

Inside the Data Centers That Train A.I. and Drain the Electrical Grid

A data center, which can use as much electricity as Philadelphia, is the new American factory, creating the future and propping up the economy. How long can this last?

By [Stephen Witt](#) October 27, 2025

"I do guess that a lot of the world gets covered in data centers," Sam Altman, the C.E.O. of OpenAI, has said. Illustration by Jun Cen

Drive in almost any direction from almost any American city, and soon enough you'll arrive at a data center—a giant white box rising from graded earth, flanked by generators and fenced like a prison yard. Data centers for artificial intelligence are the new American factory. Packed with computing equipment, they absorb information and emit A.I. Since the launch of ChatGPT, in 2022, they have begun to multiply at an astonishing rate. "I do guess that a lot of the world gets covered in data centers over time," Sam Altman, the C.E.O. of OpenAI, recently said.

The leading independent operator of A.I. data centers in the United States is CoreWeave, which was founded eight years ago, as a casual experiment. In 2017, traders at a middling New York hedge fund decided to begin mining cryptocurrency, which they used as the entry fee for their fantasy-football league. To mine the crypto, they bought a graphics-processing unit, a powerful microchip made by the company Nvidia. The G.P.U. was marketed to video gamers, but Nvidia offered software that turned it into a low-budget supercomputer. "It was so successful, from a return-of-capital perspective, that we started scaling it," Brian Venturo, one of CoreWeave's co-founders, told me. "If you make your money back in, like, five days, you want to do that a lot."

Within a year, the traders had quit the hedge-fund business and bought several thousand G.P.U.s, which they ran from Venturo's grandfather's garage, in New Jersey. After the cryptocurrency market crashed in 2018, CoreWeave acquired more microchips from insolvent miners. Before long, the firm had built a platform that allowed outside customers to access the G.P.U.s. Then, in 2022, Venturo came upon Stable Diffusion, an image-generation A.I. He fed the A.I. descriptions of different scenes, and it returned accurate, beautiful illustrations. "This is going to enrapture the entire world," Venturo remembers thinking.

Get our Classics newsletter to discover timeless gems from

Stable Diffusion had been trained on Nvidia equipment that was similar to CoreWeave's. Venturo and his co-founders sensed the business opportunity of a lifetime. CoreWeave raised a hundred million dollars, and used almost all of it to buy Nvidia hardware. Soon, Jensen Huang, Nvidia's C.E.O., arranged a meeting with the group. "He spent about ten minutes making fun of me for being from New Jersey," Venturo said. But in time Nvidia bought a portion of the company. By the middle of 2022, CoreWeave was running a new kind of business, connecting A.I. developers with warehouses full of Nvidia equipment.

Modern data-center construction began in the nineties, with the arrival of the commercial internet. Data centers hosted websites, coordinated e-mail, processed payments, and streamed video and music. Amazon was particularly aggressive about building data centers—so many were constructed in Loudoun County, Virginia, that the area became known as Data Center Alley. Even before the A.I. boom, data centers were profitable; in some years, Amazon's web-services division earned more than the company's retail operation, on a fraction of the sales.

But the arrival of Nvidia's G.P.U.s and the onset of large-scale A.I. training transformed the data-center business. ChatGPT launched in November, 2022, and exploded in

popularity. "The world goes bananas," Venturo said.

Microsoft partnered with OpenAI to provide the data-center capacity that ChatGPT needed to function. When Microsoft couldn't keep up with the demand, it turned to CoreWeave.

Working with Nvidia hardware has become a status symbol—a sign that one is serious about A.I. Talking with engineers about the equipment, I was reminded of the time I saw a snaking line of young men standing in the cold to buy sneakers from the streetwear brand Supreme.

Earlier this year, CoreWeave went public. Venturo and his co-founders are now billionaires. The company owns several hundred thousand G.P.U.s, and its platform trains models for Meta and other leading labs, in addition to OpenAI.

This summer, I visited a CoreWeave facility on the outskirts of Las Vegas. The building, a large warehouse, was surrounded by a thick fence and dotted at regular intervals with security cameras. I went through a turnstile, where I was greeted by a security guard wearing a bulletproof vest and a holstered Taser. After surrendering my phone, I took two lime-green earplugs from a dispenser and entered the facility.

I was joined by three CoreWeave engineers, geeks who had adapted to hyper-scale capitalism as Darwin's finches had to the Galápagos Islands. Jacob Yundt, from corporate, was

lean and eloquent, with a swooping part in his hair. Christopher Conley, an enthusiastic explainer with sunglasses and a beard, oversaw the hardware. Sean Anderson, a seven-foot-tall former college-basketball center, wore a shirt that read "*MOAR NODES.*"

The nodes in question were shallow trays of computing equipment, each weighing around seventy pounds and holding four water-cooled G.P.U.s along with an array of additional gear. Eighteen of these trays are stacked, then connected with cables to a control unit, to form the Nvidia GB300 computing rack, which is a little taller than a refrigerator and costs a few million dollars. In a busy year, a typical rack will use more electricity than a hundred homes. Dozens of them stretched into the distance.

CoreWeave keeps its racks in white metal cabinets, to help them stay cool and to dampen noise. Conley unlatched a door to show me a rack in action, and I was buffeted with air. The noise was unholy, as if I'd opened a broom closet and found an active jet engine inside. I watched the blinking lights and the spinning of the fans. "Tinnitus is an occupational hazard," Conley shouted at me.

I looked around. There were hundreds of identical cabinets in the facility. Above us was a metal catwalk, lined with power distributors for the computing equipment. I thought of monks in cloisters, soldiers in barracks, prisoners in cells.

What type of person voluntarily worked in such a place, I wondered. "I was told by H.R. that I can't ask this kind of question anymore, but I like to hire people that can endure a lot of pain," Yundt later said. "Endurance athletes, that sort of thing."

CoreWeave wouldn't tell me which customer was using its technology that day, although Yundt suggested that the training run we were witnessing was a modest one. He began to detail the configuration of the rack. Unable to hear what he was saying, I nodded sagely, as if in a conversation at a night club. Even with the plugs in, my ears were starting to ring, and I was developing a headache. Yundt turned to me. "Sometimes a customer will tie up this entire place for weeks at a time," he shouted. His parted hair began to flap in the fan exhaust. "We call those 'hero runs.' "

CoreWeave's hardware can train an A.I. from scratch to completion. Software developers, typically at a workstation in Silicon Valley, upload to the data center a file of numbers known as "weights" and a vast array of training data, which might be text or images or medical records or, really, anything at all. In their initial configuration, the weights are random, and the A.I. has no capabilities.

The A.I. is then exposed to a slice of the training data, and asked to offer a prediction about what should ensue—the next few letters in a sentence, say. An untrained A.I. will

invariably get this prediction wrong, but at least it will learn what *not* to do. The weights must be modified to absorb this new piece of information. The math is unwieldy, and is especially dependent on an operation known as matrix multiplication.

“Beauty is the first test: there is no permanent place in the world for ugly mathematics,” the mathematician G. H. Hardy wrote, in 1940. But matrix multiplication, to which our civilization is now devoting so many of its marginal resources, has all the elegance of a man hammering a nail into a board. It is possessed of neither beauty nor symmetry: in fact, in matrix multiplication, a times b is not the same as b times a . As the matrices increase in size, the arithmetic requires great computational power to solve. The latest large language models can involve about a trillion individual weights. A weeks-long hero run for such a model can use tens of thousands of G.P.U.s and require ten trillion trillion operations, which is more than the number of observable stars in the universe.

Data centers must cooperate with local electric utilities to manage these training runs. The water coursing above CoreWeave’s microchips enters at room temperature but leaves warmer than a hot bath. It is cooled in a storage tank before being recycled into the system. The temperature, humidity, and particulate count of the air inside the room are

also carefully monitored. "Condensation is our enemy," Conley said, gravely.

All these microchips, all this electricity, all these fans, all this money, all these data, all these water-cooling pumps and cables—all of it is there to tune the weights, this little file of numbers, which is small enough to fit on an external hard drive. A great deal depends on this well-tempered collection of synthetic neurons. The money spent to develop it, and others like it, represents one of the largest deployments of capital in human history.

When the finished product is ready, clones of the weights are distributed to data centers around the country, where they can be accessed through the internet, a process known as "inference." Users ask questions, prompting the A.I. to produce individual units of intelligence called "tokens." A token might be a small square of pixels or a fragment of a word. To write a college term paper, an A.I. might produce about five thousand tokens, consuming enough electricity to run a microwave oven at full power for about three minutes. As A.I. fields increasingly complex requests—for video, for audio, for therapy—the need for computing power will increase many times over.

Multiply that by the more than eight hundred million people who use ChatGPT every week, and the data-center explosion makes sense. ChatGPT is now more popular than

Wikipedia; young people refer to it as “Chat,” which has come to signify A.I. in the way that “Google” signifies internet search. I spoke to a data-center executive with Microsoft who thinks that we will demand A.I. constantly in the future, just as we demand the internet or electricity, and that current data-center construction may be insufficient. “I am more concerned that we are underbuilding rather than overbuilding,” the executive said.

Microsoft is one of the dominant operators of data centers, and this business has become the primary driver of growth for the American economy. Although the company still makes operating systems and office software, it is investors' excitement about data centers that has propelled Microsoft to around a four-trillion-dollar valuation, making it the world's second most valuable firm. Nvidia, which makes microchips that Microsoft uses, is No. 1.

“I'm absolutely listening to you—my cellphone is turned over!”

Cartoon by Jeremy Nguyen

It is difficult to get inside a Microsoft data center, for the same reason that it is difficult to get inside Fort Knox. The A.I.s under development in these facilities are worth a fortune. “Traditionally, when you want to steal something that's worth an insane amount of money, it's, like, ‘Back up the truck,’ ” Peter Salanki, the chief technology officer of CoreWeave, told me. “Here, you have somebody get in with

a thumb drive, and it can fit the entirety of OpenAI's I.P. on it."

But this fall, after what felt like two hundred phone calls, I was invited to tour an enormous Microsoft data-center campus under construction. I agreed to take no photographs, to leave my phone outside, to limit what I described of the interior, and not to reveal where in the United States the facility was situated. In September, I took a long drive to the middle of nowhere. The data center was surrounded by farmland, and at least three other companies were building data centers in the area. The fields were crisscrossed by a tangle of wires from high-voltage electric towers, and large, hideous boxes were popping up everywhere.

The exterior of the site did not display any Microsoft branding—or any signage at all. Behind a fence, and past several vehicle checkpoints, the campus was a spacious expanse of nothing, except for one corner, which was populated by a row of numbered sheds. The sheds were white, narrow, tall, and several football fields in length; they reminded me of the livestock barns I visited as a child at the Minnesota State Fair. Flanking each shed was a row of diesel generators and industrial air-conditioners.

At the time of my visit, there were five sheds, and the plan called for roughly ten. There were construction vehicles

everywhere: cherry pickers, earthmovers, trucks carrying spools of cable. Someone had done a bit of landscaping in front of the shed I was visiting, and a few small plants grew in the shade.

Inside, I met with Judy Priest and Steve Solomon, both Microsoft executives; they have spent their professional lives managing warehouse-size computers. Priest, an electrical engineer, is a graduate of M.I.T., with high, sculpted eyebrows and wild blond hair. Solomon, a mechanical engineer, responded to my questions with long, technical monologues. Both seemed thrilled to participate in the new industrial revolution. Priest excitedly described a recent medical visit, after which she'd been sent an A.I.-assisted summary of her conversation with her doctor. Solomon, who'd been having trouble with his stereo, had taken a picture of the connections in the back and uploaded it to Chat. The A.I., he told me, had returned seven possible troubleshooting solutions. Here his voice took on a slight intonation, which I took to mean that he was expressing an emotion. "No. 3 worked," he said.

After donning a pair of steel-toed boots and watching a PowerPoint presentation, I passed through a security checkpoint and into the inner sanctum. The facility was quieter, tidier, and more spacious than the CoreWeave center. Hundreds of identical blinking banks of servers and

computing equipment, attached to cooling stations and humming noisily, took up much of the floor. Zip-tied bundles ran along the ceiling: wires for electricity, cables for data, pipes for water and air. The cables connected to a larger bundle of cable, which linked with the other sheds, allowing all of them to act as a single, unified computer. Across all five sheds, the area dedicated to computing was the equivalent of twenty football fields.

Priest explained that an advanced training run could tie up the entire system for a month. I stood with a technician at a control center, monitoring the electrical draw. We watched as the power spiked—the computer was processing training data. Then it receded—now it was writing the results to the file. These pulses repeated as the A.I. moved from one checkpoint to the next. Somewhere inside the building, the model was improving. Somewhere inside the building, the computer was learning how to think.

Leaving the data center, I found myself desperate for human contact. A half mile down the road, the top of a grain silo peeked out from behind a data-center construction site. I drove through a landscape of gray buildings, irrigation canals, power lines, and verdant fields before arriving at a dusty yard crowded with tractors and pickup trucks. There, I found a fourth-generation alfalfa farmer wearing bluejeans, a plaid shirt, and a baseball cap with a tanker truck

embroidered on it.

The farmer gestured to the power lines cutting through his field, which the local utility had installed in the nineteen-forties. "We always considered those things a liability," he said. "We thought they depressed the value of the land." But now, he said, access to a power substation was worth a fortune—one of his neighbors claimed that he had sold a plot of farmland to a data-center developer for more than a million dollars an acre, more than the farm would return in a lifetime. Piece by piece, the farmer said, his family was doing the same.

There was a new data center to the north of the farm, and another under construction to the east. Microsoft's sprawling facility dominated the horizon; it sat on a patch of dirt his family had worked since 1979. He told me that he was planning on moving out soon—the surroundings felt unfit for agriculture, or even for human life.

I asked the farmer if he had noticed any environmental effects from living next to the data centers. The impact on the water supply, he told me, was negligible. "Honestly, we probably use more water than they do," he said. (Training a state-of-the-art A.I. requires less water than is used on a square mile of farmland in a year.) Power is a different story: the farmer said that the local utility was set to hike rates for the third time in three years, with the most recent proposed

hike being in the double digits. The biggest loss was the nutrient-rich topsoil, which his family had maintained with careful crop rotations. "Microsoft brought in an excavator and ripped it all out in a day!" he said, as if speaking of a lost heirloom. "Six to ten feet of it, all gone."

We watched as a yellow dog got up, walked in a small circle, then went back to sleep in the shade of a tree. Behind the tree, and dwarfing it, was a giant rectangular warehouse. I asked the farmer if he ever used A.I. "I use Claude," he said. "Google sucks now."

Data centers are beginning to put intense pressure on America's electrical grid. In 1999, Constellation Energy purchased the sole functioning reactor at Three Mile Island, and operated it for the next twenty years. In 2019, the company shut the reactor down, concluding that it was economically unviable. Bryan Hanson, the executive who oversees Constellation's nuclear-generation fleet, threw a farewell party for employees. "There was food there, but nobody ate it," he said. "The mood was like a funeral."

Parties may soon return to Three Mile Island. Constellation has announced that it will reopen the facility in 2027, rebranding it as the Crane Clean Energy Center. A large contract with Microsoft made the difference. "If you told me we'd be reopening this plant just eight years later, I never would have believed you," Hanson said. (The second reactor

at the site, which released a cloud of radioactive gas into the atmosphere in 1979, will remain dormant.)

Energy executives such as Hanson have been bombarded by requests for more power. Data centers “are perhaps bigger, by an order of magnitude, than anything we’ve connected to the grid before,” he said. “If you think about the city of Philadelphia, its load is about one gigawatt. Now imagine adding one-gigawatt-sized data centers to the grid, and not just one, but multiples of them.”

When a data center comes online, retail customers usually help to foot the electric bill: American utilities sought almost thirty billion dollars in retail rate increases in the first half of 2025. This spring, utilities requested almost double the rate hikes from the same period a year earlier. An analysis by *Bloomberg* estimated that, in areas near data centers, wholesale electricity costs have risen by more than two hundred per cent in the past five years. And rates will probably continue to increase—power plants can’t produce nearly enough electricity to meet the demand. Eric Schmidt, the former C.E.O. of Google, has said that the U.S. must add ninety-two gigawatts of power to the national supply to meet data-center demand—some ninety-two Philadelphias. If there isn’t enough power, American A.I. developers might lose out to the Middle East and China, where enormous data-center projects are already under way.

Data centers must operate twenty-four hours a day to be economically viable. (The Microsoft facility I visited is permitted five minutes and fifteen seconds of unscheduled downtime per year.) Renewable energy sources like wind and solar, which depend on the weather, can currently meet only a fraction of this demand. Nuclear power won't save us, either, at least not anytime soon; Hanson said that it would be years before any new large-scale nuclear reactors could be constructed in the U.S. With envy in his voice, he told me, "China is building twenty-six nuclear reactors."

"I wake up, I do all these exercises designed to cultivate positivity, then I go to sleep."

Cartoon by Bruce Eric Kaplan

In the near term, new data centers will largely be powered by fossil fuels. Developers are purchasing land near natural-gas deposits such as the Marcellus Shale, a gigantic underground gas reservoir in Appalachia. In April, Homer City Redevelopment, a group based in Pennsylvania, announced that it intended to convert a mothballed coal plant outside Pittsburgh into the largest natural-gas power plant in the country, dedicated almost exclusively to data centers and capable of producing roughly four and a half gigawatts of electricity. According to an environmental nonprofit, the Homer City plant could release as much as four million pounds of carbon dioxide into the atmosphere every hour, about the same as four million idling cars.

The Earth is now warming at an estimated three-tenths of a degree Celsius per decade—roughly ten times faster than at the end of an ice age. After the last ice age ended, the oceans rose four hundred feet. Adding plants like Homer City—and scores more worldwide—will speed this catastrophic timeline. The Trump Administration has responded by restricting the use of the phrase “climate change” in government communications.

Data centers also cause local pollution. Elon Musk’s xAI has built a natural-gas-powered data center in Memphis, near the Black neighborhood of Boxtown. The area, which already had the highest rate of emergency-room visits for asthma in Tennessee, saw levels of nitrogen dioxide, which exacerbates the condition, spike as much as nine per cent after the plant moved in. Wealthier areas have tried to block the construction of data centers. In November, 2024, voters in Warrenton, Virginia, a wealthy exurb of Washington, D.C., replaced town-council members who supported a new Amazon data center with an anti-development slate. (Ann Wheeler, a Democrat in a neighboring Virginia county who lost her seat over data centers, complained about what she called activists’ “*BANANA*” mind-set: Build Absolutely Nothing Anywhere Near Anyone.)

Data-center construction is projected to represent two to three per cent of U.S. gross domestic product in the coming

years. In the nineteenth century, the building of the railroads contributed an estimated six per cent. Railroads transformed America and generated tremendous—if unevenly distributed—prosperity, but the frenzy also produced one of the largest speculative bubbles in history. The Panic of 1893 followed: unemployment soared, hundreds of banks went out of business, and a surge of populist sentiment destabilized the U.S. political environment.

The financier Jon Gray, the president of the alternative-asset manager Blackstone, brought up Ron Chernow's biography of John D. Rockefeller, Sr. "Many of the railroads went bust!" he said. "You're trying to avoid this problem, because you don't know what the endgame looks like." Blackstone has issued debt to build data centers; not wanting to be among those who go bust, Gray can hedge the risk by securing a fifteen-year lease agreement from a tech giant such as Microsoft or Amazon, which are some of the most creditworthy customers in existence. Blackstone typically won't invest in a data center unless it has such a customer lined up. "It's not like condos in Miami or Dubai," Gray said.

The premise of continued data-center construction is that stuffing more Nvidia chips into the sheds will result in better A.I. So far, this has proved true: the latest generation of A.I. is the most capable ever produced. OpenAI's GPT-5 can even build other, more primitive A.I.s. Still, it is not an immutable

law that more chips equals more intelligence, and researchers are not entirely sure why this scaling effect even exists. "It's an empirical question whether we will hit a brick wall," the A.I. pioneer Demis Hassabis has said of scaling. "No one knows."

It's also possible that a technological innovation might render hyper-scaling obsolete. Earlier this year, when DeepSeek, a Chinese company, unveiled what seemed to be a more efficient training paradigm for A.I., Nvidia's stock plummeted, wiping out almost six hundred billion dollars of value in a single day. (It has since recovered.)

Donald Trump has made the construction of data centers a national priority; it has become a kind of ritual for tech executives to announce new projects from the White House. But pandering to Trump might mean stretching the truth. At a White House dinner in September, Mark Zuckerberg said that Meta would spend six hundred billion dollars on data centers and related infrastructure in the next few years. With his microphone still live, Zuckerberg leaned toward Trump. "Sorry, I wasn't ready," he whispered. "I wasn't sure what number you wanted to go with." Kerry Person, who manages global data-center operations for Amazon, told me that electrical utilities were skeptical of some of the newer data-center developers filing requests for power. "If I look at the amount of demand that is in these queues, and I look at the

amount of money that would be required to build that out, that amount of money does not exist," Person said.

A.I., for all its wondrous capabilities, may disappoint investors. It may prove to be an unprofitable commodity: Claude, Grok, Gemini, and ChatGPT all have similar capabilities, and technological innovations are quickly copied by competitors. The tech giants do not in fact have unlimited funds: as companies like Microsoft and Meta pour money into the data-center race, their reserves of cash are shrinking. Investors might have unrealistic expectations: the U.S. stock market is approaching valuation ratios last seen during the dot-com era, and the venture-capital market has grown frothy. "Investors don't usually give a team of six people a couple billion dollars with no product. It's rare, and that's happening today," Jeff Bezos recently said.

Then again, it could be that the hype is justified. Nvidia's C.E.O., Jensen Huang, whom I published a biography of recently, is a world-class computer scientist who is producing the microchips that make the A.I. age possible. "We used to get silicon every two years," Priest, the Microsoft engineer, said. "Now we get silicon every few months." Nvidia accounts for roughly eight per cent of the market capitalization of the S. & P. 500, the highest concentration in any one stock in at least forty-five years. A lot is riding on Huang's ability to keep producing better

chips. If Americans want to retire comfortably, Nvidia has to succeed.

Water, power, and land are scarce resources, but the most valuable commodity for a data center, as the name suggests, is data. Claude trained on LibGen, a voluminous corpus of pirated e-books that can be downloaded by torrent. In September, Anthropic, the developer of Claude, agreed to pay one and a half billion dollars to the copyright holders of these books, or about three thousand dollars per infringement—the largest class-action copyright-infringement settlement in history. (I and others at this magazine are among the claimants.) Similar lawsuits against OpenAI and Nvidia are pending.

Microsoft does not know what its customers are uploading to its data centers—the data are proprietary. It is difficult to judge the scale of copyright infringement in the A.I. era, but my guess is that it makes Napster look like a mixtape swap. The modern approach to A.I. development has been to vacuum up any online data available—including audio, video, practically all published work in English, and more than three billion web pages—and let lawyers sort through the mess.

But there is now talk of a data shortage. There are thought to be about four hundred trillion words on the indexed internet, but, as the OpenAI co-founder Andrej Karpathy has noted, much of that is “total garbage.” High-quality text is harder to

find. If trends continue, researchers say, A.I. developers could exhaust the usable supply of human text between 2026 and 2032. Since A.I. chatbots are recycling existing work, they rely on cliché, and their phrasing grows stale quickly. It's difficult to get fresh, high-quality writing out of them—I have tried.

Priest, of Microsoft, told me that she wasn't concerned about running out of data: there is a universe beyond text, and A.I. developers are just beginning to explore it. The next frontier is "world model" data, which will be used to train robots. Streams of video and spatial data will be fed into the data centers, which will be used to develop autonomous robots. Huang, of Nvidia, wants to be in this market, too, and last year appeared onstage with two mobile androids. In Los Angeles, I have paused behind driverless cars and recently stumbled into an autonomous delivery wagon, but it was on a recent visit to Beijing that I began to understand what the robot revolution was going to look like.

Robots are everywhere in China. I saw them stocking shelves and cleaning floors at a mall. When I ordered food to my hotel room, it was delivered by a two-foot-tall wheeled robot in the shape of a trash can, with the voice of a child. I opened my door, nonplussed, to find it standing in front of me, decorated with an ersatz butler's outfit and chirping in Mandarin. A hatch on the front of the robot popped open,

and a tray of noodles slid out. The machine chirped again. I took my food, the hatch closed, and the robot wheeled away. I stood there for a time, holding the tray, wondering if I would ever talk to a human again. ◆