BANK OF AMERICA INSTITUTE

BANK OF AMERICA

Sustainability

Data center construction creates a resource shock

02 June 2026

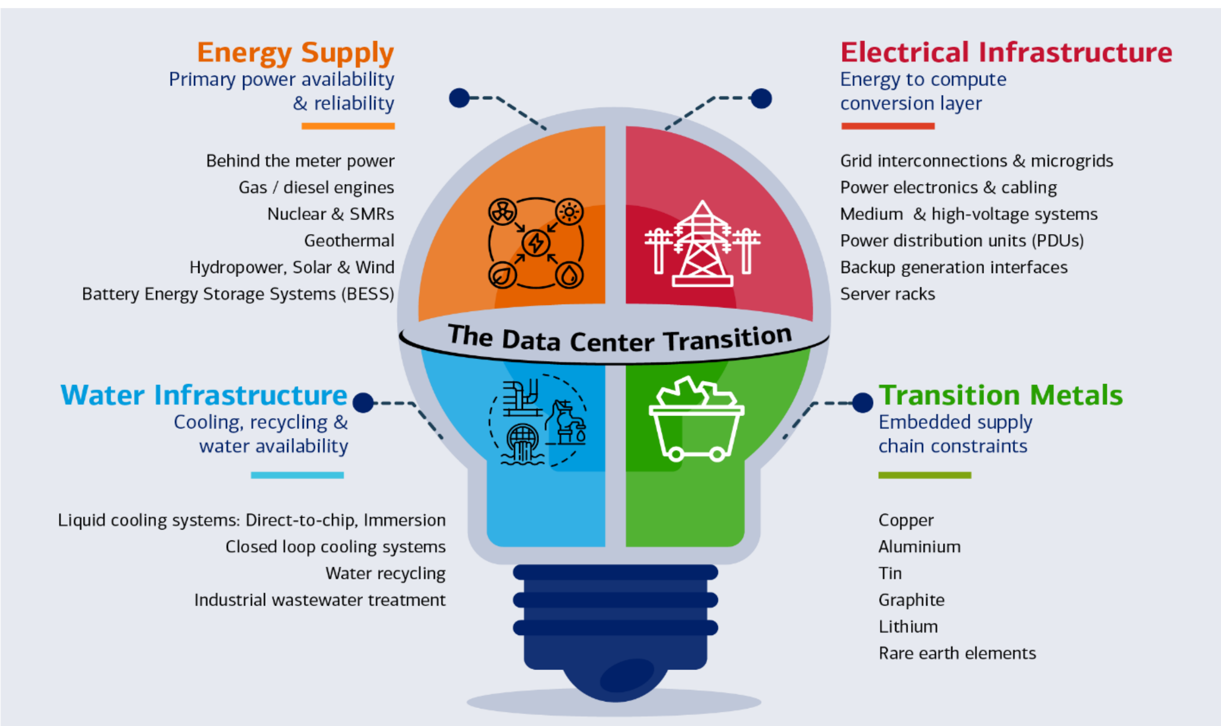
Key takeaways

- AI's water footprint is a substantial resource drain. Up to 75% of a data center's total water use happens off-site, largely through electricity generation rather than on-campus cooling. At scale, even small AI interactions add up: millions of liters of water per day are indirectly tied to routine AI use, according to BofA Global Research.
- Electricity demand from graphics processing unit-based servers is growing at roughly 30% per year, driven increasingly by AI inference. This shift stresses local grids, where the challenge is no longer total energy supply but the ability to deliver firm, continuous power at the right location.
- Each incremental megawatt of data center capacity embeds roughly 60-75 tons of metals, particularly copper, according to BofA Global Research. As data centers scale, they steadily pull on water, power, and materials that local infrastructure was never designed to supply at this pace, intensifying pressure on already-constrained regions.

Breaking down a data center's value chain

AI is turning data centers into some of the world's fastest-growing industrial water users. Every graphics processing unit (GPU) cycle adds heat; every watt of power demands cooling and cooling demands water.

Exhibit 1: AI data center growth is increasingly shaped by energy supply, electrical and water infrastructure and critical transition metals
Infographic depicting the AI data center value chain



Source: BofA Global Research

BANK OF AMERICA INSTITUTE

When compute heats up, the water runs out

AI is becoming one of the fastest-growing industrial users of water, and most of that demand sits outside the data center itself.

For a typical data center, only around a quarter of total water use is direct (Scope 1): cooling towers, humidification and other visible thermal systems. Most water is used off-site, split between electricity generation (Scope 2): particularly fossil and nuclear power plants and upstream hardware manufacturing, including semiconductor and chip fabrication. In aggregate, off-site water use can account for 75% of a data center's total water footprint.

> Using a 100-word AI prompt? This will need a half-liter of water. Scale that up, and mid-sized data centers are drinking 1-2 million (mn) liters a day. By 2030, the data center annual water tab could hit 1.2 trillion (tn) liters – equivalent to New York City's entire annual water use.

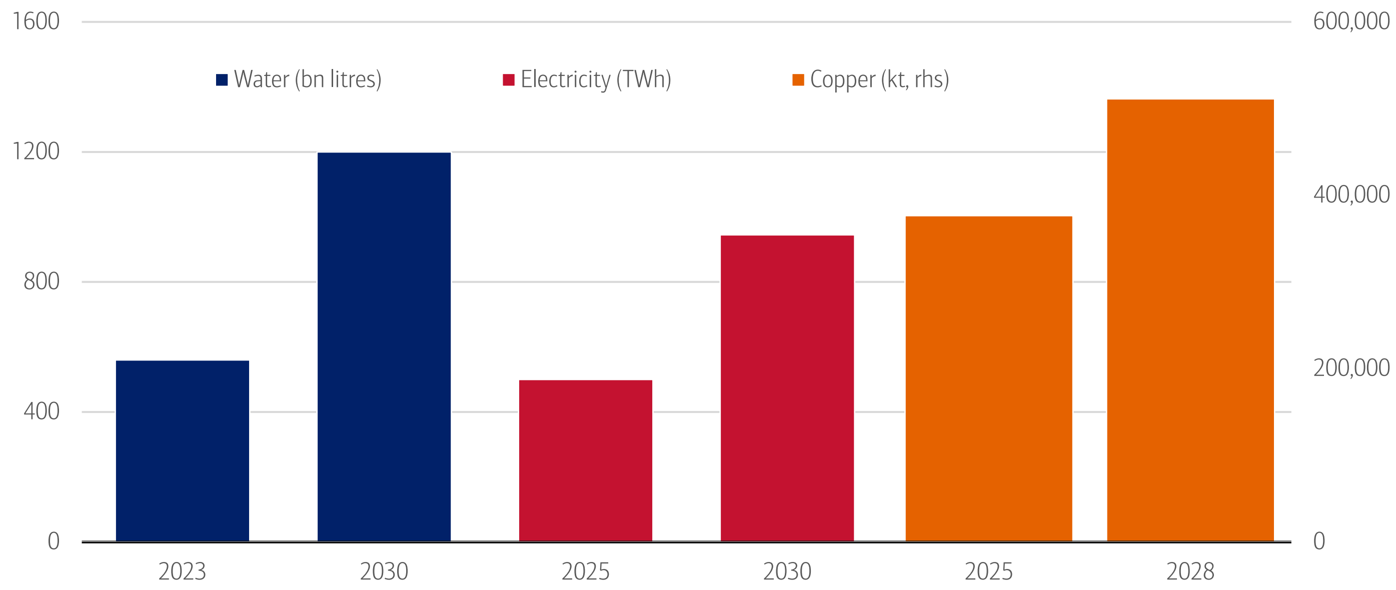
AI value chain is getting thirstier: Far larger with hidden footprint

Cooling loads are exploding as GPUs run hotter and denser, forcing operators toward increasingly water-hungry thermal systems. As power demand for AI soars, most grids still rely on water-intensive fossil and nuclear plants, meaning every new megawatt (MW) of AI compute carries significant hidden water use. Upstream manufacturing is an extreme user: semiconductor fabrication plants are water behemoths, consuming roughly 10 million gallons of ultrapure water per day to produce advanced chips.

The net result is that AI's water footprint is not just growing, but increasingly system-wide, opaque and geographically misaligned with where risks are measured or managed. For example, a data center can report an excellent Water Usage Effectiveness (WUE) – annual water consumption for humidification and cooling by the total energy consumption for information technology (IT) equipment – and still consume vast amounts if it is plugged into a water-intensive grid. In most markets, grid water intensity, not cooling system design, is the dominant driver of operational water risk.

Exhibit 2: Global data center capacity is expected to exceed 200GW (gigawatt) by 2030 and triple by 2035 versus 2024, driving a near doubling of electricity and water demand by 2030

Data center resource consumption by 2030



Source: International Energy Agency, Bloomberg NEF, BofA Global Research

Note: bn = billion; TWh = terawatts per hour; kt = kilotons; rhs = right-hand side

BANK OF AMERICA INSTITUTE

Cooling sits at the center of AI's water challenge

From an operator's perspective, cooling is the only lever of immediate, engineering-level control over water use. As a result, cooling strategies are evolving fast. Data-center cooling operates at both the server and building level and next-generation cooling is shaping one of the biggest capex waves in hyperscale infrastructure. Closed-loop liquid cooling, immersion cooling and free-air cooling are emerging as competitive differentiators and a fast-growing capex category for original equipment manufacturers (OEMs).

Next-generation designs are fueling a new wave of capex

Data centers rely on a mix of cooling systems – from room-level air conditioners and chillers to liquid systems that cool chips directly – to manage the rising heat generated by powerful AI chips (Exhibit 3). Traditional air-based cooling often uses evaporation, which consumes significant water, while liquid-based cooling systems circulate water in closed loops and are far more water-efficient.

As AI models become more advanced, the chips that run them use more power and generate more heat. Servers are also being packed more tightly, concentrating that heat in smaller spaces. Effective cooling is therefore essential to keep systems running smoothly; without it, excess heat can degrade performance, shorten equipment lifespan and increase the risk of failures.

Inside the server rack, liquid cooling is emerging as the preferred solution. BofA technology analysts expect direct liquid cooling – where liquid removes heat directly from the chip – to remain the dominant approach, as it can handle high heat loads reliably within limited space.

At the building level, data centers typically rely on one of two methods to release heat: air-based cooling, which uses little water but more electricity, or evaporative cooling, which reduces energy use but requires much more water. Local climate conditions often determine which option is practical.

Exhibit 3: Data centers rely on a mix of cooling systems – from room-level air conditioners and immersion cooling to liquid systems
Types of server cooling system

Factor	Air Cooling	Evaporative Air Cooling	Direct-to-Chip Liquid Cooling	Rear-Door Heat Exchangers	Immersion Cooling
How it works	Air conditioning units use chilled air to cool the facility. Cold air circulates to the racks, while hot air returns to the air handling units for cooling and recirculation.	Air cooling where incoming air temperature is reduced via water evaporation (adiabatic process) before or during supply to IT equipment.	Liquid coolant circulates through cold plates on CPUs/GPUs and returns to a cooling distribution unit in a closed loop.	Liquid coolant circulates through heat exchangers mounted on the rear of each rack.	Servers are submerged in dielectric fluid that absorbs heat and circulates through a heat exchanger.
Water use	Low (no evaporation).	Evaporative systems consume water by design.	Very low, small, fixed volume in a closed loop; water use depends on heat-rejection method.	Low as closed-loop coolant; minimal water use at building level.	Minimal as synthetic fluids replace water entirely.
Energy efficiency at higher rack densities	Lower efficiency at higher rack densities.	More energy-efficient than mechanical air cooling; attractive in power-constrained grids.	Liquids have orders-of-magnitude higher heat capacity than air, enabling much higher rack densities.	Moderately efficient, improves rack-level thermal performance.	Highest efficiency; supported at rack/system level.

Source: BofA Global Research
Note: This is not an exhaustive list of technologies and factors; it is provided for illustrative purposes only.

BANK OF AMERICA INSTITUTE

The $58bn water infrastructure gap

Most US water utilities have yet to fully account for the water implications of AI-driven infrastructure growth. A 2025 survey of 680 US water-sector stakeholders points to significant planning gaps.

- **Limited visibility of data-center demand**: Only 14% of respondents reported direct water demand from data centers including AI facilities in their service areas, even as development expands into secondary and rural regions.
- **Little impact on reuse planning so far**: Around two-thirds of respondents reported no increase in demand for recycled or repurposed water linked to data centers or advanced manufacturing; just 12% replied in the affirmative, but that figure is likely to rise as reuse becomes more viable for cooling.
- **Water planning lagging infrastructure growth**: 54% of respondents have not incorporated data center and advanced manufacturing water needs into resource planning, while only 13% have done so.[1]

This lack of preparedness contrasts sharply with demand projections. Academic research estimates that US utilities may require an additional 697mn to 1.45bn gallons per day of peak water-supply capacity by 2030 to support AI-driven data-center cooling. The associated investment needed to expand water infrastructure is estimated at $10bn to $58bn, depending on the pace and concentration of data center growth.[2]

---

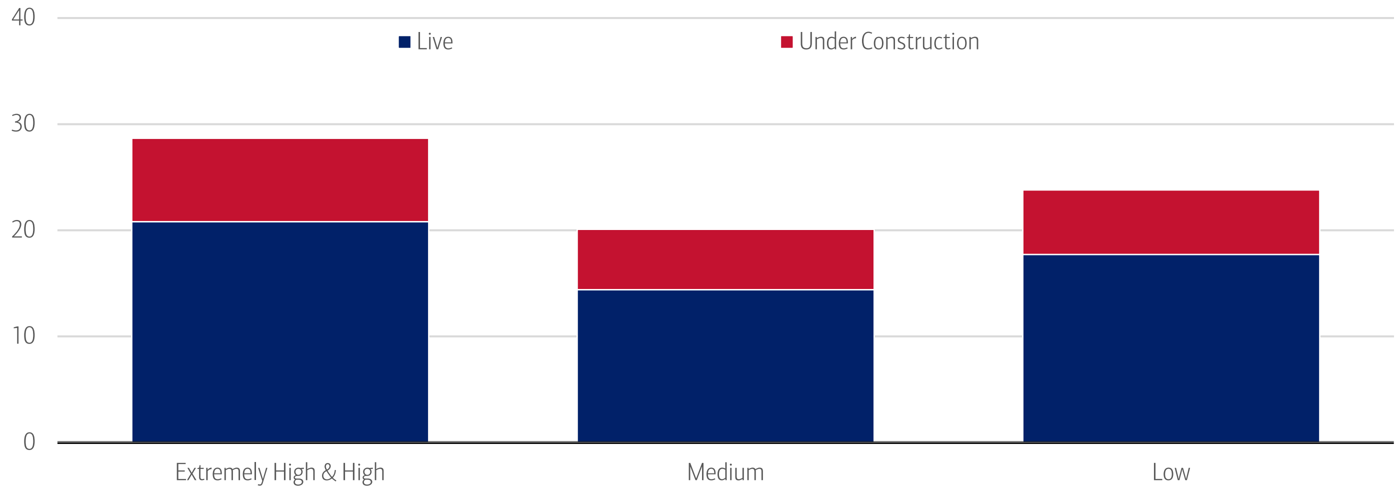
[1] Black & Veatch. (2025). Black & Veatch 2025 Water Report.
[2] Han, et. al. (2026, March 3). *Small Bottle, Big Pipe: Quantifying and Addressing the Impact of Data Centers on Public Water Systems.* University of California, Riverside.

Data centers are being built where the water isn't

Water availability is increasingly seen as an important factor in deciding where new AI computing capacity should be built. In practice, however, data centers continue to cluster in water-stressed regions, where access to power, available land and more permissive regulations often outweigh local water constraints, according to BofA Global Research.

Exhibit 4: Over 20GW of live global data center capacity is in extremely high or high water-stress regions

Total capacity by water stress level (in MW). Medium =10-20%, Low <10%



Source: Bloomberg, BofA Global Research

Note: Bloomberg's Data Centers and Water Stress Exposure Dataset maps global IT capacity to baseline and projected water stress levels.

BANK OF AMERICA INSTITUTE

Adaptation needs: Wastewater and non-potable water solutions

Water impacts do not end at withdrawal. Alternative and non-potable water sources account for less than 5% of total water used by a typical data center.[3] This highlights how early the transition remains, even in regions exposed to water stress. Cooling systems fundamentally alter both the quantity and quality of water downstream, creating challenges that extend beyond freshwater scarcity.

In evaporative cooling systems, around 75-80% of withdrawn water is irreversibly consumed through evaporation. The remainder is discharged as blowdown wastewater, containing elevated dissolved solids, heat load and treatment chemicals such as biocides and corrosion inhibitors. Most municipal wastewater treatment plants were not designed to handle this industrial chemistry at hyperscale volumes, creating friction as data center density increases.

When combined with water-intensive electricity generation, this produces a water-energy-wastewater feedback loop that utilities are increasingly scrutinizing. As a result, wastewater handling is becoming as important as freshwater withdrawal in permitting decisions, with emerging requirements around pre-treatment, discharge quality and developer-funded infrastructure upgrades.

How cooling water is managed today and where efficiency gains lie

Most cooling systems recirculate water through cooling towers or heat exchangers until quality thresholds are reached. Some facilities discharge used water to municipal treatment plants or directly to receiving waters subject to temperature and chemical limits. Evaporation remains the dominant source of water loss, while poor maintenance can also lead to leakage, drift and overflow, creating hidden inefficiencies.

Advanced approaches including closed-loop cooling, wastewater recycling and rainwater harvesting can reduce freshwater demand by an estimated 50-70%, depending on climate and system design, but require rigorous treatment and monitoring.[4]

Policy signals heat up

In the United States, policy responses remain fragmented and largely local. According to BofA Global Public Policy, the White House is largely focused on accelerating data center development through permitting reform, infrastructure prioritization and support for global AI leadership. At the Congressional level, proposals reflect growing concern over ratepayer protection, environmental impacts and the need for greater transparency and accountability while prioritizing global competitiveness.

---

[3] Rosenfield, et. al. (2025, July 14). *Data centers and water.* Norton Rose Fulbright Project Finance Newswire.

[4] Ramboll and Grundfos. (2024, May 15). *Whitepaper: Water circularity can reduce water stress and boost EU industry.* Grundfos

In contrast, states and localities are advancing a wide range of approaches. Southern Nevada has prohibited evaporative cooling in new developments due to severe water stress, while California's proposed AB 93 would require businesses to disclose water and energy use and could link efficiency standards to operating licenses. Florida has also begun to publicly debate water use by AI data centers – some of which reportedly consume up to 500,000 gallons per day – prompting discussions around new water-management legislation, a signal that oversight is likely to intensify, according to BofA Global Research.

Policy is also beginning to condition incentives based on reclaimed-water use. The proposed Virginia Senate Bill 417 (2026 session) would require data-center operators receiving state infrastructure grants to progressively increase the share of reclaimed water used for cooling from 60% in 2027 to 100% by 2031 if enacted. A small number of large data center campuses already rely on reclaimed or treated wastewater for cooling. In parts of Georgia, multiple developments access treated municipal wastewater, reducing the reliance on potable supplies.

Not just a drain on water: The data center transition is electrical

By 2030, data centers could exceed 3% of global electricity demand. Electricity use from accelerated (GPU-based) servers is growing c.30% per year, versus c.9% for conventional central processing units (CPU) servers. Accelerated servers account for nearly half of the total increase in data-center electricity demand; conventional servers contribute c.20%, with the remaining c.30% coming from cooling and infrastructure.

The key driver is no longer model training alone, but inference, especially reasoning-intensive and agentic workloads. As AI use expands from text into image, video and multimodal tasks, energy per task increases, utilization becomes near continuous, and workloads become harder to interrupt (read more on this in On the clock: Agentic AI in the workplace).

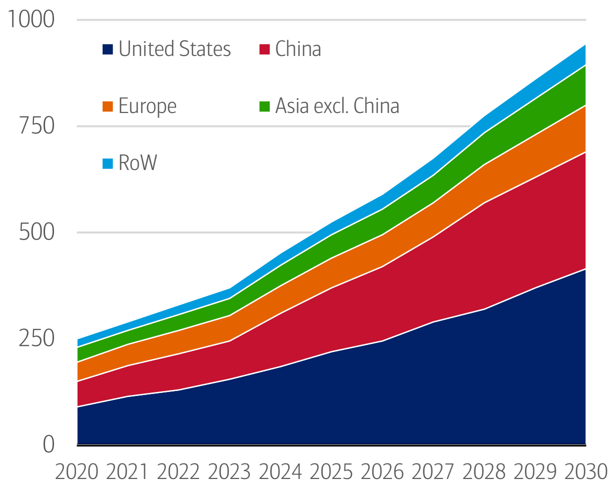
Higher power density, sustained load and infrastructure intensity

Cooling alone accounts for c.7% of electricity use in efficient hyperscale sites and >30% in less-efficient enterprise facilities. The US Department of Energy (DOE) explicitly identifies AI servers and cooling as key drivers of rising electricity demand.

The broader implication is that data center growth should not be analyzed as a single number in terawatt-hour (TWh); it should be analyzed as a stack: server power, cooling, UPS, batteries, conversion losses, backup generation and network reinforcement. This is where the next leg of value creation is building.

Exhibit 5: Data centers to consume close to 1,000TWh in 2030

Data center electricity consumption, 2020-2030, IEA base case



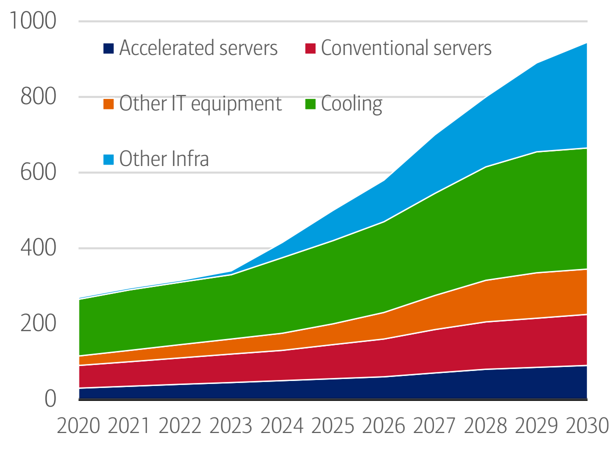
Source: International Energy Agency (IEA), BofA Global Research estimates

Note: RoW = rest of world.

BANK OF AMERICA INSTITUTE

Exhibit 6: Cooling and servers account for 80% of power consumption

Source of data center electricity consumption, 2020-2030, IEA base case



Source: International Energy Agency (IEA), BofA Global Research estimates

Note: Other infra = other infrastructure.

BANK OF AMERICA INSTITUTE

The bottleneck has shifted to firm, deliverable power

Power systems have historically been evaluated around energy adequacy or whether there is enough megawatt-hour (MWh) for the year. Data centers increasingly force the system to be evaluated around deliverability: location, firmness and controllability. That is why assets tied to access, not just generation, are being re-rated.

Once deliverability becomes a bottleneck, data centers stop being a simple demand driver and become a system-design problem: how much generation, wires, redundancy, storage and onsite infrastructure needs to be built; where does it sit; and who finances it. The DOE has already pointed to onsite generation, storage, innovative rate structures, transmission expansion and re-use of retired coal infrastructure as part of the response to data-center growth.

According to BofA Global Research, nuclear fits the data center load profile unusually well: very high-capacity factors, dense power close to site, long-duration contracts and low-carbon firmness. Data centers may therefore become the first credible commercial market for small modular reactors (SMRs), not because they are the cheapest source of power, but because they offer certainty of access in the most constrained markets, according to BofA Global Research (read publication [Q&A quick guide: Small modular reactors] for more).

Metals caught in the data center construction crosshairs

As AI reshapes data center architecture, metals are shifting from incidental construction inputs to strategic enablers of deployment. Rising compute density, advanced cooling architectures, and tighter uptime and redundancy standards mean metals demand is now driven by engineering constraints and system reliability, not floor space or headline capex.

Where metals exposure really sits

Although servers dominate the narrative, the bulk of metals in data centers sits outside the server hall, embedded in power delivery, redundancy and cooling infrastructure (Exhibit 9). Each incremental megawatt of capacity incorporates roughly 60-75 tons of metals, with power systems and cooling together accounting for 55-75% of total metals intensity.

Exhibit 7: Data centers would not work without metals

Mined raw materials used in data centers

Metals	Application
Structural and Rack Materials	
Steel	Used for server racks, frames, and building structures because of its strength and durability.
Aluminium	Common in lightweight racks and enclosures; also used for heat sinks due to its good thermal conductivity.
Electrical Conductors	
Copper	The primary metal for power cables, busbars, and networking cables because of its excellent electrical conductivity.
Aluminium	Sometimes used in power distribution for cost savings, though less conductive than copper.
Cooling and Heat Management	
Aluminium	Widely used in heat sinks and cooling components.
Copper	Used in high-performance heat exchangers and liquid cooling systems because of superior thermal conductivity.
Electronic Components	
Gold	Found in connectors and circuit boards for corrosion resistance and reliable conductivity.
Silver	Used in high-end contacts and solder for its excellent conductivity.
Tin	Common in solder for printed circuit board (PCB) assembly.
Nickel	Often used as a plating material for connectors.
Batteries and Backup Systems	
Lead	In lead-acid batteries for uninterruptable power system (UPS) systems.
Lithium	In lithium-ion batteries for modern backup solutions.
Specialty Metals	
Rare Earth Elements	Used in magnets for hard drives and cooling fans.

Source: BofA Global Research

BANK OF AMERICA INSTITUTE

Copper: Small share, outsized marginal impact

Metals typically account for less than 5% of total data center capex, making demand highly price inelastic. As data center construction accelerates, copper demand is being pulled forward into an already tightening market, with AI projects acting as volatility amplifiers rather than marginal buyers.

By 2030, AI data centers could account for c.2% of global copper demand, up from <1% today, and add 2-3% to the North American copper demand compounded annual growth rate (CAGR) through 2035, driven mainly by grids and power equipment, according to IEA. In China, direct AI demand may reach 5-6%, while indirect grid-related demand could account for roughly half of copper demand growth.

Beyond copper: Aluminum, tin and specialty metals

But it's not just copper: aluminum demand from data centers is projected to more than double, from c.330 kiloton (kt) in 2025 to c.695 kt by 2030, according to BofA Global Research. Even small increases in demand can drive outsized price volatility when end demand is structurally inelastic and tied to fixed infrastructure build-out.

Beyond basic construction materials, data centers rely on a small number of specialized metals that are hard to source and often concentrated in a few countries. Supplies of key inputs like gallium, germanium and rare earths are sensitive to geopolitics and trade policy, which can quickly drive up prices and cause delays. Even though these materials are used in tiny amounts, they are essential to high-performance computing components (i.e. chips, fiber optics, power systems, cooling equipment) making the data-center supply chain unusually exposed to disruptions.

Geopolitics turns metals into strategic assets

Geopolitics has reclassified data-center metals into strategic infrastructure assets, according to BofA Global Research. In 2025, the US expanded Section 232 tariffs, lifting steel and aluminum duties to 50% and signaling potential further action on copper, increasing uncertainty around delivered costs and procurement. At the same time, China's dominance of refining and midstream capacity across copper, aluminum, electrical steel and silver has sharpened concerns over security of access.

Which metals bind first varies by region: transformers in the US, grid reinforcement in Europe, midstream concentration in China (Exhibit 10). Beneath the near-term volatility though, the structural story remains intact: global grid expansion, electrification and renewables-led power growth continue to push key markets like copper and aluminum into medium-term deficit.

Exhibit 8: No region can yet achieve full security of supply

Security of supply of energy transition metals and minerals across the transition metals outlook regions

	China		Europe		US		Japan and South Korea		Southeast Asia		Rest of world	
	2025	2035	2025	2035	2025	2035	2025	2035	2025	2035	2025	2035
Aluminum	100%	100%	67%	50%	70%	57%	78%	57%	100%	100%	100%	100%
Cobalt	9%	15%	9%	28%	3%	11%	1%	4%	100%	100%	100%	100%
Copper	17%	18%	100%	98%	86%	62%	36%	30%	100%	56%	100%	100%
Graphite	100%	100%	19%	19%	10%	19%	98%	64%	96%	33%	28%	16%
Lithium	49%	92%	4%	22%	9%	22%	2%	4%	0%	0%	100%	100%
Manganese	7%	7%	0%	0%	0%	2%	0%	1%	0%	0%	100%	100%
Nickel	10%	12%	55%	64%	39%	42%	21%	23%	100%	100%	100%	100%
PGMs	20%	30%	2%	0%	31%	40%	15%	20%	27%	26%	100%	100%
Steel	100%	82%	80%	72%	92%	76%	100%	100%	70%	100%	100%	100%

Key
Regional supply meets total demand
Regional supply meets most demand
Regional supply meets some demand
Regional supply meets little demand
Regional supply meets less than 40% of demand

Source: BloombergNEF. BofA Global Research

Note: The table assumes that all supply is consumed domestically. Demand is total demand (energy transition and other) under BNEF's Economic Transition Scenario. Supply is both primary and secondary supply. Cobalt, copper, lithium, manganese, nickel and platinum group metals (PGMs) are mined supply. Aluminum, graphite and steel are refined supply.

BANK OF AMERICA INSTITUTE

Contributors

Taylor Bowley

Economist, Bank of America Institute

Vanessa Cook

Content Strategist, Bank of America Institute

Sources

Menka Bajaj

Sustainability Strategist, BofA Global Research

Haim Israel

Equity Strategist, BofA Global Research

Juliana Cantrell

Senior Vice President, BofA Public Policy

Disclosures