


Claude Opus 4.8 : 実務特化型エージェントへの進化

Anthropicの最新モデル「Claude Opus 4.8」の技術的特徴、ベンチマーク性能、実運用における課題。


主要な技術仕様と最適化



1M TOKEN CONTEXT WINDOW
API, Bedrock, Vertex AI support standard TM tokens for large document & code handling.



TOOL OVERHEAD REDUCED UP TO 50%
System prompt token usage significantly decreased, improving agent cost-efficiency & speed.



ADAPTIVE THINKING EVOLVED
Automatically adjusts thinking depth; defaults to "high" effort for complex reasoning.



FAST INFERENCE MODE COST CUT TO 1/3
Fast mode pricing reduced to \$10/\$50 (input/output) from previous \$30/\$150.

ベンチマーク比較

エージェント & 知識タスク性能




人類の最後の試験



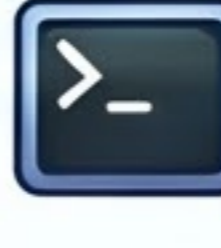
OSWorld-検証済み



Terminal-Bench 2.1



HONESTY REDUCES HALLUCINATIONS:
Selects "does not answer" in uncertain knowledge queries, lowering factually incorrect responses.



TERMINAL AUTOMATION BEHIND GPT-5.5:
Terminal-Bench 2.1 shows 74.6% vs GPT-5.5 78.2%; not dominant across all areas.

進化のタイムライン

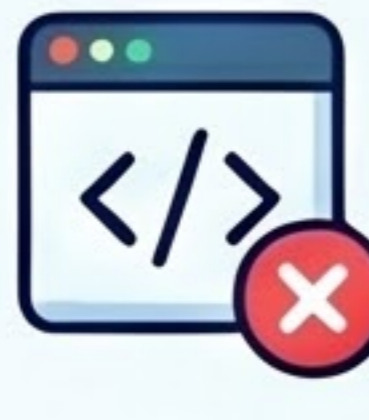
2026/02/05:
Claude Opus 4.6 Released
Base high-function model

2026/04/16:
Claude Opus 4.7 Released
Performance & feature expansion

2026/05/28:
Claude Opus 4.8 Released
Agent optimization & price freezes

2026/05/29:
Real-world Issues Emerge
Post-launch GitHub issues like Claude Code reported


安全性と信頼性の現在地



CODE FLAW OVERSIGHT REDUCED BY 75%
Probability of passing bugs in own code dramatically decreased compared to previous generation.




ALIGNMENT APPROACHES "CLAUDE MYTHOS PREVIEW"
Inappropriate behaviors like deception or misuse cooperation are lower than 4.7, highest safety level yet.




THOUGHT PROCESS HIDDEN (OMITTED THINKING)
Thinking trace is hidden by default (summarize only); explicit request needed.


⚠️ リリース直後の課題 (GitHub Issueより)



TOOL CALL (TOOL_USE) STRUCTURE BREAKAGE
JSON format failure causing conversation stops reported, unstable workflows.



PREMATURE "HALLUCINATED" ANSWERS
Fabricates plausible numbers before actual tool results return, requires strict verification.



DEVELOPMENT ENVIRONMENT BOUNDARY (WORKTREE) EDITING
Risk of Claude Code modifying files outside designated areas, attention needed for operation.