Claude BY ANTHROP\C

Frontier intelligence built for enterprise AI



Claude Fable 5.1 is Anthropic's most capable model for putting frontier intelligence into production across your code, documents, and research. Fable 5.1 was built for long-running, high-stakes work run by enterprises, developers, and power users.

A step change for production workloads

Claude Fable 5.1 is a clear improvement over Claude Fable 5, setting new benchmarks in coding, knowledge work, and the sciences. Running at $10/$50 mTok, the model offers significant gains on long-running, multi-step workflows. Fable 5.1 charges 75% less than Fable 5 for cache reads—the inputs to the model that have already been processed and stored. In practice, Anthropic estimates this will reduce the cost of Fable 5.1 for typical workloads by around 25%. For highly agentic workloads, the cost savings will often be much higher—as high as around 45%.

Agents

Built for jobs that take hours and span many applications, including working through a backlog of projects or running unattended as a managed agent.

Coding

Carries a software project from end to end, including features that span an entire codebase, code review, and performance work. It is more honest and less likely to take shortcuts than its predecessor.

Enterprise workflows

Takes a piece of analysis from the first question to the finished document. Claude Fable 5.1 does the research, builds the spreadsheet, writes the memo, and checks its numbers as it goes.

Improved judgement

Reasons about long, multi-step problems in competition mathematics and graduate-level engineering and sciences. Exhibits better judgement on ambiguous tasks than previous models.

Financial analysis

Reads dense filings and charts with sharper judgment and fewer confident errors, and holds the thread across an entire deal or reporting cycle.

Science

Drives research from idea through validation across many scientific fields. Surveys literature, forms hypotheses, builds computational models, designs experiments, and interprets what lab results.

When to use what model

Model	Use cases
Claude Fable 5.1	Your hardest problems: long-running agents in production, code migration, multi-step reasoning, and tasks that require creative thinking and full autonomy.
Claude Opus 5	Reasoning-intensive enterprise work (legal, financial analysis, research, and other complex domains) and production coding.
Claude Sonnet 5	General-purpose workloads at scale across coding and knowledge work.
Claude Haiku 4.5	High-frequency, latency-sensitive tasks; sub-agents in orchestration

Learn more about **model** and **effort level** selection best practices on **Claude Academy**.

Early users describe Claude Fable 5.1 as the first model to pull ahead on their hardest problems: it found a crash one team couldn't explain for years, ran unattended for hours without losing the plot, and beat Claude Fable 5 on benchmarks while using fewer tokens.

Announcing Enterprise Frontier Safeguards

Anthropic worked with more than 100 companies across highly regulated industries to develop Enterprise Frontier Safeguards (EFS), a solution that combines the privacy of zero data retention (ZDR) with state-of-the-art safeguards for detecting misuse. EFS works by storing data in cloud infrastructure controlled by the customer, not Anthropic. Learn more.

Claude Fable 5.1

	Fable 5.1	Fable 5	Opus 5	GPT-5.6 Sol
Agentic scientific research Terminal-Bench-Science 0.1	**52.6%**	24.7%[1]	29.0%[1]	22.4%
Agentic coding Terminal-Bench 4.0	**55.8%** **60.9%** Mythos 5.1	42.0%	52.3%	37.3%
Knowledge work GDPval-AA v2	**1853**	1723	1824	1711
Computer use OSWorld 2.0	**77.9%** partial **41.7%** strict	72.9% partial 36.1% strict	75.4% partial 39.6% strict	—
Multidisciplinary reasoning Humanity's Last Exam	**60.9%** no tools **65.0%** with tools	57.8% no tools 63.8% with tools	56.6% no tools 63.6% with tools	—
Business workflows AutomationBench	**31.4%**[2]	17.1%	26.9%	19.6%
Agentic coding CursorBench 3.2.0	**73.4%**	70.5%	70.0%	67.2%

[1] Terminal-Bench Science: The standard error is ±3.5–4.5 pts per model. The public leaderboard (3 trials/task, Claude Code harness) reports Claude Opus 5 at 30.0% and Claude Fable 5 at 21.4%; our setup reproduces them at 29.0% and 24.7% respectively, both within noise.

[2] AutomationBench: Fable 5.1 was evaluated with automated classifier fallback to Claude Opus 5 enabled. Other scores reflect standalone model runs.