MODELLE

Grok 4.5 - technische Daten, Benchmarks und Preise

Grok 4.5 ist das erste Grok-Modell seit der Übernahme von xAI durch SpaceX. Was zu Technik, Preis und Verfügbarkeit belegt ist, was am Kontextfenster ungewöhnlich ist - und warum der kursierende Wert 54 kein Benchmark-Ergebnis ist.

Stand: 04.08.2026 · 4 Min. Lesezeit

Anfang Juli 2026 hat xAI mit Grok 4.5 die nächste Ausbaustufe seiner Grok-4-Reihe freigegeben - das erste Modell der Marke, seit sie zum Raumfahrtkonzern SpaceX gehört. Der Umbau dahinter ist mehr als eine Fußnote, denn hinter dem vertrauten Anbieternamen steht kein eigenständiges Unternehmen mehr. Technisch tritt das Modell die Nachfolge von Grok 4.3 an und geht dabei einen Weg, den man selten sieht: Das Kontextfenster ist kleiner geworden statt größer. Belegte Benchmark-Einzelwerte fehlen dagegen bis heute - auch das gehört zum Bild.

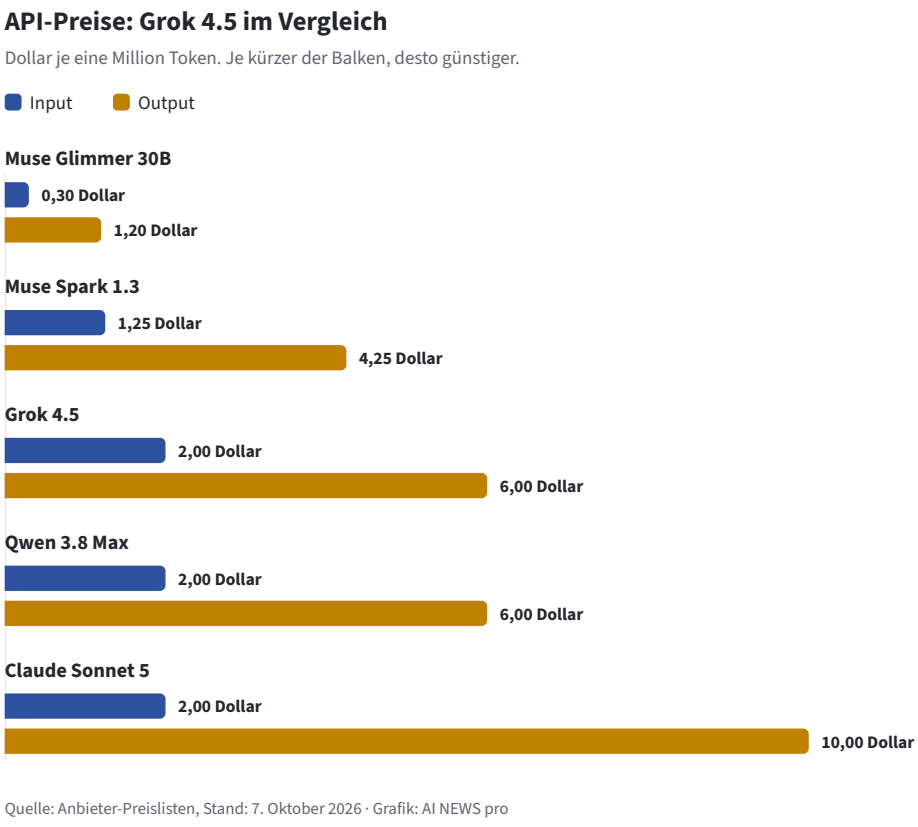
Grok 4.5 - Überblick und technische Daten

Anbieter	xAI
Familie	Grok 4
Erschienen	8. Juli 2026 8. Juli 2026 für Entwickler, 9. Juli 2026 für die Allgemeinheit.
Lizenz	Proprietär, ausschließlich über Schnittstellen nutzbar.
Weights	closed weights, kein Selbstbetrieb
Nutzbar über	xAI, OpenRouter
Preis ab	2,00 USD je Mio. Input-Token
Architektur	nicht veröffentlicht Vom Anbieter nicht veröffentlicht.
Parameter gesamt	nicht veröffentlicht Vom Anbieter ausdrücklich nicht veröffentlicht.

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

Kontextfenster	500.000 Token Bemerkenswert: kleiner als beim Vorgänger Grok 4.3, der 1 Mio. Token bot. Eine Verkleinerung des Kontextfensters zwischen zwei Generationen ist ungewöhnlich.
Input	Text Belegt ist nur Text. Ob das Modell auch Bilder entgegennimmt, geht aus den verfügbaren Angaben weder hervor noch wird es ausgeschlossen.
Output	Text
Tool Calls	Ja
Structured Output	Ja
Prompt Caching	Ja
Kennung	grok-4.5

Preise



Quelle: Anbieter-Preislisten, Stand: 7. Oktober 2026 · Grafik: AI NEWS pro

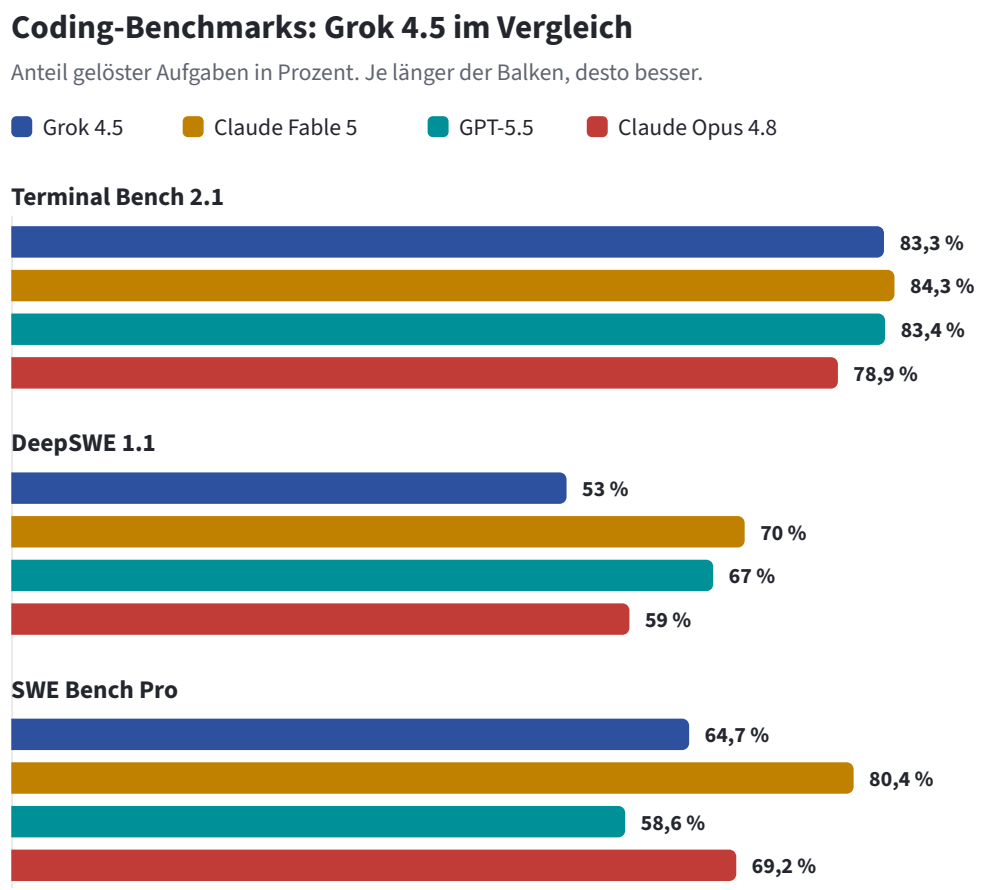
AI NEWS pro - KI-News ohne Hype! · © 2026 SOFTCLICK Labs · Stand des Beitrags: 04.08.2026

Stufe	Input	Output	Cache (Lesen)	Stand
standard (xAI)	2,00	6,00	0,50	4. August 2026

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

standard: Für Anfragen über 200.000 Token Kontext fällt ein Aufschlag an, dessen Höhe zum Stichtag nicht beziffert vorliegt. Der Cache-Lesepreis entspricht einem Nachlass von 75 Prozent auf den Eingabepreis.

Benchmarks



Quelle: xAI (Herstellerangabe), Stand: 8. Juli 2026 · Grafik: AI NEWS pro

Stand der Daten: 4. August 2026

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

Grok 4.5

Grok 4.5 gehört zur Grok-4-Familie, tritt die Nachfolge von Grok 4.3 an und wird über die Schnittstelle von xAI sowie über den Vermittler OpenRouter ausgeliefert - unter einer Marke, die kein eigenes Unternehmen mehr bezeichnet.

Die Firma xAI gibt es als eigenständige Gesellschaft nämlich nicht mehr, sie ist Anfang 2026 im Raumfahrtkonzern SpaceX aufgegangen. Grok 4.5 ist das erste Modell aus dieser neuen Struktur. Dass der Anbietername trotzdem bleibt, hat einen einfachen Grund: In den Schnittstellen werden die Modelle weiter unter der Marke xAI geführt.

Vom Start-up zur Konzerneinheit

2. Februar 2026	SpaceX übernimmt xAI in einem reinen Aktientausch. Die verbundene Gesellschaft wird mit rund 1,25 Billionen US-Dollar bewertet.
Mai 2026	Grok und die Plattform X laufen fortan unter der SpaceX-Einheit "SpaceXAI".
12. Juni 2026	Börsengang an der Nasdaq.
8. und 9. Juli 2026	Grok 4.5 erscheint - erst für Entwickler, einen Tag später für alle.

Über das Innenleben sagt der Anbieter wenig. Architektur und Parameterzahl sind ausdrücklich nicht veröffentlicht - das ist eine bewusste Entscheidung, keine Recherchelücke, und bei geschlossenen Modellen dieser Klasse inzwischen eher Regel als Ausnahme. Belegt sind dafür Tool Calls, Structured Output und Prompt Caching, also die Grundausstattung für den Einsatz in automatisierten Abläufen. Bei den Eingabeformaten führen die geprüften Quellen nur Text auf. Ob das Modell auch Bilder verarbeitet, ist damit weder belegt noch ausgeschlossen.

Der auffälligste Abstand zum Vorgänger betrifft das Kontextfenster, und zwar in ungewohnter Richtung. Bei Grok 4.3 lag die Grenze bei einer Million Token Eingabe, bei Grok 4.5 liegt sie bei der Hälfte. Eine solche Verkleinerung zwischen zwei Generationen ist ungewöhnlich, üblich ist der Weg nach oben. Eine Erklärung dafür liefern die geprüften Quellen nicht.

Einordnung

Vom belegten Funktionsumfang her zielt Grok 4.5 auf Werkzeugketten und automatisierte Abläufe. Tool Calls und Structured Output gehören dazu, und das Prompt-Caching senkt die Kosten wiederholt gesendeter Eingaben um drei Viertel gegenüber dem regulären Eingabepreis. Für Anwendungen, die denselben Systemkontext bei jeder Anfrage mitschicken, ist das der wirksamste Hebel in der Preisliste.

Das verkleinerte Kontextfenster bleibt für die meisten Zwecke komfortabel - ganze Codebasen oder Dokumentsammlungen passen weiterhin in einen einzelnen Aufruf. Wer öfter an die Obergrenze geht, plant allerdings mit einer Unbekannten: Oberhalb von 200.000 Token Kontext wird ein Aufschlag fällig, dessen Höhe

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

zum Stichtag nicht beziffert vorliegt. Beim Listenpreis selbst bewegt sich das Modell in derselben Liga wie Alibabas Qwen 3.8 Max und damit im mittleren Feld der aktuellen Generation.

Klar gezogen ist die Grenze beim Betrieb. Grok 4.5 gibt es ausschließlich über Schnittstellen, die Gewichte sind nicht offen, ein Betrieb auf eigener Hardware scheidet aus. Für Umgebungen, die eine lokale Installation verlangen, fällt das Modell damit aus der Auswahl - unabhängig von allen anderen Eigenschaften.

Am dünnsten ist die Beleglage bei den Messwerten. Knapp vier Wochen nach dem Erscheinen liegt keine nachprüfbare Einzelmessung aus einer der gängigen Benchmark-Suiten vor, weder vom Anbieter noch von unabhängiger Seite. Die Benchmark-Tabelle dieser Seite bleibt deshalb leer, bis sich das ändert.

> **Häufiger Irrtum**
>
> Für Grok 4.5 kursiert der Wert 54, gern als Benchmark-Ergebnis gelesen. Tatsächlich stammt er aus dem Gesamtindex von Artificial Analysis - einem gewichteten Sammelmaß aus neun Einzelevaluationen, in dem das Modell zum Stand 4. August 2026 Platz 4 von 168 geführten Modellen belegt. Als grobe, unabhängig erhobene Standortbestimmung taugt der Index, als Messwert einer einzelnen Suite nicht.

Verwandte Modelle

Innerhalb der eigenen Reihe ist Grok 4.3 der direkte Vorgänger - gleiche Familie, gleiches Auslieferungsmodell, aber mit doppelt so großem Kontextfenster. Nachgerückt ist am 12. August 2026 Grok 4.6, fünf Wochen nach Grok 4.5 - mit unverändert 500.000 Token Kontextfenster und dem Schwerpunkt Selbstkontrolle bei langen Aufgaben.

Unter den geschlossenen Modellen derselben Generation sind OpenAIs GPT-5.6-Familie und Anthropics Claude-Reihe die nächsten Vergleichsgrößen. Die drei GPT-Stufen Sol, Terra und Luna unterscheiden sich technisch nur im Preis, bei Anthropic sind Claude Opus 5 und Claude Sonnet 5 die aktuellen Gegenstücke. Preislich am nächsten steht Grok 4.5 dem Qwen 3.8 Max von Alibaba.

Wer offene Gewichte sucht, wird in dieser Generation woanders fündig: DeepSeek V4 Flash 0731 steht unter MIT-Lizenz, Moonshots Kimi K3 unter einer eigenen Anbieterlizenz mit kommerzieller Schranke - offen im Sinne von "open weight", nicht von Open Source.

Quellen

1. OpenRouter: Grok 4.5 - Preise und Benchmarks · https://openrouter.ai/x-ai/grok-4.5
2. Artificial Analysis: Modellübersicht · https://artificialanalysis.ai/models
3. DataCamp: Grok 4.5 - Features, Benchmarks, Pricing · https://www.datacamp.com/blog/grok-4-5

Themen: Modelle · xAI · Grok

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