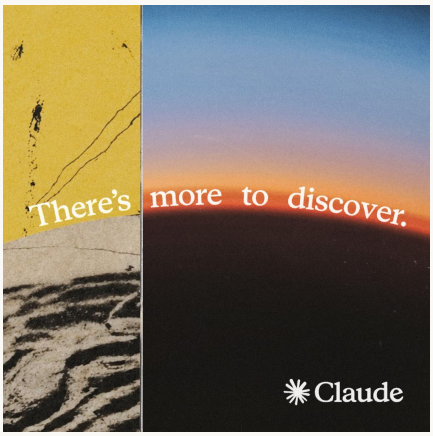
Claude BY ANTHROP\C

Opus 5.5: Raising the bar for enterprise work



Claude Opus 5.5 is Anthropic's most capable Opus model, matching or beating Fable 5.1 on coding tasks at less than half the price. Better at coding, knowledge work, and long-running tasks, Opus 5.5 is a clear step up from Opus 5. It's a better communicator, reporting back on what it did, what it found, and what it needs next.

Frontier performance at a lower cost

Running at $4/$20 per MTok (compared to $5/$25 MTok with Opus 5), Opus 5.5 does more with fewer tokens, and this new pricing passes those gains straight to customers. Lower per-token prices and much cheaper cache reads ($0.20 per MTok, compared to $0.50 per MTok with Opus 5) stack on top of these efficiency gains. For workloads billed by token it **will cost 40% less than Opus 5,** and because Opus 5.5 needs fewer tokens for higher quality work, it runs more than 30% faster.

Agents

Orchestrates complex multi-tool tasks reliably, plans deliberately, and drives long-running work forward with minimal oversight.

Coding

Handles long-running work in large codebases. It finds the root cause before changing anything, checks its work as it goes, and explains its changes in plain language.

Enterprise workflows

Built to be the enterprise workhorse, powering agents that run projects end-to-end, following instructions precisely, producing professional-grade spreadsheets, slides, and docs.

Computer use

Brings high-resolution vision and deep reasoning to computer use, handling multi-step tasks that span multiple applications and require planning and judgment.

Financial analysis

Reasons about filings, models, and charts accurately, and carries context across an entire deal or reporting cycle, with clear summaries of what it found and how it got there.

Vision

Reads dense documents, charts, screenshots, and diagrams at high fidelity, making it reliable for document extraction, visual analysis, and interpreting complex real-world imagery.

When to use what model

Model	Use cases
Claude Fable 5.1	Frontier problems: long-running agents, net-new use cases, hardest coding & reasoning, full autonomy
Claude Opus 5.5	Your daily driver for complex work. Reasoning-intensive enterprise work (legal, financial analysis, complex knowledge work), complex and production coding
Claude Sonnet 5	General purpose workloads at scale. For capable, efficient intelligence across everyday work
Claude Haiku 4.5	High-frequency, latency-sensitive tasks; sub-agents in orchestration

Early users shared that Claude Opus 5.5 matches or beats Opus 5 at a fraction of the tokens, writes clearer and more concise output that needs little editing, and handles long, unattended tasks on its own, checking its work along the way.

New capabilities launching with Opus 5.5

Alongside Opus 5.5, we're releasing new features and improvements that make Claude more proficient with longer-running tasks:

New compaction options (beta)

The API can now summarize older turns while keeping recent ones intact, and do it in the background without pausing the agent.

Communication improvements

Opus 5.5 has significant improvements in how it communicates progress, updates, and outcomes with users compared to recent Opus models.

Claude Opus 5.5

	Opus 5.5	Fable 5.1	Opus 5	GPT-6 Astra	GPT-5.6 Sol
Agentic coding Terminal-Bench 4.0[1]	**66.4%**	55.8%	52.3%	57.9%	37.3%
Agentic coding FrontierCode v1.1 (Main)	**54.4%**	50.3%	48.0%	53.3%	47.5%
Agentic coding CursorBench 4.0	**57.8%**	51.8%	46.6%	—	41.7%
Knowledge work GDPval-AA v2.1	**1846**	1735	1708	1542	1588
Business workflows AutomationBench[2]	40.0%	31.4%	26.9%	**41.4%**	28.8%
Multidisciplinary reasoning Humanity's Last Exam	**67.7%** with tools	65.6% with tools	63.6% with tools	57.2% with tools	—
Agentic scientific research Terminal-Bench-Science 0.1[3]	58.7%	52.6%	29.0%	**64.6%**	22.4%
Computer use OSWorld 2.0	**81.8%** partial	80.7% partial	74.0% partial	—	—
Visual chart recognition Chartography	**89.0%** with tools	88.4% with tools	83.4% with tools	—	—

Unless otherwise noted, all Claude Opus 5.5 results use adaptive thinking at max effort. Terminal-Bench 4.0: Claude Opus 5.5 at xhigh effort and GPT-6 Astra as reported by OpenAI at high effort. Claude Opus 5.5 was evaluated with its production safeguards enabled. When they intervened, cybersecurity tasks were completed by Claude Opus 4.8, and biology and frontier LLM development tasks by Claude Opus 5. This likely reduces Claude Opus 5.5's performance on these benchmarks.

[1] **Terminal-Bench 4.0:** The standard error is ±2.6 pts for Claude Opus 5.5 and ±1.6–2 pts for the other Claude models. The public leaderboard (5 trials/task, Claude Code harness) reports Claude Opus 5 at 51.8%; our setup reproduces it at 52.3%, within noise. GPT-6 Astra and GPT-5.6 Sol figures are as reported by OpenAI.

[2] **AutomationBench:** AutomationBench results were run and reported by Zapier. Claude Opus 5.5 results come from Zapier's own evaluation during early access. Results for Opus 5, GPT-5.6 Sol, and GPT-6 Astra come from Zapier's public leaderboard.

[3] **Terminal-Bench-Science 0.1:** The standard error is ±3.5–5 pts per model. The public leaderboard (3 trials/task, Claude Code harness) reports Claude Opus 5 at 30.0%; our setup reproduces it at 29.0%, within noise. The GPT-6 Astra figure is as reported by OpenAI.