INTELLIGENCE UPDATE

AI uncertainty: bubble trouble brewing


Andy Lawrence

30 Oct 2025

OPINION: AI uncertainty (part 2)

In a recent report (see *AI uncertainty: more adoption, more caution*), Uptime Intelligence discussed how data center owners and operators are attempting to de-risk their AI deployment strategies until customer demand and technology roadmaps are clearer. They are doing this by adopting flexible, upgradeable designs where possible, by extending the life of existing infrastructure, and by deferring major investments until they are clearly needed.

The risks these organizations are trying to manage are essentially technical and practical. Most operate within carefully allocated budgets, and they know that any significant misstep will be costly.

But not all AI investment is approached with the same diligence. Throughout 2025, Uptime Institute network members as well as others in the industry have asked us repeatedly about another risk: is there an AI bubble? And if so, how would a market correction affect the overall data center sector — including those operators that are only indirectly, or minimally, invested in AI?

At a headline level, the demand for AI-supporting infrastructure is stratospheric — far beyond what is already announced or planned — and this is set to drive strong growth for at least the next decade. Global data center power demand is forecast to reach 219 GW by 2030, at a cost of $7 trillion, according to McKinsey & Company. This would represent, Uptime Intelligence loosely estimates, a threefold increase in global data center power consumption compared with mid-2025 levels (based on an average of various forecasts).

But throughout 2025, a number of specific concerns have been mounting alongside the hype and excitement. These include the level and concentration of investment in AI-fueled tech stocks, the dizzying valuations of certain companies, and growing doubts about the viability, scale and audacity of some of the investments being made.

Many in the industry wonder whether innovation or shifting demands will render a lot of infrastructure either unnecessary, uneconomic or obsolete — some of it even before it is deployed. And in the background, the big question still nags: “How will all this be paid for?” To

that, we might add another: “When will it all be paid for?” The further out the return, the greater the financial stress on many of the players involved.

Addressing such questions is difficult and multilayered. Almost everyone consulted by Uptime Intelligence admits to uncertainty about how much and what type of AI infrastructure will be built and operated in the coming two, five and ten years.

Enterprise data center owners, investors, suppliers, colocation companies and their tenants each have their own specific concerns. Among these are volatile demand and supply of equipment, sharp price swings, unpredictable equipment supply chains, tenants running out of funds or canceling contracts, falling asset values, and the risk of being locked into inflexible or punitive power contracts.

Uncertainty and doubt

One of the biggest uncertainties about the infrastructure ramp-up concerns not just the future, but the present. How much of the announced AI infrastructure — the same projects that drive regulators, sentiment and even public perceptions and policy — is actually being built, or will be built, in the next three to five years, or even at all?

Uptime Intelligence research shows that a significant proportion of the major projects announced over the past two years have been delayed, canceled or remain uncertain — a finding that is unsurprising to industry insiders, given the projects’ technical and financial complexity, huge scale, shortages of power, equipment and expertise, and, in some cases, unproven demand (see *Hyperscale data center plans at unsustainable levels*). Many of these large projects have only provisional financing, power arrangements and planning permission.

What does this mean? In theory, it points to a coming correction: expected orders may not materialize or may be postponed, and some investments will fail or be canceled, leading to a fall in market demand and financial returns.

But these are extraordinary times. Even if a third of the projects announced to date were canceled, the remainder would still likely exceed the ability of suppliers and many power companies to meet demand comfortably. Most suppliers and established operators would still find their underlying businesses are booming, even as other investors lose money on failed or stalling projects.

Another widely cited concern is the concentration of investment, expertise and capacity within a small group of giant technology companies. The leading pioneers in AI have spent or raised enormous sums of money — and many are investing in or supporting one another, creating an appearance of circularity. These companies include Nvidia, OpenAI, Microsoft, Oracle and, most recently, AMD (though others are also participating; see **Table 1**).

Table 1 AI’s circular economy: some of the big partnerships

Date (approx)	Companies involved	Agreement	Headline
2019 (and later)	Microsoft, OpenAI	Microsoft invests $14 billion in OpenAI (deep use of each other's platform).	**$14 billion**
2023	Microsoft, CoreWeave	Microsoft to buy over $10 billion of CoreWeave AI cloud capacity.	**$10 billion**
Mar 2025	CoreWeave, OpenAI	CoreWeave provides AI cloud capacity for OpenAI (includes Google).	**$22 billion**
Jan and Sep 2025	Nvidia, OpenAI	OpenAI to deploy 10 GW of data centers with Nvidia GPUs. Nvidia to invest up to $100 billion in OpenAI.	**$100 billion**
Jun 2025	Oracle, AMD	Oracle to purchase 130,000 AMD GPUs to build data center AI infrastructure. AMD to collaborate on agentic AI.	**>$10 billion**
Jul 2025	OpenAI,Oracle	Oracle to provide 10 GW data center infrastructure to OpenAI (falls under Stargate).	**$300 billion**
Jan 2025	OpenAI, Oracle, Softbank, Nvidia, CoreWeave, MGX, G42, Nscale (and others)	Stargate: US led project to build 10 GW of data center capacity.*	**$500 billion**
Mar 2025	Blackrock, Mircosoft, MGX, Nvidia, xAI	AI Infrastructure partnership: a project to build AI capacity using xAI (similar to Stargate).	**$30-100 billion**
Oct 2025	OpenAI, AMD	OpenAI to buy up to 6 GW of AMD processors/ systems (>$90 billion). OpenAI granted up to $96 billion in equity.	**$90 billion**
Sep 2025 (and earlier)	CoreWeave, Nvidia	CoreWeave to buy up to $6 billion of processors. Nvidia to (re)purchase unsold capacity (Nvidia owns 7% CoreWeave equity).	**$6 billion**
Sep 2025	Nvidia, Intel	Nvidia buys 4% stake in Intel.	**$5 billion**
Sep 2025	Microsoft, Nebius	Microsoft to buy over $17 billion of Nebius AI cloud capacity.	**$17 billion**

** Stargate is a general announcement and incorporates other deals and projects in this table. There is therefore signficant double counting. All data is approximate and based on company announcements.*

UPTIME INSTITUTE 2025

uptime INTELLIGENCE

When companies invest in or underwrite their customers and partners, or offer buyback contracts and indemnities, it can help support shared goals, foster collaboration, provide some guaranteed demand and financial stability, and project confidence and faith in their shared markets and mission.

The way a small group of powerful companies cluster together and cross-invest is not unique — comparisons can be made with the Keiretsu in Japan and the Chaebol in South Korea. Both of these networks went on to create global industrial powerhouses in the 20th century.

More recently, however, analysts have begun to highlight the downsides: such concentration can create systemic risk, shared or single points of failure and vulnerabilities (e.g., GPU prices) and may inflate sales and profit figures, hiding real demand from customers outside the inner group.

In this AI build-out, the role of neoclouds — an estimated 150-190 companies globally that are buying GPUs and offering access to them as a service — further complicates the picture. These companies are competing with the major hyperscalers but also partnering with them (see below).

Payback time

Despite near-universal agreement that AI represents a long-term transformational advance, it is still unclear how and when all the investment (estimated by some to exceed $1 trillion by the end of 2025) will be paid back. Currently, only the equipment vendors — particularly Nvidia and its partners — are seeing strong commercial returns from their investment. Others, including Microsoft, OpenAI, Google, CoreWeave and Oracle, have invested, and continue to invest, far more than their current AI-related revenues.

For the infrastructure sector — including data centers and cloud providers — the returns will depend on two factors: large-scale consumer adoption and widespread use of AI by large enterprises. These, in turn, may repay the investment now being made in infrastructure.

It is clear that this will take time. Consumers are voraciously using AI, yet they are not buying commercial AI services at levels sufficient to recoup the investments in foundational model infrastructure. According to some sources, ChatGPT — one of the largest general-purpose LLM services — has around 800 million users, of whom only 5% are paid subscribers. OpenAI is reportedly considering introducing Google-style advertising to generate revenues.

Corporate adoption of AI, meanwhile, is also at an early stage. Most of Uptime Institute's network members, representing large corporations, say their organizations, while enthusiastic about AI adoption over time, are taking a gradual, cautious approach. They are mostly investing in pilot projects; only in a few cases are they making major infrastructure commitments, usually where there is a corporate mandate or a clear commercial benefit.

At present, enterprises (including software-as-a-service companies) are mostly relying on cloud and neocloud providers to access GPU time. However, high costs and limited availability are currently deterring stronger, full-scale uptake. Ironically, a price collapse would likely encourage more widespread adoption — but when that happens, the financial returns for the cloud and neocloud providers, while more reliable, may be lower and take longer to materialize.

Most enterprises are cautiously testing concepts and different technologies by using provisional, lower-density designs and/or cloud services. Globally, the success of models such as DeepSeek — built using older midrange Nvidia and other processors — has encouraged enterprises to seek lower-cost solutions, avoiding, for now, high-end, liquid-cooled Nvidia infrastructure.

The coming correction

There is a growing consensus that significant downward market corrections in AI and AI infrastructure are coming. However, there is an almost equally strong belief that the data center sector will weather any disruption due to the underlying strength of demand — both for data centers and for AI applications in particular. This confidence is despite the potential scale of a correction, which could reverberate across entire markets. (As of early October 2025, tech stocks accounted for more than 30% of the US S&P 500 index.)

A commonly used term is “creative destruction” — the idea that it is necessary to clear out an initial, over-capitalized group, so that new companies, technologies and models can emerge, building on the infrastructure already in place. Advocates of this view include David Solomon, chief executive of Goldman Sachs; Jamie Dimon, chief executive of JPMorgan Chase; and Jeff Bezos, founder of Amazon, who described this as a “good kind of bubble.”

Executives at companies in the cloud and data center sectors are aware of the current risks and are taking steps to manage them. The largest organizations— the hyperscalers — are best positioned to absorb valuation swings and large losses as they build their market positions over time. Even so, many are moving some of their investments off their balance sheets and into third-party infrastructure, such as neoclouds and colocation facilities, while using more debt to protect cash flow and margins.

At the recent Platform Global conference in Europe, neoclouds came under particular scrutiny, with GPU utilization identified as an important metric that their executives need to manage carefully (see *Neoclouds: a cost-effective AI infrastructure alternative*). It is widely speculated that Neoclouds will be dangerously exposed if demand falters or GPU prices fall sharply — and if they fail, it could pose problems both for the processor suppliers that support and supply them and for the colocation companies that house them (see *Neoclouds: AI’s shock absorbers*).

Meanwhile, many colocation companies consulted by Uptime Intelligence are conducting rigorous financial checks on neocloud tenants and extending contracts with hyperscalers to ensure that infrastructure investments are repaid. Colocation providers are working to avoid being left with underutilized data centers if tenants fail to use their allocated power or space. Bad debt and bankrupt tenants caused many data center operators severe difficulties during the 2001 crash, and memories of that period remain strong. Some companies are even in discussions with power providers on the best way to mitigate the risk of large penalties for failing to meet minimum usage levels.

Enterprise IT operators are perhaps the best protected, as few have yet spent heavily on AI infrastructure. Both those with and without data centers may even benefit from a cooling of the AI hype wave. GPU and cloud prices (AI) are likely to ease, some colocation space previously allocated to AI and hyperscale workloads may become available, and staff or consulting prices could level off. Equipment shortages and supply chain lead times may also improve. With less hype, enterprises will also have more time to assess whether their AI projects can deliver

tangible financial benefits.

One thing is clear: few in either the corporate IT or data center sectors believe a correction represents a hard reversal or a collapse in AI demand. Rather, it is more likely to be a realignment that leads to some restructuring in the ecosystem. This will likely cause problems for the companies and investors that have taken the biggest risks — but it will also give organizations on all sides the opportunity to deploy AI in ways that are both commercially and technically sound.

ABOUT THE AUTHOR

Andy Lawrence

Andy is a founding member and the Executive Director of Research for Uptime Institute Intelligence, which analyzes and explains trends shaping the critical infrastructure industry. He has extensive experience analyzing developments in IT, emerging technologies, data centers and infrastructure, and advising companies on technical and business strategies.

alawrence@uptimeinstitute.com

About Uptime Institute

Uptime Institute is the Global Digital Infrastructure Authority. Its Tier Standard is the IT industry's most trusted and adopted global standard for the proper design, construction, and operation of data centers – the backbone of the digital economy. For over 25 years, the company has served as the standard for data center reliability, sustainability, and efficiency, providing customers assurance that their digital infrastructure can perform at a level that is consistent with their business needs across a wide array of operating conditions.

With its data center Tier Standard & Certifications, Management & Operations reviews, broad range of related risk and performance assessments, and accredited educational curriculum completed by over 10,000 data center professionals, Uptime Institute has helped thousands of companies, in over 100 countries to optimize critical IT assets while managing costs, resources, and efficiency.