



2026

State of AI Infrastructure Report

AI-ready isn't
infrastructure-ready



Table of contents

03 Introduction

07 Section I: Power

14 Section II: Constraints converge

22 Section III: Networking constraints

28 Section IV: Tariffs and policy uncertainty

32 Section V: ROI and budgets

37 Conclusion

39 Methodology & References



What happens when organizations can afford AI, but its infrastructure can't keep up?

Two years ago, the conversation was about whether AI would matter to an enterprise. Last year, it was about how to implement it. This year, attention has turned to the systems required to support it.

Power determines **where** AI can be sited. Networking determines **how long** it takes data to reach compute. Policy and trade conditions determine **what hardware costs** and **when** it arrives. These pressures are no longer separate budget lines or departmental concerns. Businesses are discovering that they form a **single, interconnected system** of limits.

For our **third annual State of AI Infrastructure Report**, we surveyed more than 350 **IT decision-makers** at the director level or above at organizations with more than \$100 million in **annual revenue** — including 100 respondents from those exceeding \$2 billion. Our goal was to understand how businesses are **deploying AI** and how the **infrastructure** supporting it's holding up.

Power has emerged as the **central constraint**. It determines where and when AI can actually be deployed. Access to **reliable grid power** is one of the most important factors when deciding where to put AI workloads, according to 89% of respondents, and 72% express moderate or extreme concern about electricity price volatility affecting their AI operating costs.

Who's behind these insights?



350
IT decision-makers



\$100M
in annual revenue



100
respondents from
\$2B+ organizations

The surrounding infrastructure is also **struggling to keep up** and is a primary obstacle to expanding deployments, according to 40% of respondents (Fig. 1). **Limited access** to high-performance AI compute moderately or severely constrains 82% of organizations, and 71% report excessive latency in their AI workloads.

Fig. 1

Which of the following represents the greatest barrier to expanding your organization's AI initiatives?



Despite these limitations, respondents **remain confident** about where their efforts are headed. **Excitement** climbed from 73% in 2024 to 82% this year, and the share of respondents who are **proud** of their organization's AI initiatives increased from 47% to 71% (Fig. 2). The **negative reactions** that accompanied early adoption have **faded** as well. The share who reported feeling **overwhelmed** fell from 29% last year to 11%, nervous from 13% to 3%, and **uncertain** from 10% to 2%.

Fig. 2

Which of the following best describes your attitude toward the implementation of AI applications and initiatives in your organization?

(Select no more than three responses.)

	2024	→	2025	→	2026
Excited	73%	→	75%	→	82%
Inspired	49%	→	47%	→	73%
Proud	47%	→	55%	→	71%
Overwhelmed	12%	→	29%	→	11%
Apathetic	10%	→	12%	→	8%
Nervous	16%	→	13%	→	3%
Uncertain	11%	→	10%	→	2%

Organizations can largely solve the problems within their control. What they can't — securing **power**, procuring **chips**, and laying **fiber** — is converging into a **single bottleneck** on AI and digital growth that's tightening faster than most can adapt.

Most enterprises aren't set up to handle it on their own. What happens next depends on **access to the right infrastructure** and partners who can help bridge the gap.

Five biggest takeaways

AI maturity has accelerated and budgets are in place, but respondents are contending with:

01 **Rising energy costs** and constrained power availability

04 **Compute scarcity** despite record capital investment

02 **Network performance** that can't match AI workload demands

05 **A widening gap** between AI ambitions and measurable returns

03 **Tariffs and trade policy** disrupting procurement timelines

Eight key findings

- 1 Power cost or pricing differences** ranked as the top consideration influencing where organizations locate AI workloads (**55%**), ahead of data residency, security, performance, and compute availability.
- 2 96% of respondents reported at least one network-related performance issue** affecting their AI workloads over the past 12 months.
- 3 Fiber availability, carrier diversity, and low-latency connectivity** have been a limiting factor for **91%** of organizations when evaluating sites for AI workloads.
- 4 Cost reduction and operational efficiency** replaced revenue growth as the top measure of AI success, rising to **55%**, while revenue growth fell from **58%** to **42%**.
- 5 The share of organizations expecting measurable financial returns from AI** within a year fell from **51%** to **36%**, while the share already seeing returns held flat at **20%**.
- 6 54% of respondents said tariffs** have already increased their reliance on domestic suppliers, while **40%** have delayed or scaled back purchasing.
- 7 AI compute and AI data** are heading in opposite directions, with GPU deployment moving toward the public cloud (**54%**, up from **30%** in 2024) and data moving away from it (**26%**, down from **59%**).
- 8 98% of respondents view nuclear** as a viable power source for AI infrastructure within five to 10 years.



Energy constraints have become the defining limit on AI ambition

For many organizations, one of the first questions about a new AI deployment is whether enough power is available to support it.

The United States power grid is **struggling** to provide an answer. PJM Interconnection, which covers 13 states from Virginia to Illinois, is the largest wholesale electricity market in the country. Its **capacity auction prices have risen** from \$28.92 per MW-day for the 2024-25 delivery year to \$333.44 per MW-day in 2027-28, and they would have been higher had they not hit the price cap approved by the Federal Energy Regulatory Commission (FERC) for a second consecutive year.^{1,2}

Although several factors drive up rates, including **demand from other industries** and **market rule and modeling changes**, data center energy demand accounted for an estimated 63% of the price increase in the 2025-26 auction, demonstrating just how much power AI infrastructure now requires.³

In response, recent federal and state actions directed PJM in January to hold an **emergency**

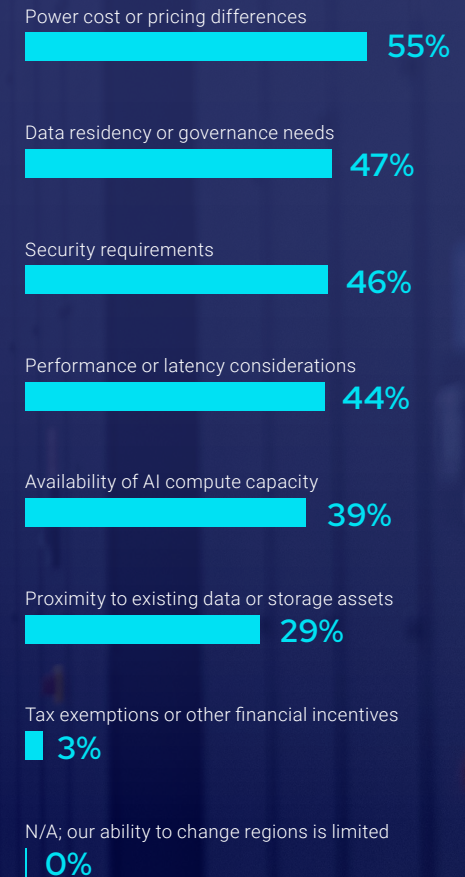
auction requiring technology companies to **bid for new power**, while leading hyperscale operators signed a voluntary ratepayer **protection pledge** in March in which they committed to **build or buy their own** electricity generation and **cover infrastructure upgrade costs**.^{4,5} Whether these promises can slow rising energy prices remains uncertain, but it's clear that power for AI has become a prominent **public policy issue**.

The organizations answering our survey operate inside this environment. For most of them, power supply is the **central factor** in AI workload placement. When asked what determines new or expanded deployments, 89% of respondents say the **availability of reliable grid power** is the deciding factor (27%) or one of several major factors (62%). **Power cost or pricing differences** ranks as the top consideration influencing regional selection (55%), ahead of **data residency** (47%), **security** (46%), **performance** (44%), and **compute availability** (39%) {Fig. 3}.

Fig. 3

Which factors influence your organization's choice of region for AI workloads?

(Select no more than three responses.)

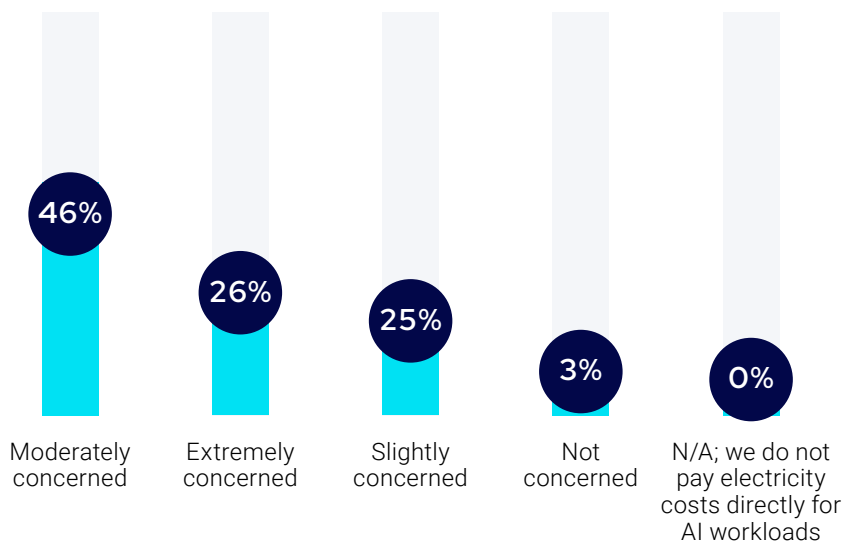


Looking ahead, 72% of respondents express **moderate** or **extreme concern** about **electricity price volatility affecting AI operating costs** {Fig. 4}, and 42% consider **power availability** to be the sustainability-related constraint posing the **greatest long-term risk to AI infrastructure growth** {Fig. 5}.

If power is the constraint, **energy sourcing is the response**. Respondents have clear expectations about where the megawatts will come from – and how quickly.

Fig. 4

How concerned is your organization about future electricity price volatility affecting AI operating costs?



72%

of respondents express moderate or extreme concern

Fig. 5

Which sustainability-related constraint poses the greatest long-term risk to AI infrastructure growth?



Power availability is the dominant long-term constraint.

How to resolve the power problem

Renewable energy (solar and wind) is the way forward, according to 67% of respondents; 42% say **renewables paired with battery storage systems** will supply most power for AI within five years, while 25% believe solar and wind alone can do so {Fig. 6}.

The strong preference for storage-backed renewables aligns with recent market dynamics. The **global benchmark cost** for a four-hour battery project fell around 27% last year, and 10.9 GW of utility- or **grid-scale energy storage was deployed** in the first three quarters of the year, matching all of 2024 in nine months.^{6,7}

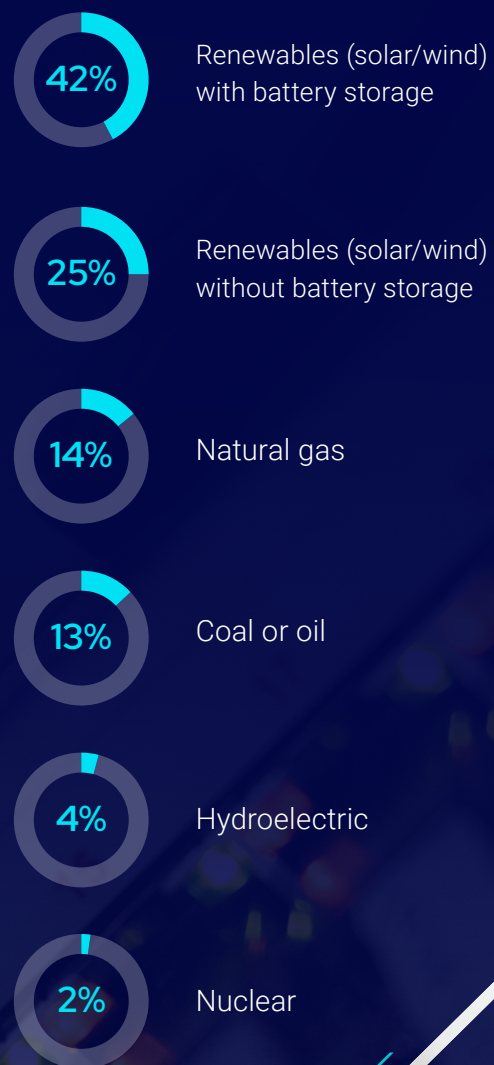
Battery storage is also increasingly valued as an interconnection accelerator — a way to bring data center campuses online years earlier by **bypassing traditional utility upgrade timelines**. One proposed Minnesota data center illustrates the integrated model with plans for 1,400 MW of wind, 200 MW of solar, and a 300 MW iron-air battery designed to dispatch up to 100 hours of continuous power — enough to cover extended periods of lower solar generation and multiday weather events that conventional lithium-ion systems cannot.⁸

But the policies that have long made renewable energy affordable and feasible are changing, as tax credits for wind and solar facilities are ending.⁹ Battery storage, however, will **remain eligible for full federal tax credits** through 2033.

This divergence creates a near-term advantage for storage while introducing significant uncertainty into the renewable generation pipeline that respondents expect to dominate AI power supply.

Fig. 6

Which power source do you believe will play the largest role in supporting AI workloads over the next five years?



Confidence in nuclear, but zero expectations

98% view nuclear as viable for AI infrastructure within five to 10 years

Respondents express strong confidence in the prospects of nuclear power but have **little belief** that it will arrive in time to ease near-term demand. All but six respondents (98%) view nuclear as viable for AI infrastructure within five to 10 years, but virtually none (2%) believe it will play the largest role over the next five.

Their outlook aligns with realistic timelines. Data centers can be built in one to two years, but **new nuclear plants typically take a decade to complete**. The first commercial small modular reactors (SMRs), touted as a compromise solution for nuclear, aren't projected to deliver power until the early 2030s.¹⁰ As a result, hyperscalers have reached a wave of offtake agreements with **existing nuclear plants** to bypass the new-build timeline and secure energy in the near future. Recent agreements include a 20-year, 1.1 GW power purchase agreement (PPA) in Illinois between a utility and a hyperscaler; a restructured \$18 billion, 1,920 MW contract in Pennsylvania; and a hyperscaler-backed restart of a previously retired 835 MW reactor.^{11,12,13}

New reactors are only beginning to move forward. In early March, the Nuclear Regulatory Commission (NRC) granted the **first construction permit for a new U.S. commercial reactor in nearly a decade** — one located in Wyoming and expected to come online in 2031.¹⁴ In January, a hyperscaler signed deals with three reactor developers to support projects that could add up to 6.6 GW of generation capacity by 2035.¹⁵ These are real commitments, but on timelines that underscore why respondents see nuclear as viable without seeing it as imminent.

Natural gas becomes an emerging solution

While respondents look to renewables for the near term and nuclear for the long term, **natural gas is the energy source** attracting additional investment today

Only 14% of respondents identify natural gas as the source that will play the **largest role in supporting AI workloads** over the next five years, yet gas is projected to remain a leading source of data center power through 2030.¹⁶ In addition, a research firm estimated that natural gas generation costs roughly five times less per kilowatt than new nuclear construction.¹⁷

Behind-the-meter gas generation has surged because simple-cycle turbines can be installed in 12 to 18 months, compared with five or more years for utility interconnection through the traditional grid. One study identified 46 **data center projects moving their energy source behind the meter**.¹⁸ A natural gas pipeline operator is investing \$5.1 billion to deliver 6 GW of behind-the-meter generation by mid-2027.¹⁹ A multinational energy corporation joined a venture to develop 4 GW of gas-fired generation and, separately, is building a 2.5 GW gas plant for data centers, with both projects set for West Texas because of its abundant natural gas supply.^{20,21}

Renewables, nuclear, and natural gas offer **different combinations** of speed and cost. What organizations are willing to pay for cleaner options reveals how those tradeoffs are landing.



gas is projected to remain a leading source of data center power through 2030

Sustainability pressure rises as willingness to pay shrinks

The share of respondents reporting increased pressure to make IT infrastructure more sustainable has climbed from 79% last year to 90% this year (Fig. 7), but their **willingness to absorb the cost** of doing so is moving in the opposite direction.

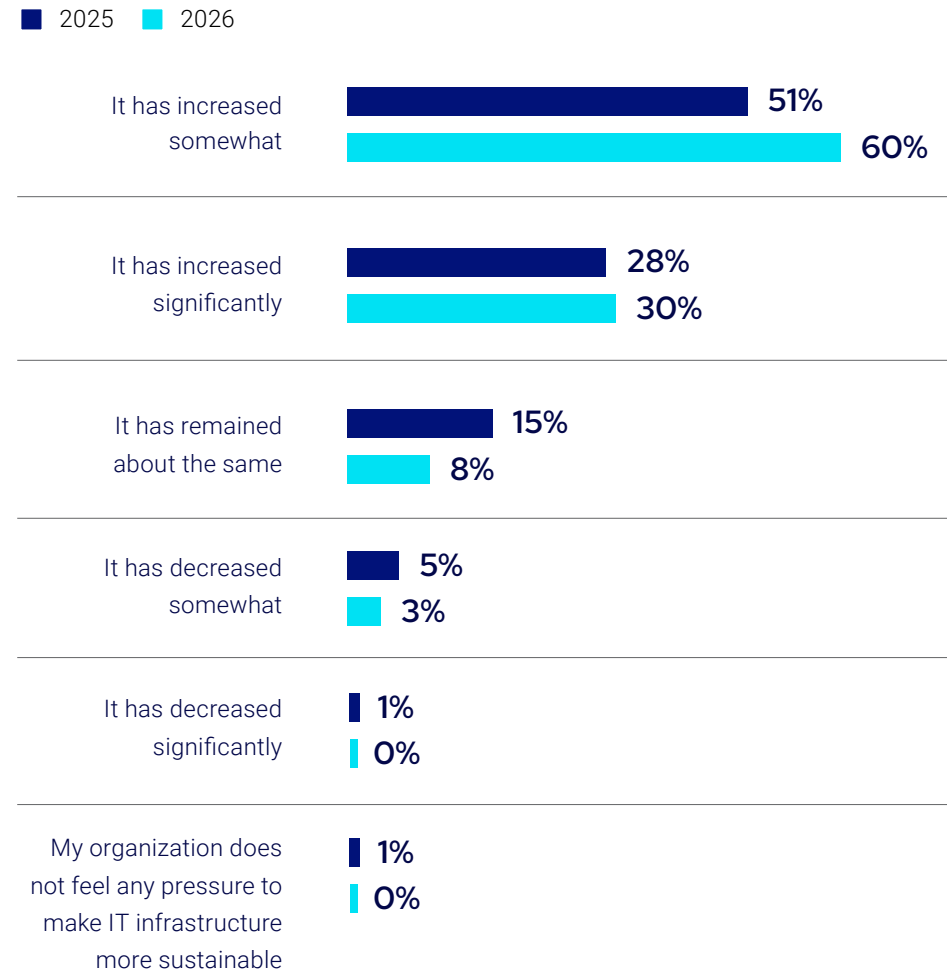
Only 51% of respondents said they are willing to pay more than a 10% premium for their data centers or third-party cloud vendors to use **clean or renewable energy**. A year ago, 78% were willing to do so; in 2024, 45% were. Meanwhile, the share open to paying more than 20% dropped from 27% last year down to 13%.

Several factors are at play. As noted, **wholesale electricity prices have surged**, raising baseline operating costs. Businesses may have long-term renewable PPAs or other sustainability commitments already in place. But **ESG spending is under scrutiny**, and there's growing uncertainty about the regulatory treatment of carbon accounting.²²

Organizations that may have considered paying extra for clean energy in the past — and how much — must decide whether the **instruments available to them** actually deliver meaningful environmental outcomes.

Fig. 7

Compared to a year ago, how much pressure do you feel to make IT infrastructure more sustainable?



90% of organizations report increased pressure to make IT infrastructure more sustainable

A mismatch between preference and procurement

Our survey data collectively paints a picture of a **market caught between** the stated **preferences** of consumers and the **options** available to them.

Nearly every respondent is willing to **pay more** for clean energy, and two-thirds believe renewables will supply most of the power for AI workloads within five years.

However, the **federal renewable tax credits** that have underwritten the clean energy buildout are being **phased out**, nuclear power remains years from contributing at scale, and natural gas, the **fastest-growing source** of new generation, is the one chosen by the fewest respondents.

Power may be the most **visible constraint** on AI deployment, but it doesn't operate in isolation.

Businesses are contending with a **system of limitations**, including compute access, networking capacity, siting and permitting, and regulatory timelines, that make any single bottleneck harder to resolve on its own.

“

A market caught between stated preferences and available options.”

SECTION I: POWER

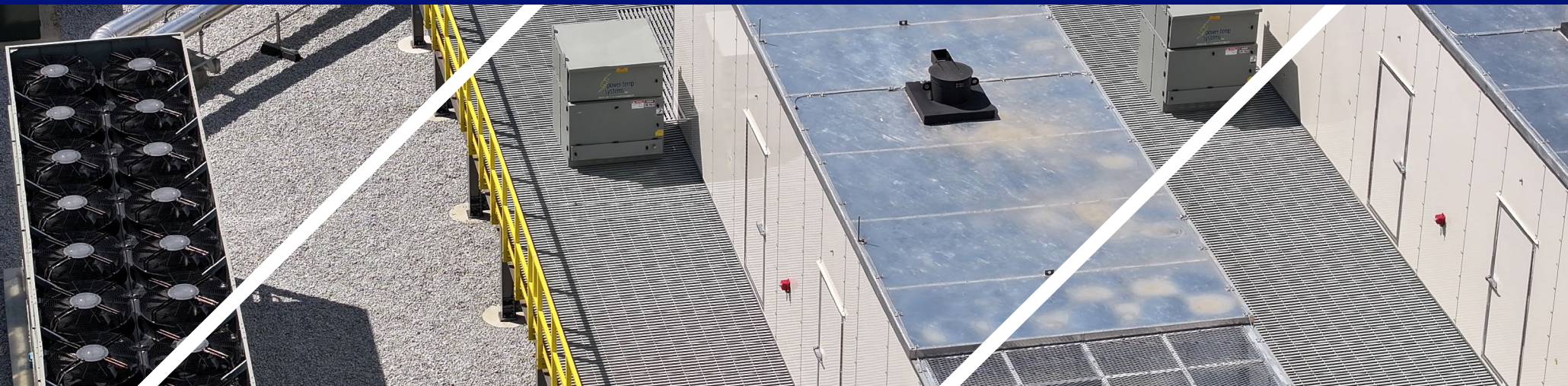
Key takeaways

01

Power availability and electricity pricing now shape where AI workloads can run. Reliable grid access influences deployment decisions for **89%** of respondents, and **55%** rank power cost differences as the top factor in choosing where workloads reside. At the same time, **72%** express concern about electricity price volatility, which will push organizations to treat energy access as a central constraint in infrastructure planning.

02

Most organizations expect renewable energy to supply AI workloads within five years, with **67%** identifying solar and wind as the leading sources and **42%** expecting them to be paired with battery storage to deliver most power. However, federal tax credits for new renewable projects are being phased out, so enterprises must plan deployments with this in mind and evaluate how energy sourcing influences where future infrastructure can operate.



Section II: Constraints Converge

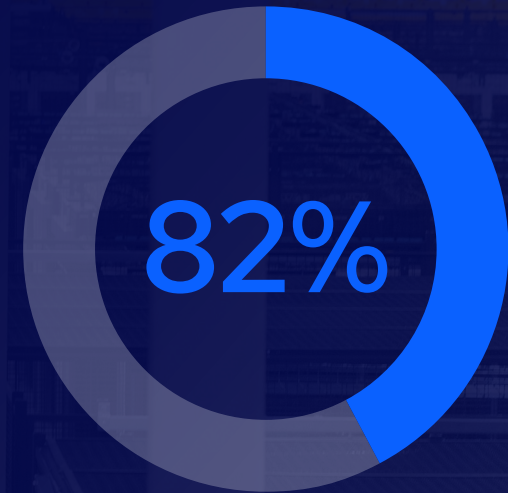
AI infrastructure is a system under pressure from every direction

The respondents in our survey aren't cautious observers of AI. These people are writing the checks and making the calls.

In all, 87% describe their organizations as **mature or leading-edge** in AI adoption, 82% say they are excited about the technology, and 86% are formal decision-makers in their company's AI buildouts.

And nearly every answer they gave describes a **system under pressure**. Over the past 12 months, 82% of respondents were moderately or severely constrained by the **availability of high-performance AI compute**, and 71% experienced excessive network latency. Meanwhile, 40% name IT infrastructure as the **single greatest barrier** to expanding AI initiatives, while 1% — just four respondents — pointed to their budget. Considering four major hyperscalers are devoting \$650 billion to AI investments during their 2026 fiscal years, the ability to spend **clearly** exists.²³

It is the supply side of that equation that's **tightening**. Recent earnings materials from a major hyperscaler indicate that demand for cloud and AI services continues to **exceed what's available**.²⁴ U.S. data center capacity under construction fell to 5.99 GW at the end of last year from 6.35 GW in 2024, the first decline since 2020.²⁵ Between 30% and 50% of the 140 data center projects slated for this year will be delayed into at least next year.²⁶



82% of respondents
were constrained by
high-performance AI compute

The national conversation about AI infrastructure constraints has been dominated by utility executives, grid operators, policymakers, and community activists. The responses to our survey come from the other side of that conversation – enterprise leaders trying to deploy AI within the constraints those groups describe.

A year ago, 61% of respondents identified the **management of specialized computing infrastructure** as their greatest constraint. {Fig. 8}. That figure fell to 35% this year, with all nine categories presented – including data science, security, employee training, and adoption willingness – among businesses’ top three constraints. No matter where organizations look, **obstacles** are present.

The clearest example runs through the heart of the stack – where organizations are putting their GPUs and where they are keeping their data.



No matter where organizations look, obstacles are present.

Fig. 8

In the past year, has your organization encountered constraints in any of the following areas related to AI?

(Select all that apply.)



The architecture mismatch driving a performance crisis

Over the past three years, **GPU deployment has consolidated toward the public cloud**. In 2024, 30% of respondents deployed most of their GPUs there; last year, 34% did, and this year, 54% do {Fig. 9}. GPU-as-a-service fell back to 30% after reaching 40% last year. GPU-as-a-service fell back to 30% after reaching 40% last year.

Fig. 9

Which of the following best describes where your organization deploys the most GPUs?

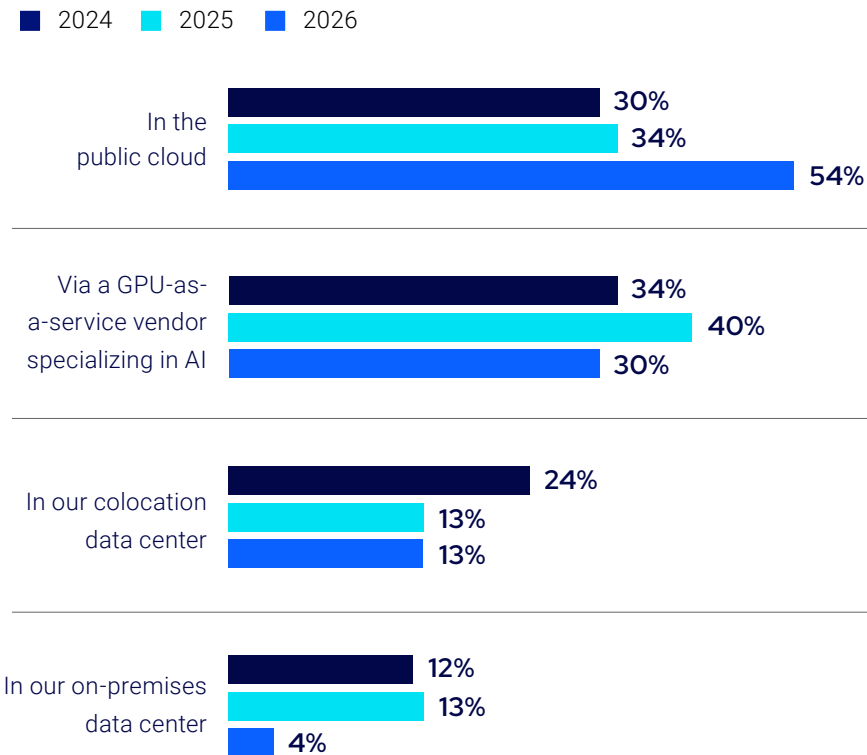


Fig. 10

Where is your AI data housed?

(Select all that apply.)

	2025	2026
In a hybrid cloud environment	48%	→ 56%
In a private cloud	60%	→ 52%
In a colocation data center	20%	→ 34%
In an on-premises data center	20%	→ 30%
In the public cloud	47%	→ 26%

AI data has been **migrating in the opposite direction**.

The share of respondents housing data in the public cloud dropped from 47% last year to 26% this year {Fig. 10}. Hybrid cloud now leads at 56%, private cloud follows at 52%, and colocation rises to 34% — up from 20% in one year. The environment where most GPUs run is increasingly the environment where the least data resides. That, by definition, is a **networking problem**.

Fig. 11

In the past 12 months, have you encountered any of the following performance issues with your AI applications or workloads?

(Select all that apply.)



Latency has moved faster than any other metric in our survey's three-year history.

The lowest-ranked performance issue in 2024 at 32%, it's now respondents' top AI performance issue by a wide margin at 71% {Fig. 11}. Bandwidth shortages remain elevated at 56%, and unreliable connections affect 45%. When asked how network availability has influenced deployment decisions, 59% say it has impacted their timing or sequencing.

The physical infrastructure facilitating those connections is **under its own pressure**. AI data centers have procured enough fiber-optic connectivity to **stretch lead times beyond 12 months**, with at least one manufacturer's inventory committed through this year.²⁷ Some GPU architectures for large-scale AI training and inference require **16 times as much fiber per rack** as traditional cloud configurations.²⁸

Enterprises competing for connectivity against hyperscalers who reserve supply years in advance **aren't bidding on the same timeline**. These aren't problems that get resolved in one procurement cycle.

Short-cycle planning meets long-cycle reality

Purchasing decisions that once held for months now expire in weeks. The infrastructure that those purchases depend on **takes years to build**.

Roughly four of five (81%) respondents are planning their IT infrastructure capacity less than four years out, with 64% targeting one to three years and 17% planning under a year {Fig. 12}. The share planning beyond three years dropped from 23% to 19%, while the five-year-plus cohort halved from 6% to 3%.

Even those timelines **may assume more stability** than today's environment provides. Over the past 18 months, the regulatory, commercial, and labor conditions surrounding AI infrastructure have each changed in ways that **can invalidate a procurement plan** between approval and execution.

AI-related regulatory uncertainty spanning data governance, AI oversight, and data centers is **meaningfully shaping enterprise planning**, with 94% of respondents reporting moderate (59%) to significant (35%) impact.

At the federal level, **policy direction has changed across administrations**, reflecting differing approaches to balancing innovation, safety, and state-level flexibility. At the same time, **momentum at the state level has accelerated**, with all but four states advancing or enacting AI-related legislation over the past 15 months. This regulatory environment underscores the growing importance of adaptable, compliant infrastructure strategies as organizations adopt AI.

Commercial terms are changing abruptly, too. Major IT vendors have **shortened quote-validity windows** from 30 days to 14 days and revised compute-order terms to allow pricing changes tied to component costs, tariffs, and exchange rates, with cancellation permitted up to 45 days before shipment.^{29,30}

The **workforce required to overcome these hurdles** is itself constrained. The scarcity of skilled AI and infrastructure staff has intensified, as the share of respondents naming it as their greatest barrier to AI expansion grew from 10%

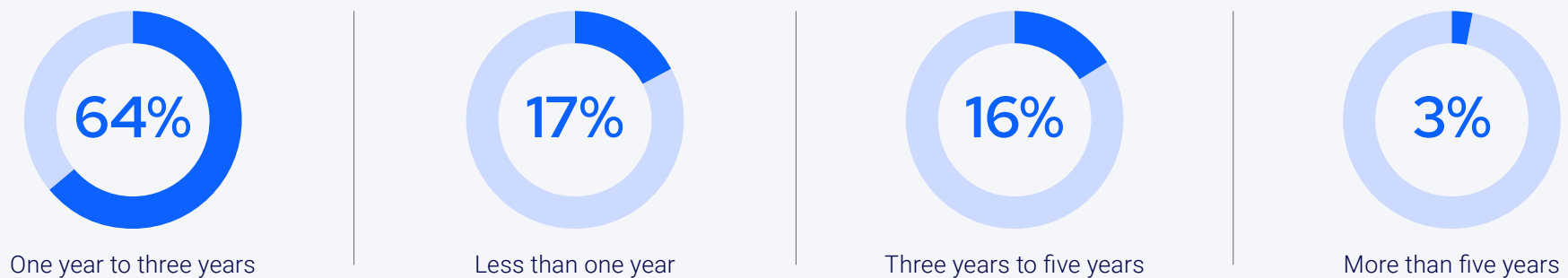
last year to 22% this year. Outside organizations, the physical construction of data centers and grid infrastructure faces its own workforce challenges. **The construction industry** will need to attract roughly 349,000 new workers this year – and 456,000 next year – **to meet overall demand**.³¹ Electrical work, which accounts for 45% to 70% of total data center construction costs, is also facing a labor shortage.³²

These aren't separate problems. Hyperscalers, utilities, renewable energy developers, and grid projects are **all drawing from the same labor pool**, placing AI infrastructure construction in direct competition with the power systems meant to support it – and with grid upgrades, energy projects, and industrial buildouts competing for the same skilled trades.

When the constraints on building, powering, connecting, and staffing AI infrastructure **trace back to the same scarce resources**, it becomes a question of whether they can be addressed independently at all.

Fig. 12

How far in advance is your organization planning its IT infrastructure and data center capacity needs?



Constraints often don't occur in a vacuum

Multiple **infrastructure pressures are converging** for the 42% of respondents who believe access to high-capacity fiber and multiple carrier options is equally as important as power and space when selecting locations for AI workloads.

This group treats **connectivity** as a **top constraint** alongside the two resources most commonly associated with data center site selection – and their responses **diverge sharply** from the rest of those surveyed.

Just over half (52%) report that networking very often limits AI deployment, compared with 23% of respondents who don't place fiber on equal footing with power and space – a 29-point gap {Fig. 13}. In addition, 59% cite power as a sustainability constraint compared with 30% among others, and 51% say regulatory uncertainty significantly impacts their AI planning, compared with 23% of those outside the group.

A similar pattern emerges around **compute scarcity** and **geography**. Of the 37% of respondents who say vendor capacity limits their access to AI compute, 67% report that compute availability influences where they deploy workloads, compared with 23% of those who don't cite vendor capacity as a limiting factor – a 44-point gap {Fig. 14}. In the same group, 57% move workloads to alternate regions when capacity is unavailable versus 23% among the rest.

When **one** of these **pressures is severe**, the others usually are, too.

Fig. 14

Compute scarcity is reshaping geography

Among the 131 respondents who say vendor capacity limits their access to AI compute ...

67% | 88 (67%) say availability of AI compute influences where they deploy workloads, compared with 51 of 223 (23%) of all other respondents.

57% | 75 (57%) say they shift workloads to alternate regions when capacity is constrained, versus 51 (23%) of others.

Fig. 13

Networking and power move together

Among the 147 respondents who say fiber connectivity is as important as power and space ...

- ▶ **77 (52%)** report that networking very often limits AI deployment, compared with 47 of 207 (**23%**) of others.
- ▶ **87 (59%)** cite power as a sustainability constraint, versus 62 of 207 (**30%**) among the others.
- ▶ **75 (51%)** say policy uncertainty significantly affects planning, compared with 48 of 207 (**23%**) of those outside the group.

Confidence reflects the plan, not the ambition

Nearly every respondent (97%) expressed **at least some** confidence in their ability to **secure sufficient AI capacity over the coming year**, and 43% describe themselves as very confident. Those figures are difficult to reconcile given the concerns they share.

When respondents encounter compute constraints, identify infrastructure as their greatest barrier to AI expansion, experience high latency more frequently, and reduce scope or phase deployments, **the infrastructure isn't keeping pace with the strategies it's supposed to support**. The hardware may exist, but it's not where these organizations need it, nor is it available at terms they can absorb.

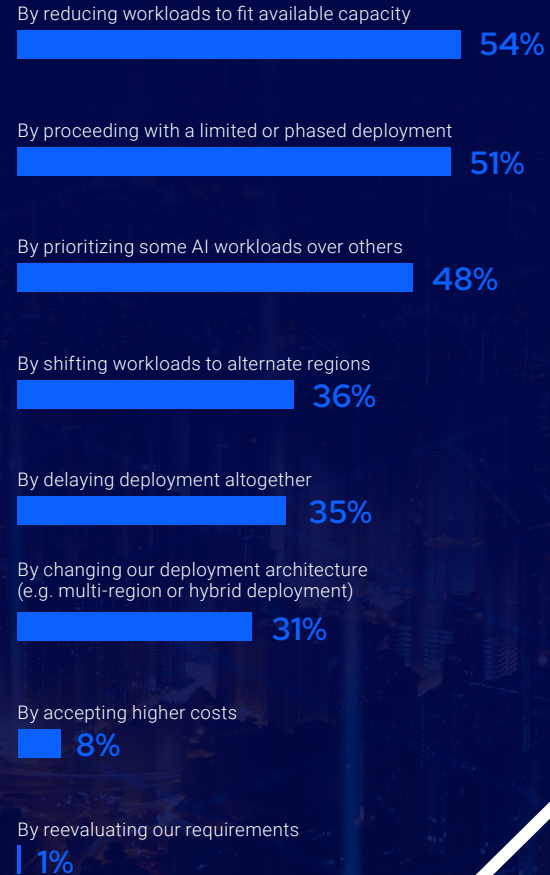
The high level of confidence may stem from respondents having already **adjusted their ambitions to reflect what's realistic**. For example, if a preferred region lacks capacity, 54% of respondents reduce workloads to fit what's available, 51% proceed with a limited or phased deployment, and 48% prioritize certain AI workloads over others. Only 8% accept higher costs {Fig. 15}. If they define "sufficient" as what their business can realistically secure rather than what its AI strategy initially required, confidence and constraint can coexist.

Whether **expectations remain low** depends on conditions that businesses don't control and, in several cases, actively work against them.

Fig. 15

How does your organization typically respond when a preferred region lacks capacity?

(Select no more than three responses.)



SECTION II:

Key takeaways

01

Physical infrastructure now limits AI expansion more than funding does, as **40%** of respondents identify infrastructure as their organization's primary barrier to growth. As a result, most report moderating deployments, with **54%** reducing workloads, **51%** phasing rollouts, and **48%** prioritizing specific applications over others when constraints arise.

02

Compute and data are drifting into separate environments, placing heavy strain on the networks that connect them. A majority of organizations now deploy most GPUs in the public cloud, rising from **30%** in 2024 to **54%** today, while only **26%** store most AI data there. As hybrid, private, and colocation environments gain traction, latency concerns now affect **71%** of respondents and influence deployment timing for **59%**.



Section III: Network Constraints

AI's other infrastructure crisis is hiding in plain sight

Ask where AI infrastructure is failing, and the answer, overwhelmingly, is between facilities — not inside them.

When respondents were presented with **performance issues** that may have affected their AI workloads over the past 12 months, 96% pointed a finger at network constraints. Excessive latency impacted 71% of those surveyed, 56% reported bandwidth shortages, and 45% encountered unreliable connections. One-quarter (25%) had difficulty scaling data center space and power, but only 13 survey respondents (4%) cited this as their sole issue. Although this category has **dominated public attention** and captured the **largest share of capital investment**, it's the one where respondents report the least friction.

That disparity is **consistent with how AI workloads have matured**. As Section II documented, compute is consolidating in public cloud environments while data is **migrating** toward hybrid, private, and colocation infrastructure — a divergence that places enormous pressure on the connections between them.

Fiber-optic connectivity is **in the middle** of that pressure. It determines whether the data volumes required for training, inference, and storage can **move at the speed** AI workloads demand — and, increasingly, where and when they can be deployed at all.

96% point to network constraints

71% experienced excessive latency

56% report bandwidth shortages

Fiber as a constraint

Although the national AI infrastructure conversation has **focused heavily on power**, respondents consider fiber **central** to deployment decisions. Half (50%) said fiber is among several major considerations, and 42% said it's as important as power and space, with only 9% downplaying its role.

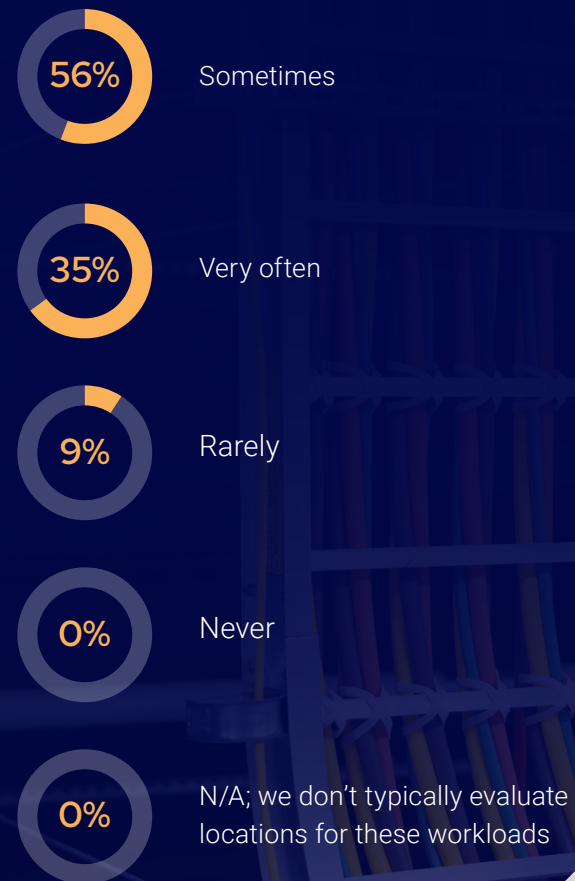
However, the supply chain **can't yet support** that priority. AI-oriented data centers have 10 to 36 times more fiber than traditional facilities, and modern GPU training clusters require thousands of individual fiber connections to move traffic between accelerators during model training.³³ Procurement lead times have **extended from two or three months to more than a year**, with ribbon fiber, which is required for dense AI clusters, in especially short supply because manufacturing processes can't keep up.

The U.S. has roughly **159.6 million fiber miles** installed and will need **213.3 million more by 2029**, with fiber route miles nearly doubling from 95,000 to 187,000.³⁴ One supplier of optical networking infrastructure entered the year with approximately \$5 billion in backlog.³⁵ In January, a major hyperscaler signed a \$6 billion multiyear agreement with a fiber and connectivity supplier to secure fiber and connectivity hardware for its U.S. AI data center buildouts.³⁶

Even when power is available, **network readiness frequently intervenes**. Fiber availability, carrier diversity, interconnection options, and low-latency connectivity have been a limiting factor for 91% of respondents, with 35% saying they are "very often" present at sites under consideration {Fig. 16}.

Fig. 16

When evaluating locations with favorable power availability for AI workloads, how often does network readiness (fiber availability, carrier diversity, interconnection options, and low-latency connectivity) become a limiting factor?



The consequences have already materialized, as 59% say network availability has **affected their deployment timing or sequencing** and 29% have been **directly limited** in where they can run workloads {Fig. 17}.

Over the past 12 months, 54% report that fiber availability increased their costs, 54% say it delayed their AI rollouts, and 50% confirmed it changed their deployment locations {Fig. 18}.

Because many respondents are experiencing some of these consequences **simultaneously**, like absorbing higher costs while also accepting delays or relocating workloads, their response has had to adjust.

Fig. 17

Which of the following best describes how network availability has affected your organization's ability to deploy AI workloads?

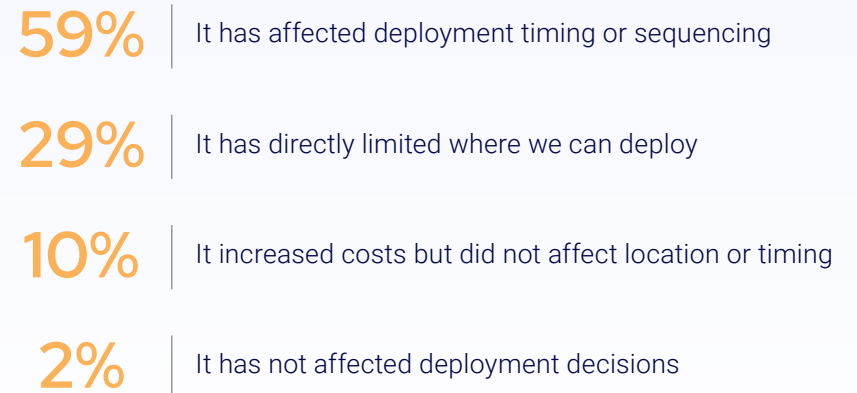
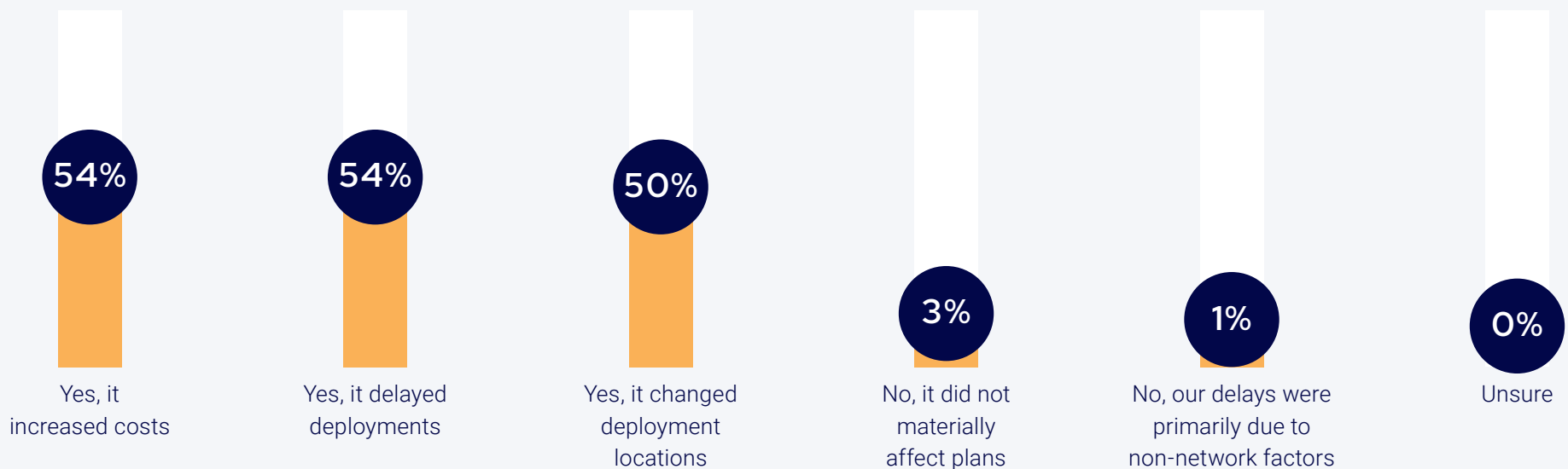


Fig. 18

In the past 12 months, has fiber availability or network delivery timelines affected your organization's AI deployment plans?

(Select all that apply.)



How organizations are restructuring their approach

These pressures are producing **measurable changes** in how enterprises architect their AI infrastructure.

Third-party colocation for edge processing rose to 46%, up from 37% last year {Fig. 19}. On-premises AI hardware fell from 32% to 24%. Software-defined networking reached 58% – 20 percentage points higher than it was in 2024 – for the **steepest three-year increase** among the six approaches we have monitored. Enterprises are concentrating on distributed, software-managed infrastructure operated by third parties.

Last year, **Ethernet overtook InfiniBand** in AI back-end network deployments for the first time, reversing InfiniBand’s roughly 80% market share from two years earlier.³⁷ The Ultra Ethernet Consortium released its 1.0.1 specification last summer with backing from leading semiconductor, networking, and hyperscale companies.³⁸ This transition reflects **growing industry support for Ethernet-based AI fabrics**, which are more widely available and generally less expensive than specialized interconnect technologies.

Carrier diversity, meanwhile, is declining primarily because of industry consolidation. Several major acquisitions in the telecommunications sector over the past year collectively exceeded \$35 billion in value and reduced the number of independent network operators serving U.S. markets, even if they expand fiber reach.

Genuine diversity **depends on where fiber physically runs**, because connections that appear independent on a network map can share the same conduit or chokepoint.³⁹ Without transparency from providers into the physical routing of their infrastructure, a redundancy plan may be built on inherited risk that only becomes visible during a failure.

As AI infrastructure pushes into secondary markets like Milwaukee and Wichita, Kansas, the fiber depth, route independence, and carrier presence that established hubs built over decades may not exist – and **building it takes longer** than securing power or land.^{40,41}

Fig. 19

Which of the following tactics is your organization implementing to reduce performance issues for its AI applications or workloads?

	2024	2025	2026
Implementing software-defined networking (SDN)	38%	→ 55%	→ 58%
Using 5G networks (e.g., 5G-enabled IoT devices, etc.)	54%	→ 65%	→ 56%
Using Wi-Fi 6 or Wi-Fi 7	N/A*	→ 48%	→ 55%
Using network function virtualization (NFV)	45%	→ 51%	→ 47%
Using third-party colocation data centers to process data closer to the edge of the network	51%	→ 37%	→ 46%
Deploying AI hardware on-premises	24%	→ 32%	→ 24%
None of the above	4%	→ 0%	→ 0%

*Using Wi-Fi 6 or Wi-Fi 7 not presented as an option this year

Workloads change faster than infrastructure

The applications driving enterprise adoption, including customer-facing assistants and autonomous decision engines, are the ones where **network delay most directly degrades performance**.

Three-quarters of enterprise decision-makers expect to need sub-30-millisecond latency within two to three years, but no U.S. state had latency below that threshold at any point last year.^{42,43} A single large language model (LLM) call takes 800 milliseconds, but an Agentic AI workflow, which chains multiple calls together, can take 10 to 30 seconds.

It's no surprise, then, that **constraints like these carry direct financial weight**. Organizations are already being forced to **shoulder delayed deployments** and **rising fiber procurement costs**, and the workloads now entering production will place greater demands on infrastructure that's already falling short.

How enterprises **absorb** those costs, and whether their AI investments are generating returns that are sufficient to justify them, is coming into focus just like the constraints themselves.



The workloads now entering production will place greater demands on infrastructure that's already falling short."

SECTION III:

Key takeaways

01

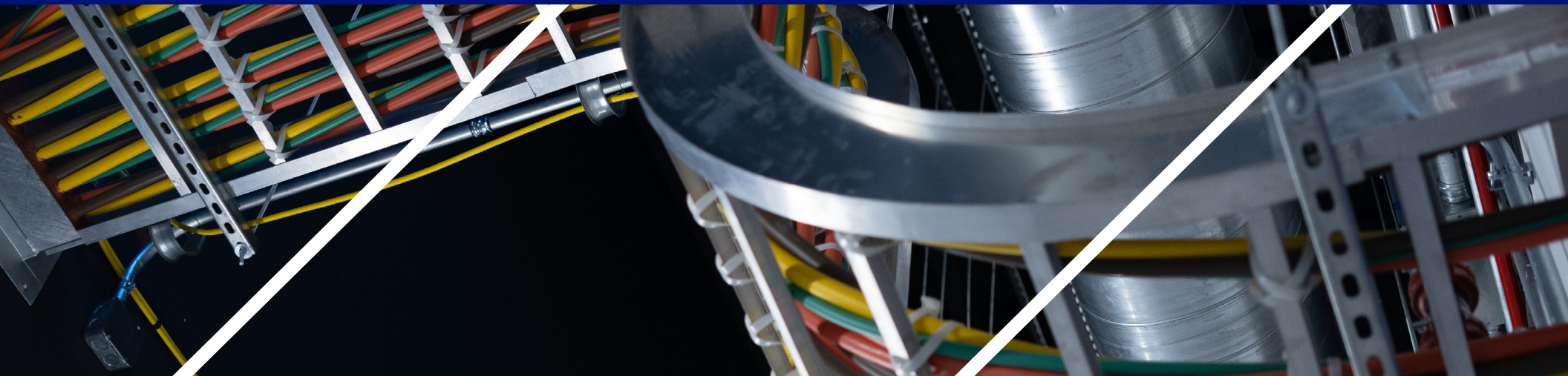
Network performance issues have become the most widespread operational challenge facing AI deployments.

During the past year, **96%** of respondents experienced at least one network-related problem, with latency affecting **71%**, bandwidth shortages reported by **56%**, and unreliable connections impacting **45%**. Difficulties scaling data center space or power were cited as the sole issue by only **4%**.

02

Fiber availability increasingly shapes where AI workloads can operate.

Connectivity constraints influenced deployment decisions for **91%** of organizations evaluating new sites, while **54%** reported higher costs and **54%** endured rollout delays linked to fiber supply. Limited carrier diversity, long procurement lead times, and expanding demand for dense AI clusters now make connectivity a deployment constraint.



Section IV: Tariffs and policy uncertainty

Organizations are building long-term AI strategies against short-term rules

For years, the obstacles to AI expansion were primarily technical. Now they include customs duties and procurement attorneys.

U.S. tariffs and import **restrictions have become a threat** for AI infrastructure buyers. More than half (54%) of respondents say tariffs have already increased their reliance on **domestic suppliers**, while 40% have delayed or scaled back their purchasing and an equal share have locked into long-term contracts (Fig. 20). With 92% of respondents identifying multiple strategies their organizations have taken, they have indicated that **no single defensive posture is sufficient** and that hedging is the most rational approach to insulating themselves from the financial fallout.

That approach arrives at a moment when respondents are increasingly focused on **whether they can deliver on their AI commitments at all**. Delayed time to market, cited by 37% of respondents, has become the most consequential risk of failing to achieve the goals organizations set out in their AI roadmap (Fig. 21). Lost market share, which led at 28% last year, dropped to 21%.

Respondents aren't standing still. Their procurement strategies have already changed in measurable ways, though the direction of those changes raises questions about what the domestic supply chain can realistically deliver.

Fig. 20

How have recent U.S. tariffs and related import restrictions on technology and industrial goods impacted your organization's AI vendor or supplier strategy?

(Select no more than two responses.)

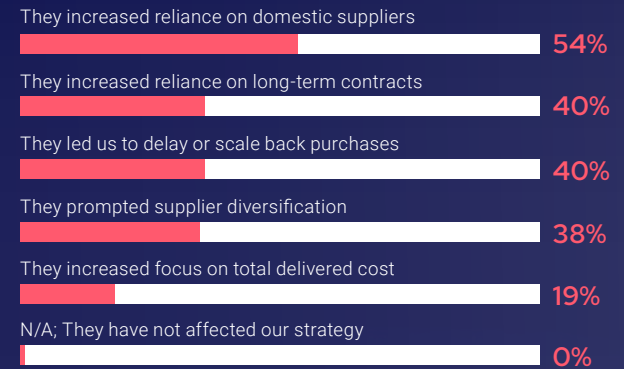


Fig. 21

What is the most impactful consequence if your organization does not achieve the goals laid out in its AI roadmap?



The domestic pivot and its limits

Of the 54% who said their organizations increased their reliance on domestic suppliers as an immediate reaction to tariffs, 86% didn't select domestic sourcing as their most important long-term strategy {Fig. 22}. They selected something else — primarily **diversification across suppliers and geographies**, which was chosen by one-third (33%) of all respondents.

This behavior suggests that organizations are doing what the rules incentivize, and that going domestic is simply the **fastest available response** to the current **tariff conditions** — not a lasting commitment to onshoring.

The semiconductor supply chain illustrates why. The U.S. accounts for around 10% of global semiconductor manufacturing capacity, leaving it heavily reliant on overseas production, and the effort to increase that figure is **moving more slowly than procurement timelines demand**.⁴⁴ Recent federal legislation intended to spur semiconductor manufacturing committed \$30.9 billion across 19 companies and 40 projects, but as of July, only one \$6.6 billion grant had been disbursed because funding is tied to completion.⁴⁵ That went to a manufacturer that opened the first of several planned fabrication plants in Arizona last summer.⁴⁶ Even if organizations wish to onshore their sourcing and manufacturing, the U.S. supply base may not yet be in place.

But diversification across suppliers and geographies **doesn't always eliminate supply constraints**, either. Businesses that consider this their most important tariff-mitigation strategy are finding that the companies dominating memory production have prioritized allocation to hyperscale and AI clients and left other buyers with longer lead times and volatile pricing.⁴⁷

DRAM prices rose more than 50% in some categories in 2025, then increased 90% to 95% in the first quarter of this year.^{48,49} One manufacturer reported that supplier inventories fell from 13 to 17 weeks in late 2024 to two to four weeks by the end of last year, and all memory set to be produced by the end of this year has already been purchased.⁵⁰

Cost and availability are two dimensions of the problem respondents are managing. The third is predictability, and the policy environment has made it the scarcest resource of all.

Fig. 22

The most important long-term tariff strategy

Among the 191 respondents who say their organizations increased their reliance on domestic suppliers as an immediate reaction to tariffs ...

- ▶ **27 (14%)** say domestic sourcing and manufacturing is most important for managing long-term tariff exposure for AI hardware and infrastructure purchases.
- ▶ **164 (86%)** chose anything else:
 - **55 (29%)** Supplier and geography diversification
 - **47 (25%)** Increased operational efficiency
 - **44 (23%)** Long-term pricing agreements
 - **16 (8%)** Financial hedging
 - **2 (1%)** Cost pass-through strategies

The rules themselves are still being written

Organizations building a multi-year AI infrastructure strategy are forced to do so under **economic and legislative conditions** that may change – perhaps significantly – by the time they begin implementing it.

Nearly all respondents (94%) say policy uncertainty affects their AI planning, with 35% describing the effect as significant. In addition, 83% consider clarity around AI governance and accountability to be very important or essential. Organizations with AI ambitions are caught between a **federal government that's promoting AI acceleration** while **state and local jurisdictions hold authority to restrict** AI and/or the physical facilities it requires.

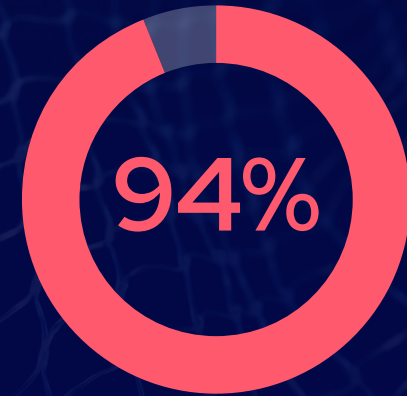
The practical result is that businesses looking to **optimize** their AI deployments depend on which branch or level of government moves first – and they are limited in their ability to predict or control that sequence.

Planning in an environment that changes by the quarter

Any one of these pressures – rising costs, constrained supply, or unstable rules – would **complicate an AI infrastructure strategy**. The challenge respondents face is that they are managing all three simultaneously, and successfully addressing one often depends on the stability of the others.

Domestic sourcing becomes harder when domestic capacity lags years behind demand. Long-term contracts become riskier when the tariff framework may not survive the quarter. Diversification becomes more complex when suppliers allocate their output to a different class of customer.

Businesses are implementing **ambitious** AI strategies. The environment they are building them in hasn't demonstrated that it will hold still long enough to see them through.



Nearly all respondents (94%)
say policy uncertainty
affects their AI planning.

SECTION IV:

Key takeaways

01

Tariffs and trade restrictions have already changed how organizations purchase AI infrastructure. More than half of respondents (**54%**) report increasing reliance on domestic suppliers in response to tariff pressure, **40%** have delayed or scaled back purchasing, and **40%** have locked in long-term contracts. Procurement strategies now prioritize cost containment as organizations seek to manage rising hardware and supply chain costs.

02

Policy volatility has become a structural challenge for long-term infrastructure planning. Nearly all respondents report that regulatory uncertainty affects their AI strategies, with **94%** acknowledging some level of impact and **35%** describing it as significant. Enterprises building multi-year AI roadmaps must overcome conditions that can change within a single procurement cycle.



AI spending continues, but how returns are assessed is being reevaluated

One of our survey's primary findings a year ago was that around half of enterprise AI decision-makers expected measurable financial returns within 12 months.

That confidence has eroded — **not into pessimism**, but into **patience**. The share of organizations expecting returns within a year fell from 51% to 36% (Fig. 23). The one- to three-year window absorbed much of that movement, growing from 23% to 32%, while the three- to five-year horizon **nearly doubled** from 5% to 9%. Meanwhile, the share **already seeing financial benefits** held essentially flat, moving from 21% to 20%.

Expectations have moved **outward**, but outcomes **haven't moved forward**. Organizations have reassessed when they believe the payoff will arrive. That recalibration is consistent with what researchers at MIT Sloan found when studying AI use in U.S. manufacturing: A 1.33 percentage-point productivity decline in the period following adoption before longer-run gains took hold.⁵¹ One survey reinforced

that pattern as firms reported that **AI had produced "no change" in outcomes** over the prior 12 months, even as they expected improvement in the next 12.⁵²

And yet, no respondent to our survey said they don't expect AI to generate **measurable financial benefits**. One consultancy found that Two-thirds (67%) of leaders would **maintain** AI spending even during a recession.⁵³ Industry research forecasts \$2.52 trillion in global AI spending this year, a 44% increase over last year.⁵⁴ The commitment to AI is deepening even as returns are proving slower to materialize.

Enterprise leaders sit downstream of those same economics, and the pressure to demonstrate returns from AI deployments only **intensifies** as the investment above them grows.

Fig. 23

When does your organization expect to generate measurable financial benefits from its AI investments?

	2025	2026
Within the next year	51%	→ 36%
In one to three years	23%	→ 32%
We are already seeing financial benefits	21%	→ 20%
In three to five years	5%	→ 9%
More than five years from now	0%	→ 2%
We do not expect to generate measurable financial benefits	0%	→ 0%

Fig. 24

Approximately what percentage of your total IT budget is allocated to AI initiatives (software, hardware, and networking)?

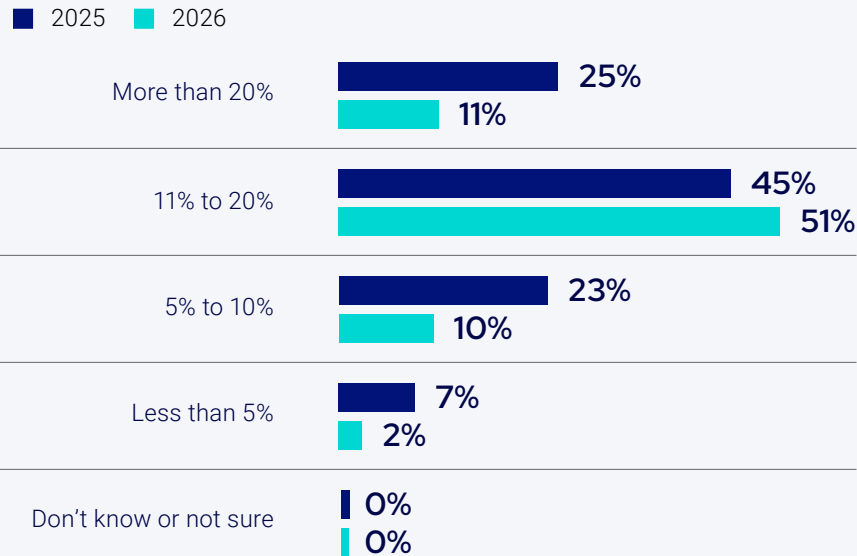


Fig. 25

What types of AI or ML use cases are you deploying or planning to deploy?

(Select all that apply)

	2025	2026
Generative AI <i>(e.g., content generation, code suggestions, chatbots, AI agents)</i>	90%	52%
AI-driven cybersecurity <i>(e.g., threat detection, fraud prevention)</i>	62%	51%
Quality control and inspection <i>(e.g., process monitoring, defect detection, compliance)</i>	N/A*	42%
Decision support and optimization <i>(e.g., pricing or supply chain optimization, resource allocation)</i>	N/A*	40%
Predictive analytics <i>(e.g., forecasting, anomaly detection)</i>	69%	38%
NLP-based applications <i>(e.g., audio or voice recognition, transcription, sentiment analysis)</i>	50%	36%
Autonomous systems <i>(e.g., robotics, self-driving technology, document and workflow automation)</i>	60%	36%
Recommendation and personalization <i>(e.g., product recommendations, content ranking, next-best-action)</i>	N/A*	32%
Computer vision <i>(e.g., image analysis, facial recognition, video surveillance)</i>	57%	32%

*Quality control, decision support, and recommendation were not presented as options in 2025

AI budgets stabilize but demands multiply

The distribution of AI spending across IT budgets has **consolidated**. The number of organizations allocating more than 20% of their IT budget to AI dropped from 25% to 11% {Fig. 24}, while those devoting 11% to 20% of their budget grew from 45% to 51%, and the 5% to 10% range expanded from 23% to 36%. At the bottom, the share spending less than 5% shrank from 7% to 2%. The early big spenders have **pulled back**, the laggards have ramped up, and the market is treating AI as a defined, durable budget category.

What that budget now must cover, however, has **broadened considerably**. A year ago, generative AI dominated the use-case portfolio at 90% of respondents {Fig. 25}. That figure, now 52%, still leads the category, but it's within a single percentage point of AI-driven cybersecurity at 51%. Quality control and inspection (42%) and decision support and optimization (40%) – categories that weren't presented in our survey a year ago – appear in the upper tier. Organizations are **distributing AI investment across cybersecurity, quality control, predictive analytics, autonomous systems, and decision optimization**, with each carrying different requirements for compute, latency, and networking.

What follows is a **mismatch** between ambition and budget. Most organizations are missing their AI cost forecasts, with overruns severe enough to erode stakeholder confidence.⁵⁵ GPU pricing, reserved capacity, data transfer, storage, and engineering time are cost categories that organizations routinely underestimate.⁵⁶ And the money to cover those costs is coming, in part, from within. Headcount growth expectations fell from 6% last year to 2% today, with HR budgets absorbing the steepest cuts.⁵⁷

AI spending is growing partly by **displacing the resources** that once went to hiring.

ROI has moved from expansion to efficiency

The financial pressure visible in those budget dynamics affects how organizations **measure AI's contribution**. Last year, 58% of respondents assessed its impact through revenue growth or an increased market share. This year, however, just 42% included those measurements in their top three, dropping them from first place to fourth {Fig. 26}. Cost reduction and operational efficiency, grouped together, were identified by 55% of respondents as their strongest indicator of AI's success, and customer satisfaction tied for second with time to market or process improvement at 49%.

Employee productivity, at 15%, fell from an already modest 20% — a decline that occurred in the same period when one consultancy found 46% of leaders experienced employee resistance to AI adoption, and 76% were willing to pay candidates with demonstrated AI skills a salary premium of 10% or more.⁵⁸ Organizations are **spending more to attract AI talent** while deprioritizing the metric that would most effectively measure AI's contribution.

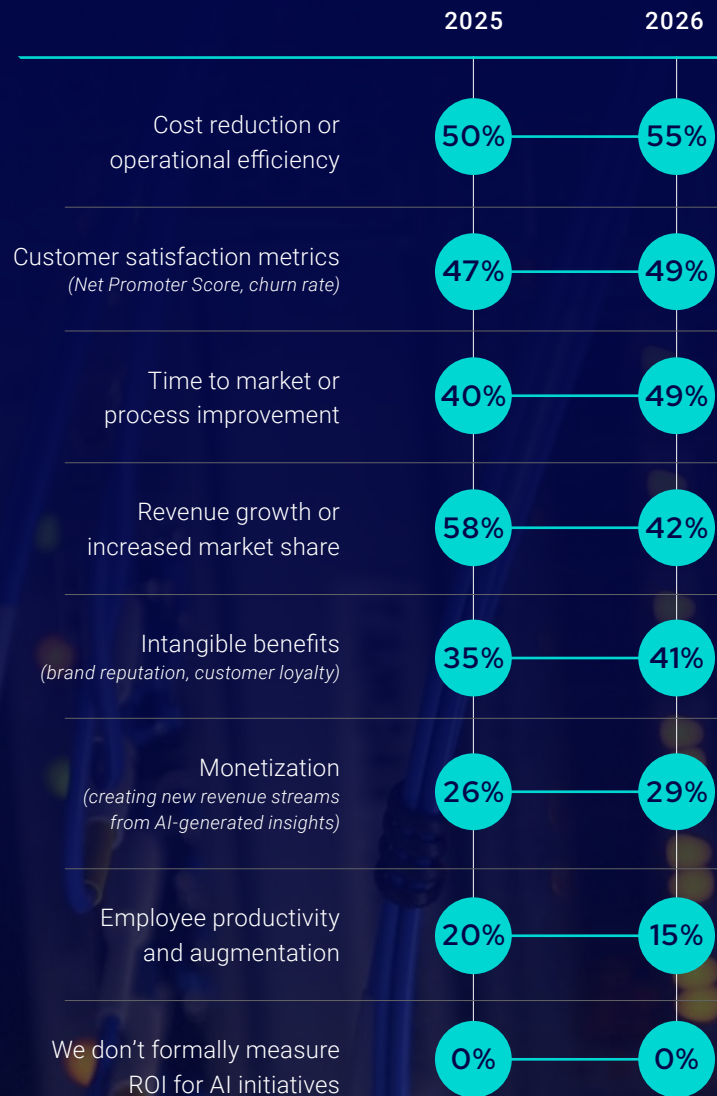
The broader cost environment helps **explain the difference** in approach. Electricity and natural gas prices were the largest drivers of inflation last year, rising 7% and 11%, respectively.⁵⁹ Utilities requested a record \$31 billion in rate hikes, more than double the prior year.⁶⁰ A 25% tariff on steel imports raised material costs across the supply chain, while tariffs on imported semiconductors, networking gear, and power equipment compounded procurement pressure.⁶¹ Enterprises, by multiple estimates, underestimate AI infrastructure costs by roughly 30%.⁶²

When hardware, energy, and materials costs all escalate simultaneously, the case for AI increasingly rests on what it can save rather than what it can earn.

Fig. 26

How do you measure the return on investment of your AI initiatives?

(Select no more than three responses.)



Revenue-driven AI demands high-performance infrastructure

A closer look at our respondents' behaviors **exposes the differences** in how organizations pursue AI value and how constrained they are by physical infrastructure. Of the 139 respondents who say compute availability influences their siting decisions, 58% measure AI ROI primarily through revenue growth, while among the remaining 215, only 31% do {Fig. 27}. The same compute-sensitive group, meanwhile, is 31 points more likely to have **relocated workloads across geographic regions** in response to capacity constraints.

Their reasoning is architectural. Revenue-generating AI applications like **real-time personalization**, **dynamic pricing**, and **autonomous decision-making** require sustained, high-performance compute with low-latency connectivity. On the other hand, cost-reduction applications like process automation, document handling, and back-office workflows can tolerate higher latency and run on less demanding infrastructure. Organizations with the most **aggressive financial ambitions** for AI are the **most exposed** to the physical realities of where compute, power, and network capacity actually exist.

A survey of enterprise IT decision-makers found that 80% of businesses missed their AI infrastructure forecasts by more than 25%, which shows that AI spending is proving hard to predict. When it comes to AI budgets, 66% goes to data, storage, and processing power, with only 33% directed toward AI software and SaaS. When infrastructure accounts for a considerable amount of spending, where to run workloads becomes a financial decision, albeit one with geographic consequences.

ROI claims face a growing credibility test

There's a further **complication beneath the confidence** these respondents express. Although 72% of organizations formally measure generative AI ROI, only 10% of CFOs fully trust the quality and reliability of their enterprise data and 86% identify technical debt as a barrier to AI readiness.^{63,64}

The distance between "we measure ROI" and "we have data infrastructure reliable enough to measure it credibly" raises a question about the 20% of respondents to our survey who say they are already seeing financial benefits: Have those benefits **survived the kind of financial audit** that CFOs and boards are now beginning to demand?

The organizations demonstrating AI returns are evaluating their investments against earnings contributions, margin improvements, and customer retention — the same standards applied to any other capital allocation.⁶⁵ One industry research firm strongly cautioned that **ROI must be more predictable** before AI adoption can expand.⁶⁶

The change in how organizations define AI success reflects an enterprise-wide reckoning. Today, AI investments are being held to the same evidentiary standards as every other line in the budget.

Fig. 27

Compute availability aligns with a revenue-driven AI posture

Among the 139 respondents who say AI compute availability influences siting decisions ...

- ▶ **81 (58%)** say AI ROI is primarily driven by revenue growth, compared with **66 (31%)** of others.
- ▶ **76 (54.7%)** report shifting regions in response to capacity constraints, versus **50 (23.3%)** among the rest.

SECTION V:

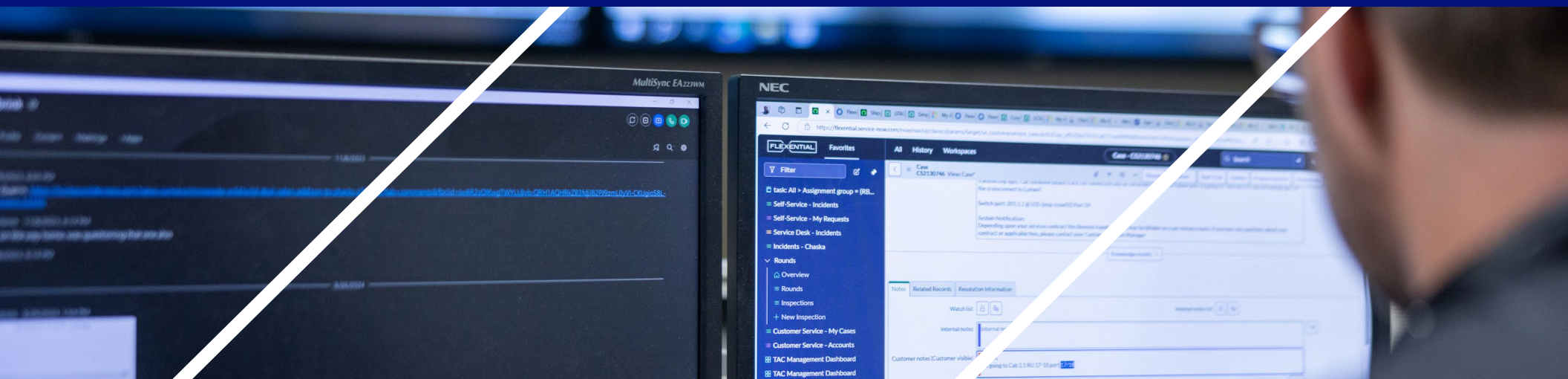
Key takeaways

01

Expectations around AI returns are moving further into the future as spending increases. Organizations expecting measurable financial benefits within a year declined from **51%** to **36%**, while the share projecting returns within one to three years rose to **32%**. With roughly **20%** of organizations reporting financial benefits for the second consecutive year, adoption timelines appear to be lengthening rather than accelerating.

02

Financial pressure is reshaping how organizations define AI success. Measures tied to revenue growth or market share declined from **58%** last year to **42%**, while cost reduction and operational efficiency became the most cited performance indicator at **55%**. Customer satisfaction and time to market followed closely, reflecting a growing emphasis on operational improvements.



Conclusion

The era of treating AI infrastructure as a purely technical challenge is over

Even the best-planned AI strategy encounters an infrastructure challenge. This year's findings suggest most organizations have already run into one.

Power availability, regulatory policy, facility siting, sustainability commitments, and compute demands are **no longer variables** that can be addressed in isolation. They have converged into a single, interconnected **system of constraints**, and the organizations that overcome it will lead the next phase of AI-driven innovation.

The imperative for IT leaders is to:

- **Secure power capacity before it becomes a bottleneck.** AI workloads are pushing electrical demands well beyond what legacy data centers were designed to support. Organizations need infrastructure partners with access to reliable, high-density power that's available today – not just promised for the future.
- **Choose locations strategically.** Siting decisions now involve more than latency and proximity. Access to power and fiber will determine which markets can actually support AI-scale deployments. A nationwide infrastructure partner provides the optionality that single-market strategies cannot.
- **Treat sustainability as an operational requirement, not a reporting exercise.** Advanced cooling technologies, energy-efficient facility design, and sustainable operating practices are prerequisites for long-term infrastructure viability and stakeholder confidence.
- **Align infrastructure strategy with a regulatory environment that won't stand still.** Data sovereignty mandates, energy use disclosure requirements, and AI governance frameworks are constantly in flux. IT leaders must ensure their infrastructure decisions account for where and how regulations are tightening, and partner with providers that understand what these changes mean for their business.
- **Build for the AI workloads of tomorrow.** The pace of model development, inference demand, and data gravity means that infrastructure must be designed to scale – in density, in connectivity, and in flexibility – well beyond today's requirements.

A power constraint in one market creates a **siting problem**.

A policy change rewrites the **economics of sustainability**.

A spike in compute demand exposes **connectivity** issues that take months to resolve.

The organizations that come out ahead will work with partners who understand these interdependencies and can address them across a broad, interconnected footprint.

That's exactly what Flexential was built to do.

With more than 40 data centers across 18 U.S. markets, Flexential provides the **geographic reach, high-density compute capacity, and interconnection fabric** that AI-driven enterprises need to scale with confidence. The FlexAnywhere Platform integrates colocation, cloud, networking, and data protection into a unified infrastructure environment, enabling organizations to place workloads where power, policy, and performance align.

Whether you're **scaling AI training environments, deploying inference across distributed locations, or future-proofing your hybrid IT architecture**, Flexential has the infrastructure, expertise, and national presence to help you move forward.

Let's build what
comes next – together.



Methodology

In February 2026, Flexential surveyed 354 IT decision-makers at the director level or above at organizations with over \$100 million in annual revenue. All respondents had knowledge of their organizations' AI implementation and related infrastructure buildouts. Respondents came from a range of industries.

Annual Revenue	
\$100,000,001 to \$500 million	43%
More than \$2 billion	29%
\$500,000,001 to \$2 billion	28%

Job Level	
Vice President	40%
Director	38%
C-Suite	22%

Industry	
Technology/IT services/software	44%
Financial services	12%
Manufacturing	10%
Retail/wholesale	10%
Telecoms/networks	6%
Healthcare	5%
Transportation/logistics	3%
Education	2%
Construction	2%
Entertainment/media	1%
Food service/hospitality	1%
Utilities	1%
Agriculture/mining	1%
Nonprofit	1%

References

- ¹ Howland, Ethan. "PJM Capacity Prices Hit Record High as Grid Operator Falls Short of Reliability Target." Utility Dive, December 18, 2025. <https://www.utilitydive.com/news/pjm-interconnection-capacity-auction-data-center/808264/>
- ² PJM Interconnection. "2027/2028 Base Residual Auction Report." December 17, 2025. <https://www.pjm.com/-/media/DotCom/markets-ops/rpm/rpm-auction-info/2027-2028/2027-2028-bra-report.pdf>
- ³ Monitoring Analytics. "Analysis of the 2025/2026 RPM Base Residual Auction Part G." June 3, 2025. https://www.monitoringanalytics.com/reports/reports/2025/IMM_Analysis_of_the_20252026_RPM_Base_Residual_Auction_Part_G_20250603_Revised.pdf
- ⁴ Maruf, Ramishah, and Kit Maher. "Trump And Northeastern Governors Push for Massive Electricity Auction to Make Tech Giants Defray Costs." CNN, January 16, 2026. <https://www.cnn.com/2026/01/16/business/pjm-electricity-auction-ai>
- ⁵ Colman, Zack, Gabby Miller, and Katherine Long. "Trump and AI Leaders Tout His 'Build Your Own Power Plant' Pledge." Politico, March 4, 2026. <https://www.politico.com/news/2026/03/04/trump-and-ai-leaders-tout-his-build-your-own-power-plant-pledge-00812891>
- ⁶ BloombergNEF. "Battery Storage Costs Hit Record Lows as Costs of Other Clean Power Technologies Increased: BloombergNEF." BloombergNEF, February 18, 2026. <https://about.bnef.com/insights/clean-energy/battery-storage-costs-hit-record-lows-as-costs-of-other-clean-power-technologies-increased-bloombergnef/>
- ⁷ Bonner, April. "US Energy Storage Installations in 2025 Have Already Surpassed Last Year's Total: Wood Mac." Energy Storage News, December 17, 2025. <https://www.energy-storage.news/us-energy-storage-installations-in-2025-have-already-surpassed-last-years-total-wood-mac/>
- ⁸ Peters, Adele. "Google's New Minnesota Data Center Comes with the World's Largest Battery – and Won't Raise Electric Bills." Fast Company, March 2, 2026. <https://www.fastcompany.com/91500104/google-minnesota-data-center-electric-bills>
- ⁹ SEIA. "Explained: The Clean Energy Provisions in the 'One Big Beautiful Bill.'" SEIA, July 21, 2025. <https://seia.org/research-resources/clean-energy-provisions-big-beautiful-bill/>
- ¹⁰ World Nuclear Association. "Small Modular Reactors." World Nuclear Association, March 24, 2026. <https://world-nuclear.org/information-library/nuclear-power-reactors/small-modular-reactors/small-modular-reactors>
- ¹¹ Constellation. "Constellation, Meta Sign 20-Year Deal for Clean, Reliable Nuclear Energy in Illinois." Constellation, June 3, 2025. <https://www.constellationenergy.com/news/2025/constellation-meta-sign-20-year-deal-for-clean-reliable-nuclear-energy-in-illinois.html>
- ¹² Talen Energy. "Talen Energy Expands Nuclear Energy Relationship with Amazon." Talen Energy, June 11, 2025. <https://ir.talenenergy.com/news-releases/news-release-details/talen-energy-expands-nuclear-energy-relationship-amazon>
- ¹³ Sinmaz, Emine. "Trump officials announce \$1bn loan to restart Three Mile Island nuclear plant." The Guardian, November 19, 2025. <https://www.theguardian.com/us-news/2025/nov/19/three-mile-island-nuclear-loan-microsoft-datacenter>
- ¹⁴ Plumer, Brad. "A Nuclear Reactor Backed by Bill Gates Gets Federal Approval to Start Building." The New York Times, March 4, 2026. <https://www.nytimes.com/2026/03/04/climate/nuclear-power-terrapower-construction-permit.html>
- ¹⁵ Vanian, Jonathan. "Meta Signs Nuclear Energy Deals to Power Prometheus AI Supercluster." CNBC, January 9, 2026. <https://www.cnbc.com/2026/01/09/meta-signs-nuclear-energy-deals-to-power-prometheus-ai-supercluster.html>
- ¹⁶ International Energy Agency. "Energy and AI: World Energy Outlook Special Report." April 2025. <https://iea.blob.core.windows.net/assets/de9dea13-b07d-42c5-a398-d1b3ae17d866/EnergyandAI.pdf>
- ¹⁷ Hardin, Kate, Patricia Tuite, Martin Stansbury, and Jaya Nagdeo. "Nuclear Energy's Role in Powering Data Center Growth." Deloitte Insights, December 24, 2025. <https://www.deloitte.com/us/en/insights/industry/power-and-utilities/nuclear-energy-powering-data-centers.html>
- ¹⁸ Thomas, Michael. "Bypassing the Grid: How Data Centers Are Building Their Own Power Plants." Cleanview, February 3, 2026. <https://newsletter.cleanview.co/p/bypassing-the-grid-how-data-centers>
- ¹⁹ Clark, Kevin. "Williams Pushes Deeper Into Power Generation as Data Center Demand Accelerates." Power Engineering, November 6, 2025. <https://www.power-eng.com/gas/williams-pushes-deeper-into-power-generation-as-data-center-demand-accelerates/>
- ²⁰ Chernicoff, David, Matt Vincent, and DCF Staff. "Chevron, GE Vernova, Engine No.1 Join Race to Co-Locate Natural Gas Plants for U.S. Data Centers." Data Center Frontier, February 10, 2025. <https://www.datacenterfrontier.com/energy/article/55264592/chevron-ge-vernova-engine-no1-join-race-to-co-locate-natural-gas-plants-for-us-data-centers>
- ²¹ Kong, Vince. "Chevron to Build Its First Data Center Power Plant in Texas Permian Area." Engineering News-Record, November 14, 2025. <https://www.enr.com/articles/61972-chevron-to-build-its-first-data-center-power-plant-in-texas-permian-area>
- ²² Peiserich, John. "Fractured & Fraught – but Still Potentially Profitable: The State of ESG in 2025." Corporate Compliance Insights, November 5, 2025. <https://www.corporatecomplianceinsights.com/the-state-of-esg-2025/>

References Cont.

- ²³ Bratton, Laura. "Big Tech Set to Spend \$650 Billion in 2026 as AI Investments Soar." Yahoo Finance, February 6, 2026. <https://finance.yahoo.com/news/big-tech-set-to-spend-650-billion-in-2026-as-ai-investments-soar-163907630.html>
- ²⁴ Microsoft. "Microsoft Cloud and AI Strength Drives Second Quarter Results." Microsoft, January 28, 2026. <https://www.microsoft.com/en-us/investor/earnings/fy-2026-q2/press-release-webcast>
- ²⁵ CBRE. "North America Data Center Trends H2 2025." CBRE, February 25, 2026. <https://www.cbre.com/insights/reports/north-america-data-center-trends-h2-2025>
- ²⁶ Wang, Olivia. "Data Center Outlook: Half of 2026 Pipeline May Not Materialize." Sightline Climate, February 24, 2026. <https://www.sightlineclimate.com/research/data-center-outlook>
- ²⁷ Morris, Iain. "'Perfect Storm' in Fiber Supply Threatens US Broadband Targets." Light Reading, December 15, 2025. <https://www.lightreading.com/fttx/-perfect-storm-in-fiber-supply-threatens-us-broadband-targets>
- ²⁸ Hardesty, Linda. "Fiber Vendors Strategize for Huge Demand from AI in 2026." Fierce Network, December 23, 2025. <https://www.fierce-network.com/broadband/major-fiber-vendors-strategize-huge-demand-ai-2026>
- ²⁹ Burke, Steven. "HPE Warns Partners of 'Price Adjustments' on Server, GreenLake Orders in Wake of Memory Price Increases." CRN.com, February 13, 2026. <https://www.crn.com/news/data-center/2026/hpe-warns-partners-of-price-adjustments-on-server-greenlake-orders-in-wake-of-memory-price-increases>
- ³⁰ Narcisi, Gina. "Cisco Revises Partner Contract Terms in Response to Soaring Memory Prices." CRN.com, February 13, 2026. <https://www.crn.com/news/networking/2026/cisco-revises-partner-contract-terms-in-response-to-soaring-memory-prices>
- ³¹ Associated Builders and Contractors. "ABC: Construction Industry Must Attract 349,000 Workers in 2026 Despite Macroeconomic Headwinds." Associated Builders and Contractors, January 15, 2026. <https://www.abc.org/News-Media/News-Releases/abc-construction-industry-must-attract-349000-workers-in-2026-despite-macroeconomic-headwinds>
- ³² International Brotherhood of Electrical Workers. "The Data Center Surge: A New Generation of IBEW Jobs." International Brotherhood of Electrical Workers, May 1, 2025. https://ibew.org/electrical_worker/the-data-center-surge-a-new-generation-of-ibew-jobs/
- ³³ Hardesty, "Fiber Vendors Strategize for Huge Demand from AI in 2026."
- ³⁴ Fiber Broadband Association. "The Underappreciated Need to Enable AI and Data Center Growth: Increased and More Strategic Fiber Interconnections." July 2025. https://fiberbroadband.org/wp-content/uploads/2025/07/FBA-103_AI_Datacenter_WhitePaper_lv6.pdf
- ³⁵ Chauhan, Harsh. "This AI Infrastructure Stock Grew 176% Last Year. Is It Too Late to Buy in 2026?" The Motley Fool, February 27, 2026. <https://www.fool.com/investing/2026/02/27/this-ai-infrastructure-stock-grew-176-last-year-is/>
- ³⁶ Snider, Shane. "Meta Inks \$6B Fiber Optic Deal With Corning for US Data Centers." Data Center Knowledge, January 27, 2026. <https://www.datacenterknowledge.com/business/meta-inks-6b-fiber-optic-deal-with-corning-for-us-data-centers>
- ³⁷ Boujelbene, Sameh. "Data Center Networking in 2025-2026: Milestones and Opportunities Amid Supply Risk." Dell'Oro Group, January 7, 2026. <https://www.delloro.com/2026-predictions-data-center-switch-frontend-ai-backed-networks/>
- ³⁸ Shilov, Anton. "Ultra Ethernet: The Data Center Interconnection of Tomorrow Detailed." Tom's Hardware, February 9, 2026. <https://www.tomshardware.com/networking/ultra-ethernet-the-data-center-interconnection-of-tomorrow-detailed>
- ³⁹ Bryan, Greg. "Building Network Resiliency with Transparency and Smart Strategy." TeleGeography, January 13, 2026. <https://resources.telegeography.com/building-network-resiliency-with-transparency-and-smart-strategy>
- ⁴⁰ Wodecki, Ben. "Lumen calls for 'network reset' to meet AI and Cloud 2.0 demands." SDxCentral, October 24, 2025. <https://www.sdxcentral.com/news/lumen-calls-for-network-reset-to-meet-ai-and-cloud-20-demands/>
- ⁴¹ Snider, Shane. "New 'AI Pods' Bring Low-Latency Compute to Smaller US Cities." Data Center Knowledge, January 15, 2026. <https://www.datacenterknowledge.com/modular-data-centers/firms-launch-ai-pods-for-smaller-us-cities-inference-needs>
- ⁴² Mahajan, Pallavi. "No Network, No AI Evolution." Politico, February 10, 2026. <https://www.politico.com/sponsor-content/2026/02/10/no-network-no-ai-evolution>
- ⁴³ Ookla. "United States Speedtest Connectivity Report." February 2026. <https://www.ookla.com/research/reports/united-states-speedtest-connectivity-report-h2-2025>
- ⁴⁴ Semiconductor Industry Association. "State of the U.S. Semiconductor Industry 2025." Semiconductor Industry Association, July 10, 2025. <https://www.semiconductors.org/wp-content/uploads/2025/07/SIA-State-of-the-Industry-Report-2025.pdf>

References Cont.

- ⁴⁵ Shepardson, David. "US Finalizes \$6.6 Billion Chips Award for TSMC Ahead of Trump Return." Reuters, November 15, 2024. <https://www.reuters.com/technology/us-finalizes-66-billion-chips-award-tsmc-ahead-trump-return-2024-11-15/>
- ⁴⁶ Butts, Dylan, and Emily Tan. "TSMC Is Set to Expand Its \$165 Billion U.S. Investment — Here's What We Know." CNBC, January 16, 2026. <https://www.cnbc.com/2026/01/16/tsmcs-arizona-chip-expansion-isnt-done-after-us-investment-cfo.html>
- ⁴⁷ Sourceability. "Memory Shortage and Tariff Risks Collide in AI Chip Supply." Sourceability, February 6, 2026. <https://sourceability.com/post/memory-shortage-and-tariff-risks-collide-in-ai-chip-supply>
- ⁴⁸ TrendForce. "Memory Price Outlook for 1Q26 Sharply Upgraded; QoQ Increases of All Product Categories to Hit Record Highs, Says TrendForce." TrendForce, February 2, 2026. <https://www.trendforce.com/presscenter/news/20260202-12911.html>
- ⁴⁹ James, Luke. "AI Memory Crunch Forces DRAM Market Into 'Hourly Pricing' Model, Report Claims — Small and Medium-Sized Businesses Fighting for Survival." Tom's Hardware, March 3, 2026. <https://www.tomshardware.com/pc-components/ram/memory-prices-now-shifting-hourly-as-smaller-firms-fight-over-scrap>
- ⁵⁰ Kaur, Dashveenjit. "2025's AI Chip Wars: What Enterprise Leaders Learned About Supply Chain Reality." AI News, January 6, 2026. <https://www.artificialintelligence-news.com/news/ai-chip-shortage-enterprise-ctos-2025/>
- ⁵¹ Burnham, Kristin. "The 'Productivity Paradox' of AI Adoption in Manufacturing Firms." MIT Sloan, July 9, 2025. <https://mitsloan.mit.edu/ideas-made-to-matter/productivity-paradox-ai-adoption-manufacturing-firms>
- ⁵² The Federal Reserve Bank of Richmond. "CFO Outlook for 2026: Tariffs, Hiring, Prices, and AI Impact." December 17, 2025. https://www.richmondfed.org/research/national_economy/cfo_survey/data_and_results/2025/20251217_data_and_results
- ⁵³ KPMG. "AI At Scale: How 2025 Set the Stage for Agent-Driven Enterprise Reinvention in 2026." KPMG, January 15, 2026. <https://kpmg.com/us/en/media/news/q4-ai-pulse.html>
- ⁵⁴ Gartner. "Gartner Says Worldwide AI Spending Will Total \$2.5 Trillion in 2026." January 15, 2026. <https://www.gartner.com/en/newsroom/press-releases/2026-1-15-gartner-says-worldwide-ai-spending-will-total-2-point-5-trillion-dollars-in-2026>
- ⁵⁵ Goss, Grant. "AI cost overruns are adding up — with major implications for CIOs." CIO.com, October 2, 2025. <https://www.cio.com/article/4064319/ai-cost-overruns-are-adding-up-with-major-implications-for-cios.html>
- ⁵⁶ Shah, Agam. "GPU Pricing, a Bellwether for AI Costs, Could Help IT Leaders at Budget Time." Computerworld, December 15, 2025. <https://www.computerworld.com/article/4104332/gpu-pricing-a-bellwether-for-ai-costs-could-help-it-leaders-at-budget-time.html>
- ⁵⁷ Gartner. "Gartner Research Reveals CFOs' Budget Plans Prioritize Growth Functions, Technology and AI in 2026." February 10, 2026. <https://www.gartner.com/en/newsroom/press-releases/2026-02-10-gartner-research-reveals-cfos-budget-plans-prioritize-growth-functions-tech-and-ai-in-2026>
- ⁵⁸ KPMG. "AI Quarterly Pulse Survey Q4 2025." January 2026. <https://kpmg.com/kpmg-us/content/dam/kpmg/corporate-communications/pdf/2026/Q4%20Pulse%20Survey%20Deck%20FINAL.pdf>
- ⁵⁹ Blum, Jordan. "Your Utility Bills Keep Going up. Here's Everyone You Can Blame — AI Data Centers Included." Fortune, March 1, 2026. <https://fortune.com/2026/03/01/utility-bills-keep-rising-everyone-blame-ai-data-centers-included/>
- ⁶⁰ Powerlines. "Utility Bills Are Rising: 2025 Review." January 26, 2026. https://powerlines.org/wp-content/uploads/2026/01/0126_PowerLines_Rising-Utility-Bills-Q4-Update-FINAL.pdf
- ⁶¹ Vincent, Matt, and DCF Staff. "How Tariffs Could Impact Data Centers, AI, and Energy Amid Supply Chain Shifts." Data Center Frontier, April 3, 2025. <https://www.datacenterfrontier.com/hyperscale/article/55279670/how-tariffs-could-impact-data-centers-ai-and-energy-amid-supply-chain-shifts>
- ⁶² Blum, "Your Utility Bills Keep Going up. Here's Everyone You Can Blame — AI Data Centers Included."
- ⁶³ Korst, Jeremy, Stefano Puntoni, and Prasanna Tambe. "Accountable Acceleration: Gen AI Fast-Tracks Into the Enterprise." The Wharton School of the University of Pennsylvania, October 2025. https://ai.wharton.upenn.edu/wp-content/uploads/2025/10/2025-Wharton-GBK-AI-Adoption-Report_Full-Report.pdf
- ⁶⁴ RPG. "RGP CFO Survey Shows Growing Divide Between AI Ambition and AI Readiness." RGP, December 12, 2025. <https://rgp.com/press/rgp-cfo-survey-shows-growing-divide-between-ai-ambition-and-ai-readiness/>
- ⁶⁵ Pratt, Mary K. "2026: The year AI ROI gets real." CIO.com, January 13, 2026. <https://www.cio.com/article/4114010/2026-the-year-ai-roi-gets-real.html>
- ⁶⁶ Gartner. "Gartner Says Worldwide AI Spending Will Total \$2.5 Trillion in 2026."