5C.



5C Industry Report
Volume 1.

Inside the Minds of AI Infrastructure Builders.



Contents.

Executive Summary 04.

About this Report 04.

Power takes center stage in AI infrastructure 06.

A new era of integrated system design 08.

Inference is unlocking a new architecture for speed 10.

Smarter cooling, stronger performance 12.

Speed is becoming a builder's biggest advantage 14.

Sustainability is becoming a source of shared value 16.

The permanent foundation 16.

Acknowledgments 18.

Closing 19.

Executive Summary.

The industry is entering an exciting new phase in which power, cooling, compute, networking, and physical space are coming together as a fully integrated system, often described as an "AI factory."

This evolution is enabling infrastructure leaders to think more holistically, plan ahead for next-generation hardware, and proactively address power requirements that are shaping deployment strategies.

Insights from leaders across the AI ecosystem highlight a strong sense of momentum and alignment around several key priorities: disciplined power planning, deeper collaboration across technology and construction partners, flexible facility design, advanced cooling innovations, and infrastructure built to support scaling AI workloads.

While much of the public conversation around AI focuses on algorithms, model capabilities, and software, these breakthroughs are made possible by a rapidly expanding and highly sophisticated physical infrastructure layer.

Here's a look at what's currently reshaping the industry, and what makes this moment such a significant one to be building in.

About this Report.

This report draws on interviews conducted with 16 leaders across AI infrastructure, including technology providers, engineers, builders, and operators.

It reflects the perspectives shared by those contributors and is intended as an industry discussion rather than an independent market study. Quotations may have been lightly edited for clarity and length without changing their meaning.



We need to become much more sophisticated on tracking power... building hundreds of gigawatts of power when we only need 20% less is hard to recover from.

Alon Gavrielov
VP of Infrastructure Strategy

together.ai

Power takes center stage in AI infrastructure.

Leaders across the AI ecosystem are improving how they manage power, making better use of available capacity while working to minimize potential impacts on surrounding communities.

The competition to deploy AI infrastructure is no longer just about buying the fastest GPUs. It's also about securing gigawatt-scale energy, and the companies leading that race are thinking further ahead than ever before. This next phase isn't just about chips, racks, or data centers in isolation — it calls for a strong ecosystem that brings advanced compute and infrastructure execution together. For years, the bottleneck was silicon availability. Today, companies are planning infrastructure and securing grid access many years ahead of hardware deployment, a level of foresight that is rapidly becoming an industry standard.

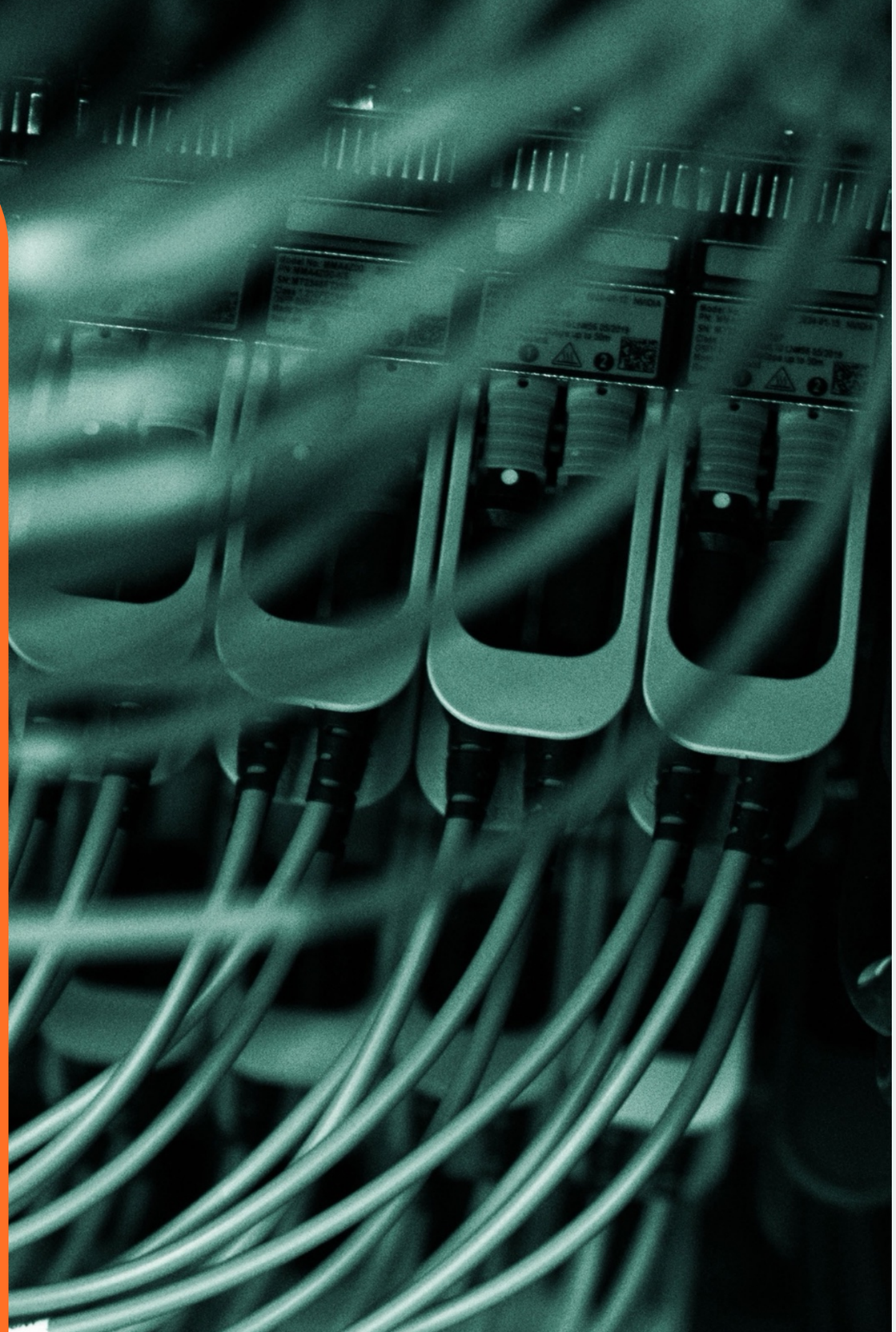
As demand forecasts evolve, infrastructure providers are placing greater emphasis on phased capacity commitments, utilization visibility, and closer alignment between power procurement and deployment schedules. As Alon Gavrielov, VP of Infrastructure Strategy at Together AI, points to exactly the discipline the industry is now building toward: "We need to become much more sophisticated on tracking power... building hundreds of gigawatts of power when we only need 20% less is hard to recover from."

That sophistication is exactly what leaders like Brad Nimri, Chief Revenue Officer at 5C, are already putting into practice. "What people are planning for is not necessarily the current generation of GPU," Nimri notes. "It's how do I secure power one, two, three, four years from now?" It's this kind of demand-matched, forward-looking planning that positions the industry to turn its biggest constraint into its clearest competitive edge.

What people are planning for is not necessarily the current generation of GPU, it's how do I secure power one, two, three, four years from now?

Brad Nimri
Chief Revenue Officer

5C.



You no longer design a data center as single elements. All of that has to be designed as a system.

Jim Simonelli
Chief Technology Officer

Schneider Electric

A new era of integrated system design.

This shift is opening the door to a more powerful way of building.

Jim Simonelli, Chief Technology Officer at Schneider Electric, sums up the new reality well: “You no longer design a data center as single elements. All of that has to be designed as a system.” That system-level thinking is what’s enabling the AI factory to perform at a level piece-by-piece design never could.

A modern AI cluster acts as one massive distributed machine, thousands of GPUs, storage nodes, and network switches operating synchronously to train a single model. That level of coordination is a remarkable engineering achievement, and it’s also raising the bar for what “reliable” means: if one node lags or one cooling loop falters, the whole system feels it. Alon Horev, Chief Technology Officer at VAST Data, puts it plainly: “It’s enough for one laggard to slow the entire application down.”

Recognizing this is pushing the industry toward exactly the kind of integration it needs. “You can’t just think about the chips in a vacuum,” notes Alon Gavrielov, VP of Infrastructure Strategy at Together AI, making the case for soup-to-nuts design.

An AI factory isn’t meant to be built piece by piece, and the best teams are embracing that. It functions as a single, interdependent system, one where getting the plumbing right matters just as much as getting the silicon right. Many conventional data centers were designed to support a broad mix of relatively independent workloads. By contrast, an AI factory is increasingly engineered as a tightly coordinated system in which compute, networking, storage, power, and cooling must operate together.

Co-design is quickly becoming the new standard, and for good reason. When hardware vendors, mechanical engineers, and utilities align from day one, the facility that results runs better on every level. Infrastructure built with strong pre-construction alignment among technology providers, engineering teams, construction partners, and utilities is best positioned to reach real operational efficiency, and it’s this kind of collaboration, not siloed procurement, that’s setting the pace for the industry going forward.

Inference is unlocking a new architecture for speed.

AI is entering an exciting new chapter, and the infrastructure that carried it through training is now evolving to meet it.

The next generation of data centers is set to prioritize disaggregated hardware, immediate data retrieval, and secure multi-tenant access to enterprise data, unlocking a level of responsiveness the industry hasn't delivered before.

The last several years were defined by the massive throughput required to train foundation models, a remarkable feat of engineering in its own right. Now, as AI moves from the lab into enterprise production and autonomous agents, the dominant workload is shifting to inference, opening up new opportunities for innovation. As AI moves from development into enterprise production and agent-based applications, inference is becoming an increasingly important part of infrastructure demand. While training typically places a strong emphasis on throughput, many real-time inference workloads introduce greater sensitivity to latency, data access, memory performance, and response time.

With training, the emphasis is on high throughput... With AI inference, there's also a critical emphasis on latency," explains Griffin Lacey, Head of Solutions Architecture at NVIDIA Canada.

Mahadev Konar, SVP of Infrastructure at Together AI, points to how this shift is sharpening hardware design: "Training requires the interconnect to be much more throughput-oriented... inference requires much more latency-sensitive workloads."

This is exactly the kind of evolution the strongest facilities are already leaning into, and the ones that adapt their architecture for latency alongside throughput are well positioned to lead the market as it moves further into the inference era.



> With training, the emphasis is on high throughput... With AI inference, there's also a critical emphasis on latency.
>
> **Griffin Lacey**
> Head of Solutions Architecture
> NVIDIA Canada

NVIDIA

The warmer the liquid gets, the more efficient it is to reject that heat to the outside air.

Jim Simonelli
Chief Technology Officer

Schneider Electric



Smarter cooling, stronger performance.

Cooling architecture is turning into one of the biggest performance levers the industry has. High-temperature direct liquid cooling can support greater rack densities and reduce mechanical cooling requirements when facility design, climate conditions, and equipment specifications are properly aligned.

Thermal management has evolved from a facilities maintenance issue into a major driver of compute density and economic viability, and the leaders getting it right are proving just how much upside is on the table.

"The use of liquid cooling in this space really is around kind of when you start exceeding at 30 to 40 kilowatt barrier, where it becomes impractical, inefficient to blow air through a server," explains Jim Simonelli, CTO of Schneider Electric. Simonelli points to a promising development from NVIDIA which has discussed supplying water at 45°C (113°F): "In that type of environment you almost don't need chillers anymore in that space."

"The warmer the liquid gets, the more efficient it is to reject that heat to the outside air," notes Simonelli, underscoring how much smarter these systems are becoming.

Guillaume Lemieux, Director of Engineering at 5C, highlights the direct business impact this is having: "By removing the chiller from the design, it allowed us to reduce that mechanical load drastically, which allowed us to power on more racks for the same footprint."

These advances demonstrate how high-temperature liquid cooling can improve both infrastructure density and operating efficiency in appropriately designed facilities.

Speed is becoming a builder's biggest advantage.

Builders are rising to meet one of the industry's most exciting challenges: keeping construction in step with the pace of AI innovation.

To do it, they're compressing timelines through an integrated approach, advancing design, procurement, prefabrication and construction in parallel, while maintaining the flexibility to adapt as AI requirements evolve.

Construction timelines that once stretched over 18–24+ months has to evolve fast to keep up with an AI innovation cycle that moves quickly, and the industry is responding with real ingenuity. If a building is designed perfectly for today's highest-end GPU, there's a risk it falls behind by the time the ribbon is cut, so the strongest builders are designing for what's next, not just what's now. The simultaneous global surge in demand has also put pressure on supply for equipment ranging from switchgear to backup generators, pushing the industry toward smarter, more resourceful solutions.

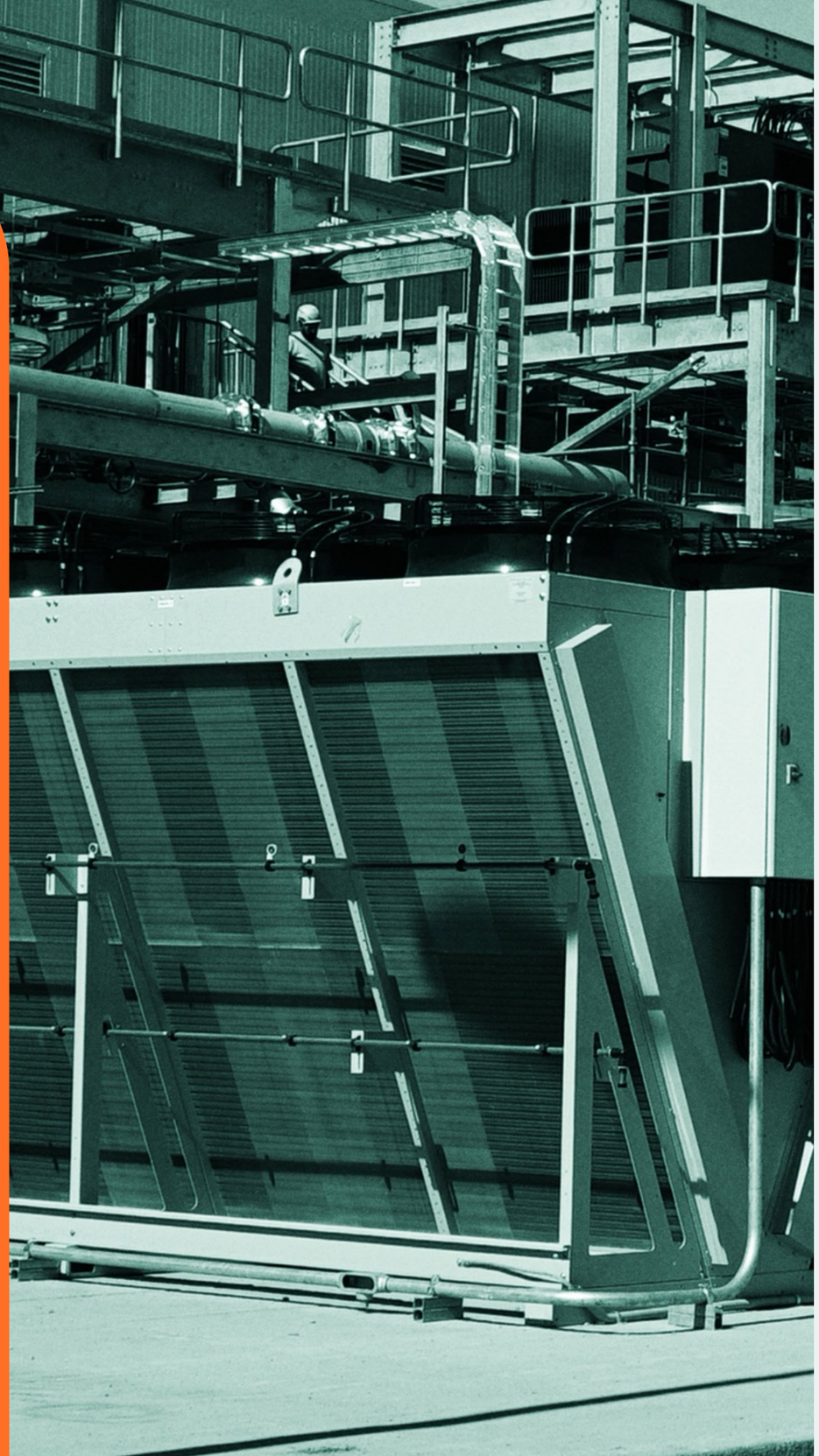
"If you think about a multi-year horizon halfway through your project, the entire landscape has changed," notes Simon Ahdoot, CEO of Hypertec, a reminder of just how much foresight this industry now requires and how well its best builders are adapting to it.

On the construction side, Eliot Ahdoot, President of Hypertec Construction, points to a smarter shift away from traditional sequential building: "Taking away labor from the site and doing it in factories helps tremendously on quality and on speed."

Future-proofing is becoming less of a luxury and more of an advantage. Data centers benefit from being deployed modularly, using adaptable designs that can accommodate changes in rack form factors and power densities as projects evolve, without tearing down the existing structure, giving builders a clear path to stay ahead of the curve rather than chasing it.



> If you think about a multi-year horizon halfway through your project, the entire landscape has changed.
>
> **Simon Ahdoot**
> CEO
>
> Hypertec



How do you achieve sustainability in data centers? You have megawatts and megawatts of heat that can be reused.

Eliot Ahdoot
President

Hypertec

Sustainability is becoming a source of shared value.

AI factories are opening up an exciting new role for themselves in the communities they operate in: not just consumers of power, but genuine contributors to the local ecosystem.

A high-density AI factory generates significant thermal energy, creating opportunities to recover and reuse that heat where local infrastructure, demand, and project economics allow.

This is driving the industry toward greater resource circularity, and traditional sustainability metrics like PUE (Power Usage Effectiveness) are evolving to keep pace with just how much more these facilities are now capable of giving back. As AI factories continue to scale, they're increasingly positioned not as isolated draws on the grid, but as active, integrated participants in the physical and economic ecosystem around them, a shift that's turning sustainability from a compliance checkbox into a genuine source of shared value.

Eliot Ahdoot puts it well: "How do you achieve sustainability in data centers? You have megawatts and megawatts of heat that can be reused." He points to some genuinely exciting applications already taking shape: "Carbon capture... desalination... hydrogen generation. We can take greenhouses and farm across these large swaths... allowing 365 farming."

This points to a real opportunity for the industry. Facilities that engineer circular resource loops help data centers become more integrated contributors to local energy and industrial ecosystems by generating tangible value for the communities around them. As AI infrastructure continues to mature, community integration is becoming one of its strongest calling cards. Operators are exploring practical ways to recover and reuse waste heat thus helping advance a more resource-efficient and locally integrated model for data center development.

The permanent foundation.

The future of AI may be digital, but it's being built on remarkably solid physical foundations, and that work is already well underway.

We're witnessing an exciting shift from piecemeal data centers to more integrated AI factories, a transformation that's setting the stage for everything AI will become over the next decade.

The progress of AI over the next decade will depend in large part on the infrastructure decisions being made today, from power and cooling to the physical systems required to support compute at scale. The builders who can translate rapid hardware advances into reliable, scalable infrastructure are doing more than providing capacity, they're helping determine how quickly AI moves from possibility to practical use.

Acknowledgments.

This report was made possible by the insights of sixteen remarkable leaders who shape the physical reality of artificial intelligence, engineers, operators, builders, and strategists working at the exciting intersection of compute,power, and infrastructure.

Their expertise and willingness to speak openly about what it takes to build at this scale made this work possible. We are deeply grateful.

Contributors.

Jim Simonelli
CTO
Schneider Electric

Alon Gavrielov
VP Infrastructure Strategy
together.ai

Mahadev Konar
SVP Engineering
together.ai

Alon Horev
Co-Founder and CTO
VAST

Griffin Lacey
Head of Solutions Architecture, NVIDIA Canada
NVIDIA

Simon Ahdoot
CEO
Hypertec

Don Schlidt
President, HPC & AI
Hypertec

Mike Marracino
President, HSP
Hypertec

Eliot Ahdoot
President, Construction
Hypertec

Brad Nimri
Chief Revenue Officer
5C.

Michael Heath
SVP Operations
5C.

Tanya Naami
Director of Customer Success
5C.

Guillaume Lemieux
Director of Engineering
5C.

Nathalie Bergeron
Director of AI Product & Marketing
5C.

Darcy Letemplier
VP Hardware Engineering
5C.

Closing.

The infrastructure decisions being made today will help shape the trajectory of artificial intelligence for the next decade, and that's an exciting thing to witness.

We hope this report serves as a valuable and inspiring lens for those navigating the same terrain.

Contact us at: 5C.ai | **Follow us on:**

Media Resources

Access approved 5C brand resources, including logos, color palettes, typography, and visual identity guidelines for consistent representation across all media and communications.

Find out more

Media contact

Nathalie Bergeron
Director, AI Product and Marketing

© 2026 5C AI Group Inc. All rights reserved.