INTELLIGENCE UPDATE

# AI to trigger radical overhaul of data center electrification

Daniel Bizo 18 Dec 2024

As we enter 2025, the IT and data center industry remains latched onto generative AI. The influence of a single type of workload reshaping an entire industry is unprecedented. The closest comparison is Bitcoin, but this attracted specialist operators' development of "shadow" facilities, rather than a transformation in mainstream data centers.

While some remain skeptical about the economic viability and utility of massive generative AI models, many see the past two years as just the beginning. For now, supersized generative AI models are here to stay, placing onerous demands on infrastructure — both in terms of IT and data center facilities. Initially, the industry focused on solving cooling issues, but the more pressing challenge for next-generation AI infrastructure will be power, forcing operators to explore new electrification architectures.

## The crushing forces of AI

The technical direction of AI hardware development centers on densification. The current generation of AI compute racks in mass deployment (built around Nvidia H-series products) are typically around 40 kW per standard 19-inch rack — already well above typical (average) rack densities. Most organizations do not have a single high-density rack above 30 kW. There are even higher density deployment options for operators that want to compress cluster footprint and optimize cabling. But the current generation of AI compute racks (mostly) remain within the realm of standard power distribution equipment — notably busways and breakers.

However, AI hardware product roadmaps project densities to jump with each generation. Nvidia is spearheading this trend, pushing its generative AI hardware to adopt supercomputing-style system architecture, that prioritizes high-speed data sharing directly between its GPUs. This objective requires tight integration of many GPUs, linked via copper interconnects in a local mesh. This has a direct consequence for rack power density — more than any other factor.

Nvidia is on track to begin volume shipments of its first rack-scale GPU-systems in 2025. The high-end configurations with 72 GPUs per rack are rated at 130 kW with half-sized versions in the 60 kW to 70 kW range. Subsequent chip upgrades demand even greater power in 2025, with the 2026 generation projected to push thermal power ratings beyond 2 kW per GPU.

But the real power challenge that lies ahead will come from even tighter integration of compute chips. Nvidia's product roadmap calls for doubling the number of GPUs per rack — and then doubling again. These rack-scale systems, 300 kW and above, are slated to start shipping as early as 2026. These extreme levels of densities have only ever been seen before in the leading edge of supercomputing.

## Data center power needs a step change

Demand for such AI training-focused rack systems remains uncertain. However, many operators and equipment suppliers consider hundreds of kilowatts of power per rack to be a likely scenario. While most organizations developing large AI models may not require the most advanced hardware, the AI market continues to be dominated by a handful of technology firms locked in an infrastructure development race. Even for mainstream organizations, typical AI racks are expected to start at around 80 kW, with many versions reaching the 100 kW to 200 kW range within a few years.

The prospect of this level of power densities becoming widespread is forcing a rethink in the data center industry. An order of magnitude jump in rack power presents several challenges, particularly the proportion of space required for low-voltage distribution (automatic transfer switches, switchboards, UPS systems, distribution boards, batteries) relative to the technical space. Without changes to the power architecture, many data centers risk becoming electrical plants built around a relatively small IT room, where every aisle (or row even) requires a corresponding large UPS system, energy storage and other associated electrical equipment.

Another issue is the size and weight of the conductors delivering power to the IT space. Each row of future high-performance racks will have equivalent power demand to entire megawatt-scale data centers of the past, concentrating busways and cables into compact areas. This also mandates tight coupling of low-voltage (LV) electrical rooms and IT rooms to minimize costs and distribution losses. As the size of electrical rooms increases, site planning and design may become more difficult, especially when retrofitting existing buildings or operating in urban environments.

## Transforming data center power systems

Major electrification suppliers are developing several product innovations that will help data center operators optimize their power distribution:

- Medium-voltage (MV) distribution to the IT space.
- Novel IT power distribution topologies.
- Solid-state transformers.

A key solution to the challenges outlined above is switching to MV UPS systems and downstream distribution — generally defined as any voltage level between 1,000 volts and 35 kilovolts. There are several benefits:

- Dramatically reduces conductor sizes and distribution losses.
- Helps site planning by enabling greater distances between the UPS rooms and the IT load, without severely impacting cost and losses.
- Compresses total electrical plant footprint.

Some MV UPS systems are currently available from select vendors, including both diesel-rotary and static — but choice is limited. For MV distribution to become more appealing, a larger number of established UPS vendors need to enter the segment, which Uptime Intelligence views as a likely outcome in coming years.

MV equipment is not new in data centers: MV engine generators and switchgears are common in multi-megawatt facilities. But as infrastructure power densification continues, the closer MV distribution is to the IT load, the better — providing segregation can be maintained between the MV and LV sides to comply with electrical safety codes. Potentially, future data center designs will incorporate novel power topologies that deliver MV power immediately next to power racks in the data hall to which IT racks connect.

MV power distribution designs are currently constrained by the need to install coupling transformers, which require considerable space and cannot be installed directly within the IT space. The development of sold-state transformers (SSTs), however, which is in progress at several power electronics suppliers for commercial deployment in a range of applications, may be a solution.

SSTs promise to be much smaller and lighter than a comparable core-based transformer, but also offer other power quality benefits. Furthermore, SSTs also make it possible to produce direct current output, simplifying IT power supply units and reducing total conversion losses in the power chain.

The overhaul of data center power chains will take several years to unfold, with 2025 shaping up to be a pivotal year. If investments into AI and data center infrastructure continue unabated, funding of the commercialization of MV UPS systems, SSTs and other innovations will be secure.

## The Uptime Intelligence View

Current data center power chains and electrical equipment are highly efficient and can support a wide range of capacities. However, despite best efforts by facility designers and equipment manufacturers to enhance LV systems, they remain suboptimal for a high-density future. Introducing MV power close to the IT space will deliver a much-needed step change to facility electrification by relaxing layout constraints and gradually improving infrastructure economics. Such a profound technical change will require substantial investment from across the data center industry, driven by confidence in future demand. Spending on infrastructure for generative AI is providing that confidence.

## ABOUT THE AUTHOR


### Daniel Bizo

2 May 2026

Over the past 15 years, Daniel has covered the business and technology of enterprise IT and infrastructure in various roles, including industry analyst and advisor. His research includes sustainability, operations, and energy efficiency within the data center, on topics like emerging battery technologies, thermal operation guidelines, and processor chip technology.

**dbizo@uptimeinstitute.com**

## About Uptime Institute

Uptime Institute is the Global Digital Infrastructure Authority. With over 4,000 awards issued in over 122 countries around the globe, and over 1,100 currently active projects in 80+ countries, Uptime has helped tens of thousands of companies optimize critical IT assets while managing costs, resources, and efficiency. For over 30 years, the company has established industry-leading benchmarks for data center performance, resilience, sustainability, and efficiency, which provide customers assurance that their digital infrastructure can perform across a wide array of operating conditions at a level consistent with their individual business needs. Uptime's Tier Standard is the IT industry's most trusted and adopted global standard for the design, construction, and operation of data centers.

Offerings include the organization's Tier Standard and Certifications, Management & Operations reviews and assessments including SCIRA-FSI financial sector risk assessment, the Sustainability Assessment, and a broad range of additional risk management, performance, availability, and related offerings. Uptime Education training programs have been successfully completed by over 100,000 data center professionals, such as the much-valued ATD (Accredited Tier Designer) and AOS (Accredited Operations Specialist). The Uptime Education curriculum has been expanded by the acquisition of CNet Training Ltd. In 2023.

Uptime Institute is headquartered in New York, NY, with offices in London, Sao Paulo, Dubai, Riyadh, and Singapore, and full-time Uptime professionals based in over thirty-four countries around the world.

For more information, visit www.uptimeinstitute.com