

Interview

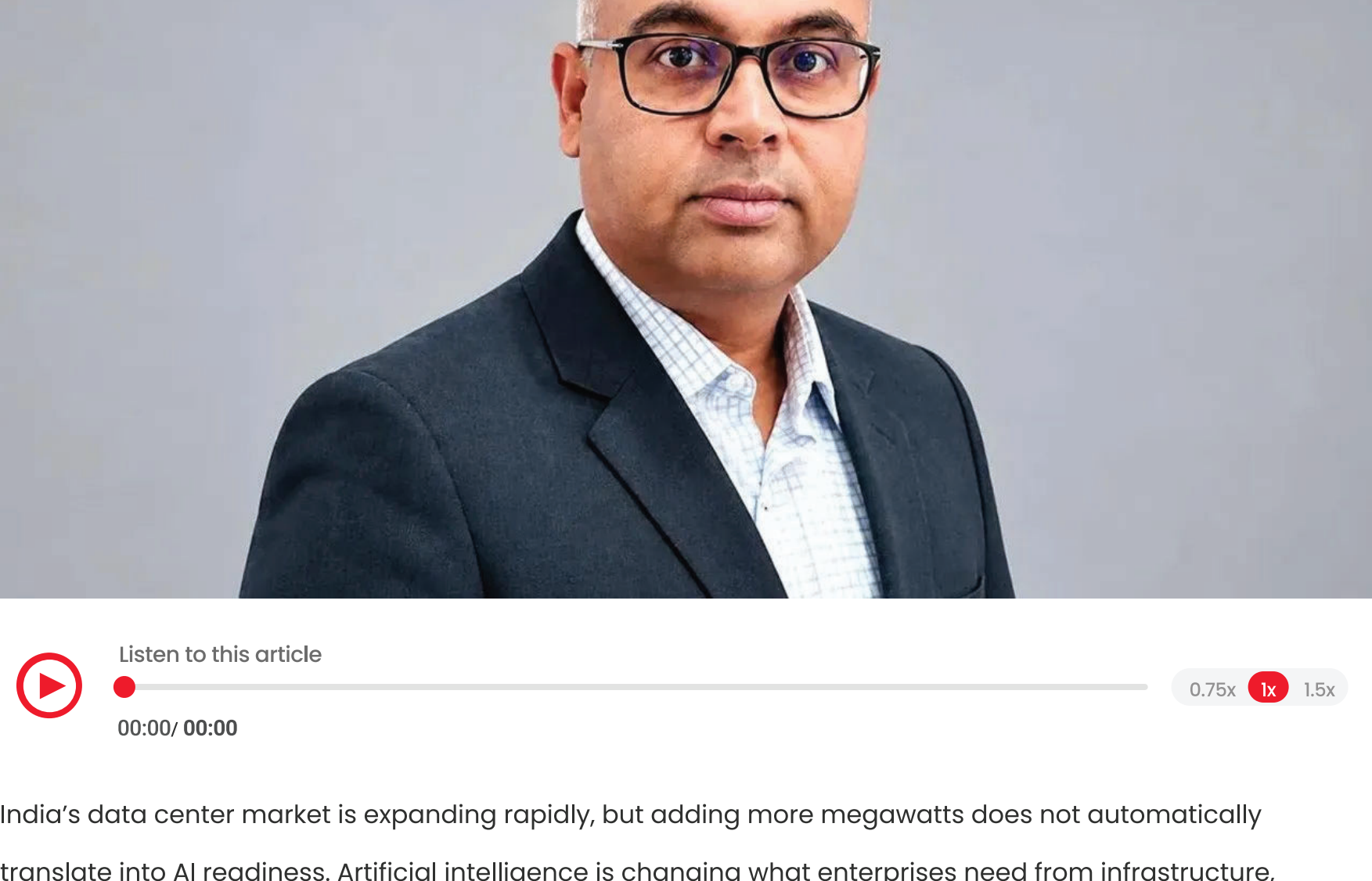
India’s AI challenge is not just GPUs, but turning compute into value

Swastik Chakraborty, VP Technology at Netweb Technologies, explains why India's AI infrastructure debate needs to move beyond GPU capacity to business outcomes, sovereign compute, power, cooling and accessible AI infrastructure.

Manisha Sharma

28 Sep 2026 17:51 IST

Follow Us



Listen to this article

00:00/00:00

0.75x1x1.5x

India's data center market is expanding rapidly, but adding more megawatts does not automatically translate into AI readiness. Artificial intelligence is changing what enterprises need from infrastructure, pushing the conversation beyond rack space and conventional compute towards high-density processing, specialised accelerators, storage, networking, cooling, and power.

More importantly, enterprises themselves are beginning to ask a different question: not simply how much AI infrastructure they can acquire, but what business value that infrastructure can generate.

According to Swastik Chakraborty, VP of Technology at Netweb Technologies, the industry is moving from a “time-to-market” mindset towards “time-to-success.” That transition could determine how effectively India converts its growing data center and compute capacity into a broader AI economy.

In an interview with **Dataquest**, Chakraborty discusses why GPUs are only one part of the compute equation, what sovereign compute should mean for India, why power could shape the next generation of AI data centers, and what it will take for India to move from consuming global AI infrastructure to building capabilities for the world.

India is adding significant data center and AI infrastructure capacity. But what does building a real compute strategy require beyond GPUs and servers?

Computing is much broader than one particular type of processor or accelerator. We have CPUs, GPUs, FGAs, and other accelerators, and each serves different kinds of computing requirements.

The larger change happening today is around how compute, storage, and other parts of the infrastructure work together. Earlier, the discussion was largely about bringing more storage closer to compute. But as computing cores become more powerful and data requirements increase, the ecosystem needs to deliver value as a whole rather than relying on one specific accelerator or computing architecture.

AI is obviously driving a significant amount of compute adoption. Within AI itself, there are three broad requirements: training, fine-tuning, and inference. Each has a different infrastructure requirement depending on the size of the model, the amount of enterprise data involved, and the scale at which the service needs to be delivered.

But AI is still only one part of computing. Enterprises continue to need infrastructure for databases, web applications, application servers, and high-performance computing. The bigger question is how much business value an organisation gets from its investment. If an enterprise invests ₹100 crore in infrastructure, what business value does that generate?

We are not completely there yet because every technology goes through a maturity cycle. The challenge is not simply infrastructure availability. It is about making that infrastructure investment deliver meaningful business outcomes.

Has the industry focused too heavily on GPUs when discussing AI infrastructure?

GPUs have become almost synonymous with AI because of the way modern AI models have evolved. They offer large numbers of computing cores and memory, and much of the model development and training ecosystem has been built around them.

But the ecosystem is gradually expanding. There are neural processing units, LLM processing units, and other specialised accelerators being developed for particular workloads. Models can also run on CPUs and other processing architectures depending on the requirement.

The challenge is that customers already have familiarity, expertise, and software ecosystems around GPUs. Moving to another specialised accelerator requires additional time, skills, and investment. That makes GPUs a relatively safer investment for many customers because they can use them across training, fine-tuning, and inference.

Going forward, however, customers will increasingly ask whether GPUs are the only solution for their AI workload or whether another accelerator could provide lower power consumption, better efficiency, or stronger alignment with the business requirement. That is where the industry is heading. It will not necessarily be one solution for every workload.

What should “sovereign compute” actually mean for India?

Sovereignty initially started with the idea of data locality—keeping your data within the country. But it is no longer only about where the data resides. It now extends across hardware, software, solutions, models, and data.

Two important elements are our model and our data. Foundation models developed for India need to understand Indian data and requirements. Take healthcare as an example. Disease patterns and datasets in India can be different from those in the US, UK, China, or other markets.

India also has enormous diversity in its datasets. So owning and developing the data ecosystem becomes an important part of sovereignty. But sovereignty goes beyond data and models. The question is whether Indian organisations can build the infrastructure, deploy the software, create solutions, and deliver outcomes such as inference and fine-tuning for public and enterprise use cases.

Wherever there is complete dependence on a third party, there is also a dependency that needs to be understood. So I see sovereignty as a much broader concept. It is not only about manufacturing. It is about building enough capability across hardware, software, datasets, models, and solutions to reduce strategic dependencies.

Which parts of that stack should India build domestically, and where will global dependence remain?

Today, India continues to source high-end processors and GPUs from global companies such as Intel, AMD, Qualcomm, and NVIDIA. But a chip does not operate in isolation. You still need to design motherboards, daughter cards, power-conditioning systems, and complete server platforms. Indian companies can design, manufacture, and build those systems domestically even when some components are sourced globally.

Over time, as Indian semiconductor capabilities develop, domestic processors can also become part of these platforms. Software is another important area. We can build software around Indian datasets, including data ingestion, processing, governance, and lineage. On top of that, we can develop or fine-tune models using our own data.

India is also strong in translating business objectives into software and increasingly into AI agents. That is an area where we can build indigenous capabilities without necessarily depending on foreign software. So, at present, our dependence is greater at the high-end processor and GPU layer, while there is considerable opportunity to build sovereign capabilities across software, services, data, and solutions.

How has the enterprise AI buyer changed over the past 12 to 18 months?

I see two major changes. Earlier, customers talked about time to market. Today, they increasingly talk about time to success. Time to market is largely about acquiring technology, experimenting with it, and understanding how it can be used. Time to success asks a different question: how quickly can that technology deliver a measurable business outcome?

The second change is that customers are moving beyond using AI simply for chatbots. They are looking at how AI and agentic AI can transform actual business processes. That changes the infrastructure discussion as well. Customers now need to decide whether they should use an existing model, fine-tune a model, or deploy dedicated infrastructure for a particular requirement.

They also want to understand the economics: How many customers can this infrastructure serve? What is the cost per customer? What is the inference cost? How does that justify the infrastructure investment? AI infrastructure is therefore becoming increasingly tied to specific business requirements rather than experimentation alone.

Are enterprises overprovisioning AI infrastructure because they are planning around GPUs rather than workloads?

The problem occurs when organisations cannot connect infrastructure investment with business success.

Our approach is to move from ideation to iteration to implementation as quickly as possible. The faster an organisation completes that cycle, the easier it becomes to understand the business value and justify further investment.

Enterprises should identify two or three important business problems, establish their current baseline, get leadership sponsorship, and then demonstrate value. If organisations remain in the experimentation phase without moving towards implementation, infrastructure investment becomes difficult to justify.

CFOs, CIOs, and CEOs ultimately want to understand what an investment will deliver. If an organisation invests ₹100 crore, leadership wants to know how that investment contributes to business growth, productivity, or another measurable outcome. AI adoption therefore needs to move alongside business transformation and employee adoption rather than simply increasing access to GPU resources.

Could power availability become a bigger constraint than GPU availability for India?

Power is certainly becoming one of the critical considerations. The challenge is not necessarily India's overall ability to generate power. The issue is whether sufficient power is available at a particular location, along with the required transmission infrastructure and redundancy.

AI data centers can require very high power density. If you cannot provide enough power at one location, it limits how much compute—and consequently how many GPUs—can be deployed in a single cluster.

One approach we are seeing is the concept of federated data centers, where infrastructure is distributed across multiple locations rather than trying to build all the required capacity at one site. We are also seeing early discussions around alternative and redundant energy sources, including combinations of solar and wind, alongside developments in backup power.

Cooling is equally important. As compute density rises, direct liquid cooling will become increasingly relevant because traditional air cooling alone may not be sufficient for future high-density systems. So power availability, cooling, and data center architecture increasingly need to be considered together.

Should future AI data centres follow power availability rather than traditional data centre hubs?

Power availability has to become an important factor in deciding where future compute infrastructure is built. At the same time, you also need to consider cooling, connectivity and other infrastructure requirements. The challenge is getting sufficient power at a single location while maintaining the redundancy required for data centre operations.

As compute becomes denser, liquid cooling and federated data centre architectures can help address some of these constraints. The design of IT infrastructure and the design of the data centre can no longer be treated as completely separate decisions. What you deploy inside the facility increasingly influences how the facility itself needs to be designed.

How can India make AI compute accessible beyond large enterprises and hyperscalers?

This is where initiatives such as the IndiaAI Mission become important. Shared compute infrastructure can give startups, researchers and institutions access to AI resources without requiring every organisation to make a large upfront infrastructure investment.

But democratising AI cannot depend only on providing everyone with access to the highest-end GPUs. There are workloads that can run on CPUs, smaller clusters and alternative accelerators. Small language models, for example, may not require the same infrastructure as very large foundation models.

Organisations should be able to experiment, develop applications and validate their business case on smaller infrastructure. Once they establish that the application delivers value, they can move it to larger systems.

We also need to develop AI skills at universities, colleges and research institutions. Access to compute is important, but people also need to understand data, AI and how these technologies can be applied to real problems. That broader ecosystem is necessary if we want to democratise AI.

What would it take for India to move from consuming AI compute to exporting AI capabilities?

It requires the entire ecosystem. At Netweb, we describe our approach as “India for India and India for the world.” The first step is to build what India needs. Once those capabilities are demonstrated domestically, they can potentially be offered globally.

That opportunity is not limited to shipping hardware. Today, India consumes models and AI services developed elsewhere. In the future, we should also be able to develop models here, host them on infrastructure in India and make those capabilities available globally through APIs and inference services.

Training a model happens periodically, but inference is where massive ongoing consumption occurs once applications scale. So India can participate in two ways. We can manufacture and export infrastructure, but we can also build infrastructure in India on which globally consumed AI models and inference services run.

That includes hardware, data platforms, cooling, data centre design, models and the software layer. The opportunity is therefore much broader than simply assembling AI servers.

What is the biggest capability India still needs to build?

At the high-end compute level, we continue to depend on global processor and GPU companies. That dependency will not disappear immediately. But we should not assume that every problem necessarily requires a GPU. The first step should be identifying the business problem and then deciding what computing architecture is best suited to solve it.

For some workloads, GPUs will be essential. For others, specialised accelerators or alternative computing architectures may be more efficient. That is why India needs to look at AI infrastructure from the perspective of business outcomes rather than simply measuring the number of GPUs available. If we cannot align AI infrastructure with specific business value, we risk missing an important part of the AI transformation.

ai

datacenters

Netweb Technologies

Latest Stories

RBI Data Centre in Bhubaneswar: How will it protect Digital payments?

Jensen Huang calls AI as a five-layer cake: Nvidia GTC Berlin

Synopsys at Nvidia GTC Berlin: Delivering engineering precision with speed

Subscribe to our Newsletter!

Be the first to get exclusive offers and the latest news

Your Name

Email address

Subscribe Now