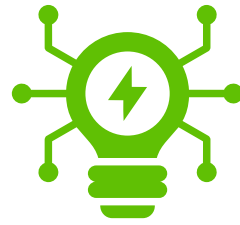


KI UND ENERGIE



Die große offene Frage einer neuen Technologie

KI verändert unsere Welt — aber zu welchem Preis? Das Problem der KI ist nicht nur ihre Intelligenz, mit der wir uns als Menschen neu arrangieren müssen, sondern ihre materielle Basis: Strom, Wasser, Rohstoffe und Rechenzentren. Während der Ausbau dieser Infrastruktur weltweit rasant voranschreitet, wächst auch die Frage nach ihren ökologischen und politischen Folgen. Wie lässt sich der steigende Energiebedarf von KI verantworten – und welche Ansätze gibt es, ihn wirksam zu begrenzen?

Was verbraucht KI? Vier Problemfelder – vom Training bis zur Kühlung

KERNDATEN ZUM ENERGIEVERBRAUCH VON RECHENZENTREN

415 TWh

Globaler Stromverbrauch
Rechenzentren 2024

945 TWh

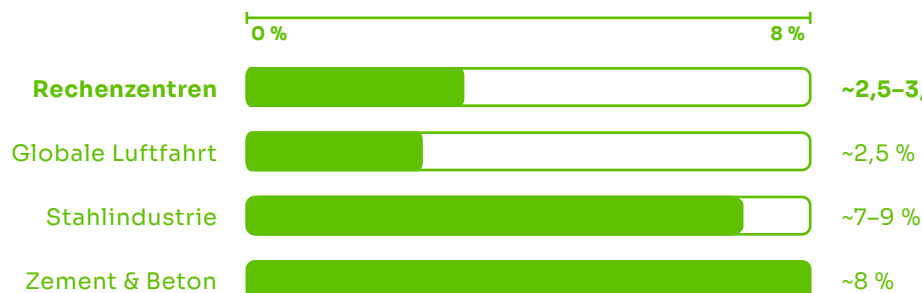
Prognose 2030 – entspricht dem
Verbrauch ganz Japans
IEA – International Energy Agency, 2025

2,5–3,7 %

Anteil Rechenzentren
am globalen CO₂-Ausstoß –
mehr als die Luftfahrt

CO₂-EMISSIONEN IM VERGLEICH

Anteil an globalen Treibhausgasen:

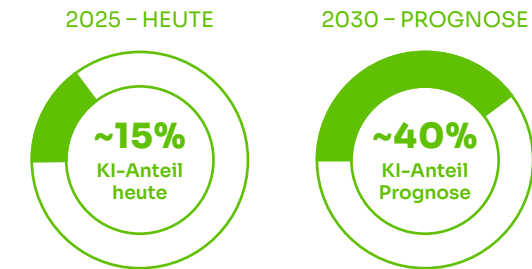


KI wächst 15 %/Jahr – andere Sektoren nur 2–4 %. Trend entscheidend.

Balkenbreite dient der Orientierung; Sektoren sind nicht direkt proportional vergleichbar.

WIE VIEL DAVON VERBRAUCHT KI?

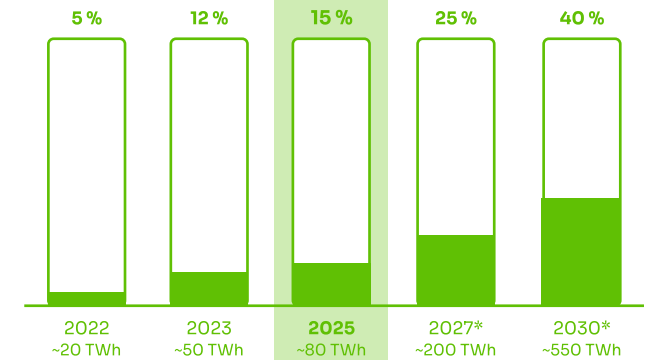
KI-Anteil am globalen Rechenzentrumsverbrauch



- KI-spezifische Rechenzentren (~15 %, heute):** Training großer Modelle, Inferenz (tägliche Nutzung), KI-Cloud-Dienste. Wächst 15 %/Jahr – viermal schneller als andere Sektoren.
- Klassische Rechenzentren (~85 %):** Cloud-Dienste, Streaming, E-Mail, Unternehmensdaten, Backups. Wachstum moderat; Effizienzgewinne teilweise sichtbar.

*Prognosen

Wachstum des KI-Anteils (grün) am Gesamtverbrauch aller Rechenzentren (weiß)



Prognose 2030: KI-Anteil steigt auf ~40 %. KI-Rechenzentren verbrauchen dann ~550 TWh, klassische ~850 TWh. Gesamt: ~1.400 TWh – das entspricht fast dem dreifachen deutschen Jahresstromverbrauch.

Quellen: Öko-Institut 2025 · Bitkom 2025 · IEA 2025 · Gartner 2024

DIE VIER HAUPTTREIBER DES ENERGIEVERBRAUCHS

1 TRAINING

Das einmalige Training großer Modelle

Wochen- oder monatelanger Hochleistungsbetrieb in Rechenzentren. Chat GPT-3-Training: ca. 1.300 MWh – Jahresverbrauch von 120 Haushalten. GPT-4: schätzungsweise 50.000 MWh. Der Energieverbrauch des Training der Chat GPT-5-Familie ist öffentlich nicht bekannt. Schätzungen bewegen sich für einzelne Frontier-Trainingsläufe im Bereich von einigen GWh bis deutlich über 100 GWh.

Kontext: Training ist die sichtbarste Phase – aber nicht die dominante. Einmal durchgeführt, bearbeitet ein Modell Milliarden von Anfragen. Der CO₂-Fußabdruck des Trainings sollte deshalb stets auf die Lebensdauer des Modells verteilt betrachtet werden.

3 HARDWARE & LIEFERKETTE

Chips, seltene Erden, Elektronikschrott

Chipproduktion ist extrem energieintensiv (+125 % von 2015–2023). Bis 2030 drohen durch KI-Ausbau 5 Mio. Tonnen zusätzlicher Elektronikschrott pro Jahr.

Kontext: Die vollständige Bilanzierung muss den Lebenszyklus umfassen: Rohstoffabbau → Chipfertigung → Betrieb → Entsorgung. Viele Unternehmen berichten nur den Betriebsstrom – das unterschätzt den Fußabdruck um bis zu 50 %.

2 INFERENZ

Jede einzelne Anfrage – in Billionen

Jede KI-Antwort kostet ca. 0,3–3 Wh (das, was ein Smartphone Display drei Minuten leuchten lässt, oder was es braucht, um eine E-Mail zu versenden und empfangen) – für sich genommen kaum spürbar. Multipliziert mit Milliarden täglicher Anfragen dominiert die Inferenz den Gesamtverbrauch: 60–70 % des KI-Energiefußabdrucks.

Kontext: Facebook verarbeitete 2022 täglich Billionen von Inferenzen. ChatGPTs monatlicher Inferenz CO₂-Fußabdruck entspricht ca. 260 transatlantischen Flügen. Bilder-KI verbraucht ca. 10× mehr als Textanfragen.

4 KÜHLUNG & WASSER

Der unsichtbare Ressourcenverbrauch

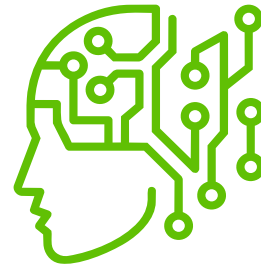
Rechenzentren kühlen mit Wasser. US-Datacenter verbrauchten 2023 bereits 64 Mrd. Liter – Prognose 2028: 257 Mrd. Liter. Schätzungsweise verbrauchen 20–50 ChatGPT-Anfragen ca. 0,5 Liter Wasser – vergleichbar mit dem Ansehen einer einzelnen Netflix-Episode.

Kontext: Wasser wird auf drei Ebenen verbraucht: bei der Chip-Produktion, indirekt durch fossile Kraftwerke und direkt zur Rechenzentren-Kühlung. Irland: Rechenzentren verbrauchen bereits mehr Strom als alle Wohngebäude zusammen.



OFFENE FRAGE DIESER SEITE: Wer ist verantwortlich für Transparenz? Unternehmen halten Verbrauchsdaten geheim. Ohne Messung gibt es keine Steuerung – und keine demokratische Debatte über die tatsächlichen Kosten der KI-Nutzung.

Wie wird KI effizienter? Sechs Lösungsfelder – von der Hardware bis ins Weltall



EFFIZIENZ-KERNDATEN

26 %

Jährliche Effizienzsteigerung
Standardserver 2017–2022

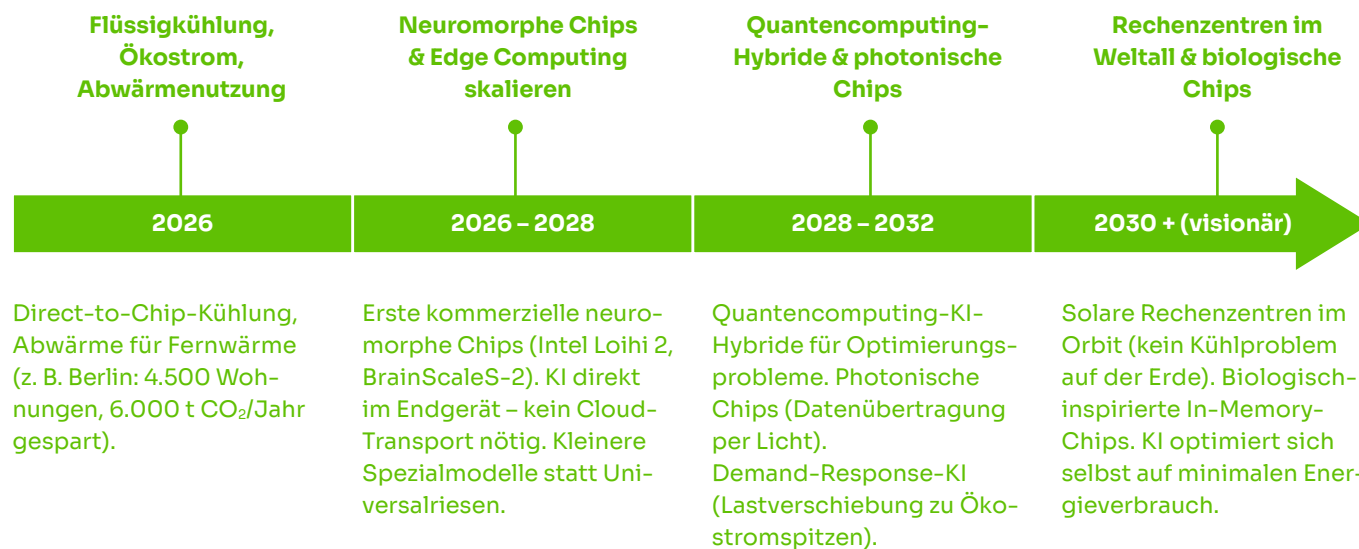
1.000 ×

Potenzielle Effizienz
neuromorpher Chips gegen-
über GPUs

11 ×

Wachstum KI-Stromverbrauch
2023–2030 ohne Gegenmaß-
nahmen

WANN KOMMT WAS – EINE ROADMAP



SECHS LÖSUNGSFELDER IM DETAIL

1 HARDWARE

Effizientere KI-Prozessoren und Chips

Neue GPU-Generationen verdoppeln Leistung/Watt ca. alle 2 Jahre. KI-spezifische ASICs (z. B. Googles TPUs) 10–100x effizienter als Standard-CPUs für spezifische Aufgaben.

Kontext: Das „Moore'sche Gesetz der Effizienz“: Nvidia, AMD, Google und Apple investieren massiv in spezialisierte KI-Chips. Kehrseite: Effizienz schafft Nachfrage – der „Jevons-Effekt“ könnte den Gesamtverbrauch trotzdem steigen lassen.



DAS JEVONS-PARADOXON – wenn Effizienz den Verbrauch erhöht

Der britische Ökonom William Stanley Jevons (1835–1882) beobachtete im viktorianischen England, dass effizientere Dampfmaschinen den Kohleverbrauch nicht senkten – sondern steigerten. Günstigere Nutzung schuf mehr Nachfrage. Dieses Paradox trägt bis heute seinen Namen – und ist aktueller denn je: Werden KI-Chips effizienter, wächst auch ihr Einsatz.

2 ARCHITEKTUR

Neuromorphe und photonische Chips

Chips, die das menschliche Gehirn nachahmen (Spiking Neural Networks), verbrauchen Energie nur bei aktiven Signalen – Potenzial: 1.000x effizienter. Intel Loihi 2 in Erprobung.

Kontext: Spiking Neural Networks feuern nur bei relevanten Ereignissen – wie echte Neuronen. Photonische Chips übertragen Daten mit Licht statt Elektronen. Großes Potenzial, aber noch nicht massentauglich.



3 MODELDESIGN

Kleinere, spezialisierte Modelle und Kompression



Statt eines Universalmodells: viele kleine Spezialmodelle für definierte Aufgaben. Modellkompression (Quantisierung, Pruning) reduziert Parameterzahl ohne große Leistungseinbußen. „Mixture of Experts“-Architekturen aktivieren bei jeder Anfrage nur einen Bruchteil aller Parameter.

Kontext: DeepSeeks sparsameres Trainingsverfahren zeigte: Ähnliche Leistung mit deutlich weniger Rechenaufwand ist möglich. Massives Einsparpotenzial – wenn die Branche es nutzt, statt immer größere Modelle zu bauen.

5 ENERGIE

Erneuerbare Quellen und smarte Lastverteilung



Training zu Zeiten hoher Solarproduktion verlagern senkt CO₂ ohne Mehraufwand. Google testet Demand-Response-Modelle. DE-Gesetz: Rechenzentren ab 2027 zu 100 % Ökostrom (bilanziell).

Kontext: Das Energieeffizienzgesetz (EnEfG) Deutschlands ist europäisches Vorbild. Abwärmennutzung: Ein geplantes Berliner Quartier soll 4.500 Wohnungen mit Rechenzentrumsabwärme heizen (6.000 t CO₂/Jahr gespart). Atomkraft als KI-Stromquelle ist kontrovers – Microsoft und Google investieren in kleine Modulreaktoren (SMRs).

4 INFRASTRUKTUR

Edge Computing und dezentrale KI



Verlagerung der Berechnung direkt auf Endgeräte (Smartphone, Auto, Sensor) – kein Datentransport zur Cloud nötig. Neuromorphe Chips ermöglichen KI auf Batterie in Echtzeit. Edge AI eliminiert den Übertragungsweg – oft 30–70 % der gesamten Systemenergie.

Kontext: Wachstumsmarkt: Globale Edge-Infrastruktur 2025 bei 261 Mrd. USD, bis 2028 auf 380 Mrd. USD. Zusätzlicher Vorteil: Daten bleiben auf dem Gerät – Datenschutz ohne Cloud-Abhängigkeit.

6 REGULIERUNG

Transparenz, Messung und Regulierung



Ohne Offenlegungspflichten keine Steuerung. EU-KI-Akt enthält erste Ansätze. Öko-Institut fordert verpflichtende Umweltkennzahlen für KI-Dienste – analog zu Energielabels für Haushaltsgeräte.

Kontext: Aktuell halten die meisten Unternehmen Verbrauchsdaten geheim. Der EU AI Act und das EnEfG sind erste Schritte – aber Durchsetzung und Vollständigkeit bleiben offen.

DIE GROSSE OFFENE FRAGE:



HOLEN EFFIZIENZ- GEWINNE DIE WACHSTUMSDYNAMIK JE EIN?

Jede Effizienzverbesserung erzeugt neue Nachfrage – das sogenannte Jevons-Paradox. Reicht technischer Fortschritt allein – oder braucht es politische Grenzen und gesellschaftliche Entscheidungen über die Nutzung von KI?



KI verbraucht heute so viel Energie wie ein mittelgroßes Industrieland – und die Kurve zeigt steil nach oben. Gleichzeitig gibt es echte technische Fortschritte: Chips werden effizienter, Modelle kompakter, Infrastrukturen smarter. Ob das reicht, hängt nicht allein von Ingenieurinnen und Ingenieuren ab – sondern auch davon, wie Gesellschaft, Politik und Wissenschaft die Nutzung von KI gestalten.

Welche Fragen stellst du dir?

Wo siehst du in deinem Fachbereich den größten Handlungsbedarf?