MODELLE

Claude Sonnet 5 - technische Daten, Benchmarks und Preise

Claude Sonnet 5 im Datenblatt: Anthropics Standardmodell mit technischen Daten und Preisen samt Stichtag - 2 und 10 US-Dollar je Million Token, neuer Tokenizer mit Folgen für die Kosten je Anfrage.

Stand: 04.08.2026 · 3 Min. Lesezeit

Claude Sonnet 5 ist Anthropics Standardmodell: Seit dem 30. Juni 2026 steckt es als Voreinstellung in den Endkunden- und Entwicklerprodukten des Anbieters, und wer dort mit Claude arbeitet, arbeitet in aller Regel mit diesem Modell. Den Platz hatte zuvor Sonnet 4.6 inne. In der Claude-5-Familie ist Sonnet 5 die günstigste der drei Stufen, unterhalb des auf agentische Programmierarbeit zugeschnittenen Opus 5 und der Spitzenstufe Fable 5. Wer die Kosten kalkuliert, schaut dabei besser zweimal hin: Der Listenpreis je Token entspricht dem des Vorgängers, doch ein neuer Tokenizer verschiebt die Kosten je Anfrage nach oben.

Claude Sonnet 5 - Überblick und technische Daten

Anbieter	Anthropic
Familie	Claude 5
Erschienen	30. Juni 2026 Ersetzte am selben Tag Sonnet 4.6 als Standardmodell in den Endkunden- und Entwicklerprodukten des Anbieters.
Lizenz	Proprietär, ausschließlich über Schnittstellen nutzbar.
Weights	closed weights, kein Selbstbetrieb
Nutzbar über	Anthropic, Anthropic API, Amazon Bedrock, Google Cloud Vertex AI, Microsoft Foundry
Preis ab	2,00 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
Maximaler Output	128.000 Token
Input	Text, Bild, PDF
Output	Text
Reasoning	Wahlweise Reasoning läuft standardmäßig im adaptiven Modus, auch wenn der Parameter weggelassen wird - beim Vorgänger Sonnet 4.6 war es dann aus. Abschalten ist uneingeschränkt möglich.
Tool Calls	Ja
Structured Output	Ja
Prompt Caching	Ja, ab 1.024 Token
Kennung	claude-sonnet-5

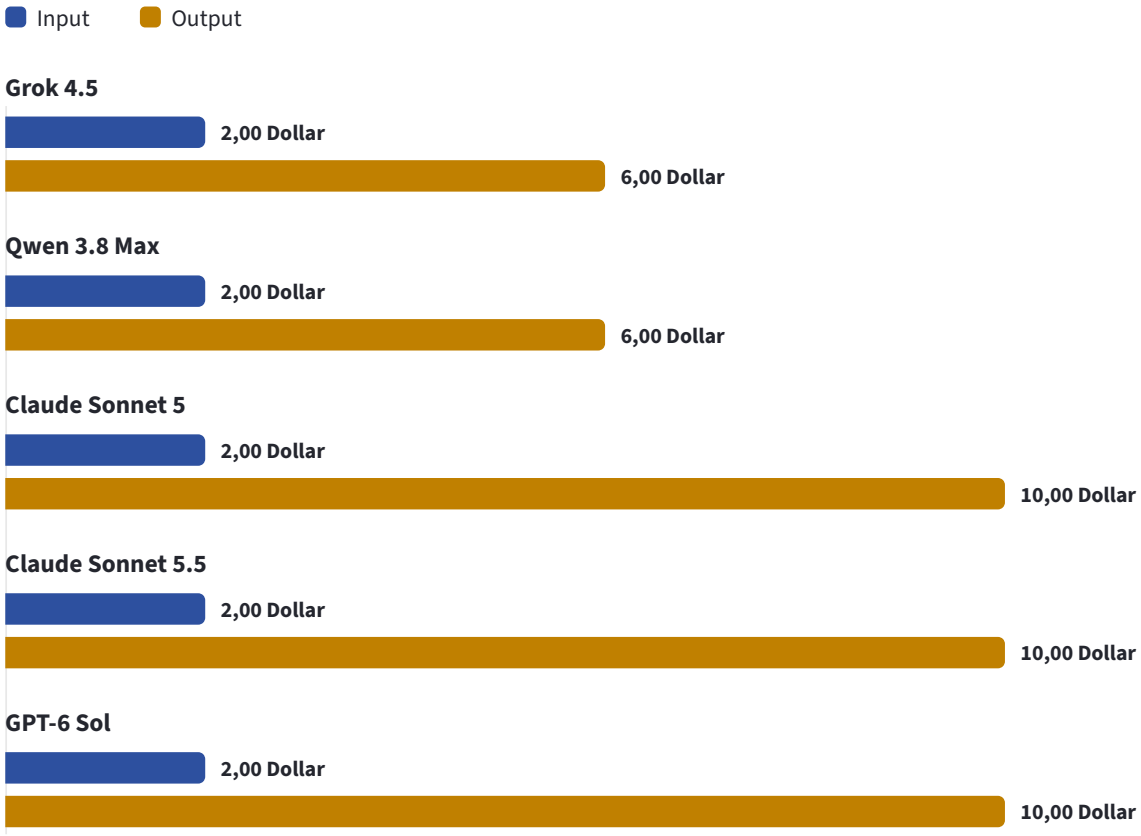
- Neuer Tokenizer: Derselbe Text ergibt rund 30 Prozent mehr Token als bei Sonnet 4.6. Der Preis je Token blieb gleich, die Kosten einer gleichwertigen Anfrage steigen dadurch trotzdem.
- Erstes Modell der Sonnet-Reihe mit hochauflösender Bildverarbeitung bis 2.576 Pixel Kantenlänge (zuvor 1.568).
- Erstes Modell der Sonnet-Reihe mit der Aufwandsstufe "xhigh".

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

Preise

API-Preise: Claude Sonnet 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

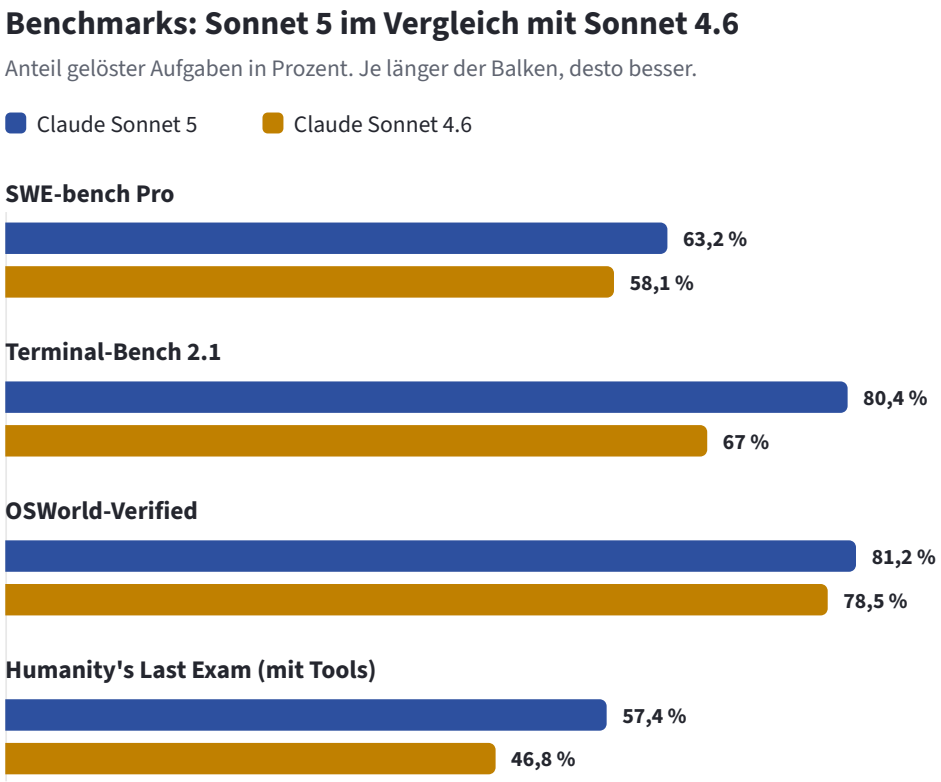
Stufe	Input	Output	Cache (Lesen)	Stand
standard (Anthropic)	2,00	10,00	0,20	10. September 2026

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

standard: Der zum Start als befristeter Einführungspreis angekündigte Satz von 2 bzw. 10 US-Dollar ist der Standardpreis. Die für den 1. September 2026 vorgesehene Erhöhung auf 3 bzw. 15 US-Dollar findet nicht statt.

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

Benchmarks



Quelle: Anthropic (Herstellerangabe), Stand: 30. Juni 2026 · Grafik: AI NEWS pro

Stand der Daten: 6. Oktober 2026

Claude Sonnet 5

Sonnet 5 ist als zweites Modell der Claude-5-Familie erschienen und darin für den Regelfall zuständig. Während Opus 5 für agentische Programmierarbeit positioniert ist und Fable 5 die Spitzenstufe bildet, beantwortet Sonnet 5 den Alltag - jede Anfrage, für die kein anderes Modell gewählt wurde.

Zwei Ausstattungsmerkmale gab es in der Sonnet-Reihe vorher nicht. Das Modell verarbeitet Bilder erstmals hochauflösend, mit bis zu 2.576 Pixeln Kantenlänge statt bisher 1.568. Dazu kennt es als erstes Sonnet die Aufwandsstufe "xhigh", eine zusätzliche Einstellung für die Denktiefe oberhalb von "high". Beim Kontextfenster gibt es ohnehin keinen Klassenunterschied mehr - die günstigste Stufe bekommt dieselbe aktuelle Größe wie die teureren Geschwister.

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

Auch die Voreinstellung fürs Reasoning hat sich gedreht. Bei Sonnet 4.6 blieben die Denkschritte aus, wenn der Aufruf den zuständigen Parameter nicht enthielt. Sonnet 5 wechselt in diesem Fall in den adaptiven Modus und entscheidet selbst, wie viel Überlegung eine Anfrage braucht. Ganz abschalten lässt sich das Denken aber weiterhin, ohne Bedingungen - innerhalb der Familie ist das die Ausnahme.

Den tiefsten Schnitt macht der neue Tokenizer. Er zerlegt denselben Text in rund 30 Prozent mehr Token als der des Vorgängers, und weil nach Token abgerechnet wird, wandert diese Differenz direkt in die Rechnung.

Einordnung

Das Profil ist Breite statt Spitze. Tool Calls, Structured Output, Bilder und PDF-Dokumente als Input - das alles bringt Sonnet 5 mit, ohne in eine bestimmte Richtung zugeschnitten zu sein. Genau darin liegt der Sinn eines Standardmodells, das nicht weiß, was als Nächstes kommt. Die hochauflösende Bildverarbeitung macht sich überall dort bemerkbar, wo Vorlagen fein aufgelöst sind, etwa bei Bildschirmfotos, Scans oder dicht beschrifteten Diagrammen.

> **Gleicher Listenpreis heißt nicht gleiche Kosten**
>
> Der reguläre Preis je Token ist derselbe wie bei Sonnet 4.6. Weil der neue Tokenizer denselben Text aber in rund 30 Prozent mehr Token zerlegt, wird eine gleichwertige Anfrage trotzdem teurer. In jeden Preisvergleich mit dem Vorgänger gehört dieser Faktor hinein.

Für Umsteiger von Sonnet 4.6 kommen sogar zwei Kostentreiber zusammen. Neben dem Tokenizer erzeugt auch das standardmäßig aktive Reasoning Token an Stellen, an denen der Vorgänger ohne den Parameter gar nicht gedacht hat. Beim Listenpreis gibt es dagegen Entwarnung: Die 2 und 10 US-Dollar je Million Token waren zum Start als befristeter Einführungspreis angekündigt, die für den 1. September 2026 vorgesehene Erhöhung auf 3 und 15 US-Dollar hat Anthropic nicht vollzogen. Es ist der reguläre Satz (Stand 10. September 2026).

Dass sich das Denken uneingeschränkt abschalten lässt, verschafft Sonnet 5 zugleich die größte Spannweite der Familie: unten der Betrieb ganz ohne Denkschritte, oben die neue Aufwandsstufe "xhigh", dazwischen der adaptive Modus, der die Tiefe selbst wählt. Für ein Modell, das als Voreinstellung sehr unterschiedliche Lasten trägt, ist genau diese Regelbarkeit das passende Profil.

Bei der Leistung bleibt das Datenblatt vorerst wortkarg. Belegte Messwerte aus den gängigen Benchmark-Suiten liegen für Sonnet 5 zum Stichtag nicht vor, und ohne sie lassen sich Vergleiche mit den Geschwistern oder der Konkurrenz nicht seriös anstellen.

Die Standardrolle bindet zugleich an den Anbieter. Wer Sonnet 5 einsetzt, mietet Rechenzeit - herunterladen und im eigenen Haus betreiben lässt sich das Modell nicht, und am Ende jeder Anfrage steht Text. Beim Bezugsweg gibt es dagegen Auswahl: Neben Anthropics eigener Schnittstelle führen auch die Cloud-Plattformen von Amazon, Google und Microsoft das Modell - für Häuser, die ihre Verträge ohnehin dort

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

bündeln, fällt damit eine Beschaffungshürde weg.

Verwandte Modelle

Direkter Vorgänger ist Claude Sonnet 4.6, das Sonnet 5 am Erscheinungstag als Standardmodell der Anthropic-Produkte abgelöst hat. Die beiden Punkte, die den Umstieg prägen, stehen oben: der neue Tokenizer und das standardmäßig aktive Reasoning.

Gemessen an den regulären Preisen ordnet sich Sonnet 5 unten in der eigenen Familie ein. Claude Opus 5 ist auf agentisches Programmieren und Rechnersteuerung per Bildschirmfoto ausgelegt, kostet regulär rund zwei Drittel mehr und lässt sein Reasoning nur bis zur Aufwandsstufe "high" abschalten. Claude Fable 5 hat die Familie im Juni eröffnet, hervorgegangen aus einer nur auf Einladung zugänglichen Vorschau, und bleibt deren Spitzenstufe - mehr als der dreifache Sonnet-Preis, bei dauerhaft aktivem Reasoning.

Quellen

1. Anthropic - Modellübersicht und Preise · https://platform.claude.com/docs/en/about-claude/models/overview
2. Anthropic: Preisangabe Claude Sonnet 5 (standard) · Stand 10.09.2026 · https://platform.claude.com/docs/en/about-claude/pricing
3. Anthropic: Benchmarks: Sonnet 5 im Vergleich mit Sonnet 4.6 · Stand 30.06.2026 · https://www.marktechpost.com/2026/07/13/anthropic-claude-sonnet-5-vs-sonnet-4-6-vs-opus-4-8-agentic-coding-benchmarks-api-pricing-and-cost-performance-tradeoffs-compared/

Themen: Modelle · Anthropic

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