

DEIN DIREKTER EINSTIEG

Opus 5.5: Fakten, Vergleiche & Prompts

Echte Grafiken von Anthropic und Artificial Analysis, verständliche Kostenvergleiche und zwei kurze Praxis-Prompts für Angebote und Website-Fehler.



- 01 Lies die wichtigsten Neuerungen und unterscheide Modellkosten von deinem Abopreis.
- 02 Vergleiche Originalgrafiken und beachte Testaufbau und Denkstufe.
- 03 Kopiere einen Praxis-Prompt und teste ihn mit eigenen Unterlagen oder deinem Website-Projekt.

[Anleitung öffnen ↗](#)[Mehr Anleitungen ↗](#)

Jeden Sonntag: KI-News, ein Trick zum Ausprobieren und ein Dokument zum Nachlesen.

[Zum Newsletter ↗](#)[Zur Community-Warteliste ↗](#)

Download und Anleitung bleiben frei zugänglich. Die folgenden Seiten enthalten die vollständige Anleitung und die Original-Links. Gestalterisch überarbeitet am 20.10.2026; Inhaltsbasis 22.09.2026; ausgewählte Aktualisierungen im Text.



Opus 5.5. Was du davon hast.

Mehr Leistung für anspruchsvolle Aufgaben. Günstigere API-Nutzung. Mehr Spielraum im Claude-Abo.

-40 %

typische Aufgabenkosten

>30 %

schnellere Textausgabe

Herstellerangaben gegenüber Opus 5. Die Kostenersparnis gilt für typische Aufgaben mit Standardeinstellungen, nicht für den Abopreis.

Pricing

Prices per 1M tokens	Claude Opus 5.5	Claude Opus 5
Input tokens	\$4	\$5
Output tokens	\$20	\$25
Cache reads	\$0.20	\$0.50
Cache writes	\$5	\$6.25

Originalgrafik aus dem Claude-Thread · US-Dollar je 1 Mio. Tokens. Input = Eingabe, Output = Ausgabe, Cache = wiederverwendeter Kontext.

Für dein Abo: Höhere Fünf-Stunden-Limits für Pro, Max, Team und sitzbasierte Enterprise-Pläne. Abonnenten erhalten außerdem einen speicherbaren Limit-Reset.

Sonnet 5.5 ist seit 28.09.2026 verfügbar. Prüfe das Modellangebot und die aktuellen Nutzungslimits in deinem Konto. [Aktuelle Sonnet-Ankündigung](#)

[Ankündigung & Zugang](#)



Wo Opus 5.5 punktet.

Die Originaltabelle zeigt starke Resultate beim Coden und bei Wissensarbeit. Sie zeigt auch: Opus gewinnt nicht jeden Test.

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 results are reported for Claude Opus 5.5 at xhigh effort and GPT-6 Astra at high effort, as reported by OpenAI; these represent each model's highest score. 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 were completed 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. These runs were performed without fallback models, so safeguard interventions were considered failures—this resulted in a lower score than Claude Opus 5.5 would achieve in practice. 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.

So liest du den Vergleich: Bei AutomationBench liegt GPT-6 Astra mit 41,4 % vor Opus mit 40,0 %. Unterschiedliche Testaufbauten und Denkstufen sind in den Originalfußnoten ausgewiesen.

[Originalpost mit Grafik und Fußnoten](#)



Mit deiner Arbeit testen.

Sinnvolle Einsätze: Angebote vergleichen, Berichte aus Unterlagen erstellen oder Fehler in einer vorhandenen Website beheben.

01 / Angebote klar vergleichen

In Claude zwei eigene Angebotsdateien anhängen. Opus 5.5 wählen, falls dein Zugang es anbietet.

Du bist mein Analyse-Assistent. Ich lade zwei Angebote für dieselbe Leistung hoch und muss eines auswählen. Vergleiche nur die Dokumente: Umfang, Gesamtpreis, laufende Kosten, Fristen und Lücken. Liefere eine Vergleichstabelle, drei Rückfragen an die Anbieter und eine begründete Empfehlung mit Fundstellen. Erfinde keine Angaben; markiere Fehlendes. Rechne alle Summen nach und gleiche jede Aussage mit den Quellen ab. Korrigiere Fehler und prüfe erneut. Prüfe aus Sicht eines unbeteiligten Lesers, ob die Entscheidung nachvollziehbar ist. Nenne durchgeführte Prüfungen und offene Punkte getrennt.

02 / Website-Fehler beheben

In Claude Code im bestehenden Projekt arbeiten. Browser- und Testwerkzeuge müssen verfügbar sein.

Du bist mein Webentwickler. Im geöffneten Projekt liegt meine bestehende Website. Prüfe das Kontaktformular und repariere belegte Funktionsfehler. Bewahre Gestaltung und übrige Funktionen; veröffentliche nichts. Teste gültige und ungültige Eingaben sowie die Bedienung auf Mobilgeräten und mit Tastatur. Nutze für Versandtests ausschließlich eine Testumgebung; sende keine echten Anfragen. Korrigiere Fehler und wiederhole die Tests. Prüfe anschließend aus Sicht eines neuen Besuchers, ob Fehlermeldungen verständlich sind. Liefere Änderungen, tatsächlich ausgeführte Tests und verbleibende Einschränkungen. Eine simulierte Besuchersicht ist kein echter Nutzertest.

Eigene Prompt-Vorlagen. Redaktionell geprüft; kein vollständiger Opus-5.5-Produkttest durchgeführt.

[Prompts bequem kopieren](#) ↗