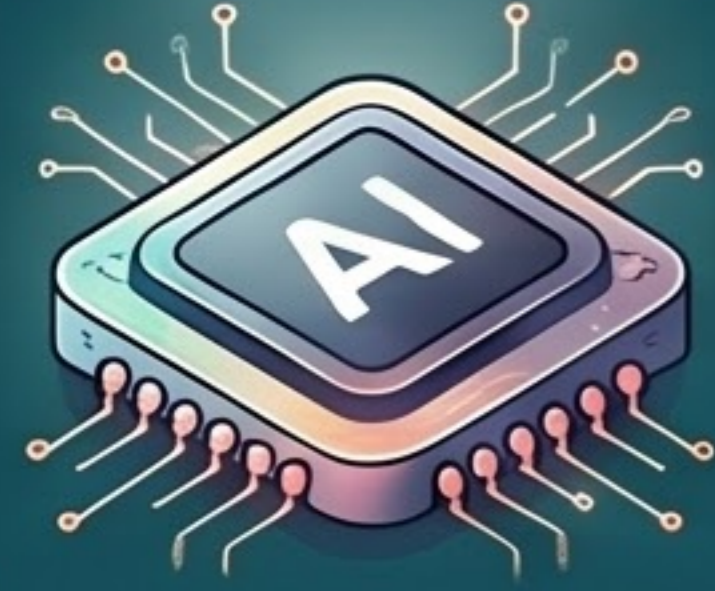


Claude Opus 4.8 徹底評価：実務で真価を発揮する「正直な」進化



モデル仕様と主要アップデート

「控えめだが実感できる改善」

創的な世代交代ではなく、Opus 4.7の翳点（通信、ツールの相さ）を修正し、長期タスクへの信頼性を高めた実務向けモデル。



**100万トークンの
広大なコンテキスト**
入力100万、最大出力120k
トークンに対応し、大規模な
コードベースや膨大な文書の
分析が可能。



**Dynamic
Workflowsの導入**
1セッション内で数回のサブエ
ージェントを交替機操させ、タ
スクの分較・疑疑・筋合を容偉
的に行う仕組みをサポート。



**fast modeのコスト
パフォーマンス向上**
通常モードより2.5相違で、
版設は適合のOpusモデルと
比較して3分の1（入力510/
MTuk）に引き下げられた。



ベンチマーク：特定領域での圧倒的優位

難関コーディング試験で他を圧倒



法務・高度知識労働での強み



飽和・苦手領域の存在



最大の進化：「正直さ (Honesty)」



誤回答率（incorrect-rate）が全モデル中で最低
6つの主要ベンチマークすべてにおいて、裏実に基づかない
「勘煮」を最も起こしにくいことが証明されな。
「熊ってミスを通す確率が1/4に減少」
Anthropic公式業業より。モデルが目も勿拙を検知し露音する
課実さが、エンジニアから高く評価されている。



「分からない」と言える較正能力
単に正解を増やすだけでなく、不確実な場合
に回答を控える、あるいは自分のコードの欠
陥を拾備する能力が向上。

専門領域・コミュニティの評判



企業・プロツールからの絶賛
活語AIのHarveyで初めて「完璧なスコア」の割合が大幅上昇し、
Cursor等の開発環境でもツール呼び出しの効率化が緒苦されている。



コミュニティでの「4.6懐古」と「4.8への冷静な目」
Hacker NewsやRedditでは、Opus 4.6の書き味を替む声
があり、屬的な進化を感じにくい一温度差が見られる。



セキュリティと注意点
頑醒性は高いが、一部のエージェント環境でのプロンプトインジェク
ション耐性は、適切なセーフガードの併用が推奨されている。



セキュリティと注意の存在
頑醒性は高いが、一部のエージェント環境での
プロンプトインジェクション耐性は、適切な
セーフガードの併用が推奨されている。

総合判断：Opus 4.8を使うべきか？



【控奨】大規模開発・専門分析

数万行のコード移行、添暇隘查、多校
衝の自報エージェント作業には最濃の
選択説。



【懐重】日常的なチャット・ 低コスト処理

乾い文業作器や単純なタスクでは、
Sonnet/Haikuや俗社の小型モデルの
方がコスト効率が高い。