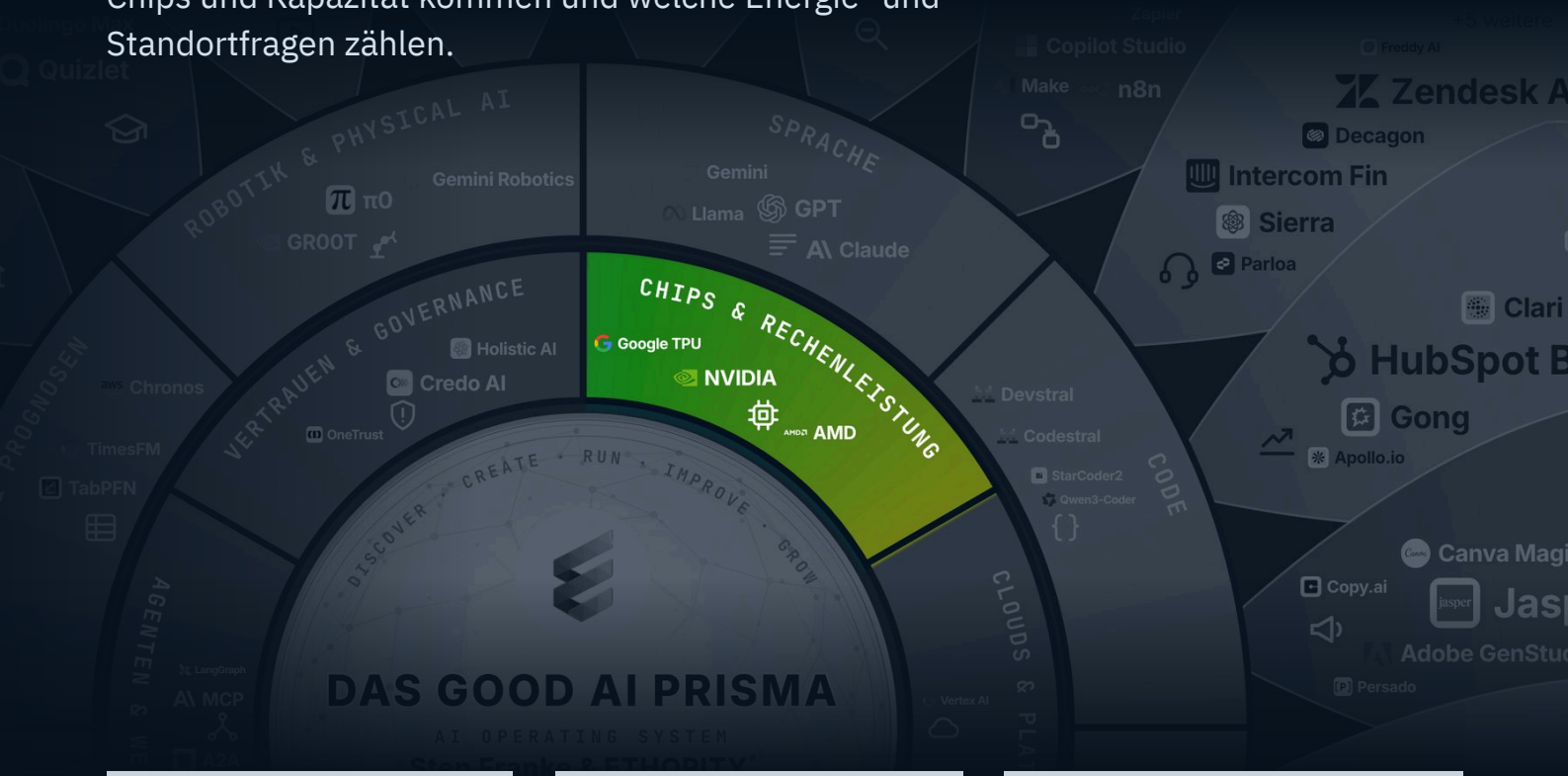
ETHORITY®
AI BUSINESS OPERATING SYSTEM

TECHNOLOGIEBRIEF

GOOD AI PRISMA · TECHNOLOGIEBRIEF 2026/27 · CHIPS & RECHENLEISTUNG

2.980 MW in deutschen Rechenzentren. Rechenleistung sicher planen.

KI wirksam einsetzen mit sicherer Rechenleistung: woher Chips und Kapazität kommen und welche Energie- und Standortfragen zählen.



2.980 MW

IT-Leistung der Rechenzentren in Deutschland 2025, über 5.000 MW bis 2030 erwartet

BITKOM, NOV. 2025

21

Angebote für KI-Chips und Rechenleistung kartiert

GOOD AI PRISMA, OKT. 2026

12. Nov. 2026

Frist des EU-Aufrufs für bis zu sieben KI-Gigafabriken

EUROHPC, JULI 2026

DEUTSCHE AUSGABE · DE · v1.1 · OKT. 2026

Erstellt von Sten Franke mit einer Perspektive auf Menschen und Governance von Severin L. Wilson

Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

TECHNOLOGIEBRIEF 2026/27 · DEUTSCHE AUSGABE

In diesem Brief.

Woher KI-Chips und Rechenkapazität kommen, was der Markt bietet und welche Regeln, Energie- und Standortfragen in Deutschland und der EU zählen.



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · Panoramastreifen · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

01	**Die Nachfrage läuft dem Angebot voraus.**	4
02	**Preise laufen auseinander, Strom und Standort entscheiden.**	8
03	**Export, Energie und Souveränität bestimmen die Bedingungen.**	12
04	**21 Angebote für KI-Rechenleistung, vom Chip bis zur Cloud.**	16
05	**Genug Kapazität sichern, zu planbaren Bedingungen.**	20
Überblick	**Auf einen Blick: sechs Zahlen zur KI-Rechenleistung**	3
ETHORITY	**Das Problem, wie ETHORITY hilft, über ETHORITY**	26
Methode	**Methode und Quellen**	29
Impressum	**Impressum, Lizenz und Herkunft**	31

ÜBER DIESEN BRIEF

Die Marktzahlen stammen aus dem Datensatz der Good AI Prisma Studie 2026/27: Jeder Eintrag ist mit einer datierten Quelle von den Seiten des Anbieters erfasst. Die Aufgabengruppen sind eine Einordnung von ETHORITY. Externe Zahlen nennen Quelle und Datum und stammen aus Unternehmensberichten, von der IEA, EuroHPC, Bitkom und Marktbeobachtern. Dieser Brief ist keine Rechts- oder Anlageberatung.

AUF EINEN BLICK

Rechenleistung ist der knappe Rohstoff der KI-Wirtschaft.

Jede KI-Anwendung läuft auf Chips, Speicher, Strom und Rechenzentrumsfläche. Die Nachfrage nach allen vieren wächst schneller als das Angebot. Wo und zu welchen Bedingungen ein Unternehmen Rechenleistung bezieht, bestimmt deshalb Kosten, Tempo und Souveränität.

Dieser Brief wendet die Good AI Prisma Marktstudie 2026/27 auf eine Technologieebene an: Chips und Rechenleistung. Er verbindet die Marktkarte der Studie mit 21 Angeboten mit aktueller Evidenz zu Angebot, Preisen, Energie und Regeln in Deutschland und der EU, mit Vergleichen aus den USA und Asien, und endet mit den Fragen vor einer Zusage und damit, wie ETHORITY hilft.

WEITERLESEN

Der Bereich Chips & Rechenleistung online, der Technologiebrief KI-Modelle und die Good AI Prisma Marktstudie.

2.980 MW

IT-Leistung der Rechenzentren in Deutschland 2025, 9 % mehr als im Vorjahr

BITKOM

40 %

der Rechenzentrumsleistung in Deutschland sollen 2030 auf KI entfallen, nach etwa 15 % 2025

BITKOM

21,3 TWh

Stromverbrauch der Rechenzentren in Deutschland 2025

BITKOM

1,0 EF

Rechenleistung von JUPITER in Jülich, Europas erstem Exascale-System

TOP500, NOV. 2025

10.000

NVIDIA-Blackwell-GPUs maximal in der Industrial AI Cloud der Telekom in München

NOV. 2025

bis zu 7

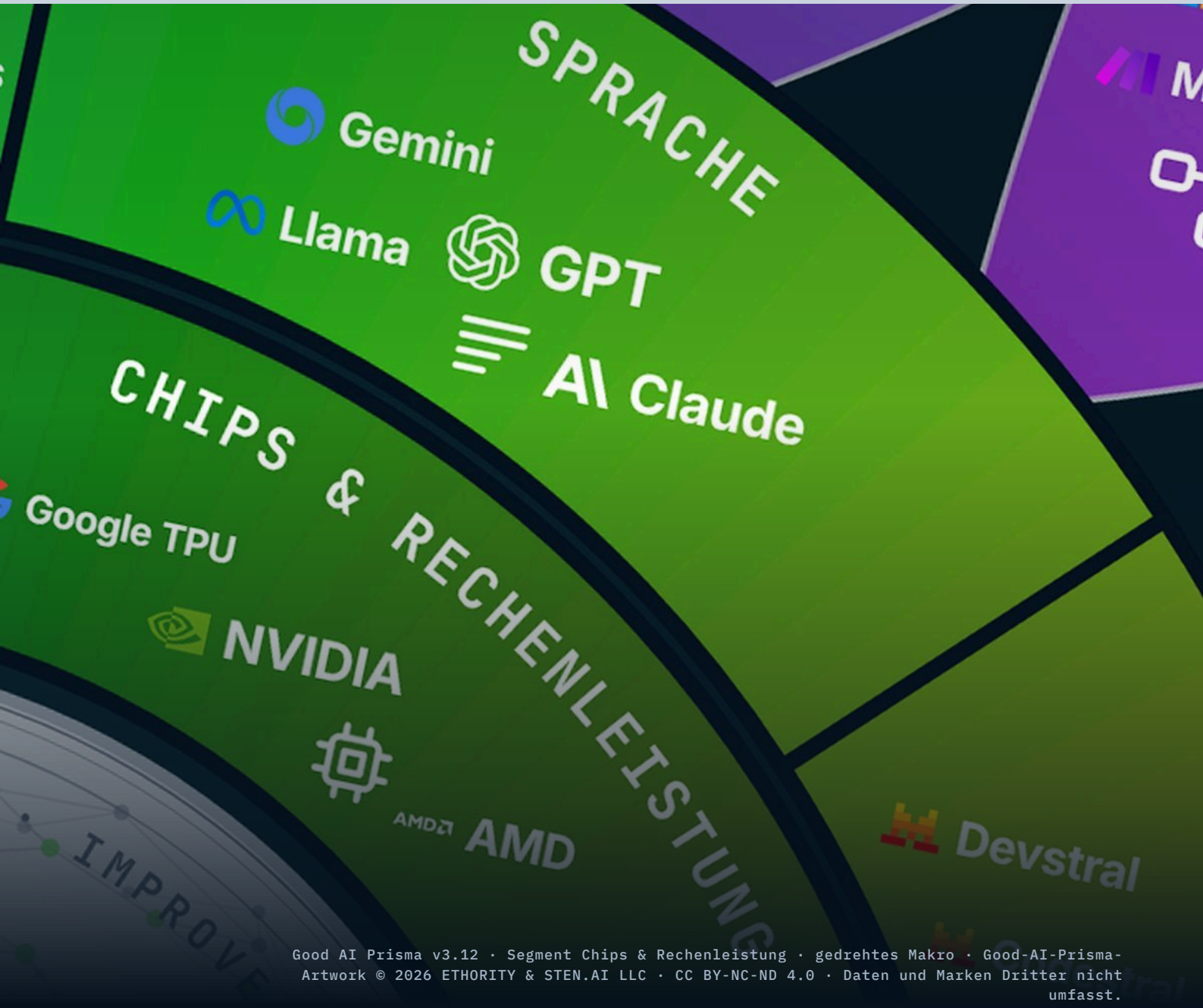
Standorte für KI-Gigafabriken sucht die EU, mit mehr als 20 Mrd. € privaten Investitionen

EUROHPC, JULI 2026

01

Die Nachfrage läuft dem Angebot voraus.

In Deutschland wuchs die IT-Leistung der Rechenzentren 2025 um 9 % auf 2.980 MW, und der KI-Anteil soll bis 2030 auf etwa 40 % steigen. Weltweit wächst die Nachfrage schneller als das Angebot: ASML erwartet, dass das geplante Angebot 2027–28 die Nachfrage nicht vollständig deckt.



Good AI Prisma v3.12 · Segment Chips & Rechenleistung · gedrehtes Makro · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

40 %

der Rechenzentrumsleistung in Deutschland sollen 2030 auf KI entfallen, nach etwa 15 % 2025

BITKOM, RECHENZENTREN IN DEUTSCHLAND, 10.11.2025

WO DER MARKT STEHT

Deutschland baut KI-Kapazität aus.

2.980 MW

IT-Leistung der Rechenzentren in Deutschland 2025

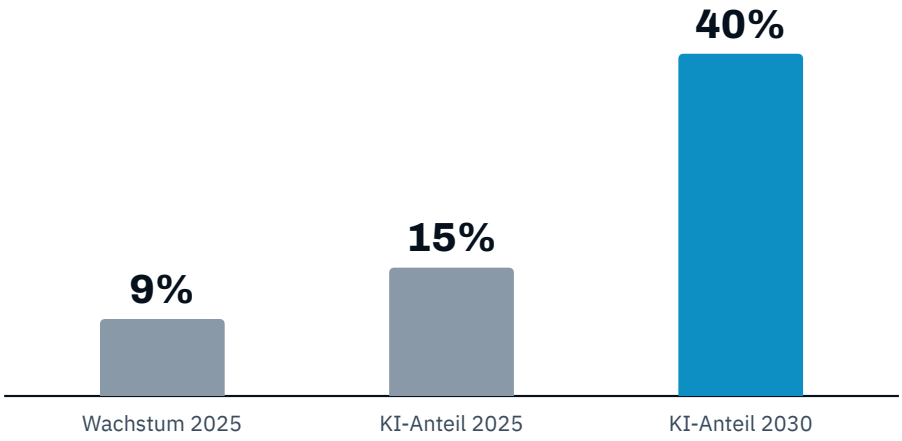
BITKOM

In Deutschland hatten Rechenzentren 2025 eine IT-Leistung von 2.980 MW, 9 % mehr als im Vorjahr; bis 2030 erwartet Bitkom mehr als 5.000 MW. Der KI-Anteil soll von etwa 15 % auf etwa 40 % steigen. Mehr als ein Drittel der Leistung steht in Hessen, rund um Frankfurt sind über 1.800 MW in Planung.

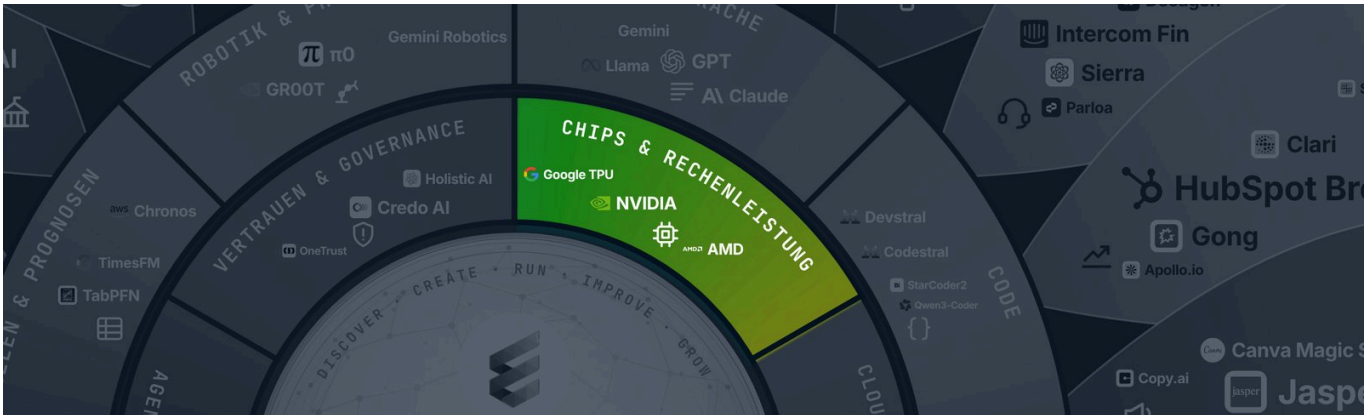
ABBILDUNG 1.1

In Deutschland könnte KI 2030 rund 40 % der Rechenzentrumsleistung beanspruchen.

Wachstum der IT-Leistung 2025; ungefährer KI-Anteil an der Leistung 2025 und 2030 (Prognose)



Quelle · Bitkom, Rechenzentren in Deutschland, 10.11.2025



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · Panorama durch das Segment und seinen Ring · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

WO DER MARKT STEHT

Weltweit läuft die Nachfrage dem Angebot voraus.

Weltweit wächst die Nachfrage schneller als das Angebot: NVIDIAs Umsatz mit Rechenzentren hat sich binnen eines Jahres auf 89,0 Mrd. US$ je Quartal mehr als verdoppelt, und ASML erwartet, dass das geplante Angebot 2027–28 die Nachfrage nicht vollständig deckt. Rechenzentren verbrauchten 2024 weltweit 415 TWh, davon 15 % in Europa (IEA, Energy and AI).

89,0

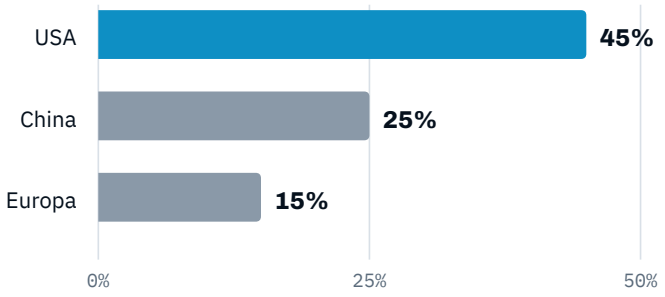
Mrd. US$ NVIDIA-Umsatz mit Rechenzentren in einem Quartal (Q2 GJ2027)

NVIDIA, AUG. 2026

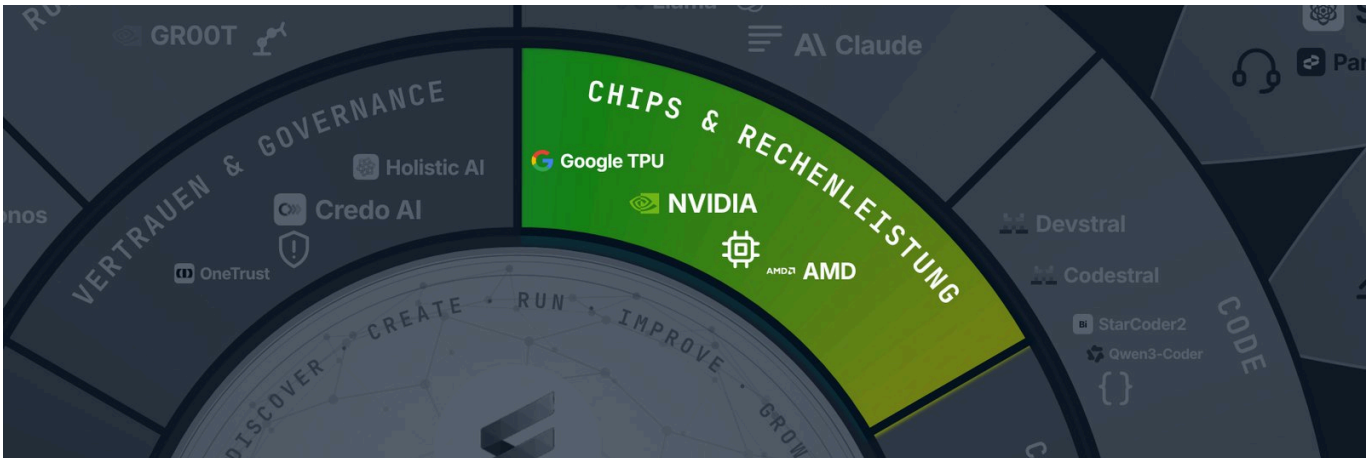
ABBILDUNG 1.2

Die USA verbrauchen 45 % des Rechenzentrumsstroms.

Anteil am weltweiten Stromverbrauch der Rechenzentren nach Region, 2024



Quelle · IEA, Energy and AI, 2025 · weltweiter Verbrauch der Rechenzentren 2024: 415 TWh



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · Segment mit Anbieterlogos · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

02

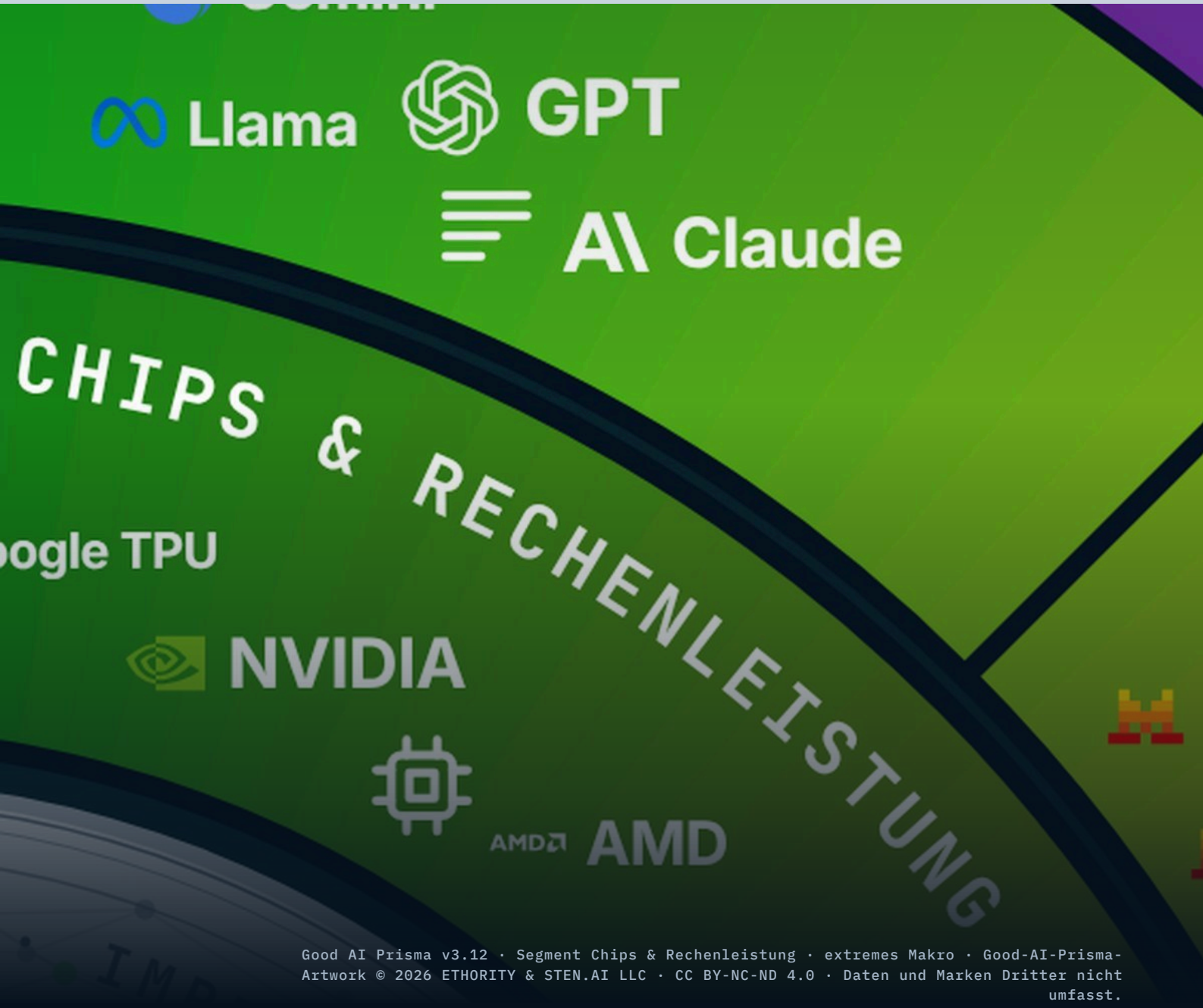
Preise laufen auseinander, Strom und Standort entscheiden.

Inferenz wird stetig günstiger, GPU-Miete folgt dem nicht: Eine H100-Stunde kostet bei Neoclouds rund ein Drittel der On-Demand-Preise der Hyperscaler. Für Deutschland zählen Strom und Netzanschluss, und das Energieeffizienzgesetz stellt Anforderungen an Effizienz, Ökostrom und Abwärme.

Chips & Rechenleistung

IN DIESEM KAPITEL

INFERENZ WIRD GÜNSTIGER, · 10
STROM UND NETZANSCHLUSS · 11



Good AI Prisma v3.12 · Segment Chips & Rechenleistung · extremes Makro · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

2,53 US$

je H100-GPU-Stunde bei Neoclouds, rund ein Drittel der On-Demand-Preise der Hyperscaler

SILICON-DATA-INDEX VIA IMPLICATOR, JULI 2026

PREIS, STROM UND STANDORT

Inferenz wird günstiger, GPU-Miete nicht.

15–18 %

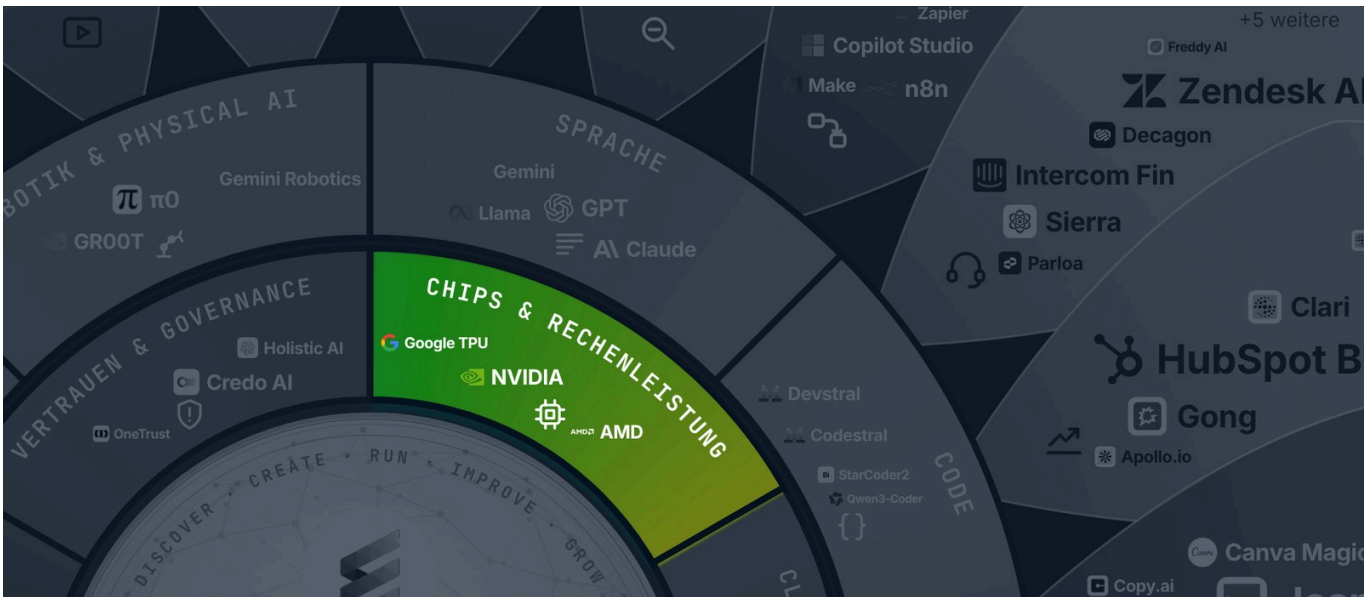
Anstieg der Speicher-Kontraktpreise in einem Quartal

TRENDFORCE, JULI 2026

Inferenz wird stetig günstiger: Der Preis für ein bestimmtes Leistungsniveau sank je nach Aufgabe um das 9- bis 900-Fache pro Jahr (Epoch AI price trends). GPU-Miete folgt dem nicht: Eine H100-Stunde kostet bei Neoclouds etwa 2,53 US$, rund ein Drittel der On-Demand-Preise der Hyperscaler, und die Preise blieben zuletzt stabil, während Speicherpreise in einem Quartal um 15–18 % stiegen (TrendForce).

Der Preis je Leistungsniveau sank um das 9- bis 900-Fache pro Jahr.

EPOCH AI, MÄRZ 2025



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · Segment und die Anwendungsbereiche dahinter · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

PREIS, STROM UND STANDORT

Strom und Netzanschluss entscheiden.

Für Deutschland zählen Strom und Netzanschluss: Rechenzentren verbrauchten 2025 21,3 TWh (Bitkom-Rechenzentrumsstudie). Das Energieeffizienzgesetz stellt Anforderungen an Effizienz, Ökostrom und Abwärmenutzung; eine Novelle ist in Arbeit.

Neue Gigafabrik-Kapazität kommt ab 2028.

EUROHPC, JULI 2026

23.752

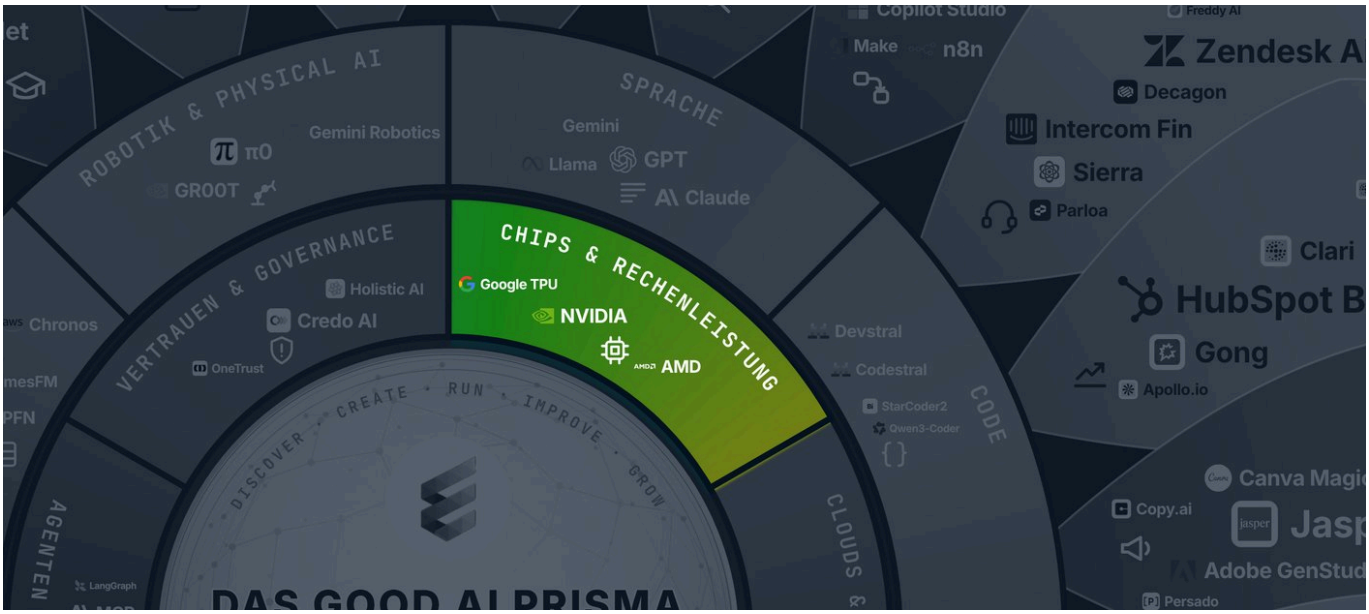
GH200-Superchips in JUPITER in Jülich, Europas erstem Exascale-System

TOP500, NOV. 2025

48 GW

Rechenzentrumsleistung in den USA 2024, gegenüber rund 3 GW in Deutschland

BITKOM



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · Segment und Nachbarringe · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

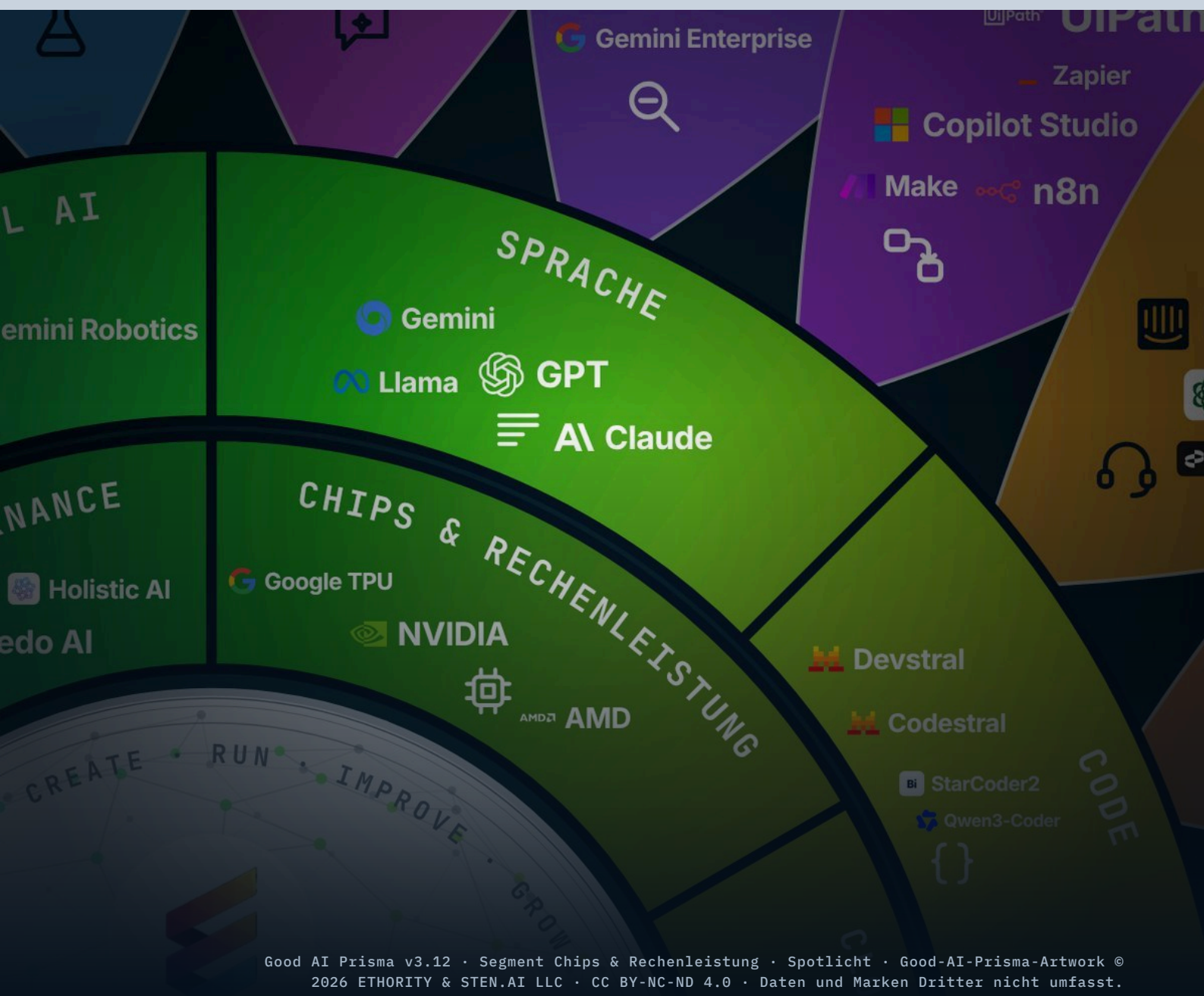
03

Export, Energie und Souveränität bestimmen die Bedingungen.

Für Rechenleistung kommen die Regeln aus drei Richtungen: Exportkontrollen entscheiden, wer welche Chips kaufen darf, Energie- und Rechenzentrumsregeln, wo Kapazität entsteht und zu welchen Kosten, und der AI Act setzt Pflichten für KI-Modelle mit allgemeinem Verwendungszweck.

Chips & Rechenleistung

IN DIESEM KAPITEL
EUROPÄISCHE UNION: · 14
DEUTSCHLAND ERGÄNZT · 15



Good AI Prisma v3.12 · Segment Chips & Rechenleistung · Spotlicht · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

12. Nov. 2026

Frist des EU-Aufrufs für bis zu sieben KI-Gigafabriken

EUROHPC JU, AUFRUF KI-GIGAFABRIKEN, 30.7.2026

DIE REGELN

Europäische Union: die Termine.

ABBILDUNG 3.1

Europas neue KI-Kapazität kommt ab 2028.

Termine für Käufer von Rechenleistung; Cyan: neue Kapazität



Quelle · EuroHPC JU, Aufruf KI-Gigafabriken, 30.7.2026; Delegierte Verordnung (EU) 2024/1364; CNAS zur H200-Regel

TERMINE

Pflichten für Anbieter von KI-Modellen mit allgemeinem Verwendungszweck gelten seit 2. Aug. 2025; das KI-Büro setzt sie seit 2. Aug. 2026 direkt durch (AI Act in der Fassung des Digital Omnibus, Verordnung (EU) 2026/1744). Im EuroHPC-Netz bestehen bereits 19 KI-Fabriken und 13 Antennen.

15. Sep. 2024 · EED	Rechenzentren ab 500 kW berichten jährlich Energie- und Nachhaltigkeitskennzahlen (Energieeffizienzrichtlinie).
13. Jan. 2026 · US-Export	US-Regel für Chips der H200-Klasse nach China: Einzelgenehmigung, Obergrenze von 50 % des US-Volumens und 25 % Abgabe; Blackwell bleibt gesperrt.
12. Nov. 2026 · Aufruf	Der EuroHPC-Aufruf für KI-Gigafabriken endet: bis zu sieben Standorte, mehr als 20 Mrd. € private Investitionen, Auswahl Anfang 2027.
ab 2028 · Kapazität	Neue Gigafabrik-Kapazität kann starten: Ausgewählte Standorte sollen binnen 18 Monaten nach der Auswahl in Betrieb gehen.

DIE REGELN

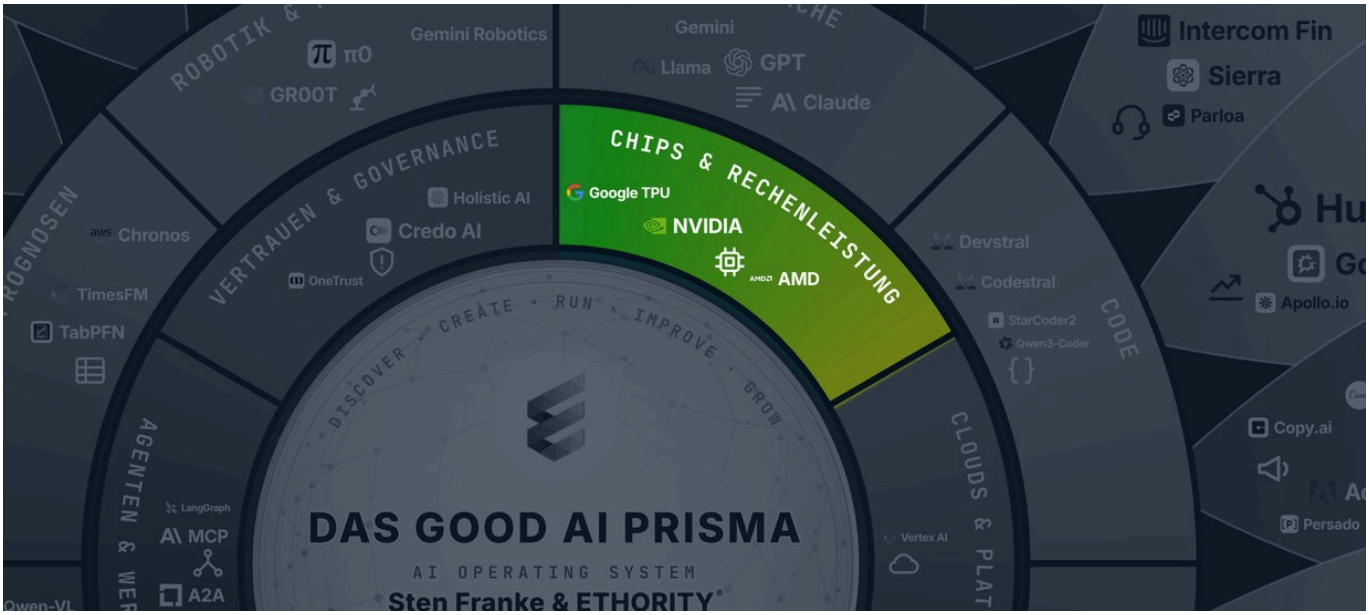
Deutschland ergänzt eigene Bausteine.

- Energieeffizienzgesetz: Effizienz-, Ökostrom- und Abwärmevorgaben für Rechenzentren; eine Novelle ist in Arbeit.
- JUPITER in Jülich: Europas erstes Exascale-System (1,0 Exaflops, 23.752 GH200-Superchips), Anker der EuroHPC-KI-Fabrik in Jülich.
- Industrial AI Cloud der Telekom in München: bis zu 10.000 NVIDIA-Blackwell-GPUs für die Industrie, Investition 1 Mrd. €.

1,0 EF

Rechenleistung von JUPITER in Jülich, Europas erstem Exascale-System

TOP500, NOV. 2025



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · innere Ringe und Kern · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

04

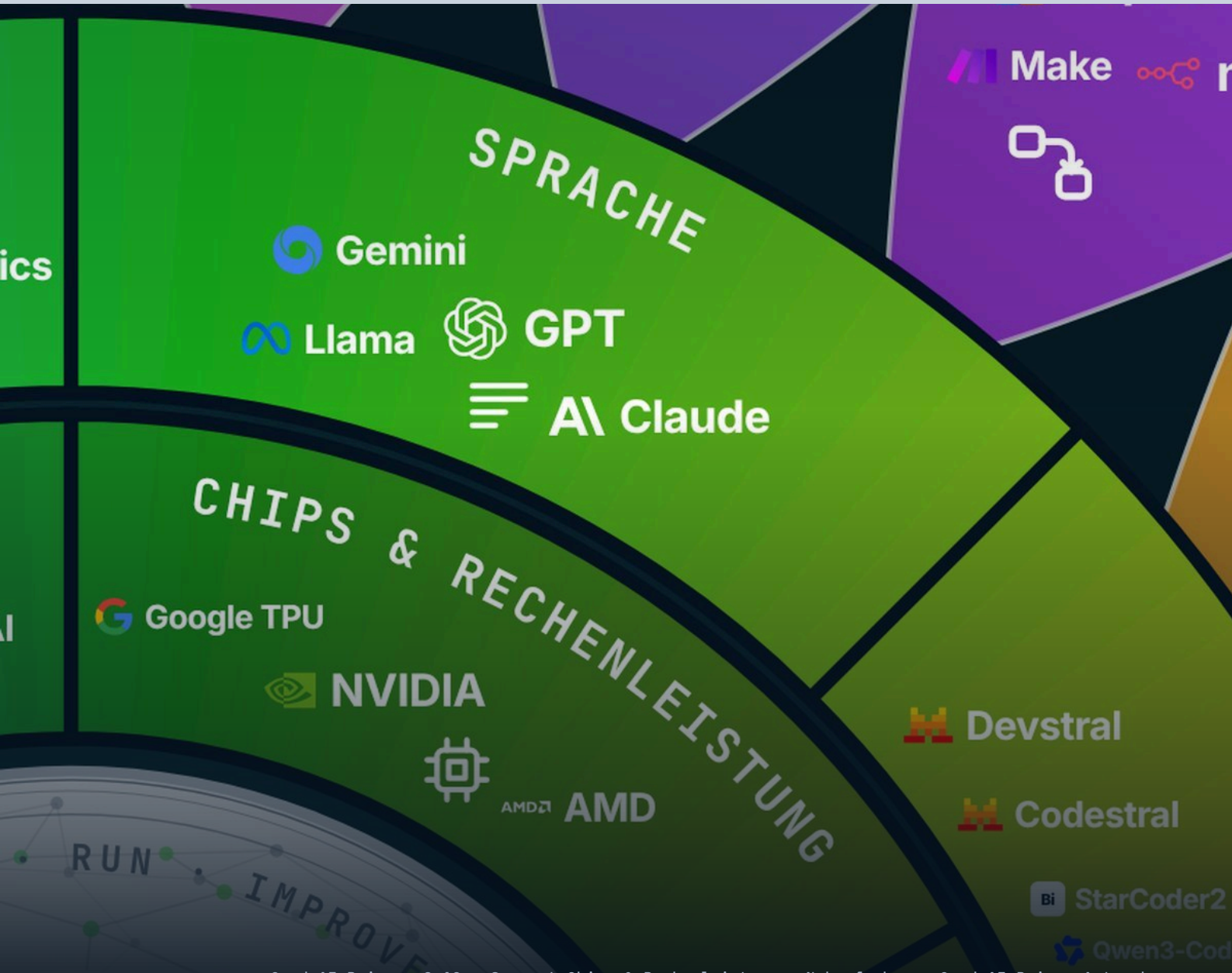
21 Angebote für KI-Rechenleistung, vom Chip bis zur Cloud.

Das Good AI Prisma 2026/27 ordnet 21 Angebote dem Bereich Chips und Rechenleistung zu Fünf sind GPUs und allgemeine Beschleuniger, fünf GPU-Clouds und Neoclouds, die Kapazität vermieten Vier sind eigene KI-Chips der Hyperscaler, vier Inferenz-Chips; zwei bringen KI an den Rand des Netzes, eines verbindet Beschleuniger mit Licht.

Chips & Rechenleistung

IN DIESEM KAPITEL

DIE KARTE CHIPS & RECHENLEISTUNG · 18
DER BEREICH CHIPS & RECHENLEISTUNG · 19



Good AI Prisma v3.12 · Segment Chips & Rechenleistung · Nahaufnahme · Good-AI-Prisma-Artwork
© 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

5

der 21 kartierten Angebote sind GPU-Clouds und Neoclouds, die Kapazität vermieten

DATENSATZ DER GOOD AI PRISMA STUDIE 2026/27, OKT. 2026

DIE MARKTKARTE

Die Karte Chips & Rechenleistung nach Aufgabe.

ABBILDUNG 4.1

GPUs und Neoclouds führen: 10 von 21 Angeboten.

Zahl der Angebote je Aufgabengruppe; Namen alphabetisch, kein Ranking. Logos wie im Artwork des Good AI Prismas v3.12.

5 — **GPUs und Beschleuniger**

AMD · Cerebras · Intel Gaudi · SambaNova RDU · Tenstorrent Blackhole

AI Act · US-Exportregeln bestimmen, welche Chips nach China dürfen; Lizenz- und Endnutzerklauseln prüfen

5 — **GPU-Clouds und Neoclouds**

CoreWeave · Industrial AI Cloud · Lambda · Nscale · Verda

AI Act · Energieregeln für Rechenzentren (EED-Berichte, Energieeffizienzgesetz) prägen Standort, Preis und Weitergabeklauseln

4 — **KI-Chips der Hyperscaler**

AWS Inferentia · AWS Trainium · Google TPU · Maia 200

AI Act · Nutzung in den Cloud-Regionen des Anbieters; Region, Datenresidenz und Portabilität prüfen

4 — **Inferenz-Chips**

Corsair · Etched · Groq · Qualcomm Cloud AI 100 Ultra

AI Act · Export-, Endnutzer- und Portabilitätsklauseln prüfen; bei der Inferenz wächst das Volumen

2 — **Edge-KI-Chips**

Hailo-10H · Metis AIPU

AI Act · In Geräten verbaut: AI-Act-Pflichten folgen Produkt und Einsatz, Anhang I ab 2.8.2028

1 — **Optische Verbindungstechnik**

Passage M1000

AI Act · Baustein von Beschleunigersystemen; Lieferung und Bedingungen folgen dem System, in das er eingebaut wird

Quelle · Datensatz der Good AI Prisma Studie 2026/27 · Aufgabengruppen: Einordnung von ETHORITY · Regeln knüpfen an Chip, Standort und Einsatz an: Exportregeln folgen dem Chip und seinem Käufer, Energieregeln dem Rechenzentrum, der AI Act dem Modell und seinem Einsatz.

Stand Oktober 2026 folgen auf erfasste Flaggschiffe Qualcomms AI200- und AI250-Racks, Axeleras Europa und SambaNovas SN50; NVIDIA hat Groqs Technologie lizenziert (Dez. 2025), GroqCloud läuft weiter.

ONLINE-ANSICHT

Der Bereich Chips & Rechenleistung im Good AI Prisma.

Der Bereich Chips und Rechenleistung im Good AI Prisma

Das Good AI Prisma online ordnet jedes Angebot seinem Geschäftsfeld zu und zeigt Quelle und Datum hinter jedem Eintrag. Es zeigt den Bereich Chips und Rechenleistung im Stand Oktober 2026.

ONLINE-ANSICHT

Einträge mit Quelle und Datum. Dieser Brief zählt alle 21 für die Ausgabe erfassten Einträge. (Good AI Prisma, Okt. 2026)

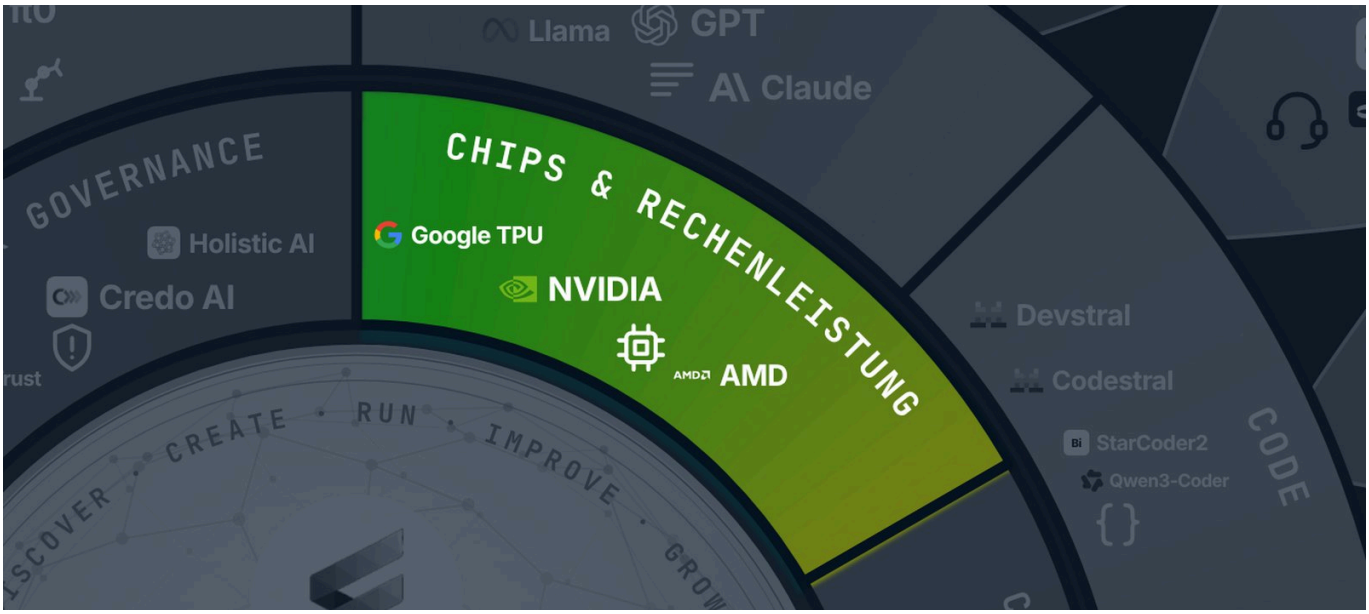
SZENARIO

So lesen Sie einen Bereich

Wählen Sie einen Bereich, um die dort eingeordneten Angebote zu sehen, und öffnen Sie ein Profil, um die Angabe des Anbieters, ihr Datum und die erfassten Entscheidungsfakten zu prüfen, etwa Preisbasis, Regionen und Datenhosting, sofern dokumentiert. Angebote direkt vergleichen.

Jeder Eintrag zeigt Quelle und Datum.

GOOD AI PRISMA, OKT. 2026



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · Logos im Segment · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

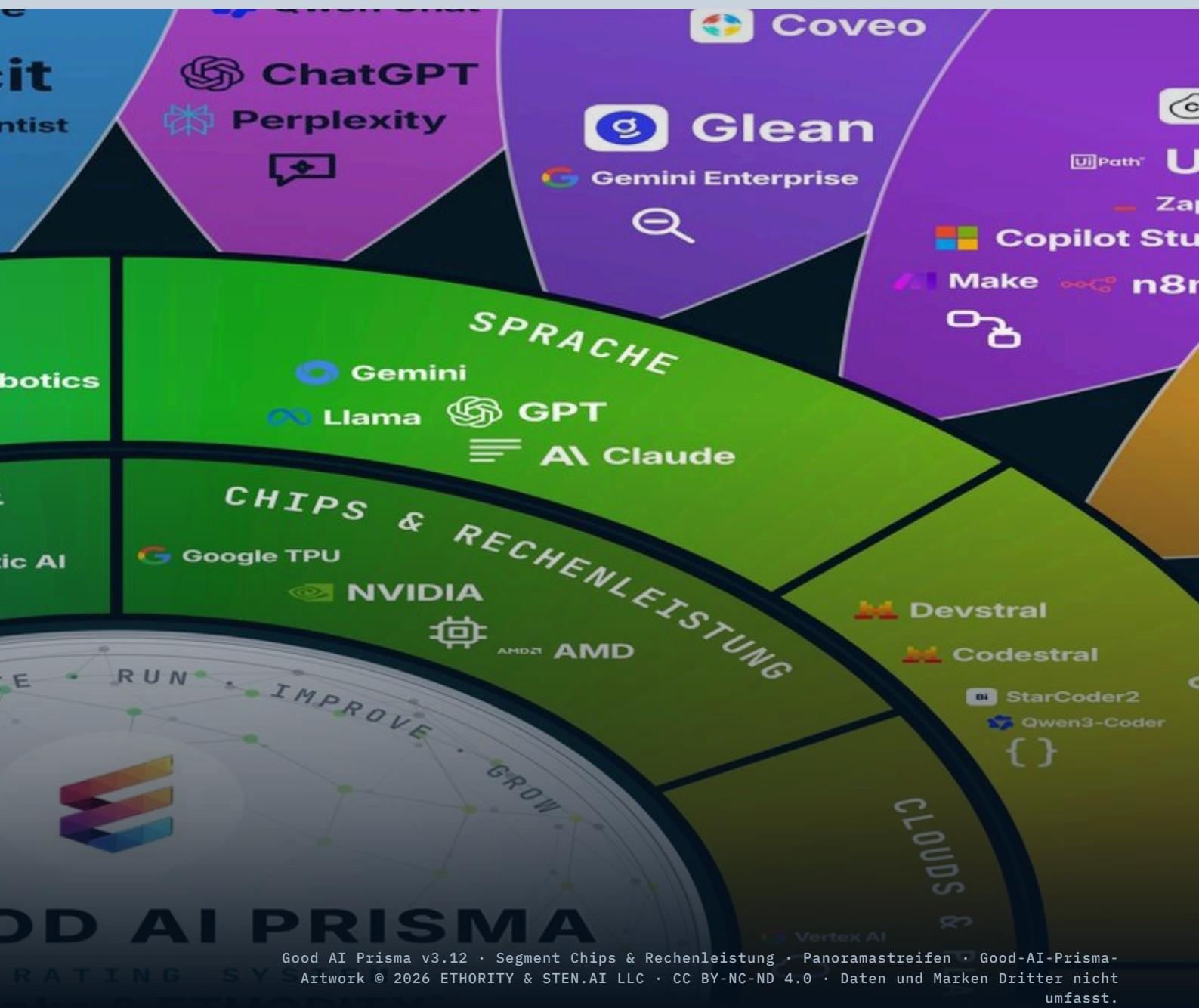
05

Genug Kapazität sichern, zu planbaren Bedingungen.

Rechenleistung ist zu einem strategischen Rohstoff geworden. Die Frage ist nicht mehr, welcher Chip am schnellsten ist, sondern wie sich genug Kapazität am richtigen Ort zu planbaren Bedingungen sichern lässt. Fünf Schlussfolgerungen für Entscheidungen jetzt.

IN DIESEM KAPITEL

WAS KÄUFER VON RECHENLEISTUNG BRAUCHEN · 22
DIE MENSCHLICHE SEITE · 23
EIN 90-TAGE-START · 24
WAS DIE EVIDENZ · 25



Good AI Prisma v3.12 · Segment Chips & Rechenleistung · Panoramastreifen · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

20 %

der geplanten Rechenzentren drohen Verzögerungen beim Netzanschluss

IEA, ENERGY AND AI, 2025

BEDARF UND ENTSCHEIDUNGEN

Was Käufer von Rechenleistung brauchen und was vor einer Zusage zu klären ist.

Was die Evidenz zeigt.

KAPAZITÄT

ASML und SK hynix rechnen mit Engpässen bis 2027; Speicher könnte bis 2030 knapp bleiben.

PREIS

Eine H100-Stunde kostet bei Neoclouds rund ein Drittel der On-Demand-Preise der Hyperscaler, während Inferenzpreise je Aufgabe weiter fallen (Epoch AI price trends).

ENERGIE

Rund 20 % der geplanten Rechenzentren drohen Verzögerungen beim Netzanschluss; Netze, Turbinen und Transformatoren sind Engpässe (IEA, Energy and AI).

STANDORT

Neue Gigafabrik-Kapazität in der EU kommt ab 2028; bis dahin bieten Cloud-Regionen, souveräne Clouds, Neoclouds und EuroHPC-KI-Fabriken Rechenleistung in der EU.

Sieben Fragen, bevor Sie sich auf Rechenleistung festlegen

1. Wofür kaufen wir: Training, Feinabstimmung oder Inferenz, und wie wächst das Volumen in drei Jahren?
2. Welche Region und welche Gesellschaft stellen die Kapazität, und erfüllt das unsere Anforderungen an Datenresidenz und Souveränität?
3. Was ist zugesagt: Kapazität, Liefertermin, Hardwaregeneration und Servicelevel, und was gilt bei Verzug?
4. Wie werden Preisänderungen geregelt: Reservierung, Weitergabe von Speicher- oder Energiekosten, Preisgleitklauseln?
5. Was ist die Energiebasis: Stromquelle, Effizienz (PUE) und die Kennzahlen für unsere eigene Berichterstattung?
6. Wie portabel ist unsere Software beim Wechsel von Chip oder Cloud, und was kostet der Wechsel?
7. Begrenzen Export- und Endnutzerklauseln, wo wir einsetzen dürfen, und wer zeichnet sie ab?

PLAN FÜR DIE BESCHÄFTIGTEN

Die menschliche Seite mit der Technik planen.

Wenn sich Agenden schnell ändern, bricht zuerst die menschliche Seite: halb eingeführte Werkzeuge, Kompetenzen, die dem System hinterherhinken, Beschäftigte, die nicht einbezogen wurden, und Entscheidungen ohne Verantwortliche. Planen Sie Einführung, Schulung, Beteiligung der Beschäftigten und verantwortliche Aufsicht als Teil jeder stufenweisen Zusage ein, nicht erst danach.

SEVERIN L. WILSON, PERSPEKTIVE BELEGSCHAFT IN DER GOOD AI PRISMA MARKTSTUDIE 2026/27

Anhang Belegschaft zur Pilot-Charta

Betroffene Rollen	Welche Rollen in ML, Plattform, Rechenzentrum, Einkauf und Finanzen die Änderung betrifft
Ausgangslage Arbeitslast	Auslastung, Wartezeiten und Kosten je Trainingslauf oder je Million Tokens heute, gemessen vor dem Wechsel
Benötigte Kompetenzen	Betrieb eines neuen Chip- oder Cloud-Stacks, Portierung von Modellen, Kapazitäts- und Kostenplanung; Schulungsplan und verantwortliche Person
Beteiligung	Betriebsrat und Beschäftigtenvertretung, sofern vorhanden
Verantwortlich für die Aufsicht	Die namentlich benannte Person, verantwortlich für Zusagen zu Kapazität, Kosten und Energie
Eskalation	Wie Teams Engpässe, Ausfälle oder Kostenüberschreitungen melden und die Reaktionszeit
Kapazitätsplan	Wohin gewonnene Ingenieurszeit fließt: neue Modelle, Optimierung oder Qualifizierung; wer entscheidet
Gemessene Qualität	Auslastung, Kosten und Energie je Ergebnis sowie Vorfälle, erfasst zusammen mit der Ergebniskennzahl

BEST PRACTICE

Gewonnene Zeit zählen, nicht eingesparte

Gebuchte Kapazität ist noch kein realisierter Wert. Zählen Sie ihn erst, wenn die Rechenleistung nachweislich in genutzte Modelle, Produkte oder Dienste geflossen ist.

NÄCHSTE SCHRITTE

Ein 90-Tage-Start für Entscheidungen zur Rechenleistung.

01

Erfassen

Jede KI-Arbeitslast mit Rechenquelle, Region Vertragslaufzeit und monatlichen Kosten in einem Register erfassen.

02

Prognostizieren

Bedarf für Training und Inferenz für drei Jahre schätzen, mit einem niedrigen und einem hohen Fall.

03

Beteiligen

IT, Einkauf, Finanzen, Nachhaltigkeit, Recht und Betriebsrat früh einbinden.

04

Vergleichen

Mindestens zwei Chip- oder Cloud-Optionen mit eigenen Modellen auf Kosten, Tempo, Energie und Portabilität testen.

05

Vereinbaren

Kapazität, Region, Preisanpassung, Energieberichte und Ausstieg regeln, bevor Sie sich binden.

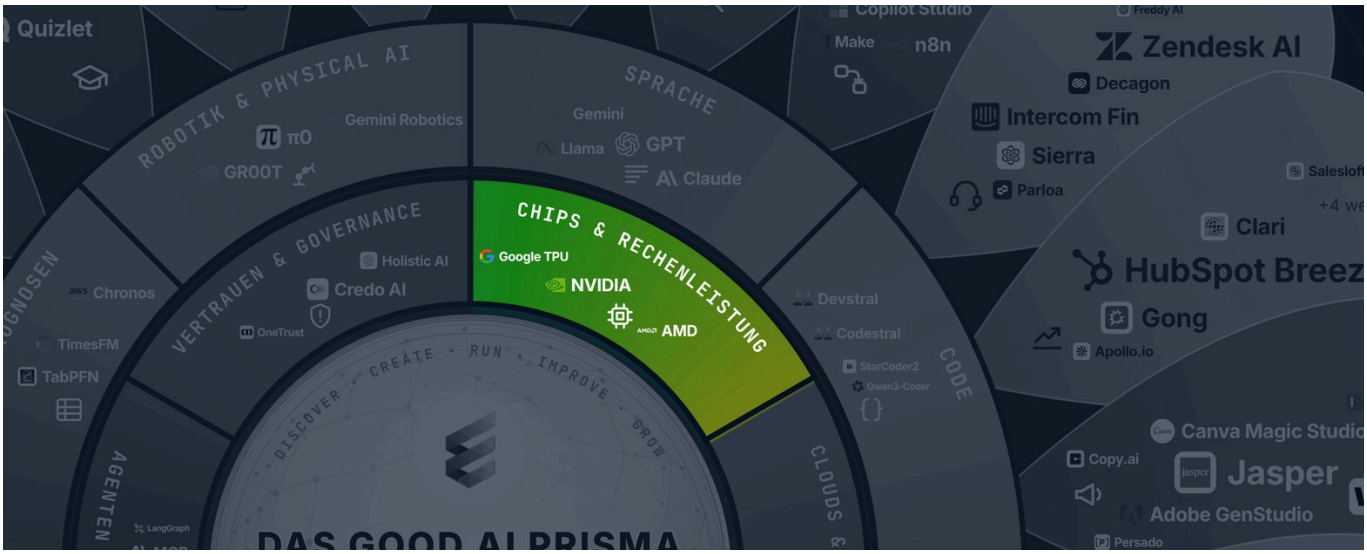
06

Entscheiden

Nur auf Basis gemessener Kosten je Ergebnis, Auslastung, Energie und Lieferung skalieren.

EMPFEHLUNG VON ETHORITY

Abgeleitet aus den Regeln und Befunden dieses Briefs.



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · Segment und Nachbarringe · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

SCHLUSSFOLGERUNGEN

Was die Evidenz ergibt.

Rechenleistung ist zu einem strategischen Rohstoff geworden. Die Frage ist nicht mehr, welcher Chip am schnellsten ist, sondern wie sich genug Kapazität am richtigen Ort zu planbaren Bedingungen sichern lässt. Fünf Schlussfolgerungen für Entscheidungen jetzt.

1 Kapazität planen, nicht nur Chips

EVIDENZ ASML und SK hynix rechnen mit Engpässen bis 2027; Speicher könnte bis 2030 knapp bleiben.

WAS ZU TUN IST Bedarf für drei Jahre prognostizieren und Kapazität früh sichern, mit Ausweichoptionen.

SIEHE · WO DER MARKT STEHT

2 Nach Bedingungen kaufen, nicht nach Listenpreis

EVIDENZ Eine H100-Stunde kostet bei Neoclouds rund ein Drittel der Hyperscaler-Preise; Speicherpreise stiegen in einem Quartal um 15–18 %.

WAS ZU TUN IST Gesamtkosten je Ergebnis vergleichen und Reservierung, Preisanpassung und Ausstieg verhandeln.

SIEHE · PREIS, STROM UND STANDORT

3 Energie in den Vertrag schreiben

EVIDENZ Rund 20 % der geplanten Rechenzentren drohen Verzögerungen beim Netzanschluss; das Energieeffizienzgesetz setzt Vorgaben.

WAS ZU TUN IST Stromquelle, Effizienzbasis und Regeln zur Weitergabe von Energiekosten schriftlich vereinbaren.

SIEHE · PREIS, STROM UND STANDORT

4 Den Standort bewusst wählen

EVIDENZ Neue Gigafabrik-Kapazität in der EU kommt ab 2028; US-Exportregeln bestimmen, wer welche Chips kaufen darf.

WAS ZU TUN IST Jede Arbeitslast einer Region und einem Anbieter zuordnen, die Residenz-, Souveränitäts- und Exportanforderungen erfüllen.

SIEHE · DIE REGELN

5 Software portabel halten

EVIDENZ Hyperscaler-Chips, AMD, Inferenz-Chips und Neoclouds erweitern die Auswahl; die Software ist der größte Wechselaufwand.

WAS ZU TUN IST Eigene Modelle auf mindestens zwei Plattformen testen, bevor Sie sich langfristig binden.

SIEHE · DIE MARKTKARTE

DAS PROBLEM

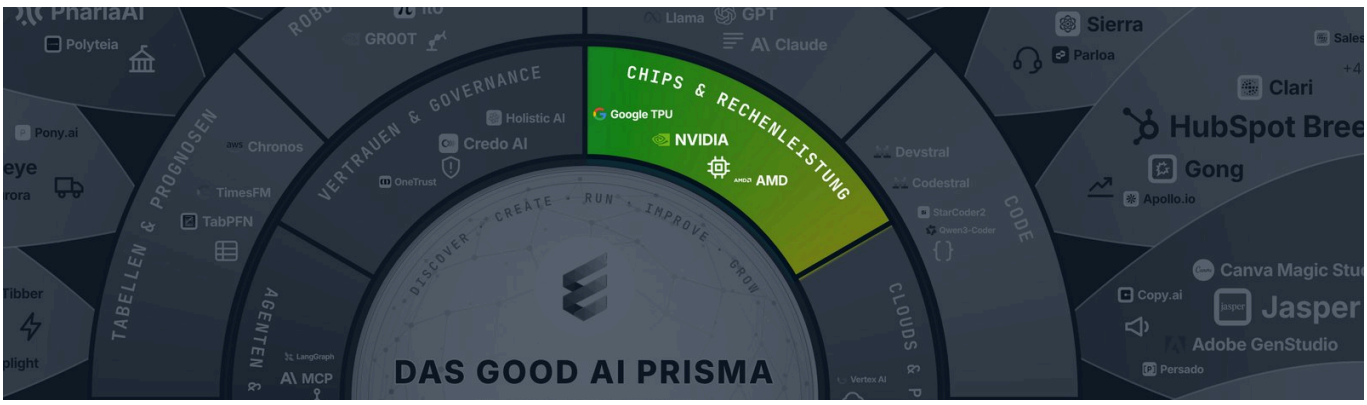
Der KI fehlen nicht Ideen. Es fehlt Rechenleistung, die sich planen, bezahlen und begründen lässt.

Unternehmen müssen genug Rechenleistung in der richtigen Region zu planbaren Preisen sichern, mit Liefer- und Energierisiken, die sie verstehen und gegenüber Vorstand und Aufsichtsrat begründen können.

DAS PROBLEM	WAS DIE EVIDENZ ZEIGT	WAS ES KOSTET
Knappe Kapazität	Modernste Chips, Speicher und Packaging werden zugeteilt, mit langen Lieferzeiten.	Verzögerte Projekte und wartende Teams.
Schwankende Kosten	GPU-Miete, Speicher- und Strompreise entwickeln sich unterschiedlich und sind schwer vergleichbar.	Budgetüberschreitungen und ungünstige Verträge.
Energie und Netz	Netzanschlüsse und Stromversorgung begrenzen, wo Kapazität entstehen kann.	Höhere Preise und verzögerte Standorte.
Regeln und Souveränität	Exportkontrollen, Energieregeln und Datenresidenz gelten gleichzeitig.	Eingeschränkte Optionen und Nacharbeit.
Abhängigkeit	Software, die an einen Chip oder eine Cloud gebunden ist, macht den Wechsel teuer.	Schwächere Verhandlungsposition.

Die Frage hinter jedem Projekt: Was soll besser werden, welche Optionen sind glaubwürdig, welche Evidenz fehlt, was verlangt die Umsetzung, und wer verantwortet das Ergebnis?

ETHORITY



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · innere Ringe und Kern · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

WIE ETHORITY HILFT

Vom Rechenbedarf zu Entscheidungen, die bestehen.

ETHORITY hilft Unternehmen, KI-Rechenleistung zu planen, einzukaufen und zu steuern: von unabhängiger Marktevidenz zu Kapazitätsentscheidungen, die vor Vorstand und Aufsichtsrat bestehen, und Ergebnissen, die sich messen lassen.

01
Discover
Chance finden und prüfen: Optionen, Evidenz, Rahmenbedingungen und Regeln.

02
Create
Lösung mit Ihren Daten, Systemen und Menschen gestalten und bauen.

03
Run
Verantwortlich betreiben, mit Verantwortlichen, Kontrollen und Support.

04
Improve
Am Ziel messen und nachsteuern.

05
Grow
Aus dem Gelernten neue Fähigkeiten und neuen Wert entwickeln.

Wo ein Projekt beginnen kann

KI-Compute-Case
Eine Arbeitslast, etwa Modelltraining, Feinabstimmung oder ein Inferenzdienst mit hohem Volumen, geprüft gegen Evidenz, Kosten, Energie und Regeln, mit klarer Entscheidung: bauen, mieten, umgestalten oder stoppen.

Anbieter- und Evidenzprüfung
Rechercheprofile und Vergleiche der 21 Angebote aus dem Good AI Prisma, mit strukturierter Prüfung von Aussagen zu Leistung, Verfügbarkeit und Preis.

KI-Bestandsaufnahme
Ein Register aller KI-Arbeitslasten und Rechenquellen, mit Verantwortlichen, Regionen, Kosten, Energieverbrauch und Überschneidungen.

Governance und Vertragsbereitschaft
Jede Arbeitslast zugeordnet zu Export-, Energie-, Datenresidenz- und AI-Act-Anforderungen, mit angemessenen Kontrollen und Vertragsklauseln.

KI-Kompetenz für Technologieteams
Rollenbasierte Qualifizierung für IT, Einkauf, Finanzen und Führung zu Optionen, Kosten und Risiken von Rechenleistung.

Vom Pilot in den Betrieb
Eine Pilot-Charta mit Ausgangsmessung, Schwelle und Abbruchbedingungen, danach Aufbau und Betrieb auf dem AI Business Operating System von ETHORITY.

MIT WEM WIR ARBEITEN

CIOs und CTOs, IT- und Cloud-Einkauf, Rechenzentrumsbetreiber, KI-Teams in Industrie, Forschung und öffentlicher Hand sowie Investoren in KI-Infrastruktur. Beginnen Sie mit einer Frage. Sprechen Sie mit uns: sten@ethority.ai · severin@ethority.ai

ÜBER ETHORITY

Vom Marktverständnis zur KI-Fähigkeit.

2009

das erste Social Media Prisma, der Beginn der Marktkarten von ETHORITY

ETHORITY, 21. SEP. 2009

HERKUNFT

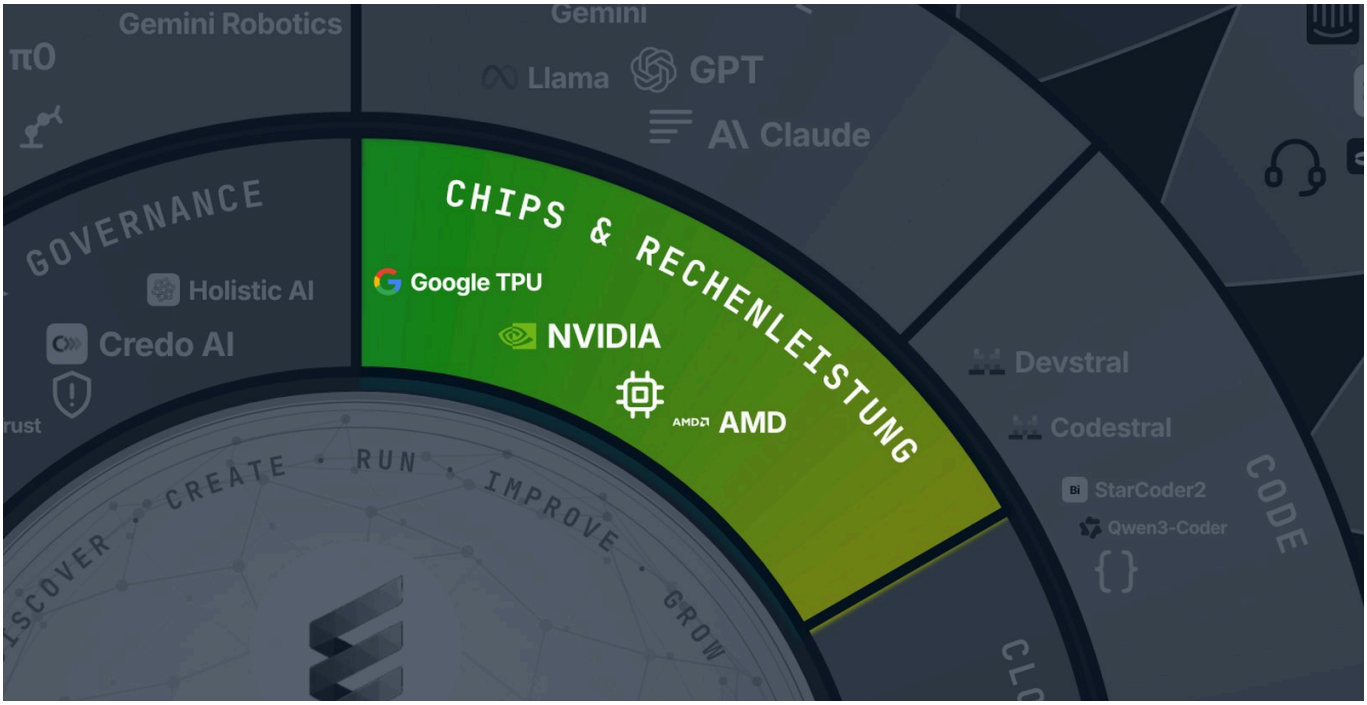
Das Good AI Prisma von Sten Franke & ETHORITY® · seit dem ersten Social Media Prisma 2009 · Märkte kartieren seit 1997 · Wurzeln: ETHORITY.AI und STEN.AI

ETHORITY verbindet Marktintelligenz, Forschung und praktische Umsetzung, damit Organisationen KI wirksam einsetzen. ETHORITY ist als AI1st-Organisation aufgebaut: Menschen und KI-Agenten arbeiten auf einem AI Business Operating System, mit Forschung, Engineering und Betrieb, um KI auch nach dem Pilot zu bauen und zu betreiben. Die Wurzeln liegen im Digital-Intelligence-Geschäft, das Sten Franke 2005 gegründet hat und das Monitoring, Forschung und Beratung verband, um Organisationen den Wandel ihrer Märkte verständlich zu machen.

Das Good AI Prisma ist der Einstieg: Rechercheprofile, dokumentierte Beziehungen und Vergleiche helfen, Optionen zu prüfen. Beratung, Governance und Transformation verbinden diese Optionen mit Zielen, Daten, Systemen, Menschen und Betriebsbedingungen einer Organisation. Mehr über ETHORITY, das Unternehmen.

AUSGEWÄHLTE REFERENZEN UND PARTNER SEIT 2005

Bosch · Bayer · BMW · Siemens · Mercedes-Benz · Coca-Cola · Deutsche Bank · T-Mobile · Telefónica o2 · DHL · EDEKA · AIDA Cruises. Ausgewählte historische Referenzen und Partner aus der Marktintelligenz- und Beratungstradition der Gruppe, in Projekten oder über frühere Gruppengesellschaften.



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · Logos im Segment · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

METHODE

Wie dieser Brief entstanden ist.

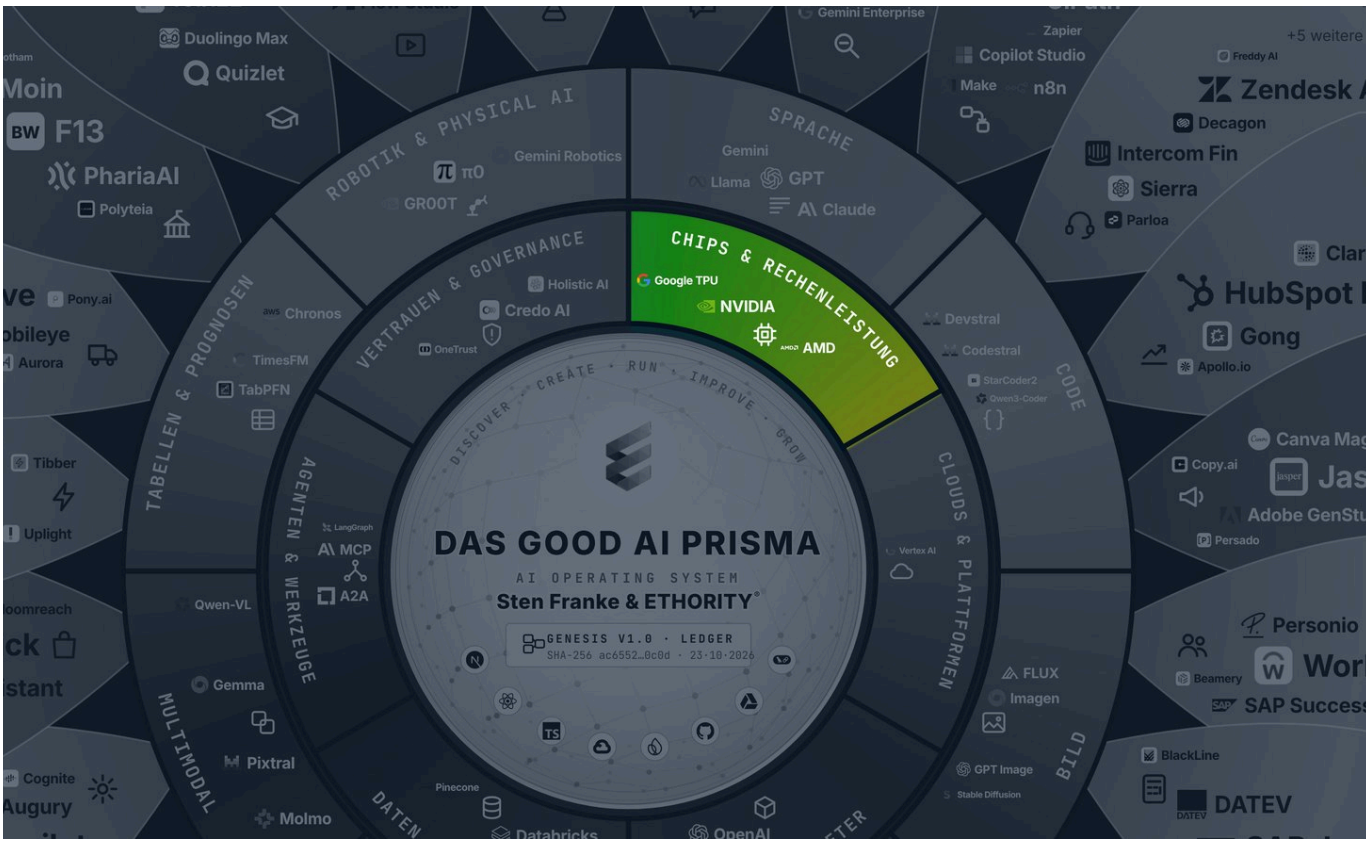
Marktzahlen: Datensatz der Good AI Prisma Studie 2026/27, Bereich Chips und Rechenleistung, 21 Einträge, jeder mit einer datierten Quelle von den Seiten des Anbieters erfasst (Oktober 2026). Aufgabengruppen und Regelbezug sind eine Einordnung von ETHORITY; Hinweise auf neuere Flaggschiffe haben den Stand Oktober 2026. Regeln und Marktdaten: Stand 22. Oktober 2026 aus Unternehmensberichten, IEA, EuroHPC, Bitkom und Politikanalysen; spätere Quartalszahlen sind nicht enthalten. Externe Zahlen wie zitiert und nicht kombiniert. Keine Rechts-, Anlage- oder Beschaffungsberatung. (siehe Quellen)

DATENSATZ

Datensatz der Good AI Prisma Studie 2026/27, Bereich Chips & Rechenleistung, 21 Einträge, Oktober 2026.

ABBILDUNGEN

Die Diagramme zeigen die in diesem Brief gedruckten Zahlen; jedes nennt seine Quelle unter der Grafik.



Good AI Prisma v3.12 · Segment Chips & Rechenleistung hervorgehoben · Segment und Kern des Prismas · Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.

QUELLEN

Wichtigste Quellen.

01 NVIDIA, financial results for Q2 fiscal 2027, 26 Aug 2026; NVIDIA GTC 2026 roadmap (DCD, Mar 2026)

02 TSMC, Q3 2026 results (Form 6-K), 15 Oct 2026; ASML, Q3 2026 results (Form 6-K), 14 Oct 2026

03 SK hynix CEO on memory supply (Reuters via Tom's Hardware), 11 Jul 2026; TrendForce on HBM share (Mar 2026) and DRAM contract prices (Jul 2026)

04 IEA, Energy and AI, 2025; IEA, Key Questions on Energy and AI, 16 Apr 2026

05 Bitkom, Rechenzentren in Deutschland, 10 Nov 2025; TOP500 (JUPITER), Nov 2025

06 EuroHPC JU, AI Gigafactories call, 30 Jul 2026; Deutsche Telekom Industrial AI Cloud (Telecoms.com, DCD, Nov 2025)

07 CNAS, H200 export policy, Jan 2026; Epoch AI, Stargate sites (Apr 2026) and inference price trends (Mar 2025); Silicon Data H100 index (Implicator, Jul 2026)

08 Microsoft, Maia 200, 26 Jan 2026; AWS Trainium3 (DCD, Dec 2025); Google Ironwood (TechWire Asia, Nov 2025); Jon Peddie Research on Intel, Oct 2026

09 UK Compute Roadmap (Jul 2025); Japan AI and chip package (Reuters, Nov 2024); IndiaAI Mission (Oct 2026); Stargate UAE (Oct 2025); Humain and Qualcomm (Oct 2025); Huawei Ascend roadmap (Caixin, Sep 2025)

10 Regulation (EU) 2024/1689 (AI Act) and 2026/1744 (Digital Omnibus); Delegated Regulation (EU) 2024/1364 (EED data centres); Energieeffizienzgesetz (EnEfG)

11 Good AI Prism Market Study 2026/27 and study dataset (ETHORITY, October 2026)

IMPRESSUM

Impressum, Lizenz und Herkunft

DEUTSCHE AUSGABE · DE · v1.1 · OKT. 2026

PUBLIKATION

Good AI Prisma Technologiebrief: Chips & Rechenleistung · 2026/27 · Deutsche Ausgabe · Deutsch · Version 1.1 · 25. Okt. 2026.

AUSGABEN

v1.0: 22. Okt. 2026. v1.1: 25. Okt. 2026, neues redaktionelles Design, Befund-zuerst neu gegliedert; Inhalte und Zahlen der v1.0. Frühere Ausgaben bleiben verfügbar.

URHEBER

Erstellt von Sten Franke mit einer Perspektive auf Menschen und Governance von Severin L. Wilson.

METHODE UND DATEN

Datensatz der Good AI Prisma Studie 2026/27, Bereich Chips & Rechenleistung, 21 Einträge, Oktober 2026. Externe Zahlen wie zitiert; siehe Quellen.

ZITIERVORSCHLAG

ETHORITY (2026). Good AI Prisma Technologiebrief: Chips & Rechenleistung. Deutsche Ausgabe, v1.1, Oktober 2026. ethority.ai/de/research/publications/compute-germany-de

HERAUSGEBER UND KONTAKT

ETHORITY · Hauptsitz San Francisco · Europäische Niederlassung: Hamburg, Deutschland · ethority.ai · sten@ethority.ai

LIZENZ DES GOOD-AI-PRISMA-ARTWORKS

Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC, von Sten Franke & ETHORITY®. Lizenziert unter CC BY-NC-ND 4.0 (creativecommons.org/licenses/by-nc-nd/4.0).

ETHORITY gestattet zusätzlich die unveränderte redaktionelle und Presse-Nutzung mit Quellenangabe. Kommerzielle Lizenzen: ETHORITY.

Daten, Software sowie Logos und Marken Dritter sind nicht umfasst.

TEXT UND ABBILDUNGEN

© 2026 ETHORITY & STEN.AI LLC. Abbildungen dürfen mit Quellenangabe zitiert werden. Die Lizenz oben gilt nur für das Prisma-Artwork, nicht für dieses Dokument als Ganzes.

MARKEN DRITTER

Anbieternamen und Logos im Prisma-Artwork und in der Marktkarte bleiben Marken ihrer Inhaber.

BERATUNG

Dieser Brief ist keine Rechtsberatung.

Das Good AI Prisma von Sten Franke & ETHORITY®

SEIT DEM ERSTEN SOCIAL MEDIA PRISMA 2009 · MÄRKTE KARTIEREN SEIT 1997 · WURZELN: ETHORITY.AI UND STEN.AI

ETHORITY®

AI BUSINESS OPERATING SYSTEM

VON DER GESCHÄFTSAMBITION ZU KI, DIE WIRKT

KI wirksam einsetzen, mit Evidenz.

ETHORITY hilft Organisationen, KI auszuwählen, zu steuern und ihre Wirkung nachzuweisen: von unabhängiger Marktevidenz zu Entscheidungen, die sich begründen lassen, und Ergebnissen, die sich messen lassen.

Sten Franke
Gründer, Principal Architect & Lead Creator
sten@ethority.ai
linkedin.com/in/stenfranke

Severin L. Wilson
Chief Business & Transformation Officer
severin@ethority.ai
linkedin.com/in/severinwilson

01 **KI wirksam einsetzen: Globale Good AI Prisma Marktstudie 2026/27**
Globale Ausgabe · DE und EN

02 **Good AI Prisma Branchenbriefe 2026/27**
Personal · Gesundheit · Finanzdienstleistung & Versicherung · Öffentlicher Sektor · Bildung · Energie · Mobilität & Logistik · DE und Global EN

03 **Good AI Prisma Technologiebriefe 2026/27** SIE LESEN
KI-Modelle · Chips & Rechenleistung · Robotik & physische KI · DE und Global EN

04 **KI für die Pflege in alternden Gesellschaften**
Studie 2026/27 · DE, EN, US- und Global-Ausgabe

05 **KI in Pharma, Biotech und Langlebigkeit**
Studie 2026/27 · DE und Global EN

06 **Das Social Media Prisma 2009–2025**
Historisches Papier · DE und EN

ethority.ai/de/research/publications/compute-germany-de

DEUTSCHE AUSGABE · DE · v1.1 · OKT. 2026

Good-AI-Prisma-Artwork © 2026 ETHORITY & STEN.AI LLC · CC BY-NC-ND 4.0 · Daten und Marken Dritter nicht umfasst.
© 2026 ETHORITY & STEN.AI LLC · Hauptsitz San Francisco · Europäische Niederlassung: Hamburg, Deutschland · ethority.ai