Datacenters 2027: The Year of Acceleration, +50GW in a Single Year

Datacenters 2027: The Year of Acceleration, +50GW in a Single Year 60GW Under Construction, SemiAnalysis Leading Indicators, 1.3T of Lease Contracts Signed YTD, The Shortage Could Move to Chips

OCT 19 https://newsletter.semianalysis.com/p/datacenters-2027-the-year-of-acceleration

View in browser

Datacenters 2027: The Year of Acceleration, +50GW in a Single Year

60GW Under Construction, SemiAnalysis Leading Indicators, 1.3T of Lease Contracts Signed YTD, The Shortage Could Move to Chips

QQ:1413221435

JEREMIE ELIAHOU ONTIVEROS, REYK KNUHTSEN, AND DYLAN PATEL

OCT 18 • PAID

READ IN APP

Not a week goes without another piece of datacenter FUD. We've already debunked depth why moratoriums aren't as impactful as one may think. Despite that, many still misunderstand the pace of the global AI datacenter buildout. Just the other week, Brad Gerstner from Altimeter showed our data on the All-In Podcast and stated that "SemiAnalysis' forecast of +43GW in 2027 is too aggressive" and that he thought only about 25GW would get built, close to half of our number, based on his "hunch". But let's not operate on hunches.

It's very tempting to cede to the FUD given all the constraints on the datacenter buildout: power, labor, supply chains, capital, and many more. But at SemiAnalysis, we care about the truth, not about narratives. And here's the truth for you: one picture, >24GW of datacenters in advanced stage of construction, all to be delivered in 2027. Most of them are more than halfway done, with exterior cooling and electrical equipment already installed. That's just 40 campuses, a tiny share of the

QQ:1413221435

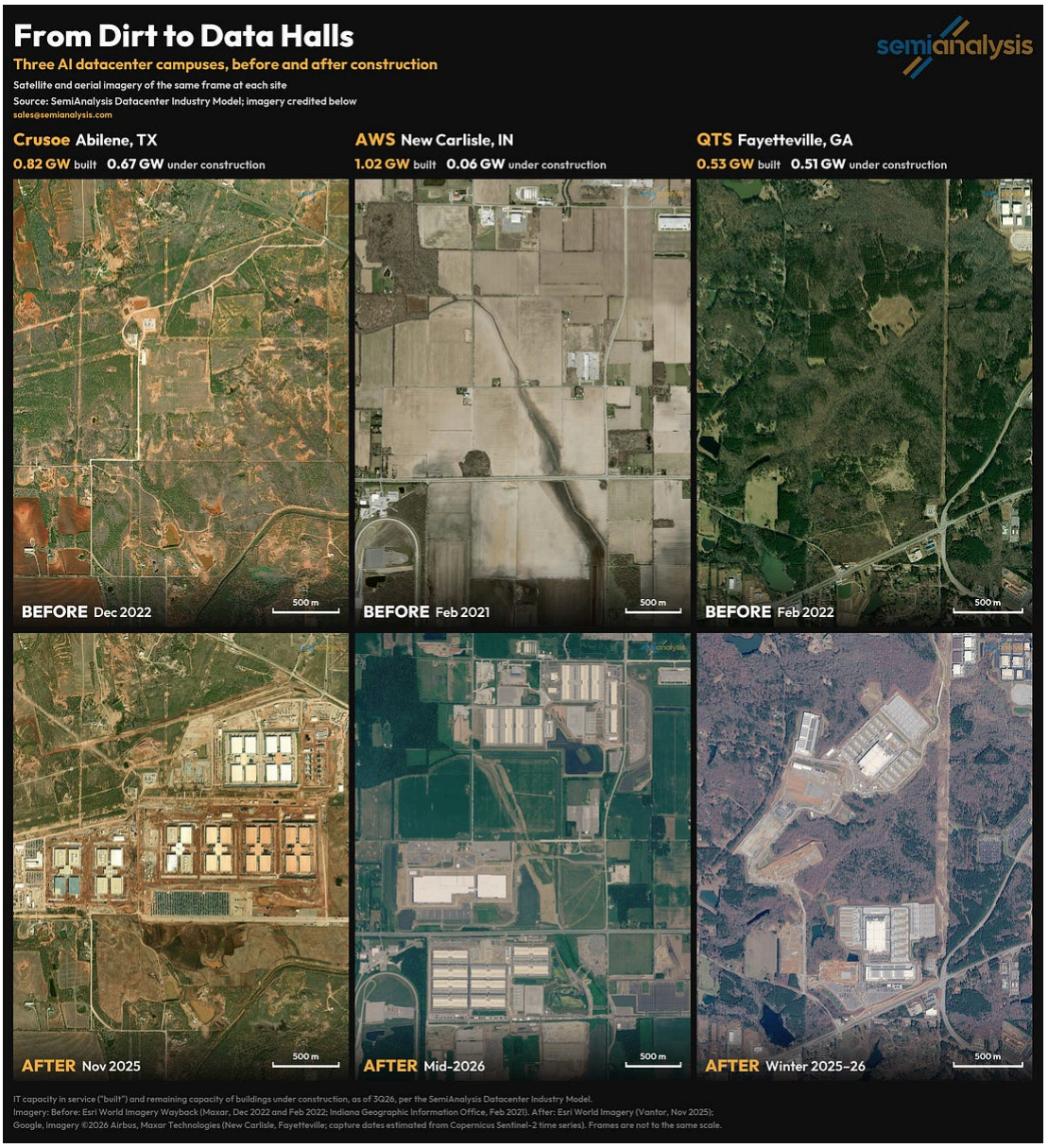
multiple hundred sites in advanced stage of construction. We track them all, and for the full granularity, use our Datacenter Industry Model.



Source: SemiAnalysis Datacenter Industry Model

Of course, there are plenty of challenges and shortages. Behind these pictures, there are labor issues, supply chain delays, fundraising challenges, among many others. We've written in-depth about Modular Datacenters and Labor, Datacenter Electrical systems, Cooling systems, Behind-The-Meter power, the US Grid Constraints, and Nvidia's ecosystem financial support. But the fact is simple: gigantic sites are simply getting built, and many are on time. For example, just go look at what Crusoe built in Abilene, what AWS built in Indiana, or QTS/Microsoft's giant site in Georgia. These are the sites that have produced SOTA technology like GPT-6 and Fable 5.5. Looking ahead, despite all the constraints, an accelerating number of giant sites keep getting built, as shown by our Top 40 picture

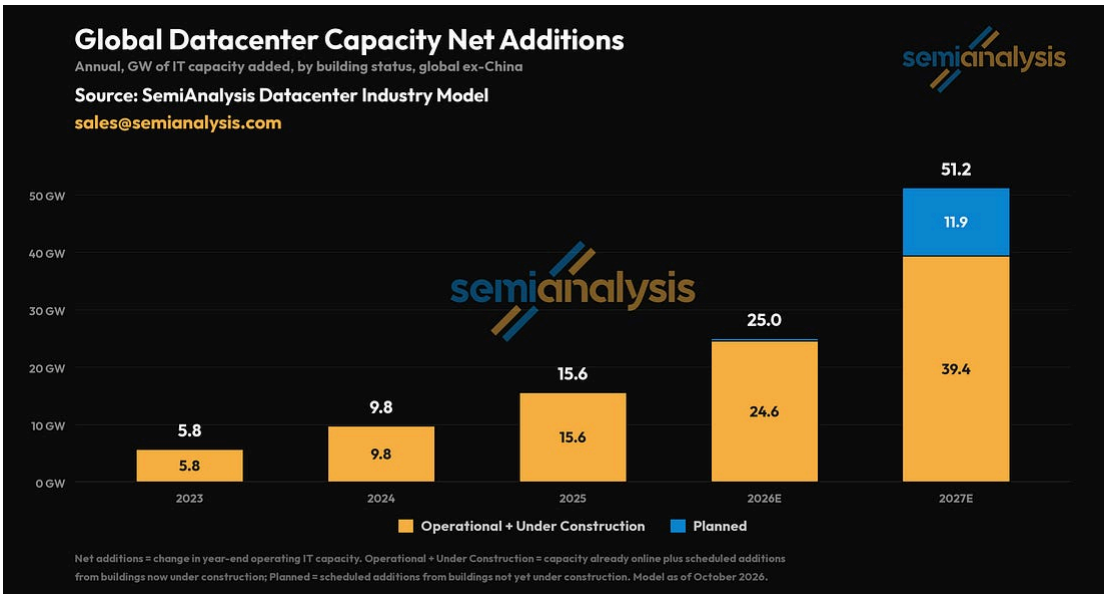
QQ:1413221435
QQ:1413221435



Source: SemiAnalysis Datacenter Industry Model

If you ask any datacenter builder, they'll tell you how hard it is, how some workers aren't showing up, how vendors are delayed, and many other issues. All of this is accurate. And that's why we've successfully called out countless datacenter delays and keep calling out developer timelines that appear overly aggressive in our Datacenter model, including the Oracle New Mexico Force Majure site delay in April. To be fair, +50GW is essentially the global cumulative datacenter market up through 2024 being built in one year! But our Datacenter Industry Model adjusts for all of that and still points to a material acceleration in datacenter capacity in 2027. So, what are we seeing that others can't?

QQ:1413221435
QQ:1413221435



Source: SemiAnalysis Datacenter Industry Model

SemiAnalysis is a reader-supported publication.
To receive new posts and support our work,
consider becoming a subscriber.

Subscribed

Let’s dig in. We’ll explore the following topics in this report:

- SemiAnalysis proprietary leading indicators: we’ll show a sneak peak of our proprietary data supporting our forecast, such as self-build construction starts and datacenter leasing (in MWs and USD). Year-to-date, **slightly under USD 1T of binding / non-cancellable datacenter leases have been signed** – that’s key to supporting our aggressive 2027-28 forecast.
- The shape of the buildout: how can we support +50GW in a single year? Where, what design, what power solutions?
- The “real” question: are we going to face a datacenter oversupply? As we explained in The Great Silicon Shortage, datacenters & power may be the tip of the iceberg, with memory and logic possibly becoming the new #1 constraint in 2027-28.

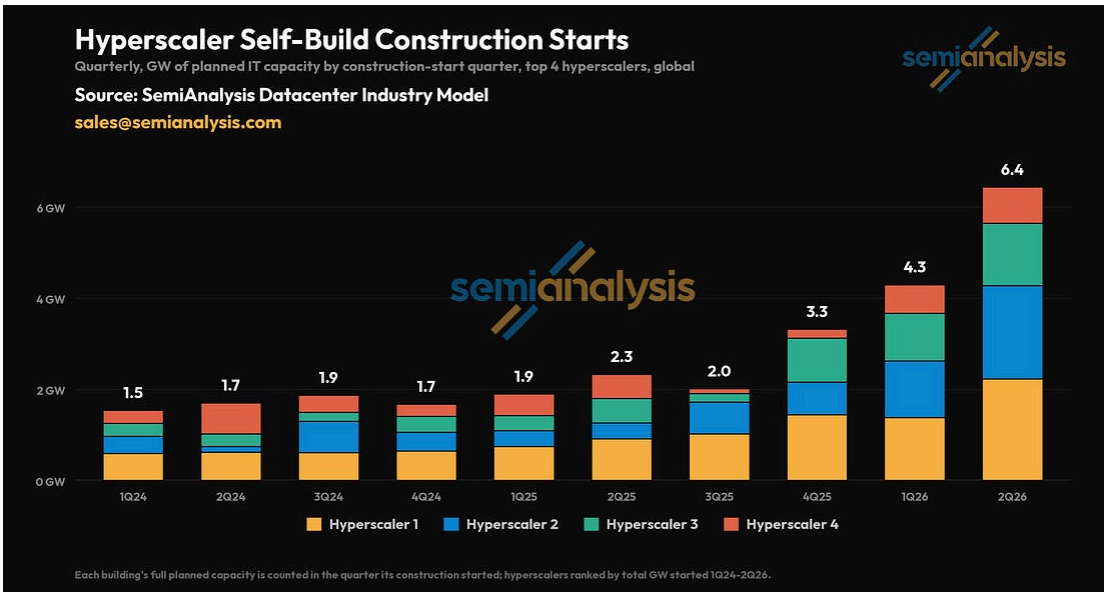
Overall, we are very excited about 2027. The natural implication of the acceleration will be significantly more compute supply. That means cheaper and smarter AI, faster pace of AI model progress, and likely material revenue acceleration for leading AI labs and other token vendors. That also means unthinkable capex numbers. Reflexively rates will have to shoot up to enable the buildouts from a capital demand perspective. Let’s dive in!

QQ:1413221435

SemiAnalysis Datacenter Leading Indicators: how we see the buildout before everyone else

Today we want to dive into two crucial indicators that give us strong confidence on our 2027 forecast. The first one is quarterly self-build construction starts. To accurately forecast US hyperscalers, we track the exact construction evolution of every single building. And the conclusion is simple: US hyperscalers have never launched more construction of datacenters. Each quarter, we track how many MWs start construction, and that number is going parabolic, driven in particular by two hyperscalers. If you want to know who or track the granularity of our data, subscribe to our Datacenter Industry Model.

QQ:1413221435



Source: SemiAnalysis Datacenter Industry Model

The second indicator is Datacenter Leasing. We've talked about this a lot in prior reports, but as a quick recap: given the capital-intensity of building multi hundred MW datacenters, developers typically only spend 5-10% of project capital without a customer. They need a "pre-lease" to fund and develop the whole project. That's typically a 15-20 year contract, with a TCV of $30-50B per GW. Check out our HPC Lease Tracker as part of the Datacenter Model to dive into actual contract economics.

Subscribed

QQ:1413221435

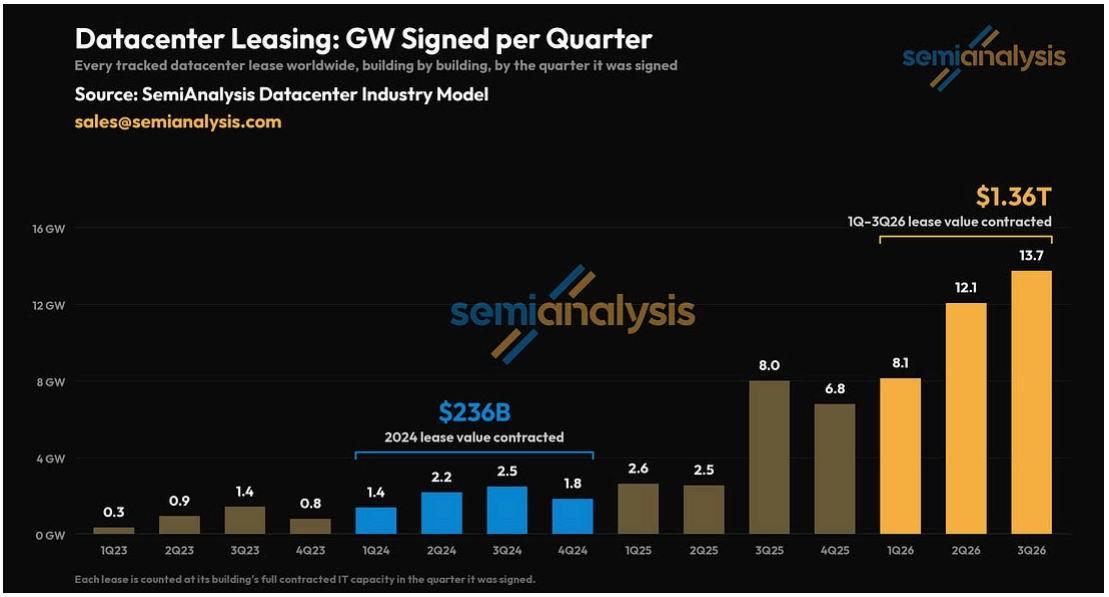
As such, tracking these lease contracts is a key leading indicator, as it leads debt raises, which itself leads the actual buildout. A big reason why 2026 datacenter

growth isn't "that" high relative to 2025, is that quarterly datacenter leasing remained relatively stable between H2 2023 and H1 2025. Of course, the key reason was the Microsoft Datacenter Freeze, which we called out before anyone else in December 2024 for Datacenter model subscribers: the Redmond giant single-handedly drove the leasing market in 2023 and 2024, then paused, and others could only pick up what was left behind.

Since H2 2025, things have dramatically changed. Oracle kicked off the era of gigantic leases, after closing on ~4.5GW of datacenter IT capacity across three sites between August and October 2025. Everyone else followed, and the pool of buyers has never been broader, with Meta, AWS, Microsoft, Google, Nvidia, AMD, Anthropic and OpenAI all being in the market for GW-scale leases. There's also a long tail of Neoclouds signing hundred-MW-scale deals.

In USD terms, this means an unprecedented amount of lease value that builders can use to raise debt. **Just in 2026 year-to-date, we've identified over a trillion USD of lease value that was signed and closed**. That trillion is supporting a gigantic datacenter buildout in 2027-28. Our new Compute, Capital & Markets Model is now tracking all the debt raises and can confirm that most of this debt has been raised.

QQ:1413221435



Source: SemiAnalysis Datacenter Industry Model

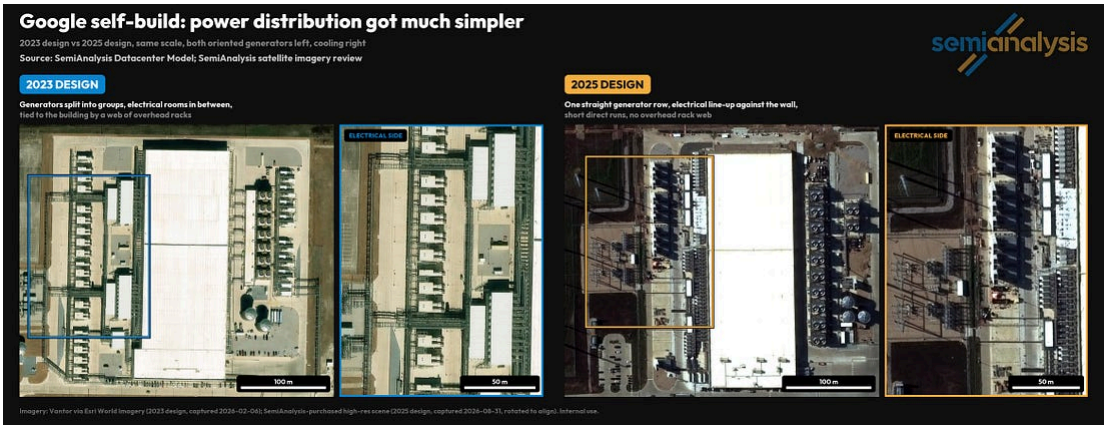
The shape of the buildout: how to build 50GW in a year

QQ:1413221435

Tracking all of these leading indicators has enabled us to forecast, before anyone else, the upcoming surge in orders for many industrial suppliers like Vertiv, ABB, Sterling Infrastructure. Check out our Industrials Model for in-depth coverage. However, the

challenge is now going from orders to delivery. And resources have obviously become constrained, as power, labor, and sheer regulatory overhead are enormous burdens for developers. But the industry has promptly reacted with innovation.

The first step is to simplify datacenter designs. We show below how Google has modularized their electrical architecture, removing a significant amount of outdoors cabling that requires onsite assembly, replacing them with prefabricated electrical modules. We track all the equipment changes (including suppliers) for datacenter designs and forecast orders in our Industrials Model



Source: SemiAnalysis Industrials Model

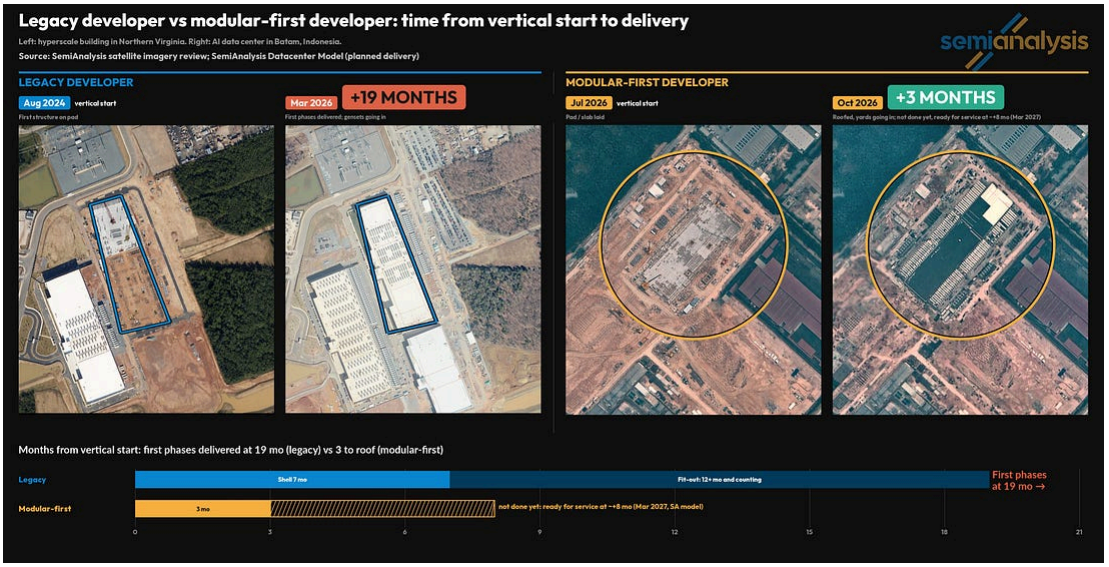
Others have opted for more radical design changes. One example is Meta with its tent design, that we exclusively unveiled over a year ago. Another one is AWS which has materially simplified their shell design and launched a new rapid design dubbed “SAMDC”. Data halls are also increasingly prefabricated, an effort called Project Houdini at AWS. Construction timelines are now under 10 months. Hence our “LEGO Datacenters” title for the modular movement, you literally slot them in and go, and doing so allows for MEP labor (7-11 months of work) to drop up to 63%! We will have a full deep dive on the labor of the 2027 buildout coming soon.

QQ:1413221435
QQ:1413221435



Source: SemiAnalysis Industrials Model

In addition to hyperscale selfbuild, datacenter developers promoting prefabricated designs have also been remarkably successful. DayOne is likely the best example. With a total of ~3GW leased since their first deal in early 2023, the Singaporean company has been a rocket ship in terms of construction speed. While traditional developers have >18 months build timelines, DayOne cut this to ~9 months, which was key to signing huge deals with Firmus, OpenAI and Nvidia.

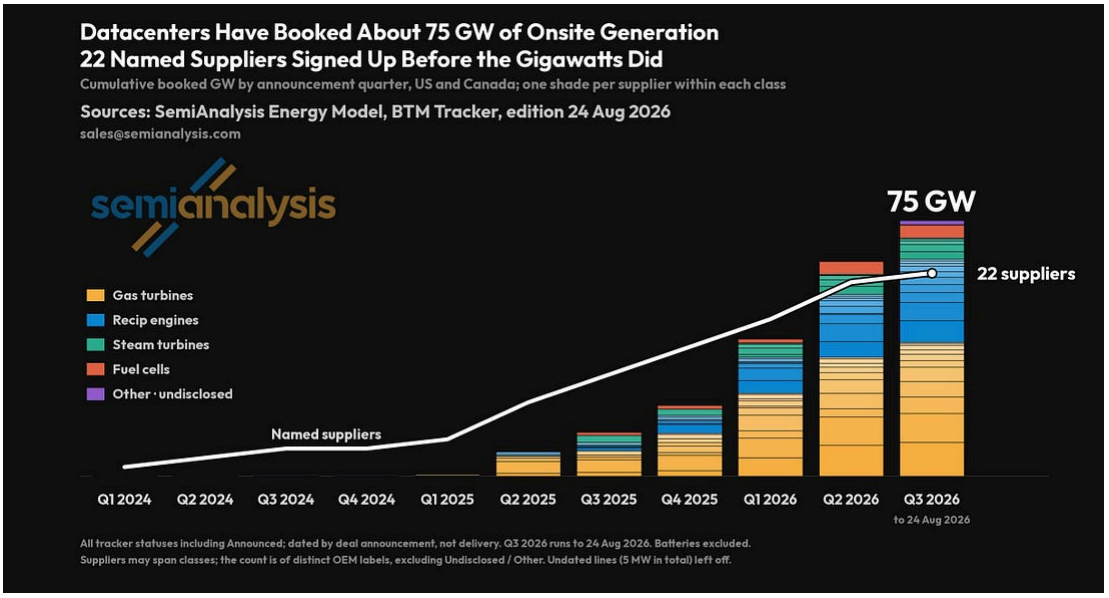


Source: SemiAnalysis Datacenter Industry Model

Subscribed

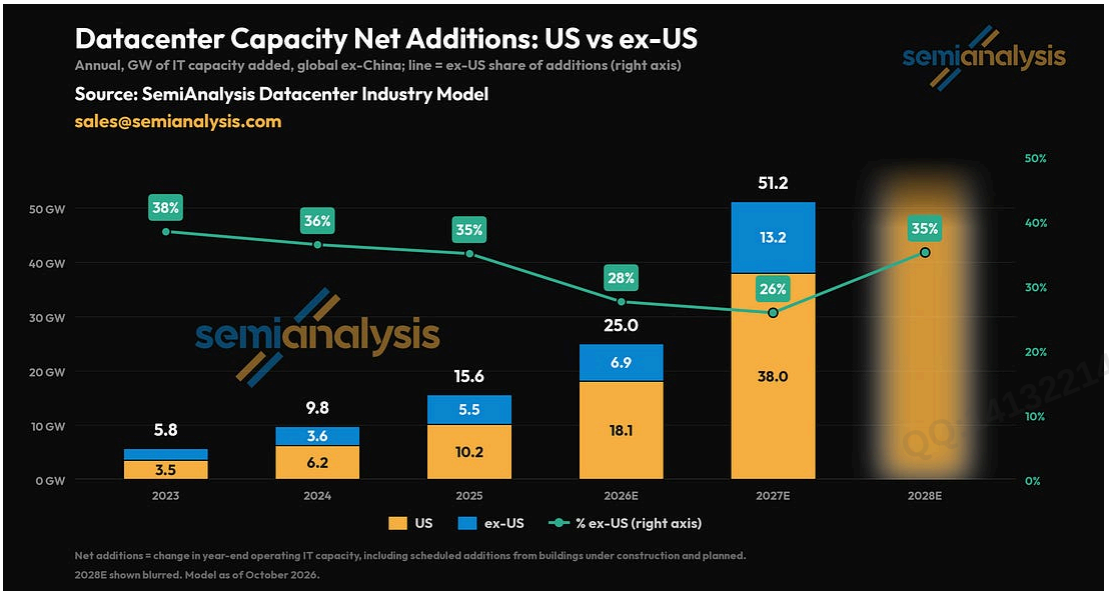
QQ:1413221435

Another core issue is power, and again we think the issue is largely solved. We've been the first to talk about all issues on power, since March 2024 and subsequently given the grid headroom dips negative at the end of 2027. This ended up driving AI Labs and hyperscalers to build their own power plants, better known as Behind-the-meter generation, bypassing 5-10year grid interconnection timelines. And while turbine incumbents like GEV quickly faced production bottlenecks, the industry quickly reacted and we've since seen a surge in new OEMs and new solutions supplying the market. From fuel cells to truck engines to refurbished turbines, the power generation problem has broadly been solved.



Source: SemiAnalysis Energy Model

In addition, the buildout is now geographically diversifying. This naturally bypasses US labor, power, and (in some countries) permitting constraints. And as shown below, the ex-US buildout is sharply accelerating in 2027, and we expect it to gain a lot of market share in 2028, largely driven by US challenges.

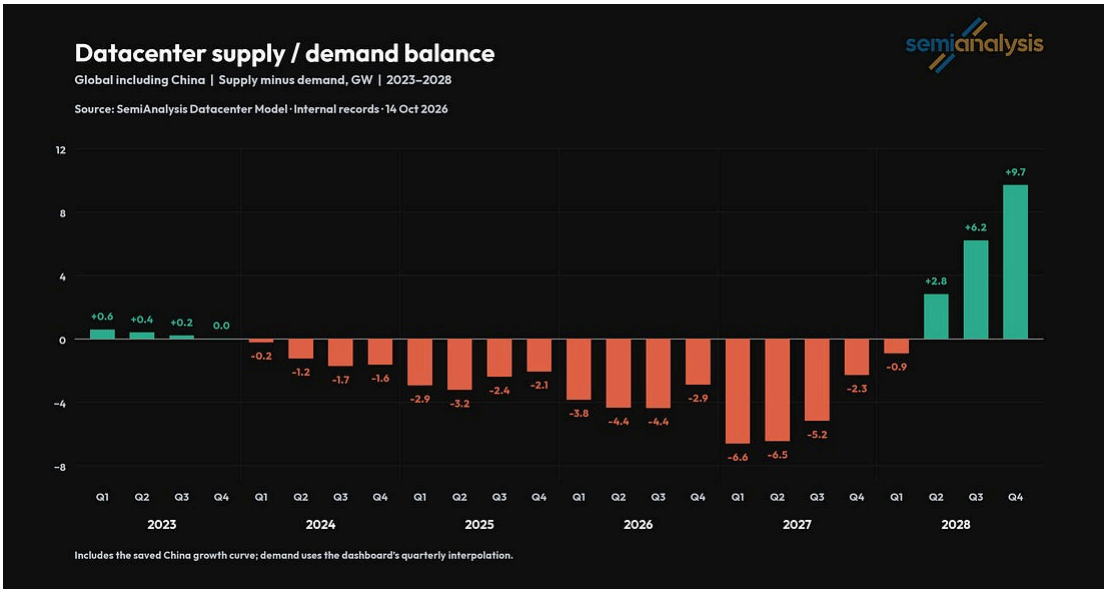


Source: SemiAnalysis Datacenter Industry Model

The Real Question: Will We Have an Oversupply of Datacenters?

While everyone is worried about datacenter supply constraints driving a temporary global AI slowdown, we are more worried about oversupply of datacenters relative to chips. In 2028, our Accelerator Model team puts up their potential chip supply against our bottoms-up datacenter supply count. We see a risk of a ~10GW datacenter oversupply by year end 2028. That could mean a slowdown in the overall datacenter & power supply chains, which we would expect to materialize towards the end of 2027 or early 2028 (in terms of new leases / new orders) as silicon constraints start becoming more obvious. But the situation is dynamic and there may be workarounds.

QQ:1413221435



Source: SemiAnalysis Datacenter Industry Model

What does this mean? There could temporarily be empty shells as companies build or lease datacenters but can't procure the XPU servers. This happened for a few months at the middle of 2025, as supply chains couldn't produce GB200 racks fast enough. But this one may be more structural, as we explained in The Great AI Silicon Shortage.

How is this possible? Many new entrants have arrived, and their proposed massive capacities begin to bear fruit in 2028. Like the Cryptominers bringing GWs, neoclouds now self-building, and new colocation providers bringing multi-100MW or GW-scale projects to the market. Think of this as a pendulum swing in the

QQ:1413221435

opposite direction from 2024-2025, where datacenter capacity was restrictive, and tales of Oracle warehousing chips were spread like wildfire.

This broadly applies to the datacenter infrastructure sector, like the BtM solutions, modular construction solutions, even powered land folks providing sites pre-permitted with grid connections. All of these folks enable the rapid buildout, including the new datacenter provider entrants, the means to bring these projects alive.

Now, everyone trying to answer this bottleneck arrives, and it may be too much.

Will There Be Enough GPUs?

As we laid out in The Great AI Silicon Shortage, we are firmly in the “silicon shortage” phase of the AI buildout. The two binding constraints are memory (DRAM, and specifically HBM) and leading-edge logic (TSMC N3). CoWoS, the bottleneck of 2023–24, is now “tight but easing.” TSMC is deliberately sizing packaging to front-end wafer supply rather than overbuilding it.

Let’s begin with memory. HBM takes roughly 3x the wafer area of commodity DRAM per bit, and that grows to nearly 4x with HBM4 and more with HBM4E. Our model has HBM bit supply growing only by +27% in 2027. No new fabs contribute until 2028. Doubling GPU output would require ~1.2 exabytes of extra HBM. That’s ~22% of the entire 2027 market and more than Micron’s whole 2027 output. It would also need ~0.8M more CoWoS wafers, about Broadcom, AMD and AWS’s combined 2027 allocation.

The answer so far has been to despec memory, like Rubin Ultra shipping with 192GB of HBM per GPU instead of 288GB on Rubin and B300. This is far from the initial 1TB, and may be even further with the 4-hi push from AI Labs. This is for two reasons really:

- There aren’t enough wafers for the 12-hi cubes needed to match the N3 logic and CoWoS

- 4-Hi resolves the “stranded” memory capacity issue in which capacity is bottlenecked by the static, limited bandwidth of the total stack (regardless of height)

A future article will dive deeper on the amount of capacity available when despec-ing memory vs logic supply, but for now, this is a source of upside in GPU shipments that is not fully digested yet in our supply-demand analysis.

QQ:1413221435
QQ:1413221435

Like Comment Restack

© 2026 Dylan Patel
548 Market Street PMB 72296, San Francisco, CA 94104
Unsubscribe

Start your Substack

QQ:1413221435

QQ:1413221435