HELMUT KOHL
Management GmbH

Markt-, Infrastruktur- und Investitionsstudie

Rechenzentren in Deutschland

Schwerpunkte:

DOKUMENTTYP	Branchenstudie zur deutschen Rechenzentrumswirtschaft
ZIELBILD	Belastbares Whitepaper für Investoren, Betreiber, Politik und Verwaltung
ERSTELLT VON	Helmut Kohl, Gerd Simon Erstellt mit Hilfe von Perplexity (Sonar 2), ChatGPT (5.5) und Claude (Opus 4.7)
STAND	4. Mai 2026

Die vorliegende Studie analysiert Marktstruktur, Investitionen, Energiebedarf, regulatorischen Rahmen und Zukunftsperspektiven der Rechenzentrumsbranche in Deutschland. Im Fokus stehen die Wachstumstreiber – insbesondere KI-Workloads, Hyperscaler-Investitionen und das Energieeffizienzgesetz – sowie konkrete Ausblicke für die kommenden 3, 5 und 10 Jahre. Quellengrundlage sind Bitkom/Borderstep 2025, die Rechenzentrumsstrategie der Bundesregierung vom März 2026, BMWE/EnEfG-Dokumente sowie Branchenanalysen führender Marktforschungshäuser.

Helmut Kohl Management GmbH
Niddastraße 49
D-61440 Oberursel

Nassauische Sparkasse
IBAN DE69 5105 0015 0159 0785 42
BIC/Swift NASS DE 55

Ust-IdNr. DE349599735
HRB 15809
Amtsgericht Bad Homburg

HELMUT KOHL
Management GmbH

Inhaltsverzeichnis

1 Executive Summary 2
2 Einleitung und Zielsetzung 4
3 Methodik und Datengrundlage 6
4 Marktüberblick: Rechenzentren in Deutschland 2026 7
5 Geografische Verteilung und regionale Cluster 10
6 Investitionen und Investitionsumfeld 13
7 Infrastruktur und Technologie 17
8 Energiebedarf und Nachhaltigkeit 20
9 Regulatorischer Rahmen: EnEfG und Rechenzentrumsstrategie 24
10 Künstliche Intelligenz als Wachstumstreiber 27
11 Herausforderungen und Risiken 29
12 Internationaler Vergleich 32
13 Ausblick: Entwicklungsperspektiven 2029, 2031 und 2036 35
14 Handlungsempfehlungen 41
15 Fazit 44
Anhang A: Quellenverzeichnis 45
Anhang B: Verzeichnisse und Abkürzungen 49

1 Executive Summary

Der deutsche Rechenzentrumsmarkt hat im Jahr 2026 einen historischen Wendepunkt erreicht. Mit erstmals mehr als 3.000 Megawatt installierter IT-Anschlussleistung – ein Wachstum von rund neun Prozent gegenüber 2024 – hat sich Deutschland als der größte Rechenzentrumsstandort Europas weiter konsolidiert. Hinter diesem Wachstum stehen rund 2.000 Rechenzentren mit jeweils mehr als 100 Kilowatt Anschlussleistung, davon rund 100 Großanlagen mit über 5 Megawatt Leistung. Frankfurt am Main bleibt mit weitem Abstand der bedeutendste Standort und beheimatet allein über ein Drittel aller deutschen Kapazitäten.

Treibende Kraft des Wachstums ist die Künstliche Intelligenz. Während HPC-/KI-Kapazitäten 2025 nach Bitkom/Borderstep bei rund 15 Prozent der Gesamtkapazität lagen (rechnerisch ca. 450 MW auf Basis von 2.980 MW) der Gesamtkapazität ausmachten, soll dieser Anteil nach Bitkom/Borderstep-Prognose bis 2030 auf rund 40 Prozent steigen; bei rund 5.000 MW Gesamtleistung entspräche dies etwa 2.000 MW und damit auf 40 Prozent der gesamten installierten Leistung anwachsen. Parallel dazu hat sich der Energiebedarf der Branche von rund 10,5 TWh im Jahr 2010 auf 21,3 TWh im Jahr 2025 verdoppelt. Bis 2030 erwartet die Deutsche Energie-Agentur (dena) einen Anstieg auf über 30 TWh, langfristig bis 2045 auf rund 80 TWh.

Die Hyperscaler – allen voran Amazon Web Services, Microsoft und Google – haben für die kommenden Jahre Investitionen in zweistelliger Milliardenhöhe angekündigt. AWS allein investiert 8,8 Milliarden Euro im Großraum Frankfurt sowie weitere 7,8 Milliarden Euro in die European Sovereign Cloud in Brandenburg. Microsoft errichtet im Rheinischen Revier Anlagen für 3,2 Milliarden Euro, Google plant 5,5 Milliarden Euro für den Zeitraum 2026–2029. Insgesamt fließen 2025 nach Schätzung des Bitkom rund 12 Milliarden Euro in IT-Hardware und weitere 3,5 Milliarden Euro in Gebäude und Gebäudetechnik – ein Allzeithoch.

Trotz dieser Dynamik steht der Standort Deutschland vor erheblichen Herausforderungen. Die Strompreise zählen im internationalen Standortvergleich zu den wichtigsten Kostennachteilen Deutschlands; Genehmigungsverfahren dauern in Einzelfällen bis zu sieben Jahre, und für neue Großanschlüsse an das 380-kV-Höchstspannungsnetz gibt es laut Tennet aktuell kaum noch freie Schaltfelder. Die im November 2023 in Kraft getretenen verschärften Anforderungen des Energieeffizienzgesetzes (EnEfG) wurden Anfang 2026 in einem Referentenentwurf zwar gelockert (PUE-Grenzwert für Neuanlagen von 1,2 auf 1,3 angehoben), bleiben aber im EU-Vergleich anspruchsvoll. Im internationalen Wettbewerb bleibt Deutschland weit hinter den USA zurück.

Mit der am 18. März 2026 vom Bundeskabinett verabschiedeten Nationalen Rechenzentrumsstrategie hat die Bundesregierung den politischen Rahmen geschaffen, um die Kapazitäten bis 2030 mindestens zu verdoppeln und die KI- und HPC-Rechenleistung zu vervierfachen. 28 konkrete Maßnahmen adressieren die Handlungsfelder Energie und Nachhaltigkeit, Standort und Fläche sowie Technologie und Souveränität. Im Sondervermögen wurden 805 Millionen Euro als deutscher Finanzierungsbeitrag an einer der bis zu fünf europäischen KI-Gigafabriken bereitgestellt, deren Standortauswahl im Sommer 2026 entschieden werden soll.

Der vorliegende Bericht analysiert die Marktstruktur, die wesentlichen Trends und die Herausforderungen des deutschen Rechenzentrumsmarkts und gibt einen detaillierten Ausblick auf die Entwicklung in den nächsten Jahren. Im Fokus stehen dabei die Wechselwirkungen zwischen technologischer Entwicklung, regulatorischem Rahmen, Energiewende und globalem Wettbewerb.

Schlüsselzahlen 2026 auf einen Blick

Kennzahl	Bestand 2025	Prognose 2030
Anzahl Rechenzentren > 100 kW	ca. 2.000	ca. 2.300–2.500
IT-Anschlussleistung gesamt (MW)	**2.980**	**5.000–6.000**[1]
davon KI-/HPC-Kapazität (MW)	ca. 450	ca. 2.020
Anteil KI an Gesamtkapazität	15 %[2]	40 %
Energieverbrauch (TWh/a)	21,3	> 30
Investitionen IT-Hardware in 2025 (Mrd. €)	12,0	>18
Investitionen Gebäude/TGA in 2025 (Mrd. €)	3,5	>6
Abgeleiteter ITK-Hardware-Kapitalstock (Mrd. €)	180[3]	>360[3]
Abgeleiteter Gebäude-/TGA-Kapitalstock (Mrd. €)	45[3]	>90[3]
Anteil Erneuerbare Energien (Strom)	> 65 % (bilanziell)	100 % (gesetzl. Pflicht)

Tabelle 1: Schlüsselkennzahlen des deutschen Rechenzentrumsmarktes - Stand und Prognose
Quelle: Bitkom/Borderstep 2025; eigene Berechnungen

[1] **Konservativ geschätzter Wert der Bundesregierung**

[2] **Die Prozentangaben folgen Bitkom/Borderstep. Absolute MW-Werte wurden rechnerisch aus der Gesamtleistung abgeleitet.**

[3] **Eigene überschlägige Ableitung; nicht mit jährlichem CapEx verwechseln. Annahme: Gebäude/TGA 0,015 Mrd. € je MW installierter Leistung, Verhältnis ITK-Hardware zu Gebäude/TGA gerundet 4:1.**

2 Einleitung und Zielsetzung

2.1 Bedeutung von Rechenzentren für Wirtschaft und Gesellschaft

Rechenzentren sind das physische Rückgrat der digitalen Wirtschaft. Ohne die in ihnen gebündelten Kapazitäten an Server-, Speicher- und Netzwerkleistung wären moderne Anwendungen wie Cloud-Computing, Streaming, E-Commerce, autonomes Fahren, industrielle Automatisierung und insbesondere Künstliche Intelligenz nicht denkbar. Während Telekommunikationsnetze die Datenautobahnen der digitalen Welt darstellen, sind Rechenzentren ihre Knotenpunkte und Produktionsstätten – Bundesdigitalminister Karsten Wildberger bezeichnete sie im April 2026 vor dem Deutschen Bundestag als die „neuen Raffinerien" einer Wirtschaft, in der Rechenleistung zum strategischen Rohstoff geworden ist.

Mit der zunehmenden Digitalisierung und insbesondere mit der breiten industriellen Nutzung Künstlicher Intelligenz hat sich die ökonomische und politische Bedeutung von Rechenzentren in den letzten Jahren deutlich verändert. Sie sind nicht mehr nur ein Infrastrukturthema der IT-Branche, sondern ein Standort- und Souveränitätsthema ersten Ranges. Wer eigene leistungsfähige Rechenzentren betreibt, behält die Kontrolle über die in ihnen verarbeiteten Daten, kann KI-Modelle nach eigenen Maßstäben entwickeln und trainieren und ist nicht von ausländischen Cloud-Anbietern abhängig, die ihrerseits unter dem Recht und gegebenenfalls dem politischen Einfluss ihrer Heimatländer stehen.

Die deutsche Volkswirtschaft profitiert in mehrfacher Hinsicht von der Existenz starker Rechenzentrumsstandorte. Erstens schaffen sie die Grundlage für einen leistungsfähigen Wirtschaftsstandort, dessen Industrie auf datenintensive Anwendungen wie Industrial Internet of Things, digitale Zwillinge und Predictive Maintenance angewiesen ist. Zweitens generieren sie selbst erhebliche Wertschöpfung durch Investitionen in Bau, technische Ausrüstung und Betrieb. Allein AWS rechnet damit, dass die eigenen Investitionen in Deutschland zusätzliches BIP-Wachstum von 15,4 Milliarden Euro generieren und jährlich 15.200 Arbeitsplätze in lokalen Unternehmen unterstützen. Drittens sind sie ein zentraler Hebel für die Erreichung von Klimazielen, da effiziente Anwendungen aus Cloud und KI in vielen Sektoren – Verkehr, Gebäude, industrielle Fertigung – CO_2-Einsparungen ermöglichen, die ein Vielfaches des eigenen Energieverbrauchs übersteigen.

2.2 Zielsetzung und Forschungsfragen

Die vorliegende Studie verfolgt das Ziel, einen umfassenden, datenbasierten Überblick über den Status quo des deutschen Rechenzentrumsmarktes im Jahr 2026 zu geben und einen fundierten Ausblick auf die Entwicklung in den nächsten Jahren bereitzustellen. Im Mittelpunkt stehen dabei sechs zentrale Forschungsfragen:

- Wie ist der deutsche Rechenzentrumsmarkt strukturiert, welche Marktsegmente und welche Akteure dominieren ihn, und wie verteilen sich die Kapazitäten geografisch?
- Welche Rolle spielt Deutschland im internationalen Vergleich, und welche Stärken und Schwächen prägen den Standort gegenüber den USA, China und konkurrierenden europäischen Standorten?
- Welche technologischen Trends – insbesondere im Bereich der Hochleistungskühlung, der Energieversorgung und der KI-Beschleuniger – werden den Markt in den kommenden Jahren prägen?
- Wie wirken sich die regulatorischen Rahmenbedingungen, insbesondere das Energieeffizienzgesetz und die Nationale Rechenzentrumsstrategie, auf Investitionen und Innovationsdynamik aus?
- Welche Risiken und Herausforderungen können das prognostizierte Wachstum gefährden – von Stromnetzanschluss über Fachkräftemangel bis zu gesellschaftlicher Akzeptanz?

- Welche konkreten Handlungsempfehlungen lassen sich für Politik, Wirtschaft und Verwaltung ableiten, um die Wettbewerbsfähigkeit des Standorts Deutschland in den nächsten zehn Jahren zu sichern?

2.3 Struktur und Aufbau der Studie

Die Studie gliedert sich in vierzehn Kapitel. Nach einer kurzen Einführung in Methodik und Datengrundlage (Kapitel 3) folgt ein quantitativer Marktüberblick mit den zentralen Kennzahlen zu Anzahl, Leistung, Umsatz und Wachstum (Kapitel 4). Die Kapitel 5 bis 7 behandeln die räumliche Verteilung der Rechenzentren in Deutschland, die laufenden und angekündigten Investitionen sowie die zugrundeliegende Infrastruktur und Technologie. Anschließend werden in den Kapiteln 8 bis 10 die zentralen Querschnittsthemen Energie und Nachhaltigkeit, der regulatorische Rahmen sowie die KI-Treiber detailliert beleuchtet. Die Kapitel 11 und 12 widmen sich Herausforderungen und internationalem Vergleich, bevor die Kapitel 13 bis 15 mit dem Ausblick, den Handlungsempfehlungen und einem Fazit abschließen.

3 Methodik und Datengrundlage

Die in der vorliegenden Studie verwendeten Daten und Analysen stützen sich auf eine systematische Auswertung primärer und sekundärer Quellen. Als Hauptquelle für die quantitative Marktbeschreibung dient die jährlich aktualisierte Bitkom-Studie „Rechenzentren in Deutschland“, die das Borderstep Institut für Innovation und Nachhaltigkeit erstellt. Die zugrundeliegende Methodik kombiniert eine Vollerhebung großer Rechenzentren mit einem statistischen Hochrechnungsmodell für kleinere Anlagen und mit primären Befragungen von Betreibern und Branchenexperten.

Ergänzend wurden amtliche Quellen wie die am 18. März 2026 vom Bundeskabinett verabschiedete Nationale Rechenzentrumsstrategie sowie das im April 2026 veröffentlichte Referentenentwurf-Papier des Bundesministeriums für Wirtschaft und Energie zur Novelle des Energieeffizienzgesetzes herangezogen. Daten der Deutschen Energie-Agentur (dena), Veröffentlichungen der German Datacenter Association (GDA) sowie Marktanalysen von Beratungshäusern wie Cushman & Wakefield, JLL und CBRE flossen ebenfalls ein. Internationale Vergleichsdaten stammen unter anderem von Synergy Research Group, Uptime Institute, IEA und den jeweiligen nationalen Branchenverbänden.

Sämtliche Investitionsdaten wurden direkten Pressemitteilungen der Hyperscaler und weiterer Anbieter (insbesondere AWS, Microsoft und Google), öffentlichen Investorenpräsentationen sowie Berichten der Wirtschaftsförderungen der Bundesländer entnommen. Für die regionale Verteilung der Kapazitäten wurden zusätzlich die Datenbestände von Cloudscene, datacentermap.com und der Eco-Verbandsbroschüre konsultiert.

Die zukunftsbezogenen Aussagen der Studie sind explizit als Prognosen oder Szenarien zu verstehen und basieren auf den derzeit erkennbaren Wachstumstreibern, Investitionsankündigungen und politischen Rahmensetzungen. Strukturbrüche – etwa eine plötzliche Verlangsamung des KI-Booms, geopolitische Verwerfungen oder regulatorische Eingriffe – können von den hier beschriebenen Entwicklungen abweichen. Die quantitativen Prognosen orientieren sich daher in der Regel an Korridoren mit unterer und oberer Schranke.

4 Marktüberblick: Rechenzentren in Deutschland 2026

4.1 Anzahl, Größe und IT-Anschlussleistung

Der deutsche Rechenzentrumsmarkt umfasst Anfang 2026 rund 2.000 Rechenzentren mit einer IT-Anschlussleistung von jeweils mehr als 100 Kilowatt. Hinzu kommen rund 50.000 kleinere IT-Installationen unterhalb dieser Schwelle, die zwar einzeln eine geringe, in Summe aber relevante Last darstellen. Innerhalb der Gruppe der größeren Anlagen gibt es eine ausgeprägte Konzentration: Rund 100 Großrechenzentren mit jeweils mehr als 5 Megawatt Anschlussleistung tragen rund die Hälfte zur gesamten installierten Leistung von 2.980 Megawatt im Jahr 2025 bei. Anfang 2026 wird die symbolisch wichtige Marke von 3.000 Megawatt erstmals überschritten.

Im Vergleich zum Vorjahr ist die installierte Leistung 2025 um rund neun Prozent oder 250 Megawatt gewachsen. Damit setzt sich die Wachstumsdynamik der vergangenen Jahre auch in der aktuellen Phase erhöhter regulatorischer Anforderungen und steigender Energiekosten ungebrochen fort. Die Wachstumstreiber haben sich allerdings in ihrer Zusammensetzung verschoben: Während in den 2010er Jahren vor allem das klassische Cloud-Computing den Ausbau dominierte, kommt seit 2023/2024 ein wachsender Anteil aus dem KI-Segment hinzu, das mittlerweile rund 15 Prozent der installierten Leistung ausmacht.

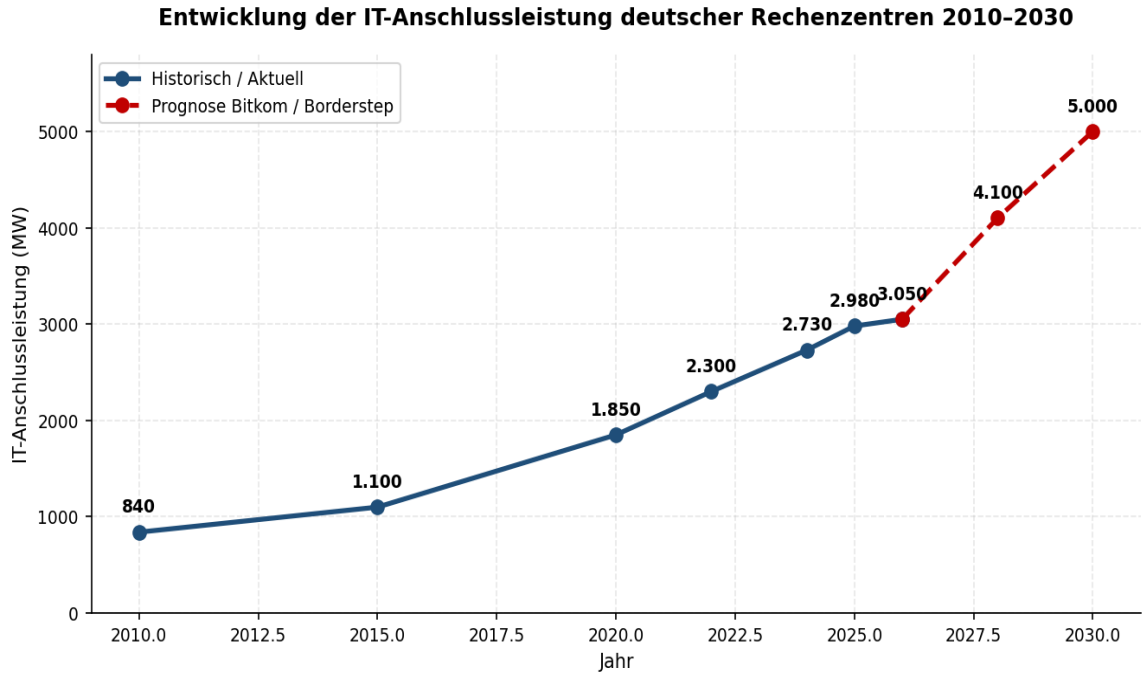


Abbildung 1: Entwicklung der IT-Anschlussleistung deutscher Rechenzentren 2010–2030 (in MW)
Quelle: Bitkom-Studie „Rechenzentren in Deutschland 2025“ (Borderstep Institut), eigene Darstellung

4.1.1 Klassifikation und Marktsegmente

Der Markt lässt sich entlang mehrerer Dimensionen segmentieren. Nach Bereitstellungsmodell wird zwischen Enterprise- bzw. On-Premises-Rechenzentren (Eigenbetrieb durch Unternehmen für eigene Anwendungen), Colocation-Rechenzentren (Vermietung von Stellplätzen, Strom und Kühlung an mehrere Kunden) und Cloud-Rechenzentren (Bereitstellung skalierbarer IT-Ressourcen) unterschieden. Hinzu kommen die KI-Rechenzentren, die je nach Geschäftsmodell als spezialisierte Cloud, Colocation oder Eigenbetrieb organisiert sein können, sowie Edge-Rechenzentren mit typischerweise weniger als 5 MW Leistung in räumlicher Nähe zum Endnutzer.

Bei der quantitativen Verteilung dominieren mittlerweile die Cloud- und Hyperscale-Rechenzentren. Während ihre Kapazitäten zwischen 2016 und 2021 um rund 150 Prozent gewachsen sind, stagnierten die klassischen Enterprise-Rechenzentren weitgehend. Cloud-Installationen machen 2025 rund 49 Prozent der deutschen Rechenzentrumskapazitäten aus; Colocation-Rechenzentren stellen nach Bitkom/Borderstep mehr als 50 Prozent der Kapazität bereit. Beide Blickwinkel sind teilweise überlappend, da Hyperscaler in Deutschland häufig Colocation-Angebote nutzen.

Marktstruktur nach analytischen Kategorien (teilweise überlappend)

Marktsegment	Anteil an Kapazität (MW, 2025)	Wachstum p.a. 2022–2025	Beispielhafte Anbieter
Cloud-Installationen	ca. 49 % (1.450 MW)	ca. 16,5 %	AWS, Microsoft, Google, Oracle, IBM
Colocation-Rechenzentren (inkl. Cloud-Nutzung)	> 50 % der Kapazität (überlappend mit Cloud)	ca. 14 %	Equinix, Digital Realty/Interxion, NTT, Cloud HQ
On-Premises/Enterprise + traditionell	rückläufiger Anteil; zusammen mit Edge > 1/3	ca. 1 %	DAX-/MDAX-Konzerne, Banken, Versicherungen
HPC/KI-Workloads	ca. 15 % der Gesamtleistung	ca. 30 %	FZ Jülich (JUPITER), HLRS Stuttgart, LRZ Garching
Edge & kleinere Installationen	ca. 8 % Edge; kleinere IT-Installationen zusätzlich	ca. 18 %	Telekommunikationsanbieter, Industrie, Mobilfunk

Tabelle 2: Marktstruktur des deutschen Rechenzentrumsmarktes nach Anbieterkategorien (Schätzung)
Quelle: Eigene Berechnung auf Basis Bitkom/Borderstep, Synergy Research, Cushman & Wakefield

4.2 Wachstumstreiber

Das Wachstum des Marktes wird auch 2026 durch ein Bündel sich gegenseitig verstärkender Faktoren angetrieben. Den dominierenden Treiber stellt mittlerweile die Künstliche Intelligenz dar, deren Modelle für Training und Inferenz beträchtliche und in den vergangenen drei Jahren exponentiell gestiegene Mengen an spezialisierter Rechenleistung benötigen. Hinzu kommt die fortschreitende Migration in die Cloud, die durch die Konsolidierung dezentraler IT-Landschaften, die Modernisierung von Legacy-Systemen sowie die zunehmende Verbreitung von Software-as-a-Service-Modellen befeuert wird.

Auf der Nachfrageseite tragen weitere strukturelle Faktoren zum Wachstum bei. Die zunehmende Digitalisierung der öffentlichen Verwaltung im Rahmen des sogenannten Deutschland-Stacks, die Forderung nach digitaler Souveränität und die damit verbundene Repatriierung sensibler Workloads aus außereuropäischen Clouds, regulatorische Anforderungen wie DORA für den Finanzsektor sowie die digitale Transformation industrieller Wertschöpfungsketten bilden ein robustes Fundament der Marktdynamik. Auch die Ausbreitung von Streaming-Diensten, Gaming und insbesondere generativen KI-Anwendungen für den Endkonsumenten erzeugen einen kontinuierlichen Lastzuwachs.

Die hohe Nachfrage zeigt sich insbesondere an den sehr niedrigen Leerstandsquoten in den führenden Colocation-Märkten: Für Frankfurt/Rhein-Main weist Bitkom/Borderstep unter Bezug auf CBRE 2025 eine Leerstandsquote von nur rund 3 % aus – ein Wert, der den dringenden Ausbaubedarf unterstreicht und in einigen Marktsegmenten und Regionen bereits zu Wartelisten und Ausweichbewegungen führt. Frankfurt am Main weist mit einer Leerstandsquote von rund drei Prozent die niedrigste verfügbare Kapazität aller führenden europäischen Standorte auf, was zu Preissteigerungen und einer Ausdehnung des Frankfurter Rechenzentrums-Ökosystems in die Regionen Rhein-Main, Rhein-Neckar und das südliche Hessen führt.

4.3 Umsatz, Wertschöpfung und Beschäftigung

Die wirtschaftliche Bedeutung des Rechenzentrumsmarktes lässt sich an mehreren Kennzahlen ablesen. Die Investitionen in IT-Hardware lagen in 2025 nach Bitkom-Angaben bei rund 12 Milliarden Euro, ergänzt durch weitere 3,5 Milliarden Euro für Gebäude und technische Gebäudeausrüstung. Diese Werte stellen ein Allzeithoch dar und übertreffen den Vorjahreswert um rund 18 Prozent. Auf den deutschen Cloud-Markt entfallen dabei 2026 nach Bitkom-Berechnungen rund 20 Milliarden Euro mit einem Wachstum von 17 Prozent gegenüber 2025.

Die unmittelbare Beschäftigung in deutschen Rechenzentren wird auf rund 130.000 Vollzeitäquivalente geschätzt. Über vor- und nachgelagerte Branchen – Bauunternehmen, Anlagenhersteller, Stromversorger, IT-Dienstleister, Beratungs- und Wartungsfirmen – ergibt sich ein Beschäftigungseffekt in der Größenordnung von 250.000 bis 300.000 Personen. Die direkte und indirekte Bruttowertschöpfung der Branche liegt bei geschätzt 30 bis 35 Milliarden Euro pro Jahr.

5 Geografische Verteilung und regionale Cluster

5.1 Bundesländer im Vergleich

Die Verteilung der Rechenzentrumskapazitäten ab 40 MW in Deutschland ist hochgradig konzentriert. Mit über einem Drittel aller installierten Leistung dominiert das Bundesland Hessen mit der Metropolregion Frankfurt/Rhein-Main den Markt. Der Vorsprung gegenüber den Verfolgern Bayern, Nordrhein-Westfalen und Baden-Württemberg ist dabei deutlich. Pro Einwohner gerechnet ist die Rechenzentrumsdichte in Hessen rund fünfmal höher als in den anderen großen Flächenstaaten und etwa dreimal so hoch wie in den Stadtstaaten Hamburg und Berlin.

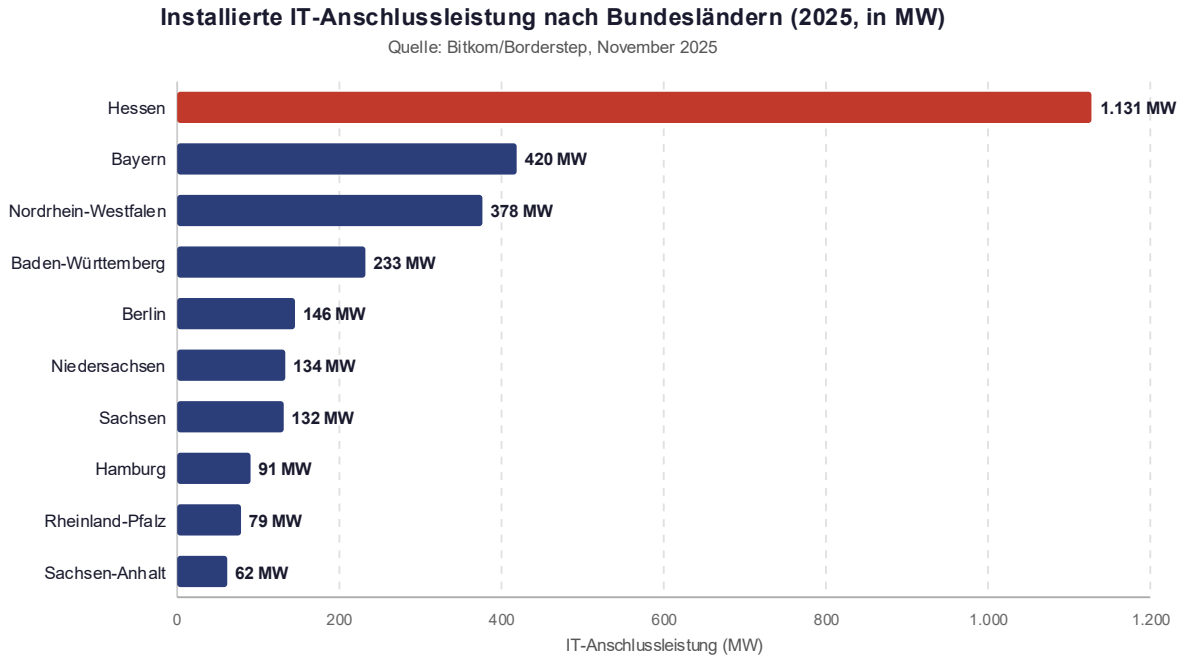


Abbildung 2: Installierte IT-Anschlussleistung nach Bundesländern 2025 (in MW)
Quelle: Bitkom-Studie „Rechenzentren in Deutschland 2025“ (Borderstep), eigene Darstellung

Detaildaten der Top-10-Bundesländer

Bundesland	MW (2025)	Anteil DE	Wachstum p.a.	Standortmerkmale
Hessen	**1.131**	**38 %**	**ca. 14 %**	DE-CIX, Finanzplatz, RZ-Cluster
Bayern	420	14 %	ca. 10 %	Industrie, München, LRZ Garching
Nordrhein-Westfalen	378	13 %	ca. 12 %	Rheinisches Revier, FZ Jülich
Baden-Württemberg	233	8 %	ca. 8 %	Stuttgart HLRS, Industrie, Forschung
Berlin	146	5 %	ca. 18 %	Start-up-Szene, Verwaltung, Tor zum Osten
Niedersachsen	134	5 %	ca. 12 %	Wind-Strom, große Flächen
Sachsen	132	4 %	ca. 10 %	Silicon Saxony, TSMC-Halbleiter-Cluster
Hamburg	91	3 %	ca. 9 %	Seekabel-Anlandung, Logistik, Medien
Rheinland-Pfalz	79	3 %	ca. 22 %	480-MW-Projekt Nierstein, Mainz
Sachsen-Anhalt	62	2 %	ca. 22 %	T-Systems Twin-Core, Cloud-/Sicherheits-RZ-Ökosystem

Tabelle 3: Top-10-Bundesländer nach installierter IT-Anschlussleistung 2025
Quelle: Bitkom/Borderstep, eigene Berechnungen.

5.2 Frankfurt: Europas führender Standort

Die Metropolregion Frankfurt/Rhein-Main bildet das mit Abstand bedeutendste Rechenzentrums-Cluster Deutschlands und ist zugleich der größte Standort in ganz Europa. Über 100 kommerzielle Rechenzentren werden hier von internationalen Anbietern wie NTT, Equinix, Digital Realty (Interxion), Telehouse, Global Switch, Cloud HQ und zahlreichen spezialisierten Betreibern unterhalten. Die installierte Kapazität liegt bei über 1.131 Megawatt – mehr als ein Drittel aller deutschen Rechenzentrumskapazitäten konzentrieren sich damit in einem Radius von rund 50 Kilometern um den Frankfurter Flughafen.

Die Anziehungskraft Frankfurts beruht auf einem dichten Geflecht aus Standortvorteilen. Den Kern bildet der DE-CIX, mit einem Rekordwert von 18,73 Terabit pro Sekunde Spitzenverkehr (2025) einer der größten Internet-Knoten der Welt, der Frankfurt zur zentralen Drehscheibe des europäischen Datenverkehrs macht. Hinzu kommen die Nähe zum Finanzplatz mit seinen Latenzanforderungen, ein etabliertes Ökosystem aus Stromversorgern, spezialisierten Bauunternehmen und Servicepartnern, die historisch gute Stromversorgung, deren Anschlusskapazitäten bei neuen Großprojekten inzwischen zunehmend an Grenzen stoßen sowie die zentrale geografische Lage in Europa.

Frankfurt wächst mit einer Rate von rund 14 Prozent pro Jahr schneller als jeder andere große europäische Standort und weist mit nur drei Prozent zugleich die niedrigste Leerstandsquote auf. Diese hohe Auslastung führt zu einem Auspressungseffekt: Neue Großprojekte weichen zunehmend in Metropolregion Rhein-Main (NTT Nierstein mit 480 MW geplant, STACKIT Infrastructure in Liederbach mit 80 MW), nach Bayern (Hyperscale-Projekte im Großraum München) oder in das südliche Hessen aus. Innerhalb der Stadt Frankfurt selbst hat die Stadtverwaltung mit einem Rechenzentrenkonzept die Ansiedlung neuer Anlagen an industrielle Standorte wie Höchst oder Fechenheim gelenkt, um Konflikte mit Wohngebieten zu minimieren.

Das Rechenzentrums-Ökosystem in Frankfurt ist eng mit der sogenannten FLAP-D-Region verknüpft – einem Akronym für Frankfurt, London, Amsterdam, Paris und Dublin, das die fünf bedeutendsten europäischen Rechenzentrumsstandorte zusammenfasst. Innerhalb dieser Gruppe ist Frankfurt der wachstumsstärkste Markt, knapp gefolgt von London. Die Gesamtkapazität der FLAP-D-Märkte liegt bei rund 4,7 Gigawatt. Während London mit verschärften Genehmigungsverfahren kämpft, Amsterdam in 2019 ein zeitweises Moratorium für neue Rechenzentren verhängt hat, das auch heute noch nachwirkt, und Dublin wegen Stromnetzengpässen Wartezeiten von mehreren Jahren aufweist, profitiert Frankfurt aktuell von vergleichsweise stabilen Rahmenbedingungen.

5.3 Aufstrebende Standorte

Über die etablierten Standorte hinaus zeichnet sich eine substanzielle Verschiebung der Wachstumsdynamik in die sogenannte „zweite Reihe“ ab. Insbesondere das Rheinische Revier entwickelt sich mit der Ansiedlung von Microsoft und QTS zunehmend zu einem wichtigen, neuen Standort. Auch Berlin wächst infolge der Standortentwicklung durch Vantage Data Centers.

Konkret sind in Brandenburg neue Projekte in erheblicher Größenordnung angekündigt, dominiert vom AWS-European-Sovereign-Cloud-Projekt (7,8 Mrd. Euro Investitionsvolumen). In Rheinland-Pfalz entsteht in Nierstein ein Rechenzentrum mit 480 Megawatt, in Mainz auf der Ingelheimer Aue auf dem Gelände der KMW AG bis 2026 ein moderner Standort mit 80 Megawatt Gesamtleistung.

Die Verschiebung in die Tier-2-Regionen folgt einem europaweiten Muster, das Branchenvertreter als möglichen Beginn vom „Ende des Akronyms FLAP“ beschreiben: An die Stelle weniger dominanter Hubs tritt zunehmend ein gleichmäßiger verteiltes Netz von Rechenzentrums-Standorten, das durch hochleistungsfähige Glasfaserverbindungen miteinander vernetzt ist. Für Deutschland eröffnen sich dadurch Wachstumschancen in strukturschwächeren Regionen, deren regionalpolitische Bedeutung erheblich ist.

6 Investitionen und Investitionsumfeld

6.1 Gesamtinvestitionen und Investitionsdynamik

Die Investitionen in den deutschen Rechenzentrumsmarkt haben 2025 ein neues Allzeithoch erreicht. Nach Berechnungen des Bitkom flossen in 2025 rund 12 Milliarden Euro in IT-Hardware (Server, Speicher, Netzwerktechnik) und weitere 3,5 Milliarden Euro in Gebäude und technische Gebäudeausrüstung (Strom, Kühlung, Sicherheit, USV). Das gesamte Investitionsvolumen liegt damit bei rund 15,5 Milliarden Euro – ein Wachstum von rund 18 Prozent gegenüber 2024.

Treiber dieser Investitionsdynamik ist neben der allgemeinen Cloud-Expansion vor allem die KI-Infrastruktur. Spezialisierte GPU-Server, deren Kosten pro Einheit ein Vielfaches herkömmlicher Server-Hardware betragen, machen einen rasch wachsenden Anteil an den IT-Hardware-Investitionen aus. Mit der Ablösung der NVIDIA H100/H200-Generation durch die Blackwell-Familie (B100, B200, GB300) und den geplanten Übergang auf die nächste Generation Vera Rubin im Jahr 2027 wird sich der Investitionszyklus weiter verkürzen.

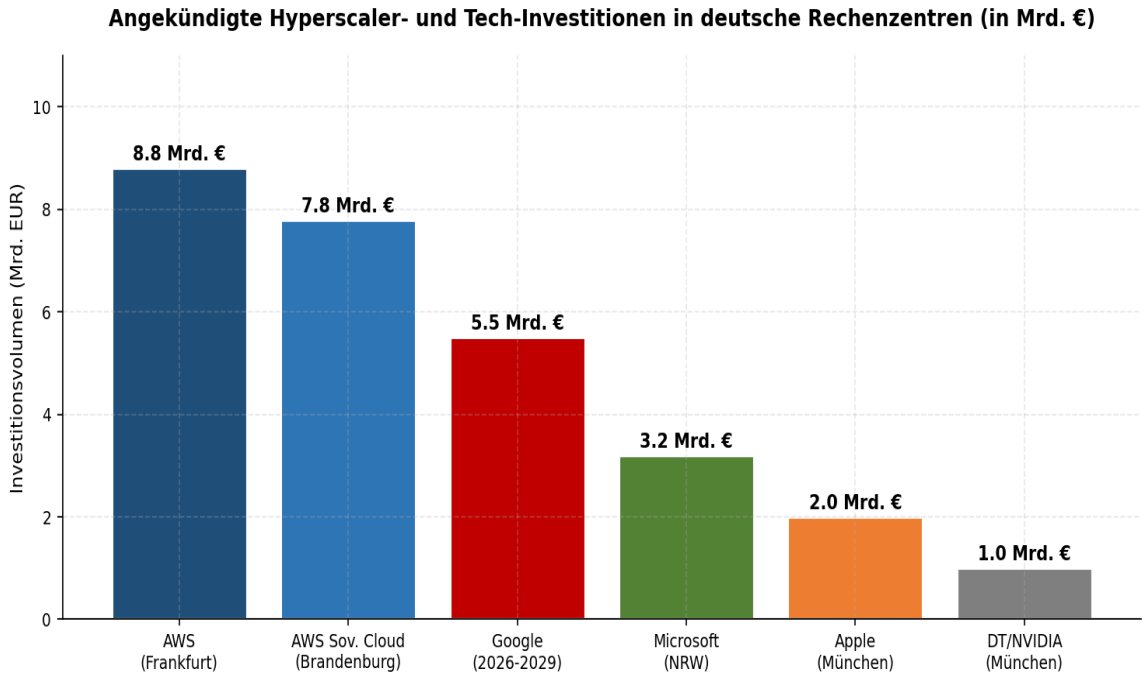


Abbildung 3: Angekündigte Hyperscaler- und Tech-Investitionen in deutsche Rechenzentrums-, Cloud- und angrenzende Tech-Infrastruktur (in Mrd. Euro)
Quelle: Pressemitteilungen der jeweiligen Unternehmen, Stand März 2026

6.2 Hyperscaler-Investitionen im Detail

Die Investitionssummen der internationalen Tech-Konzerne in den deutschen Rechenzentrumsmarkt haben in den vergangenen achtzehn Monaten neue Dimensionen erreicht. Die wichtigsten Vorhaben im Überblick:

Amazon Web Services (AWS)

Amazon Web Services hat für die AWS Europe (Frankfurt) Region Investitionen in Höhe von 8,8 Milliarden Euro bis 2026 angekündigt. Die Region umfasst aktuell drei Availability Zones im Großraum Frankfurt und ist die wichtigste AWS-Region in der Europäischen Union nach der irischen Region. Parallel dazu baut AWS in Brandenburg eine eigenständige „European Sovereign Cloud" auf, die Kunden mit besonderen Anforderungen an die Datensouveränität und das EU-Recht eine vollständig in Europa betriebene und von europäischem Personal gesteuerte Alternative bietet. Das Investitionsvolumen liegt bei 7,8 Milliarden Euro bis 2040, es wird erwartet, dass der Dienst in 2026 in den produktiven Betrieb gehen wird. Insgesamt summieren sich die AWS-Investitionen in Deutschland damit auf 16,6 Milliarden Euro – die höchsten ausländischen Direktinvestitionen eines einzelnen Unternehmens in Deutschland in den letzten zehn Jahren.

Microsoft

Microsoft errichtet im Rheinischen Revier in Bedburg neue Rechenzentren und plant insgesamt 3,2 Milliarden Euro für die Verdopplung seiner KI- und Cloud-Kapazitäten in Deutschland. Die neuen Anlagen sollen auch die bestehende Microsoft-Region Frankfurt ergänzen. Die Standortwahl im Rheinischen Revier ist strukturpolitisch motiviert, da der Bau in einer vom Kohleausstieg betroffenen Region erfolgt und dort neue, hochwertige Arbeitsplätze entstehen sollen.

Google / Alphabet

Google hat im November 2025 angekündigt, im Zeitraum 2026 bis 2029 insgesamt 5,5 Milliarden Euro in Cloud- und KI-Infrastruktur sowie Büros in Deutschland zu investieren. Geplant sind Cloud- und KI-Rechenzentren in Dietzenbach und Hanau im Großraum Frankfurt, die die Google Cloud Region Frankfurt erheblich erweitern werden. Damit reagiert Google auf die wachsende europäische Nachfrage nach KI-Diensten und positioniert sich gegen die etablierten Hyperscaler AWS und Microsoft Azure.

Deutsche Telekom & NVIDIA

Im Herbst 2025 haben die Deutsche Telekom und NVIDIA die Errichtung eines gemeinsamen KI-Rechenzentrums in München mit einem Investitionsvolumen von rund einer Milliarde Euro angekündigt. Das Projekt firmiert als „Industrial AI Cloud" und wird als weltweit erstes seiner Art beworben – mit dem Anspruch, speziell auf die Bedürfnisse der europäischen Industrie für die Entwicklung und den Betrieb industrieller KI-Anwendungen zugeschnitten zu sein. Erster Großkunde ist SAP. Das Projekt unterstreicht die strategische Bedeutung lokaler KI-Infrastruktur für die deutsche Industrie und schafft eine Alternative zur Abhängigkeit von US-amerikanischen Hyperscalern.

Apple und weitere Akteure

Apple hat insgesamt rund zwei Milliarden Euro in sein European Silicon Design Center in München investiert, in dem über 2.000 Ingenieure an der Entwicklung der eigenen Chip-Architekturen arbeiten – eine Chip- und F&E-Investition und damit kein Rechenzentrumsprojekt im engeren Sinne. Hinzu kommen Investitionen weiterer internationaler Betreiber: Equinix investiert bis 2027 über 730 Millionen US-Dollar in den Ausbau seiner Frankfurter xScale- und Retail-Kapazitäten (u. a. FR10x.1 mit 14 MW); Digital Realty entwickelt den Digital Park Fechenheim zu einem Campus mit bis zu 200 MW Gesamtkapazität (Investitionssumme: ca. 1 Mrd. Euro); Vantage Data Centers plant allein im Großraum Frankfurt mit Bad Vilbel (160 MW) und Groß-Gerau (2,5 Mrd. Euro Gesamtinvestition) zwei weitere Großprojekte; und Oracle kündigt für seine Cloud Region Frankfurt Investitionen von zwei Milliarden US-Dollar über fünf Jahre an.

Neben den klassischen Hyperscalern tritt seit 2024 eine neue Anbieterkategorie in den Vordergrund: sogenannte Neocloud-Anbieter. Diese auf GPU-as-a-Service spezialisierten Infrastrukturanbieter – darunter CoreWeave (USA), Nebius (Niederlande) und Nscale (Großbritannien) – bauen zweckoptimierte KI-Rechenzentren und vermieten GPU-Rechenkapazität an Hyperscaler und KI-Entwickler. Nach Daten der Synergy Research Group erzielte das Segment 2025 weltweit einen Umsatz von über 25 Milliarden US-Dollar – eine Verdreifachung gegenüber dem Vorjahr.

Bemerkenswert aus deutscher Perspektive: Keiner der führenden Neocloud-Anbieter hat bislang ein konkretes Infrastrukturprojekt in Deutschland angekündigt. CoreWeave investiert 2,2 Milliarden US-Dollar in Rechenzentren in Norwegen, Schweden und Spanien; Nebius betreibt seinen europäischen Leitstandort in Finnland; Nscale fokussiert auf Norwegen, das Vereinigte Königreich und Portugal. Mehr als die Hälfte aller europäischen Neocloud-Kapazitätsabschlüsse in 2025 entfiel auf die nordischen Länder – Regionen mit günstigem Strom aus Wasserkraft und niedrigen Grundstückspreisen. Deutschlands strukturell höhere Energiekosten und komplexe Genehmigungsverfahren erweisen sich damit nicht nur als Wettbewerbsnachteil gegenüber klassischen Hyperscalern, sondern auch gegenüber dieser wachstumsstarken neuen Anbieterkategorie.

Eine Übersicht der größten angekündigten Einzelprojekte zeigt die Tabelle 4.

Übersicht der größten angekündigten Einzelprojekte

Projekt / Standort	Investor	Leistung (MW)	Volumen	Zeitplan
Brandenburg (Sov. Cloud)	AWS	**n. v. (öffentlich nicht genannt)**	7,8 Mrd. €	bis 2040
Nierstein, Rheinland-Pfalz	NTT	**480**	ca. 4 Mrd. €	ab 2027
Frankfurt-Höchst (FRA03)	Cloud HQ	**ca. 276**	> 2 Mrd. €	ab 2026
Frankfurt-Region (gesamt)	AWS	**n. v. (öffentlich nicht genannt)**	8,8 Mrd. €	bis 2026
Rheinisches Revier / Bedburg	Microsoft	**n. v. (öffentlich nicht genannt)**	3,2 Mrd. €	Spatenstich 12.03.26
Dietzenbach/Hanau	Google	**n. v. (öffentlich nicht genannt)**	5,5 Mrd. €	2026–2029
München (Industrial AI)	DT/NVIDIA	**n. v. (KI-Cluster; ca. 10.000 NVIDIA-GPUs angekündigt)**	1,0 Mrd. €	ab 2026

Tabelle 4: Größte angekündigte Rechenzentrums-Einzelprojekte in Deutschland
Quelle: Unternehmenspressemitteilungen, Stadtverwaltungen, Branchenpresse (Stand März 2026); MW-Angaben, soweit nicht offiziell publiziert, als indikative Annahmen behandeln.

Ergänzende Analyse zur Kapazitätsplanung (MW):

Da führende Hyperscaler (insb. AWS, Google und Microsoft) die exakte IT-Anschlussleistung ihrer Standorte aus Wettbewerbsgründen und zum Schutz geschäftskritischer Infrastruktur in der Regel nicht offiziell publizieren, wurden in Tabelle 4 für diese Projekte die Angaben „n. v.“ (nicht veröffentlicht) gewählt.

Marktgängige Analysen und öffentlich einsehbarere Genehmigungsverfahren (u. a. Netzanschlussbegehren und Bebauungspläne) erlauben jedoch eine Eingrenzung der Kapazitäten. Basierend auf aktuellen Branchendaten ergeben sich für die Projektentwicklungen folgende indikative Richtwerte:

- Region Brandenburg (AWS): Die langfristige Zielkapazität wird im Vollausbau (Prognosehorizont 2040) auf ca. 700 MW geschätzt.
- Frankfurt-Höchst (Cloud HQ/FRA03): Hier deuten Planungsunterlagen auf eine Gesamtkapazität von bis zu 350 MW hin.

- Rheinisches Revier / Bedburg (Microsoft): Für den Initialausbau der Cloud-Region West wird mit einer Kapazität von ca. 180 MW kalkuliert.
- Rhein-Main / Dietzenbach & Hanau (Google): Die kumulierte Leistung der neuen Standorte wird auf ca. 200 MW taxiert.

Diese Werte stellen Schätzgrößen dar und können je nach modularer Ausbaustufe und tatsächlicher Serverdichte (insb. bei KI-spezifischen Workloads) variieren.

6.3 Investitionsumfeld und Wertschöpfungseffekte

Die Wirtschaftseffekte dieser Investitionen sind beträchtlich. AWS allein schätzt, dass die geplanten Investitionen 15,4 Milliarden Euro zusätzliches Bruttoinlandsprodukt generieren und jährlich 15.200 Arbeitsplätze in lokalen Unternehmen unterstützen werden. Insgesamt könnten die in der oben stehenden Tabelle aufgeführten Großprojekte über die nächsten zehn Jahre eine direkte und indirekte Bruttowertschöpfung von 80 bis 100 Milliarden Euro auslösen – ein erheblicher Beitrag zur Modernisierung und Diversifizierung der deutschen Wirtschaft.

Trotz der hohen absoluten Investitionssummen zeigen sich im internationalen Vergleich jedoch signifikante Standortnachteile. Während Deutschland bei ausländischen Direktinvestitionen nach dem aktuellen EY-Standortindex 2025 den größten Rückgang (minus 17 Prozent) aller großen europäischen Länder verzeichnet, fließen US-amerikanische und chinesische Investitionen zunehmend in Tier-1-Märkte mit niedrigeren Strompreisen, kürzeren Genehmigungszeiten und mehr verfügbarer Fläche. Die Bundesregierung hat mit dem Sondervermögen Infrastruktur und der nationalen Rechenzentrumsstrategie versucht, gegenzusteuern, doch der Wettbewerb um Hyperscale-Investitionen bleibt intensiv.

7 Infrastruktur und Technologie

7.1 Server-Hardware und Leistungsdichte

Die technologische Entwicklung im Rechenzentrum wird seit etwa drei Jahren maßgeblich von zwei sich gegenseitig verstärkenden Trends bestimmt: dem rasanten Anstieg der Leistungsdichte pro Rack und der zunehmenden Spezialisierung der Hardware auf KI-Workloads. Während ein klassisches Enterprise-Rack 2020 typischerweise 10 bis 15 Kilowatt Leistungsaufnahme aufwies, liegen aktuelle KI-Cluster bei 80 bis 150 Kilowatt pro Rack. Für 2026 erwarten Experten Rack-Dichten von 240 Kilowatt, 2028 von einem Megawatt, und die Forschung untersucht bereits die Realisierbarkeit von 1,5 Megawatt pro Rack.

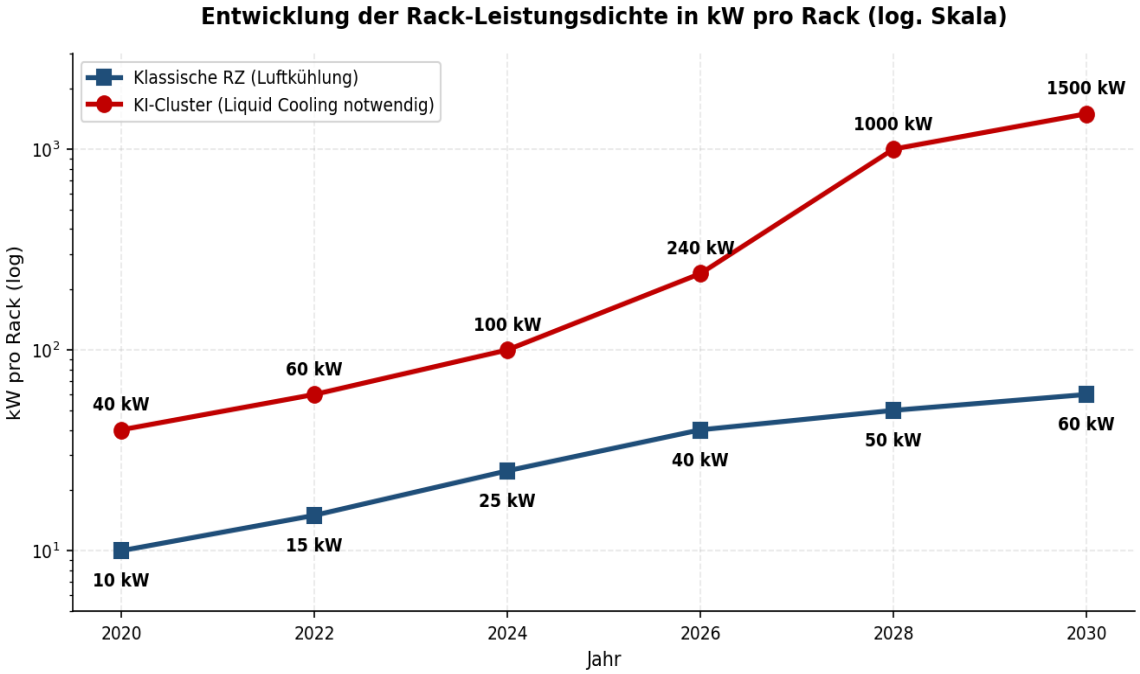


Abbildung 4: Entwicklung der Rack-Leistungsdichte für klassische und KI-Workloads (in kW pro Rack)
Quelle: Schneider Electric, Vertiv, Rittal, eigene Auswertung

Die Treiber dieser Entwicklung sind die GPUs der jüngsten Generationen. Eine NVIDIA H100 hat eine Leistungsaufnahme von rund 700 Watt, eine B200 bereits 1.000 Watt, und ein vollständig bestückter NVIDIA GB200 NVL72-Rack mit 72 Beschleunigern setzt rund 120 Kilowatt um. Die geplanten GB300- und Vera-Rubin-Generationen werden diese Werte weiter steigern. Damit ergeben sich neuartige Anforderungen an die gesamte Infrastruktur: Stromverteilung, Kühlung, Brandschutz, Tragfähigkeit der Böden und nicht zuletzt die Bauphysik der Gebäude.

Die Marktstruktur teilt sich nach Schneider Electric bis 2030 in drei klar abgegrenzte Segmente auf: Rund 25 Prozent der Neubauten werden auf weniger als 40 Kilowatt pro Rack ausgelegte Anlagen für den produktiven Betrieb klassischer Workloads sein. Etwa 50 Prozent entfallen auf Hybrid-Rechenzentren im Bereich von 40 bis 80 Kilowatt pro Rack, die sowohl produktive als auch Trainings-Workloads bewältigen. Die verbleibenden 25 Prozent sind Hochleistungs-Trainings-Cluster mit über 100 Kilowatt pro Rack, die ausschließlich für KI-Modelltrainings im Großmaßstab eingesetzt werden.

7.2 Kühlung: Vom Luftstrom zur Flüssigkeitskühlung

Die klassische Luftkühlung über Doppelböden und Kalt-/Warmgang-Konzepte stößt im Bereich von 50 bis 60 Kilowatt pro Rack an ihre physikalischen Grenzen. Da KI-Cluster diese Dichte bereits heute deutlich überschreiten, hat sich die Flüssigkeitskühlung – Liquid Cooling – innerhalb weniger Jahre von einer Nischenanwendung zum De-facto-Standard für KI- und HPC-Anwendungen entwickelt. Marktforscher gehen von einem rasant wachsenden weltweiten Markt für Liquid Cooling aus, der in den nächsten Jahren zweistellige Wachstumsraten verzeichnen dürfte

Im Wesentlichen haben sich drei Liquid-Cooling-Architekturen etabliert. Bei der Direct-to-Chip-Kühlung wird eine von Flüssigkeit durchströmte Kühlplatte direkt auf den Hauptwärmequellen wie CPU und GPU montiert; die Wärme wird über geschlossene Kühlkreisläufe und sogenannte Coolant Distribution Units (CDUs) effizient abgeführt. Single-Phase- oder Two-Phase-Immersionskühlung taucht komplette Server in eine elektrisch nichtleitende, dielektrische Flüssigkeit ein und nutzt entweder reine Wärmeleitung oder einen Phasenwechsel zur Wärmeabfuhr. Liquid-to-Air-Hybridlösungen kombinieren beide Welten und ermöglichen den schrittweisen Umbau bestehender, primär luftgekühlter Rechenzentren zu KI-fähigen Anlagen.

Die Vorteile der Flüssigkeitskühlung gegenüber der Luftkühlung sind erheblich. Wasser hat eine erheblich bessere Wärmeübertragungseigenschaft als Luft. Dadurch lassen sich höhere Vorlauftemperaturen (typischerweise 40 bis 52 Grad Celsius) realisieren, was zum einen die Energieeffizienz der Kühlung verbessert und zum anderen die Voraussetzungen für eine effektive Abwärmenutzung schafft. Die Geräuschemission sinkt deutlich, da weniger und kleinere Ventilatoren benötigt werden.

Auf der Anbieterseite haben sich neben den etablierten Spezialisten Vertiv, Rittal und Stulz internationale Player wie Schneider Electric (durch die Übernahme von Motivair), Eaton, ZutaCore und CoolIT positioniert. Rittal bietet beispielsweise CDUs mit 500 bis über 1.000 Kilowatt Kühlleistung an, Schneider/Motivair eine MCDU-70 mit 2,5 Megawatt zentralisierter Steuerung und nahtloser Skalierbarkeit über 10 Megawatt hinaus.

7.3 Stromversorgung und Resilienz

Die Stromversorgung ist die dominierende Variable bei Planung, Bau und Betrieb moderner Rechenzentren. Sie umfasst die externe Anbindung an das Übertragungs- und Verteilnetz, die unterbrechungsfreie Stromversorgung (USV) zur Überbrückung kurzer Spannungseinbrüche, die Notstromversorgung über Dieselaggregate für längere Ausfälle sowie zunehmend auch lokale Erzeugungs- und Speicherkonzepte.

Hyperscale-Rechenzentren benötigen heute typischerweise einen direkten Anschluss an das 110-kV-Hochspannungsnetz, große Anlagen mit mehr als 100 MW IT-Anschlussleistung an das 380-kV-Höchstspannungsnetz. Hier zeigen sich gravierende Engpässe: Nach Aussagen der Tennet TSO GmbH gibt es derzeit kein einziges freies Schaltfeld im gesamten Übertragungsnetzgebiet, sodass für jeden neuen Großanschluss zunächst Schaltanlagen erweitert werden müssen. Wartezeiten von vier bis sieben Jahren für einen Hochleistungsanschluss sind heute keine Ausnahme mehr.

Als Reaktion auf diese Engpässe gewinnen lokale Stromerzeugungskonzepte an Bedeutung. Photovoltaik-Anlagen auf Dächern und Freiflächen, Gasturbinen mit CO_2-Abscheidung, Blockheizkraftwerke, Brennstoffzellen sowie HVO-betriebene Notstromaggregate werden zunehmend zur Überbrückung von Anschlussverzögerungen und zur Stabilisierung der Versorgung eingesetzt. Internationale Vorreiter wie Microsoft setzen sogar auf Small Modular Reactors (SMR) als langfristige Stromquelle für ihre Rechenzentren – ein Konzept, das in Deutschland aufgrund des Atomausstiegs derzeit nicht realisierbar ist.

Für die unterbrechungsfreie Stromversorgung dominieren weiterhin batteriebasierte USV-Systeme, die zunehmend von Lithium-Ionen-Akkus mit höherer Energiedichte und längerer Lebensdauer als die früher üblichen Bleiakkus profitieren. Die Notstromversorgung erfolgt fast durchgängig über Diesel-Aggregate, deren Kraftstoffversorgung für mehrere Tage autonomen Betrieb gewährleistet ist. Eine zunehmende Anzahl neuer Rechenzentren stellt auf den biogenen Dieselersatz HVO (Hydrotreated Vegetable Oil) um, der CO_2-Einsparungen von bis zu 90 Prozent gegenüber konventionellem Diesel ermöglicht.

7.4 Konnektivität und Netzwerk-Infrastruktur

Die Leistungsfähigkeit eines Rechenzentrumsstandorts hängt maßgeblich von seiner Anbindung an das globale Glasfasernetz und an die zentralen Internet-Knotenpunkte ab. In Europa ist der DE-CIX in Frankfurt mit einem Rekordwert von 18,73 Terabit pro Sekunde Spitzenverkehr (2025) der mit Abstand größte Internet-Knoten. Er fungiert als zentrale Drehscheibe für den europäischen Datenverkehr und ist einer der Hauptgründe für die dominante Stellung Frankfurts im europäischen Rechenzentrumsmarkt. Weitere bedeutende Internet-Knoten in Deutschland sind der BCIX in Berlin, der Ecix in München und der KleyReX in Frankfurt.

Auf der überregionalen Ebene wird die Konnektivität durch ein dichtes Netz aus dedicated dark fiber-Verbindungen, Cloud-On-Ramps zu den großen Hyperscalern und privaten Interconnection-Diensten ergänzt. Frankfurt ist nach Angaben des Global Interconnection Index unter den FLAP-Märkten mit einer durchschnittlichen jährlichen Wachstumsrate von rund 50 Prozent der am schnellsten wachsende Standort für private Interconnection in Europa.

Innerhalb der Rechenzentren selbst hat sich die Netzwerk-Architektur in den letzten Jahren grundlegend verändert. Klassische dreistufige Topologien wurden durch flachere, hochgradig vernetzte Spine-Leaf-Architekturen ersetzt, die für die Ost-West-Verkehrsmuster moderner KI- und Big-Data-Workloads optimiert sind. Auf der physikalischen Ebene haben sich 400-Gigabit- und zunehmend 800-Gigabit-Ethernet sowie Infiniband-Lösungen mit bis zu 800 Gbps für KI-Cluster durchgesetzt. Die nächste Stufe – Ethernet mit 1,6 Terabit pro Sekunde – steht für 2026/2027 vor der Marktreife.

8 Energiebedarf und Nachhaltigkeit

8.1 Energieverbrauch der deutschen Rechenzentren

Der Energieverbrauch der deutschen Rechenzentren ist in den vergangenen anderthalb Jahrzehnten kontinuierlich gestiegen. Lag er 2010 noch bei 10,5 Terawattstunden pro Jahr, erreichte er 2020 bereits 16 TWh, 2024 rund 20 TWh und 2025 etwa 21,3 TWh. Damit entfällt auf die Branche aktuell ein Anteil von rund vier Prozent am gesamten deutschen Stromverbrauch. Der Verbrauch übersteigt die Jahresproduktion aller deutschen Wasserkraftwerke und entspricht etwa dem doppelten Stromverbrauch der Bundeshauptstadt Berlin.

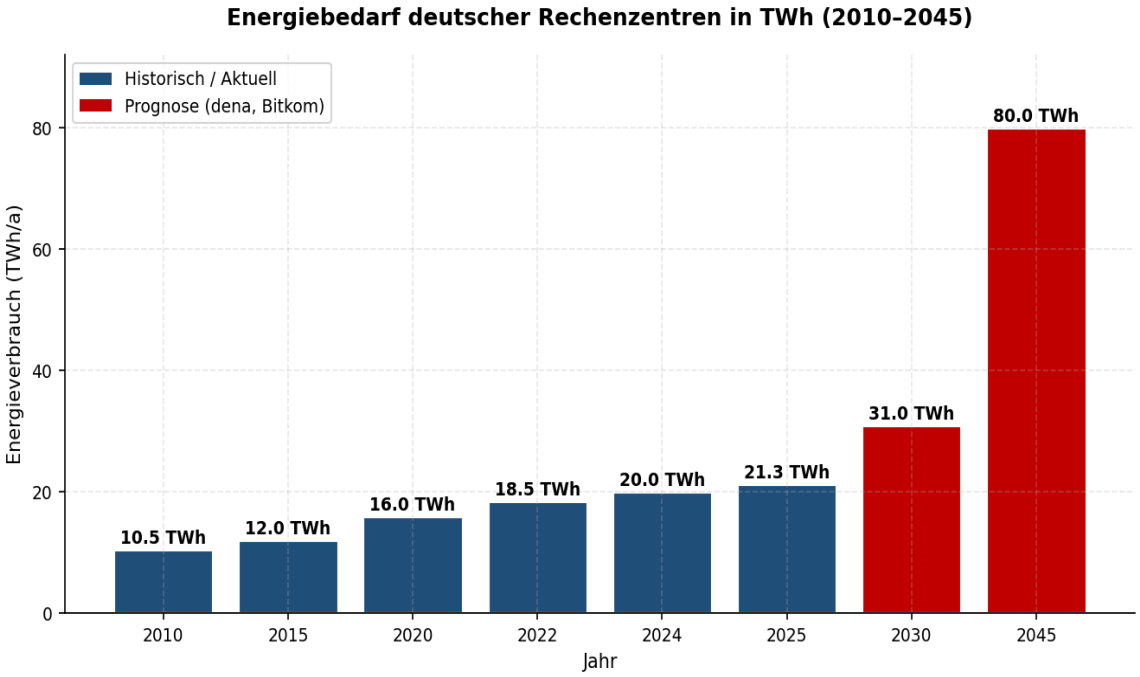


Abbildung 5: Energiebedarf deutscher Rechenzentren in TWh (2010–2045)
Quelle: Bitkom/Borderstep, dena, eigene Darstellung

Etwa zwei Drittel dieses Verbrauchs entfallen direkt auf die IT-Infrastruktur, also Server, Speicher und Netzwerktechnik. Das verbleibende Drittel verteilt sich auf die Gebäudeinfrastruktur, vor allem auf die Klimatisierung und Kühlung sowie auf die unterbrechungsfreie Stromversorgung. Diese Aufteilung verschiebt sich mit dem zunehmenden Einsatz von Liquid Cooling und besser konzipierten Wärmemanagementsystemen sukzessive zugunsten der IT-Anteile.

Bemerkenswert ist die parallele Entwicklung von absolutem Verbrauch und spezifischer Effizienz. Während der absolute Stromverbrauch sich seit 2010 verdoppelt hat, ist die in den Rechenzentren installierte Rechenkapazität pro verbrauchter Kilowattstunde Strom seit 2010 fast um den Faktor fünf gewachsen. Die Energieeffizienz bei Standardservern stieg laut bitkom in den Jahren 2017 bis 2022 um jährlich rund 26 Prozent. Diese Effizienzsteigerung wird allerdings durch die noch stärkere Steigerung der Nachfrage nach Rechenleistung mehr als kompensiert – ein klassischer Jevons-Effekt.

Für die Zukunft erwartet die dena einen Anstieg des Stromverbrauchs auf über 30 TWh bis 2030 und – bei gleichbleibender Wachstumsdynamik – auf knapp 80 TWh bis 2045. Diese Prognose impliziert, dass Rechenzentren mittelfristig zu den größten industriellen Stromverbrauchern Deutschlands aufrücken werden, vergleichbar mit der Aluminium- oder Stahlindustrie. In der politischen Debatte wird deshalb bereits vor einer Vervielfachung des Strombedarfs durch Rechenzentren bis 2045 gewarnt.

8.2 Power Usage Effectiveness (PUE) und Effizienz

Die zentrale Kennzahl zur Beurteilung der Energieeffizienz von Rechenzentren ist die Power Usage Effectiveness (PUE), definiert als das Verhältnis des Gesamtstromverbrauchs zum Stromverbrauch der IT-Infrastruktur. Ein PUE-Wert von 1,0 würde bedeuten, dass keinerlei Energie für Kühlung, USV oder andere Infrastruktur aufgewendet werden müsste – ein theoretischer Idealwert. Die durchschnittlichen PUE-Werte deutscher Rechenzentren haben sich in den letzten 15 Jahren erheblich verbessert: von rund 2,0 im Jahr 2010 auf 1,43 im Jahr 2025.

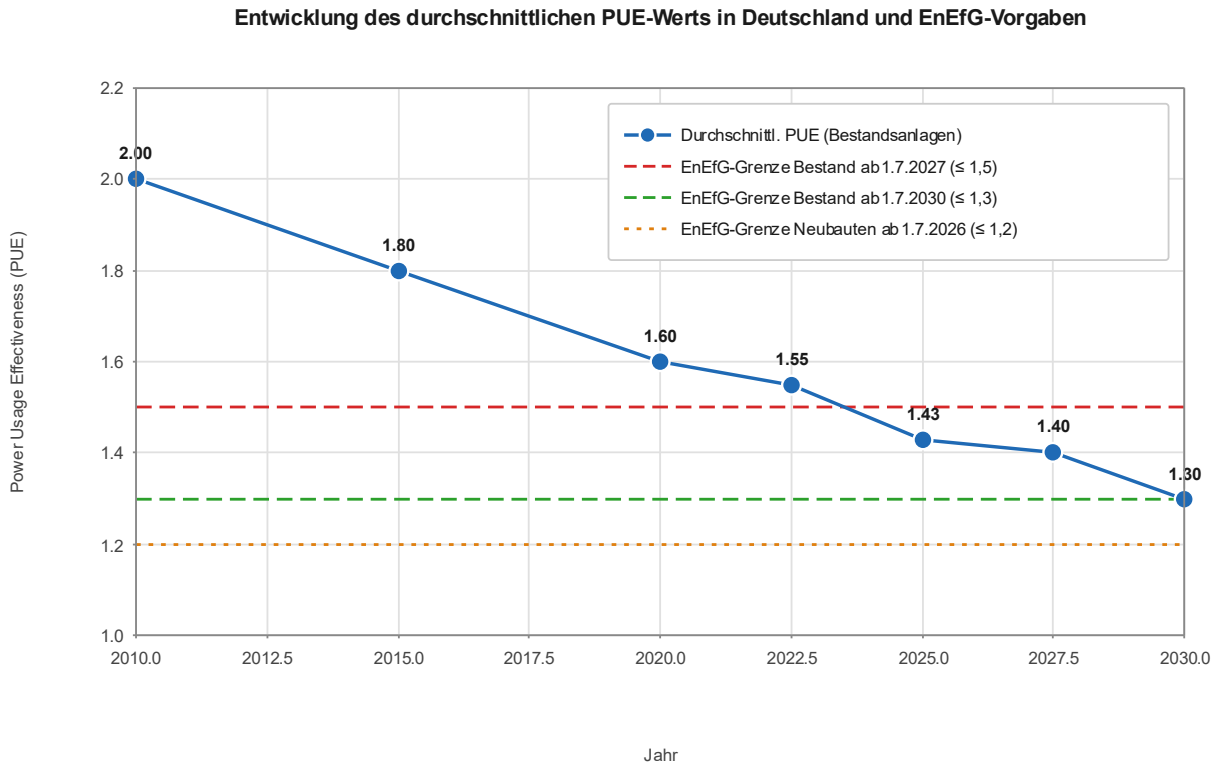


Abbildung 6: Entwicklung des durchschnittlichen PUE-Werts in Deutschland und gesetzliche Vorgaben des EnEfG
Quelle: Borderstep, EnEfG, eigene Darstellung

Diese Verbesserung wurde durch eine Vielzahl technischer Maßnahmen erreicht: konsequente Trennung von Kalt- und Warmgang, Anhebung der zulässigen Server-Eingangstemperaturen, Einsatz indirekter freier Kühlung, Optimierung der USV-Topologien und nicht zuletzt die Skaleneffekte großer Hyperscale-Anlagen, die typischerweise PUE-Werte zwischen 1,1 und 1,2 erreichen. Das im November 2023 in Kraft getretene Energieeffizienzgesetz (EnEfG) hat diese Entwicklung mit verbindlichen PUE-Vorgaben weiter beschleunigt: Neuanlagen, die nach dem 1. Juli 2026 in Betrieb gehen, müssen einen PUE von 1,2 oder besser erreichen, Bestandsanlagen ab Juli 2027 einen PUE von 1,5 und ab Juli 2030 von 1,3.

Der im April 2026 vom Bundesministerium für Wirtschaft und Energie vorgelegte Referentenentwurf zur Novelle des EnEfG sieht eine moderate Lockerung dieser Grenzwerte vor: Für Neuanlagen wird der zulässige PUE-Wert von 1,2 auf 1,3 angehoben, für Bestandsanlagen ab 2027 auf 1,6 und ab 2030 auf 1,4. Befürworter dieser Lockerung argumentieren, dass die ursprünglichen Werte technisch und wirtschaftlich nicht durchgängig realisierbar gewesen seien, insbesondere für KI-Rechenzentren, deren spezifische Lastprofile eine vollständige Optimierung erschweren. Kritiker – etwa die Deutsche Unternehmensinitiative Energieeffizienz (DENEFF) – sehen in der Lockerung einen Verstoß gegen die EU-Energieeffizienzrichtlinie und einen Rückschritt bei der Erreichung der Klimaziele.

8.3 Erneuerbare Energien und Stromherkunft

Das EnEfG schreibt vor, dass Rechenzentren ihren Stromverbrauch ab 2024 zu 50 Prozent und ab 2027 zu 100 Prozent bilanziell aus erneuerbaren Quellen decken müssen. Diese Vorgabe ist unter den Hyperscalern weitgehend unstrittig: AWS, Microsoft, Google und Meta haben seit Jahren öffentliche Zusagen zur 100-prozentigen Versorgung mit erneuerbarem Strom abgegeben und Power Purchase Agreements (PPAs) mit deutschen und europäischen Wind- und Solarparkbetreibern in erheblichem Umfang abgeschlossen.

Die bilanzielle Versorgung mit erneuerbarem Strom darf jedoch nicht mit einer physischen Versorgung verwechselt werden. Da Wind und Solar wetterabhängig fluktuieren, beziehen Rechenzentren in der Realität zu jedem Zeitpunkt den Strommix des öffentlichen Netzes. Erst durch den Erwerb von Herkunftsnachweisen oder durch synchronisierte PPAs wird die bilanzielle Bilanz auf 100 Prozent erneuerbar gestellt. Eine echte 24/7-Übereinstimmung von Verbrauch und erneuerbarer Erzeugung – das von Google verfolgte sogenannte 24/7 Carbon-Free Energy-Konzept – ist aktuell weder technisch noch ökonomisch flächendeckend umsetzbar.

Weltweit stammen nach Daten der Internationalen Energie-Agentur derzeit rund 27 Prozent des von Rechenzentren verbrauchten Stroms physisch aus erneuerbaren Energien, überwiegend aus Wind-, Solar- und Wasserkraft. Die Gesamtstromerzeugung aus erneuerbaren Energien soll bis 2030 jährlich um 22 Prozent wachsen und damit fast die Hälfte des prognostizierten Anstiegs des Rechenzentrums-Strombedarfs abdecken. Für die verbleibende Lücke werden Erdgas mit CO_2-Abscheidung, Brennstoffzellen und – international – Kernkraft (insbesondere Small Modular Reactors) als Brückenoptionen diskutiert.

8.4 Abwärmenutzung

Ein zentraler Hebel für die Nachhaltigkeit von Rechenzentren ist die Nutzung der entstehenden Abwärme. Die Server eines Rechenzentrums setzen praktisch die gesamte aufgenommene elektrische Energie als Wärme wieder frei – ein Potenzial, das in den meisten Fällen bislang ungenutzt in die Atmosphäre abgegeben wird. Das EnEfG hat hier mit verbindlichen Quoten erstmals einen Rahmen geschaffen: Neue Rechenzentren, die ab dem 1. Juli 2026 in Betrieb gehen, müssen mindestens 10 Prozent ihrer Abwärme nutzen oder einer wirtschaftlich verwertbaren Nutzung zuführen. Für Inbetriebnahmen ab Juli 2027 steigt diese Quote auf 15 Prozent, ab Juli 2028 auf 20 Prozent.

Die technische Umsetzung erfolgt typischerweise durch die Einspeisung der erwärmten Kühlflüssigkeit in lokale oder regionale Fernwärmenetze, durch direkte Wärmeversorgung benachbarter Gebäude (Bürokomplexe, Wohnhäuser, Gewächshäuser, Fischzuchten) oder – bei Niedertemperatur-Wärme – über Wärmepumpen, die das Temperaturniveau auf die für Heizungszwecke erforderlichen 65 bis 80 Grad Celsius anheben. Die zunehmende Verbreitung von Liquid Cooling, das Vorlauftemperaturen von 50 bis 52 Grad Celsius ermöglicht, macht die Abwärmenutzung deutlich wirtschaftlicher als bei luftgekühlten Anlagen, deren Abwärme typischerweise bei 25 bis 35 Grad Celsius anfällt.

Die praktische Umsetzung steht jedoch vor erheblichen Hürden. Die meisten Rechenzentren befinden sich nicht in der unmittelbaren Nähe von Fernwärmenetzen, und der Bau einer Anbindungsleitung über mehrere Kilometer ist ohne öffentliche Förderung selten wirtschaftlich. Hinzu kommen die saisonale Schwankung der Wärmenachfrage (im Sommer ist die Nachfrage minimal) sowie die unterschiedlichen technischen Standards. Die Novelle des EnEfG sieht daher eine Ausnahme von der Abwärmenutzungspflicht vor, wenn im Umkreis von fünf Kilometern keine technisch und wirtschaftlich zumutbare Anschlussmöglichkeit besteht.

Vorzeigeprojekte zeigen jedoch, was technisch möglich ist. Das Frankfurter Cloud HQ-Projekt im Industriepark Höchst nutzt die Synergie mit den dort betriebenen 120 Produktions- und Verbrennungsanlagen. Das geplante Hyperscale-Rechenzentrum in Frankfurt (Oder) wird einen Zuschuss von 15 Millionen Euro für die Anbindung an das Fernwärmenetz der Stadtwerke leisten. In Stockholm, Helsinki und Kopenhagen sind bereits seit Jahren Modelle etabliert, bei denen Rechenzentren einen erheblichen Beitrag zur Fernwärmeversorgung der Innenstädte leisten – ein Vorbild, an dem sich Deutschland orientieren sollte.

8.5 Wasserverbrauch und WUE

Ein zunehmend in den Fokus rückendes Nachhaltigkeitsthema ist der Wasserverbrauch von Rechenzentren. Während luftgekühlte Anlagen praktisch kein Wasser verbrauchen, benötigen verdunstungsbasierte Kühltürme erhebliche Mengen an Frischwasser. Die Water Usage Effectiveness (WUE) – das Verhältnis des Wasserverbrauchs zur IT-Energie in Litern pro Kilowattstunde – liegt bei großen Hyperscale-Anlagen typischerweise zwischen 0,5 und 2,0 l/kWh. In Regionen mit Wasserknappheit – Spanien, Italien, Teile Frankreichs – ist dieser Verbrauch zunehmend Gegenstand öffentlicher Kritik.

In Deutschland ist die Wasserversorgung in den meisten Regionen weniger kritisch, doch auch hier setzen Betreiber zunehmend auf wassersparende Konzepte. Die Bundesregierung hat in der Nationalen Rechenzentrumsstrategie wassersparende Kühlsysteme als zentralen Hebel für die Nachhaltigkeit identifiziert. Adiabate Kühlsysteme mit deutlich reduziertem Wasserverbrauch, geschlossene Kühlkreisläufe und der Einsatz von aufbereitetem Brauchwasser oder Regenwasser werden zunehmend Standard. Das in Mainz auf der Ingelheimer Aue entstehende Rechenzentrum soll nach Level-3-Standards zu einem der zuverlässigsten und umweltfreundlichsten in Europa werden.

9 Regulatorischer Rahmen: EnEfG und Rechenzentrumsstrategie

9.1 Energieeffizienzgesetz (EnEfG)

Das Energieeffizienzgesetz, kurz EnEfG, ist seit dem 18. November 2023 in Kraft und stellt das zentrale Bundesgesetz zur Steigerung der Energieeffizienz in Deutschland dar. Es setzt die EU-Energieeffizienzrichtlinie (EED) in nationales Recht um und enthält in Abschnitt 4 spezifische Vorgaben für Rechenzentren. Diese Vorgaben gelten grundsätzlich für alle Rechenzentren mit einer nicht redundanten elektrischen Nennanschlussleistung ab 300 Kilowatt; strengere Pflichten greifen ab einer Anschlussleistung von 1 Megawatt.

Die zentralen Anforderungen des EnEfG für Rechenzentren lassen sich in fünf Kategorien gliedern:

- Energieverbrauchseffektivität (PUE): Verbindliche Grenzwerte je nach Inbetriebnahmezeitpunkt, abgestuft zwischen Bestand- und Neuanlagen.
- Anteil wiederverwendeter Energie (ERF): Pflicht zur Abwärmenutzung mit gestaffelten Mindestanteilen von 10 bis 20 Prozent für Neuanlagen.
- Bilanzielle Stromversorgung aus erneuerbaren Quellen: 50 Prozent ab 2024, 100 Prozent ab 2027.
- Energie- oder Umweltmanagementsystem nach ISO 50001 oder EMAS, Zertifizierungspflicht ab 1. Januar 2026.
- Meldepflicht: Jährliche Übermittlung von Daten zu Energieverbrauch, PUE-Werten, Kühlsystemen und Abwärmenutzung an das Energieeffizienzregister beim BAFA bzw. der BfEE.

PUE-Grenzwerte und Abwärmequoten im Detail

Inbetriebnahme	PUE-Vorgabe (urspr.)	PUE-Vorgabe (Novelle)	Abwärmequote (ERF)
Bestand vor 1.7.2026	≤ 1,5 ab 1.7.2027	≤ 1,6 ab 1.7.2027	Kosten-Nutzen-Analyse
Bestand vor 1.7.2026	≤ 1,3 ab 1.7.2030	≤ 1,4 ab 1.7.2030	Kosten-Nutzen-Analyse
Neuanlagen ab 1.7.2026	**≤ 1,2**	**≤ 1,3**	min. 10 %
Neuanlagen ab 1.7.2027	**≤ 1,2**	**≤ 1,3**	min. 15 %
Neuanlagen ab 1.7.2028	**≤ 1,2**	**≤ 1,3**	min. 20 %

Tabelle 5: PUE- und Abwärmequoten-Vorgaben nach EnEfG (Stand vor und nach geplanter Novelle April 2026)
Quelle: § 11 EnEfG, BMWE-Referentenentwurf 9.4.2026

9.1.1 Die Novelle des EnEfG

Am 9. April 2026 hat das Bundesministerium für Wirtschaft und Energie einen Referentenentwurf zur Novellierung des EnEfG veröffentlicht. Ziel der Überarbeitungen ist es, auf die signifikant gestiegene Bedeutung energieintensiver Infrastrukturen zu reagieren und den hohen Umsetzungsaufwand für Unternehmen zu reduzieren. Der Entwurf sieht im Wesentlichen folgende Anpassungen vor: eine Anhebung der PUE-Grenzwerte um 0,1 für alle Inbetriebnahmekategorien; eine Lockerung der Abwärmenutzungspflicht durch eine neue Ausnahme bei fehlender Anschlussmöglichkeit im Umkreis von fünf Kilometern; eine Umstellung der Abwärmevermeidungspflicht auf eine Kosten-Nutzen-Analyse bei Planung oder erheblicher Modernisierung; sowie eine Anhebung der Schwellenwerte für verpflichtende Energiemanagementsysteme von 7,5 GWh auf 23,6 GWh Endenergieverbrauch pro Jahr.

Die Branche hat die geplanten Änderungen weitgehend begrüßt, da sie eine deutliche Reduzierung des Umsetzungsaufwands bedeuten. Insbesondere für KI-Rechenzentren mit ihren spezifischen Lastprofilen sind die ursprünglichen Grenzwerte schwer einzuhalten. Kritik kommt von Umweltverbänden und der DENEFF, die eine Verletzung der EU-Vorgaben sehen. Mit einer Verabschiedung der Novelle ist voraussichtlich erst im dritten oder vierten Quartal 2026 zu rechnen, sodass für Neuanlagen, die bis dahin in Betrieb gehen, weiterhin die ursprünglichen Grenzwerte gelten.

9.2 Nationale Rechenzentrumsstrategie 2026

Das Bundeskabinett hat am 18. März 2026 die erste Nationale Rechenzentrumsstrategie verabschiedet. Sie geht zurück auf einen Beteiligungsprozess, den das Bundesministerium für Digitales und Staatsmodernisierung (BMDS) im August 2025 gestartet hatte, und verbindet die bestehenden Initiativen mit konkreten neuen Maßnahmen. Die Strategie wurde im Bundestag am 16. April 2026 erstmals debattiert und an den Ausschuss für Digitales und Staatsmodernisierung überwiesen.

Ziel der Strategie ist es, die Rechenzentrumskapazitäten in Deutschland bis 2030 mindestens zu verdoppeln und die Rechenkapazitäten für High-Performance-Computing und Künstliche Intelligenz mindestens zu vervierfachen. Sie umfasst 28 Maßnahmen in drei zentralen Handlungsfeldern:

- Energie und Nachhaltigkeit: Förderung nachhaltiger Rechenzentren mit erneuerbarem Strom, energieeffizienten Lösungen und wassersparenden Kühlsystemen. Schaffung verlässlicher und bezahlbarer Stromversorgung.
- Standort und Fläche: Beschleunigung von Planungs- und Genehmigungsverfahren, Ausweisung von Vorzugsflächen, Anpassung der Netzanschlussverfahren und Einführung von Reservierungsmechanismen für Netzanschlusskapazitäten.
- Technologie und Souveränität: Aufbau hochleistungsfähiger KI-Rechenkapazitäten, Unterstützung der Ansiedlung mindestens einer europäischen KI-Gigafabrik in Deutschland, Stärkung von HPC- und Forschungsnetzwerken sowie Aufbau einer souveränen Cloud-Plattform für Behörden-KI im Rahmen des Deutschland-Stacks.

9.2.1 Netzpaket und Netzanschlussverfahren

Eng verzahnt mit der Rechenzentrumsstrategie ist das im März 2026 als Arbeitsentwurf veröffentlichte Netzpaket, das verschiedene Änderungen im Energiewirtschaftsgesetz (EnWG) vorsieht. Zentrale Elemente sind: die Möglichkeit zur Priorisierung und Freihaltung von Netzanschlusskapazitäten durch die Übertragungsnetzbetreiber (§ 17b EnWG-Entwurf), die digitale Ausweisung freier Kapazitäten im Verteilernetz (§ 17c EnWG-Entwurf) sowie ein vereinheitlichter Reservierungsmechanismus für Netzanschlusskapazitäten (§ 17f EnWG-Entwurf).

Trotz dieser Maßnahmen wird ein wesentliches strukturelles Problem von Branchenexperten als nicht ausreichend adressiert angesehen: Auf Übertragungsnetzebene fehlt es nach Aussagen der Tennet TSO GmbH derzeit nicht primär an Netzanschlusskapazitäten, sondern an Anschlussmöglichkeiten – konkret an freien Schaltfeldern in den bestehenden Umspannwerken. Der Bau neuer Schaltanlagen oder die Erweiterung bestehender Anlagen erfordert mehrjährige Genehmigungs- und Bauprozesse.

9.3 EU-Vorgaben und internationale Rahmensetzung

Über das nationale Recht hinaus prägen verschiedene EU-Regelungen die Rahmenbedingungen für deutsche Rechenzentren. Die EU-Energieeffizienzrichtlinie (EED) bildet die Grundlage für das EnEfG. Die EU-Taxonomie definiert Kriterien für nachhaltige Investitionen, die zunehmend Einfluss auf die Finanzierungskosten von Rechenzentrumsprojekten haben. Die EU-Verordnung über digitale Betriebsstabilität (DORA) verpflichtet Finanzinstitute zur Resilienz ihrer IT-Infrastrukturen. Die Datenschutz-Grundverordnung (DSGVO) bleibt der Goldstandard für den Umgang mit personenbezogenen Daten.

Mit dem 2024 verabschiedeten EU AI Act treten zudem für KI-Systeme spezifische Anforderungen in Kraft, die sich indirekt auf die Rechenzentrumslandschaft auswirken. Die Einstufung bestimmter KI-Modelle als „General Purpose AI with Systemic Risk“ ist mit zusätzlichen Transparenz-, Audit- und Risikomanagementpflichten verbunden, deren operative Umsetzung in der Regel im Rechenzentrum erfolgt.

Eine zentrale neue Initiative auf EU-Ebene ist die InvestAI-Fazilität mit einem Volumen von 20 Milliarden Euro für die Errichtung von bis zu fünf europäischen KI-Gigafabriken. Diese werden im Rahmen der EuroHPC Joint Undertaking ausgeschrieben und sollen ab 2028 in Betrieb gehen. Deutschland hat im Sondervermögen 805 Millionen Euro für einen Finanzierungsbeitrag zu einer KI-Gigafabrik in Deutschland reserviert. Die Standortauswahl wird voraussichtlich im Sommer 2026 entschieden.

10 Künstliche Intelligenz als Wachstumstreiber

10.1 KI-Rechenzentren in Deutschland

Künstliche Intelligenz ist im Jahr 2026 unstrittig der dominierende Wachstumstreiber des deutschen Rechenzentrumsmarkts. Aktuell entfallen rund 15 Prozent der gesamten installierten IT-Anschlussleistung in Deutschland auf KI-Rechenzentren – rechnerisch entspricht dies auf Basis von 2.980 MW Gesamtleistung etwa 450 Megawatt. Bis 2030 soll sich diese Kapazität auf rund 2.020 Megawatt vervierfachen, was dann einem Anteil von 40 Prozent an der Gesamtkapazität entsprechen würde.

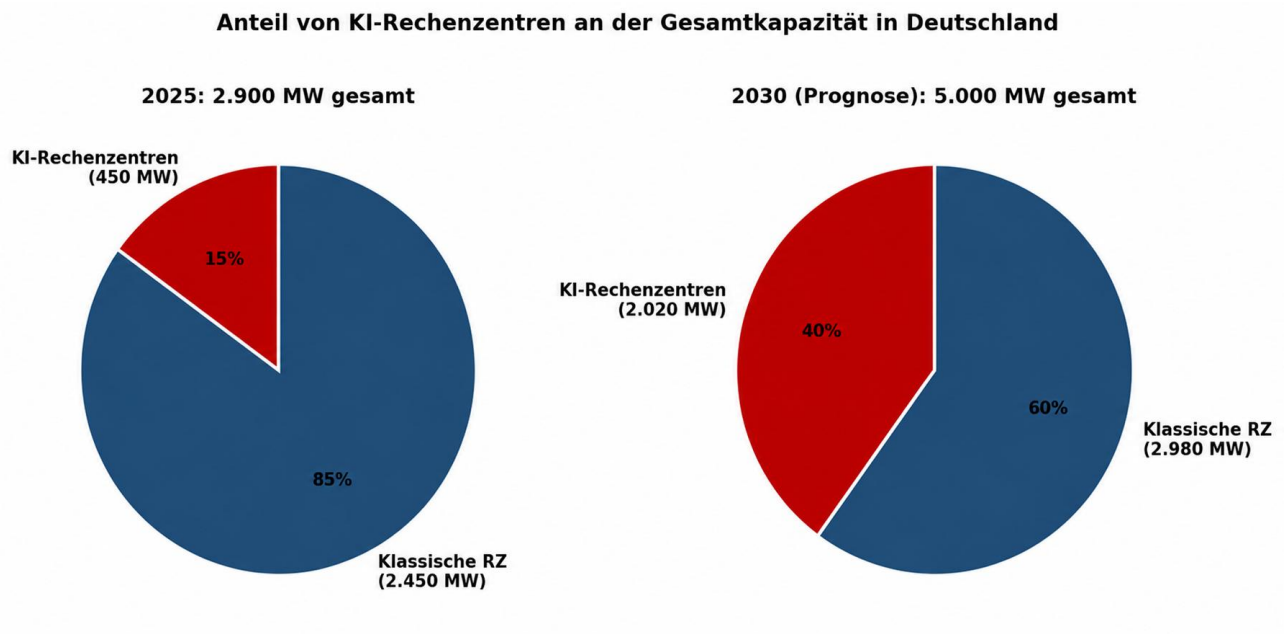


Abbildung 7: Anteil von HPC-/KI-Kapazitäten an der Gesamtkapazität in Deutschland 2025 vs. Prognose 2030
Quelle: Bitkom-Studie 2025, eigene Darstellung

KI-Rechenzentren unterscheiden sich in mehrfacher Hinsicht von klassischen Cloud-Rechenzentren. Auf der Hardware-Seite dominieren spezialisierte GPU-Beschleuniger – aktuell vor allem die NVIDIA-Familie H100, H200, B100, B200 und GB200 – sowie zunehmend Konkurrenzprodukte von AMD (Instinct MI300, MI325), Intel (Gaudi 3, Gaudi 4) und kundeneigene ASICs (Google TPU, Amazon Trainium/Inferentia). Diese Beschleuniger weisen eine extrem hohe Leistungsdichte auf und benötigen entsprechende Kühl-, Strom- und Netzwerk-Infrastruktur.

KI-Workloads zerfallen typischerweise in zwei Hauptkategorien: Training (das einmalige oder periodisch wiederholte Erlernen eines Modells aus großen Datenmengen) und Inferenz (die laufende Anwendung des trainierten Modells auf neue Eingaben). Während Training extrem rechen- und energieintensiv ist und in großen, dedizierten Clustern erfolgt, kann Inferenz dezentral – auch in kleineren Rechenzentren oder am Edge – stattfinden. Mit der Einführung der nächsten Generation großer Sprachmodelle, die teils zehn- bis hundertfach mehr Parameter haben werden als die aktuellen Spitzenmodelle, wird der Trainingsaufwand weiter explodieren.

10.2 EuroHPC AI Factories und Gigafactories

Die Europäische Union hat mit dem EuroHPC Joint Undertaking eine zentrale Plattform für die Förderung von Hochleistungs- und KI-Rechenkapazitäten geschaffen. Ergänzend zur bereits vorhandenen Supercomputer-Infrastruktur werden seit 2024 sogenannte AI Factories aufgebaut – KI-optimierte Hochleistungsrechenzentren, die in Verbindung mit Datenbanken, Beratungs- und Schulungsangeboten europäischen Start-ups, KMU und Forschern den Zugang zu modernster KI-Infrastruktur ermöglichen sollen.

Im Dezember 2024 wurden sieben erste AI Factories ausgewählt – darunter mit der „Jupiter AI Factory" am Forschungszentrum Jülich auch ein deutscher Standort. Die JUPITER ist Europas erster Supercomputer der Exascale-Klasse und verfügt über mehr als 24.000 NVIDIA GH200 Superchips. Im März 2025 folgten sechs weitere AI Factories in Österreich, Bulgarien, Frankreich, Polen, Slowenien und einem zweiten deutschen Standort: dem vom Höchstleistungsrechenzentrum Stuttgart koordinierten „Hammer HAI". Im Oktober 2025 wurden sechs weitere AI Factories in Tschechien, Litauen, den Niederlanden, Rumänien, Spanien und Polen sowie 13 sogenannte AI Factory Antennas in weiteren Mitgliedsstaaten und assoziierten Ländern angekündigt.

Die nächste Stufe des europäischen KI-Infrastrukturausbaus bilden die KI-Gigafabriken (AI Gigafactories). Diese werden mit drei- bis viermal so vielen Beschleunigern ausgestattet sein wie die aktuellen AI Factories – die EU-Kommission spricht von rund 100.000 KI-Spezialprozessoren pro Gigafabrik – und sind explizit auf die Entwicklung der nächsten Generation hochmoderner KI-Modelle ausgerichtet. Die Investitionskosten je Gigafabrik werden auf vier bis fünf Milliarden Euro geschätzt, wovon die EU bis zu 35 Prozent über die InvestAI-Fazilität fördert. Das offizielle Bewerbungsverfahren startet im ersten Quartal 2026, die Auswahl der bis zu fünf Standorte ist für Sommer 2026 vorgesehen.

Bis zum Ende der Interessenbekundungsphase im Juni 2025 sind 76 Bewerbungen aus 16 EU-Mitgliedsstaaten für 60 verschiedene Standorte eingegangen. Deutschland ist mit mehreren Konsortien vertreten, an denen unter anderem die Deutsche Telekom, die Schwarz-Gruppe (Schwarz Digits) und Ionos beteiligt sind. Im Koalitionsvertrag der 21. Legislaturperiode wurde festgehalten, dass Deutschland Standort für mindestens eines der geplanten KI-Zentren sein soll. Die Bundesregierung hat im Sondervermögen 805 Millionen Euro für einen deutschen Finanzierungsbeitrag bereitgestellt. Mit dem Bau der ersten KI-Gigafactories ist frühestens ab 2028 zu rechnen.

10.3 Souveränitäts- und Industrial-AI-Initiativen

Parallel zum EU-getriebenen Aufbau der KI-Gigafabriken haben sich in Deutschland mehrere Initiativen für industrielle und souveräne KI-Infrastruktur etabliert. Die im Herbst 2025 zwischen der Deutschen Telekom und NVIDIA angekündigte „Industrial AI Cloud" in München ist das prominenteste Beispiel: Mit einem Investitionsvolumen von rund einer Milliarde Euro und einer eigenständigen Betreibergesellschaft mit europäischem Personal soll sie speziell auf die Bedürfnisse der europäischen Industrie zugeschnitten sein. SAP ist erster Großkunde.

Die Schwarz Digits, der IT-Arm der Schwarz-Gruppe (Lidl, Kaufland), hat sich mit der Schwarz Cloud als europäischer Cloud-Anbieter positioniert und betreibt eigene KI-Rechenzentren mit deutscher Datenresidenz. Ionos, der größte deutsche Webhosting-Anbieter, hat mit der „Ionos Cloud" eine wettbewerbsfähige Alternative zu den US-Hyperscalern aufgebaut, die zunehmend auch KI-Workloads bedient. Das deutsche Start-up Aleph Alpha aus Heidelberg entwickelt mit dem Sprachmodell „Pharia" eine konkurrenzfähige europäische Alternative zu OpenAI und Anthropic, deren Training auf der heimischen Infrastruktur erfolgt.

Auf der Forschungsseite koordiniert das Forschungszentrum Jülich mit der JUPITER AI Factory die deutschen Beiträge zur europäischen KI-Infrastruktur. Zu den Partnern gehören das Center für Künstliche Intelligenz der RWTH Aachen, die Fraunhofer-Institute für Angewandte Informationstechnik und für Intelligente Analyse- und Informationssysteme sowie das Hessian Center for Artificial Intelligence. Das Höchstleistungsrechenzentrum Stuttgart koordiniert mit „Hammer HAI" einen weiteren deutschen Standort der EuroHPC-Initiative.

11 Herausforderungen und Risiken

11.1 Strompreise und Energiekosten

Die mit Abstand größte strukturelle Herausforderung für den Standort Deutschland sind die im internationalen Vergleich extrem hohen Strompreise. Industriestrompreise von über 15 Cent pro Kilowattstunde liegen in Deutschland um das Zwei- bis Dreifache über den Werten konkurrierender Standorte wie den USA (5–8 ct/kWh), Skandinavien (4–7 ct/kWh) oder bestimmten Regionen Frankreichs (Atomstromangebote unter 6 ct/kWh). Da die Energiekosten nach Angaben der German Datacenter Association (GDA) etwa 50 Prozent der Betriebsausgaben eines Rechenzentrums ausmachen, schlägt dieser Standortnachteil unmittelbar auf die Wettbewerbsfähigkeit durch.

Eine zentrale Forderung der Branche ist daher die Einbeziehung von Rechenzentren in die Entlastungsmaßnahmen für energieintensive Industrien, etwa die Strompreiskompensation oder den geplanten Industriestrompreis. Bislang werden Rechenzentren – im Gegensatz zur klassischen Industrie – nicht als produzierendes Gewerbe anerkannt und sind von vielen Entlastungsregelungen ausgeschlossen. Ein vom Bitkom in Auftrag gegebenes Rechtsgutachten kommt zu dem Schluss, dass die europäischen Beihilferahmen die Einbeziehung von Rechenzentren als digitale Infrastruktur grundsätzlich erlauben würden.

11.2 Stromnetzanschluss und Verfahren

Die Verfügbarkeit von Stromnetzanschlüssen entwickelt sich zur zweiten zentralen Engpassgröße. Insbesondere im Übertragungsnetz, das für Hyperscale-Anlagen mit mehr als 100 MW Anschlussleistung zwingend erforderlich ist, stehen nach Aussagen der Tennet TSO GmbH derzeit kaum noch freie Schaltfelder zur Verfügung. Die Wartezeit für einen neuen 380-kV-Anschluss kann vier bis sieben Jahre betragen – ein Zeithorizont, der für viele Investitionsentscheidungen prohibitiv lang ist.

Die Bundesregierung adressiert diese Herausforderung mit dem im März 2026 vorgelegten Netzpaket und dem in der Rechenzentrumsstrategie verankerten Reservierungsmechanismus. Branchenexperten kritisieren jedoch, dass diese Maßnahmen das eigentliche Problem – die fehlende physische Schaltfeld-Infrastruktur – nicht lösen. Eine Beschleunigung der Genehmigungsverfahren für Schaltanlagen und Trafostationen wäre ebenso erforderlich wie die finanzielle Unterstützung der Übertragungsnetzbetreiber beim Ausbau der Knotenpunkte.

11.3 Genehmigungsverfahren und Bürokratie

Neben dem Stromnetzanschluss bremsen langwierige Genehmigungsverfahren den Ausbau der Rechenzentrumskapazitäten. Die Bauleitplanung, der Immissionsschutz, der Brandschutz und der Wasserrechtsbescheid erfordern in vielen Bundesländern Verfahrensdauern von 18 bis 36 Monaten, in Einzelfällen deutlich länger. Hinzu kommen Verfahren nach BImSchG für die Genehmigung von Notstromaggregaten und Kühltürmen, die je nach Anlagentyp und Standort weitere 12 bis 18 Monate beanspruchen können.

Die Nationale Rechenzentrumsstrategie sieht im Handlungsfeld „Standort und Fläche“ eine Reihe von Maßnahmen zur Beschleunigung der Verfahren vor: Ausweisung von Vorzugsflächen durch die Länder, Vereinheitlichung von Genehmigungsverfahren, Digitalisierung der Antragsstellung und gebündelte Behördenzuständigkeiten. Die konkrete Umsetzung obliegt jedoch den Ländern und Kommunen, sodass die Effektivität dieser Maßnahmen erst in den nächsten Jahren überprüfbar sein wird.

11.4 Fachkräftemangel

Der Fachkräftemangel ist eine branchenübergreifende Herausforderung, die Rechenzentren in besonderer Weise trifft. Der Betrieb moderner Rechenzentren erfordert eine breite Palette spezialisierter Qualifikationen: Elektroingenieure für die Stromversorgung, Versorgungstechniker für die Klimatisierung, Netzwerkspezialisten, Sicherheitsexperten, Liquid-Cooling-Spezialisten, Cloud-Architekten und KI-Infrastruktur-Engineers. In Europa beschäftigen Rechenzentren rund 600.000 Vollzeitäquivalente; bis 2025 wurde ein Anstieg um 14 Prozent erwartet, der bis 2030 noch deutlich übertroffen werden dürfte.

Für Deutschland ist schon heute ein erheblicher Fachkräftemangel feststellbar. Die geplante Verdoppelung der Kapazitäten bis 2030 setzt eine entsprechende Erweiterung der personellen Basis voraus, die ohne gezielte Aus- und Weiterbildungsinitiativen sowie eine erleichterte Zuwanderung qualifizierter Fachkräfte nicht zu leisten ist. Die Branche fordert die Etablierung eines spezialisierten Berufsbildes „Rechenzentrumstechniker“ sowie Investitionen in die Modernisierung der Curricula in den relevanten Studiengängen.

11.5 Gesellschaftliche Akzeptanz und Standortkonflikte

Die zunehmende Größe und Sichtbarkeit von Rechenzentren führt in einigen Regionen zu wachsenden Akzeptanzproblemen. Bürgerinitiativen kritisieren den hohen Energie- und Wasserverbrauch, die Geräuschemissionen der Notstromaggregate und Kühlanlagen, die optische Beeinträchtigung des Landschaftsbildes sowie die häufig geringe lokale Wertschöpfung pro Quadratmeter Fläche. In Amsterdam hat eine entsprechende Bürgerbewegung 2019 zu einem zeitweisen Moratorium für neue Rechenzentren geführt, dessen Effekte sich bis heute auswirken.

In Deutschland sind bislang vergleichbare Moratorien nicht erlassen worden, doch zeichnen sich in einigen Kommunen verschärfte Ansiedlungsregeln ab. Die Stadt Frankfurt hat beispielsweise mit ihrem Rechenzentrenkonzept klare Vorgaben für die Standortwahl getroffen, um Konflikte mit Wohngebieten zu vermeiden. Die proaktive Einbindung der lokalen Bevölkerung, transparente Informationen über Energie- und Wasserverbrauch sowie konkrete Beiträge zur lokalen Wertschöpfung – etwa über Abwärmenutzung oder Gewerbesteuern – sind zentrale Voraussetzungen für die anhaltende Akzeptanz.

11.6 Cybersicherheit und physische Sicherheit

Mit der zunehmenden Konzentration kritischer Daten und Anwendungen in Rechenzentren wächst auch das Risiko von Cyberangriffen, Sabotage und physischen Angriffen. Die EU-NIS2-Richtlinie und das deutsche IT-Sicherheitsgesetz 2.0 stellen erhöhte Anforderungen an die Resilienz, die durch das ab 2025 anwendbare DORA-Regelwerk für den Finanzsektor weiter verschärft werden. Hinzu kommen geopolitische Risiken: Die Sabotage von Seekabeln in der Ostsee 2023 und 2024 hat die Verwundbarkeit kritischer Datenverbindungen unterstrichen.

11.7 Internationale Konkurrenz und Datenkolonialisierung

Der schärfste Wettbewerb für den Standort Deutschland kommt aus den USA und China. Die USA verfügen 2025 über zehnmal so viele Rechenzentrumskapazitäten wie für Deutschland bis 2030 geplant sind. In den USA lagen sowohl Bestand als auch jährlicher Neuzubau Ende 2025 um den Faktor 16 bis 17 über dem deutschen Niveau: Laut FERC waren in den USA Ende 2025 mehr als 50 GW in Betrieb[1] während in Deutschland 2,98 GW installiert waren. Der jährliche Neuzubau betrug in den USA 2025 rund 4,0 GW (Colocation-Segment; Cushman & Wakefield)[2] gegenüber 250 MW in Deutschland. Bemerkenswert ist dabei auch die durchschnittliche Projektgröße: Neu in Betrieb gehende US-Rechenzentren wiesen 2025 eine mittlere Kapazität von knapp 80 MW auf – gegenüber deutlich kleineren Projekten im deutschen Markt. Beide Kennzahlen sind auf Basis installierter IT-Anschlussleistung direkt vergleichbar; Unterschiede in der PUE und der statistischen Erfassungsgrenze (US: alle Größenklassen; Deutschland: ab 100 kW) sind zu beachten. Auch China baut seine Kapazitäten in einem Tempo aus, das Europa weit hinter sich lässt. Bitkom-Hauptgeschäftsführer Bernhard Rohleder warnte in diesem Zusammenhang, Deutschland riskiere zur „Datenkolonie“ zu werden, wenn nicht massiv in eigene Kapazitäten investiert werde.

[1] **FERC, State of the Markets Report, März 2026**

[2] **Cushman & Wakefield, Americas Data Center Update H2 2025, Februar 2026**

12 Internationaler Vergleich

12.1 Globale Marktstruktur

Der globale Rechenzentrumsmarkt ist hochgradig konzentriert auf wenige Länder. Die USA dominieren mit einer installierten Kapazität von über 50 Gigawatt – mehr als sechszehnmal so viel wie Deutschland. Hinter den USA folgen China mit etwa 12 GW, Japan und Großbritannien mit jeweils 1,5 bis 1,7 GW sowie Deutschland mit knapp 3 GW. In Europa rangiert Deutschland damit unangefochten an der Spitze, gefolgt von Großbritannien, Frankreich (ca. 950 MW), den Niederlanden (ca. 870 MW) und Irland (ca. 700 MW).

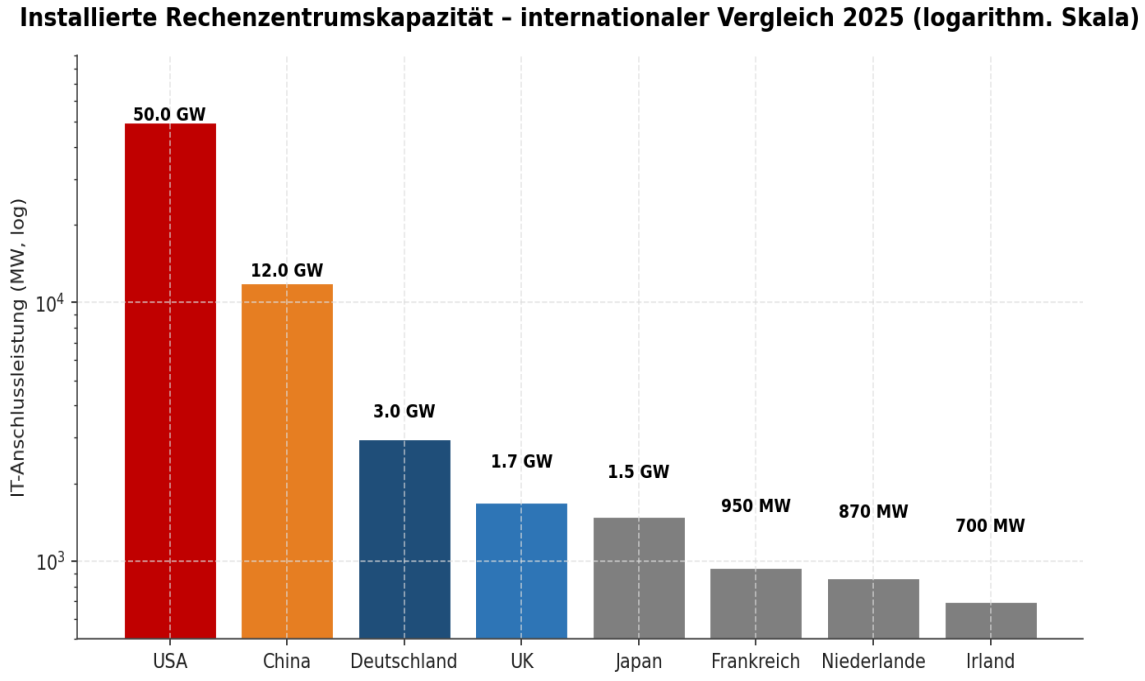


Abbildung 8: Installierte Rechenzentrumskapazität – internationaler Vergleich 2025 (in MW, logarithmische Skala)
Quelle: FERC, Synergy Research, Bitkom, eigene Auswertung

12.2 Standortvergleich der FLAP-D-Region

Innerhalb Europas dominiert die sogenannte FLAP-D-Region das Marktgeschehen: Frankfurt, London, Amsterdam, Paris und Dublin bilden zusammen mehr als 70 Prozent der europäischen Rechenzentrumskapazitäten. Die Standorte unterscheiden sich jedoch erheblich in ihren Wachstumsraten, Leerstandsquoten und regulatorischen Rahmenbedingungen.

Standort	Kapazität (MW)	Wachstum p.a.	Leerstand	Spezifische Standortmerkmale
Frankfurt	ca. 1.131	**ca. 14 %**	ca. 3 %	DE-CIX, Finanzplatz, etabl. Ökosystem
London	ca. 1.100	ca. 9 %	ca. 4 %	Brexit-Effekte, hohe Strompreise
Amsterdam	ca. 870	ca. 6 %	ca. 5 %	AMS-IX, Moratorium-Folgen, Wassermangel
Paris	ca. 950	ca. 12 %	ca. 4 %	Atomstrom, KI-Strategie, Förderung
Dublin	ca. 700	ca. 4 %	ca. 6 %	Stromnetzengpässe, Steuervorteile

Tabelle 6: Vergleich der FLAP-D-Standorte in Europa (Stand 2025)
Quelle: CBRE, JLL, Cushman & Wakefield, eigene Auswertung

12.3 Aufstrebende europäische Standorte

Außerhalb der FLAP-D-Region etablieren sich seit einigen Jahren neue Standorte, die teils erhebliche Standortvorteile bieten. Die skandinavischen Länder – Schweden, Finnland und Norwegen – punkten mit kostengünstigem grünen Strom (überwiegend Wasserkraft), natürlichen Kühlungsvorteilen, transparenten Genehmigungsverfahren und großzügigen verfügbaren Flächen. SAP betreibt aus diesen Gründen seit 2018 ein großes Rechenzentrum in Schweden. Microsoft hat in Finnland mit der Heumos-Anlage einen weiteren Großstandort errichtet.

Polen entwickelt sich zu einem wichtigen Tier-2-Standort für Unternehmen, die einen geographisch diversifizierten, aber dennoch kostengünstigen und gut vernetzten Standort innerhalb der EU suchen. Microsoft, Google und AWS haben in den letzten Jahren neue Rechenzentren in Polen errichtet. Analysten prognostizieren ein Wachstum auf etwa 181 MW im Jahr 2026. Die Latenzen zwischen Frankfurt und den wichtigsten polnischen Wirtschaftsmetropolen Posen und Warschau liegen im einstelligen bis niedrigen zweistelligen Millisekunden-Bereich – Werte, die für die meisten Anwendungsfälle unproblematisch sind.

Spanien (Madrid), Italien (Mailand, Rom), die Schweiz (Zürich) und Österreich (Wien) gewinnen ebenfalls an Bedeutung, insbesondere als Standorte für KI-Gigafabriken und für regionale Cloud-Regionen der Hyperscaler. Wien hat sich beispielsweise mit Unterstützung der österreichischen Bundesregierung um eine der bis zu fünf europäischen KI-Gigafabriken beworben.

12.4 USA und China: Die Maßstäbe globaler Skalierung

Der Vergleich mit den USA und China verdeutlicht die Größenordnungen, in denen sich der globale Wettbewerb um Rechenzentrumskapazitäten abspielt. In den USA dominieren die fünf großen Hyperscaler-Cluster Northern Virginia (mit über 7 GW der weltweit größte Hub), Dallas, Phoenix, Silicon Valley und Atlanta. Allein in Northern Virginia werden in den nächsten zwei Jahren weitere mehrere Gigawatt an Kapazität installiert – mehr als die gesamte aktuelle deutsche Kapazität.

Die Investitionssummen sind in einer anderen Größenordnung. Microsoft hat 2025 angekündigt, bis 2028 30 Milliarden US-Dollar in die KI-Infrastruktur und bestehende Betriebe in Großbritannien zu investieren – allein dieses Vorhaben übersteigt die gesamten deutschen Hyperscaler-Investitionen der letzten zehn Jahre. Google plant in den nächsten zwei Jahren rund 6,8 Milliarden US-Dollar in KI, Forschung und Entwicklung in Großbritannien. In den USA selbst hat OpenAI gemeinsam mit SoftBank und Oracle das „Stargate“-Projekt mit einem Investitionsvolumen von 500 Milliarden Dollar angekündigt.

China verfolgt mit dem nationalen Programm „East-Data-West-Computing" (EDWC, 中国东数西算) eine zentralstaatliche Rechenzentrumsstrategie, die explizit darauf abzielt, die Datenproduktion in den östlichen Ballungsräumen mit der Stromerzeugung in den westlichen Provinzen zu verbinden – insbesondere mit Wasserkraft in Sichuan und Solar- sowie Windkapazitäten in Xinjiang und der Inneren Mongolei. Die strategische Dimension wird bei einem Blick auf die Kapazitätszahlen deutlich: China verfügte Ende 2025 über eine installierte IT-Anschlussleistung von rund 32 GW – dem Elfachen der deutschen Gesamtkapazität. Bis 2030 soll dieser Wert auf über 60 GW steigen, während Deutschland für denselben Zeitraum eine Kapazität von rund 5 GW anstrebt. [1]

Zum Vergleich: US-amerikanische Rechenzentren kamen bereits 2024 auf eine Kapazität von 48 GW – das 16-fache Deutschlands. Allein die zehn leistungsstärksten Anlagen in den USA entsprechen der gesamten deutschen Rechenzentrumsinfrastruktur.

Die staatliche Steuerung des EDWC-Programms – mit verbindlichen Vorgaben für Standortwahl, Energieeffizienz (Pflicht: mindestens 80 % erneuerbare Energien für alle Neubauten in den acht Nationalen Computing-Hubs) und Anwendungsbereiche – steht in deutlichem Kontrast zum stärker marktgetriebenen Modell in den USA und Europa. Für Investoren bedeutet dies: Deutschland und Europa wachsen dynamisch, operieren aber in einer anderen Größenordnung als die führenden Digitalwirtschaften..

[1] Die Zahlen sind auf Basis installierter IT-Anschlussleistung (Watt) direkt vergleichbar. Zwei Einschränkungen sind zu beachten: Erstens liegt Chinas durchschnittlicher PUE bei ca. 1,49 gegenüber 1,43 in Deutschland, sodass die Gesamtstromentnahme bei gleicher IT-Kapazität in China höher ausfällt. Zweitens erfasst die Rystad-Zahl (32 GW China) alle Rechenzentren einschließlich Enterprise-Anlagen; die Bitkom-Zahl (2,98 GW Deutschland) bezieht sich auf Rechenzentren ab 100 kW. Die Vergleichbarkeit ist methodisch vertretbar, aber nicht vollständig.

13 Ausblick: Entwicklungsperspektiven 2029, 2031 und 2036

Der folgende Ausblick prognostiziert die Entwicklung des deutschen Rechenzentrumsmarktes über drei Zeithorizonte: kurzfristig bis 2029 (3 Jahre), mittelfristig bis 2031 (5 Jahre) und langfristig bis 2036 (10 Jahre). Die Prognosen basieren auf den Datenreihen der Bitkom/Borderstep-Studie, der 2026 vorgelegten Rechenzentrumsstrategie der Bundesregierung, den angekündigten Investitionsplänen der Hyperscaler sowie internationalen Vergleichswerten. Naturgemäß unterliegen Langfristprognosen erheblichen Unsicherheiten – insbesondere die Geschwindigkeit der KI-Adoption, regulatorische Änderungen und energiepolitische Weichenstellungen können die Entwicklung deutlich beschleunigen oder verlangsamen.

13.1 Kurzfristiger Ausblick: Bis 2029 (3-Jahres-Horizont)

Im kurzfristigen Zeithorizont bis 2029 wird sich die hohe Investitions- und Wachstumsdynamik der Jahre 2024–2026 nahtlos fortsetzen. Die bereits angekündigten und in Bau befindlichen Großprojekte werden in den Jahren 2027 bis 2029 sukzessive ans Netz gehen und die installierte IT-Anschlussleistung in Deutschland deutlich erhöhen.

13.1.1 Kapazitätsentwicklung bis 2029

Auf Basis der bereits genehmigten und in Bau befindlichen Projekte sowie der historischen Wachstumsraten ist eine Steigerung der installierten IT-Anschlussleistung von rund 3.000 MW Anfang 2026 auf 4.000 bis 4.500 MW im Jahr 2029 zu erwarten. Dies entspricht einer durchschnittlichen jährlichen Wachstumsrate von 10 bis 14 Prozent und einer absoluten Kapazitätssteigerung von 1.000 bis 1.500 MW innerhalb von drei Jahren.

Wesentliche Treiber dieser Entwicklung sind: (1) die zweite und dritte Phase der AWS-Investitionen mit zusätzlichen Modulen am Standort Frankfurt sowie der ersten Inbetriebnahme der Sovereign-Cloud-Region in Brandenburg, (2) der Anlauf des Microsoft-Standorts in Nordrhein-Westfalen mit der vollen Cloud-Region und der lokalen KI-Recheninfrastruktur, (3) die Inbetriebnahme der ersten Module des Telekom/NVIDIA-Industrial-AI-Cloud-Projekts in München, (4) der Start mehrerer großer Colocation-Projekte in Frankfurt und im Berliner Raum sowie (5) die Realisierung von Apple-Standorten in Bayern.

Hessen wird seine Spitzenposition mit hoher Wahrscheinlichkeit ausbauen und 2029 voraussichtlich rund 1.500 bis 1.600 MW installierte Anschlussleistung aufweisen. Bayern wird durch die KI-Großprojekte – insbesondere das Industrial-AI-Cloud-Projekt von Telekom/NVIDIA und ergänzende Chip- und F&E-Investitionen Apples – stark wachsen und sich als zweitgrößter Standort verfestigen, wobei NRW auf Augenhöhe folgen dürfte. Brandenburg wird durch das AWS-Projekt zum wichtigsten neuen Wachstumsstandort und könnte bis 2029 die 200-MW-Marke erreichen.

13.1.2 KI als dominanter Wachstumstreiber

Die KI-getriebene Nachfrage wird im 3-Jahres-Horizont der wesentliche Wachstumstreiber bleiben und einen immer größeren Anteil der neu installierten Kapazitäten ausmachen. Während 2025 rechnerisch etwa 450 MW bzw. 15 Prozent der Anschlussleistung KI-spezifisch waren, dürfte dieser Anteil bis 2029 auf 30 bis 35 Prozent steigen, was etwa 1.300 bis 1.500 MW KI-Kapazität entsprechen würde. Die Größenordnungen einzelner KI-Projekte werden weiter zunehmen: Während heute 50 bis 100 MW pro Standort als typische Hyperscale-Größe gelten, werden 200 bis 400 MW pro Standort zur neuen Normalität.

Bis 2028/29 ist mit der Inbetriebnahme der ersten KI-Gigafabrik im Rahmen der EuroHPC-Initiative zu rechnen. Das im Sondervermögen Infrastruktur vorgesehene Volumen von 805 Millionen Euro für eine deutsche KI-Gigafabrik soll bis dahin ausgeschöpft sein und einer der bewerbenden Standorte – diskutiert werden München, Stuttgart, Köln/Bonn und Magdeburg/Sachsen-Anhalt – wird voraussichtlich realisiert sein. Diese Anlage wird auf eine Rechenleistung im Bereich mehrerer Hundert MW ausgelegt sein und über zigtausend GPUs der jeweils aktuellsten Generation verfügen.

13.1.3 Energieverbrauch und Effizienz

Der Energieverbrauch der deutschen Rechenzentren wird sich auf die Marke von rund 28 bis 30 TWh im Jahr 2029 zubewegen, ausgehend von 21,3 TWh im Jahr 2025. Dies entspricht einem Anteil von etwa 5 Prozent am gesamten deutschen Stromverbrauch und einer Steigerung um etwa ein Drittel innerhalb von vier Jahren. Die Energieeffizienz wird durch das Wirksamwerden der EnEfG-Vorgaben spürbar verbessert: Der durchschnittliche PUE wird von 1,43 (2025) auf etwa 1,40 (2029) sinken, getrieben durch Pflichten für Neuanlagen, freiwillige Modernisierungen und den steigenden Einsatz von Flüssigkühlungssystemen.

Die Anteile erneuerbarer Energien werden zunehmen, wobei die EnEfG-Vorgabe „100 Prozent erneuerbar ab 2027“ rein bilanziell durch den Bezug grüner Stromzertifikate erreicht werden wird. Die physische Versorgung mit erneuerbarem Strom – insbesondere zu Zeiten geringer Sonneneinstrahlung und Windaufkommen – wird hingegen weiterhin teilweise auf Strom aus dem Verbundnetz mit fossilen Anteilen zurückgreifen. Eigenversorgungsmodelle (PPAs, Eigenanlagen) werden zunehmen, sind aber insbesondere für mittelgroße Standorte mit erheblichem Aufwand verbunden.

13.1.4 Infrastruktur und Genehmigungen

Die strukturellen Engpässe – Stromnetzanschluss, Bauleitplanung und Glasfaseranbindung – werden im Drei-Jahres-Zeitraum nur teilweise entschärft werden können. Die Maßnahmen der Bundesregierung aus der Rechenzentrumsstrategie 2026 werden ab 2027/28 Wirkung entfalten, jedoch sind die strukturellen Probleme – insbesondere im Hochspannungsnetz – nicht binnen drei Jahren lösbar. Tennet und die anderen Übertragungsnetzbetreiber werden mit erheblichen Investitionen die Netzkapazitäten ausbauen, die vollständige Aufhebung der heutigen Anschlussengpässe wird jedoch erst in den 2030er Jahren erreicht werden.

Das positive Signal: Die im April 2026 vorgelegte EnEfG-Novelle und die parallele Beschleunigungsinitiative der Bundesregierung könnten die Genehmigungszeiten für Neuanlagen von heute durchschnittlich 24 bis 36 Monaten auf 12 bis 18 Monate reduzieren – allerdings nur, wenn die in der Strategie vorgesehenen Verfahrensvereinfachungen tatsächlich umgesetzt werden.

13.2 Mittelfristiger Ausblick: Bis 2031 (5-Jahres-Horizont)

Im 5-Jahres-Horizont bis 2031 wird sich die strukturelle Transformation des deutschen Rechenzentrumsmarktes vertiefen. Die KI-Nachfrage wird zur dominierenden Marktkraft, neue Standortmuster setzen sich durch, und die Energie- und Wärmeintegration wird zum integralen Bestandteil der Standortplanung.

13.2.1 Kapazität: Verdoppelung gegenüber 2025

Bis 2031 wird die installierte IT-Anschlussleistung in Deutschland voraussichtlich auf 5.000 bis 6.000 MW anwachsen. Dies entspricht annähernd einer Verdoppelung gegenüber dem Niveau von 2025 (3.000 MW) und liegt in der Größenordnung des Ziels der Bundesregierung „Verdoppelung der Rechenzentrumskapazität bis 2030“. Die durchschnittliche jährliche Wachstumsrate liegt damit über die fünf Jahre bei etwa 10 bis 13 Prozent – ein historisch beispielloses Niveau, das nur durch die kombinierte Dynamik aus KI-Boom, Cloud-Migration, regulatorischen Anforderungen (DSGVO, NIS2) und industrieller Digitalisierung erreichbar ist.

Bitkom und Borderstep gehen in ihren Prognosen davon aus, dass der KI-Anteil bis 2030 auf rund 40 Prozent (entsprechend etwa 2.020 MW) ansteigt – eine Vervierfachung gegenüber 2025. Bis 2031 wird dieser Anteil voraussichtlich auf 45 bis 50 Prozent steigen. Die übrigen Kapazitäten verteilen sich auf klassische Cloud-Workloads (etwa 30 Prozent), unternehmenseigene und Colocation-Anwendungen (etwa 15 Prozent) sowie Edge- und spezielle Rechenzentren (etwa 10 Prozent).

13.2.2 Liquid Cooling als neuer Standard

Bis 2031 wird die Flüssigkühlung – insbesondere Direct-to-Chip-Cooling – im Hyperscale- und HPC-Segment zum Industriestandard avancieren. Während heute weniger als 20 Prozent der neu installierten Hyperscale-Kapazität flüssiggekühlt sind, wird dieser Anteil bis 2031 auf über 60 Prozent ansteigen. Die Rack-Dichten werden sich entsprechend entwickeln: Während heute 80 bis 150 kW pro Rack als typisch für KI-Workloads gelten, werden bis 2028 1 MW pro Rack möglich, und bis 2031 sind durchschnittliche Werte von 500 bis 800 kW pro Rack im KI-Segment realistisch.

Diese Entwicklung hat mehrere wichtige Implikationen. Erstens: Die Anforderungen an die Gebäudeinfrastruktur ändern sich grundlegend – statt großer luftgeführter Klimakanäle dominieren Flüssigkeitsleitungen, Pumpenstationen und Wärmetauscher. Zweitens: Die Wärmerückgewinnung und -nutzung wird signifikant einfacher, da die abgeführten Temperaturen mit 50 bis 70 °C deutlich höher sind als bei Luftkühlung (typisch 25 bis 30 °C). Drittens: Die Wartungs- und Servicekonzepte ändern sich erheblich – Leckagen, Wasserqualität und Pumpentechnik werden zu zentralen Themen.

13.2.3 Wärmeintegration als Standardanforderung

Mit den ab 2027 wirksamen EnEfG-Vorgaben zur Abwärmenutzung – 15 Prozent ab Mitte 2027, 20 Prozent ab Mitte 2028 – und der durch Liquid Cooling deutlich erleichterten Wärmeauskopplung werden bis 2031 die Wärmeintegration zum Standardbestandteil neuer Rechenzentrumsprojekte. Frühe Vorzeigeprojekte wie das Telehouse-Rechenzentrum in Frankfurt (Wärmelieferung an die Frankfurter Gärtnerei und das Schwimmbad), das Eurotheum-Hochhaus oder die Wärmeauskopplung in Norderstedt werden zur Regel statt Ausnahme.

Die wirtschaftliche Tragfähigkeit hängt jedoch entscheidend von der Verfügbarkeit geeigneter Wärmeabnehmer ab – Wohnquartiere, Gewerbeparks, Schwimmbäder, Treibhausanlagen oder industrielle Niedertemperaturprozesse. In ländlichen Regionen ohne entsprechende Abnehmer wird die 20-Prozent-Quote weiterhin nur unter erheblichem Aufwand zu erreichen sein. Die Bundesregierung plant daher in der Strategie 2026 die Förderung von „Energy Hubs“ – integrierten Konzepten aus Rechenzentrum, Quartierswärme und ggf. Wasserstoffproduktion.

13.2.4 Geographische Diversifizierung

Die starke Konzentration auf den Frankfurter Raum wird sich bis 2031 abschwächen. Während Hessen 2025 noch knapp 38 Prozent der gesamten Anschlussleistung auf sich vereinte, dürfte dieser Anteil bis 2031 auf etwa 30 bis 32 Prozent sinken. Treiber dieser Entwicklung sind: (1) Stromnetzengpässe und gestiegene Bodenpreise im Frankfurter Raum, (2) gezielte Ansiedlungspolitik strukturschwacher Bundesländer wie Mecklenburg-Vorpommern, Sachsen-Anhalt und Brandenburg, (3) Latenzanforderungen für Edge- und Industrial-AI-Anwendungen, die eine geographische Verteilung notwendig machen, und (4) das politische Ziel, Rechenzentrumskapazität in Regionen mit hohem Erneuerbaren-Anteil zu konzentrieren.

Konkret werden bis 2031 voraussichtlich folgende neue Rechenzentrums-Cluster substantielle Bedeutung erlangt haben: das Cluster Berlin/Brandenburg (AWS Sovereign Cloud, lokale Hyperscaler), die Region München (Industrial-AI-Cloud; Apple als angrenzende Chip-/F&E-Infrastruktur), die Achse Köln/Bonn–Düsseldorf (Microsoft, Google) sowie als neuer Standort der Norden Deutschlands mit Großprojekten in Mecklenburg-Vorpommern (Dummerstorf) und Schleswig-Holstein.

13.2.5 Stromverbrauch nähert sich der 40-TWh-Marke

Der Energieverbrauch der deutschen Rechenzentren wird bis 2031 bei einem Wachstum gemäß den genannten Kapazitätsannahmen die Marke von 35 bis 40 TWh erreichen. Damit wird der Anteil am gesamten deutschen Stromverbrauch auf 6 bis 7 Prozent steigen. Die spezifische Energieintensität pro MW installierter IT-Last sinkt zwar durch verbesserte Effizienz, dies wird jedoch durch das absolute Kapazitätswachstum mehr als kompensiert.

Diese Entwicklung wird die Diskussion um die Energieinfrastruktur weiter verschärfen. Die Bundesnetzagentur und die Übertragungsnetzbetreiber werden gezwungen sein, in ihren Netzentwicklungsplänen größere Reserven für Rechenzentren auszuweisen. Möglicherweise werden bis 2031 auch erste Pilotprojekte für die direkte Nutzung von Grundlastenergie aus Geothermie oder kleinen modularen Reaktoren (SMR) im Konzeptstadium sein – wenngleich tatsächliche Inbetriebnahmen erst nach 2031 zu erwarten sind.

13.3 Langfristiger Ausblick: Bis 2036 (10-Jahres-Horizont)

Der 10-Jahres-Horizont bis 2036 erlaubt nur noch Aussagen mit erheblichen Unsicherheitsbandbreiten. Dennoch lassen sich aus heutiger Sicht plausible Trajektorien skizzieren, die den Korridor möglicher Entwicklungen abstecken.

13.3.1 Kapazität: 8 bis 12 GW als Zielkorridor

Bis 2036 wird die installierte IT-Anschlussleistung in Deutschland in einer Bandbreite von 8 bis 12 GW liegen. Die untere Grenze von 8 GW entspricht einer Fortschreibung des historischen Wachstumstrends mit moderater Abflachung; die obere Grenze von 12 GW würde ein anhaltend hohes Wachstum von 10 bis 12 Prozent jährlich voraussetzen, getrieben durch eine breite Adoption von KI in allen Wirtschaftsbereichen, signifikante Nachfrage für Edge-Computing in industriellen Anwendungen und einen Erfolg der politisch gewollten europäischen Souveränitätsstrategie.

Das untere Szenario (8 GW) entspräche einer Versechsfachung gegenüber 2010 und etwa einer Verdreifachung gegenüber 2025. Selbst in diesem konservativen Szenario würde Deutschland damit das vom Branchenverband eco im Update der Studie für 2030 prognostizierte Niveau von etwa 4,8 GW übertreffen. Im oberen Szenario läge die deutsche Kapazität auf einem Niveau, das heute Northern Virginia (USA) entspricht und Deutschland zum klar führenden Rechenzentrumsstandort Europas machen würde.

13.3.2 Post-FLAP: Strukturwandel der europäischen Hub-Landschaft

Die heute dominante FLAP-D-Konstellation (Frankfurt, London, Amsterdam, Paris, Dublin) wird bis 2036 ihren Charakter ändern. Mehrere Faktoren tragen dazu bei: (1) Stromnetzengpässe und politische Ansiedlungsmoratorien in Amsterdam, Dublin und teilweise Frankfurt zwingen Wachstum in Sekundärstandorte; (2) die KI-getriebene Nachfrage erfordert standortmäßig flexiblere und größere Anlagen, die in den dichten innerstädtischen Standorten nicht mehr realisierbar sind; (3) die strategische Diversifizierung – Stichwort Cloud-Souveränität – führt zur bewussten Verteilung der Rechenkapazität auf mehr Standorte.

Bis 2036 wird Europa wahrscheinlich ein zweistufiges Modell entwickeln: An der Spitze stehen weiterhin die FLAP-D-Standorte als Konnektivitäts- und Latenzhubs, ergänzt um Madrid, Mailand und Berlin als gleichrangige neue Konnektivitätsknoten. Darunter etabliert sich eine zweite Schicht aus Compute-orientierten Standorten – im deutschen Kontext sind dies vor allem Mecklenburg-Vorpommern, Brandenburg, Bayern und Sachsen-Anhalt – die primär durch Energieverfügbarkeit, Flächen und politische Förderprogramme attraktiv sind.

13.3.3 KI-Souveränität als langfristige strategische Größe

Bis 2036 wird die Frage der europäischen KI-Souveränität – die Fähigkeit, KI-Modelle eigenständig zu trainieren und zu betreiben – weitgehend entschieden sein. Die heute mit erheblichen öffentlichen Mitteln initiierten KI-Gigafabriken (EuroHPC-Initiative mit bis zu fünf Standorten in Europa) werden bis dahin entweder etabliert sein und eigenständige Modelle und Anwendungen ermöglichen, oder sie werden durch die Dynamik der amerikanischen und chinesischen Hyperscaler überholt worden sein.

Im Erfolgsszenario verfügt Deutschland 2036 über drei bis fünf KI-Gigafabriken im GW-Bereich, die in europäische Forschungs- und Industrieanwendungen eingebunden sind und auch kommerziell genutzt werden. Im pessimistischen Szenario hingegen sind diese Anlagen mit US-amerikanischer Hardware (NVIDIA) und chinesischer Software (oder umgekehrt) gefüllt und dienen primär als Rechenkapazität für globale Hyperscaler-Modelle. Welches Szenario eintritt, hängt entscheidend von der politischen Konsequenz ab, mit der die in der Rechenzentrumsstrategie 2026 angekündigten Souveränitätsmaßnahmen umgesetzt werden.

13.3.4 Energieinfrastruktur: Neue Konzepte und SMR-Diskussion

Bei einem angenommenen Wachstum auf 8 bis 12 GW IT-Last bis 2036 wird der Stromverbrauch der deutschen Rechenzentren auf 60 bis 90 TWh ansteigen – das entspricht 10 bis 15 Prozent des prognostizierten gesamten deutschen Stromverbrauchs nach Umsetzung der Energiewende. Die langfristige Borderstep-Prognose von 80 TWh bis 2045 würde damit deutlich übertroffen.

Diese Größenordnungen werden zu einer Neubewertung der Energieinfrastruktur zwingen. Mehrere Konzepte werden bis 2036 wahrscheinlich diskutiert oder bereits in Pilotanwendungen umgesetzt sein: (1) dedizierte Onshore-Windparks oder Solaranlagen direkt für Rechenzentrumscampus, oft kombiniert mit Wasserstoff-Speicherung; (2) Geothermie-Großprojekte zur Grundlastversorgung großer Standorte, insbesondere in geologisch geeigneten Regionen wie Bayern und Brandenburg; (3) Pilotprojekte für kleine modulare Reaktoren (SMR) – diese werden in Deutschland angesichts des Atomausstiegs politisch heikel bleiben, sind aber in Großbritannien, Frankreich und den USA bereits in der Realisierung; (4) integrierte Konzepte aus Rechenzentrum, Wasserstoffproduktion (Elektrolyse mit Eigenstrom) und Wärmenetz, die als „Energy Hubs“ mehrere Versorgungsfunktionen kombinieren.

In der politischen Debatte werden Rechenzentren bis 2036 fast sicher zu einem zentralen Thema der Energiewende avancieren. Die heute in Spitzen schon sichtbaren Konflikte zwischen industrieller Wettbewerbsfähigkeit, Klimaschutzzielen und Versorgungssicherheit werden sich verschärfen, wenn die Branche tatsächlich zweistellige Prozentanteile am Stromverbrauch beansprucht.

13.3.5 Technologische Disruption: Quantencomputing und neuromorphe Chips

Auf der zehnjährigen Zeitachse können auch technologische Sprünge die Marktstruktur verändern. Quantencomputer werden bis 2036 für ausgewählte Anwendungen (Kryptographie, Materialforschung, bestimmte Optimierungsprobleme) kommerziell relevant sein und werden in spezialisierten Rechenzentren betrieben werden – wahrscheinlich als Cloud-Service über die etablierten Hyperscaler. Die heute in Forschungseinrichtungen wie dem Forschungszentrum Jülich oder dem Leibniz-Rechenzentrum installierten ersten kommerziellen Quantenrechner werden zu größeren Anlagen ausgebaut werden.

Neuromorphe Chips – nach dem Vorbild des menschlichen Gehirns konstruierte Spezialprozessoren mit deutlich höherer Energieeffizienz für bestimmte KI-Aufgaben – könnten bis 2036 die heute dominanten GPU-Architekturen für Inference-Workloads teilweise verdrängen. Erste kommerzielle Anbieter wie Intel (Loihi), IBM (TrueNorth) oder europäische Akteure wie SpiNNaker (Universität Manchester) sind bereits am Markt. Ein flächendeckender Durchbruch würde die Energiebilanz der KI-Inferenz drastisch verbessern und damit den Wachstumsdruck auf die Rechenzentrumskapazität reduzieren.

13.3.6 Konsolidierung und neue Geschäftsmodelle

Auf der Zehnjahresperspektive ist auch eine Marktkonsolidierung zu erwarten. Während heute über 2.000 Rechenzentren existieren und der Markt durch zahlreiche regionale und mittelgroße Anbieter geprägt ist, werden bis 2036 drei Trends die Struktur verändern: (1) die fortschreitende Konzentration auf Hyperscaler und große Colocation-Anbieter, die Skaleneffekte realisieren können; (2) die schrittweise Stilllegung kleiner unternehmenseigener Rechenzentren, deren Workloads in die Cloud migriert werden; (3) die Spezialisierung mittelgroßer Anbieter auf Nischen wie Edge-Computing, Hochsicherheits-RZs (z.B. für Behörden) oder Branchenlösungen (Finanzwesen, Gesundheit).

Neue Geschäftsmodelle werden entstehen. Die Verschmelzung von Energieerzeugung und Datenverarbeitung – „Compute-to-Power“ oder „Power-to-Compute“ – wird neue Akteure auf den Plan rufen. Energieversorger werden in den Rechenzentrumsmarkt einsteigen (RWE, EnBW haben bereits Interesse signalisiert), Rechenzentrumsbetreiber werden eigene Erzeugungsanlagen aufbauen. Industrieunternehmen werden ihre digitalen Zwillinge und KI-Anwendungen in zunehmendem Maße auf eigenen Edge-Rechenzentren betreiben statt vollständig in der Cloud.

13.4 Zusammenfassende Prognose

Die nachfolgende Tabelle fasst die Kernpunkte der Ausblicksszenarien für die drei Zeithorizonte zusammen.

Indikator	2025 (Ist)	2029 (3 J.)	2031 (5 J.)	2036 (10 J.)
IT-Anschlussleistung	3.000 MW	4.000–4.500 MW	5.000–6.000 MW	8.000–12.000 MW
Energieverbrauch	21,3 TWh	28–30 TWh	35–40 TWh	60–90 TWh
Anteil am Stromverbrauch	ca. 4 %	ca. 5 %	6–7 %	10–15 %
KI-Anteil an Kapazität	15 %	30–35 %	45–50 %	55–65 %
Durchschnittl. PUE	1,43	ca. 1,40	ca. 1,30	ca. 1,20
Hessen-Anteil	38 %	ca. 35 %	30–32 %	25–28 %
Liquid-Cooling-Anteil (Hyperscale)	< 20 %	ca. 35 %	ca. 60 %	> 80 %
KI-Gigafabriken in DE	0	1 (in Bau/Inbetriebn.)	1–2 betriebsbereit	3–5 betriebsbereit

Tabelle 7: Prognose-Korridor für die Entwicklung des deutschen Rechenzentrumsmarktes 2025–2036
Quellen: Bitkom/Borderstep 2025, Rechenzentrumsstrategie der Bundesregierung 2026, eigene Hochrechnungen

14 Handlungsempfehlungen

Aus den vorangehenden Kapiteln lassen sich konkrete Handlungsempfehlungen für die zentralen Akteursgruppen ableiten. Die Empfehlungen sind nach Adressaten gegliedert und priorisieren Maßnahmen, die kurzfristig (innerhalb von 12 bis 24 Monaten) Wirkung entfalten können oder die strategisch für die mittel- bis langfristige Wettbewerbsposition Deutschlands von zentraler Bedeutung sind.

14.1 Empfehlungen für die Politik

Die politischen Rahmenbedingungen entscheiden in erheblichem Maße darüber, ob Deutschland seine Position als zweitgrößter Rechenzentrumsstandort Europas verteidigen, ausbauen oder verlieren wird. Die im März 2026 vorgelegte Rechenzentrumsstrategie der Bundesregierung enthält die richtigen Stoßrichtungen, ihre Wirksamkeit hängt jedoch von der konsequenten Umsetzung der einzelnen Maßnahmen ab.

14.1.1 Beschleunigung von Genehmigungsverfahren

Die durchschnittliche Genehmigungsdauer von 24 bis 36 Monaten muss auf maximal 12 bis 18 Monate reduziert werden. Konkrete Maßnahmen sind: gesetzliche Festlegung verbindlicher Bearbeitungsfristen für Behörden mit „Genehmigungsfiktion“ bei Überschreitung; Einrichtung von One-Stop-Shops auf Landesebene, in denen alle relevanten Genehmigungen (Baurecht, Immissionsschutz, Wasserrecht, Stromnetzanschluss) gebündelt bearbeitet werden; Vorbereitung „RZ-tauglicher“ Bauleitpläne in Kooperation mit Kommunen, sodass Investoren auf vorgeprüfte Flächen zurückgreifen können; Anerkennung von Rechenzentren als „im überragenden öffentlichen Interesse“ stehende Infrastruktur, vergleichbar zu Stromtrassen oder Erneuerbare-Energien-Anlagen.

14.1.2 Stromnetzausbau und Anschlusskapazitäten

Der Engpass beim Stromnetzanschluss ist die zentrale Hürde der nächsten Jahre. Erforderliche Maßnahmen umfassen: vorgezogenen Ausbau der Übertragungs- und Verteilnetze in den Hauptrechenzentrumsregionen; Etablierung eines transparenten Anschlusswartelisten-Systems mit klaren Prioritätskriterien; Förderung lokaler Eigenversorgungsmodelle (PPAs mit Direktleitungen, eigene Erzeugungsanlagen); Weiterentwicklung des regulatorischen Rahmens für „Compute-to-Power“-Konzepte, bei denen Rechenzentren ihren Strom direkt vom Erzeuger beziehen ohne Belastung des öffentlichen Netzes.

Insbesondere die Tennet-Kapazitätssituation in Hessen erfordert ein dezidiertes politisches Vorgehen. Die im Bundestag bereits diskutierten Vorschläge zur Bevorzugung „kritischer Digitalinfrastruktur“ beim Netzanschluss sollten zügig umgesetzt werden.

14.1.3 Förderung der KI-Souveränität

Die EuroHPC-KI-Gigafabrik-Initiative bietet eine einmalige Chance, die deutsche und europäische KI-Souveränität substantiell zu stärken. Empfehlenswert ist eine verstärkte Anstrengung Deutschlands, mehr als nur eine der bis zu fünf europäischen Anlagen anzuwerben. Die im Sondervermögen vorgesehenen 805 Millionen Euro sollten gezielt durch private Mittel und EU-Ko-Finanzierung gehebelt werden, sodass eine Anlage im einstelligen Milliardenbereich realisiert werden kann.

Parallel ist die Finanzierung von KI-Modellen und KI-Anwendungen aus europäischer Provenienz zu fördern – ohne diese ist die Hardware allein wertlos. Die Forschungsförderung für deutsche und europäische KI-Akteure (Aleph Alpha, Mistral AI, Helsing, Black Forest Labs) sollte deutlich intensiviert werden.

14.1.4 Energiepolitische Flankierung

Die EnEfG-Vorgaben sind grundsätzlich ambitioniert und richtig, einzelne Aspekte – insbesondere die starre Abwärmenutzungsquote – sollten überprüft werden. Die im April 2026 vorgelegte Novelle zur leichten Aufweichung des PUE-Ziels für Neuanlagen ist sachgerecht. Empfehlenswert ist eine strikte Trennung zwischen technisch sinnvollen Effizienzanforderungen und politischen Symbolzielen.

Mittelfristig muss die Diskussion um die Energiebasis der Rechenzentren – einschließlich Geothermie, Wasserstoff und ggf. SMR – ergebnisoffen geführt werden. Eine ideologisch verengte Debatte würde die Wettbewerbsposition Deutschlands gefährden, da andere europäische Länder (Frankreich, Großbritannien, Schweden) hier bereits handeln.

14.2 Empfehlungen für die Wirtschaft

14.2.1 Für Rechenzentrumsbetreiber

Betreiber sollten frühzeitig in flüssigkeitsgekühlte Infrastruktur investieren, da diese in den nächsten Jahren zum Marktstandard wird und Bestandsanlagen mit reiner Luftkühlung an Wert verlieren werden. Bei Neubauten sollten die EnEfG-Anforderungen nicht nur als Mindestmaßstab, sondern als Ausgangspunkt für ambitioniertere Effizienzziele dienen, da regulatorische Verschärfungen in den 2030er Jahren wahrscheinlich sind.

Die Erschließung neuer Standorte sollte aktiv jenseits der etablierten FLAP-D-Hubs verfolgt werden. Frühe Position in aufstrebenden Regionen wie Brandenburg, Sachsen-Anhalt oder Mecklenburg-Vorpommern verspricht langfristige Wertsteigerungen, sofern die Infrastrukturanbindung gewährleistet werden kann. Partnerschaften mit lokalen Energieerzeugern und Wärmeabnehmern sollten von Beginn an in die Standortplanung einbezogen werden.

14.2.2 Für Anwenderunternehmen

Industrieunternehmen sollten ihre KI- und Cloud-Strategien auf Resilienz und Souveränität ausrichten. Eine Multi-Cloud-Strategie mit gezielter Nutzung europäischer Cloud-Anbieter (T-Systems, IONOS, OVHcloud, Schwarz Digits, STACKIT) reduziert Abhängigkeiten von den US-Hyperscalern und ist im Lichte der politischen Risiken (CLOUD Act, geopolitische Spannungen) geboten. Insbesondere für sensitive Daten und systemkritische Anwendungen sollten europäische Lösungen bevorzugt werden.

Edge-Computing wird in industriellen Anwendungen (Industrie 4.0, autonome Systeme, vernetzte Produktion) wachsende Bedeutung gewinnen. Unternehmen sollten frühzeitig Pilotprojekte aufsetzen und ihre Datenarchitekturen auf hybride Modelle aus Edge, Private Cloud und Public Cloud auslegen.

14.2.3 Für Energieversorger

Die Rechenzentrumsbranche bietet Energieversorgern erhebliche strategische Chancen. Die Entwicklung dezidierter Angebote – langfristige Lieferverträge, Direktleitungen, integrierte Wärmelösungen, eigene RZ-Beteiligungen – sollte intensiviert werden. Energieversorger mit erheblichen Erneuerbare-Erzeugungskapazitäten in strukturschwachen Regionen können diese gezielt für die Ansiedlung von Rechenzentren nutzen und damit beide Geschäftsfelder synergistisch verbinden.

14.3 Empfehlungen für Verwaltung und Kommunen

Kommunen und Landesbehörden sind oft die ersten Ansprechpartner für Investoren und damit entscheidend für die Wettbewerbsfähigkeit deutscher Standorte. Empfehlenswert ist die Einrichtung dezidierter Ansiedlungsteams mit Fachexpertise für Rechenzentren in den Wirtschaftsförderungen, die Investoren über alle Phasen des Genehmigungs- und Realisierungsprozesses begleiten. Die proaktive Bereitstellung von vorgeprüften Industrieflächen mit gesicherten Stromanschlüssen reduziert die Realisierungszeiten erheblich.

Die kommunale Beteiligung am wirtschaftlichen Erfolg der Rechenzentren – etwa durch Gewerbesteuern, Wärmeintegration in kommunale Netze oder Beteiligungen an Energiegesellschaften – sollte transparent und vorhersehbar gestaltet werden. Dies erhöht die Akzeptanz in der Bevölkerung und schafft langfristige Bindungen zwischen Investoren und Standorten.

In ländlichen Regionen sollten Wärmeintegrationskonzepte aktiv entwickelt werden, um die EnEfG-Anforderungen für Investoren erfüllbar zu machen. Dies kann durch kommunale Wärmenetze, Gewerbeparks mit integrierten Wärmeabnehmern oder Kooperationen mit landwirtschaftlichen Betrieben (Treibhäuser, Trocknungsanlagen) erfolgen.

14.4 Empfehlungen für Forschung und Bildung

Der Fachkräftemangel ist ein strukturelles Problem, das die Wachstumsperspektiven der gesamten Branche limitiert. Empfehlenswert sind ein gezielter Ausbau dualer Ausbildungsgänge im Bereich Rechenzentrumstechnik, die Etablierung spezialisierter Studiengänge an Hochschulen (Data Center Engineering, Sustainable Computing), die Förderung von Weiterbildungsangeboten für Quereinsteiger sowie verstärkte internationale Anwerbung von Fachkräften.

Die Forschung an energieeffizienten Rechenzentrumstechnologien sollte intensiviert werden. Themen wie immersive Kühlung, Wärmepumpen mit hohen Hub-Temperaturen, neuromorphe und photonische Computing-Architekturen oder integrierte Energie-Compute-Systeme bieten erhebliche Innovationspotenziale, in denen deutsche und europäische Forschung führend werden kann.

15 Fazit

Der deutsche Rechenzentrumsmarkt befindet sich Anfang 2026 in einer historischen Umbruchphase. Die im Frühjahr 2026 erstmals überschrittene Schwelle von 3.000 MW installierter IT-Anschlussleistung markiert nicht nur einen quantitativen Meilenstein, sondern dokumentiert auch einen qualitativen Wandel: Vom überwiegend latenzgetriebenen Cloud-Standort entwickelt sich Deutschland zu einem KI-zentrierten, energieintensiven und politisch gestalteten Rechenzentrumsmarkt, der Spitzenpositionen in Europa anstrebt.

Die Untersuchung der vorliegenden Studie zeigt ein differenziertes Bild. Auf der Habenseite stehen ein historisch nie dagewesenes Investitionsvolumen von rund 12 Milliarden Euro IT-Investitionen und 3,5 Milliarden Euro Bauinvestitionen im Jahr 2025, eine starke Marktposition im internationalen Vergleich (Frankfurt als größter europäischer Hub), erhebliche Standortvorteile (zentrale Lage, hohe Konnektivität, Rechtssicherheit, qualifizierte Arbeitskräfte) sowie eine deutlich verbesserte politische Aufmerksamkeit, die in der Rechenzentrumsstrategie 2026 ihren Ausdruck gefunden hat.

Auf der Sollseite stehen erhebliche strukturelle Herausforderungen: die langwierigen Genehmigungsverfahren, die Engpässe im Stromnetzanschluss (insbesondere in Hessen), die hohen Energiepreise im internationalen Vergleich, der Fachkräftemangel sowie die unterausgeprägte Kapazität in den Bereichen KI-Hardware und souveräne KI-Modelle. Diese Schwächen sind im Wesentlichen lösbar, erfordern jedoch konsequente politische und unternehmerische Anstrengungen über die kommenden Jahre.

Der Ausblick auf die kommenden drei, fünf und zehn Jahre ist verhalten optimistisch. Eine Verdoppelung der Kapazität bis 2030/31 erscheint realistisch, eine Verdreifachung bis 2036 möglich. Der KI-Anteil wird sich vervierfachen oder verfünffachen, die Energieeffizienz wird sich deutlich verbessern, neue Standortmuster jenseits der traditionellen FLAP-D-Region werden sich etablieren. Die Frage ist nicht, ob Deutschland wachsen wird, sondern in welcher Geschwindigkeit und mit welcher strategischen Tiefe.

Im internationalen Wettbewerb bleibt Deutschland gegenüber den USA und China in einer Position, die nur durch europäische Bündelung aussichtsreich gestaltet werden kann. Die EuroHPC-KI-Gigafabrik-Initiative, die Cloud-Souveränitätsbestrebungen sowie die Förderung europäischer KI-Ökosysteme sind hierfür die zentralen Hebel. Eine rein nationale Strategie wäre angesichts der Größenordnungen, in denen sich der globale Wettbewerb abspielt, von vornherein zum Scheitern verurteilt.

Für die deutsche Wirtschaft – sowohl die Rechenzentrumsbetreiber als auch die anwendenden Industrieunternehmen – ergeben sich daraus klare strategische Imperative: frühzeitige Investitionen in zukunftsweisende Technologien (Liquid Cooling, KI-Infrastruktur, Wärmeintegration), Diversifizierung der Standorte über die etablierten Hubs hinaus, Aufbau souveräner europäischer Cloud- und KI-Kompetenzen sowie aktive Mitgestaltung der politischen Rahmenbedingungen.

Die kommenden zehn Jahre werden über Deutschlands Position im globalen Rechenzentrumsmarkt entscheiden. Die Voraussetzungen sind gut, die Chancen erheblich, die Herausforderungen jedoch ebenfalls beträchtlich. Ob Deutschland die historische Chance zur Konsolidierung und zum Ausbau seiner führenden europäischen Position nutzt, wird davon abhängen, ob die in der Rechenzentrumsstrategie skizzierten Vorhaben zügig und konsequent umgesetzt werden, ob die Energieinfrastruktur den Anforderungen der digitalen Transformation gewachsen ist und ob der politische Wille zur europäischen Souveränität auch in Zeiten sich verändernder geopolitischer Konstellationen erhalten bleibt.

Die vorliegende Studie zeichnet das Bild einer Branche im Aufbruch – mit erheblichen Chancen, klar erkennbaren Risiken und einem gestaltbaren Zukunftspfad. Die nächsten Jahre werden die Weichen stellen.

Anhang A: Quellenverzeichnis

Die folgende Aufstellung dokumentiert die für diese Studie verwendeten Primär- und Sekundärquellen. Die Quellen sind nach Kategorien gegliedert.

Studien und Marktberichte

Bitkom Research / Borderstep Institut (2025): Rechenzentren in Deutschland 2025 – Aktuelle Marktentwicklungen, Update der Studie. Berlin, November 2025.

Borderstep Institut für Innovation und Nachhaltigkeit (2024): Energiebedarf der Rechenzentren in Deutschland 2010–2030. Studie im Auftrag der Bitkom e.V., Berlin.

eco – Verband der Internetwirtschaft e.V. / Arthur D. Little (2024): Rechenzentren in Deutschland 2024. Köln/Frankfurt.

CBRE Research (2025): European Data Centre Market Outlook H1 2025. London.

Cushman & Wakefield (2025): Global Data Center Market Comparison 2025. Chicago.

Cushman & Wakefield (2026): Americas Data Center Update H2 2025. Chicago: Cushman & Wakefield, Februar 2026. cushmanwakefield.com

JLL (2025): European Data Centre Report Q4 2024. London.

DLA Piper (2025): Europäischer Rechenzentrumsmarkt – Rechtliche und regulatorische Rahmenbedingungen. Frankfurt am Main.

Synergy Research Group (2025): Hyperscale Data Center Capacity Outlook 2025–2030. Reno, Nevada.

Cloudscene (2026): Germany Data Center Market Report – Infrastructure Deployment and Capacity Estimates for Hyperscale Regions. Brisbane/Berlin

Politische und behördliche Dokumente

Bundesministerium für Digitales und Staatsmodernisierung – BMDS (2026): Rechenzentrumsstrategie der Bundesregierung. Vorgelegt durch Bundesminister Karsten Wildberger am 18. März 2026 im Bundeskabinett, Berlin.

Bundesministerium für Wirtschaft und Energie – BMWE (2026): Referentenentwurf einer Verordnung zur Änderung des Energieeffizienzgesetzes (EnEfG). Berlin, 9. April 2026.

Energieeffizienzgesetz (EnEfG) vom 4. November 2023 (BGBl. I, Nr. 309), zuletzt geändert durch das Erste Gesetz zur Änderung des Energieeffizienzgesetzes vom Dezember 2024.

Bundesnetzagentur (2025): Versorgungssicherheitsbericht Strom 2025. Bonn.

Bundesregierung (2026): Antwort auf die Kleine Anfrage zur Lage der deutschen Rechenzentrumsbranche. Bundestags-Drucksache 21/4910. Berlin.

Bundesregierung (2026): Antwort auf die Kleine Anfrage zur Umsetzung des Energieeffizienzgesetzes. Bundestags-Drucksache 21/5271. Berlin.

Bundesregierung (2026): Antwort auf die Kleine Anfrage zur Förderung von KI-Gigafabriken. Bundestags-Drucksache 21/4817. Berlin.

Deutsche Energie-Agentur (dena) (2025): Studie zur Wärmeintegration von Rechenzentren in Deutschland. Berlin.

Tennet TSO GmbH (2025): Netzentwicklungsplan Strom 2037/2045 – Stellungnahme zum Lastzuwachs durch Rechenzentren. Bayreuth.

Chinesische Nationalregierung / Xinhua (2024): China invests over 6.1 billion USD in major computing hubs. Peking: Xinhua News Agency, 29. August 2024. [english.www.gov.cn]

Federal Energy Regulatory Commission (FERC) (2026): 2025 State of the Markets. Washington D.C.: FERC, März 2026. ferc.gov

Europäische Quellen

EuroHPC Joint Undertaking (2025): Call for Hosting Entities of AI Factories and AI Gigafactories. Luxemburg.

Europäische Kommission (2024): Verordnung (EU) 2024/1244 über die Berichterstattung des Energieverbrauchs und Wassernutzung von Rechenzentren. Brüssel.

Europäische Kommission (2025): AI Act – Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz, Anwendungsleitfäden 2025. Brüssel.

European Data Centre Association (2025): Climate Neutral Data Centre Pact – Annual Report 2024. Brüssel.

Unternehmens- und Branchenmeldungen

Amazon Web Services (2024): Pressemitteilung zur Investition von 8,8 Mrd. Euro in die AWS Region Frankfurt bis 2026. Seattle/Frankfurt am Main.

Amazon Web Services (2024/2026): Pressemitteilungen zur European Sovereign Cloud in Brandenburg, Investitionsvolumen 7,8 Mrd. Euro bis 2040. Seattle/Berlin.

Microsoft Corporation (2024): Pressemitteilung zur Investition in eine neue Cloud- und KI-Region in Nordrhein-Westfalen, Volumen 3,2 Mrd. Euro. Redmond/Berlin.

Google Cloud / Alphabet Inc. (2025): Pressemitteilung zu Investitionen in Höhe von 5,5 Mrd. Euro in Deutschland 2026–2029. Mountain View/Berlin.

Deutsche Telekom AG / NVIDIA Corporation (2025): Gemeinsame Pressemitteilung zur Industrial-AI-Cloud in München, Partnerschaft im Wert von rund 1 Mrd. Euro. Bonn/Santa Clara.

Apple Inc. (2023): Pressemitteilung zur Erweiterung des European Silicon Design Center in München; insgesamt 2 Mrd. Euro Zusage für den Standort München. Cupertino.

CloudHQ (2025): FRA Campus Höchst – Project Overview. cloudhq.com/campus/fra-campus-hochst [Stand: Dezember 2025]

NTT Global Data Centers (2025): Pressemitteilung Nierstein/Rhein-Selz-Park, zitiert nach SWR [Stand: November 2025]

Amazon Web Services (2024/2026): AWS European Sovereign Cloud Brandenburg. press.aboutamazon.com/de [Stand: Januar 2026]

Microsoft Deutschland (2026): Spatenstich Rechenzentrums-Cluster Rheinisches Revier. news.microsoft.com/source/emea [Stand: März 2026]

Google / Hessische Landesregierung (2025): Pressemitteilung Investitionspaket Deutschland. hessen.de/presse [Stand: November 2025]

Equinix, Inc. (2025): Equinix investiert über 700 Millionen US-Dollar in den Ausbau der digitalen Infrastruktur in Deutschland. Pressemitteilung vom 2. September 2025. newsroom.equinix.com

Digital Realty (2025): Digital Realty erweitert Digital Park Fechenheim – FRA20. Pressemitteilung + Projektbeschreibung. datacenter-insider.de, November 2025

Vantage Data Centers (2025): Rechenzentrum Bad Vilbel – 160 MW. Immobilien Zeitung, 2. Juli 2025. iz.de

Oracle Deutschland (2025): Oracle investiert zwei Milliarden US-Dollar in KI- und Cloud-Infrastruktur in Deutschland. Pressemitteilung vom 15. Juli 2025. oracle.com/de/news

Rystad Energy (2026): China's data center capacity set to top 60 GW by 2030, driving a doubling of power demand. Oslo: Rystad Energy, April/Mai 2026. [rystadenergy.com/news]

Bitkom / Borderstep Institut (2025): Rechenzentren in Deutschland – KI treibt Wachstum. Pressemitteilung vom 10. November 2025. Berlin: Bitkom. [bitkom.org/Presse]

Fachartikel und Branchenberichte

datacenter-insider.de (2025–2026): Verschiedene Artikel zu Rechenzentrumsprojekten in Deutschland. Vogel Communications Group, Würzburg.

Taylor Wessing (2025): Rechtliche Aspekte der Rechenzentrumsentwicklung in Deutschland – Mandantenbriefing. Berlin/Frankfurt.

Frankfurter Allgemeine Zeitung (2025–2026): Diverse Artikel zur Entwicklung des Frankfurter Rechenzentrumsclusters. Frankfurt am Main.

Handelsblatt (2025–2026): Diverse Artikel zu Hyperscaler-Investitionen und KI-Infrastruktur in Deutschland. Düsseldorf.

Heise Online / iX Magazin (2025–2026): Diverse technische Berichte zu Liquid Cooling, KI-Infrastruktur und Rechenzentrumstechnologie. Hannover.

Computerwoche (2025–2026): Diverse Artikel zur Cloud-Marktentwicklung in Deutschland. München.

Wirtschaftsförderung Brandenburg / NRW.Global Business (2025/2026): Projektberichte zur Ansiedlung von Cloud-Infrastrukturen und regionalen Netzlastprognosen.

Internationale Vergleichsstudien

McKinsey & Company (2024): Investing in the rising data center economy. New York.

Goldman Sachs Research (2024): AI is poised to drive 160 % increase in data center power demand. New York.

International Energy Agency – IEA (2025): Energy and AI – World Energy Outlook Special Report. Paris.

Boston Consulting Group (2024): The Cost of Compute – Strategic Implications of the AI Infrastructure Boom. Boston.

Synergy Research Group (2025): Hyperscale Cloud Provider CAPEX and Data Center Footprint Analysis. Reno, NV

Methodischer Hinweis

Die in dieser Studie genannten Zahlen, Trends und Prognosen wurden auf Basis der zum Stichtag 4. Mai 2026 verfügbaren Informationen erstellt. Bei Investitionsangaben handelt es sich teilweise um angekündigte, noch nicht vollständig realisierte Vorhaben; tatsächliche Umsetzungen können in zeitlicher Streckung, abweichender Höhe oder veränderter Konfiguration erfolgen. Prognoseangaben unterliegen den üblichen Unsicherheiten von Trendfortschreibungen und expertenbasierten Schätzungen. Wo möglich, wurden Datenpunkte aus mehreren unabhängigen Quellen abgeglichen. Die Studie erhebt keinen Anspruch auf Vollständigkeit der Quellenlage; sie stellt vielmehr eine integrierte Analyse auf Basis der wesentlichen verfügbaren Informationen dar.

Anhang B: Verzeichnisse und Abkürzungen

Tabellenverzeichnis

Tabelle 1: Schlüsselkennzahlen des deutschen Rechenzentrumsmarktes - Stand und Prognose (Kapitel 1)

Tabelle 2: Marktstruktur des deutschen Rechenzentrumsmarktes nach Anbieterkategorien (Schätzung) (Kapitel 4)

Tabelle 3: Top-10-Bundesländer nach IT-Anschlussleistung 2025 (Kapitel 5)

Tabelle 4: Größte angekündigte oder in Bau befindliche Einzelprojekte in Deutschland (Kapitel 6)

Tabelle 5: PUE- und Abwärmenutzungsvorgaben des EnEfG nach Anlagentyp und Datum (Kapitel 9)

Tabelle 6: Vergleich der FLAP-D-Standorte in Europa (Stand 2025) (Kapitel 12)

Tabelle 7: Prognose-Korridor für die Entwicklung des deutschen Rechenzentrumsmarktes 2025–2036 (Kapitel 13)

Abbildungsverzeichnis

Abbildung 1: Entwicklung der IT-Anschlussleistung deutscher Rechenzentren 2010–2030 (in MW) (Kapitel 4)

Abbildung 2: Installierte IT-Anschlussleistung nach Bundesländern 2025 (in MW) (Kapitel 5)

Abbildung 3: Angekündigte Hyperscaler- und Tech-Investitionen in deutsche Rechenzentrums-, Cloud- und angrenzende Tech-Infrastruktur (in Mrd. Euro) (Kapitel 6)

Abbildung 4: Entwicklung der Rack-Leistungsdichte für klassische und KI-Workloads (in kW pro Rack) (Kapitel 7)

Abbildung 5: Energiebedarf deutscher Rechenzentren in TWh (2010–2045) (Kapitel 8)

Abbildung 6: Entwicklung des durchschnittlichen PUE-Werts in Deutschland und gesetzliche Vorgaben des EnEfG (Kapitel 8)

Abbildung 7: Anteil von HPC-/KI-Kapazitäten an der Gesamtkapazität in Deutschland 2025 vs. Prognose 2030 (Kapitel 10)

Abbildung 8: Installierte Rechenzentrumskapazität – internationaler Vergleich 2025 (in MW, logarithmische Skala) (Kapitel 12)

Abkürzungsverzeichnis

AI / KI – Artificial Intelligence / Künstliche Intelligenz

AWS – Amazon Web Services

BMDS – Bundesministerium für Digitales und Staatsmodernisierung

BMWE – Bundesministerium für Wirtschaft und Energie

CapEx – Capital Expenditure (Investitionsausgaben)

DC – Data Center / Rechenzentrum

DSGVO – Datenschutz-Grundverordnung

EnEfG – Energieeffizienzgesetz

ERF – Energy Reuse Factor / Anteil wiederverwendeter Energie

EuroHPC – European High Performance Computing Joint Undertaking

FLAP-D – Frankfurt, London, Amsterdam, Paris, Dublin

GPU – Graphics Processing Unit

HPC – High Performance Computing

IaaS – Infrastructure as a Service

KMU – Kleine und mittlere Unternehmen

MW / GW – Megawatt / Gigawatt

NIS2 – Network and Information Security Directive 2 (EU-Richtlinie)

PaaS – Platform as a Service

PPA – Power Purchase Agreement (langfristiger Stromliefervertrag)

PUE – Power Usage Effectiveness (Effizienzkennzahl Rechenzentrum)

RZ – Rechenzentrum

SaaS – Software as a Service

SMR – Small Modular Reactor (Kleiner modularer Kernreaktor)

TWh – Terawattstunde