

Wann platzt die KI-Blase? Warum der 30. September 2027 zum Prüfdatum wird

KI ist real – offen ist die Rendite auf Chips, Rechenzentren, Strom und Netze. Der 30. September 2027 ist kein Crashtermin, sondern ein transparenter Stichtag, an dem sich die Infrastrukturwette prüfen lässt.

AUTOR: KI-MEINT.DE · KI-REDAKTION

STAND: 23. JULI 2026 UM 15:15 UHR

DIESER BEITRAG WURDE KI-GENERIERT.



Illustration: KI-meint.de · KI-generiert · Bildnachweis: ki-meint.de/bildrechte/#1777d92d-c6a6-48e3-9778-14ad1ad87947

THESE DER AUSGABE

Nicht die Nützlichkeit von KI entscheidet über eine mögliche Korrektur, sondern ob Kapazitätsausbau, Abschreibung, Stromkosten und konzentrierte Nachfrage bis 2027 tragfähige Renditen ergeben.

Hinweis zur Einordnung: Dieser Beitrag wurde vollständig mit künstlicher Intelligenz (KI) erstellt und anhand der angegebenen Quellen redaktionell geprüft. Trotzdem können Angaben fehlerhaft, unvollständig oder nicht mehr aktuell sein. Prüfen Sie wichtige Aussagen bitte selbst; eine Gewähr für Richtigkeit, Aktualität und Vollständigkeit übernehmen wir nicht. So arbeitet die KI-Redaktion →

KURZ EINGEORDNET

KI-Nutzen und Kapitalrendite sind verschiedene Fragen. Dieser Artikel zeigt, an welchen fünf Signalen sich bis zum 30. September 2027 erkennen lässt, ob die gebaute Infrastruktur ihre Kosten verdient.

Die zentrale These

Der 30. September 2027 ist kein vorhergesagter Crash-Tag. Er ist ein bewusst gewähltes Prüfdatum für die Wette auf KI-Infrastruktur.

Kurzthese: Generative KI schafft bereits in einzelnen Anwendungen messbaren Wert.^{[1][2]} Offen ist, ob die heute für Chips, Rechenzentren, Netze und Stromanschlüsse eingesetzten Summen rechtzeitig ausreichende und dauerhafte Erlöse erwirtschaften. Bis zum 30. September 2027 sollten Auslas-

tung, Preise, Abschreibungen und Finanzierung deutlich genug sichtbar sein, um diese Wette strenger zu beurteilen.

Die fünf unten definierten Signale sind kein Crash-Orakel, sondern ein Prüfmodell. Bleiben sie aus, wird die Überbau-These schwächer. Treten mehrere gleichzeitig auf, würde eine Korrektur wahrscheinlich zuerst Anbieter und Finanzierer der physischen KI-Infrastruktur treffen – und danach Stromprojekte, Bauketten, Private Credit, Teile des Softwaremarkts und große Aktienindizes.

Was hier mit „KI-Blase“ gemeint ist

Eine Blase ist nicht einfach ein hoher Aktienkurs. Sie entsteht, wenn Preise und Investitionen länger auf Erlöse setzen, die erst noch bewiesen werden müssen. Bei der generativen KI laufen drei Wetten gleichzeitig:

1. Die Leistungswette: Größere oder bessere Modelle lösen künftig mehr wirtschaftlich wertvolle Aufgaben.
2. Die Nutzungswette: Unternehmen bauen ihre Prozesse schnell genug um, damit aus einzelnen Hilfsmitteln wiederkehrende Budgets werden.
3. Die Kapitalwette: Die heute gebaute Rechen-, Netz- und Stromkapazität verdient vor ihrer technischen und ökonomischen Entwertung ausreichend Geld.

Die erste Wette ist zum Teil schon gewonnen: Für einzelne Anwendungen zeigen Feldstudien messbare Produktivitätsgewinne.[1][2] Der zweite Punkt ist offen, aber nicht leer. Der dritte Punkt ist der gefährliche. Selbst ein nützliches Produkt kann für seine Kapitalgeber enttäuschend sein, wenn zu viele Anbieter gleichzeitig dieselbe Infrastruktur bauen oder wenn Preise schneller fallen als die Auslastung steigt.

Das unterscheidet die heutige Lage von der populären Rede über eine „KI-Blase“. Der Kern ist nicht die Frage, ob ein Sprachmodell denken kann. Der Kern ist ein Investitionszyklus in langlebige Ge-

bäude, kurzfristiger veraltende Server, knappe Stromanschlüsse und hoch konzentrierte Lieferketten. Die Risiken sitzen in Bilanzen und Verträgen.

Der Befund zuerst: Der Ausbau ist real, die Erlösdeckung noch nicht vollständig sichtbar

Die Größenordnung lässt keinen Zweifel daran, dass hier mehr als ein Softwaretrend läuft. Die Internationale Energieagentur beziffert die Investitionen in Rechenzentren im Jahr 2024 auf rund 500 Milliarden US-Dollar; seit 2022 hätten sie sich nahezu verdoppelt. Diese Zahl betrifft nicht nur KI. Sie misst aber den gemeinsamen Infrastrukturzyklus, dessen stärkster Treiber KI-Rechenlast ist.[3]

Die Firmenmeldungen passen dazu. Alphabet plant für 2026 Investitionen von 175 bis 185 Milliarden Dollar; etwa 60 Prozent der Investitionen des Jahres 2025 entfielen auf Server, der Rest überwiegend auf Rechenzentren und Netze.[4] Microsoft erwartet für das Kalenderjahr 2026 etwa 190 Milliarden Dollar Investitionsausgaben und weist zugleich auf höhere Komponentenpreise.[5] Amazon kündigte rund 200 Milliarden Dollar für 2026 an. Der freie Cashflow der vergangenen zwölf Monate fiel 2025 gegenüber dem Vorjahr deutlich, vor allem wegen zusätzlicher Käufe von Sachanlagen für KI.[6] Meta gab 2025 einschließlich Finanzierungsleasing 72,22 Milliarden Dollar aus und stellte für 2026 einen nochmals größeren Zuwachs in Aussicht.[7]

Das ist nicht automatisch irrational. Die vier Konzerne erwirtschaften hohe operative Cashflows und können einen Teil des Ausbaus aus dem laufenden Geschäft zahlen.[4][5][6][7] Die entscheidende Frage lautet deshalb nicht: „Sind die Hyperscaler überschuldet?“ Sie lautet: Wird die Grenzrendite der nächsten hundert Milliarden Dollar noch hoch genug sein, wenn die Kapazität ans Netz geht?

Eine zweite Beobachtung schützt vor Alarmismus. Rechenleistung ist nicht bloß Vorrat. Microsoft meldete für das Geschäftsjahr 2025 ein Wachstum des Segments Intelligent Cloud um 21 Prozent; Azure und verwandte Cloud-Dienste wuchsen um 34 Prozent. Gleichzeitig sank die Cloud-Bruttomarge auf 69 Prozent, ausdrücklich wegen des Ausbaus der KI-Infrastruktur.[8] Das ist das richtige Bild der Lage: Nachfrage und Umsatz wachsen, aber die Infrastruktur drückt bereits auf die Marge. Der Markt bezahlt damit heute für einen Ertrag, der erst künftig dauerhaft höher sein muss als Kapital-, Strom- und Abschreibungskosten.

Die einfachste Rechnung ist die härteste

Die KI-Debatte verdeckt häufig eine banale ökonomische Tatsache. Ein Rechenzentrum erzeugt nicht deshalb Wert, weil es gebaut wurde. Es erzeugt Wert, wenn Kunden seine Rechenzeit über Jahre zu Preisen kaufen, die nach Strom, Kühlung, Netz, Wartung, Personal, Fremdkapital und Abschreibung noch eine ausreichende Rendite lassen.

Man kann die Prüfung auf eine vereinfachte Gleichung verdichten:

$$\begin{array}{rcl} \text{Erforderlicher} & \text{Jahresumsatz} & = \\ \text{Kapitalkosten + Abschreibung + Strom} & & \\ \text{und Betrieb + Zielmenge.} & & \end{array}$$

Das klingt trivial, ist aber die Stelle, an der sich die Prognosen unterscheiden. Anbieter sprechen oft über „Tokens“, Nutzerzahlen, Modellparameter oder reservierte Grafikprozessoren. Für die Kapitalseite zählen andere Größen: bezahlte Auslastung, Preis je Recheneinheit, Vertragslaufzeit, Wiederverkaufswert der Hardware und die Geschwindigkeit, mit der ein neuer Chip die alte Generation entwertet.

Die technische Seite verschärft die Rechnung. Gebäude und Stromanschlüsse können Jahrzehnte nutzbar sein. KI-Server nicht. Meta verlängerte 2025 die geschätzte Nutzungsdauer bestimmter

Server- und Netzwerkanlagen auf 5,5 Jahre. Allein für bereits Ende 2024 im Betrieb befindliche Anlagen sollte diese Bilanzierungsänderung die Abschreibung des Jahres 2025 um etwa 2,9 Milliarden Dollar senken.[9] Das ist kein Vorwurf an Meta und keine Behauptung falscher Bilanzierung. Es zeigt aber, wie stark der ausgewiesene Gewinn von Annahmen über die ökonomische Lebensdauer abhängt. Je schneller Hardware, Architektur oder Energieeffizienz voranschreiten, desto anspruchsvoller wird diese Annahme.

Die Gegenseite lautet: Effizienz kann die Nachfrage erhöhen. Wenn ein Modell pro Aufgabe billiger wird, entstehen neue Anwendungen. Die IEA zeigt den Effekt in ihrem Hoch-Effizienz-Szenario: Dort liegt der Strombedarf von Rechenzentren im Jahr 2035 um 20 Prozent unter dem Basisszenario.[10] Preisverfall zerstört also nicht zwingend den Markt. Er verschiebt ihn. Für Investoren ist das nur dann gut, wenn die neue Nachfrage schneller wächst als die Erlöse pro Recheneinheit fallen.

Genau hier liegt das Problem eines Infrastrukturbooms. Die Nachfrage kann elastisch sein, die Kapazität kommt aber in Schüben. Stromanschlüsse, Grundstücke, Transformatoren und Gebäude werden lange geplant.[11] Wenn sie fertig sind, können viele Projekte nahezu gleichzeitig Kapazität freisetzen. Dann wird aus der Frage „Gibt es Nachfrage?“ die schärfere Frage „Gibt es genug zahlende Nachfrage zu einem Preis, der die gesamte Kette trägt?“

Warum die Lage an den Rändern fragiler ist als im Zentrum

Die großen Plattformen haben Cashflows, Kundenbeziehungen und mehrere Erlösquellen.[4][5][6][7] Ihr Risiko ist eher eine Renditeenttäuschung als eine unmittelbare Zahlungsunfähigkeit. Fragiler sind die Bereiche, die ihre Investitionsentscheidung an die Bestellungen dieser Plattformen hängen und selbst wenig Puffer haben.[12]

Der IWF beschreibt die Kette inzwischen bemerkenswert offen. In seinem Finanzstabilitätsbericht vom April 2026 warnt er, die „AI value stack“ könne bei nachlassender Erwartung zugleich bei Bauunternehmen, Energieversorgern, Chip- und Softwareentwicklern, Hyperscalern und nachgelagerten Betreibern an Umsatz und Unternehmenswert verlieren.[12] Der Bericht unterscheidet dabei ausdrücklich: Hyperscaler und Chipentwickler hätten bislang solide finanziert; „Energizers“ und Betreiber seien wegen des raschen Ausbaus bilanziell empfindlicher.[12] Seit 2025 hätten zirkuläre Finanzierungsarrangements zugenommen, bei denen KI-Entwickler, Chiphersteller und große Technologiefirmen gleichzeitig Kunden, Investoren und Finanzierer seien.[12]

„Zirkulär“ heißt nicht automatisch fingiert. Ein Chiphersteller kann sinnvoll in einen Cloud-Anbieter investieren, der wiederum seine Chips kauft. Ein Modellanbieter kann langfristige Rechenkapazität bei einem Cloud-Anbieter buchen, der damit den Bau weiterer Infrastruktur rechtfertigt. Ökonomisch entsteht trotzdem ein Risiko: Derselbe erwartete Endkunde kann in mehreren Bilanzen als Nachfrage, Vertragswert, Finanzierungssignal und Bewertungsgrundlage auftauchen. Fällt seine Zahlungsbereitschaft oder sein Finanzierungszugang weg, trifft der Rückschlag mehrere Stationen zugleich.[12]

Nvidia zeigt, wie konzentriert der Zyklus ist. Laut 10-K für das Geschäftsjahr 2026 entfielen 22 Prozent des Umsatzes auf einen direkten Kunden und weitere 14 Prozent auf einen zweiten. Nvidia weist selbst darauf hin, dass ein bedeutender Teil des Geschäfts direkt oder indirekt von wenigen Partnern und Abnehmern abhängt; die Verfügbarkeit von Rechenzentren, Energie und Kapital bei Kunden sei für die eigene Entwicklung entscheidend. [13] Die Firma ist nicht nur Chipproduzent. Sie ist ein Hebel auf die Investitionsbereitschaft einer kleinen Zahl sehr großer Käufer.

Auch der Aktienmarkt bündelt diesen Hebel. Der IWF beschreibt die Dominanz weniger großer Unternehmen in der Marktbewertung als histori-

schen Extrembereich.[14] Das ist keine Aussage, dass diese Unternehmen sämtlich „KI-Aktien“ seien. Es bedeutet aber: Eine enttäuschte KI-Erwartung bei wenigen Mega-Caps würde weit über spezialisierte Technologiefonds hinausgehen. Indexfonds, Altersvorsorgeportfolios und internationale Anleger tragen diese Konzentration mit.

Was der bisherige Verlauf wirklich beweist – und was nicht

Ein häufiger Denkfehler lautet: Weil Unternehmen KI bereits einsetzen, seien alle Infrastrukturanahmen bestätigt. Die Daten sprechen für eine schnellere Verbreitung, nicht für eine bereits gesicherte gesamtwirtschaftliche Rendite.

Das U.S. Census Bureau meldete für Dezember 2025 bis Mai 2026, dass 17 bis 20 Prozent der US-Unternehmen KI in ihren Geschäftsprozessen einsetzten; bei Unternehmen mit mindestens 250 Beschäftigten waren es 37 Prozent. Zugleich blieb die Quote in vielen Branchen über sechs Monate weitgehend stabil. Im Handel lagen Nutzung und erwartete Nutzung unter dem nationalen Durchschnitt.[15] Die breite Masse ist also nicht bei null. Aber der Sprung vom Werkzeuggebrauch zur tiefen Prozessumstellung ist keineswegs vollzogen.

Die methodisch präzisere Census-Auswertung kommt für November 2025 bis Januar 2026 auf 18 Prozent der Firmen mit KI in mindestens einer Geschäftsfunktion; beschäftigungsgewichtet sind es 32 Prozent. Große, wissensintensive Unternehmen liegen weit darüber.[16] Daraus folgt zweierlei. Erstens kann die Nutzungsbasis rasch wachsen, weil große Firmen einen hohen Teil der Wirtschaftsleistung ausmachen. Zweitens ist eine gleichmäßige, breite Nachfrage, auf der jede neue GPU-Farm rechnen könnte, noch nicht belegt.

Die Produktivitätsforschung liefert dasselbe Doppelbild. Ein vorregistriertes Feldexperiment bei Procter & Gamble mit 776 Fachkräften fand deutliche Leistungsgewinne; Einzelpersonen mit KI erreichten die Leistung von Teams ohne KI.[1] Eine

andere Feldstudie in 66 Firmen und mit 7.137 Wissensarbeitern ermittelte bei tatsächlichen Nutzern eines integrierten Werkzeugs rund zwei Stunden weniger E-Mail-Zeit pro Woche. Die Autoren fanden daneben keine nachweisbare Verschiebung der Menge oder Zusammensetzung von Arbeitsaufgaben durch die individuelle Bereitstellung.[2]

Das ist keine Widerlegung des Produktivitätspotenzials. Es ist ein Warnhinweis gegen eine falsche Multiplikation: Einsparungen in einer Teilaufgabe werden nicht automatisch zu zusätzlichem Umsatz, weniger Personal oder höheren Margen im gesamten Unternehmen. Dazwischen liegen Datenqualität, Haftung, IT-Integration, Schulung, Prozessänderung und die Frage, ob Kunden den besseren Dienst bezahlen.

Die OECD benennt genau diese Lücke. Ihr Bericht zur KI-Einführung in Unternehmen behandelt nicht nur Technik, sondern Daten, Kompetenzen, Organisationsgestaltung, Finanzierung und Vertrauen als Bedingungen der Diffusion.[17] Das ist für den Blasenbegriff zentral. Eine Technik kann sich an der Oberfläche extrem schnell verbreiten und doch länger brauchen, bis sie den Kapitalstock einer Volkswirtschaft produktiver macht.

Die Gegenposition verdient mehr als einen Absatz

Die stärkste Gegenposition lautet: Wir erleben keine Blase, sondern den frühen Bau einer allgemeinen Technologie. Viele frühere Infrastrukturbooms wirkten in ihrer Bauphase übertrieben und waren später trotzdem die Grundlage produktiver Wirtschaft. Eisenbahn, Elektrizität, Mobilfunk, Cloud und Glasfaser lieferten unterschiedliche Mischungen aus Überbau, Pleiten und bleibendem Nutzen. Der Fehlschluss wäre, aus einer Korrektur einzelner Finanzierungen auf das Scheitern der Technologie zu schließen.

Für diese Gegenposition gibt es harte Fakten. Die IEA sieht den Stromverbrauch von Rechenzentren im Basisszenario von 415 Terawattstunden 2024

auf rund 945 Terawattstunden 2030 steigen.[18] Das ist eine Projektion, keine Messung künftiger Erlöse. Sie baut aber auf realen Bauvorhaben, bereits steigender Nutzung und dem Bedarf an Rechenleistung auf.

Auch die Anwendungsschicht bleibt nicht stehen. Anthropic wertet reale, datensparsam aggregierte Nutzung aus und findet eine rasche, aber geografisch und nach Aufgaben ungleiche Diffusion. Gerade diese Ungleichheit kann ein Frühstadium sein, nicht das Ende der Nachfrage.[19] Wenn Agentensysteme verlässlich in Unternehmenssoftware arbeiten, können sie Rechenlast und Zahlungsbereitschaft deutlich stärker erhöhen als die heutigen Schreib- und Suchhilfen.

Hinzu kommt ein Marktmechanismus, den Untergangserzählungen oft übersehen. Ein billigeres Modell reduziert die Kosten für den Nutzer. Das schwächt zwar die Preissetzung eines einzelnen Modellanbieters, es kann aber die Gesamtmenge der ausgeführten Aufgaben stark vergrößern. Die Reaktion des Marktes auf DeepSeek im Januar 2025 machte diese Ambivalenz sichtbar. Nach der Ankündigung eines potenziell kostengünstigeren Modells fiel der IT-Sektor des S&P 500 nach Darstellung des IWF um mehr als fünf Prozent. Der IWF hielt zugleich fest, günstigere offene Modelle könnten die Nutzung ausweiten und den Übergang von Infrastruktur zu Software und Anwendung beschleunigen.[20]

Ein Szenario ohne großen Knall ist deshalb plausibel. Dann würden nicht Rechenzentren leer stehen, sondern die Renditen würden normalisieren. Einige Chiplieferanten und spezialisierte Betreiber verlören Bewertungsprämien; die großen Plattformen würden ihre Ausgaben langsamer steigern und die Infrastruktur über mehr Jahre monetarisieren. Die Volkswirtschaft behielte die Technik. Anleger verlören die Illusion, dass jeder Teilnehmer des Stacks dieselbe außergewöhnliche Marge verdienen kann.

Warum dennoch gerade 2027 der Prüfzeitraum ist

Die Jahreszahl folgt nicht aus Astrologie. Sie folgt aus der Verzögerung zwischen Bestellung, Inbetriebnahme und Ertragsrechnung.

Erstens: Der Baukörper kommt spät. Stromanschlüsse, Umspannwerke, Transformatoren, Genehmigungen, Kühlung und Gebäude haben längere Vorläufe als ein Modellrelease. Die IEA weist auf die außergewöhnliche lokale Konzentration hin: Fast die Hälfte der US-Rechenzentrumskapazität liegt in fünf regionalen Clustern. Ein typisches KI-orientiertes Rechenzentrum verbrauche so viel Strom wie 100.000 Haushalte; die größten im Bau befindlichen Anlagen das Zwanzigfache.[11] Das macht Strom nicht zu einem Nebenkostenposten, sondern zu einer echten Angebotsgrenze.

Zweitens: Der Hardwarezyklus kommt früh. Server und Netzwerkbestandteile werden im Vergleich zu Gebäuden schnell ersetzt. Wer heute sehr teure Beschleuniger installiert, muss zeigen, dass sie vor dem nächsten Generationssprung genügend Deckungsbeitrag liefern. Die technischen Verbesserungen sind für die Gesellschaft wünschenswert, für die Restwerte älterer Kapazität aber nicht neutral.

Drittens: Ausgaben werden zunächst als Zuversicht gelesen, dann als Abschreibung. Im ersten Jahr belohnt der Markt oft den Ausbau, weil er künftige Nachfrage signalisiert. Mit Zeitversatz erscheinen höhere Abschreibungen, Energiekosten und Leasinglasten in den Gewinn- und Verlustrechnungen. Alphabet sagte selbst, dass seine Abschreibungen 2025 um fast sechs Milliarden Dollar oder 38 Prozent stiegen und 2026 deutlich stärker wachsen dürften.[4] Die Frage ist dann nicht mehr, ob der Konzern Kapazität bauen kann, sondern ob das zusätzliche Ergebnis schneller wächst als die laufende Belastung.

Viertens: Verträge müssen in Nutzung übergehen. Ein Kapazitätsvertrag kann eine starke Nachfrageanzeige sein. Er ist aber etwas anderes als wiederkehrender, breit diversifizierter Endkundenum-

satz. Microsoft betont, die Verschiebung zu GPU- und CPU-Investitionen korreliere stärker mit konkreten Kundenverträgen.[21] Das ist positiv. Es hebt aber nicht den Unterschied zwischen einem großen Ankervertrag und einer dauerhaft breiten Zahlungsbasis auf.

Die Jahre 2025 und 2026 sind in diesem Artikel Bau- und Bestelljahre. 2027 ist deshalb ein sinnvoller Prüfzeitraum: Bis dahin sollte mehr Kapazität produktiv sein; zugleich erscheinen Abschreibung, Energie- und Finanzierungskosten über mehrere Berichtsperioden. Der 30. September 2027 ist ein redaktioneller Szenariostichtag, keine Vorhersage eines Einbruchs. Nach drei Quartalen liegen genug neue Unternehmensprognosen vor, um Nachfrage, Kosten und Energieverfügbarkeit nicht an einem einzelnen Quartalsbericht oder einem einzelnen Chipanbieter zu messen.

Ein prüfbares Warnsystem bis zum 30. September 2027

Die These braucht Kriterien. Von einem Blasenbruch im engeren Sinn würde ich erst sprechen, wenn mindestens drei der folgenden fünf Signale zeitgleich auftreten. Ein einzelnes Signal kann schlicht Wettbewerb oder eine normale Konjunkturdelle bedeuten.

1. Ausgaben kehren sich breit um. Mindestens zwei große Hyperscaler senken ihre Investitionsprognose für KI-nahe Kapazität deutlich, und sie begründen dies mit schwächerer Auslastung oder schwächerer Zahlungsbereitschaft, nicht mit effizienterer Beschaffung allein.
2. Die Marge fällt trotz Umsatzwachstum. Mehrere Cloud-Anbieter melden anhaltenden Margendruck, bei dem Abschreibung, Strom oder gemietete Rechenleistung schneller wachsen als die KI-nahen Erlöse.
3. Die Preislogik bricht. Preise für GPU-Cloud, Inferenz oder Rechenzentrumskapazität sinken stärker als die Kosten je berechneter Aufgabe,

während die Auslastung nicht entsprechend steigt. Das wäre ein Überangebotssignal, nicht bloß der normale Effizienzfortschritt.

4. Finanzierung wechselt die Richtung. Betreiber, Strom- oder Bauprojekte verschieben Vorhaben, weil Fremdkapital teurer wird, Sicherheiten nachgefordert werden oder Ankerkunden nicht mehr dieselben Abnahmeverpflichtungen übernehmen.
5. Konzentration wird zur Korrektur. Chip- und Infrastrukturunternehmen reduzieren Prognosen oder bauen Bestände ab, weil wenige große Kunden ihre Bestellungen verschieben. Nvidias offengelegte Kundenkonzentration macht dieses Signal besonders aussagekräftig.[13]

Diese Kriterien sind bewusst bilanziell und vertraglich. Schlagzeilen über Modellfehler, neue Benchmarks oder einen einzelnen Rückgang einer Aktie reichen nicht. Auch eine Produktivitätsstudie, die einen positiven Effekt zeigt, widerlegt keinen Infrastrukturzyklus. Die Frage ist enger: Erwirtschaftet der real installierte Kapitalstock seine erwartete Rendite?

Das Gegenkriterium ist ebenso wichtig. Der Stichtag wäre klar gegen die Blasenthese zu lesen, wenn die Hyperscaler ihre Ausgaben halten oder erhöhen, zugleich Auslastung und margenstarke KI-Umsätze nachvollziehbar wachsen und Betreiber ihre Finanzierung zu stabilen Konditionen verlängern. Dann hätte die Nachfrage die Ausbauwelle schneller absorbiert als skeptische Szenarien annehmen.

Was eine Korrektur mitreißen würde

Zuerst die Teile des Stacks ohne eigenen Endkundenpuffer

Am Anfang stünden wahrscheinlich nicht Microsoft, Alphabet oder Amazon als operative Krisenfälle. Sie könnten Ausgaben verschieben, eigene Kapazität anders nutzen und Verluste in anderen Sparten auffangen.[4][5][6] Stärker gefährdet wären Neoclouds, GPU-Vermieter, spezialisierte Rechenzentrumsbetreiber und Zulieferer, die wenige

Großkunden haben und ihre Projekte mit hohen Fixkosten oder Fremdkapital finanzieren.[12] Ihr Geschäftsmodell ist eine Wette darauf, dass bestellte Rechenleistung auch nach Ende der ersten Vertragsperiode knapp und teuer bleibt.

Das gleiche gilt für die Bau- und Energieebene, allerdings differenziert. Eine unterzeichnete Anschlusszusage und ein belastbarer Abnahmevertrag sind etwas anderes als ein Projekt, das auf einer Pressemitteilung über einen künftigen Campus basiert. In einer Korrektur würden Projekte mit gesicherter Stromversorgung und bonitätsstarkem Kunden eher überleben. Projekte mit unklarer Netzanschlusslage, kurzer Vertragsbindung oder spekulativer Finanzierung verlören zuerst Zugang zu Kapital.

Dann die Chip- und Netzlieferanten

Bei Halbleitern kann eine kleine Änderung der Bestellpläne große Wirkungen haben. Die Lieferkette ist kapitalintensiv, die Produkte haben kurze Innovationszyklen, und die Absatzseite ist konzentriert. Das bedeutet nicht, dass ein führender Anbieter verschwindet. Es bedeutet, dass Anleger ihre Annahme einer dauerhaft außergewöhnlichen Wachstumsrate neu bewerten würden. Der IWF verweist darauf, dass schon die DeepSeek-Nachricht 2025 die Nachhaltigkeit der Infrastrukturinvestitionen in Frage stellte und eine breite, wenn auch begrenzte Korrektur im IT-Sektor auslöste.[20]

Ein Rückgang träfe zudem Netzwerktechnik, Speicher, Kühlung, Stromtechnik und Hersteller von Rechenzentrumsmodulen. Die volkswirtschaftliche Wirkung wäre regional sehr ungleich. Dort, wo mehrere neue Rechenzentren gleichzeitig Bau, Stromprojekte und Gewerbesteuererwartungen tragen, kann ein verschobenes Projekt eine größere lokale Lücke reißen als der globale Index vermuten lässt.

Private Credit und die falsche Erzählung vom neuen Subprime

Private Credit ist ein echter Übertragungskanal, aber die Analogie zur US-Hypothekenkrise 2008 wäre falsch. Die Gefahr läge nicht in einem einzel-

nen KI-Kredit, sondern in einer gleichzeitigen Neubewertung von Softwareumsätzen, Betreiberprojekten und Sicherheiten. Gerade private Kredite werden seltener und weniger transparent bepreist als börsennotierte Anleihen. Das kann Verluste später sichtbar machen und eine restriktive Finanzierung auslösen.[22]

Das ist kein Beleg für einen bevorstehenden Kollaps. Es beschreibt einen Pufferverlust. Wenn KI bestimmte Softwarefirmen unter Preisdruck setzt, können Cashflows und Kreditqualität gleichzeitig schlechter aussehen. Weil Private Credit oft weniger transparent und weniger täglich bepreist ist als Anleihen, erscheint eine Neubewertung häufig spät und dann gebündelt. Die Federal Reserve nennt steigende Rückgabewünsche bei einzelnen nicht börsennotierten BDCs bereits als Finanzierungsrisiko.[22]

Die entscheidende Unterscheidung lautet: Eine Korrektur der KI-Infrastruktur wäre nicht zwingend eine Bankenkrise. Sie kann aber über Fonds, BDCs, Leasinggesellschaften und private Kreditgeber die Finanzierung neuer Projekte abrupt verteuern.[22] Das genügt, um den Bauzyklus zu drehen.

Der breite Aktienmarkt und der reale Nachfrageeffekt

Durch die Konzentration im S&P 500 trifft eine Neubewertung weniger großer Titel passiv investierte Portfolios direkt.[14] Der IWF ergänzt eine zweite Ebene: Die wachsende internationale Exponierung gegenüber US-Aktien kann einen Rückgang über Vermögenseffekte grenzüberschreitend übertragen.[14] Das ist kein Argument gegen Indexfonds. Es ist eine Beschreibung der heutigen Verteilung des Marktrisikos.

Der reale Schaden hinge von der Art des Rückgangs ab. Eine geordnete Neubewertung, bei der die Unternehmen weiter investieren, aber mit normalen Renditeerwartungen, wäre unangenehm für Aktionäre und begrenzt für die Gesamtwirtschaft. Ein harter Rückgang mit gleichzeitiger Kreditverknappung würde dagegen Bauprojekte stoppen, Zulieferbestellungen streichen und den Kon-

sum über Vermögensverluste belasten. Die aktuelle Verwundbarkeit erhöht sich, weil hohe Bewertungen, geopolitische Unsicherheit und teils enge Kreditspreads nicht unabhängig voneinander sind.[14][22]

Was wahrscheinlich nicht mitgerissen würde

Die wichtigste Gegenbremse sind die Bilanzen der größten Plattformen. Die Hyperscaler haben diversifizierte, laufende Geschäfte.[4][5][6][7] Ihre KI-Anlagen können oft auch für klassische Cloud-Lasten, interne Produkte oder spätere Modelle verwendet werden. Deshalb ist das Grundszenario nicht „alles fällt auf null“. Es ist eine Umschichtung von Wert: von einigen Hardware- und Infrastrukturinvestoren zu Nutzern, die von günstigeren Modellen und billigerer Rechenzeit profitieren.

Eine Korrektur könnte sogar die Verbreitung beschleunigen. Wenn Rechenleistung billiger wird, können kleinere Firmen mehr Anwendungen testen. Das wäre für den gesamtwirtschaftlichen Nutzen gut und für die Rendite der Kapazitätsanbieter schmerzhaft. Technologieerfolg und Anlegererfolg können auseinanderlaufen.

Woran die These scheitert

Bleiben die fünf Warnsignale bis zum 30. September 2027 überwiegend aus, ist die Überbau-These geschwächt. Steigen Auslastung und margenstarke KI-Umsätze dagegen nachvollziehbar, während die Betreiber ihre Finanzierung zu stabilen Konditionen verlängern, hat die Nachfrage den Ausbau schneller aufgenommen als hier angenommen. Ein großer technischer Sprung, ein unerwartet starker Agentenmarkt oder anhaltende Kapazitätsknappheit würden in dieselbe Richtung weisen.

Der IWF und die Federal Reserve warnen bereits heute vor der Kombination aus Konzentration, hohen Bewertungen, KI-bezogenen Investitionen und möglichen Verstärkern im Kreditmarkt.[12][14][22]

Das eigentliche Risiko ist die falsche Frage

Die falsche Frage lautet: „Ist KI eine Blase?“ Sie zwingt zu einem Ja oder Nein, obwohl Technologie, Geschäftsmodell und Finanzierung verschiedene Dinge sind.

Die bessere Frage lautet: Welche Renditeannahmen müssen wahr werden, damit der heute gebaute Kapitalstock bezahlt ist – und wer trägt den Verlust, wenn sie nur teilweise wahr werden?

Auf diese Frage gibt es eine nüchterne Antwort. Für einzelne Anwendungen ist die Nützlichkeit belegt, die breite organisatorische Nutzung wächst aber ungleich.[15][16][1][2] Die Infrastruktur wird in einer historischen Größenordnung gebaut.[3] Ihre Finanzierung und ihre Erlösdeckung sind im Zentrum solide, an den Rändern aber konzentriert und teilweise miteinander verschränkt.[12] Das macht die Lage weder harmlos noch apokalyptisch.

Wer aus der Technik das Ende aller Arbeit ableitet, übertreibt. Wer aus hohen Bewertungen das Ende der Technik ableitet, ebenfalls. Der wahrscheinlichere Verlauf ist härter und interessanter: Eine reale, produktive Technologie wird durch eine Phase überhöhter Kapitalerwartungen gehen. Einige Unternehmen werden daran sehr viel verdienen. Andere werden Anlagen abschreiben, Kapazität billiger vermieten oder ihre Finanzierungen neu verhandeln müssen.

Bis dahin sollten die Rechnungen offen genug liegen, um die zentrale Wette zu beurteilen: Kann die KI-Infrastruktur zu den Preisen, den Stromkosten und der Geschwindigkeit des technischen Fortschritts genug verdienen, um den heutigen Bau zu rechtfertigen?

Wichtiger Hinweis: Dieser Beitrag ist keine Anlageberatung und keine Empfehlung zum Kauf oder Verkauf von Finanzinstrumenten. Kapitalanlagen können an Wert verlieren; individuelle Entscheidungen brauchen eine eigene Prüfung.

QUELLEN (22)

[1] Dell'Acqua et al., *The Cybernetic Teammate*, NBER Working Paper 33641, 2025

<https://www.nber.org/papers/w33641>, National Bureau of Economic Research, abgerufen am 18. Juli 2026

[2] Dillon et al., *Shifting Work Patterns with Generative AI*, NBER Working Paper 33795, 2025

<https://www.nber.org/papers/w33795>, National Bureau of Economic Research, abgerufen am 18. Juli 2026

[3] International Energy Agency, *Energy and AI*, 2025

<https://www.iea.org/reports/energy-and-ai/>, International Energy Agency, abgerufen am 18. Juli 2026

[4] Alphabet, *2025 Q4 Earnings Call*, 2026

https://abc.xyz/investor/events/event-details/2026/2025-Q4-Earnings-Call-2026-Dr_C033hS6/default.aspx, Alphabet Inc., abgerufen am 18. Juli 2026

[5] Microsoft, *FY 2026 Third Quarter Earnings Conference Call*, 2026

<https://www.microsoft.com/en-us/investor/events/fy-2026/earnings-fy-2026-q3>, Microsoft Corporation, abgerufen am 18. Juli 2026

[6] Amazon, *Fourth Quarter Results 2025*, 2026

<https://ir.aboutamazon.com/news-release/news-release-details/2026/Amazon-com-Announces-Fourth-Quarter-Results/>, Amazon.com, Inc., abgerufen am 18. Juli 2026

[7] Meta, *Fourth Quarter and Full Year 2025 Results*, 2026

<https://investor.atmeta.com/investor-news/press-release-details/2026/Meta-Reports-Fourth-Quarter-and-Full-Year-2025-Results/default.aspx>, Meta Platforms, Inc., abgerufen am 18. Juli 2026

[8] Microsoft, *Annual Report 2025*, 2025

<https://www.microsoft.com/investor/reports/ar25/>, Microsoft Corporation, abgerufen am 18. Juli 2026

[9] Meta, *Fourth Quarter and Full Year 2024 Results*, 2025, Bilanzierungsannahme zur Nutzungsdauer

<https://investor.atmeta.com/investor-news/press-release-details/2025/Meta-Reports-Fourth-Quarter-and-Full-Year-2024-Results/>, Meta Platforms, Inc., abgerufen am 18. Juli 2026

[10] International Energy Agency, *Energy and AI*, Executive Summary, 2025

<https://www.iea.org/reports/energy-and-ai/executive-summary>, International Energy Agency, abgerufen am 18. Juli 2026

[11] International Energy Agency, *Energy and AI*, Executive Summary, 2025

<https://www.iea.org/reports/energy-and-ai/executive-summary>, International Energy Agency, abgerufen am 18. Juli 2026

[12] International Monetary Fund, *Global Financial Stability Report*, April 2026, Kapitel 1

<https://www.imf.org/en/Publications/GFSR/Issues/2026/04/14/global-financial-stability-report-april-2026>, International Monetary Fund, abgerufen am 18. Juli 2026

[13] NVIDIA, Form 10-K für das Geschäftsjahr 2026

<https://www.sec.gov/Archives/edgar/data/1045810/000104581026000021/nvda-20260125.htm>, U.S. Securities and Exchange Commission, abgerufen am 18. Juli 2026

[14] International Monetary Fund, *Global Financial Stability Report*, April 2026

<https://www.imf.org/en/publications/gfsr/issues/2026/04/14/global-financial-stability-report-april-2026>, International Monetary Fund, abgerufen am 18. Juli 2026

[15] U.S. Census Bureau, *Large Firms With at Least 20 Employees Biggest AI Users*, 26. Mai 2026

<https://www.census.gov/library/stories/2026/05/ai-use-businesses.html>, U.S. Census Bureau, abgerufen am 18. Juli 2026

[16] Bonney et al., *The Microstructure of AI Diffusion*, U.S. Census Bureau Working Paper CES-26-25, 2026

<https://www.census.gov/library/working-papers/2026/adrm/CES-WP-26-25.html>, U.S. Census Bureau, abgerufen am 18. Juli 2026

[17] OECD, BCG und INSEAD, *The Adoption of Artificial Intelligence in Firms*, 2025

https://www.oecd.org/en/publications/the-adoption-of-artificial-intelligence-in-firms_f9ef33c3-en/full-report.html, OECD, abgerufen am 18. Juli 2026

[18] International Energy Agency, *Energy and AI*, 2025

<https://www.iea.org/reports/energy-and-ai/>, International Energy Agency, abgerufen am 18. Juli 2026

[19] Anthropic, *Anthropic Economic Index Report: Uneven Geographic and Enterprise AI Adoption*, 2025

<https://www.anthropic.com/research/anthropic-economic-index-september-2025-report>, Anthropic, abgerufen am 18. Juli 2026

[20] International Monetary Fund, *Global Financial Stability Report*, Box 1.1 „DeepSeek and the AI Revolution“, April 2025

<https://www.imf.org/-/media/files/publications/gfsr/2025/april/english/text.pdf>, International Monetary Fund, abgerufen am 18. Juli 2026

[21] Microsoft, *FY 2025 Second Quarter Earnings Conference Call*, 2025

<https://www.microsoft.com/en-us/investor/events/fy-2025/earnings-fy-2025-q2>, Microsoft Corporation, abgerufen am 18. Juli 2026

[22] Board of Governors of the Federal Reserve System, *Financial Stability Report*, Mai 2026

<https://www.federalreserve.gov/publications/2026-may-financial-stability-report-overview.htm>, Board of Governors of the Federal Reserve System, abgerufen am 18. Juli 2026

BEGRIFFE IM ARTIKEL (5)

Abschreibungen: Buchhalterische Verteilung der Anschaffungs- oder Herstellungskosten eines Vermögenswerts über seine Nutzungsdauer.

GPU: Grafikprozessor, der viele Rechenoperationen parallel ausführen kann und häufig für KI-Anwendungen genutzt wird.

Hyperscaler: Sehr große Cloud-Anbieter mit weltweit skalierter Rechen- und Speicherinfrastruktur.

Kapitalkosten: Kosten für das eingesetzte Eigen- oder Fremdkapital eines Unternehmens.

Rechenzentrum: Gebäude oder Anlage, in der Server und Netztechnik für digitale Dienste betrieben werden.

Eine Perspektive. Kein Orakel.

KI meint ist eine digitale Tageszeitung für begründete Einordnung. Jeder Beitrag entsteht vollständig mit künstlicher Intelligenz, verbindet überprüfbare Quellen mit einer sichtbar markierten These und benennt Gegenargumente sowie offene Unsicherheit. Die angegebenen Quellen werden redaktionell geprüft.

ÜBER DAS FORMAT

Wofür KI meint steht und wie eine Ausgabe einordnet.

www.ki-meint.de/ueber

KI-REDAKTION

Wie die Beiträge entstehen und redaktionell geprüft werden.

www.ki-meint.de/ueber-die-ki

GRUNDSÄTZE

Fakten, Bewertung und Unsicherheit bleiben klar erkennbar.

www.ki-meint.de/grundsatz

Impressum

Angaben gemäß § 5 Digitale-Dienste-Gesetz (DDG)

ANBIETER

TURNUS Espresso GmbH
Rathenower Straße 16b
39576 Stendal
Deutschland

Geschäftsführung: Andreas Goclik

KONTAKT UND REGISTER

Telefon: +49 162 2063808
E-Mail: info@turnus-espresso.de

Amtsgericht Stendal
HRB 34956
USt-IdNr.: DE 345769299

REDAKTIONELL VERANTWORTLICH

Andreas Goclik
Rathenower Straße 16b
39576 Stendal

Verantwortlich gemäß § 18 Abs. 2
Medienstaatsvertrag (MStV).

Vollständiges Impressum: www.ki-meint.de/impressum · Stand: 18. Juli 2026