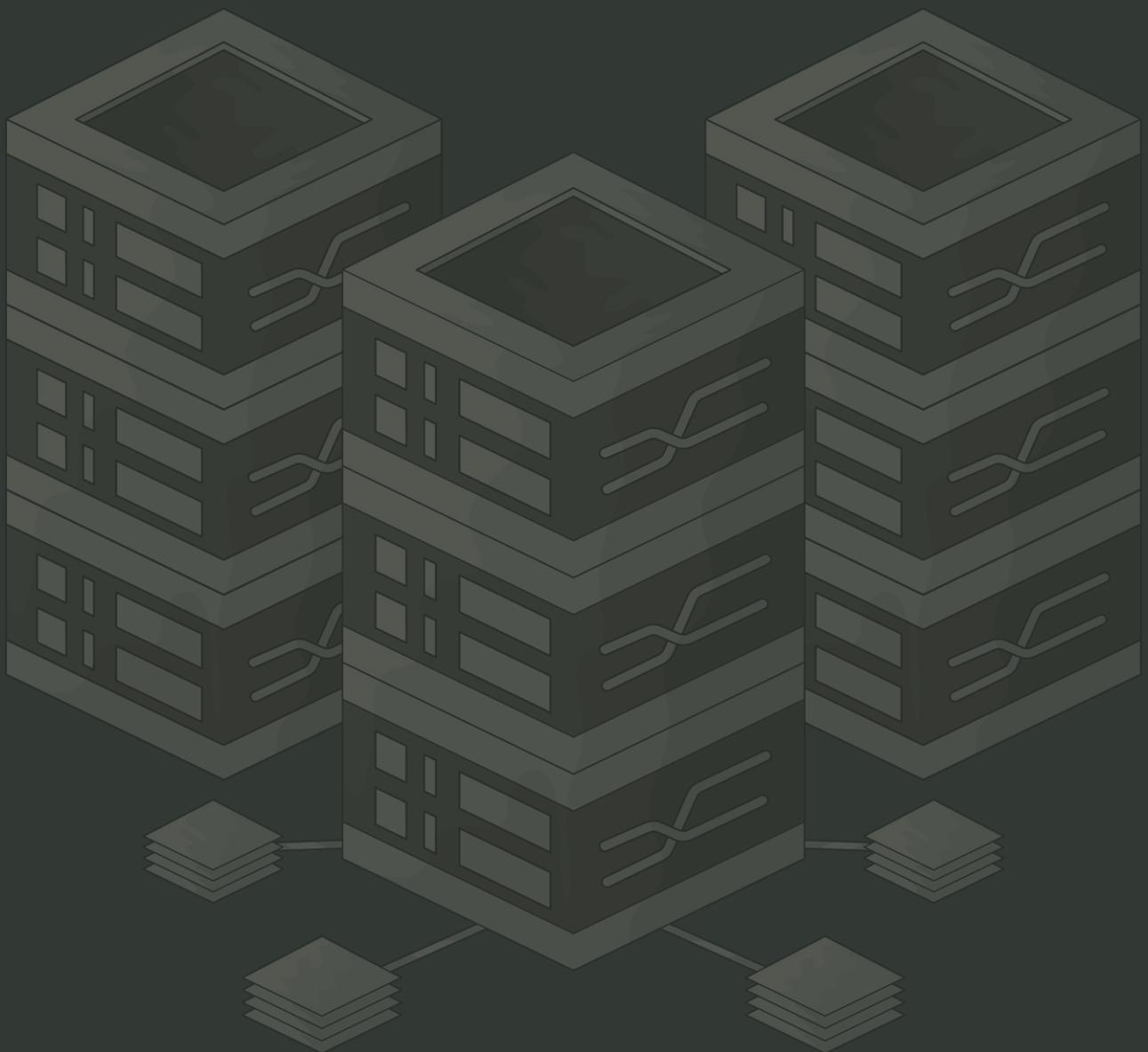


# Mapping the Neocloud Landscape



# Executive Summary

This report analyzes the emergence of the neocloud: a new category of specialized provider purpose-built for AI. Our research concludes that neoclouds are not a niche trend but a fixture in the market, forming an essential, parallel infrastructure layer alongside traditional hyperscalers. This position is contingent on two factors: the sustained, structural demand for AI workloads and the stability of the high-leverage financial models that fund their growth.

Four key findings define the market in 2025:

## **1. The Market Has Flipped: Competitors Are Now Customers.**

The most significant trend is that hyperscalers have become the neoclouds' largest clients. Faced with an internal shortage of AI-ready infrastructure, they have signed massive multi-year contracts to rent capacity, led by Microsoft's commitment of over \$60 billion. This is a tactical move to meet immediate AI demand faster than they can build.

## **2. The Bottleneck Has Moved from Silicon to Power.**

The defining constraint on AI's growth has changed. In 2024, the challenge was the supply of advanced GPUs. In 2025, the primary bottleneck is physical: securing enough power, grid access, and specialized data center capacity. The industry is now constrained by its ability to plug in the chips it already has.

## **3. The "Big 4" Lead Through Specialization.**

The market is consolidating around four distinct leaders, each with a different strategic advantage: CoreWeave (preferred access to Nvidia hardware), Lambda (developer-focused ecosystem), Crusoe (vertically integrated energy), and Nebius (European data sovereignty).

## **4. The Next Battle is for What Runs on Top.**

Looking ahead, competition will shift beyond just providing hardware. The long-term winners will be determined by their ability to compete on three new fronts: securing low-cost energy, meeting country-specific sovereignty laws, and building the high-value software and MLOps stack for developers.

# Table of Contents

Executive Summary	<b>2</b>
The Emergence of Neocloud	<b>4</b>
The Neocloud Business	<b>6</b>
The Market Dynamics	<b>8</b>
The “Big 4” Competitive Landscape	<b>11</b>
The New Industrial Constraints	<b>17</b>
Implications for Your Role	<b>19</b>
The Three Battlegrounds for 2026	<b>21</b>
About ARPU	<b>23</b>

# The Emergence of Neocloud

## *Why the “Supermarket” Cloud Broke and the “Delicatessen” Emerged*

For over a decade, the public cloud was defined by a single, successful model: the general-purpose hyperscaler. Providers like AWS, Azure, and Google Cloud built their dominance by operating as vast digital “supermarkets.” Their strength was offering a massive catalog of services, from databases to web hosting, built on a versatile, CPU-centric architecture designed for horizontal scale.

But the emergence of large-scale AI—specifically the training of foundational models with hundreds of billions of parameters—created a cloud market rupture. Unlike traditional cloud computing, which handles millions of separate, independent tasks (like a user loading a webpage or querying a database), training a single, massive AI model is a completely different kind of problem. It is a monolithic, tightly-coupled workload. It requires thousands of individual GPUs to be clustered together and operate in perfect concert as a single, virtual supercomputer, often for weeks or months at a time.

This core requirement—thousands of processors working simultaneously on one job—is what breaks the legacy data center design. The need to cluster these high-power processors creates a cascade of physical problems that infrastructure optimized for 5-10 kilowatt (kW) air-cooled racks cannot easily solve:

- **Extreme Power Density:** To enable communication, GPUs must be packed densely together. A single server rack filled with modern GPUs can demand 50-100+ kW of power—an order-of-magnitude increase that legacy power distribution systems were never designed to deliver.
- **Thermal Wall:** This extreme power density generates an immense amount of concentrated heat. Traditional air cooling is physically incapable of removing this much heat from such a small space, making a complete shift to high-performance liquid cooling a mandatory, non-negotiable requirement.
- **Networking Bottleneck:** For thousands of GPUs to act as one, they must exchange massive amounts of data with near-zero delay. Standard data center networking is far too slow and creates a crippling bottleneck. This necessitates a specialized, ultra-low-latency fabric like InfiniBand or NVLink, which is not standard in legacy facilities.

Because of these cascading problems across power, cooling, and networking, you cannot simply upgrade a traditional data center to run large-scale AI. The physics of the workload demand a new type of facility be built from the ground up. This is the architectural rupture that created the market for the neocloud.

Continuing the earlier analogy, if hyperscalers are the supermarket, neoclouds are the specialized “delicatessen.” They forgo a broad catalog of services to do one thing exceptionally well: deliver raw, high-performance, GPU-centric computing. Their entire stack—from the power grid connection to the software layer—is optimized for AI workloads, eliminating the overhead, performance bottlenecks, and prohibitive costs of running these tasks on a general-purpose platform. This shift represents a permanent bifurcation of the cloud market. AI infrastructure is becoming a fundamentally different class of utility.

### The Physical Shift from Legacy to AI Data Centers

Architectural Element	Legacy Data Centers	Modern AI Data Centers
Typical Rack Density	5-10 kW	50-100 kW
Cooling	Air-cooled	Liquid cooling required >50kW
Networking	Standard Ethernet	NVLink, InfiniBand

# The Neocloud Business Model

*Where Competitors Become Customers and Nvidia Plays Kingmaker*

The neocloud phenomenon is enabled by a fundamentally different economic engine than the one that powers traditional hyperscalers. Their entire business model is an exercise in specialization, designed to offer superior performance at a significantly lower cost.

## Transparent Pricing & The Cost Advantage

The most disruptive element of the neocloud model is its pricing. Unlike hyperscalers, who are notorious for complex, multi-layered billing, neoclouds compete on radical transparency.

- **The Model:** A simple, flat-rate, per-GPU-hour price that bundles compute, networking, and support.
- **The Advantage:** This model eliminates the “hidden fees”—such as high-cost data egress charges or API call fees—that make cloud bills from hyperscalers unpredictable. This predictability is highly valued by developers and financial operators.

This specialized, lean approach is what enables their primary value proposition: a 30-66% lower price for equivalent high-performance compute.

## The Lean Operating Model

This cost advantage is not a marketing gimmick; it's the result of a disciplined, lean operational strategy. While hyperscalers invest billions in massive, owned-and-operated campuses, most neoclouds adopt a more agile approach.

By renting space in specialized, GPU-ready colocation data centers, they avoid the massive upfront capital expenditure and long depreciation cycles of building. This allows them to scale capacity quickly in response to demand. However, this model is unforgiving; their financial viability hinges on maintaining extremely high hardware utilization, often needing to exceed 66% utilization to remain cost-competitive with the public cloud.

## The New Financial Leverage: Contract-Backed Debt

The single biggest challenge for any neocloud is financing its multi-billion dollar fleet of GPUs. While venture capital is used for initial growth, the most successful providers have pioneered a more powerful mechanism: **contract-backed debt**.

This model of financial engineering is the key to their explosive scaling and works as follows:

- A neocloud signs a massive, multi-year, multi-billion-dollar capacity agreement with an anchor tenant (e.g., Microsoft).
- They take this legally-binding contract, which represents a guaranteed future revenue stream from a highly creditworthy company, to lenders.
- Lenders then issue billions in debt, secured not just against the physical GPUs but against this guaranteed revenue.

This structure allows neoclouds to finance their growth with less-dilutive debt instead of equity. It also reveals a deeper strategic alignment: major chipmakers like Nvidia often provide additional guarantees to lenders, backstopping these deals to ensure their most important partners have the capital they need to buy more hardware. This combination of asset-backed and contract-backed financing is the financial engine that is fueling the entire neocloud build-out.



When we wanted to create a new type of cloud, we invested in a small company called CoreWeave... We helped them build themselves into a company, and they went public and we made a fortune. And that was fantastic. It was great for them. It was great for us.

**- Jensen Huang, Nvidia CEO**

***CNBC Interview***

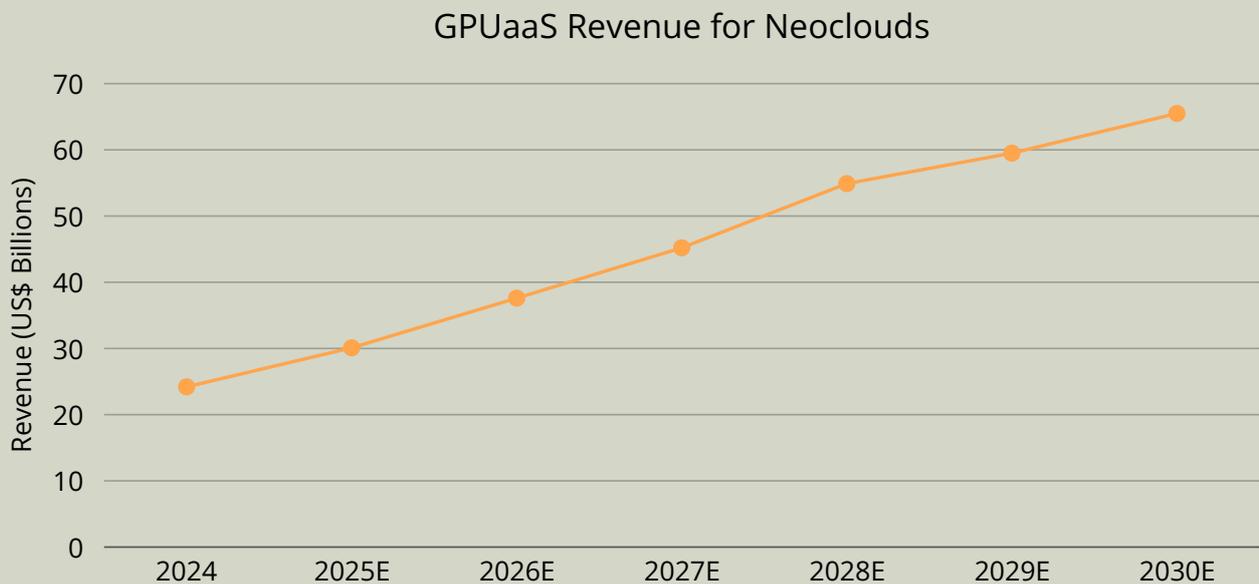
# The Market Dynamics

## *Where Competitors Become Customers and Nvidia Plays Kingmaker*

The bifurcation of the cloud is a tectonic event backed by a massive flow of long-term capital commitments and infrastructure contracts. This is the land grab that is shaping the new market.

First, it is critical to understand the scale and velocity of this new market by examining its core service: GPU-as-a-Service (GPUaaS). According to ABI Research, GPUaaS revenue from neoclouds exceeded \$24 billion in 2024, and is projected to reach \$65.5 billion by 2030.

This rapid expansion is part of a much larger ecosystem. When including related platform and software services, the total neocloud market is forecast by Synergy Research Group to approach \$180 billion by 2030, growing by an average 69% per year. This makes neocloud one of the fastest-growing segments in technology infrastructure.



Source: ABI Research

The market demand of neocloud is the direct result of three forces colliding at once.

### **Force 1: The Great Compute Hedge**

The most significant force defining the 2025 landscape is that hyperscalers have become the neoclouds' largest customers. This "frenemy" dynamic is driven by a critical capacity crisis inside the big cloud providers.

Faced with insatiable AI demand from their own customers and a multi-year lead time to build new, specialized data centers, they have turned to neoclouds to fill the gap. Microsoft is the primary example, committing over \$60 billion in multi-year deals with providers like Nebius, CoreWeave, and Nscale.

This hedge allows hyperscalers to meet immediate customer demand, accelerate their AI roadmaps, and convert a high-risk capital expenditure problem (building) into a more predictable operating expense (renting). In effect, they are paying their most agile competitors to secure their own market position.

### **Force 2: The AI Lab's Declaration of Independence**

The second force is the decoupling of frontier AI labs from their exclusive cloud partners. For years, labs like OpenAI were captive to a single provider (Microsoft). In 2025, that model ended.

This is a calculated move to prevent "compute starvation" and eliminate dependency on any single infrastructure partner. The evidence is in their multi-billion dollar deals:

- **OpenAI**, historically reliant on Azure, has signed massive, multi-year capacity agreements with **AWS** and **Oracle** to power its model development and the ambitious Stargate project.
- **Anthropic**, similarly, announced a \$50 billion plan to build its own dedicated AI infrastructure, breaking from its primary reliance on **AWS** and **Google Cloud**.

These labs are intentionally creating a multi-provider environment. Their goal is to guarantee access to the massive-scale compute they need to survive while maintaining the negotiating leverage to control their own destiny.

### Force 3: The Nvidia Nexus

The final and most critical force is Nvidia, which functions not as a passive supplier but as the central bank and architect of the entire neocloud ecosystem. With its own largest customers—the hyperscalers—all developing rival chips, Nvidia has executed a second front strategy to ensure its long-term dominance.

Nvidia is using its market power and balance sheet to act as a kingmaker for the neocloud market through three distinct mechanisms:

- **Direct Equity Investment:** It has taken strategic equity stakes in all of the “Big 4” neoclouds, directly funding their growth.
- **Debt Underwriting:** It provides guarantees to lenders who finance the neoclouds, effectively de-risking their multi-billion dollar GPU purchases.
- **Preferential Hardware Access:** Most importantly, it grants its key partners priority access to its newest and most powerful chips, a decisive competitive advantage in a supply-constrained market.

This strategy creates a loyal, competitive, and 100% Nvidia-dependent market for its hardware, ensuring robust demand for years to come, regardless of the hyperscalers’ custom silicon ambitions.



I would welcome every neocloud to become part of our marketplace. Because if they bring their capacity into our marketplace, the customer who comes through Azure will end up using that neocloud’s capacity. That’s a huge win for them, and at the same time they’ll be using Azure compute, storage, databases, and everything else.

- Satya Nadella, Microsoft CEO

*Interview with Dwarkesh Patel*

# The “Big 4” Competitive Landscape

## *Four Players, Four Different Paths to Dominance*

While the neocloud market includes nearly 200 operators, the capital, hardware allocations, and large-scale contracts are consolidating around four clear leaders. These players are not just competing on price; each is building a defensible business based on a unique competitive advantage.

The table below provides a top-level comparison of these four dominant providers.

Company (Origin)	Core Strength	Customer Sweet Spot	Main Geographical Footprint
CoreWeave (2017 Crypto-Pivot)	Nvidia Hardware Access: Preferred, early access to latest GPUs.	AI Megalabs (OpenAI, Meta, Mistral AI), AI Enterprises.	USA (Megaclusters in NJ, TX, IL, GA, OR)
Lambda (2012 AI Developer Tooling)	Developer Ecosystem: Full-stack (hardware-to-cloud) focus on ML engineers.	AI R&D (Startups, Academia), Enterprise Labs (Apple, Microsoft, DoD)	USA (AI Factories in MO, IL, TX).
Crusoe (2018 Energy-Tech)	Energy Vertical Integration: Energy-first model; low-cost power from flare gas & renewables.	Sustainable Compute (OpenAI Stargate), AI Innovators (Together AI).	USA (Energy Hubs: TX, WY etc.).
Nebius (2024 Yandex Spin-off)	European AI Sovereignty: Amsterdam HQ; EU-compliant data residency.	EU Enterprises, National AI Programs, US Hyperscalers (Microsoft, Meta).	EMEA (Finland, UK, Paris, Iceland)

## **CoreWeave: Nvidia's Chosen Partner**

CoreWeave is the current leader among the specialized neoclouds, reporting over \$2.1 billion in revenue in the first half of 2025 alone. The company originated as a large-scale crypto-mining operation, giving it deep expertise in operating high-density, power-intensive GPU infrastructure.

CoreWeave's primary strength is its close relationship with Nvidia. As an early partner and a company in which Nvidia is a direct investor, CoreWeave receives priority, early access to the newest and most powerful GPUs. In a supply-constrained market, this hardware advantage is decisive. It allows CoreWeave to service the world's largest AI labs—including OpenAI, Meta, and Mistral AI—who depend on having immediate access to the fastest chips at massive scale.

## **Lambda: The Developer's Cloud**

Lambda has established itself as a strong contender by focusing its entire strategy on the end-user: the machine learning engineer. The company's history informs this approach; it began not as a cloud provider but as a hardware company building and selling GPU workstations and servers directly to AI developers.

This gives Lambda a unique, full-stack moat built on deep credibility within the developer ecosystem. Its platform is engineered to remove friction for AI researchers, with products like its pre-configured Lambda Stack and 1-Click Clusters for rapid deployment. Lambda is particularly strong with AI-native startups, academic labs, and enterprise R&D teams that prioritize a developer-first, turnkey experience — as reflected in notable customers such as Apple, MIT, and U.S. government and defense organizations.

## **Crusoe: The Vertically Integrated Energy Model**

Crusoe's origin story defines its unique competitive position. It is not a legacy tech or crypto company, but an energy-tech company founded to solve the problem of natural gas flaring at oil fields. It pioneered a model of deploying mobile data centers to capture this wasted energy and use it to power computation.

This gives Crusoe a powerful moat built on the vertical integration of energy. Its "energy-first" approach involves co-locating its data centers directly with wasted, stranded, or curtailed energy sources, including flare gas and renewable hubs. This provides two key advantages: access to power at below-market rates and the ability to build and power its "AI Factories" significantly faster than competitors constrained by the public grid. This unique, sustainable model was a key factor in its selection as a primary partner for OpenAI's massive Stargate project.

## **Nebius: The European Sovereignty Play**

Nebius has the most complex geopolitical origin, having been spun out of the international arm of "Russia's Google," Yandex. Now a fully independent, Amsterdam-headquartered company, Nebius is leveraging this background to build a powerful and unique moat: European AI Sovereignty.

In a market where data privacy (GDPR) and digital sovereignty are paramount concerns for European enterprises and governments, Nebius is positioned as the default regional champion. It provides the compliant, EU-based, "sovereign AI" alternative to the US-based hyperscalers and neoclouds. This has allowed it to pursue a powerful dual strategy: winning sovereign-conscious European customers while also serving as a massive wholesale capacity provider to US tech giants like Microsoft and Meta, who need a compliant partner to service their own European AI ambitions.

## Deconstructing the Hyperscaler, Neocloud, and Nvidia Nexus

The deals listed below reveal a sophisticated financing strategy that is fueling the entire neocloud build-out:

- **Anchor Contracts (The Collateral):** The massive, multi-billion dollar capacity agreements from hyperscalers like Microsoft (with Nebius, Nscale, etc.) and Meta (with CoreWeave) are the foundational assets. They act as guaranteed revenue streams that make the entire model “bankable.”
- **Capital Infusion (The Fuel):** This collateral is then used to raise billions in capital through multiple channels. This includes traditional debt financing (like Crusoe’s \$750M credit facility from Brookfield) and accessing the public markets for equity (like CoreWeave’s \$1.5B IPO).
- **Kingmaker’s Backstop (The Guarantee):** Notice Nvidia’s role not just as a supplier, but as a direct strategic investor (e.g., in Lambda’s February funding round). Nvidia also provides crucial backstop guarantees (like the September arrangement with CoreWeave), de-risking the entire financial model to ensure their hardware gets sold.

### Notable Neocloud Deals (Jan 2025 - present)

Company	Date	Deal Description
CoreWeave, IBM	Jan 2025	CoreWeave partnered with IBM to deliver a new AI supercomputer equipped with Nvidia GB200 Grace Blackwell Superchips and Quantum-2 InfiniBand networking, integrated with IBM Storage Scale System for high-performance storage, to train and optimize IBM's open-source Granite AI models.
Lambda, Raison	Feb 2025	Raison invested in Lambda via a special purpose vehicle, joining early institutional backers to fuel expansion of Lambda’s AI infrastructure business.
Lambda Labs, Andra Capital, SGW, Nvidia	Feb 2025	Lambda Labs raised \$480 million in a Series D funding round led by Andra Capital and SGW with participation from Nvidia and others, to accelerate the expansion of its GPU-powered AI cloud platform.
Nebius, DataOne	Mar 2025	Nebius announced a partnership with DataOne to develop a new data center facility offering up to 300 MW of compute capacity in the New Jersey region, scheduled to open in summer 2025.

Company	Date	Deal Description
CoreWeave, OpenAI	Mar 2025	CoreWeave signed a multi-year agreement with OpenAI to provide dedicated high-performance compute capacity for training and deploying advanced AI models
Nebius, Verne	Mar 2025	Verne announced that Nebius will be colocating a 10 MW cluster of Nvidia H200 GPUs at Verne's data center campus in Iceland
CoreWeave	Mar 2025	CoreWeave completed its initial public offering on Nasdaq on March 27, 2025, raising \$1.5 billion through the sale of shares, to fund the rapid scaling of its AI hyperscale cloud infrastructure and data centers.
Lambda, Aligned	May 2025	Lambda partnered with Aligned, a data center infrastructure company to provide customers with data center infrastructure and an AI cloud platform. Lambda will occupy Aligned's newest Dallas-Fort Worth area facility, DFW-04
Crusoe, Brookfield	Jun 2025	Crusoe secured a \$750M credit facility from Brookfield that will primarily be deployed to fuel the continued growth and scaling of Crusoe's development of AI factories.
Crusoe, Redwood Materials	Jun 2025	Crusoe and Redwood Materials announced a partnership to unveil the world's largest second-life battery deployment, using repurposed EV batteries to power mobile, off-grid AI data centers
CoreWeave, Core Scientific	Jul 2025	Announced agreement to acquire Core Scientific in all-stock deal, aiming to bring 1.3 GW of US data center capacity in-house, for greater vertical control and efficiency in AI/HPC infrastructure .
CoreWeave	Jul 2025	CoreWeave announced a commitment to equip a state-of-the-art data center with an initial 100 MW capacity, expandable to 300 MW, to expand its AI infrastructure and support growing customer demand for high-density computing.
Crusoe, Tallgrass	Jul 2025	Partnership between Crusoe and Tallgrass to develop a 1.8GW AI campus in Wyoming, with expansion plans up to 10GW .
Crusoe, atNorth	Aug 2025	Crusoe and atNorth, a leading Nordic data center services provider, announced an agreement to expand its clean energy-powered AI data center in Iceland, featuring Nvidia DGX and Blackwell GPUs.

Company	Date	Deal Description
Nebius, Microsoft	Sep 2025	Nebius announced a multi-billion dollar agreement with Microsoft. Under this multi-year agreement, Nebius will deliver dedicated capacity to Microsoft from its new data center in Vineland, New Jersey starting later this year.
CoreWeave, Nvidia	Sep 2025	CoreWeave agreed to a strategic collaboration with Nvidia in a deal that guarantees that the AI chipmaker will purchase any cloud capacity not sold to customers.
CoreWeave	Sep 2025	CoreWeave committed an incremental investment to accelerate AI innovation and growth through sustainable computing initiatives
CoreWeave, OpenAI	Sep 2025	CoreWeave expanded its partnership with OpenAI through a \$6.5 billion addition to prior commitments, granting OpenAI usage rights to data centers equipped with Nvidia GPUs through 2029
CoreWeave, Meta	Sep 2025	CoreWeave announced an agreement with Meta. As a part of this agreement, CoreWeave will provide the Facebook-parent access to Nvidia's latest GB300 systems
Nscale, Microsoft	Oct 2025	Microsoft committed to a multi-year arrangement with Nscale, a British AI cloud startup, providing access to approximately 200,000 Nvidia GB300 GPUs across multiple data center sites, to secure specialized GPU capacity for Microsoft's AI workloads.
Crusoe, Valor Equity Partners, Mubadala Capital	Oct 2025	Crusoe announced a Series E funding round led by Valor Equity Partners and Mubadala Capital.
Lambda, Microsoft	Nov 2025	Microsoft announced a multibillion-dollar partnership with Lambda to deploy tens of thousands of Nvidia GPUs, including advanced GB300 NVL72 systems, building on their eight-year collaboration to construct massive AI supercomputers.
IREN, Microsoft	Nov 2025	Microsoft announced a five-year cloud services agreement with IREN Limited to access Nvidia GB300 GPUs at IREN's 750 MW data center campus.
Nebius, Meta	Nov 2025	Nebius signed a five-year agreement with Meta to deliver AI infrastructure capacity, with deployment of the required resources over the next three months.

# The New Industrial Constraints

## *A race for grid connections*

For the past two years, the AI narrative has been dominated by a single question: “Can we get enough GPUs?” In 2025, that question has become secondary. The new, more critical question is: “Can we find a place to plug them in?”

The primary bottleneck for AI’s growth has fundamentally shifted from the supply of silicon to the availability of physical infrastructure—specifically, power and cooling. This new reality is reshaping every investment decision, construction timeline, and geographic strategy in the industry.

## **The Power Bottleneck**

The most direct evidence of this shift comes from the hyperscalers themselves. Microsoft CEO Satya Nadella recently stated plainly that the company is no longer chip-constrained, but power-constrained. He noted that Microsoft has GPUs in its inventory that they simply “can’t plug in” due to a lack of available electricity.

This constraint is rooted in the new physics of AI infrastructure:

- An AI query consumes 10 times more energy than a traditional web search.
- An AI server rack demands 50 times the power density (power per square foot) of a traditional server rack.

The industry is now on a collision course with public utility infrastructure. Growth is no longer gated by the 6-9 month chip production cycle, but by the 3-5 year timeline required to secure grid connections and build new, high-density data centers.

## **The Thermal Wall**

Extreme power density creates extreme heat. The 50-100+ kW of power running through a single AI rack generates a “thermal wall” that traditional air cooling cannot overcome.

This has made high-performance liquid cooling a mandatory, non-negotiable technology for any new AI data center. The shift is so significant that it has created a new, ancillary market for specialized cooling systems projected to be worth over \$3 billion in 2025. This is not an upgrade; it is a complete architectural redesign.

## **The Financial Mismatch**

This new physical reality creates a profound and systemic financial risk that is largely invisible on company earnings reports today: a mismatch between the economic life of AI hardware and the accounting practices used to depreciate it.

Understanding this mismatch is critical to assessing the long-term sustainability of the current investment boom.

- **The Economic Reality:** The useful economic lifespan of a high-end GPU is brutally short, estimated at just 3 to 5 years. This is driven by two factors: rapid technological obsolescence (as Nvidia introduces next-generation chips like its forthcoming Rubin platform) and intense physical wear from 24/7 operation.
- **The Accounting Practice:** Despite this 3-5 year reality, many companies are depreciating these multi-billion-dollar assets over 5 to 6 years on their balance sheets.

This accounting discrepancy systematically overstates the long-run profitability of the industry. It creates a “debt-and-obsolescence spiral”: companies are forced to take on massive new debt to purchase the next generation of chips while they are still paying off the previous generation, which is already becoming obsolete.

This creates significant refinancing pressure and represents a ticking clock on the balance sheet for any operator in this space. The single greatest long-term risk in the AI infrastructure market is not a slowdown in demand, but this collision between the relentless pace of hardware innovation and the financial models used to pay for it.

# Implications For Your Role

Major market events never happen in isolation. They create cascading effects that ripple across the entire ecosystem, impacting each business and function in a unique way.

A hyperscaler's decision to rent capacity, for example, is the first domino. The cascading effects include new vendor risks for their enterprise customers, new partnership complexities for software vendors, and a new capital strategy for competing neoclouds.

While most analysis focuses on the initial event, ARPU specializes in mapping these cascading effects directly to your specific role and business model.

To demonstrate, let's deconstruct one of the most significant forces identified in this report—The Great Compute Hedge—and show how this “frenemy” dynamic between hyperscalers and neoclouds creates distinct, actionable intelligence for different tech operators.

## One Trend, Three Bespoke Implications: When Competitors Become Customers

The fact that hyperscalers like Microsoft are renting billions in capacity from neoclouds is not just a deal headline; it's a market-altering event with cascading consequences for everyone in the ecosystem.

### **For the Go-to-Market Leader at a B2B Software Vendor:**

Your partnership landscape just became more complex. When you sign a co-sell deal with Microsoft Azure, your customer's AI workload may not actually run on Microsoft's physical servers, but on CoreWeave's. This introduces new layers of technical due diligence and complicates security and compliance conversations, potentially stalling your deals.

**The Key Question:** How do you re-tool your sales and partnership strategy to navigate a multi-layered cloud environment where your primary partner is also a customer of your customer's other potential vendors?

### ⚙️ **For the VP of Platform Engineering at a Large Enterprise:**

Your vendor risk profile has changed. You signed a contract with a hyperscaler for the security, stability, and compliance guarantees their brand provides. But for your most critical AI workloads, they are now acting as a middleman, subcontracting the core infrastructure to a much younger company. You now have an indirect dependency on a neocloud's performance, security, and financial stability.

**The Key Question:** How do you update your risk management and technical due diligence framework to account for the opaque, multi-layer supply chain that now underpins your primary cloud provider's AI services?

### 💰 **For the Head of FinOps / IT Procurement at a Global Enterprise:**

You may be paying a hidden "reseller tax." If your Hyperscaler is fulfilling your GPU quota by routing workloads to a partner Neocloud, you are effectively paying a premium for the Hyperscaler's brand and SLA, while the underlying compute is being sourced from a lower-cost provider. You are paying blue chip rates for neocloud metal, creating a pricing arbitrage opportunity that is currently invisible on your invoice.

**The Key Question:** How do you audit your AI infrastructure spend to identify where you are paying "Hyperscaler margins" for third-party hardware, and can you reduce costs by bypassing the middleman and contracting directly?

**Ready to see what these trends mean for your 2026 roadmap?**

**Request a Complimentary, Personalized Sample Brief**

Contact us online at: <https://arpu.hedder.com/contact-us/>

Or email our team at: [arpu@hedder.com](mailto:arpu@hedder.com)

# The Three Battlegrounds for 2026

## *The Next Phase of the Neocloud War*

The neocloud market has moved beyond its initial "gold rush" phase. The 2025 story was about securing scarce GPUs and landing massive anchor-tenant contracts. While that race continues, the capital has been committed and the market leaders are established. The terms of competition are now changing.

The next phase of the neocloud war will not be won simply by deploying more hardware. Long-term dominance will be determined by a provider's ability to win on three new, more complex strategic battlegrounds.

### **The Battle for Energy**

The ultimate bottleneck is power. Crusoe's energy-first model is the leading indicator of a market-wide shift where access to cheap, abundant, and reliable electricity is the most critical competitive advantage. The winners will be those who can move beyond standard grid contracts and vertically integrate their energy supply, whether through partnerships with renewable hubs or by pioneering new power sources. The race for AI leadership is now a race for gigawatts.

### **The Battle for Sovereignty**

The "one-size-fits-all" global cloud is fragmenting. As demonstrated by Nebius's success in Europe, the next wave of high-value growth will come from winning regulated industries and national markets. This requires building jurisdiction-aware, compliant infrastructure that can satisfy strict data residency laws like GDPR. This is not a feature; it is a distinct business model. The providers who master this geopolitical complexity will unlock the lucrative government, finance, and healthcare sectors that their US-based competitors cannot easily reach.

## **The Battle for the Stack**

The greatest long-term risk for any neocloud is becoming a low-margin, commodity provider of GPU-hours. The most durable business is not in the infrastructure, but in the high-margin software platform that runs on top of it. The strategic imperative is to move up the stack—building the developer ecosystems, managed Kubernetes platforms, MLOps tools, and data platforms that create deep customer lock-in. The future belongs to the provider that can transition from being an "AI factory" to being the indispensable "AI operating system."

# About ARPU

ARPU provides bespoke, function-specific intelligence for tech operators. In a market saturated with generic research and noise, we deliver signal.

Our analysis translates market-wide events and trends into intelligence briefs tailored to the specific challenges and questions of your role. We exist to help you make better, faster decisions.

## Go Deeper.

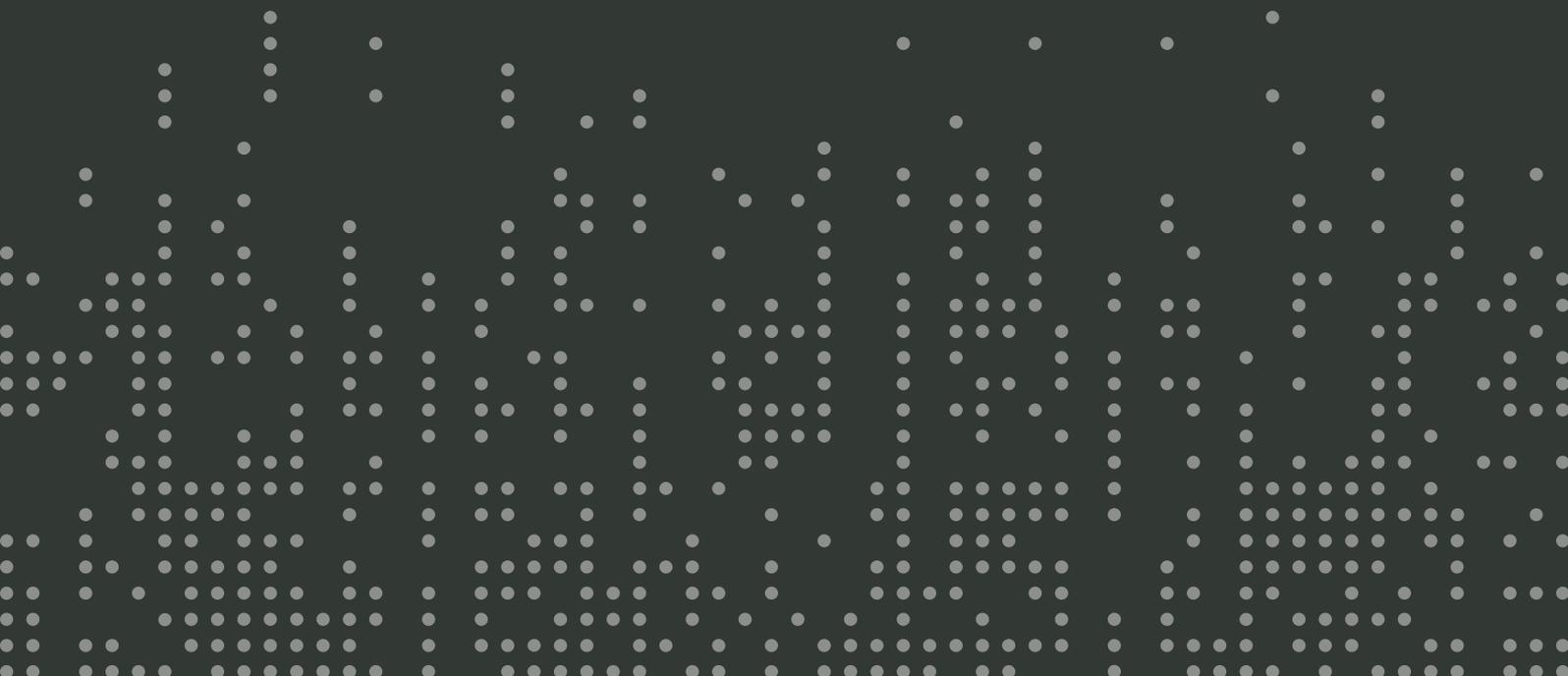
This report provided a market-wide view of the neocloud landscape. But your organization, your products, and your strategic questions are unique.

**Ready to see what these trends mean for your 2026 roadmap?**

**Request a Complimentary, Personalized Sample Brief**

Contact us directly at [arpu@hedder.com](mailto:arpu@hedder.com)

Or visit our website at <https://arpu.hedder.com/>



# Sources

1000 homes of power in a filing cabinet - rising power density disrupts AI infrastructure (<https://www.goldmansachs.com/insights/articles/rising-power-density-disrupts-ai-infrastructure>)

AI Data Center vs Traditional Data Center: Key Differences Explained (<https://www.hyperbolic.ai/blog/ai-data-center-vs-traditional-data-center-key-differences>)

AI Datacenter Liquid Cooling Market (<https://www.futuremarketinsights.com/reports/ai-datacenter-liquid-cooling-market>)

AI Infra Summit 2025: Are Neoclouds the Next Hyperscalers? (<https://datacenterpost.com/ai-infra-summit-2025-are-neoclouds-the-next-hyperscalers/>)

Announcing Our Top 10 Cloud Trends For 2025 - Forrester (<https://www.forrester.com/blogs/announcing-our-top-10-cloud-trends-for-2025/>)

Can US infrastructure keep up with the AI economy? (<https://www.deloitte.com/us/en/insights/industry/power-and-utilities/data-center-infrastructure-artificial-intelligence.html>)

CoreWeave Buying Marimo To Expand Unified AI Developer Platform (<https://pulse2.com/coreweave-buying-marimo-to-expand-unified-ai-developer-platform/>)

CoreWeave's great AI infrastructure race (<https://www.constellationr.com/blog-news/insights/coreweaves-great-ai-infrastructure-race>)

Europe Seeking Greater AI Sovereignty, Accenture Report Finds (<https://newsroom.accenture.com/news/2025/europe-seeking-greater-ai-sovereignty-accenture-report-finds>)

Industry Credit Outlook: Despite The Risks, Tech, Power, And Data Center Companies Are Going All-In On Their AI Gamble - S&P Global (<https://www.spglobal.com/ratings/en/regulatory/article/industry-credit-outlook-despite-the-risks-tech-power-and-data-center-companies-are-going-all-in-on-their-ai-gamble-s101647473>)

Lifespan of AI Chips: The \$300 Billion Question - CITP Blog (<https://blog.citp.princeton.edu/2025/10/15/lifespan-of-ai-chips-the-300-billion-question/>)

Microsoft commits \$60b to neocloud data center providers (<https://www.techinasia.com/news/microsoft-commits-60b-to-neocloud-data-center-providers>)

Microsoft Taps Nebius in \$17B AI Infrastructure Deal To Alleviate Cloud Strain (<https://rcpmag.com/articles/2025/09/12/microsoft-nebius-ai-infrastructure-deal.aspx>)

Neocloud GPUaaS Revenue (<https://www.abiresearch.com/news-resources/chart-data/gpu-as-a-service-revenue-for-neoclouds>)

Neocloud provider IREN secures US\$9.7bn AI cloud deal with Microsoft (<https://staging.w.media/neocloud-provider-iren-secures-us9-7bn-ai-cloud-deal-with-microsoft/>)

Neoclouds Currently Growing by Over 200% per Year; Will Reach \$180 Billion in Revenues by 2030 (<https://www.srgresearch.com/articles/neoclouds-currently-growing-by-over-200-per-year-will-reach-180-billion-in-revenues-by-2030>)

Neoclouds ride a runaway revenue growth train to 2030 - Fierce Network (<https://www.fierce-network.com/cloud/neoclouds-ride-runaway-revenue-growth-train-2030>)

Neoclouds: a cost-effective AI infrastructure alternative (<https://journal.uptimeinstitute.com/neoclouds-a-cost-effective-ai-infrastructure-alternative/>)

Neoclouds: The new cloud players vying for the AI crown - STL Partners (<https://stlpartners.com/articles/edge-computing/neoclouds-the-new-cloud-players/>)

Nvidia-backed Lambda lands \$480M at \$4B valuation (<https://techfundingnews.com/nvidia-backed-lambda-lands-480m-at-4b-valuation-to-scale-its-ai-cloud/>)

Profiling Six Leading Neocloud Companies (<https://www.abiresearch.com/blog/leading-neocloud-companies>)

The Future of AI & High-Performance Computing (<https://edgecore.com/ai-data-center-ifastructure/>)

The rise of Neocloud in the AI landscape - JLL (<https://www.jll.com/en-us/insights/the-rise-of-neocloud-in-the-ai-landscape>)

VAST Data and CoreWeave Seal \$1.17 Billion Partnership to Build Advanced AI Cloud Infrastructure (<https://techafricanews.com/2025/11/10/vast-data-and-coreweave-seal-1-17-billion-partnership-to-build-advanced-ai-cloud-infrastructure/>)

What Is a Neocloud? (<https://blog.equinix.com/blog/2025/10/14/what-is-a-neocloud/>)

What is a Neocloud? - NEXTDC (<https://www.nextdc.com/blog/what-is-a-neo-cloud>)

What makes AI infrastructure different from traditional data center requirements? (<https://cologix.com/resources/infographic-and-videos/faqs-what-makes-ai-infrastructure-different-from-traditional-data-center-requirements/>)