MODELLE

Claude Opus 5 - technische Daten, Benchmarks und Preise

Claude Opus 5 im Datenblatt: Anthropics Modell für agentisches Programmieren - technische Daten, Preise und Benchmarks mit Quelle und Stichtag.

Stand: 04.08.2026 · 4 Min. Lesezeit

Claude Opus 5 ist Anthropics Modell für agentische Programmierarbeit und die Bedienung von Rechnern über Bildschirmfotos. Es ist Ende Juli 2026 erschienen und hat Opus 4.8 abgelöst - zum Preis, den auch die Opus-Stufe zuvor hatte. Die Neuerungen stecken weniger in einer einzelnen Kennzahl als im Betrieb: Reasoning ist jetzt die Voreinstellung, die Schwelle für Prompt Caching ist halbiert worden, und das Modell bekommt ein eigenes Kontingent für Anfragegrenzen. Genau solche Details entscheiden in automatisierten Abläufen über Kosten und Verhalten - und damit darüber, ob sich der Umstieg lohnt.

Claude Opus 5 - Überblick und technische Daten

Anbieter	Anthropic
Familie	Claude 5
Erschienen	24. Juli 2026 Die Quellen weichen um einen Tag ab. Eine nennt den 24. Juli 2026, eine andere den 23. Juli 2026 als Tag der Bereitstellung über die Plattformanbieter.
Lizenz	Proprietär, ausschließlich über Schnittstellen nutzbar.
Weights	closed weights, kein Selbstbetrieb
Nutzbar über	Anthropic, Anthropic API, Amazon Bedrock, Google Cloud, Microsoft Foundry
Preis ab	5,00 USD je Mio. Input-Token
Architektur	nicht veröffentlicht Vom Anbieter nicht veröffentlicht.
Parameter gesamt	nicht veröffentlicht Vom Anbieter nicht veröffentlicht.

AI NEWS pro - KI-News ohne Hype! · © 2026 SOFTCLICK Labs · Stand des Beitrags: 04.08.2026

Kontextfenster	1.000.000 (rund 1 Mio.) Token Eine Million Token sind zugleich Standard und Maximum.
Maximaler Output	128.000 Token
Input	Text, Bild, PDF
Output	Text
Wissensstand	Mai 2026
Reasoning	Wahlweise Reasoning ist standardmäßig aktiv - anders als beim Vorgänger Opus 4.8, wo ein Weglassen des Parameters es abschaltete. Abschalten ist nur bis zur Aufwandsstufe "high" möglich. In Verbindung mit "xhigh" oder "max" wird die Anfrage abgewiesen.
Tool Calls	Ja
Structured Output	Ja
Prompt Caching	Ja, ab 512 Token Die Mindestlänge für Prompt Caching liegt bei 512 Token und damit halb so hoch wie bei Opus 4.8 (1.024).
Kennung	claude-opus-5 Feste Kennung ohne Datumssuffix. Konstruierte Varianten wie "claude-opus-5-20260724" existieren nicht und laufen auf einen Fehler.

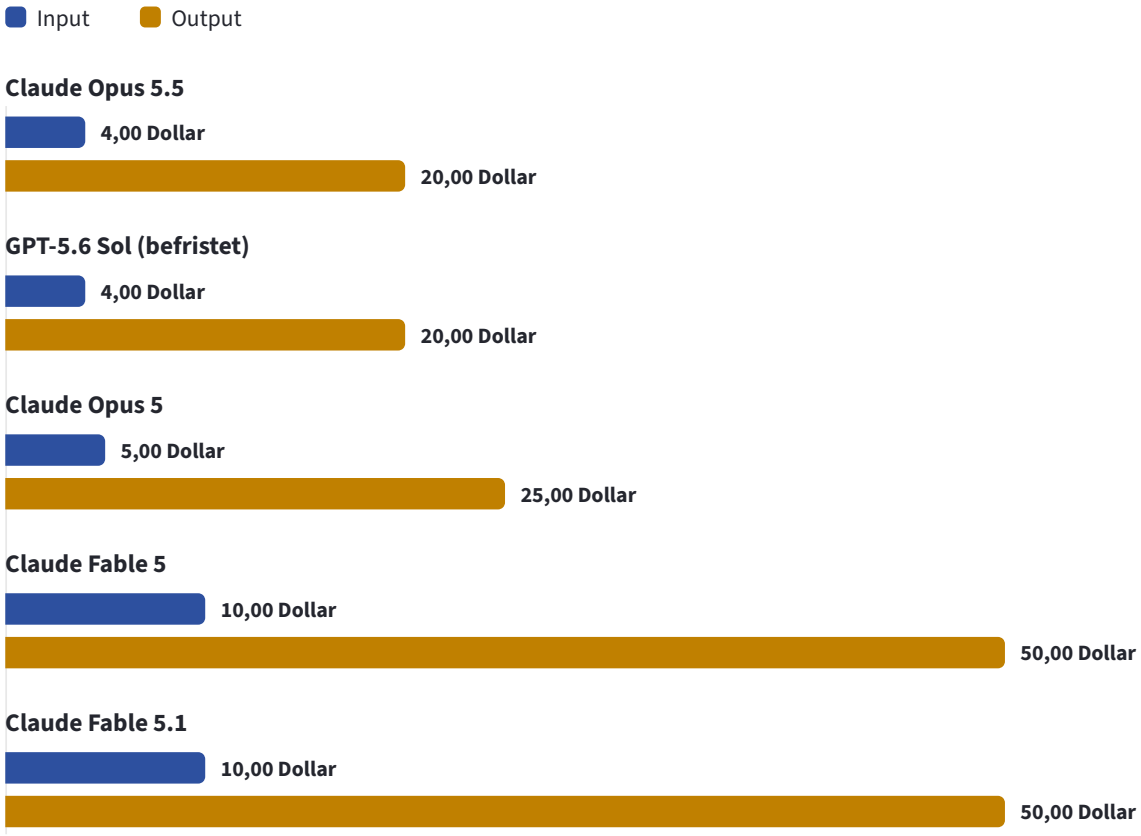
- Aufwandsstufen (low, medium, high, xhigh, max) steuern Denktiefe und Token-Verbrauch. Voreinstellung ist "high".
- Hochauflösende Bildverarbeitung bis 2.576 Pixel Kantenlänge. Koordinaten im Modellausgang entsprechen den Bildpixeln ohne Umrechnung.
- Eigenes Kontingent für Anfragegrenzen, getrennt vom gemeinsamen Topf der Opus-4er-Reihe.
- Sicherheitsklassifikatoren können eine Anfrage ablehnen. Die Antwort ist dann technisch erfolgreich, enthält aber keinen Inhalt.

AI NEWS pro - KI-News ohne Hype! · © 2026 SOFTCLICK Labs · Stand des Beitrags: 04.08.2026

Preise

API-Preise: Claude Opus 5 im Vergleich

Dollar je eine Million Token. Je kürzer der Balken, desto günstiger.



Quelle: Anbieter-Preislisten, Stand: 6. Oktober 2026 · Grafik: AI NEWS pro

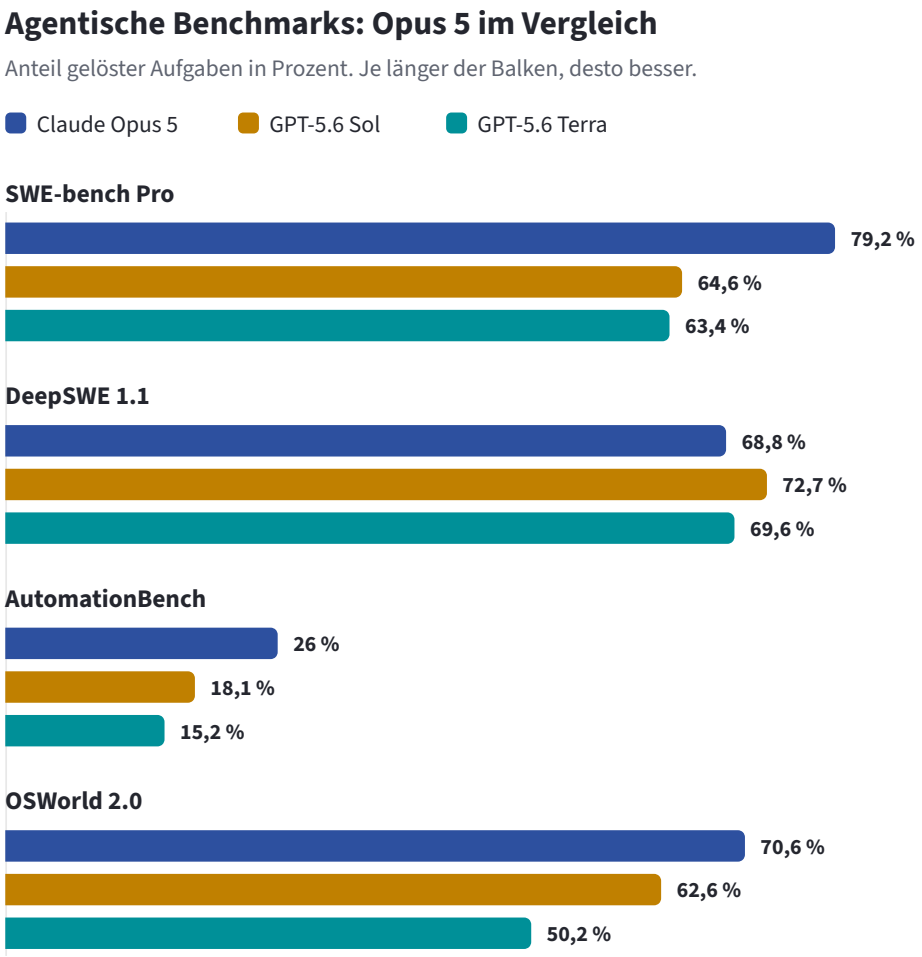
Stufe	Input	Output	Cache (Lesen)	Stand
standard (Anthropic)	5,00	25,00	0,50	6. Oktober 2026
fast (Anthropic)	10,00	50,00	-	4. August 2026

Beträge in US-Dollar je einer Million Token, nach Anbieterangabe.

fast: Schnellmodus mit bis zu 2,5-facher Ausgabegeschwindigkeit. Nur über die Anthropic-Schnittstelle, nicht bei Amazon, Google oder Microsoft.

AI NEWS pro - KI-News ohne Hype! · © 2026 SOFTCLICK Labs · Stand des Beitrags: 04.08.2026

Benchmarks



Quelle: Anthropic (Herstellerangabe), Stand: 24. Juli 2026 · Grafik: AI NEWS pro

Suite	Wert	Gemessen von	Bedingungen	Stand
SWE-bench Verified (Fassung verified)	96 %	Anbieter (Anthropic)	Bedingungen nicht angegeben	24. Juli 2026
GDPval-AA v2 (Fassung v2)	1.861	unabhängig (Artificial Analysis)	Elo-Wert, nicht Prozent	24. Juli 2026

Stand der Daten: 6. Oktober 2026

AI NEWS pro - KI-News ohne Hype! · © 2026 SOFTCLICK Labs · Stand des Beitrags: 04.08.2026

Claude Opus 5

Opus 5 ist das jüngste der drei allgemein verfügbaren Modelle der Claude-5-Familie, mit der Anthropic seine Claude-Reihe seit Juni 2026 erneuert. Positioniert ist es für agentisches Programmieren und für Aufgaben, bei denen das Modell einen Rechner über Bildschirmfotos bedient. Das aktuelle Kontextfenster teilt es mit seinen Geschwistern, und es steht jedem Aufruf in voller Größe offen - eine Staffelung nach Zugangs- oder Ausbaustufen gibt es nicht.

Der deutlichste Schnitt gegenüber Opus 4.8 betrifft das Reasoning. Beim Vorgänger war es aus, wenn der Aufruf den entsprechenden Parameter nicht enthielt. Bei Opus 5 ist es standardmäßig aktiv, und ganz abschalten lässt es sich nur noch bis zur Aufwandsstufe "high" - in Verbindung mit "xhigh" oder "max" weist die Schnittstelle die Anfrage ab. Gesteuert wird die Denktiefe über fünf Aufwandsstufen von "low" bis "max", die zugleich den Token-Verbrauch bestimmen. Voreingestellt ist "high".

> **Beim Umstieg von Opus 4.8**
>
> Wer seine Aufrufe unverändert übernimmt und den Parameter fürs Reasoning wie bisher weglässt, bekommt bei Opus 5 kein Modell ohne Denkschritte mehr, sondern die Voreinstellung "high" - mit entsprechend höherem Token-Verbrauch je Antwort.

Auch drumherum hat Anthropic Betriebsdetails geändert. Die Modellkennung der Schnittstelle kommt ohne Datumszusatz aus - wer nach altem Muster eine datierte Kennung konstruiert, produziert einen Fehler. Anfragegrenzen zählt der Anbieter für Opus 5 in einem eigenen Kontingent, getrennt vom gemeinsamen Topf der Opus-4er-Reihe, sodass der Umstieg nicht mit bestehendem Verkehr konkurriert. Und die Mindestlänge, ab der Prompt Caching greift, ist gegenüber dem Vorgänger halbiert worden - damit profitieren auch kürzere Systemprompts vom günstigeren Preis für Cache-Treffer.

Einordnung

Die Anlagen zielen auf Werkzeugketten. Opus 5 nutzt Tool Calls, liefert Structured Output und nimmt neben Text auch Bilder und PDF-Dokumente entgegen. Für die Bedienung von Rechnern zählt ein Detail der hochauflösenden Bildverarbeitung: Die ausgegebenen Koordinaten entsprechen den Bildpixeln, ein Klickziel aus einem Bildschirmfoto lässt sich also ohne Umrechnung ansteuern.

Die Benchmark-Lage verlangt einen zweiten Blick. Der Wert, den Anthropic selbst für die Programmier-Suite SWE-bench Verified meldet, kommt ohne Angabe der Messbedingungen - und er liegt nahe an der Sättigungsgrenze der Suite, in der Spitzenmodelle 2026 zwischen 88 und 96 Prozent erreichen. Ein Vorsprung von wenigen Punkten ist in dieser Zone kaum noch aussagekräftig.

Den einzigen unabhängigen Messwert steuert Artificial Analysis bei, eine Elo-Wertung für agentische Aufgaben aus realer Wissensarbeit. Zwei weitere kursierende Werte führt das Datenblatt bewusst nicht: Bei den betreffenden Suiten sind Methodik oder Herausgeber bislang nicht belegt, und ein Wert ohne belegte

AI NEWS pro - KI-News ohne Hype! · © 2026 SOFTCLICK Labs · Stand des Beitrags: 04.08.2026

Messmethode gehört in keine Tabelle.

Die Grenzen sind die eines geschlossenen Modells. Es antwortet ausschließlich mit Text, die Gewichte sind nicht offen, Selbstbetrieb scheidet aus. Der Zugang läuft über Anthropics eigene Schnittstelle sowie über die Cloud-Plattformen von Amazon, Google und Microsoft. Wer dort einkauft, verzichtet allerdings auf den Schnellmodus - die Stufe mit bis zu zweieinhalbfacher Ausgabegeschwindigkeit zum doppelten Preis gibt es nur direkt bei Anthropic. In automatisierten Strecken gehört außerdem ein Sonderfall abgefangen: Sicherheitsklassifikatoren können eine Anfrage ablehnen, die Antwort ist dann technisch erfolgreich, bleibt aber leer.

Preislich sitzt Opus 5 in der Mitte der eigenen Familie - die Hälfte von Fable 5, bei Eingabe wie Ausgabe. Gegenüber dem Standardmodell Sonnet 5 kostet Opus 5 das Zweieinhalbfache, bei Eingabe wie Ausgabe - die für den 1. September 2026 angekündigte Preiserhöhung bei Sonnet 5 hat Anthropic nicht vollzogen, der Abstand bleibt also so groß (Stand 10. September 2026). Auffällig kurz ist der Abstand zum Geschehen: Zwischen dem dokumentierten Wissensstand des Trainings und dem Erscheinen liegen nur rund zwei Monate.

Verwandte Modelle

Direkter Vorgänger ist Claude Opus 4.8. Was sich beim Umstieg ändert, steht oben - Reasoning als Voreinstellung, halbierte Schwelle für Prompt Caching, ein eigenes Kontingent für Anfragegrenzen.

In der eigenen Generation stehen zwei Geschwister daneben. Claude Sonnet 5 ist das Standardmodell der Anthropic-Produkte und die günstigere Wahl - sein neuer Tokenizer zerlegt denselben Text allerdings in rund 30 Prozent mehr Token als noch Sonnet 4.6, was Preisvergleiche je Anfrage verschiebt, und sein Reasoning lässt sich uneingeschränkt abschalten. Claude Fable 5 kostet das Doppelte von Opus 5, sein Reasoning lässt sich nie abschalten, und es ist ausschließlich direkt bei Anthropic verfügbar - vorausgesetzt, die Organisation akzeptiert eine Datenaufbewahrung von 30 Tagen.

Die Claude-5-Familie

9. Juni 2026	Claude Fable 5 eröffnet die Familie.
30. Juni 2026	Claude Sonnet 5 ersetzt Sonnet 4.6 als Standardmodell der Anthropic-Produkte.
24. Juli 2026	Claude Opus 5 löst Opus 4.8 ab.

Quellen

1. Anthropic - Modellübersicht und Preise · https://platform.claude.com/docs/en/about-claude/models/overview
2. MarkTechPost: Claude Opus 5 · https://www.marktechpost.com/2026/07/24/meet-the-new-claude-opus-5-frontier-class-agentic-coding-and-computer-use-at-unchanged-opus-pricing/
3. Anthropic: Preisangabe Claude Opus 5 (standard) · Stand 06.10.2026 · https://platform.claude.com/docs/en/about-claude/pricing
4. Anthropic: Preisangabe Claude Opus 5 (fast) · Stand 04.08.2026 · https://platform.claude.com/docs/en/pricing

AI NEWS pro - KI-News ohne Hype! · © 2026 SOFTCLICK Labs · Stand des Beitrags: 04.08.2026

5. Artificial Analysis: GDPval-AA v2 v2 für Claude Opus 5 · Stand 24.07.2026 · https://artificialanalysis.ai/models
6. Anthropic: Agentische Benchmarks: Opus 5 im Vergleich · Stand 24.07.2026 · https://www.datacamp.com/blog/claude-opus-5

Themen: Modelle · Anthropic

AI NEWS pro - KI-News ohne Hype! · © 2026 SOFTCLICK Labs · Stand des Beitrags: 04.08.2026