

India's silent AI revolution: Infra fuels the new digital core

As India's digital ambition meets reality, enterprises are quietly turning to robust AI infrastructure as the backbone of strategic transformation.



BY SHUBHENDU PARTH & DHIRENDRA PRATAP SINGH

A quiet force is reshaping India's enterprise landscape—not with the hype of chatbots or robotic automation, but with the rigour of computation, the elegance of architecture, and the foresight of system design. At the heart of this transformation lies artificial intelligence (AI), yes—but

it is the infrastructure powering it, which is driving real, scalable, and strategic impact across industries.

While AI dominates boardroom discussions and press announcements, it is the invisible but intelligent infrastructure—spanning compute, storage,

Today's AI infrastructure blends multi-GPU servers, NVMe storage, TPUs, and high-speed interconnects, managed by software like Kubernetes and MLOps tools.



India's AI infrastructure is evolving from cloud consumer to sovereign creator—powered by innovation, policy-driven scale, and ethical foresight.

HIRDEYVIKRAM
CMO & Senior VP, Netweb Technologies

and connectivity—that is quietly becoming the real differentiator in enterprise success. The GPU clusters crunching health data in hospitals, the edge servers humming silently in factories, and the hybrid clouds calculating fraud probabilities at banks—these are the unsung workhorses of India's digital transformation.

FROM HYPE TO HABITAT: INDIA'S AI EVOLUTION

India's AI adoption is entering its decisive phase. According to the EY-NAASSCOM AI Adoption Index, nearly 65% of enterprises in India have moved beyond pilot experiments to live AI deployments. Another 71% plan to ramp up AI spending over the next 12 months. This wave spans sectors—from technology and telecom to BFSI and retail, and healthcare and manufacturing—each betting on AI to redefine operations.

But this transformation cannot happen in isolation.

"India's AI infrastructure is evolving from cloud consumer to sovereign creator—powered by indigenous innovation, policy-driven scale, and ethical foresight," says Hirdey Vikram, CMO and Senior VP, Netweb Technologies. He points to platforms like Skylus and strategic investments in data centres and semiconductor ecosystems as early signs that India is building its own AI core.

The change is visible. IDC estimates AI spending in India will touch USD 5 billion by 2027, growing at a CAGR

IN BRIEF

- AI infra now drives India's enterprise shift from pilots to full-scale, cross-industry deployment with real-time and edge-ready capabilities.
- Cloud-native, hybrid, and thick-stack systems are replacing legacy IT, enabling AI-driven insights with better speed, compliance, and scalability.
- Infrastructure design is now deeply aligned to use cases, from LLM clusters in metros to edge AI in Tier-2 industries and agri-tech hubs.
- Private and hybrid clouds are rising fast amid new data protection laws, driving a balance of sovereignty, speed, and sustainability.
- MLOps practices are taking centre stage, improving lifecycle management and bridging gaps between innovation, governance, and ROI.
- Modular infrastructure with GPUs, CPUs, and AI accelerators is reducing inference latency, enhancing analytics, and boosting cost control.



Enterprises are moving beyond traditional cloud architectures toward GPU-accelerated, modular environments tailored for demanding AI workloads.

ARJUN NAGULAPALLY
Chief Technology Officer, AlonOS



Scaling AI infrastructure requires a phased approach—combining automation with human judgement to build trust at every level.

ASHISH ARORA
CEO – Nxtra by Airtel



To understand the value of AI infra, measure how it speeds up decisions, improves efficiency and service—TCO and Return on AI show its long-term impact.

JASPREET SINGH
Partner, Grant Thornton Bharat

of 30%. A significant chunk of that is earmarked for infrastructure, marking a shift in how enterprises view AI not as an app but as a foundational capability.

AI INFRA'S REAL-WORLD IMPACT ACROSS INDUSTRIES

Across India, the business impact of AI infrastructure is tangible. A multi-specialty hospital chain in Bengaluru uses GPU-powered compute clusters to interpret MRI scans in under five minutes. In Surat, textile factories have embedded edge-AI systems to monitor machine vibration, cutting downtime by 40%. And in the financial corridors of Mumbai, banks like HDFC and SBI are deploying real-time AI models to sniff out fraud and assess credit risk, supported by compute stacks that analyse millions of transactions in milliseconds.

"Enterprise infrastructure is undergoing a profound transformation—from siloed systems to intelligent, agentic platforms," says Raj K Gopalakrishnan, CEO and Co-founder, KOGO AI. "Organisations now demand thick-stack, low-code platforms that unify data, automate workflows, and scale seamlessly across environments."

INSIDE THE DATA CENTRE: AI REINVENTS THE CORE

The application of AI in data centre infrastructure is no longer aspirational—it is foundational. Airtel's data centre arm, Nxtra, for instance, exemplifies how intelligent infrastructure is being designed not just to support

AI but to evolve with it: its facility in Siruseri, Chennai, has deployed AI not just as a monitoring layer but as an active orchestrator of performance, sustainability, and operational agility.

Thousands of embedded sensors continuously track variables such as cooling, airflow, and power consumption. Using platforms like Ecolibrium's SmartSense, Nxtra's AI systems analyse this real-time data to make dynamic adjustments, optimising energy use without compromising uptime. This operational intelligence translates into a tangible impact: lower energy consumption, enhanced equipment lifespan, and smarter automation across systems.

"We are truly differentiated in our use of AI for energy efficiency," said Ashish Arora, CEO – Nxtra by Airtel. "It is not just about cost savings—it is about delivering a sustainable, high-performance digital backbone."

The implementation is already delivering measurable returns. Early estimates show a 10% reduction in non-IT power usage and a 10% increase in asset life, thanks to predictive maintenance capabilities that flag subtle performance degradation before it escalates into failure. Productivity, too, is on the rise—with automation allowing teams to focus on high-value functions rather than manual oversight. In some operations, the company anticipates up to a 25% increase in productivity.

But scaling AI across distributed infrastructure is not without complexity. Thousands of interlinked devices, varied data formats, and real-time decision requirements demand high model accuracy and rigorous human oversight. "AI is only as good as the data it learns from," Arora said. "Scaling AI infrastructure requires a phased approach—combining automation with human judgement to build trust at every level."

As AI takes on more responsibility, the role of human teams at Nxtra is evolving—from operational execution to model governance and strategic oversight. Engineers are now training and tuning AI models, ensuring alignment with security protocols, sustainability goals, and customer expectations. In essence, AI is not replacing people—it is amplifying their impact.

Nxtra's model is a microcosm of what AI-enabled infrastructure can offer the broader data economy. In a country where digital demand is outpacing traditional capacity, this approach delivers both resilience and responsibility. As AI continues to become core to digital infrastructure, Nxtra's strategy signals the direction of travel—toward greener, smarter, and more responsive data centres that power the next generation of enterprises.

EXPANDING THE ENTERPRISE AI TECH STACK

The infrastructure that supports AI today is no longer just racks of servers in cold rooms. It is a mix of multi-

GPU servers, NVMe storage, TPUs, and high-bandwidth interconnects, all orchestrated by software layers that include Kubernetes, PyTorch, TensorFlow, and Machine Learning Operations (MLOps) pipelines.

This new stack is deeply use-case specific. While large language models (LLMs) are trained in centralised clusters, their inference—particularly in industrial contexts—is pushed to the edge for low-latency response. One Surat-based textile company, for instance, deploys ruggedised edge servers to monitor machinery health and trigger predictive maintenance alerts.

"Organisations embracing GenAI and real-time analytics are reengineering their systems for performance, security, and sustainability," notes Pankaj Vyas, Head – Security and Surveillance, Syntel by Arvind. "High-performance computing, hybrid/multi-cloud setups, Edge AI, and automated CI/CD pipelines are now table stakes for AI at scale."

Regulatory shifts like the Digital Personal Data Protection Act (DPDPA) are also influencing infrastructure decisions. Private and hybrid clouds are gaining preference as enterprises seek control over data while optimising for performance and compliance.

"We are seeing a decisive move toward distributed, intelligent architectures," says Umesh Shah, Whole-Time



Legacy integration, compute scaling, and skill gaps are major hurdles in AI infra migration—demanding a phased, outcome-driven modernisation approach.

GANESH GOPALAN
Co-Founder & CEO, Gnani.ai



Intelligent data fabrics with vector databases, real-time ingestion, and high-speed pipelines are replacing data lakes to power AI at scale.

VIJAY KUMAR
Head – Engineering & Product Development, NeevCloud



Modern AI infrastructure is shifting from static stacks to intelligent orchestration—optimising cost, performance, and scalability in real time.

KUMAR VIKAS
Executive Vice President, Data & AI, Bounteous

Director, Orient Technologies. "Enterprises are not just using AI—they are building it."

INVESTMENTS DELIVERING MEASURABLE OUTCOMES

This reorientation towards intelligent infrastructure is delivering tangible business outcomes. Companies report model training times cut by 30–50%, data processing speeds up by 60%, and up to 40% cost reduction through AI-optimised workflows.

Consider the example of a logistics firm in Mumbai. With infrastructure support from Netweb, it implemented a GPU-based AI system to optimise delivery routes. The result: an 18% cut in fuel costs and a 22% improvement in delivery times. For today's CIOs, AI infrastructure is no longer an IT concern—it is a business metric.

However, challenges persist. Over 60% of companies surveyed by NASSCOM cite infrastructure cost and talent scarcity as critical bottlenecks. High-end GPUs and specialised accelerators do not come cheap. Skilled talent to manage these systems is also in short supply. "Migrating from legacy IT to AI-ready systems is not just technical—it is a strategic overhaul of infrastructure and talent," the report notes.

Compliance adds another layer of complexity. In sectors like healthcare and BFSI, infrastructure

must be resilient, traceable, and in step with evolving data regulations.

BHARAT'S TURN: AI SPREADS BEYOND METROS

One of the quiet but more promising trends is the expansion of AI beyond urban centres. From Tier-2 hospitals to agritech startups in Punjab and manufacturing units in Coimbatore, the demand for low-latency, cost-efficient AI solutions is growing.

Here, edge computing and hybrid clouds are playing a starring role—delivering intelligent processing closer to the source of data, while maintaining the flexibility and scale of cloud systems.

Ved Antani, SVP Engineering and MD, New Relic India, puts it succinctly: "Enterprises are rapidly shifting to cloud-native, data-centric, and hybrid infrastructure to power GenAI and real-time analytics—driven by scalability, compliance, sustainability, and faster decision-making."

Indian OEMs and startups are stepping up, designing platforms tailored to local constraints—power fluctuations, thermal stress, and budget limits. These platforms support both open-source frameworks and enterprise-grade integrations, bridging the affordability-accessibility gap.

The next phase of AI infrastructure will likely see standardisation of AI frameworks for MSMEs, growth in



Hybrid cloud empowers enterprises with control, low-latency AI processing, and cost efficiency—balancing data security and scalability.

ANSHUL BHIDE
Director & AI/ML Practice Head, Calsoft

domestic data centres, and the rise of India-hosted AI-as-a-Service models. According to experts, this will be key to inclusive, scalable transformation.

SCALING MLOPS AND MODULAR DEPLOYMENT MODELS

To deliver continuous value, MLOps is becoming a critical practice. Efficient lifecycle management of AI models—from training and validation to deployment and monitoring—is now a board-level agenda.

However, roadblocks remain. R&D in energy-efficient hardware is limited, there is a lack of standard enterprise adoption frameworks, and academia-industry-policy collaboration is still evolving.

Hybrid cloud, meanwhile, is emerging as the consensus favourite.

"Hybrid cloud delivers the best of both worlds—cost control, low latency, and data sovereignty," says Ramanujam Komanduri, Country Manager, Pure Storage India. "It empowers enterprises to run AI closer to data while scaling efficiently."

As Niraj Kumar, Chief Technology Officer (CTO), Onix, adds, "Industries like finance, healthcare, telecom, and retail are modernising AI infrastructure fastest—driven by regulatory needs, data complexity, and real-time insights."

And the underlying strategy is workload-based differentiation. "CPUs are better for everyday AI tasks, but GPUs are better for training large AI models quickly. For tasks that need high speed and efficiency at scale, we recommend using dedicated AI hardware like TPUs or other custom accelerators," says Jaspreet Singh, Partner, Grant Thornton Bharat.

A general rule of thumb, according to Srinath Venkatesh Nadkarni, SVP – Data and Analytics, Indium, is to use CPUs for prototyping and simple inference, GPUs for the majority of training tasks and in deep learning workloads, and AI accelerators for training, deployment, and inference workloads that demand scale and speed. "If you combine all three types of hardware, you can generally optimise overall cost, speed, and flexibility across a mixture of AI workloads."

AI INFRA STRATEGY: DRIVING BUSINESS ADVANTAGE

The silent success story of AI in India is not the flashy chatbot or voice assistant—it is the data pipeline, the inference engine, the Kubernetes pod running inference at midnight. It is also about the strategic design choices behind public versus private clouds, the nuanced understanding of latency, and the balancing act between compliance and agility.

At the heart of this transition is a deeper understanding of what truly drives enterprise performance. "To



Hybrid cloud offers a middle path—balancing regulatory compliance, control, and AI cost-efficiency where data localisation and scale both matter.

VIKRAM JEET SINGH
Partner, BTG Advaya



If you combine CPUs, GPUs, and accelerators, you can generally optimise overall cost, speed, and flexibility across a mixture of AI workloads.

SRINATH VENKATESH NADKARNI
SVP – Data & Analytics, Indium



As AI becomes operational, enterprises need infrastructure that can reduce handoffs, shorten iteration cycles, and scale horizontally without rework.

NIRAJ KUMAR

Chief Technology Officer, Onix



AI infrastructure boosts efficiency, accelerates insights with real-time analytics, and scales on demand—delivering measurable ROI and long-term value.

GIRISH SHARMA

Director – Service Offer Management, NTT DATA GDC & Cloud Infrastructure

understand the value of AI infrastructure, look at how much money it saves and how it improves everyday operations,” said Singh. “Key things to track include quicker decisions, increased efficiency, and better service for customers.”

That value is becoming more measurable. Enterprises are seeing that infrastructure directly influences the speed of insight, resilience of operations, and strategic differentiation. Girish Sharma, Director – Service Offer Management at NTT DATA Global Data Centres and Cloud Infrastructure, added further, “Modern AI infrastructure boosts efficiency through automation, accelerates insights with real-time analytics, and scales on demand—delivering both measurable ROI and long-term value.”

The ROI conversation is also shifting from cost savings to innovation velocity. According to Arjun Nagulapally, CTO at AlonOS, the real payoff lies in continuous value delivery. “AI infrastructure ROI goes beyond cost savings—true impact lies in measurable business outcomes, continuous value delivery, and the ability to scale intelligent systems that drive innovation, trust, and transformation,” he said.

This ability to scale responsibly is particularly important as enterprises embrace more decentralised architectures. For many, hybrid cloud is emerging as the model of choice, balancing performance, governance, and

cost. “Hybrid cloud empowers enterprises with control, low-latency AI processing, and cost efficiency—balancing data security and scalability,” said Anshul Bhide, Director and AI/ML Practice Head, Calsoft.

The next generation of AI-ready infrastructure is modular, cloud-native, and increasingly open. It is designed to deliver performance at scale and also meet the demands of compliance-heavy industries, sustainability mandates, and edge-first use cases. Enterprises are investing in systems that are flexible enough to support evolving workloads—from training large models in centralised clusters to pushing real-time inference to the edge.

“Enterprises are shifting to modular, cloud-native, and API-driven infrastructure—enabling secure, low-latency AI, faster innovation, and seamless integration of LLMs and real-time analytics,” said Ganesh Gopalan, Co-founder and CEO of Gnani.ai.

Today, the isolated projects or experimental pilots no longer define India’s enterprise AI story. It is being shaped by infrastructure choices that are quietly but powerfully transforming how businesses operate, innovate, and grow. And in this transformation, infrastructure is no longer a supporting actor—it is the strategic engine driving India’s digital future. 🌟

shubhendup@cybermedia.co.in