

Another capacity crunch is coming

Is the AI boom masking the looming challenge of 2030s coal retirements?

— MAEVE ALLSUP | JANUARY 26, 2026



Photo Credit: engel.ac / Shutterstock

The U.S. power sector is in an all-out sprint to modernize power markets, state and federal regulations, and permitting and interconnection processes to get data centers online faster. We're in a capacity crunch: We can't build new generation — or the poles and wires that make up the grid itself — fast enough to satiate the energy demands of the AI boom.

But there's a second capacity crunch right around the corner, one currently being masked by the data center rush — and exacerbated by the resulting delays in coal plant retirements.

Most of the U.S. generating coal fleet was built between 40 and 60 years ago. In the last decade, the country has retired more than a third of that fleet, causing capacity to drop from 300 gigawatts to roughly 180 GW, according to data from the International Energy Agency. Most of those plants have come offline after around 55 years. That's largely because it gets more expensive to maintain old equipment, explained Jesse Noffsinger, a partner at McKinsey & Company, and as other forms of generation become cheaper, it becomes less economical to do so.

But in the last few years, the amount of old coal infrastructure coming off the grid has dropped dramatically, as grid operators, states, and utilities grapple with reliability challenges; they are increasingly willing to pay high premiums to keep aging, less-reliable plants online to avoid blackouts.

In 2025, the Trump administration entered the fray, leveraging federal overrides to force aging plants to stay online despite economic or environmental pressure to close.

IEA data shows 15 GW of coal capacity was slated to retire in 2025, but just 6.2 GW actually did. Coal demand, after decreasing by roughly 6% annually for the last 15 years, rose by around 8% in 2025, thanks both to the retirement slowdown and to higher natural gas prices.

The result of all of this is that by the early 2030s, the U.S. will be leaning harder than ever on an aging coal fleet, having deferred a wave of retirements, without having done enough to build replacement capacity. The country, Noffsinger said, has “dug into a bit of a backlog.” By 2035, as much as 90 GW of capacity will be well past its typical 55-year lifespan. And that capacity, Noffsinger added, will eventually have no choice but to retire.

The capacity debt will be exacerbated in the 2030s because the U.S. grid will be in a “valley” between inevitable but delayed retirements, and the time when next-generation resources like advanced nuclear and geothermal will be ready at commercial scale. And all the while, load will continue to grow and severe weather events will be increasingly frequent and damaging.

“There’s this period where there’s a lot of challenge around how to meet reliability and affordability without choking off economic development,” Noffsinger said, in reference to the fact that winning the AI race will depend largely on what gets built in the next several years — a timeline that could imperil Big Tech’s 2030 clean energy commitments. “From here to there, it seems like the only way to keep the grid stable and ensure that we have resource adequacy is to make decisions that could fly in the face of some of those commitments.”

Without some substantive changes, there’s going to be “a lot of friction and frustration” entering the 2030s, he added, “the outcome of which could be much higher power prices, coupled with more outage events, with more emissions. And if you think about it, that’s the worst possible outcome. But it’s not driving conversations.”

Are we Over-Indexing on the Data Center Problem?

While the data center buildout is monopolizing headlines about grid challenges, it’s not the case that the AI boom is creating a capacity debt that wouldn’t otherwise have existed, said Brent Nelson, managing director of markets and strategy at Ascend Analytics. Instead, it’s acting as an accelerant on a problem that was already looming.

“At its core, what we’re trying to do is get some new supply online to meet rising demand, and even though it’s happening faster now, it was always a problem that we were going to have to solve,” Nelson explained. “We’d still see a crunch in capacity markets...whether it’s a data center or a manufacturing facility or an EV charging network.”

The fundamental issue is that U.S. markets have long struggled with balancing retirements and financing new plants. In PJM, for example, current capacity structures have failed to generate enough stable revenue for developers to justify the decade-long investment of a new power plant. The result is a holding pattern in which utilities are clinging to aging assets because there isn’t a clear path to replacing them.

Some are pushing to solve the problem of data center load by walling it off — whether by leveraging behind-the-meter solutions or implementing separate auctions, for example. But Nelson thinks that won’t necessarily fix the underlying structural problems. The risk, he suggested, is that building a parallel track for hyperscalers becomes a political release valve, quarantining data center demand without fixing the core market failures that make it hard to finance new capacity in the first place.

It’s all well and good for hyperscalers to foot the bill for speed to power, Nelson added. “Where I have some concern is that everybody’s talking about how we can protect ratepayers from data centers...and anybody that has that conversation without thinking about how the market is going to work with load growth from something else is missing the underlying challenges.”

In other words, even “shunting off” data centers doesn’t change the reality that the market presently lacks a structure to incentivize new supply for organic load growth. That said, major market reforms are likely coming.

PJM, for example, has become the test bed for protecting the grid from data center demand: Just last week, the Trump administration and a bipartisan coalition of 13 governors released a “statement of principles” calling for an emergency backstop procurement auction. Under this plan, new generation would be paid for directly by data centers through 15-year contracts — a proposal the PJM board echoed in its own filings the following day.

Preparing for 2035

The massive question hanging over all of this, of course, is what data center load will look like in that “valley” of the mid-2030s.

“What happens after 2030 is a lot harder to predict,” Noffsinger explained. It depends on things that evolve quickly, like new GPUs and new cooling equipment coming to market. Given the rapid development timelines for those technologies, 2035 is several tech cycles away from now.

That makes it tricky to finance and build new generation, thanks to the massive “error bar” in AI demand predictions, he added. The result is a grid stuck clinging to 60-year-old coal plants, not knowing what’s on the other side of 2030.

To navigate the uncertainty, Noffsinger points to needed changes in system planning, including an approach that treats demand as a flexible resource. The bulk of new clean generation will still be coming from wind and solar by that time, and the key to a stable grid that relies on variable resources is system flexibility. That includes rapid deployment of battery energy storage for short-term arbitration and demand-side flexibility, to shift load away from peak periods.

That doesn’t strictly mean data centers, despite the current hype around flexible compute. But data centers do offer a unique advantage, Noffsinger said: They consolidate massive amounts of power into single interconnection points, potentially making them easier to integrate into demand response programs than millions of individual households.

And then, of course, we need an accelerated build-out of transmission infrastructure, he added. The tension between state and federal authority regarding transmission has contributed to policymakers’ enduring failure to pass permitting reform, as has the question of whether the customers who ultimately pay for new transmission actually benefit from it. The Senate is currently gearing up to take on the challenge yet again, spurred by the impending deadline of midterm elections.

In other words, Noffsinger said, “there’s not a silver bullet” for solving the next capacity crunch — but there’s potential for data centers to help fill the mid-2030s gap, rather than merely make it bigger.