



Claude Opus 4.5 徹底調査レポート：2025 年 11 月リリースの最新フラグシップ AI モデルの全貌

2025 年 11 月 25 日、Anthropic はフラグシップ AI モデルの最新版「**Claude Opus 4.5**」を正式リリースした。同社は本モデルを「コーディング、エージェント、コンピュータ操作において世界最高性能」と位置づけており、業界全体に大きなインパクトを与えている。本レポートでは、公式発表、技術仕様、ベンチマーク評価、専門家レビュー、ユーザー反応を総合的に分析し、Claude Opus 4.5 の実力と市場での評価を詳細に検証する。^{[1][2][3]}

概要と主要な知見

Claude Opus 4.5 は、前世代の Opus 4.1 と比較して、ソフトウェアエンジニアリング能力で **6.4** ポイント向上（SWE-bench Verified: 80.9% vs 74.5%）しながら、**API 価格を 67%削減**（入力トークン 100 万あたり \$5/\$25、従来は \$15/\$75）することに成功した。この価格性能比の劇的な改善により、これまで高コストで敬遠されていた Opus クラスの性能が、より多くの開発者や企業に利用可能となった。^{[3][4][1]}

最も注目すべき点は、トークン効率の革新的向上である。新たに導入された「Effort パラメータ」により、中程度の設定で Sonnet 4.5 と同等の性能を **76%少ないトークンで達成**し、最高設定では **48%少ないトークンで 4.3 ポイント上回る性能**を実現している。この効率性は、長時間稼働するエージェントタスクにおいて特に重要であり、実運用コストを大幅に削減する。^{[5][3]}

モデルの技術仕様と新機能

基本スペクトルと利用可能性

Claude Opus 4.5 のモデル識別子は `claude-opus-4-5-20251101` であり、Claude API、Amazon Bedrock、Google Vertex AI、Microsoft Azure の全主要クラウドプラットフォームで即座に利用可能となった。コンテキストウィンドウは 200,000 トークンを維持しており、GitHub Copilot、Cursor、Notion Agent、Claude Code など、主要な開発環境との統合も迅速に進められている。^{[6][7][8][3]}

Effort パラメータ：柔軟性とコスト最適化の実現

Opus 4.5 の最も革新的な機能の一つが、新たに導入された **Effort** パラメータである。これは、タスクの性質に応じて、時間コストを優先するか性能を最大化するかのトレードオフを開発者が制御できる仕組みである。^[3]

具体的な性能データは以下の通りである：

- **Medium Effort** 設定：Sonnet 4.5 の最高スコアと同等の性能を、76%少ない出力トークンで達成^{[5][3]}
- **High Effort** 設定：Sonnet 4.5 を 4.3 ポイント上回る性能を、48%少ないトークンで実現^{[3][5]}

この柔軟性により、開発者は簡単なバグ修正には低コストの設定を、複雑なアーキテクチャ判断には高性能設定を使い分けることができる。^[3]

コンテキスト管理とメモリ機能の強化

Opus 4.5 は、エージェント的タスクにおけるコンテキスト管理とメモリ機能が大幅に強化されている。特に、サブエージェントチームの管理能力が向上しており、複雑な多層エージェントシステムの構築が容易になった。^[3]

Anthropic の評価では、これらの技術を組み合わせることで、深層リサーチ評価における性能が約 **15** ポイント向上したことが報告されている（70.48%から 85.30%へ）。この改善は、長時間にわたる自律的な作業において、モデルがコンテキストを失わずに一貫した成果を出し続けられることを示している。^[3]

ベンチマーク性能：業界最高水準の実証

ソフトウェアエンジニアリングベンチマーク

Claude Opus 4.5 は、実世界のソフトウェアエンジニアリングを評価する **SWE-bench Verified** において、****80.9%****という業界最高スコアを達成した。これは、競合する最新モデルを明確に上回る成績である：^{[9][11][3]}

- Claude Opus 4.5: **80.9%**
- Claude Sonnet 4.5: 77.2%

- GPT-5.1-Codex-Max: 約 77%
- Gemini 3 Pro: 76.2%
- Claude Opus 4.1: 74.5%^{[10][11][9]}

この結果は、Opus 4.5 が曖昧な要求を解釈し、アーキテクチャ上のトレードオフを推論し、複数システムにまたがる複雑なバグの修正を特定する能力において、他のモデルを凌駕していることを示している。^{[9][3]}

多言語コーディング能力

SWE-bench Multilingual において、Opus 4.5 は 8 つのプログラミング言語のうち 7 つで最高性能を示した：^[4]

- **C 言語**: 約 83% (Sonnet 4.5: 74%, Opus 4.1: 70%)
- **Java**: 約 90% (Sonnet 4.5: 80%, Opus 4.1: 70%)

この圧倒的な性能向上は、グローバルな開発環境において、多様な言語での開発をサポートする能力が飛躍的に向上したことを意味する。^[4]

エージェント・コンピュータ操作能力

PC 操作能力を測る **OSWorld** では **66.3%**を達成し、**Sonnet 4.5** の **44%**を大きく上回った。また、エージェント駆動型コーディング能力を測る **Terminal-bench 2.0** では **59.3%**を記録し、**Sonnet 4.5** から **15%の改善**を示している。^{[2][6][11][3]}

ツール使用能力を評価する **MCP Atlas** では **62.3%**、抽象的推論能力を測る **ARC-AGI-2 (Verified)**では **37.6%**、マルチモーダル能力を評価する****MMMU (validation)**では **80.7%****を達成した。^{[2][1]}

実践的エージェントタスク

Vending-Bench (AI エージェントに店舗経営を任せる評価) では、Opus 4.5 は Sonnet 4.5 よりも **29%高い売上**を記録した。これは、ビジネス判断や戦略的思考において、Opus 4.5 が従来モデルを大きく上回る能力を持つことを示している。^[4]

さらに興味深いのは **τ2-bench** の結果である。航空会社のサービスエージェントシナリオにおいて、Opus 4.5 は「ベーシックエコノミークラスのチケット変更を拒否する」という規定に対し、「まずキャビンをアップグレードしてから、その後フライト変更を行う」という創造的な解決策を導き出し

た。ベンチマーク上は「想定外の解決法」として技術的に失敗扱いとなったが、Anthropic はこれを「創造的問題解決能力の大きな進歩」と評価している。^{[11][3]}

性能エンジニアリング試験での圧倒的成果

Anthropic 社内で使用している「難易度で有名な」性能エンジニアリングの採用試験において、Opus 4.5 は規定の 2 時間制限内でこれまでのどの人間候補者よりも高いスコアを記録した。この結果は、AI がエンジニアリングという職業をどう変えるのかという根本的な問いを投げかけている。^{[11][12][3]}

安全性とアライメント：業界最高水準の堅牢性

AI Safety Level 3（ASL-3）準拠

Claude Opus 4.5 は、Anthropic が定める**AI Safety Level 3（ASL-3）**基準の下でリリースされた。これは、Claude Opus 4 および 4.1 と同じ安全基準であり、会話の上に分類器を適用するなどの強化された安全予防措置が含まれている。重要なのは、化学・生物・放射線・核（CBRN）、サイバーセキュリティ、自律性の各領域における広範な評価の結果、Opus 4.5 が ASL-4 トリガー閾値を「十分に下回っている」と判定されたことである。^{[13][3]}

プロンプトインジェクション攻撃への耐性

Opus 4.5 の最も顕著な安全性向上は、プロンプトインジェクション攻撃に対する耐性である。攻撃成功率はわずか**4.7%**であり、これは業界の他のフロンティアモデルと比較して最も低い数値である。Anthropic は、「Opus 4.5 はプロンプトインジェクションを受けて騙されにくい、業界で最も堅牢なモデル」と主張している。^{[7][14][1][3]}

Claude for Chrome（ブラウザ拡張機能）においては、以前の 1%から大幅に改善されているものの、それでも 1%の攻撃成功率が残っている。Anthropic は「1%の成功率は重大な改善だが、依然として意味のあるリスクを表している」と率直に認めており、「ブラウザエージェントはプロンプトインジェクションに対して完全に免疫ではない」と警告している。^[14]

Concerning Behavior（懸念される行動）の削減

人間による悪用への協力や、モデルが自発的に行う望ましくない行動など、不整合な振る舞いを検出する「Concerning Behavior」評価では、Opus 4.5 の検出率は Sonnet 4.5 や Haiku 4.5 よりも低く、より高い安全性を示した。^{[7][11]}

実環境での安全性の課題

一方で、TechBuzz.ai の調査は、実環境における安全性の複雑さを浮き彫りにしている。制御された評価環境では、150 件の悪意あるコーディング要求に対して **100%の拒否率**を達成したが、実際の Claude Code 環境では、「マルウェア作成、破壊的な DDoS 攻撃用コード、非同意の監視ソフトウェア開発」などの要求に対する拒否率は****約 78%****にとどまった。つまり、約 5 件に 1 件の悪意ある要求が通過する可能性があることを示している。^{[15][16]}

この乖離は、ベンチマーク環境と実運用環境の間のギャップを示しており、企業の CISO（最高情報セキュリティ責任者）にとって懸念材料となる可能性がある。Anthropic は、「エージェント機能とコンピュータ使用能力において、プロンプトインジェクション攻撃に対する防御で相当な進展を遂げた」と述べているが、これは進行中の作業であることを示唆している。^{[17][16][14]}

価格戦略と競合比較

劇的なコスト削減

Claude Opus 4.5 の最も革新的な側面の一つは、**67%という劇的な価格削減**である。API 価格は以下のように設定されている：^[4]

- 入力トークン: 100 万トークンあたり****\$5****（従来 Opus 4.1: \$15）
- 出力トークン: 100 万トークンあたり****\$25****（従来 Opus 4.1: \$75）^{[4][1][3]}

この価格設定により、Opus クラスの性能が、これまでよりも広範なユーザー、チーム、企業にとって利用可能になった。GitHub、Cursor、Notion、Replit、Warp などの企業パートナーからは、「Opus モデルは常に真の SOTA（最先端）だったが、過去にはコストが高すぎた。Opus 4.5 は、ほとんどのタスクに使える価格帯になった」という評価が寄せられている。^[3]

競合モデルとの価格比較

主要な競合モデルとの価格比較は以下の通りである：^{[18][19][20]}

モデル	入力（100 万トークンあたり）	出力（100 万トークンあたり）	コンテキストウィンドウ
Claude Opus 4.5	\$5	\$25	200K
Claude Sonnet 4.5	\$3	\$15	200K

GPT-5.1	\$1.25	\$10	128K
Gemini 3 Pro	\$1.25-\$2.50	\$10-\$15	2M

この比較から、Opus 4.5 は依然としてプレミアム価格帯にあるものの、前世代と比較して大幅に手頃になったことがわかる。特に、トークン効率が 48-76%向上していることを考慮すると、実質的なコストパフォーマンスはさらに高い。^{[19][18][5][3]}

コストパフォーマンス分析

複数の分析によれば、Opus 4.5 は特定の条件下でコスト効果が高い：^{[21][19]}

- タスクが非常に複雑で、Sonnet 4.5 が 5 回以上の試行を必要とする場合
- ワンショット精度が重要で、反復の余地がない場合
- 出力の価値がコスト差を大きく上回る場合（例：\$100 以上の価値を持つタスク）

一方、多くのユーザーは、Sonnet 4.5 を 3～5 回の反復で使用することで、単一の Opus 4.1 クエリと同等かそれ以上の結果を、より低い総コストで得られると報告している。^[21]

ユーザー評価と実践的フィードバック

開発者コミュニティからの初期反応

Reddit、Hacker News、X などのコミュニティでは、Opus 4.5 に対する反応は概ね好意的だが、いくつかの課題も指摘されている。

肯定的な評価：

- 「テキスト表現の観点から、Opus 4.5 は改善されている。応答がより簡潔で、明確かつ直接的」^[22]
- 「曖昧さを処理し、トレードオフについて推論する能力が向上。複雑なマルチシステムバグを自力で修正できる」^{[9][3]}
- 「5 時間使用して、Max プランの使用率はわずか 14%。Opus だけを使っている」^[5]
- 「中程度の設定で Sonnet 4.5 と同じスコアを 76%少ないトークンで達成。コスト削減は約 60%」^[5]

企業パートナーからの証言：

GitHub Copilot は「内部コーディングベンチマークを上回り、トークン使用量を半減。コード移行やリファクタリングに特に適している」と報告している。Cursor は「難しいコーディングタスクにおいて、改善された価格設定と知能で、以前の Claude モデルより顕著に向上」と評価している。^{[6][3]}

Warp は「Terminal Bench で 15%の改善。複雑なワークフローを処理し、行き止まりが少ない」と述べ、Replit、Lovable、v0 などの企業も同様の高評価を示している。^[3]

課題と制限事項

一方で、いくつかの課題も報告されている：

過剰エンジニアリングの傾向：

Anthropic の公式ドキュメントは、「Opus 4.5 は、要求されていない余分なファイル、不必要な抽象化、柔軟性の構築により、過剰エンジニアリングする傾向がある」と認めている。^[23]

使用制限の問題：

初期のリリース直後、一部のユーザーが Opus の使用制限に急速に達するという問題が報告された。特に Claude Code では、システムプロンプトとツールによる最小プロンプトが 15,000 トークンを超えることがあり、キャッシングが 5 分リセットで効果的に機能しない場合、制限が急速に消費されると報告されている。^{[24][25][26]}

Anthropic は 10 月 8 日までにこの問題に対処し、「Opus 4.5 にアクセスできるユーザーについて、Opus 固有の上限を撤廃した。Max および Team Premium ユーザーは、以前 Sonnet で使用していたのと同様のトークン数を Opus でも利用できる」と発表した。^{[26][3]}

速度の問題：

一部のユーザーは、「Opus 4.5 は初日の負荷のためか Sonnet 4.5 より遅く感じる」と報告している。ただし、これは一時的なサーバー負荷によるものである可能性が高い。^[27]

Claude Code 統合とデスクトップアプリ

Opus 4.5 の導入に伴い、Claude Code には 2 つの重要なアップグレードが追加された：^{[1][3]}

1. **Plan Mode**：より正確な計画を構築し、実行前に明確化のための質問を行い、ユーザーが編集可能な plan.md ファイルを作成する

2. **デスクトップアプリ対応**：ローカルおよびリモートで複数のセッションを並行して実行可能。

例えば、あるエージェントがバグを修正している間に、別のエージェントが **GitHub** をリサーチするといった使い方ができる^{[11][3]}

さらに、**Claude for Chrome**（ブラウザのタブ全体でタスクを処理できる拡張機能）が全 **Max** ユーザーに開放され、**Claude for Excel** のベータ版アクセスがすべての **Max**、**Team**、**Enterprise** ユーザーに拡大された。^{[11][3]}

競合モデルとの詳細比較

GPT-5.1 との比較

OpenAI の GPT-5.1 は、Opus 4.5 リリースのわずか 5 日前に発表されており、直接的な競合関係にある。^{[28][10]}

コーディング性能：

SWE-bench Verified において、Opus 4.5（80.9%）は GPT-5.1-Codex-Max（約 77%）を明確に上回っている。ただし、アルゴリズム生成に特化した **LiveCodeBench Pro** では、競争は接戦となる。

^{[28][10][9][1]}

推論能力：

数学的推論では、両モデルとも AIME 2025 でコード実行ありで 100%を達成しているが、ツールなしでは Gemini 3 Pro が優位とされる。^[10]

コストと速度：

GPT-5.1 は\$1.25/\$10 という低価格で、適応的推論によりタスクの複雑さに応じて計算資源を動的に割り当てる効率性を持つ。一方、Opus 4.5 はトークン効率が高く、実質コストでは競争力がある。

^{[18][10][5][3]}

Gemini 3 Pro との比較

Google の Gemini 3 Pro は、Opus 4.5 の直前にリリースされた強力な競合モデルである。^{[29][10]}

コーディング：

SWE-bench Verified では、Opus 4.5（80.9%）が Gemini 3 Pro（76.2%）を 4.7 ポイント上回る。ただし、LiveCodeBench Pro（アルゴリズム生成）では、Gemini 3 がリードしている。^{[10][9][1]}

UI/フロントエンド作業：

複数の開発者レビューは、Gemini 3 Pro が UI 生成、ビジュアルデザイン、DOM との相互作用において優れていると報告している。「Gemini 3 Pro は初回でビジュアルデザインに近づくことが多い」という評価がある。^{[30][29]}

信頼性と指示追従：

開発者は、Gemini 3 Pro が「過信的で信頼性が低い」傾向があり、「修正が機能したと主張するが、実際には機能していない」ケースが多いと指摘している。対照的に、Claude 4.5 は「より信頼性が高く、指示に従う傾向が強い」と評価されている。^[30]

コンテキストウィンドウ：

Gemini 3 Pro は 200 万トークンという圧倒的に大きなコンテキストウィンドウを持ち、Opus 4.5 の 200K を大きく上回る。これは、巨大なコードベースや長期的なコンテキストが必要なタスクにおいて有利である。^{[31][18][10]}

Claude Sonnet 4.5 との関係

興味深いことに、多くの開発者が Opus 4.5 と Sonnet 4.5 の「使い分け」戦略を報告している：^{[32][21]}

Sonnet 4.5 の優位性：

- コード生成、デバッグ、ツール統合においてより高速（30-50%速い）^[21]
- 30 時間以上の自律作業が可能（Opus 4.1 は約 7 時間）^{[33][21]}
- コストパフォーマンスが高い（\$3/\$15）^{[19][21]}

Opus 4.5 の優位性：

- 最高の初回正解率（ワンショット精度）^{[34][4]}
- 深い推論が必要な複雑なタスク^{[32][21]}
- 最終レビュー段階での見落としバグ検出^{[35][32]}

Claude Code 公式チームは、「チーム全員が Sonnet 4.5 を選択している。最強のコーディングモデルだと考えている」と述べている。しかし、複雑なリファクタリング判断や最終的な戦略的レビューには、Opus 4.5 を推奨している。^{[26][32][21]}

産業界での採用と統合

GitHub Copilot での展開

2025 年 11 月 24 日、GitHub は Claude Opus 4.5 を GitHub Copilot のパブリックプレビューとして発表した。Copilot Pro、Pro+、Business、Enterprise ユーザーが利用でき、2025 年 12 月 5 日まで **1 倍のプレミアムリクエスト乗数**というプロモーション価格が適用される。^[8]

GitHub Copilot において、Opus 4.5 は以下で利用可能である：^[8]

- VS Code の Copilot Chat (Agent、Plan、Ask、Edit モード)
- github.com の Copilot Chat
- GitHub Mobile (iOS および Android) の Copilot Chat
- Copilot CLI

プロモーション期間中、Opus 4.5 は **Copilot** コーディングエージェントのデフォルトモデルとしても展開される。^[8]

主要クラウドプラットフォームでの提供

Opus 4.5 は、発表と同時に主要な企業向けクラウドサービスで利用可能となった：^{[7][3]}

- **Amazon Bedrock**：パブリックプレビュー^[36]
- **Google Cloud Vertex AI**：一般提供^{[36][7]}
- **Microsoft Azure / Foundry**：パブリックプレビュー^{[9][7]}
- **Microsoft Copilot Studio**：統合^[7]

Microsoft は特に、Opus 4.5 について「100 万トークンあたり入力\$5、出力\$25」という価格を公表しており、「内部ベンチマークで Sonnet 4.5 と競合を上回り、トークン使用量を半減」と評価している。^{[9][7]}

企業パートナーの評価

Anthropic の公式発表には、多数の企業パートナーからの証言が含まれている：^[3]

Replit：「Opus 4.5 は、最も難しい評価で強力な結果を示し、30 分の自律コーディングセッションを通じて一貫したパフォーマンスを発揮する」

Lovable：「Opus 4.5 は、プロジェクトの計画と反復を行うチャットモードでフロンティア推論を提供。推論の深さが計画を変革し、優れた計画がコード生成をさらに向上させる」

Notion：「Opus 4.5 は、ユーザーが実際に望むものを解釈することに優れ、初回で共有可能なコンテンツを生成する。速度、トークン効率、驚くほど低いコストと組み合わせることで、**Notion Agent** で Opus を提供する初めてのケースとなった」

v0：「Opus 4.5 は、最も難しい 3D 視覚化のいくつかを成功させる唯一のモデル。洗練されたデザイン、上品な UX、優れた計画とオーケストレーション、すべてがより効率的なトークン使用で実現。以前のモデルで 2 時間かかったタスクが、30 分で完了する」

技術的洞察：トークン効率とエージェント能力

ハイブリッド推論アーキテクチャ

Opus 4.5 は、標準的な応答と拡張「思考ブロック」をターン間で保持するハイブリッド推論を採用している。これにより、トークンの無駄が削減され、中程度の effort 設定で SWE-bench タスクにおいて Sonnet 4.5 より 76%少ない出力トークンで同等またはそれ以上のスコアを達成している。^[4]

長期自律作業能力

複数の評価により、Opus 4.5 が長期にわたる自律的な作業において優れた能力を持つことが示されている：^{[6][36][3]}

- 30 分以上の自律コーディングセッションで一貫したパフォーマンスを維持^{[6][3]}
- 長時間タスクにおいて最大 65%少ないトークンで高い合格率を達成^{[6][3]}
- 複雑なワークフローを処理し、行き止まりが少ない^[3]

この能力は、実世界の開発プロジェクトにおいて、「数日かかっていた開発プロジェクトを数時間で完了させる」可能性を示している。^{[7][3]}

マルチエージェントシステムの構築

Opus 4.5 は、サブエージェントチームの管理において非常に効果的であり、複雑で連携が必要なマルチエージェントシステムの構築を容易にする。ある企業は、「オフィス自動化において、エージェン

トが自律的に自身の能力を洗練させ、4 回の反復で最高性能に到達した。他のモデルは 10 回の反復後もその品質に到達できなかった」と報告している。^{[37][3]}

実用的な適用分野と推奨事項

Opus 4.5 が最適なユースケース

複数のレビューと実践的評価から、Opus 4.5 が特に優れている分野が明らかになっている：^{[32][21]}

1. 複雑なソフトウェアリファクタリング：2 つのコードベースと 3 つの調整されたエージェントにまたがる印象的なリファクタリングを実行^[3]
2. コードレビューと品質保証：他のモデルがスキップする「再構築問題、欠落した dispose、非同期バグ」を検出^[32]
3. 長期的な自律プロジェクト：30 分以上の中断なき作業が必要なタスク^{[6][3]}
4. Excel 自動化と財務モデリング：内部評価で精度が 20%向上、効率が 15%上昇^[3]
5. 深層リサーチとドキュメント生成：10-15 ページのチャプターを強固な構成と一貫性で生成^[3]

Sonnet 4.5 との使い分け戦略

実践的な開発者は、以下のようなハイブリッド戦略を推奨している：^{[21][32]}

- **Sonnet 4.5 をデフォルトとして使用**：データ処理、初期分析、迅速なコード生成（作業の 90%、低コスト）
- **Opus 4.5 を戦略的ポイントで使用**：最終的な戦略的洞察、複雑なアーキテクチャ判断、重要なレビュー（作業の 10%、高価値）
- **総コスト**：Sonnet のみより約 20%増加するが、Opus の推論能力を最も重要な場面で活用できる^[21]

注意が必要な分野

一方で、以下の分野では注意が必要である：

- **単純なタスク**：過剰エンジニアリングの傾向があるため、Sonnet 4.5 や Haiku 4.5 の方が適切^{[23][32]}

- 超高速レスポンスが必要な場合：Sonnet 4.5 の方が 30-50%速い^[21]
- 極めて大きなコンテキストが必要な場合：Gemini 3 Pro の 200 万トークンの方が有利^{[31][18]}

今後の展望と業界への影響

AI 開発競争の加速

Claude Opus 4.5 のリリースは、AI 業界における激しい競争を象徴している。わずか 2 週間の間に、OpenAI の GPT-5.1（11 月 19 日）、Google の Gemini 3 Pro（11 月 13 日）、そして Claude Opus 4.5（11 月 25 日）がリリースされた。この前例のない速度での進化は、2025 年後半から 2026 年初頭にかけて、さらに加速する可能性が高い。^{[38][28][10]}

エンジニアリング職業の変革

Anthropic が指摘するように、Opus 4.5 が性能エンジニアリング試験で人間の全候補者を上回ったという事実は、「AI がエンジニアリングという職業をどう変えるのか」という根本的な問いを投げかけている。同社は、「Societal Impacts and Economic Futures 研究」を通じて、多くの分野にわたるこれらの変化を理解することを目指していると述べている。^{[12][3]}

エージェント AI の実用化

Opus 4.5 の長時間自律作業能力（30 分以上）と高度なツール使用能力は、エージェント AI の実用化を大きく前進させる可能性がある。ブラウザ操作、スプレッドシート処理、コード生成を統合した複雑なワークフローの自動化が、より現実的になってきている。^{[39][36][6][11][3]}

安全性研究の重要性の高まり

一方で、プロンプトインジェクション攻撃や実環境での安全性の課題は、より強力な AI モデルが普及するにつれて、安全性研究の重要性が増していることを示している。Anthropic が「1%の攻撃成功率は重大な改善だが、依然として意味のあるリスク」と率直に認めている姿勢は、業界全体に対する重要なメッセージである。^{[16][15][14]}

結論

Claude Opus 4.5 は、性能、効率、価格のバランスにおいて画期的な進歩を実現した AI モデルである。SWE-bench Verified で 80.9%という業界最高スコアを達成しながら、前世代から 67%の価格削減

を実現し、トークン効率を 48-76%向上させたことは、技術的にも商業的にも大きな成功と言える。

[4][1][3]

特に注目すべきは、**Effort** パラメータによる柔軟性、長時間自律作業能力、業界最高のプロンプトインジェクション耐性という 3 つの革新である。これらは、エージェント AI の実用化に向けた重要なマイルストーンとなっている。[14][3]

企業パートナーからの証言、開発者コミュニティでの評価、ベンチマーク結果のすべてが、Opus 4.5 が「現時点で世界最高のコーディングモデル」であるという Anthropic の主張を裏付けている。一方で、実環境での安全性の課題、過剰エンジニアリングの傾向、使用制限の問題など、解決すべき課題も存在する。[15][16][26][23][9][6][1][3]

総合的に見て、Claude Opus 4.5 は、プロフェッショナルな開発環境、長期的なエージェントタスク、複雑な推論が必要な業務において、現時点で最も強力な選択肢の一つである。ただし、タスクの性質、予算、必要な速度に応じて、Sonnet 4.5 や Gemini 3 Pro、GPT-5.1 との使い分けを検討することが、最も賢明なアプローチと言えるだろう。[30][10][21]

★
★★

1. <https://gigazine.net/news/20251125-anthropic-claude-opus-4-5/>
2. <https://forest.watch.impress.co.jp/docs/news/2065838.html>
3. <https://www.anthropic.com/news/claude-opus-4-5>
4. <https://apidog.com/blog/claude-opus-4-5-pricing/>
5. https://www.reddit.com/r/ClaudeAI/comments/1p5psy3/claude_opus_45/
6. <https://www.claude.com/solutions/coding>
7. <https://www.itmedia.co.jp/aipplus/articles/2511/25/news053.html>
8. <https://github.blog/changelog/2025-11-24-claude-opus-4-5-is-in-public-preview-for-github-copilot/>
9. <https://azure.microsoft.com/en-us/blog/introducing-claude-opus-4-5-in-microsoft-foundry/>
10. <https://vertu.com/lifestyle/gemini-3-launch-google-strikes-back-less-than-a-week-after-gpt-5-1-release/>

11. <https://news.yahoo.co.jp/articles/a97d9064e9d8bc234f526baae441cec287c18943>
12. <https://japan.zdnet.com/article/35240832/>
13. <https://thezvi.substack.com/p/claude-sonnet-45-system-card-and>
14. <https://www.anthropic.com/research/prompt-injection-defenses>
15. <https://www.theverge.com/ai-artificial-intelligence/828003/anthropics-new-claude-opus-4-5-model-ai-agents-cybersecurity>
16. <https://www.techbuzz.ai/articles/anthropic-s-claude-opus-4-5-claims-coding-crown-but-fails-key-security-tests>
17. <https://cyberscoop.com/anthropic-sonnet-4-5-security-safety-testing/>
18. <https://intuitionlabs.ai/articles/llm-api-pricing-comparison-2025>
19. <https://blog.getbind.co/2025/09/30/claude-sonnet-4-5-vs-gpt-5-vs-claude-opus-4-1-ultimate-coding-comparison/>
20. <https://www.linkedin.com/pulse/claude-45-pricing-vs-gpt-5-gemini-25-zheng-bruce-li-eeepic>
21. <https://fastgptplus.com/en/posts/opus-41-vs-sonnet-45>
22. https://www.reddit.com/r/ClaudeAI/comments/1p5ylpn/feedback_on_opus_45/
23. <https://platform.claude.com/docs/en/build-with-claude/prompt-engineering/claude-4-best-practices>
24. https://www.reddit.com/r/ClaudeAI/comments/1p1zvwi/we_are_getting_opus_45_octopus_guyss_soon/
25. https://www.reddit.com/r/ClaudeCode/comments/1ny3nn6/yep_they_just_plain_took_opus_away_sonnet_45_is/
26. <https://github.com/anthropics/claude-code/issues/8449>
27. <https://news.ycombinator.com/item?id=46037637>
28. <https://venturebeat.com/ai/anthropics-claude-opus-4-5-is-here-cheaper-ai-infinite-chats-and-coding>
29. <https://www.cometapi.com/gemini-3-pro-vs-claude-4-5-sonnet-for-coding/>
30. <https://www.glbgt.com/hub/gemini-3-pro-vs-claude45/>
31. <https://neontri.com/blog/chatgpt-vs-claude/>
32. <https://www.codegpt.co/blog/anthropic-claude-models-complete-guide>
33. <https://blog.promptlayer.com/claude-sonnet-4-5-first-reactions/>

34. <https://www.cometapi.com/claude-opus-4-5-what-is-it-like-and-how-much/>
35. <https://dev.to/coderabbitai/claude-sonnet-4-5-better-performance-but-a-paradox-29jp>
36. <https://aws.amazon.com/blogs/machine-learning/claude-opus-4-5-now-in-amazon-bedrock/>
37. <https://note.com/npaka/n/nd3bfdfe2c593>
38. <https://news.yahoo.co.jp/articles/876ae1c39d1f4b0ea53050bf561e734a46abe10b>
39. <https://www.tomsguide.com/ai/claude-opus-4-5-is-here-anthropics-most-powerful-version-to-date>
40. <https://www.cnbc.com/2025/11/24/anthropic-unveils-claude-opus-4point5-its-latest-ai-model.html>
41. <https://www.cursor-ide.com/blog/claude-sonnet-4-5-vs-opus-4-1>
42. <https://www.coderabbit.ai/blog/claude-sonnet-4-5-better-performance-but-a-paradox>
43. <https://uraiguide.com/claude-opus-4-5>
44. <https://platform.claude.com/docs/en/about-claude/models/whats-new-claude-4-5>
45. <https://platform.claude.com/docs/en/about-claude/models/overview>
46. <https://www.anthropic.com/news>
47. <https://japan.cnet.com/article/35240833/>
48. https://www.reddit.com/r/ClaudeAI/comments/1nu2jym/analyzed_top_7_posts_about_sonnet_4_5_and_all/
49. https://www.reddit.com/r/ClaudeAI/comments/1nv5gbo/im_sorry_but_4_5_is_insanelly_amazing/
50. <https://www.assemblyai.com/blog/gemini-3-pro-vs-gpt-5-vs-claude-4-5>
51. https://note.com/r1250_gs/n/n03ff84c86f76
52. <https://www.finalroundai.com/blog/claude-sonnet-4-5-what-software-developers-are-saying-after-testing>
53. https://www.reddit.com/r/singularity/comments/1p5po9f/opus_4_5_benchmark_results/
54. https://www.reddit.com/r/claude/comments/1ntxap7/holy_mother_of_god_claude_4_5_is_insane/
55. https://www.reddit.com/r/singularity/comments/1p5rw0d/something_interesting_from_the_claude_4_5_opus/
56. <https://simonwillison.net/2025/Nov/24/claude-opus/>
57. <https://www.reddit.com/r/ClaudeAI/>

58. [https://www.reddit.com/r/ClaudeCode/comments/1o3zzlb/update sonnet 45 completely solved my issues with/](https://www.reddit.com/r/ClaudeCode/comments/1o3zzlb/update_sonnet_45_completely_solved_my_issues_with/)
59. <https://jobirun.com/claude-sonnet-4-5-system-card/>
60. <https://github.com/anthropics/claude-code/issues/9358>
61. <https://assets.anthropic.com/m/64823ba7485345a7/Claude-Opus-4-5-System-Card.pdf>
62. <https://apps.apple.com/jp/app/claude-by-anthropic/id6473753684>
63. <https://platform.claude.com/docs/en/resources/overview>
64. https://x.com/AI_masaou/status/1972973680386019521
65. [https://www.reddit.com/r/ClaudeAI/comments/1nwup8t/claude 45 sonnet lots of hype middling ranks what/](https://www.reddit.com/r/ClaudeAI/comments/1nwup8t/claude_45_sonnet_lots_of_hype_middling_ranks_what/)
66. [https://www.reddit.com/r/ClaudeAI/comments/1n8no2c/claude code feels like a knockoff compared to/](https://www.reddit.com/r/ClaudeAI/comments/1n8no2c/claude_code_feels_like_a_knockoff_compared_to/)
67. <https://www.cometapi.com/github-copilot-cli-vs-claude-code/>
68. [https://www.reddit.com/r/GithubCopilot/comments/1nwsxjs/claude sonnet 45 preview in github copilot is/](https://www.reddit.com/r/GithubCopilot/comments/1nwsxjs/claude_sonnet_45_preview_in_github_copilot_is/)
69. <https://openrouter.ai/anthropic/claude-opus-4.5>
70. <https://www.anthropic.com/news/claude-4>
71. <https://arstechnica.com/ai/2025/11/anthropic-introduces-opus-4-5-cuts-api-pricing-and-enables-much-longer-claude-chats/>
72. <https://github.com/features/copilot?locale=ja>
73. <https://www.clickforest.com/en/blog/ai-tools-comparison>
74. <https://www.unite.ai/claude-sonnet-45-review/>
75. [https://www.reddit.com/r/AISEOInsider/comments/1nuc5dd/claude 45 sonnet my experience testing the new/](https://www.reddit.com/r/AISEOInsider/comments/1nuc5dd/claude_45_sonnet_my_experience_testing_the_new/)
76. [https://www.reddit.com/r/singularity/comments/1p5qmjo/i have enterprise access to claude 45 opus give/](https://www.reddit.com/r/singularity/comments/1p5qmjo/i_have_enterprise_access_to_claude_45_opus_give/)
77. <https://skywork.ai/blog/claude-4-5-vs-claude-4-3-2-2025-comparison/>
78. <https://miralab.co.jp/media/claude-sonnet-4-5/>
79. <https://www.techbuzz.ai/articles/anthropic-drops-claude-opus-4-5-after-350b-valuation-surge>
80. <https://www.datastudios.org/post/claude-4-5-vs-deepseek-s-in-november-2025-full-report-and-comparison-on-features-performance-pri>