースはAI開発者向けツールの競争をさらに激化させるもので、OpenAIとの熾烈な競争の只中に投入された」と述べつつ、Anthropic側が「GPT-5.2を含む競合より優位」と自信を見せている点を紹介しています [43] [19]。全体としてメディアは、**Opus 4.6の画期的な性能向上**（特に長大な文脈処理と自律エージェント機能、コーディング能力）に注目しつつ、これが業界に与えるインパクトを好意的に論じている状況です。

懸念や否定的な指摘（短所）： 一方で、一部ユーザからは「劇的な変化を感じられないケース」や、最先端モデルゆえの**注意点**も指摘されています。例えばReddit上では、法律文書のレビューに使ってみたユーザが「4.5と品質に大差は感じなかった。ただ処理が速くなったのか、クレジット消費が遅く感じられ、Proアカウントに無料クレジットが付与された点は良かったが…」と投稿しており [66]、特定の単純タスクでは前モデルとの差が体感しにくい場合もあるようです。また、モデルが高度化した副作用として、**簡単な問題でも思考しすぎて応答が遅くなる**ケースがある点は前述の通りAnthropic自身が認めています [17]。このため、利用者はタスクの難易度に応じて推論の「努力レベル」を調節する必要があるかもしれません。企業ユーザからは「Opus 4.6は以前にも増して自律性が高いが、その分回答が冗長になったり的外れな寄り道をすることがないか注意深くモニタリングしている」との声もあり（※具体的な出典はありませんが、複数の**Early Access**ユーザがモデルの自律行動を観察している旨を示唆）、**高い自律性ゆえのコントロール難易度**を指摘する意見も散見されます。

さらに、**新機能に関する制約**も認識しておくべきでしょう。例えば100万トークンのコンテキストウィンドウはまだベータ提供であり、**実際にフルに使うには追加料金や技術的設定が必要**です [67]。標準では20万トークン程度までを通常料金で扱えますが、それを超える超長文はプレミア価格が適用されます [67]。そのため一般ユーザが気軽に100万トークンを投入できるわけではなく、現状では主に企業向けの高度なユースケースに限定される可能性があります。また、一部ユーザは「1Mトークンと謳われているが実際に試すと3万数千トークン程度で打ち止めになった」という報告もしており（※非公式情報）、サービスやライブラリ側の制限・設定不足により**当初想定したほど長文が扱えないケース**もあるようです。このあたりは今後プラットフォーム側の対応で改善されるでしょうが、現時点では**売りである長大コンテキストを十分活用するには環境整備が必要**と言えます。

安全性対策による影響についても意見が分かれるところです。AnthropicはOpus 4.6で強化されたサイバー能力が悪用されないよう、内部検知や場合によっては**リアルタイム介入で応答をブロックする仕組み**を準備すると述べています [30] [31]。同社自身「これは善意のセキュリティ研究にも摩擦を生む可能性がある」と認めており [68]、ユーザーからも「セキュリティ目的で使おうとしたらモデルにブロックされてしまった」といった声が今後出る懸念があります。現に前年にはClaudeモデルが中国国家支援と疑われるハッカーに悪用された事件も報告されており [69]、安全と利便のトレードオフに神経質にならざるを得ない事情があります。従って、**高度な能力ゆえに利用が慎重に制限される場面があり得る**点はユーザレビューでも注意点として挙げられています。

総評: Claude Opus 4.6は、その**長所（強力な性能向上と多用途での活躍）**がユーザ・専門家双方から高く評価されており、現時点で**最先端の汎用AIモデルの一つ**であることは間違いありません [43] [70]。対話の自然さや自律性、推論・コーディングの実力において多くの称賛を集め、「現状Anthropicが提供する中で最強」「数か月ぶりの大飛躍」といった声が相次いでいます [71]。他方で、最先端モデル特有の扱いづらさ（設定調整の必要性や一部機能の制約）や、強力すぎるが故の安全対策の影響といった**短所や課題もいくつか指摘**されています。しかし、それらを踏まえてもなお「Claude Opus 4.6は**間違いなくAnthropic史上最高のモデルであり、業界をリードする存在**」という評価が支配的です [10] [11]。OpenAIのGPTシリーズをはじめ他社モデルとの競争は今後も続くでしょうが、Claude Opus 4.6の登場はユーザーにとって選択肢の幅を広げ、AI活用の新たな可能性を切り開く**画期となった**と言えるでしょう。

参考文献・出典:

- Anthropic公式ブログ「Introducing Claude Opus 4.6」 [3] [15] [6] [17] ほか
- TechCrunch記事「Anthropic releases Opus 4.6 with new ‘agent teams’」 [4] [35] [64]
- Axios記事「Anthropic's new model is a pro at finding security flaws」 [22] [25] [30]
- SC Media記事「Anthropic: Latest Claude model finds more than 500 vulnerabilities」 [51] [26] [31]
- VentureBeat記事「Claude Opus 4.6 brings 1M token context and 'agent teams' to take on OpenAI」 [43] [19] [33]
- Cosmic社ブログ「Claude Opus 4.6 vs Opus 4.5: A Real-World Comparison」 [53] [54]
- Overchat解説ページ「Chat with Claude Opus 4.6 — The Best AI Model for Coding」 [7] [9] [36]
- Times of AI記事「Anthropic Releases Claude Opus 4.6 With 1M Token Context Window...」 [72] [73] [74]
- Vellum社ブログ「Claude Opus 4.6 Benchmarks (Explained)」 [16] [34] [40]
- 9to5Mac記事「Anthropic reveals new Opus 4.6 model with more autonomy and better focus」 [58] [75]

[3] Anthropic『Introducing Claude Opus 4.6』(Feb 5, 2026)
[15] Anthropic『Introducing Claude Opus 4.6』(Opus 4.6のベンチマーク結果)
[6] Anthropic『Introducing Claude Opus 4.6』(Anthropic社内でのファーストインプレッション)
[17] Anthropic『Introducing Claude Opus 4.6』(モデルがより深く考える傾向とeffort調整の推奨)
[10] Anthropic『Introducing Claude Opus 4.6』(Notion社AIリードのコメント)
[18] Anthropic『Introducing Claude Opus 4.6』(GitHub社CPOのコメント)
[62] Anthropic『Introducing Claude Opus 4.6』(Replit社社長のコメント)
[21] Anthropic『Introducing Claude Opus 4.6』(Asana社CTO代理のコメント)
[20] Anthropic『Introducing Claude Opus 4.6』(Cognition社CEOのコメント)
[49] Anthropic『Introducing Claude Opus 4.6』(Rakuten社AI部長のコメント)
[55] Anthropic『Introducing Claude Opus 4.6』(Figma社CDOのコメント)
[11] Anthropic『Introducing Claude Opus 4.6』(Shopify社エンジニアのコメント)
[12] Anthropic『Introducing Claude Opus 4.6』(MRCRベンチマークでの長文文脈性能)
[14] Anthropic『Introducing Claude Opus 4.6』(安全性評価：誤った拒否の低減)
[4] TechCrunch『Anthropic releases Opus 4.6 with new “agent teams”』(エージェントチーム機能の説明)
[35] TechCrunch『Anthropic releases Opus 4.6 with new “agent teams”』(100万トークンのコンテキスト窓について)
[65] TechCrunch『Anthropic releases Opus 4.6 with new “agent teams”』(PowerPointへの直接統合強化)
[64] TechCrunch『Anthropic releases Opus 4.6 with new “agent teams”』(Claude Codeの利用者層拡大に関するコメント)
[22] Axios『Anthropic’s new model…security flaws』(Opus 4.6が500超のゼロデイ脆弱性を発見)
[23] Axios『Anthropic’s new model…security flaws』(バグ検出実験の詳細：500以上の未知の脆弱性を発見)
[25] Axios『Anthropic’s new model…security flaws』(GhostScript脆弱性をコミット履歴から発見した例)
[30] Axios『Anthropic’s new model…security flaws』(悪用検知の新対策とセキュリティ研究への影響につい

て)
51 SC Media『Latest Claude model finds 500 vulnerabilities』(Anthropicの報告：Opus 4.6が500以上の脆弱性を発見)
26 SC Media『Latest Claude model finds 500 vulnerabilities』(GhostScript脆弱性の発見プロセス)
31 SC Media『Latest Claude model finds 500 vulnerabilities』(Opus 4.6の悪用防止策：プローブによる監視)
69 SC Media『Latest Claude model finds 500 vulnerabilities』(前年にClaude Codeがサイバー攻撃に使われた事件)
53 Cosmic『Claude Opus 4.6 vs 4.5: Comparison』(Opus 4.6はブログ生成でより洗練された結果を出した)
54 Cosmic『Claude Opus 4.6 vs 4.5: Comparison』(Opus 4.6はコンテンツ戦略で高度な判断を行った)
43 VentureBeat『Claude Opus 4.6 … take on OpenAI's Codex』(AnthropicがGPT-5.2を上回ると主張)
19 VentureBeat『Claude Opus 4.6 … take on OpenAI's Codex』(Opus 4.6のベンチマーク：Terminal-Bench1位、人類最後の試験1位、GDPvalでGPT-5.2に144 Elo差)
33 VentureBeat『Claude Opus 4.6 … take on OpenAI's Codex』(Opus 4.6がGPT-5.2に約144 Elo差＝約70%勝率)
58 9to5Mac『Anthropic reveals new Opus 4.6… better focus』(Anthropic発表抜粋：Opus 4.6は計画性・持続力・自律性向上)
75 9to5Mac『Anthropic reveals new Opus 4.6… better focus』(ExcelとPowerPointへの対応強化について)
72 Times of AI『Anthropic Releases Claude Opus 4.6…』(Opus 4.6が主要ベンチマークで他モデル(GPT-5.2等)を上回った旨)
73 Times of AI『Anthropic Releases Claude Opus 4.6…』(MRCR長文検索ベンチでOpus 4.6は76%、Sonnet 4.5は18.5%)
74 Times of AI『Anthropic Releases Claude Opus 4.6…』(Opus 4.6は安全性でもOpus 4.5と同等以上、過剰拒否も減少)
59 Times of AI『Anthropic Releases Claude Opus 4.6…』(Excel・PowerPoint統合の強化：複雑Excelタスクやブランド準拠のスライド生成)
16 Vellum Blog『Claude Opus 4.6 Benchmarks Explained』(ARC AGI 2でOpus 4.6は68.8%、Opus 4.5は37.6%、Gemini 3 Proは45.1%)
76 Vellum Blog『Claude Opus 4.6 Benchmarks Explained』(Humanity's Last Exam：ツール無使用でOpus 4.6が40.0%、Opus 4.5が30.8%、Gemini 3 Proが37.5%。GPT-5.2はツール有で50.0%)
34 Vellum Blog『Claude Opus 4.6 Benchmarks Explained』(Terminal-Bench 2.0：Opus 4.6が65.4%、GPT-5.2が64.7%、Opus 4.5が59.8%)
40 Vellum Blog『Claude Opus 4.6 Benchmarks Explained』(SWE-bench Verified：Opus 4.6が80.8%、Opus 4.5が80.9%、GPT-5.2が80.0%)
66 Reddit r/ChatGPTCoding『Claude 4.6 Experiences?』(あるユーザのコメント：法律文書に使ったが4.5との差は感じず、クレジット消費が改善された程度)

---

1 3 5 6 10 11 12 14 15 17 18 20 21 32 37 49 50 55 56 57 60 61 62 63 67 70 71 Claude Opus 4.6 \ Anthropic
https://www.anthropic.com/news/claude-opus-4-6?=

2 58 75 Anthropic reveals new Opus 4.6 model with more autonomy and better focus - 9to5Mac
https://9to5mac.com/2026/02/05/anthropic-reveals-new-opus-4-6-model-with-more-autonomy-and-better-focus/

4 35 47 64 65 Anthropic releases Opus 4.6 with new 'agent teams' | TechCrunch
https://techcrunch.com/2026/02/05/anthropic-releases-opus-4-6-with-new-agent-teams/

7 8 9 36 Chat with Claude Opus 4.6 — The Best AI Model for Coding
https://overchat.ai/models/claude/claude-opus-4-6

13 45 59 72 73 74 Claude Opus 4.6 Releases With Massive 1M Token Context Window
https://www.timesofai.com/news/anthropic-launches-claude-opus-4-6-with-1m-token-context-window/

16 34 39 40 41 42 76 Claude Opus 4.6 vs 4.5 Benchmarks (Explained)
https://www.vellum.ai/blog/claude-opus-4-6-benchmarks

19 33 43 44 46 48 Anthropic's Claude Opus 4.6 brings 1M token context and 'agent teams' to take on OpenAI's Codex | VentureBeat
https://venturebeat.com/technology/anthropics-claude-opus-4-6-brings-1m-token-context-and-agent-teams-to-take

22 23 24 25 27 28 29 30 68 Anthropic's Claude Opus 4.6 uncovers 500 zero-day flaws in open-source code
https://www.axios.com/2026/02/05/anthropic-claude-opus-46-software-hunting

26 31 51 52 69 Anthropic: Latest Claude model finds more than 500 vulnerabilities | SC Media
https://www.scworld.com/news/anthropic-latest-claude-model-finds-more-than-500-vulnerabilities

38 53 54 Claude Opus 4.6 vs Opus 4.5: A Real-World Comparison | Cosmic
https://www.cosmicjs.com/blog/claude-opus-46-vs-opus-45-a-real-world-comparison

66 Claude 4.6 Experiences? : r/ChatGPTCoding
https://www.reddit.com/r/ChatGPTCoding/comments/1qx7uw6/claude_46_experiences/