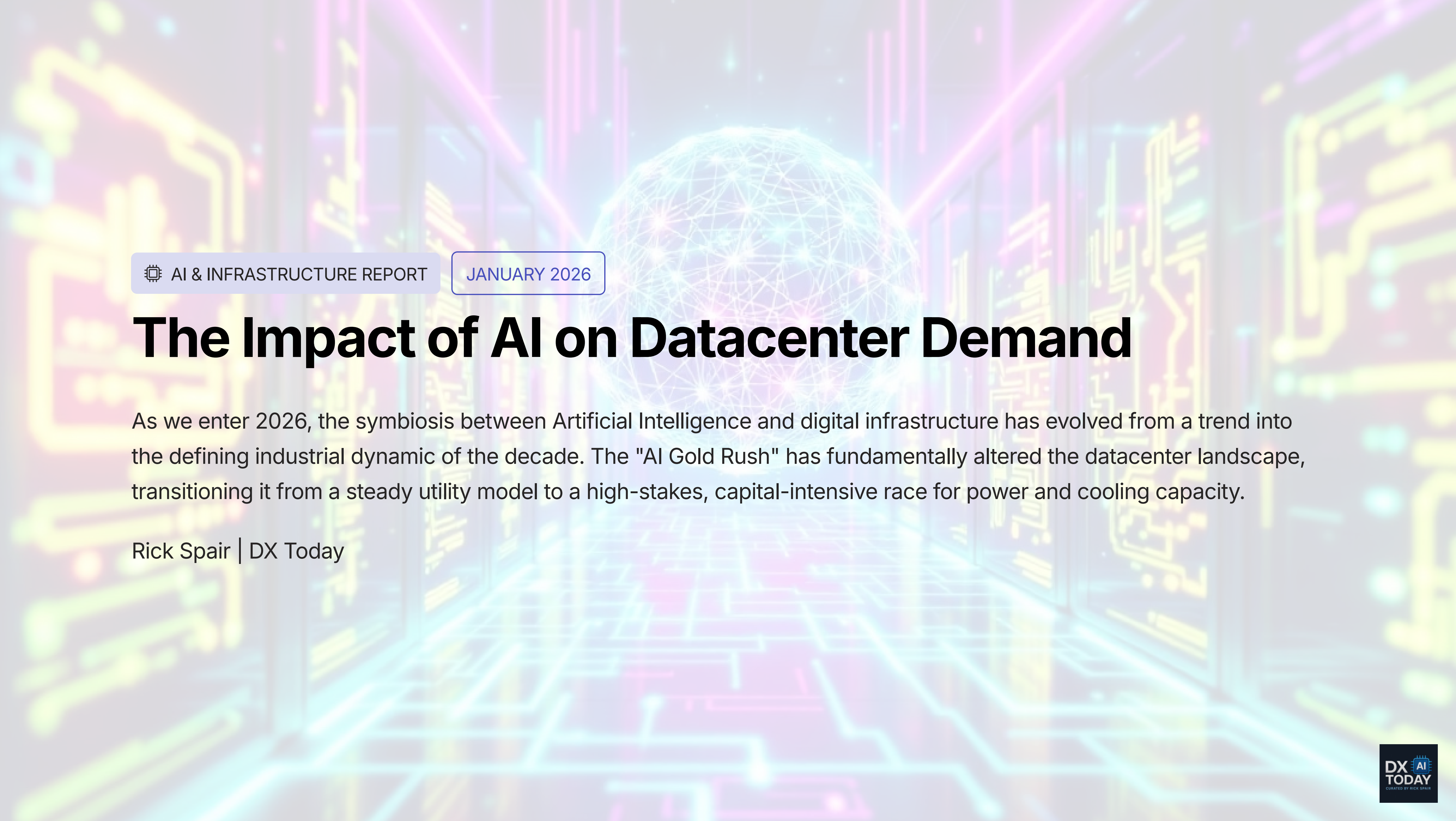


AI & INFRASTRUCTURE REPORT

JANUARY 2026

# The Impact of AI on Datacenter Demand

As we enter 2026, the symbiosis between Artificial Intelligence and digital infrastructure has evolved from a trend into the defining industrial dynamic of the decade. The "AI Gold Rush" has fundamentally altered the datacenter landscape, transitioning it from a steady utility model to a high-stakes, capital-intensive race for power and cooling capacity.

Rick Spair | DX Today

DX AI TODAY
CURATED BY RICK SPAIR

# Executive Overview: The AI Infrastructure Revolution

**$315B**

### Hyperscaler CapEx

Combined capital expenditure by Microsoft, Google, Amazon, and Meta in 2025

**165%**

### Power Demand Growth

Projected increase in global datacenter power consumption by 2030

**100kW**

### Rack Density

Average power per rack for cutting-edge AI clusters, up from 10 kW

**3-4yr**

### Transformer Lead Time

Current wait time for critical electrical components

These unprecedented figures represent more than growth—they signal a complete transformation of the datacenter industry from a utility model to a high-stakes infrastructure race.

# Key Findings: The New Landscape

### Power Bottleneck

Power availability has replaced floor space as the primary constraint, fundamentally changing site selection criteria

### Nuclear Renaissance

2025 marked historic partnerships between tech giants and nuclear power providers

### Regulatory Headwinds

FERC's rejection of behind-the-meter expansion forced strategic pivots toward grid independence

### Supply Chain Strain

Critical component prices have doubled with lead times extending to multiple years

# Defining the New Era

The datacenter industry is undergoing its most significant architectural shift since the advent of cloud computing. For fifteen years, the industry optimized for "hyperscale"—massive, standardized facilities designed to host web services and storage.

Today, the mandate is "AI-scale"—facilities designed to host high-performance computing clusters that run hot, heavy, and power-hungry. These aren't incremental improvements; they represent a fundamental reimagining of digital infrastructure.

In early 2026, we are witnessing the physical manifestation of the Generative AI boom. The training of trillion-parameter models requires "AI Factories"—single, unified supercomputers acting as one machine.

## From Hyperscale to AI-Scale

The shift impacts everything from real estate valuations in rural Wisconsin to water rights in Arizona, creating ripple effects across the entire economy.

# Historical Evolution of Datacenters

1

## Phase 1: Enterprise Era (1990s-2010)

On-premise server rooms and fragmented colocation facilities. Power density was negligible at 2-4 kW per rack. Infrastructure was distributed and inefficient.

2

## Phase 2: Cloud Era (2010-2022)

AWS, Azure, and Google Cloud consolidated compute into massive hyperscale regions. Efficiency improved significantly with PUE optimization, and densities stabilized around 8-10 kW per rack.

3

## Phase 3: AI Era (2023-Present)

ChatGPT's release kicked off an infrastructure arms race. Nvidia's H100 and Blackwell B200 chips shattered previous thermal and power ceilings, necessitating complete datacenter redesign.

# The Hyperscaler Investment Surge

The "Big Four" hyperscalers—Microsoft, Google, Amazon, and Meta—committed over $300 billion in 2025 toward infrastructure, with the vast majority allocated to servers and the buildings to house them. This represents not merely growth but a multiplication of capacity unlike anything the industry has witnessed.

### Microsoft

Leading the charge with massive Azure AI investments and the Three Mile Island nuclear partnership

### Google

Expanding TPU infrastructure and pioneering nuclear partnerships with Kairos Power

### Amazon

AWS infrastructure expansion despite regulatory setbacks on behind-the-meter projects

### Meta

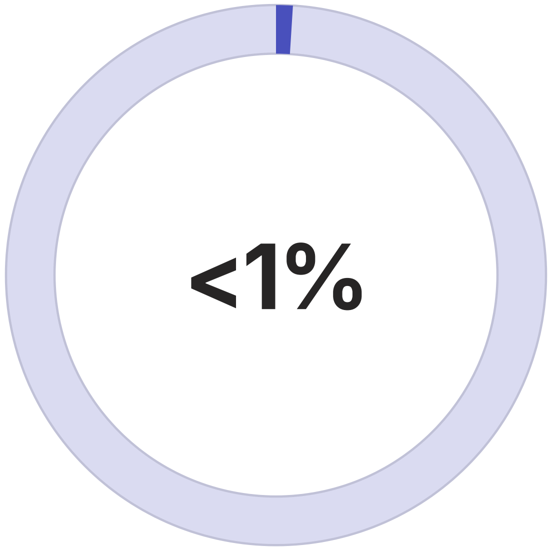
Building dedicated AI training facilities to support next-generation LLaMA models

# Market Dynamics: Vacancy Crisis

## Primary Markets Under Pressure

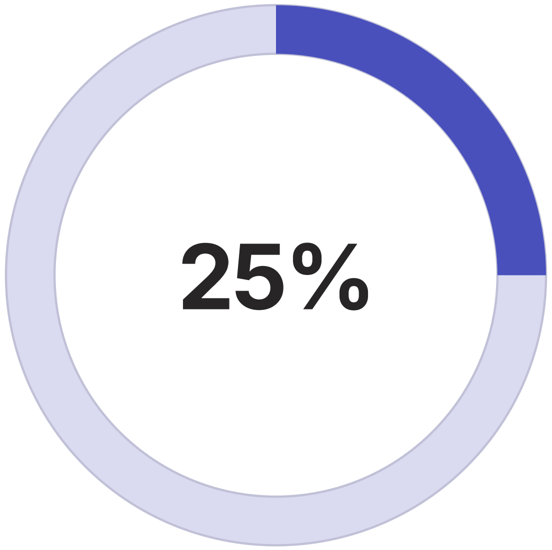
In primary markets like Northern Virginia—the world's largest datacenter hub—vacancy rates have dropped to near-zero levels below 1%. This historic scarcity has driven rental prices up by 20-30% year-over-year.

The supply-demand imbalance has created a seller's market where landlords can command premium rates and strict terms. Pre-leasing of facilities still under construction has become standard practice.



**Vacancy Rate**

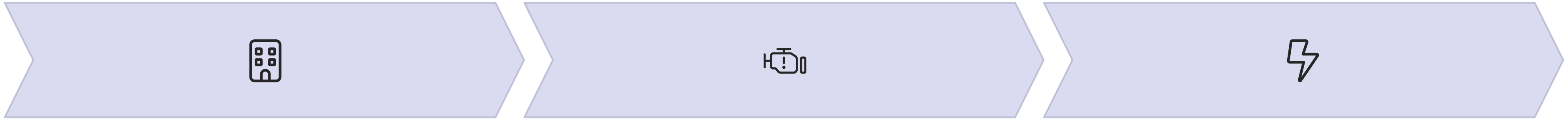
Northern Virginia market



**Price Increase**

Year-over-year rental growth

# The Power Paradigm Shift



### Traditional Model

Floor space was the primary constraint. Power was abundant and affordable.

### Transition Period

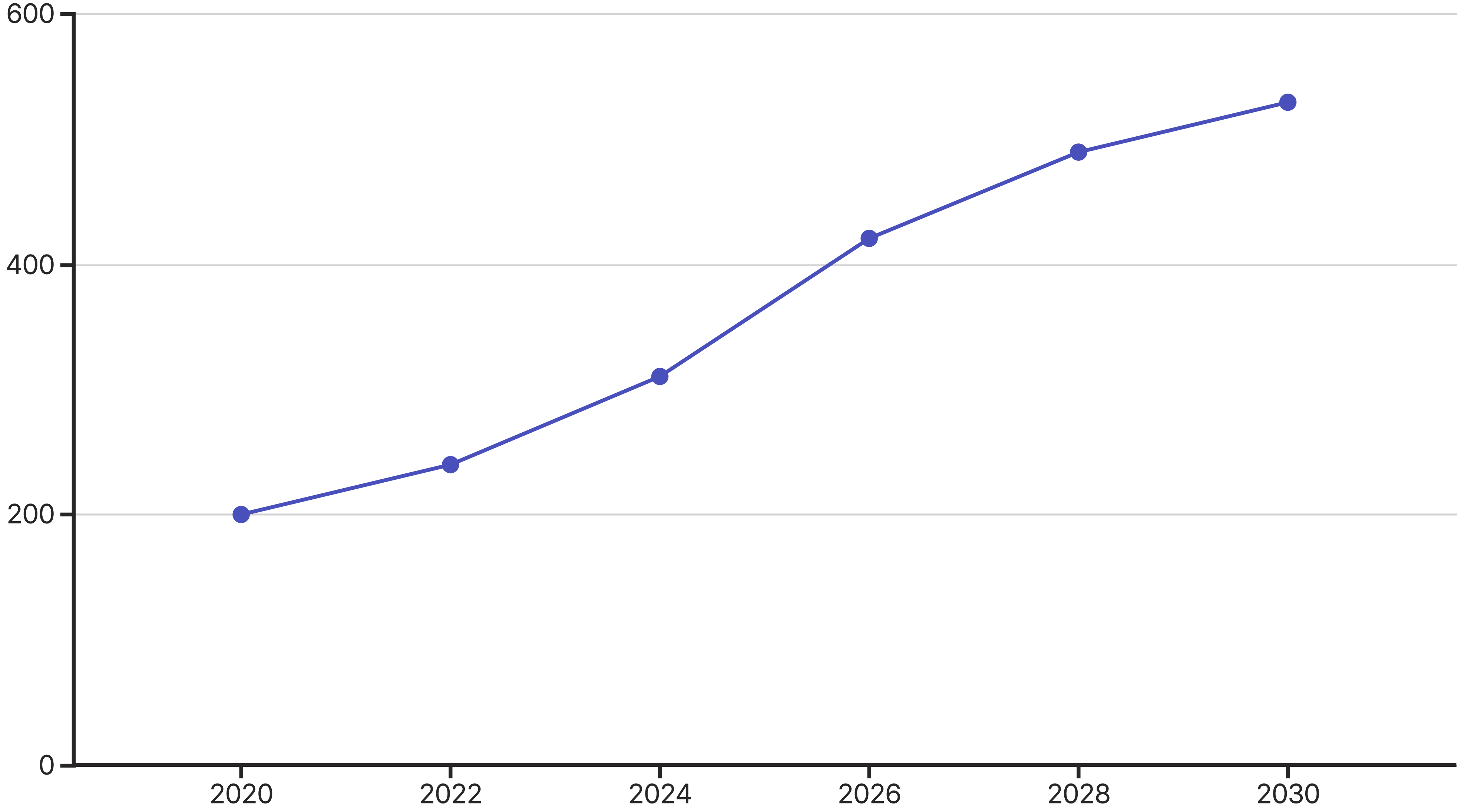
Power and space equally important. Grid capacity concerns emerging.

### AI Era Model

Power availability is the defining constraint. Sites chosen for energy access first.

This fundamental shift has upended traditional site selection criteria. Projects that once prioritized proximity to fiber and markets now prioritize access to megawatts. The question is no longer "How much space do we need?" but rather "How much power can we secure?"

# Energy Demand Projections



Global datacenter power demand is forecast to increase by 165% between 2024 and 2030, a trajectory that far outpaces grid infrastructure development. This exponential growth curve has forced the industry to seek alternative energy solutions beyond traditional grid connections.

# Rack Density Revolution

01

### Traditional Enterprise (2-4 kW)

Standard office servers with basic air cooling requirements

02

### Cloud Era (8-10 kW)

Hyperscale optimization with improved efficiency and modest density gains

03

### Early AI (20-30 kW)

First generation GPU clusters requiring enhanced cooling infrastructure

04

### Current AI-Scale (100+ kW)

Cutting-edge AI clusters with liquid cooling and extreme power requirements

The average rack density for cutting-edge AI clusters has surged from 10 kW to over 100 kW, representing a tenfold increase that demands entirely new approaches to power delivery and thermal management.

# The Nuclear Renaissance

## Historic Partnerships Reshape Energy Strategy

2025 marked the "Nuclear Renaissance" for datacenters, with major technology companies embracing nuclear power as the only viable solution to meet their energy demands at scale. These partnerships signal that traditional grid connections alone cannot support AI's exponential energy appetite.

### Microsoft & Three Mile Island

Historic deal to restart Unit 1, providing dedicated power for Azure AI infrastructure. The 20-year power purchase agreement will deliver 835 MW of carbon-free baseload power starting in 2028.

### Google & Kairos Power

Partnership to develop next-generation small modular reactors (SMRs) specifically designed for datacenter applications. First units expected online by 2030.

### Amazon Nuclear Strategy

Investments in multiple SMR developers and nuclear energy startups, positioning for long-term energy independence across AWS regions.

# Why Nuclear Makes Strategic Sense

### Baseload Reliability

Nuclear provides 24/7 power generation regardless of weather conditions, essential for AI training workloads that cannot be interrupted.

### Carbon-Free Operation

Meets corporate sustainability commitments while delivering massive power capacity, addressing both environmental and operational goals.

### Grid Independence

Dedicated nuclear facilities bypass grid congestion and regulatory constraints that have stalled traditional datacenter expansion.

### Long-Term Cost Stability

Fixed fuel costs and 60+ year operational lifespans provide predictable economics for multi-decade infrastructure investments.

# Regulatory Landscape: The FERC Rejection

## The Amazon-Talen Energy Decision

In late 2024, the Federal Energy Regulatory Commission (FERC) rejected Amazon's proposal to expand its "behind-the-meter" datacenter at the Talen Energy nuclear facility in Pennsylvania. This landmark decision sent shockwaves through the industry.

The rejection centered on concerns about grid stability and fairness. FERC argued that large behind-the-meter loads could destabilize regional grids and unfairly shift costs to other ratepayers.

**Strategic Impact:** The FERC decision has forced datacenter operators to pursue grid-independent solutions, accelerating the shift toward dedicated nuclear partnerships and on-site generation.

# Behind-the-Meter vs. Grid-Connected Models

### Behind-the-Meter Model

Direct connection to power source, bypassing grid transmission. Benefits: Lower costs, reduced transmission losses, greater control. Challenges: Regulatory scrutiny, limited scalability, fairness concerns.

### Grid-Connected Model

Traditional connection through utility infrastructure and regional grid. Benefits: Regulatory clarity, easier permitting, shared infrastructure. Challenges: Grid congestion, transmission constraints, cost volatility.

### Hybrid Emerging Model

Combination of on-site generation with grid backup. Benefits: Best of both worlds, resilience, regulatory compliance. Challenges: Higher capital costs, complex operations, dual infrastructure.

# Supply Chain Bottlenecks

The rapid acceleration of datacenter construction has exposed critical vulnerabilities in the infrastructure supply chain. Components that were once readily available now face unprecedented lead times and cost pressures.

**3-4yr**

### Transformer Lead Time

Wait time for large power transformers has tripled

**2x**

### Price Increase

Component costs have nearly doubled since 2020

**6mo**

### Switchgear Delays

Additional wait time for medium-voltage equipment

# Critical Component Constraints

1

## Power Transformers

Large transformers are manufactured by only a handful of global suppliers, creating severe bottlenecks. The specialized nature of these units—often custom-designed for specific applications—means lead times of 3-4 years.

2

## Backup Generators

Diesel and natural gas generators in the 2-5 MW range face 12-18 month lead times. AI datacenters require significantly more backup capacity than traditional facilities.

3

## Cooling Infrastructure

Liquid cooling systems, chillers, and cooling towers designed for high-density AI workloads are in high demand. Specialized equipment manufacturers are ramping production but cannot keep pace.

4

## Electrical Distribution

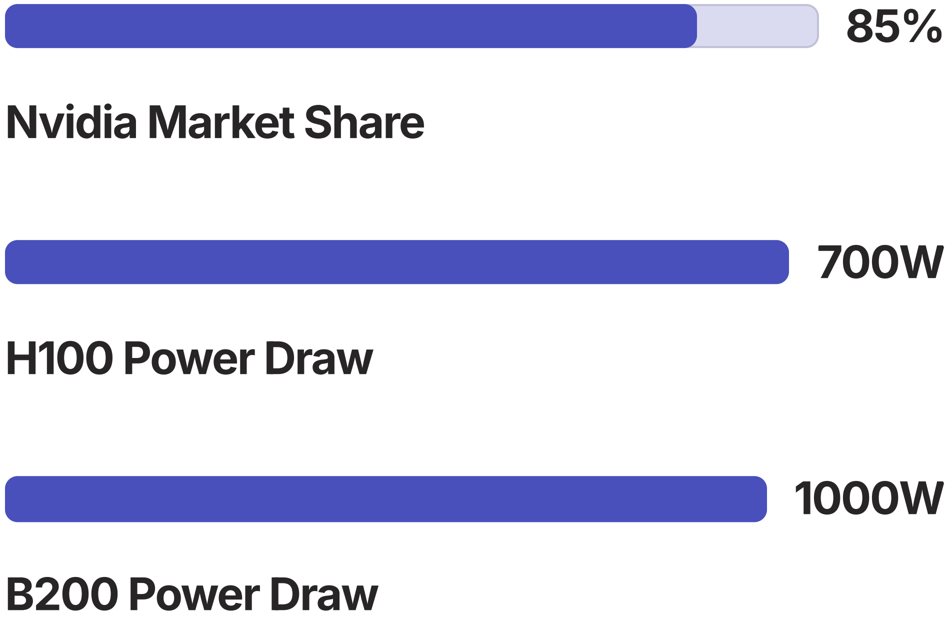
Medium and high-voltage switchgear, bus ducts, and distribution panels require specialized manufacturing. Suppliers are prioritizing orders based on long-standing relationships.

# AI Chip Architecture Drives Demand

## The Nvidia Revolution

The introduction of Nvidia's H100 and subsequent Blackwell B200 chips fundamentally changed the datacenter equation. These GPUs deliver unprecedented computational performance but at a significant power cost.

A single H100 GPU consumes up to 700 watts under full load. A typical AI training cluster requires thousands of these chips working in parallel, creating thermal and electrical challenges that legacy datacenter designs simply cannot handle.



# The AI Factory Concept

## Single-Purpose Supercomputers

The training of trillion-parameter AI models requires what industry leaders call "AI Factories"—single, unified supercomputers acting as one machine. Unlike traditional datacenters that host diverse workloads across distributed infrastructure, AI Factories are purpose-built for a single mission.

### Scale

10,000-100,000 GPUs in a single interconnected cluster

### Power

100-500 MW dedicated to a single training workload

### Networking

Ultra-low latency InfiniBand or custom interconnects

### Duration

Months-long continuous operation for model training

# Cooling Challenges in the AI Era

### Air Cooling Limitations

Traditional air cooling becomes ineffective above 20-30 kW per rack. The volume of air required creates noise, space, and efficiency problems.

### Direct Liquid Cooling

Cold plates attached directly to chips remove heat more efficiently. Allows densities up to 60-80 kW per rack but requires new infrastructure and maintenance approaches.

### Immersion Cooling

Servers submerged in dielectric fluid. Enables 100+ kW densities and near-silent operation. Still emerging technology with higher upfront costs.

### Hybrid Approaches

Combining multiple cooling technologies to optimize for different workload types and cost profiles within the same facility.

# Water Consumption Crisis

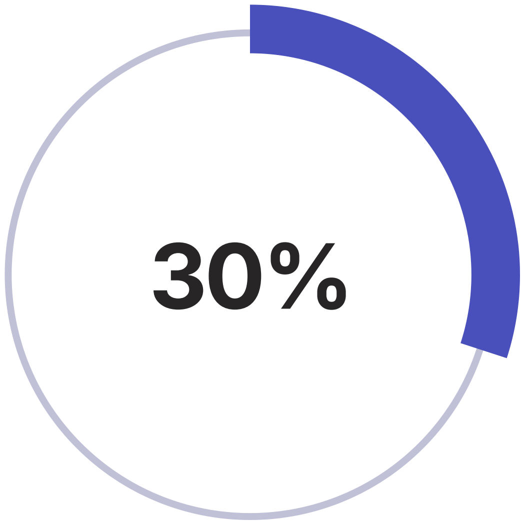
Advanced cooling systems—particularly those supporting AI workloads—consume enormous quantities of water. A single large datacenter can use 3-5 million gallons of water per day, equivalent to a city of 30,000-50,000 people.

This consumption has sparked intense debates in water-stressed regions. Arizona, Nevada, and parts of Texas face particular challenges as datacenter development collides with agricultural and municipal water needs.

Operators are exploring water recycling, air-cooled alternatives, and partnerships with water utilities to address these concerns, but solutions remain incomplete.

**Daily Water Use**

Gallons per large facility



**Water Stress Regions**

US markets facing constraints

# Geographic Shift in Datacenter Development

### Traditional Hubs

Northern Virginia, Silicon Valley, and other established markets face power and land constraints despite strong connectivity.

### Emerging Markets

The Midwest and rural areas with abundant power and lower land costs are attracting mega-projects despite connectivity challenges.

### Renewable Corridors

Texas, Arizona, and Nevada combine renewable energy potential with available land, creating new AI datacenter hotspots.

Site selection now prioritizes power availability over traditional factors like network connectivity and proximity to population centers. This geographic rebalancing is reshaping regional economies across North America.

# Economic Impact on Local Communities

The arrival of hyperscale datacenters transforms local economies, bringing construction jobs, tax revenue, and infrastructure investment. However, these benefits come with trade-offs that communities must carefully evaluate.

### Job Creation

Construction phase provides temporary employment surge. Permanent operational jobs are more limited but highly technical and well-compensated.

### Tax Revenue

Property taxes from multi-billion dollar facilities provide significant funding for schools and infrastructure, though often negotiated at reduced rates.

### Infrastructure Strain

Increased demand on electrical grid, water systems, and roads can overwhelm local capacity, requiring coordinated upgrades.

### Community Impact

Housing demand from workers, increased traffic, and changing local character create social dynamics requiring active management.

# Sustainability Paradox

## AI's Environmental Contradictions

### The Challenge

AI promises to optimize energy systems, accelerate climate research, and improve resource efficiency across industries. Yet training and operating these models requires massive energy consumption and physical infrastructure.

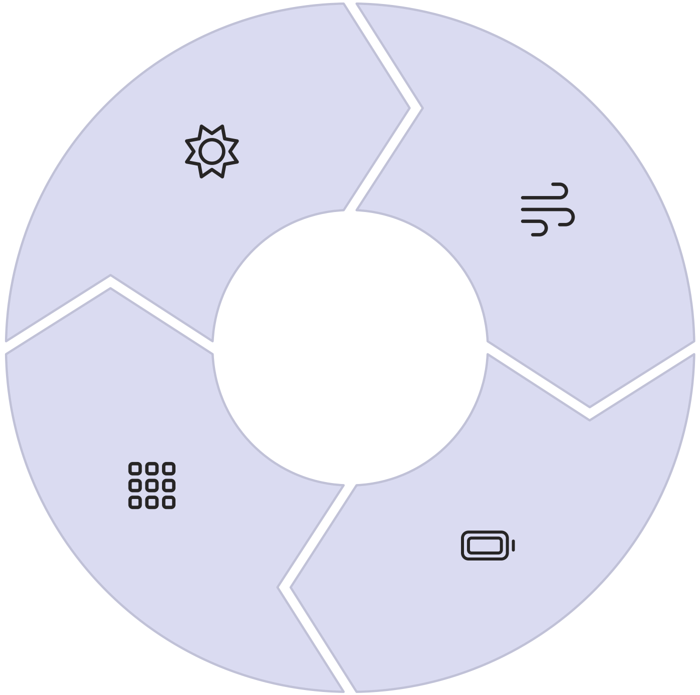
A single training run for a large language model can consume as much electricity as 100 US homes use in an entire year. Multiply this across thousands of models being developed simultaneously, and the scale becomes staggering.

### Industry Response

Hyperscalers have committed to 100% renewable energy and carbon neutrality, driving unprecedented demand for clean power. These corporate commitments are reshaping electricity markets and accelerating renewable deployment.

However, the intermittent nature of solar and wind creates mismatches with 24/7 AI workload demands, explaining the pivot toward nuclear baseload power.

# Renewable Energy Integration



### Solar Power

Daytime generation peaks align poorly with constant AI training demands

### Wind Power

Variable generation requires grid storage or backup systems

### Grid Management

Smart systems balance renewable supply with datacenter demand

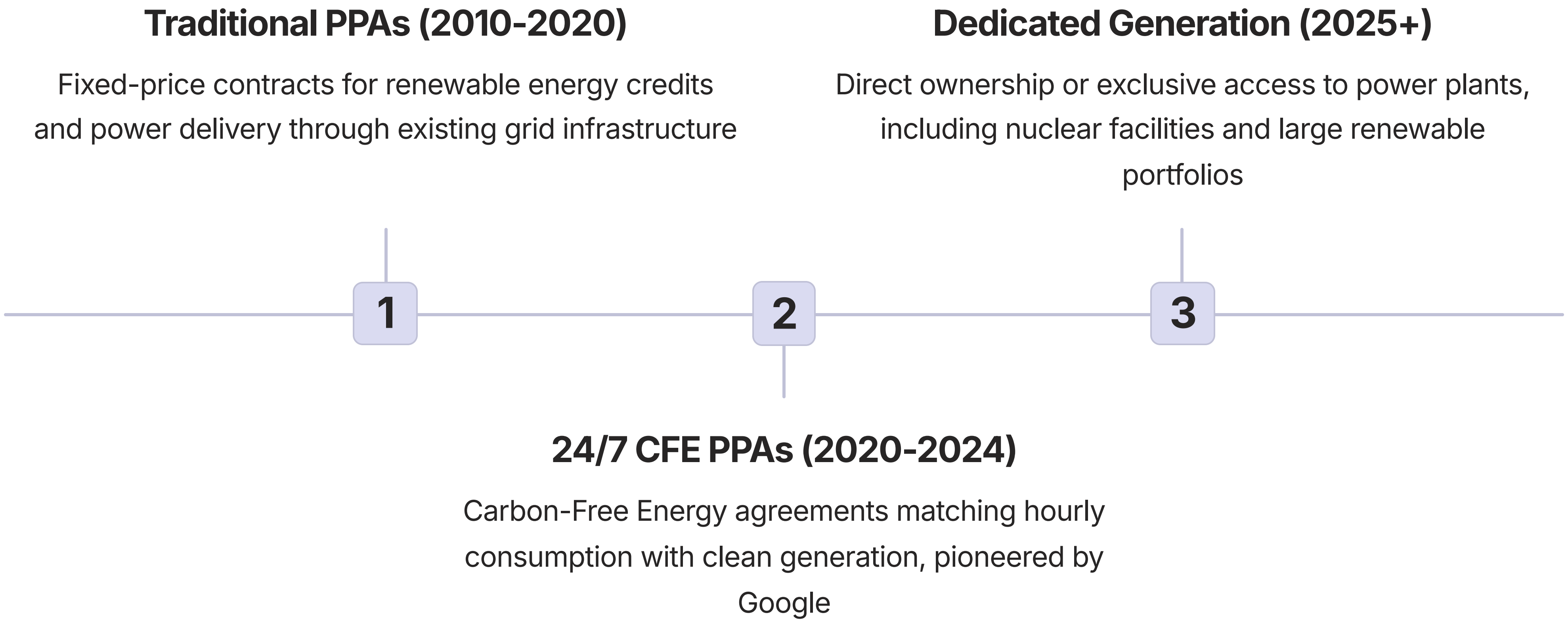
### Energy Storage

Battery systems smooth intermittency but add cost and complexity

The integration of renewable energy with AI datacenter operations requires sophisticated energy management systems that can respond to generation variability while maintaining uninterruptible service for critical AI workloads.

# Power Purchase Agreements Evolution

Power Purchase Agreements (PPAs) have evolved from simple capacity contracts to complex financial instruments that define the datacenter industry's relationship with energy markets.



### Traditional PPAs (2010-2020)

Fixed-price contracts for renewable energy credits and power delivery through existing grid infrastructure

### 24/7 CFE PPAs (2020-2024)

Carbon-Free Energy agreements matching hourly consumption with clean generation, pioneered by Google

### Dedicated Generation (2025+)

Direct ownership or exclusive access to power plants, including nuclear facilities and large renewable portfolios

# Real Estate Market Transformation

### Land Values

Rural land with power access has appreciated 200-300% in key markets

### Lease Structures

20-year terms with power guarantees now standard, replacing 5-10 year agreements

### Build-to-Suit

Hyperscalers increasingly develop their own facilities rather than leasing colocation space

### Capital Requirements

$1-2 billion minimum for modern AI-scale facilities, up from $200-400 million

The datacenter real estate market has transformed from a niche sector into a multi-hundred-billion-dollar asset class with institutional investors, REITs, and sovereign wealth funds competing for premium properties.

# Competitive Landscape: Hyperscalers vs Colocation

## Hyperscaler Strategy

Microsoft, Google, Amazon, and Meta are building their own facilities to maintain control over critical infrastructure. This vertical integration provides:

- Custom designs optimized for specific AI workloads
- Direct power procurement relationships
- Reduced dependency on third-party providers
- Long-term cost advantages at scale

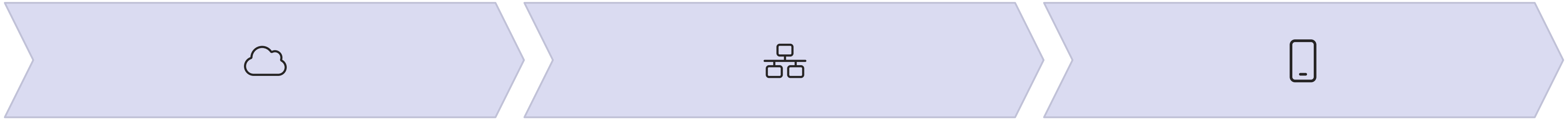
## Colocation Provider Response

Traditional colocation operators like Equinix, Digital Realty, and CyrusOne are adapting by:

- Developing AI-ready facilities with high power density
- Securing their own power generation assets
- Targeting mid-market customers and edge deployments
- Offering turnkey solutions for smaller AI companies

# Edge Computing and AI Inference

While training large AI models requires centralized AI Factories, inference—actually using these models to generate outputs—can be distributed to edge locations closer to end users. This creates a two-tier datacenter architecture.



### Core AI Training

Massive centralized facilities with 100+ MW power for model development

### Regional Inference

Mid-sized facilities (10-50 MW) in major metro areas for real-time AI applications

### Edge AI

Small facilities and on-device processing for ultra-low latency use cases

# Workforce Development Challenges

The AI datacenter boom has created acute workforce shortages across multiple skill categories. Traditional IT operations skills are insufficient for managing the complexity of modern AI infrastructure.

### Electrical Engineers

Specialists in high-voltage systems, power distribution, and generator management are in extremely high demand. Salaries have increased 30-40% in competitive markets.

### Mechanical Engineers

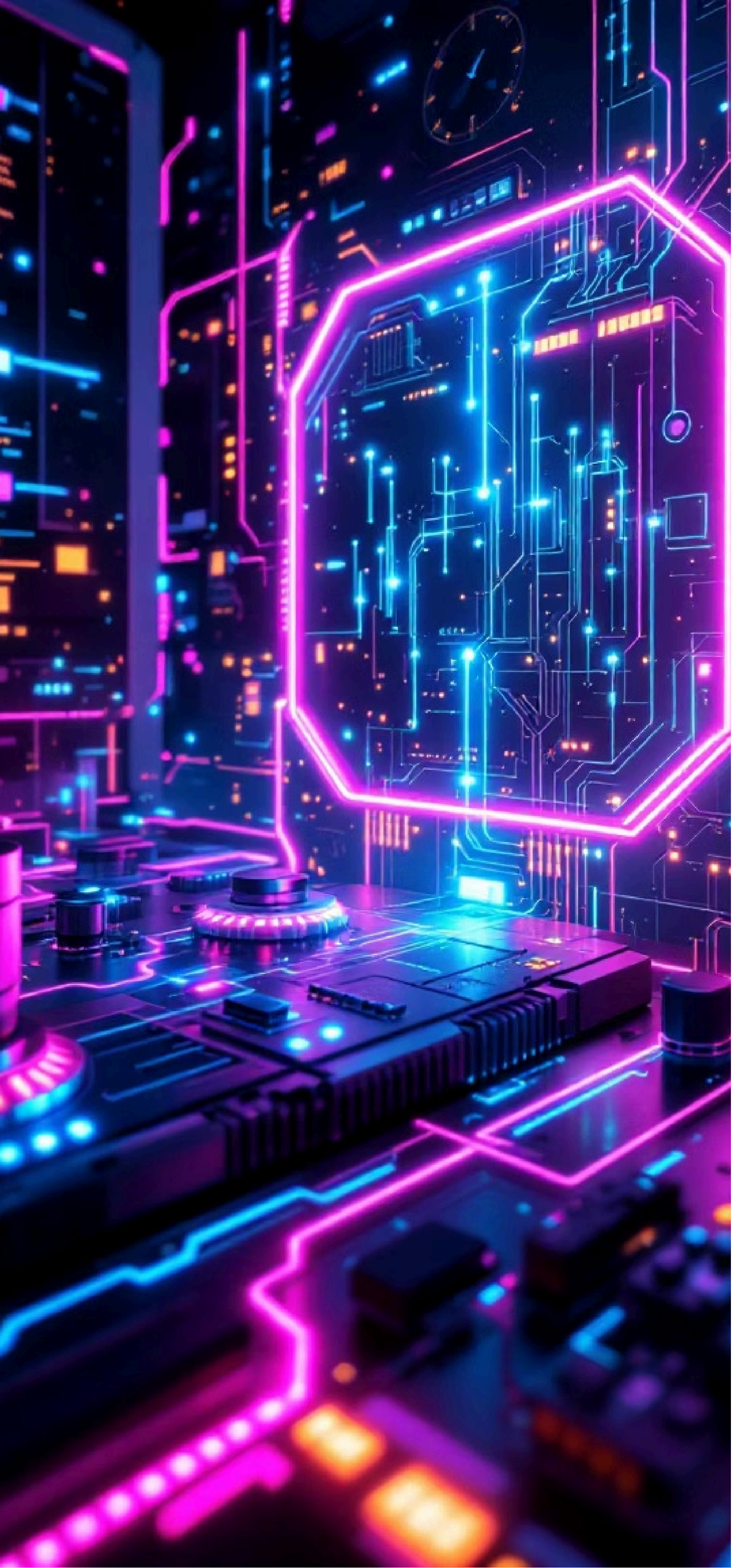
Experts in HVAC, liquid cooling, and thermal management are critical for AI-scale facilities. Universities are struggling to produce graduates fast enough.

### Data Center Technicians

Hands-on operators who can maintain 24/7 operations across increasingly complex systems. Training programs are expanding but can't keep pace with demand.

### Energy Managers

New role focused on optimizing power procurement, managing renewable portfolios, and coordinating with utility partners.



# Future Technology Trajectories

1

### Next-Generation Chips

More efficient AI accelerators from Nvidia, AMD, and custom silicon from hyperscalers will gradually reduce power per operation, though total demand continues growing.

2

### Quantum Computing Integration

Hybrid classical-quantum systems for specific AI optimization problems may emerge, requiring entirely new infrastructure paradigms.

3

### Neuromorphic Computing

Brain-inspired chips that operate at dramatically lower power levels could eventually transform AI efficiency, though commercial viability remains 5-10 years away.

4

### Advanced Cooling

Two-phase immersion cooling, on-chip liquid cooling, and other emerging technologies will enable even higher densities.

# Regulatory and Policy Outlook

Governments worldwide are grappling with how to regulate datacenter development while supporting technological innovation. The policy landscape is evolving rapidly across multiple dimensions.

### Energy Policy

New frameworks for behind-the-meter generation, interconnection standards, and priority access to grid capacity are under development in multiple jurisdictions.

### Environmental Review

Accelerated permitting for datacenter projects balanced against environmental impact assessments, particularly regarding water use and habitat protection.

### Data Sovereignty

Countries increasingly requiring that citizen data be processed within national borders, driving localized datacenter construction despite inefficiencies.

# Investment Outlook and Financial Markets

The datacenter sector has become one of the most attractive infrastructure investment categories, drawing capital from diverse sources including institutional investors, sovereign wealth funds, and infrastructure-focused private equity.

Total investment in datacenter infrastructure is projected to exceed $500 billion between 2025-2030, with AI-focused facilities commanding premium valuations due to long-term lease commitments from hyperscalers.

Public datacenter REITs have outperformed broader market indices, while private equity firms are raising dedicated funds for datacenter development.

**$500B**

**Investment Pipeline**

2025-2030 projected

**15%**

**Annual Returns**

Datacenter REIT average

# Risk Factors and Vulnerabilities

### Technology Disruption Risk

Breakthrough efficiency gains in AI chips or alternative computing paradigms could suddenly render current infrastructure obsolete, stranding billions in invested capital.

### Energy Price Volatility

Despite long-term PPAs, exposure to energy market fluctuations remains significant. A sharp spike in electricity costs could impact profitability and competitiveness.

### Regulatory Uncertainty

Changing interconnection rules, environmental standards, or tax policies could dramatically alter project economics and timelines.

### Geopolitical Factors

AI infrastructure has become strategically important, potentially subject to export controls, security requirements, or international tensions.

# Strategic Recommendations

## For Datacenter Operators and Developers

### 1 Secure Power First

Prioritize power availability over all other site selection criteria. Establish relationships with utilities, renewable developers, and nuclear operators early in the planning process.

### 2 Design for Flexibility

Build infrastructure that can adapt to evolving chip architectures and cooling technologies. Modular approaches allow for incremental upgrades without complete rebuilds.

### 3 Invest in Workforce

Develop training programs and partnerships with technical schools to build a reliable talent pipeline. Competitive compensation alone won't solve workforce shortages.

### 4 Engage Communities Early

Proactive community engagement and benefit-sharing arrangements can smooth permitting and reduce local opposition to large-scale projects.

# Conclusion: The Road Ahead

The AI revolution has fundamentally transformed the datacenter industry, creating unprecedented demand for power, specialized infrastructure, and innovative solutions to seemingly intractable constraints. The $315 billion invested by hyperscalers in 2025 represents just the beginning of a multi-year buildout that will reshape electricity markets, real estate values, and the competitive landscape of technology.

The shift from traditional hyperscale to AI-scale infrastructure is not simply an incremental evolution—it represents a complete reimagining of digital infrastructure. Power has replaced space as the primary constraint, nuclear energy has emerged as a critical enabling technology, and entire communities are being transformed by the arrival of these massive facilities.

As we look toward 2030, several key trends will define success: those who secure power capacity earliest will gain insurmountable advantages, partnerships between technology companies and energy providers will deepen, and sustainability considerations will increasingly shape both public policy and corporate strategy.

> **The future of AI depends on solving infrastructure challenges today.** The companies and countries that navigate this complex landscape successfully will lead the next era of technological innovation.