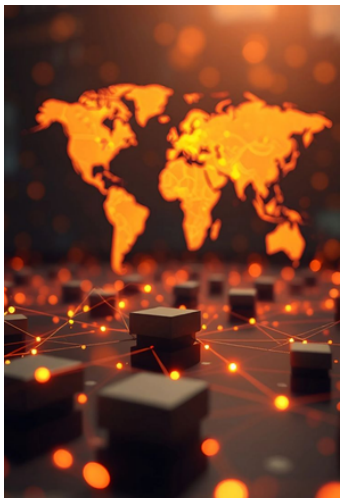
AI INFRASTRUCTURE CYCLE

INDIA'S STRATEGIC INVESTMENT OPPORTUNITY



Ellenox

ELLENOX Pvt. Ltd.

JUNE-8-2026

CONTENTS

ABOUT THIS REPORT

This report provides a comprehensive strategic analysis of **India's transition from a software-driven AI cycle to a full-scale industrial infrastructure transformation.** As global hyperscalers project investments exceeding *$660–750B in 2026*—with roughly 75% directed toward AI infrastructure—India is positioned to capture a massive opportunity fundamentally rooted in physical capacity expansion. By analyzing the ecosystem through the lenses of compute availability, power stability, and thermal management, this research identifies the critical shifts from generative AI model development toward capital-intensive hardware and execution systems. The findings underscore that India's AI scaling is no longer constrained by software demand but by how rapidly its underlying physical systems—specifically data centers, fiber backbones, and renewable energy grids—can deploy to meet accelerating transaction volumes.

Our structural assessment highlights that while India possesses a globally competitive talent pool and rapid digital adoption rates, the infrastructure ecosystem remains uneven and highly dependent on global semiconductor supply chains. With GPU lead times extending to 52 weeks and high-density AI racks requiring up to 160 kW per rack, **the report identifies power procurement and liquid cooling as the next primary strategic moats for long-term competitiveness.** The analysis concludes with a high-conviction investment outlook, asserting that durable value will increasingly accrue to those controlling the physical bottlenecks of the AI era—compute clouds, captive energy storage, and infrastructure-ready land corridors—rather than model creators alone.

1. EXECUTIVE SUMMARY

Artificial Intelligence is no longer just a software revolution... it is rapidly becoming a full-scale infrastructure transformation. Across the global economy, the AI race is increasingly being defined not only by model capability, but by access to compute, power, cooling, networking, data infrastructure, and execution capacity.

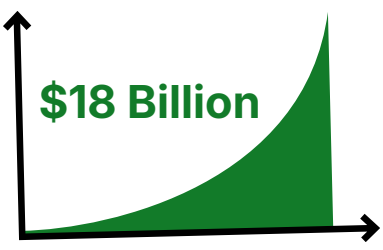
This report analyzes India's position within that transformation through three core lenses:

1. **Software Infrastructure Analysis**
2. **Hardware & Infrastructure Analysis**
3. **India Supply Map & Capacity Readiness**

The findings point to one central conclusion:
India's AI opportunity is fundamentally an infrastructure opportunity.

1.1 The Global Shift: From Models to Infrastructure

The early AI cycle (2022–2024) was dominated by generative AI applications and model development. However, by 2025, infrastructure spending reached near parity with application-layer spending. Enterprise AI infrastructure spending alone reached approximately $18B, representing nearly half of total enterprise AI investment.



AI infrastructure spending

At the same time, hyperscalers including Amazon, Microsoft, Google, Meta, and Oracle are projected to invest $660–750B+ in 2026, with roughly 75% directed toward AI infrastructure.

This shift reflects a new reality:

- AI is no longer constrained by software demand.
- It is constrained by physical infrastructure capacity

ORACLE Microsoft amazon Google Meta

1.2 India's Position

India is entering a large-scale AI infrastructure expansion phase, supported by:

- Rapid digital adoption
- Expanding data center ecosystems
- Large technical talent pools
- Lower infrastructure costs compared to global markets
- Strong government-led digital infrastructure initiatives.

India's operational data center capacity has already expanded from approximately 375 MW (2020) to ~1.5 GW (2025), with projections of 4–5 GW by 2030.
Simultaneously:

- Fiber backbone infrastructure crossed **4.24M route-km**
- Renewable energy capacity **surpassed 262 GW**
- AI/ML workforce reached **~2.75M professionals**
- Subsea bandwidth capacity grew to **309 Tbps**

However, despite rapid growth, India's infrastructure ecosystem remains uneven and highly dependent on global supply chains.

1.3 Key Structural Findings

Compute Is the Core Constraint

India's AI demand is accelerating faster than its computer availability.
While data center capacity is scaling aggressively, GPU availability remains heavily import-dependent. Advanced AI hardware... including GPUs, HBM memory, and advanced semiconductor packaging... continues to be controlled by a small number of global suppliers.
This creates a strategic vulnerability where:

- Demand exists
- Capital exists
- Talent exists
- But compute access remains constrained

The long-term importance of sovereign compute infrastructure, GPU clouds, and localized AI capacity will therefore increase significantly.

Power Is Becoming a Strategic Asset

High-density AI racks now require **$30\text{--}160\text{ kW}$ per rack**, far exceeding legacy infrastructure assumptions and making energy **30–40% of data center operating costs**. While India has sufficient national generation capacity, significant constraints remain regarding transmission bottlenecks, renewable intermittency, industrial tariffs, and storage gaps. Consequently, power economics have become central to competitiveness in the AI sector.

These challenges are accelerating a shift toward captive renewable energy, battery energy storage systems (BESS), hybrid energy architectures, and smart grid optimization. Because of these dynamics, energy infrastructure has emerged as one of the strongest long-term investment themes within the AI ecosystem.

Cooling Has Become a Primary Infrastructure Layer

AI workloads are fundamentally thermal workloads.
Traditional air-cooling systems are increasingly inadequate for modern GPU environments. Rack densities are rapidly moving from 5–15 kW toward 70–160+ kW, driving adoption of:

- Liquid cooling
- Immersion cooling
- Advanced heat exchange systems
- AI-optimized thermal management

Cooling is transitioning from a support function into a core determinant of AI scalability and operational efficiency.

Networking & Connectivity Are Emerging as Competitive Moats

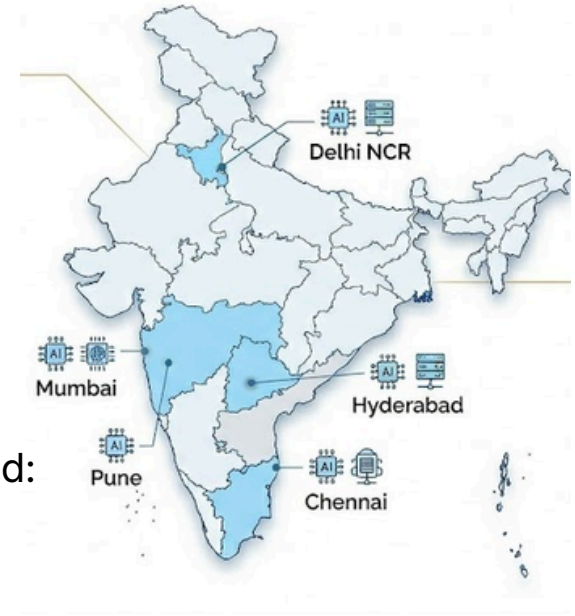
As AI becomes increasingly real-time and distributed, low-latency connectivity is becoming as important as compute itself.
India’s fiber backbone and subsea infrastructure are scaling rapidly, but important weaknesses remain:

- Concentration of internet exchanges in a few cities

- Limited Tier-2 interconnection ecosystems
- Dependence on international routing hubs
- Weak metro fiber depth in many regions
- Future AI competitiveness will increasingly depend on efficient movement of data-not simply compute ownership.

Infrastructure-Ready Land Is Scarce

India has large industrial land availability, but infrastructure-ready land remains limited.
Power access, fiber density, transport connectivity, and approval timelines now determine land value far more than raw acreage.



This has created premium digital corridors around:

- Mumbai
- Chennai
- NCR
- Pune
- Hyderabad

The report identifies industrial land corridors as a quiet but critical strategic asset class within India’s digital economy.

Services & Execution Remain India’s Largest Structural Advantage

India’s strongest long-term advantage remains execution capability.
The country already possesses:

- A globally competitive IT services ecosystem
- Large-scale engineering talent
- AI/ML workforce growth
- Cost-efficient operational scaling capability

As enterprises move from AI experimentation toward deployment, demand is shifting toward ***system integration, managed AI operations, and compliance and governance***. This shift also emphasizes ***infrastructure optimization and AI reliability engineering,*** positioning India extremely well within the global AI services and managed operations market.

1.4 Core Investment Conclusion

The report's central investment conclusion is clear:
The AI economy is becoming infrastructure-led, capital-intensive, and operationally complex.

Long-term value will increasingly accrue not simply to AI model creators, but to the companies controlling the hard constraints of the ecosystem:

- Compute infrastructure
- Power systems
- Cooling infrastructure
- Networking and fiber
- Industrial land
- GPU access
- AI operations and managed services

The strongest investment opportunities therefore sit in:

- AI-ready data centers
- Captive renewable energy & BESS
- GPU cloud infrastructure
- Cooling & thermal systems
- Fiber & edge connectivity
- Managed AI services
- Infrastructure-ready industrial corridors
- AI operations & reliability engineering

1.5 India's Infrastructure Imperative

India is not constrained by AI demand; it is constrained by how quickly physical infrastructure can scale to support that demand.
The next decade will not simply be a software cycle, but a national-scale infrastructure buildout across compute, power, networking, cooling, and execution systems.

Whoever controls the infrastructure bottlenecks will control the economics of India's AI future.

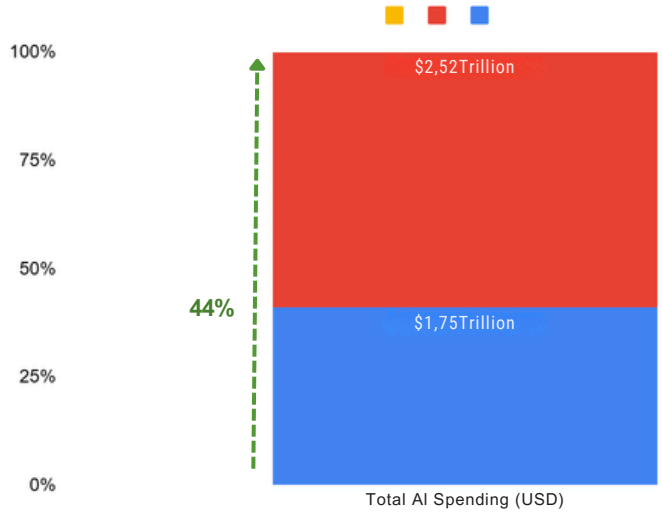
2. INDUSTRY STATUS QUO AND SHIFT

2.1 Investment Shift: Gen AI vs AI Infrastructure

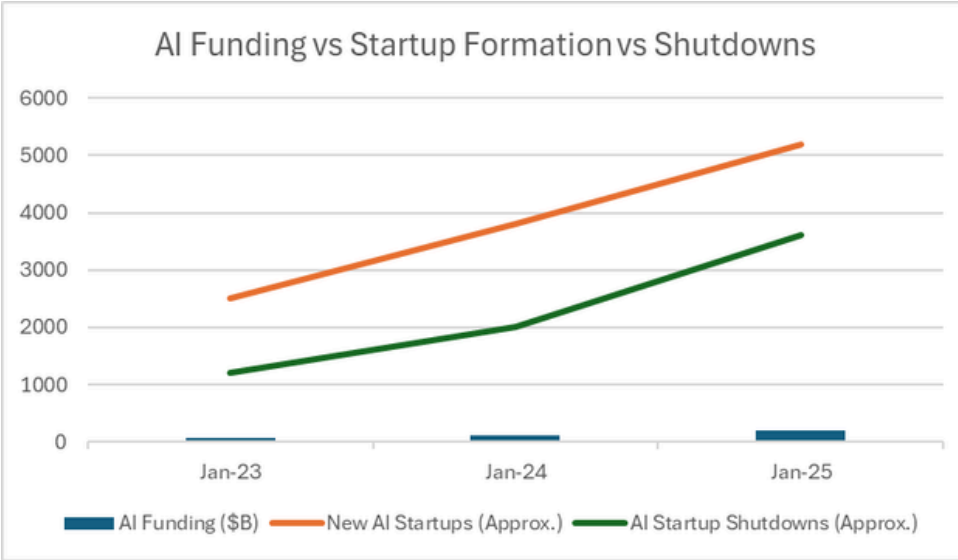
The AI market is no longer driven by models alone. By 2025, infrastructure has reached near-parity with applications, with ***$18B (49%) vs $19B (51%***) in enterprise spend. Both layers now co-drive growth.

A major driver is hyperscaler spending. Companies like Amazon, Alphabet, Meta, Microsoft, and Oracle are expected to invest ***$660–750B+ in 2026***, with ***75%+ directed to AI infrastructure,*** reflecting a 36–67% increase YoY.

On a global level, ***Gartner projects $2.52T total AI spending in 2026***, a ***44% YoY increase, alongside 49% growth in AI-optimised servers***.



Venture capital has followed aggressively. ***AI startups raised ~$114B in 2024 → ~$202B+ in 2025*** (75%+ growth), capturing ~50% of global VC funding.



(Capital is scaling rapidly- but so is failure, signaling a shift from growth to selection.)

In 2025 alone, AI accounted for 61% of total global VC investment (~$258.7B). Early-stage intensity has also increased, with 40%+ of seed/Series A rounds now exceeding $100M.
At the same time, startup formation has surged. Platforms like ***Y Combinator alone list 1,400+ AI startups,*** reflecting rapid ecosystem expansion.

AI Portfolio Companies / Rounds of the top investors

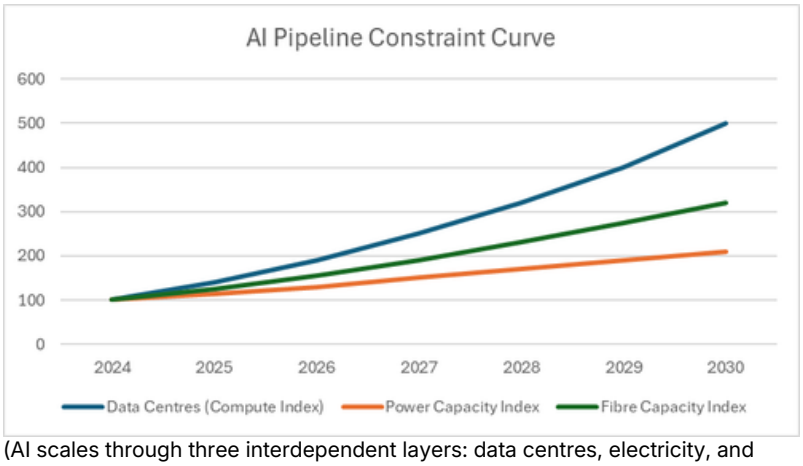
235 Y Combinator
148 Europa Capital
67 Andreessen Horowitz
65 Sequoia
52 NVIDIA
51 Lightspeed

However, failure rates are rising just as fast. ***~5,600 AI startups shut down across 2025–early 2026,*** implying **~40% failure** rates even among funded companies.

Supply remains constrained. NVIDIA controls ~90% of AI accelerators, with data centre revenue rising from $35.6B → $51.2B within three quarters. GPU lead times are 36–52 weeks.

To meet demand, McKinsey & Company estimates $5.2T in data centre capex by 2030, making this one of the largest investment cycles of the decade.

2.2 Why the AI Pipeline Matters



(AI scales through three interdependent layers: data centres, electricity, and fibre networks.)

DATA CENTRES

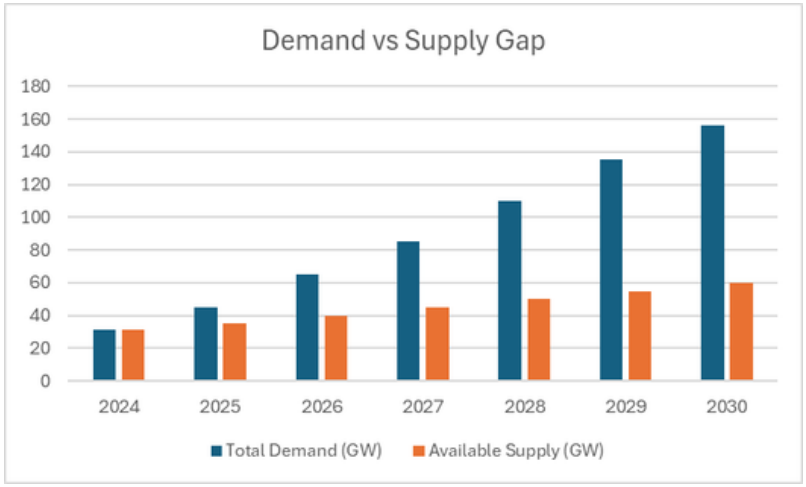
ELECTRICITY

FIBRE NETWORK

AI scales through three interdependent layers: ***data centres (compute), electricity (power), and fibre networks (connectivity)***. A constraint in any one layer limits the entire system

Power is the primary bottleneck. AI racks consume 10–50x more power per rack, while the U.S. operates at only ~31 GW of data centre capacity. Compute supply is now outpacing powered infrastructure availability.

This imbalance is already visible. Colocation vacancy has dropped from ~10% to ~2%, creating a clear supply shortage.
Looking ahead, McKinsey & Company projects 156 GW global demand by 2030, requiring 125 GW incremental capacity from 2025.



Demand is also becoming structural. Gartner forecasts AI SaaS growing from $18.3B → $37.5B (104% YoY). Enterprises adopting AI are seeing 10–25% EBITDA gains, reinforcing recurring infrastructure demand.

Early infrastructure players… especially those who secured power… hold a durable competitive moat.

2.3 Client Takeaway

AI has shifted from a model-driven cycle to an infrastructure-constrained cycle. With near-parity spend ($18B vs $19B), $660–750B+ hyperscaler capex (75%+ to infra), ~90% control by NVIDIA, and 36–52 week supply delays, the bottleneck is clear. At the same time, power constraints (10–50x usage, ~31 GW capacity, ~2% vacancy) and 156 GW demand by 2030 make infrastructure the limiting factor.

The opportunity is not in building better models... it is in owning compute, power, and capacity where scarcity is already pricing in advantage.

2.4 Risk Analysis

Compute & GPU Supply Risk

India's AI infrastructure remain**s heavily dependent on imported GPUs, advanced semiconductors, HBM memory, and hyperscaler-controlled cloud ecosystems.** Current global risks include GPU lead times of 3–7 months, HBM memory being effectively sold out through 2026, and severe shortages in advanced packaging (CoWoS). Furthermore, NVIDIA's market dominance at roughly 90% share of AI accelerators underscores a lack of domestic control over GPU manufacturing, advanced semiconductor fabrication, or high-bandwidth memory production.

Even though over 38,000 GPUs have been onboarded through government-linked frameworks, this capacity remains extremely small relative to the massive projected demand from hyperscalers, enterprises, and sovereign AI initiatives. Without domestic advanced packaging infrastructure or fabrication capabilities, India remains vulnerable to global supply chain volatility. This gap highlights a critical strategic dependency as the nation attempts to scale its compute power to meet the needs of a rapidly expanding digital economy.

Key Risk

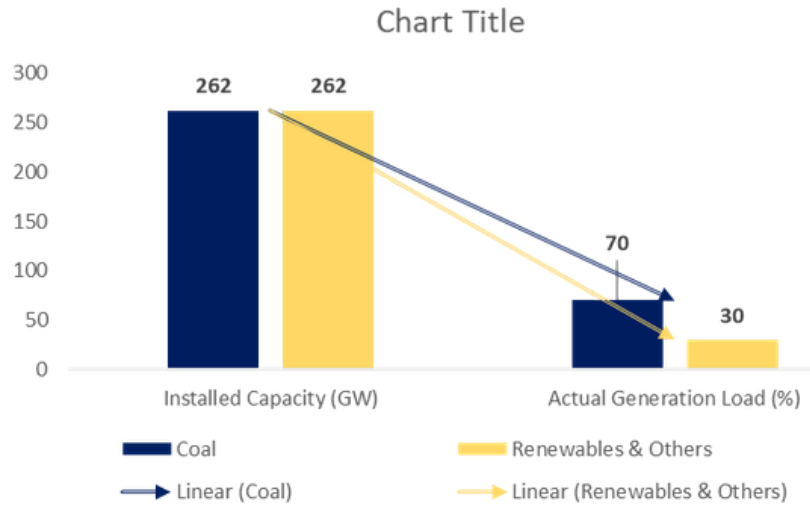
AI demand may grow materially faster than compute availability, creating:

- Rising cloud costs
- Delayed deployments

- Lower enterprise adoption speed
- Strategic dependence on foreign suppliers

Power & Grid Stability Risk

AI infrastructure is fundamentally power-intensive, with high-density racks now requiring 30–160 kW each. Although India's installed generation capacity has reached approximately 524 GW, the challenge has shifted toward transmission, distribution, and reliability constraints, evidenced by 50+ GW of stranded renewable capacity and a 42% transmission target shortfall in FY2025. With industrial tariffs ranging from ₹7–₹10/unit and an estimated 411 GWh storage requirement by 2032, the sector remains vulnerable. Furthermore, while renewable capacity has crossed 262 GW, actual generation still relies on coal for approximately 70% of the load, creating significant sustainability and reliability concerns for consistent AI operations.



India has reached a 50/50 capacity split (~262 GW each for Coal and Renewables), yet AI's 24/7 uptime forces a 70% actual reliance on coal. This gap is widened by a 42% transmission shortfall and AI rack densities surging to 30–160 kW.

Solving this bottleneck requires a mandatory shift to liquid cooling and achieving the 411 GWh storage target to sync green capacity with high-density compute.

Key Risk
Without major investments in:

- Grid modernization
- Storage infrastructure
- Transmission corridors
- Captive energy systems

AI infrastructure deployment could face severe power bottlenecks by the end of the decade.

Cooling & Thermal Risk

The AI infrastructure transition is also a thermal transition.
Traditional cooling systems are increasingly unable to support next-generation GPU densities.
Key risks include:

- AI racks exceeding 70–160 kW density
- Cooling systems consuming 30–40% of total DC energy
- CDU lead times rising from 12–16 weeks → 40–60 weeks
- Hyperscale water usage reaching 3–5M gallons/day
- Retrofit costs of $2–5M per MW

India also faces climate-related disadvantages:

- High ambient temperatures
- Water scarcity in major cities
- Increasing regulatory pressure around water usage

Key Risk

Thermal management failures could directly reduce:

- GPU efficiency
- Facility uptime
- Infrastructure ROI

Cooling infrastructure may become one of the largest hidden constraints on AI scaling.

Connectivity & Network Concentration Risk

India’s digital connectivity ecosystem remains geographically concentrated, with **only 31 Internet Exchange Points (IXPs) nationwide**, largely restricted to just seven cities. This centralized architecture is further strained by a heavy subsea cable concentration in Mumbai and Chennai, resulting in domestic internet traffic utilization at exchanges of only 5–6% and much of that traffic still being routed internationally through Singapore. These structural bottlenecks pose a significant risk as the demand for AI inference increasingly requires low-latency edge infrastructure and distributed compute environments to support real-time AI applications.

Key Risk

Weak Tier-2 and semi-urban interconnection infrastructure could slow:

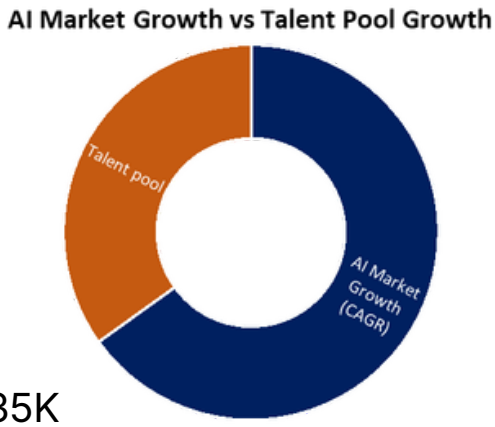
- Edge AI deployment
- Real-time AI services
- Regional digital expansion

Talent Shortage Risk

India possesses a large AI talent base, but frontier-level expertise remains limited relative to projected demand.

Key figures:

- 2.75M AI/ML professionals (broad workforce)
- Core frontier pool: 600K–650K
- AI market growth: 25–35% CAGR
- Frontier talent growth: only ~15% CAGR
- 23,000+ active SRE openings
- Data center engineering workforce: only 25K–35K



Critical shortages are emerging in:

- GPU infrastructure reliability
- CUDA optimization
- AI thermal operations
- Advanced networking
- High-density data center engineering

Key Risk

Talent shortages could become a scaling bottleneck even if physical infrastructure expands successfully.

Land & Infrastructure Execution Risk

India possesses 137,517 hectares of industrial land with 1.25 lakh+ vacant plots, yet infrastructure-ready sites remain limited. Premium data center land pricing has reached ₹20–40 crore/acre, while approval processes frequently cause 6–12 month delays.

Key Risks

- Slow approvals
- Power connectivity delays
- Fiber deployment bottlenecks
- Environmental clearances
- Urban congestion in hyperscale hubs

This creates a situation where land availability exists, but deployment readiness remains constrained.

Financial & Capital Intensity Risk

AI infrastructure remains one of the most capital-intensive sectors globally, with data center build costs ranging from ₹40–60 crore/MW. Essential specialized cooling infrastructure adds an additional ₹8–15 crore/MW, while AI-ready racks cost between ₹8–12 lakh each. These expenditures are further pressured by sharply rising costs for renewable energy and Battery Energy Storage System (BESS) deployment.

Large-scale hyperscale projects increasingly require:

Multi-billion-dollar capital commitments

Long payback periods

Stable utilization assumptions

Key Risk

If AI adoption grows slower than expected, infrastructure assets could face:

- Underutilization
- Margin compression
- Delayed ROI cycles

Regulatory & Policy Risk

The AI ecosystem remains highly exposed to policy changes regarding data localization, energy regulation, and environmental compliance. Additionally, shifts in semiconductor imports and cross-border AI governance continue to impact operational stability and strategic planning across the sector.

Emerging risks include:

- PFAS restrictions affecting cooling fluids
- Power tariff reforms
- Renewable procurement regulations
- Data sovereignty mandates
- Semiconductor export restrictions globally

Key Risk

Infrastructure planning cycles are long-term, while policy environments remain rapidly evolving.

2.5 Overall Risk Assessment

Risk Category	Severity
GPU & Hardware Dependency	**Critical**
Power & Grid Stability	**High**
Cooling & Thermal Constraints	**High**
Talent Shortage	**High**
Connectivity Concentration	**Medium-High**
Regulatory Risk	**Medium-High**
Land & Execution Delays	**Medium**
Capital Intensity	**Medium**

2.6 Final Risk Conclusion

India's AI infrastructure opportunity remains extremely strong, but the ecosystem is exposed to a common structural issue across nearly every layer: Demand is scaling faster than infrastructure readiness.

The largest risks are not around lack of AI adoption- they are around:

- Compute availability
- Power stability
- Cooling scalability
- Skilled execution
- Supply chain control

In the AI economy, resilience and execution capacity may become more valuable than raw technological capability alone.

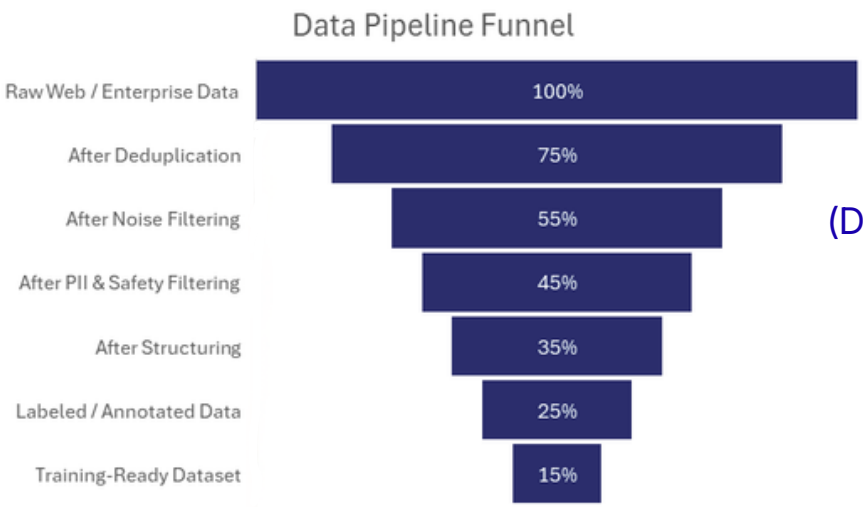
The long-term winners will therefore be the companies capable of reducing infrastructure friction and controlling bottlenecks across multiple layers simultaneously.

3. SOFTWARE INFRASTRUCTURE ANALYSIS

3.1 Data Gathering & Collection

This stage focuses on how AI systems acquire, generate, clean, and label data before training. It is one of the most critical layers in the pipeline because model performance, safety, and reliability are determined more by data quality than model architecture.

It combines multiple processes... web-scale data collection, synthetic data generation, human labeling, and data cleaning pipelines... all working together to produce a usable dataset. At the same time, this layer is becoming increasingly constrained by privacy laws, licensing restrictions, and limited availability of high-quality human-generated data.
A large portion of AI effort happens here, before any model is trained.



(Data volume decreases at every stage, but value increases sharply.)

Key Figures

Metric	Detail
Common Crawl Scale	**Pages per crawl: 2.16B–2.74B, Total Size: 364–468 TiB**
Projected Human-Generated Data Bottleneck	**Expected between 2026 and 2032**
Synthetic Data Benchmark	**38,000 sentences and 250,000 ratings**
Waymo Training Data Volume	**Approximately 200M real miles vs billions of simulated miles**
Synthetic Computer Vision (CV) Performance	**mAP@50 score of 0.910**
Data Labeling Market Size	**Projected growth from $3.77B (2024) to $10.43B (2030)**
Alternative Crowdsourcing / Labeling Market Forecast	**$17.10B by 2030, reflecting a 28.4% CAGR**
Scale AI Valuation	**$29B**
Organizations Lacking AI-Ready Data	**63% (Gartner, 2025)**
AI Projects at Risk Due to Poor Data Quality	**60% by 2026**
Data Preparation Market Size	**Projected growth from $6.5B (2024) to $27.28B (2033)**

INVESTMENT OPPORTUNITY

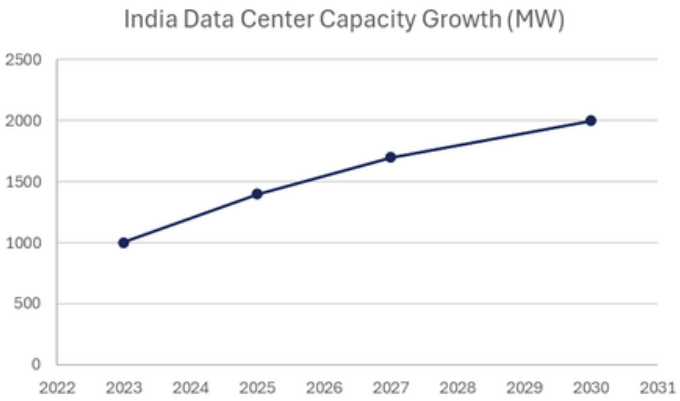
As raw data becomes abundant but usable data remains scarce, the opportunity moves to data refinement layers. With 63% of firms lacking AI-ready data and 60% of projects at risk, value lies in data labeling, synthetic data, and data governance platforms. The edge is in turning large volumes of raw data into high-quality, usable datasets

3.2 Data Storage & Management

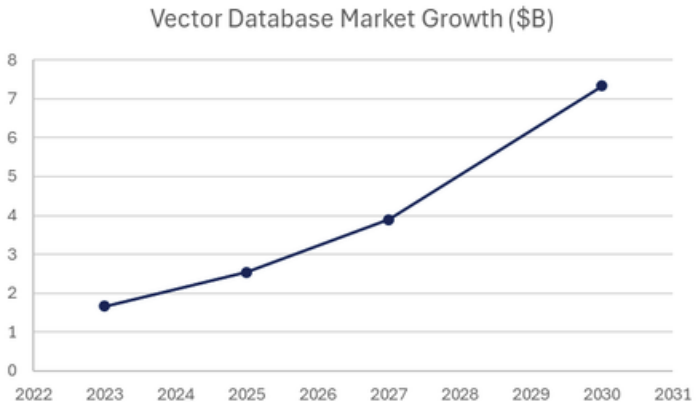
This stage covers how data is stored, organized, retrieved, and served across AI systems. It includes cloud storage, data warehouses, vector databases, and data centres. While compute gets most of the attention, this layer is the actual backbone, because every model depends on data access, movement, and storage efficiency.

AI systems require large-scale storage, fast retrieval, and lifecycle management. This layer directly impacts latency, cost, scalability, and regulatory compliance. It also needs to integrate seamlessly with compute and data pipelines. At the same time, data sovereignty, localization, and hybrid infrastructure are becoming central decisions.

Metric	Detail
Cloud Market Share (Q4 2024)	Amazon Web Services **33%**, Microsoft Azure **20%**, Google Cloud **10%** (**~64% combined**)
India Data Center Capacity	**1,000+ MW → 2,000+ MW** (projected growth)
Vector Database Market	**$1.66B (2023) → $7.34B (2030)** (**23.7% CAGR**)
Snowflake Revenue (Q4 FY2025)	**$943M**
India Cloud AI Market	**$2.73B (2024) → $100.86B (2033)**
India Cloud Services Market	**$951M (2024) → $2.3B (2033)**
Cloud Usage in India	**Dominated by hyperscalers**



(Capacity is doubling (1000 → 2000 MW), but still needs to keep pace with accelerating AI demand.)



(Vector databases are scaling rapidly (1.66 → 7.34B), becoming a critical layer for AI retrieval and GenAI applications.

INVESTMENT OPPORTUNITY

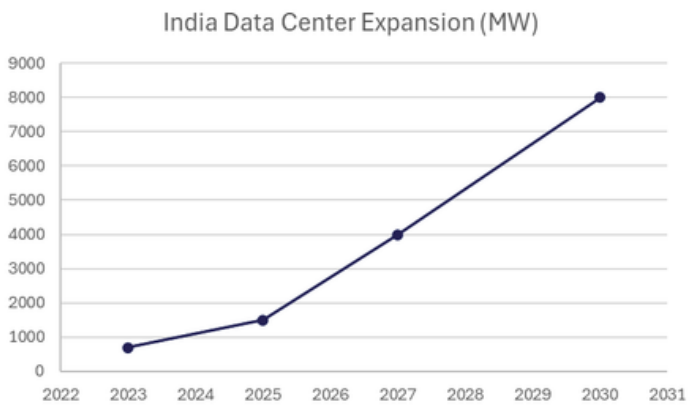
As AI systems scale, storage becomes a performance and control layer, not just infrastructure. With ~64% cloud concentration, rapid growth in vector databases ($1.66B → $7.34B), and rising data sovereignty needs, the opportunity lies in cloud platforms, data warehouses, and hybrid infrastructure models. The advantage comes from efficient data access, movement, and control.

3.3 Data Processing and Transformation

This stage converts raw data into structured, model-ready datasets and processes it through compute infrastructure. It includes ETL pipelines, data cleaning, preprocessing systems, and GPU compute. This is where data moves from being collected to being usable... and economically valuable.

It directly impacts model accuracy, speed to insight, and ROI, while shifting from manual pipelines to AI-driven automation. In India, this layer becomes even more critical due to fragmented data formats, multiple languages, and legacy systems, making efficient processing a key differentiator.

Metric	Detail
Global Data Integration Market	**$14.96B (2023) → $30.97B (2030)**
ETL Pipeline Share	**39.46%** of data integration market
India AI Data Market Growth	**24.3% CAGR**
Data Cleaning Effort	**60–80% of ML project time**
ROI from Automation	**~150% in first year**
GPU Cost Advantage (India)	**Up to 70% cheaper vs global**
India Data Center Expansion	**700 MW → ~8 GW by 2030**
India AI Market Size	**$6.05B → $17–22B by 2027 (25–35% CAGR)**
AI Contribution to GDP	**$450–500B potential**



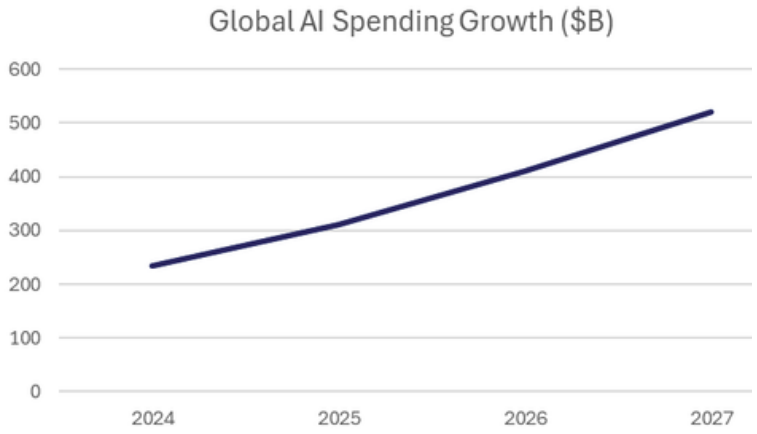
(Infrastructure is scaling ~10x (700 MW → 8000 MW), enabling large-scale AI processing capacity)

INVESTMENT OPPORTUNITY

This layer offers one of the highest ROI opportunities in the AI stack, as it directly converts data into business outcomes. With 60–80% of time spent on data preparation and ~150% ROI from automation, value lies in ETL automation, data cleaning platforms, preprocessing infrastructure, and low-cost GPU compute.

3.4 Model Development & Training

Model development and training is the most compute-intensive stage in the AI pipeline and one of the most resource-constrained areas in India. While demand for AI is rising rapidly, the availability of training-grade GPUs, cooling systems, and power-ready infrastructure remains limited.



(Demand is scaling aggressively (235B → 631B), while infrastructure capacity lags... widening the compute gap.)

Training large models requires high-performance compute, stable power, and advanced cooling, all of which are still developing domestically. As a result, India continues to rely heavily on external cloud providers and imported hardware for advanced model training.

This makes the constraint clear: demand exists, but infrastructure is insufficient to support it at scale.

Key Figures

Metric	Detail
Global AI Spending	**$235B (2024) → $631B+ (2028)**
India Data Center Capacity (2025)	**~1.53 GW**
Additional Capacity by 2030	**+2.9 GW expansion**
Infrastructure Gap	Advanced AI workloads still require **significant additional capacity**

INVESTMENT OPPORTUNITY

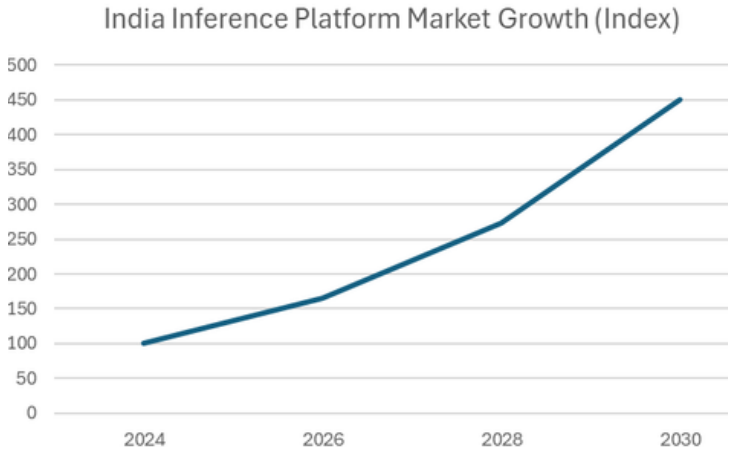
This layer presents a pure infrastructure opportunity, where demand is already validated but supply remains constrained. The strongest opportunities lie in GPU hosting, managed training clusters, AI-ready data centers, and cloud-linked compute infrastructure.
As India's capacity (~1.53 GW + 2.9 GW expansion) still falls short of advanced AI requirements, value will concentrate with players who can provide reliable, scalable, and locally accessible compute.

3.5 Model Management & Deployment

Once a model is trained, it must be reliably deployed and managed within enterprise systems. This involves MLOps platforms, model serving infrastructure, API orchestration, and governance tools.

This layer ensures models are not just built... but consistently running, updated, and integrated into real business workflows.

Enterprises require version control, uptime, security, and seamless integration with internal systems. Unlike training, which is often a one-time cost, deployment generates recurring spend, making it a long-term revenue layer. In India, while hyperscalers remain dominant, there is growing demand for tools that simplify deployment and reduce operational complexity.



(A 28.3% CAGR compounds rapidly, indicating strong and sustained demand for deployment infrastructure and MLOps tools.)

INVESTMENT OPPORTUNITY

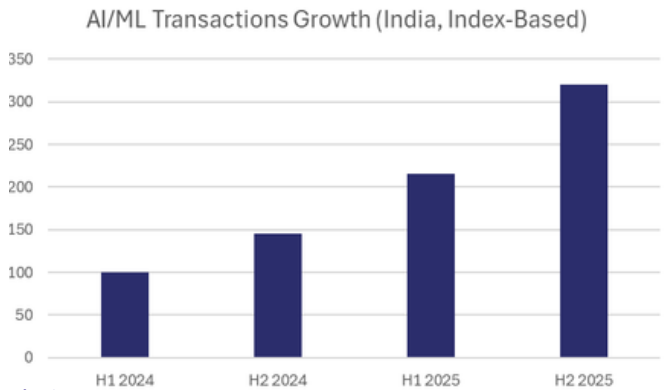
This is a high-margin, recurring revenue layer within the AI stack. The opportunity lies in MLOps platforms, model serving infrastructure, and orchestration tools that enable faster and more reliable deployment.

As enterprises scale AI adoption, demand will rise for solutions that reduce hyperscaler dependency, ensure compliance, and shorten time-to-production.

3.6 Inference & Production

Inference is where AI generates real-world value at scale, as every model interaction... queries, predictions, recommendations... counts as inference. As enterprise adoption deepens, this layer is expected to become one of the largest long-term spend categories in the AI stack

It requires low latency, high uptime, and cost-efficient serving infrastructure. Demand is expanding across BFSI, retail, telecom, healthcare, and customer service, where models are used continuously in production environments. To manage costs at scale, techniques such as edge deployment, batch processing, model compression, and inference optimization become critical.



(Inference workloads are scaling rapidly, with production usage already reaching massive transaction volumes... indicating strong recurring demand.)

82.3 billion

AI/ML Transactions in India (H2 2025)

46.2%

Share of APJ Activity

INVESTMENT OPPORTUNITY

This layer represents a shift from building AI to operating AI at scale, making it one of the most durable and recurring revenue opportunities. Value flows into model serving infrastructure, inference optimization tools, edge deployment systems, and bandwidth/storage layers.

As usage scales to billions of transactions, cost efficiency becomes critical... creating demand for solutions that reduce latency, optimize compute usage, and lower serving costs

3.7 Monitoring & Maintenance

Once AI systems are deployed, they require continuous monitoring to remain accurate, reliable, and compliant. This includes drift detection, retraining pipelines, observability, performance tracking, and A/B testing. Without this layer, models degrade as data patterns and user behavior evolve.

This is especially critical in regulated sectors like BFSI, where governance and auditability are mandatory. While less visible than training or inference, this layer creates stable, long-term enterprise spending.

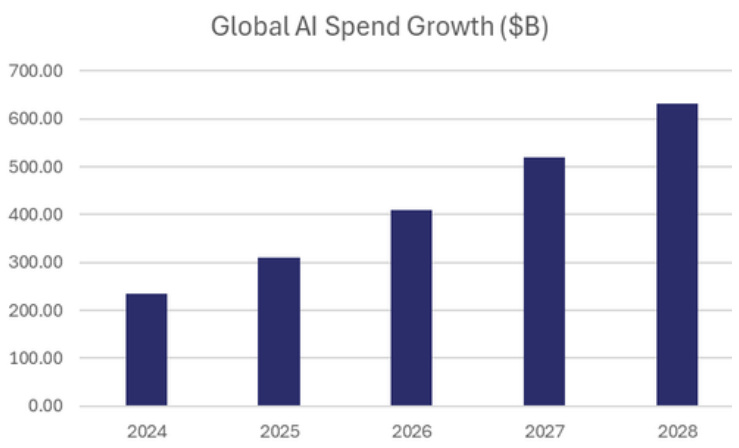
Metric	Value	Why it matters
Global AI Spend (2024)	**$235B**	Baseline market size
Global AI Spend (2028)	**$631B+**	Strong growth runway
India Data Center Capacity (2025)	**1.53 GW**	Current infrastructure depth
India Data Center Expansion	**+2.9 GW**	Ongoing capacity buildout
India Inference Platform CAGR	**28.30%**	Growth in deployment & inference
India AI/ML Transactions	**82.3B**	Large-scale production usage

1.53 GW

India Data Center Capacity (2025)

+2.9 GW

India Data Center Expansion



INVESTMENT OPPORTUNITY

Monitoring and maintenance represent a highly sticky, recurring revenue layer. As AI systems scale, enterprises need continuous solutions for observability, compliance, retraining, and governance.

The strongest opportunity lies in bundled platforms that combine monitoring with automation and regulatory support, reducing operational complexity. As production usage increases (82.3B+ transactions), demand for reliability and performance assurance will scale proportionally.

3.8 Final Investment View - Software Infrastructure Layers

Across the software stack, a clear pattern emerges: value compounds as you move from building AI to operating it continuously at scale.

- Data layers (collection, processing, storage) determine what can be built
- Training infrastructure determines what can be achieved

Deployment, inference, and monitoring determine what actually generates money.

The economics shift accordingly. Training is capital-heavy and episodic, while deployment, inference, and monitoring are recurring and sticky, capturing long-term enterprise budgets. As seen across the data:

- 60–80% of effort sits in data preparation
- Inference is already at 82.3B+ transactions (India)
- Deployment markets are growing at ~28%+ CAGR
- AI spend scales from $235B → $631B+ globally

This creates a layered opportunity:

- Upstream (data + training): high capex, foundational control
- Midstream (storage + processing): efficiency and performance advantage
- Downstream (deployment + inference + monitoring): recurring revenue and strongest margins

Investment Positioning

The most attractive strategy is to focus on layers that combine necessity with recurrence:

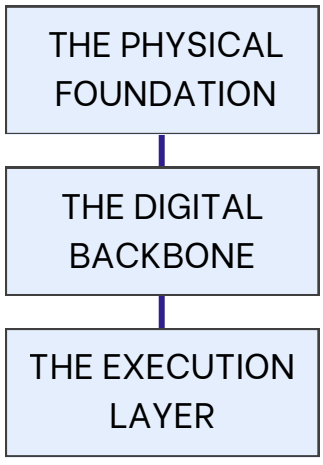
- Data processing & preparation platforms (where most effort and failure risk exists)
- MLOps and deployment infrastructure (enterprise integration layer)
- Inference optimization & serving (usage-driven revenue)
- Monitoring, observability, and retraining (long-term retention layer)

Bottom Line

AI is not monetized when models are trained... it is monetized when they are repeatedly used, optimized, and maintained. The winners will not be those who build the best models, but those who control the full software infrastructure stack that keeps AI running efficiently, reliably, and at scale.

4. HARDWARE & INFRASTRUCTURE ANALYSIS

LAYERS OF HARDWARE INFRASTRUCTURE



forms the foundation layer of AI infrastructure, enabling the physical environments required to support compute, storage, networking, and AI workloads at scale. This ecosystem includes EPC firms, cooling systems, modular infrastructure, and AI-ready data centre design, all of which are becoming increasingly critical as India enters a large-scale infrastructure expansion cycle.

India's market is currently in a hyper-growth phase, driven by hyperscalers, enterprise cloud adoption, AI workloads, and data localization requirements. The industry is dominated by a combination of global engineering firms and strong domestic execution players, including Larsen & Toubro, AECOM, Jacobs, Vertiv,

THE PHYSICAL FOUNDATION

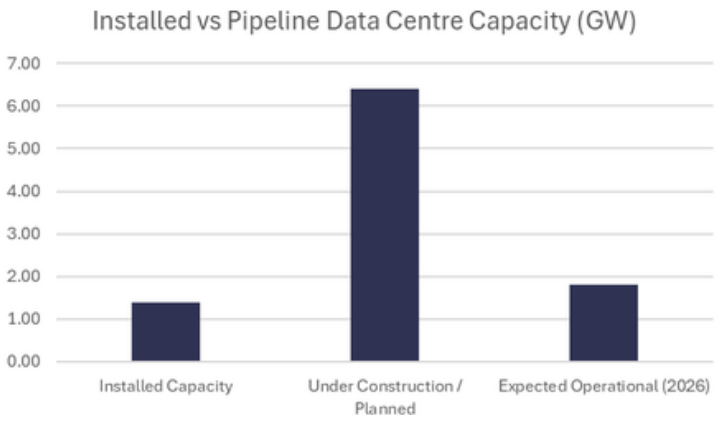
4.1 Data Centre Construction

Data centre construction

Schneider Electric, STT GDC, Sify, and AdaniConneX.
The market is rapidly shifting from traditional cloud infrastructure toward AI-first infrastructure, where data centres require significantly higher power density, advanced cooling systems, and GPU-ready environments.

AI-ready racks now demand 10–20x higher power requirements, with power density reaching 30–100 kW per rack, compared to traditional deployments.

India currently operates at approximately ~1.4 GW installed data centre capacity, while another ~6.4 GW is under construction or planned, reflecting an aggressive infrastructure buildout cycle. Major hubs include Mumbai (44% market share), Chennai, Hyderabad, and Delhi NCR, supported by projects from Reliance Industries, AdaniConneX, CtrlS, and hyperscalers such as Google, AWS, and Microsoft



SUPPORTED BY PROJECTS FROM -

Infrastructure costs remain substantial:

- Total hyperscale build cost: ₹40–60 Cr per MW
- EPC cost: ₹5–8 Cr per MW
- Cooling infrastructure: ₹8–15 Cr per MW
- AI-ready cooling systems: 20–30% additional premium

At the same time, modular data centres are emerging as a major trend due to their lower deployment timelines (6–12 months vs 18–36 months) and lower upfront cost structure. Cooling is also becoming a strategic differentiator, as it already contributes ~30–40% of total data centre energy consumption.

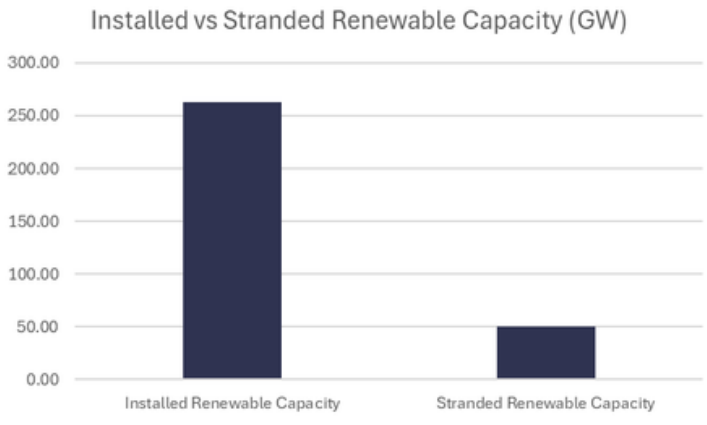
4.2 Energy Infrastructure

Energy infrastructure is becoming the single most important constraint layer in India's AI and data center expansion cycle. While compute and data center construction receive most of the attention, long-term scalability ultimately depends on reliable power generation, transmission capacity, industrial tariffs, renewable integration, and backup systems.

India reached a major milestone in 2025 with 509.64 GW installed electricity capacity, of which 262.74 GW (51.5%) came from non-fossil sources, achieving its COP-26 target five years ahead of schedule. Between January and November 2025 alone, India added 44.51 GW of renewable capacity, including 132.85 GW solar and 53.99 GW wind.

However, installed capacity does not translate into continuous, reliable power. Coal still supplies 70%+ of actual electricity generation, while renewable energy remains intermittent. For AI data centers operating 24/7, this creates a major challenge: reliable power availability matters more than headline renewable capacity figures.

Transmission infrastructure is emerging as a critical bottleneck. As of June 2025, over 50 GW of commissioned renewable energy capacity remained stranded because transmission networks could not evacuate power efficiently. India added only 8,830 circuit km of transmission lines in FY2025, against a target of 15,253 km, creating a 42% shortfall.



This mismatch is especially important as AI-ready racks now require 30–100 kW per rack, far above traditional data center requirements. Existing urban grids-particularly in Mumbai and NCR, were not designed for such high-density AI loads.

At the same time, electricity pricing creates a major economic opportunity. Industrial tariffs vary significantly by state, ranging from ₹6.90/kWh in Karnataka to ₹11.5/kWh in Maharashtra, while captive renewable procurement can reduce costs to ₹3–4/kWh for 20–25 years. Since energy accounts for 30–40% of total data center operating

costs, power procurement becomes one of the most important financial decisions for operators.

As a result, the market is increasingly shifting toward behind-the-meter strategies, including:

- Captive solar and wind PPAs
- Battery Energy Storage Systems (BESS)
- Open-access renewable procurement
- Diesel backup and hybrid systems

Large operators such as Nxtra and STT GDC already operate at 60–63% renewable energy integration, demonstrating that partial renewable transition is commercially viable today.

The ecosystem also includes critical support layers such as:

- UPS systems (Vertiv, Schneider Electric, Eaton, Numeric, Delta)
- Grid operators & transmission infrastructure (PGCIL, Grid-India, NLDC)
- Energy management software (Schneider EcoStruxure, Siemens, ABB, Honeywell)

India's average data center PUE currently sits at approximately 1.4–1.6, significantly above global hyperscaler benchmarks of 1.1–1.3, highlighting a major efficiency gap and cost optimization opportunity.

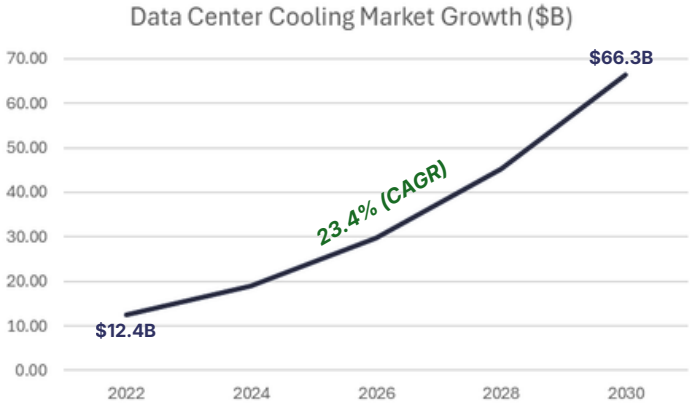
Metric	Detail
India Installed Power Capacity (2025)	**509.64 GW**
Non-Fossil Share	**262.74 GW (51.5%)**
Renewable Capacity Added in 2025	**44.51 GW**
Solar Capacity	**132.85 GW**
Wind Capacity	**53.99 GW**
Coal Share of Actual Generation	**70%+**
Stranded Renewable Capacity	**50+ GW**
Transmission Target Shortfall (FY2025)	**42%**

Industrial Tariffs Range	**₹6.90–11.5/kWh**
Captive Renewable Cost	**₹3–4/kWh**
Energy Share of DC Operating Cost	**30–40%**
AI Rack Power Density	**30–100 kW per rack**
BESS Installed Capacity	**~2.56 GW**
Projected BESS Opportunity (2025–2050)	**USD 487B**
India Data Center Investments (2019–2025)	**~USD 95B**
Commercial & Industrial RE Market Growth	**22% CAGR**

4.3 Cooling & Thermal Management

Cooling and thermal management have become one of the most critical constraint layers in AI infrastructure. Every GPU cluster, training workload, and inference system generates large amounts of heat that must be removed efficiently and continuously. If thermal management fails, performance drops, hardware degrades, and infrastructure investments become unusable.

As AI workloads push rack densities far beyond the limits of conventional air cooling, cooling is no longer a support function... it is now a primary infrastructure and capital allocation decision.



The global data center cooling market was valued at $12.4B in 2022 and is projected to reach $66.3B by 2030, growing at approximately 23.4% CAGR. This growth is being driven primarily by AI infrastructure expansion, ***where a single H100 GPU*** cluster can generate ***20–30x the thermal load*** of a traditional server environment.

The transition toward AI infrastructure is fundamentally a transition toward high-density compute environments. NVIDIA's H100 SXM5 GPU carries a 700W TDP, while the upcoming Blackwell B200 reaches 1,000W per GPU. An 8-GPU DGX H100 server generates approximately 10.2 kW of heat, and fully populated AI racks can exceed 70–100 kW, far above the ~20 kW ceiling manageable through conventional CRAC/CRAH air systems.

This has accelerated adoption of liquid cooling systems, including:

- Direct Liquid Cooling (Cold Plate)
- Rear-Door Heat Exchangers (RDHx)
- Immersion Cooling (Single-Phase & Two-Phase)
- Cooling Distribution Units (CDUs)

Immersion cooling is emerging as the most thermally efficient commercial solution, with achievable PUE levels of 1.02–1.05 and support for 200+ kW rack densities. At the same time, AI-optimized liquid cooling environments are now targeting PUE ranges of 1.03–1.08, compared to the global average data center PUE of 1.58.

The market is increasingly dominated by firms specializing in thermal management and AI-ready cooling infrastructure, including:

- Vertiv
- Schneider Electric
- Asetek
- CoolIT Systems
- LiquidStack
- nVent Electric

At the same time, the sector faces significant bottlenecks:

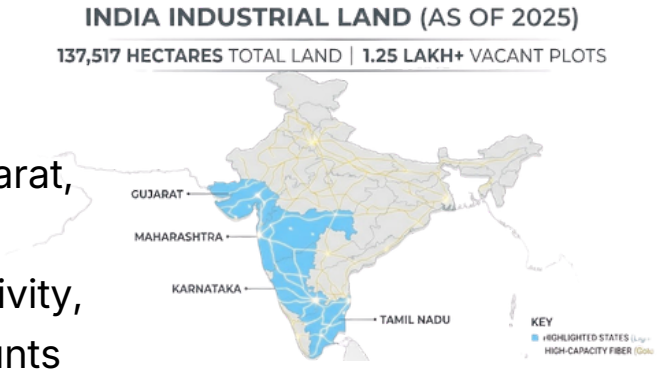
- Water usage: 100 MW hyperscale facilities can consume 3–5M gallons/day
- CDU lead times: increased from 12–16 weeks → 40–60 weeks
- Retrofit costs: $2–5M per MW converted
- Technician shortage: 35,000+ gap expected by 2026
- PFAS regulation pressure: disrupting immersion cooling fluid supply chains

4.4 Real Estate & Facilities

Industrial land forms the physical foundation of India's digital infrastructure economy. Every data center, AI campus, semiconductor facility, logistics hub, and manufacturing cluster ultimately depends on one core requirement: infrastructure-ready land with access to power, fiber, transportation, and regulatory approvals.

India's advantage lies not only in growing digital demand, but also in its relatively lower land costs compared to global markets. However, the real constraint is not land availability... it is the availability of usable, infrastructure-ready land.

As of 2025, India has approximately 137,517 hectares of industrial land, with more than 1.25 lakh vacant plots across industrial parks and estates. Yet much of this supply remains underutilized because many sites lack critical infrastructure such as reliable power, high-capacity fiber connectivity, transport access, or fast regulatory clearances.



Geography plays a decisive role in land economics. States such as Maharashtra, Gujarat, Tamil Nadu, and Karnataka dominate due to stronger industrial ecosystems, port connectivity, and policy support. Maharashtra alone accounts

for approximately 25,000 hectares, making it India's largest industrial land market.

At the same time, premium digital infrastructure corridors are experiencing significant pricing pressure. While standard industrial land ranges from ₹10–25 lakh per acre, infrastructure-ready hyperscale and data center zones command significantly higher premiums:

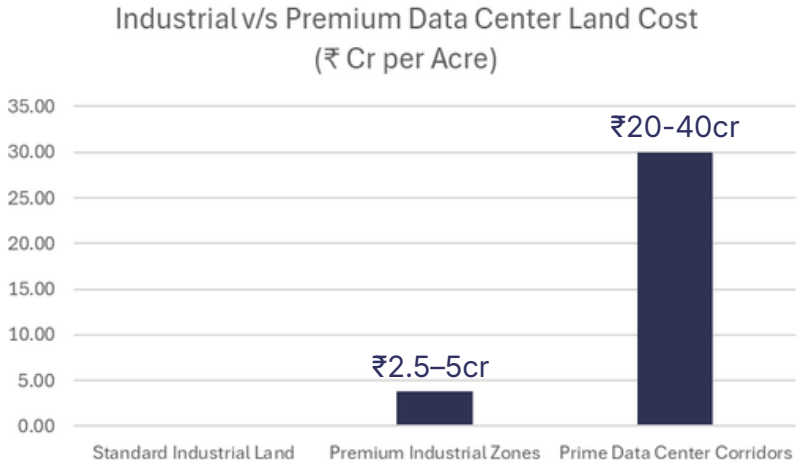
- Industrial hotspots: ₹2.5–5 crore per acre
- Prime data center corridors: ₹20–40 crore per acre

This reflects the increasing value of locations with:

- Reliable power access
- Dense fiber connectivity
- Proximity to airports and highways
- Access to hyperscaler ecosystems

Approval speed is another major investment variable. Basic approvals can now take as little as 2–4 months through single-window systems, but environmental and infrastructure clearances can still add 6–12 months, directly impacting project timelines and capital deployment.

Compared to global markets, India still retains a strong structural cost advantage. Prime data center land in India is estimated at approximately $100–130/sq ft, compared to roughly $270/sq ft in China, making India increasingly attractive for hyperscalers and global infrastructure investors.



At the same time, industrial parks are evolving into integrated digital ecosystems with advanced facility management systems covering:

- Power and HVAC optimization
- Security and operational management
- Predictive maintenance platforms
- Smart infrastructure monitoring

INVESTMENT OPPORTUNITY

India's AI data center boom is driving a premier capital allocation cycle focused heavily on infrastructure readiness over raw capacity. The immediate upside belongs to specialized EPC firms and modular infrastructure providers capable of compressing build timelines from years down to 6–12 months. In real estate, value is concentrating intensely in prime digital corridors across Mumbai, Delhi NCR, Chennai, and Pune, where operators secure competitive moats by providing immediate utility integration and fast-tracked 2–4 month regulatory approvals.

With energy commanding 30–40% of data center operating costs, power access directly limits scalability, turning energy into an asymmetric investment opportunity. Operators are capturing alpha through behind-the-meter strategies like captive renewable PPAs that slash costs to ₹3–4/kWh, transmission upgrades, and Battery Energy Storage Systems (BESS) for 24/7 reliability. This grid strain directly accelerates demand for specialized liquid and immersion cooling, shifting massive pricing power to provider ecosystems (CDUs and heat exchangers) capable of managing 70–100+ kW rack densities and driving PUE down to 1.02–1.05.

THE DIGITAL BACKBONE

4.5 Semiconductor & Hardware Supply

Semiconductors and hardware form the physical foundation of the AI ecosystem. While software, models, and applications evolve rapidly, AI progress ultimately depends on the availability of GPUs, AI accelerators, memory systems, advanced packaging, and semiconductor manufacturing capacity.

This layer has become one of the largest constraints in the global AI race because compute demand is growing faster than supply chains can scale. High-end AI chips now face 3–7 month lead times, while critical components such as HBM memory and advanced packaging capacity (CoWoS) are already fully booked into 2026.

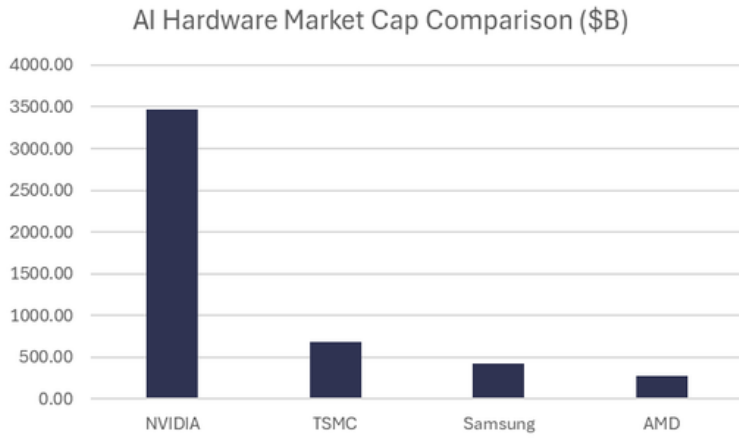
The ecosystem is dominated by a small number of strategic players across different layers:

- GPU Designers: NVIDIA, AMD, Intel
- Foundries: TSMC, Samsung
- Lithography: ASML
- Custom AI Chips: Google TPUs, Amazon Trainium, Microsoft Maia

This concentration creates extreme dependency across the AI stack.

NVIDIA alone has reached a $3.47 trillion market capitalization, reflecting how critical GPU access has become. At the same time, TSMC and Samsung remain essential because nearly all advanced AI chips depend on their manufacturing capacity.

Metric	Detail
NVIDIA Market Cap	**$3.47T**
AMD Market Cap	**$278B**
TSMC Market Cap	**$680B**
Samsung Market Cap	**$420B**
GPU Lead Time	**3–7 months**
PCB Lead Time	**6 weeks → 6 months**
India Customs Clearance	**3–4 days → <18 hours**
HBM Production Capacity	**Fully sold out for 2026**



Supply chain complexity is also increasing. AI chips require:

- HBM memory for high-speed data transfer
- Advanced packaging (CoWoS) for assembling GPU systems
- PCB manufacturing and logistics for deployment

Any bottleneck across these layers slows down AI scaling globally.
India remains highly import-dependent in this segment, although logistics efficiency has improved significantly, with customs clearance times reduced from 3–4 days to under 18 hours through the CIS system. However, the country still lacks large-scale domestic advanced semiconductor manufacturing capability.

4.6 Networking & Connectivity

Networking and connectivity form the circulatory system of the AI economy, enabling data movement between users, clouds, enterprises, and data centers. If semiconductors and compute represent the processing layer, networking determines how efficiently that compute can actually be used.
As AI workloads become increasingly real-time and distributed, performance depends less on raw compute power and more on latency, interconnection density, fiber depth, and edge infrastructure. This is driving a structural shift from centralized cloud architecture toward distributed edge and low-latency ecosystems.

India’s digital infrastructure expansion has accelerated rapidly through 5G deployment, fiber backbone expansion, interconnection growth, and subsea cable investments. The country now operates over 5.18 lakh 5G BTS, with coverage reaching 99.6% of districts (767/778). At the same time, India’s 5G user base is projected to grow from 36 crore to 42 crore, creating massive traffic growth across networks.

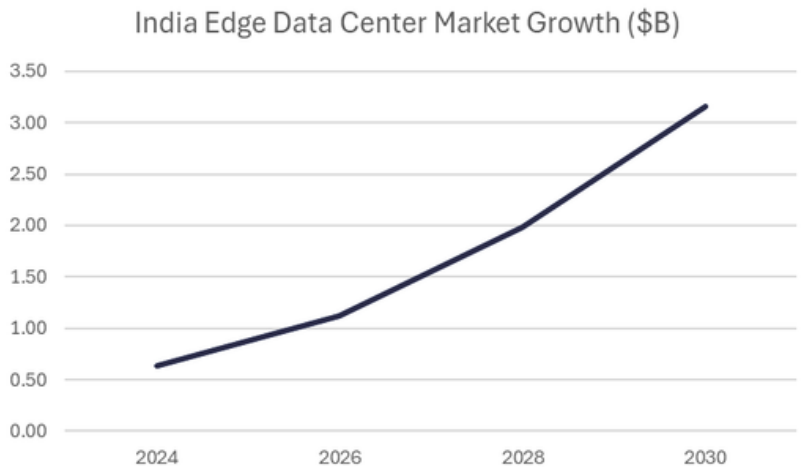
Fiber infrastructure remains the hidden backbone of this ecosystem. Players such as Airtel and RailTel operate extensive national optical fiber networks, with:

- **Airtel: *400,000+ route km***
- **RailTel: *63,000+ route km***

However, while India’s national backbone is relatively strong, metro fiber depth and last-mile connectivity remain uneven, especially in dense urban areas where permissions, road access, and building connectivity create execution bottlenecks.

At the same time, edge computing is emerging as a major growth layer. The Indian edge data center market is expected to grow from $627M to $3.16B, at a 19.1% CAGR, driven by:

- AI inference
- Gaming
- OTT streaming
- Fintech



Real-time applications

International connectivity is also becoming increasingly strategic. Recent subsea cable projects have added 320+ Tbps capacity, with Mumbai and Chennai continuing to dominate as India’s primary international data gateways.

Even inside data centers, networking is becoming more critical. AI workloads generate heavy east-west traffic (server-to-server communication), meaning poor intra-data center networking can leave GPUs underutilized despite high compute investment.

Metric	Detail
5G BTS in India	**5.18 lakh+**
5G Coverage	**99.6% districts (767/778)**
5G Users	**36 Cr → 42 Cr projected**
Edge Data Center Market	**$627M → $3.16B**
Edge DC CAGR	**19.10%**
Airtel Fiber Network	**400,000+ route km**
RailTel Fiber Network	**63,000+ route km**
Recent Subsea Capacity Addition	**320+ Tbps**
AWS Direct Connect Speed	**50 Mbps → 100 Gbps**

INVESTMENT OPPORTUNITY

The semiconductor and hardware stack represents the tightest physical constraint in global AI, making it a highly strategic value-capture layer. Exceptional upside resides in companies controlling immediate bottlenecks: high-end GPU/accelerator distribution, High-Bandwidth Memory (HBM) production, and advanced packaging infrastructure (CoWoS). Concurrently, a structural transition is underway as hyperscalers aggressively pivot toward proprietary, in-house silicon development, shifting long-term capital toward custom ASIC architectures to eliminate external dependencies.

As AI workloads become real-time and distributed, connectivity is transitioning from a basic utility into a highly defensive infrastructure moat. Value is concentrating in private interconnection ecosystems, deep metro fiber networks, and high-capacity subsea gateways where latency and density dictate performance over raw compute. The strongest opportunities lie in operators combining 5G/FWA infrastructure with distributed edge data networks—a market expanding at a 19.1% CAGR—to commercialize low-latency platforms.

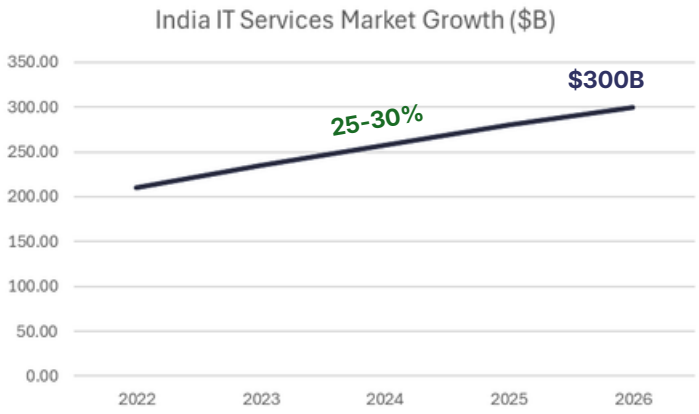
THE EXECUTION LAYER

4.7 Professional Services

The professional services layer sits between technology capability and enterprise deployment. It includes AI consulting, system integration, compliance, governance, managed services, and energy optimization, all of which help enterprises convert AI infrastructure into real business outcomes. In practical terms, technology creates potential... but professional services determine whether that potential becomes operational, scalable, and profitable.

This layer has become increasingly important because most enterprises still lack the internal expertise required to deploy AI systems effectively. Without proper execution, integration, and governance, AI projects often remain stuck at pilot stage or fail entirely. Industry estimates suggest that nearly ~60% of enterprise AI projects fail due to poor execution, weak integration, or inadequate data infrastructure.

The global IT services and consulting industry has already surpassed $1 trillion, while India's IT services market is projected to exceed $300B by 2026. At the same time, AI-focused services are growing rapidly at approximately 25–35% CAGR, reflecting the shift from traditional IT outsourcing toward AI transformation and managed AI operations.



This layer includes multiple interconnected functions:

- AI Consulting: strategy, use-case selection, transformation roadmaps
- System Integration: connecting cloud, data, APIs, AI models, and legacy systems
- Compliance & Governance: privacy, auditability, regulatory alignment
- Energy Efficiency Advisory: PUE optimization, cooling strategy, power efficiency
- Managed AI Services: monitoring, optimization, and ongoing AI operations

The economics of this layer are particularly attractive because services tend to generate:

- Higher margins
- Recurring revenue
- Long-term enterprise relationships
- High switching costs

India holds a structural advantage due to its mature IT services ecosystem, led by firms such as TCS, Infosys, and Wipro, which already possess deep enterprise integration capabilities and global delivery infrastructure.

At the same time, enterprise demand is shifting away from high-level advisory toward execution-linked outcomes. Clients increasingly expect:

- Measurable ROI
- Deployment support

- Continuous optimization
- Managed AI operations
- This transition is moving the industry from project-based consulting toward long-term AI operational partnerships.

Enterprise AI Failure Rate ~60% without proper execution/integration

INVESTMENT OPPORTUNITY

The professional services layer represents a highly scalable, recurring revenue opportunity because raw infrastructure access alone is insufficient to guarantee enterprise ROI. With approximately 60% of enterprise AI projects failing due to weak integration, poor execution, or inadequate data foundations, substantial high-margin value will be captured by full-stack firms providing specialized AI consulting, system integration, compliance platforms, and energy optimization advisory. The long-term alpha lies in firms handling managed AI operations (MLOps and AIOps) to capture sticky, recurring revenue streams, heavily leveraging India's mature IT services ecosystem and structural delivery cost advantages as a global competitive moat.

4.8 Final Investment View: Hardware & Infrastructure Layers

Across the hardware and infrastructure stack, one pattern becomes clear: AI growth is no longer constrained by software demand... it is constrained by physical infrastructure capacity.

Every layer of the ecosystem now faces bottlenecks:

- Semiconductors: GPU, HBM, and packaging shortages
- Power: grid congestion and renewable transmission gaps
- Cooling: thermal limits at high-density AI racks
- Networking: edge scaling, fiber depth, and interconnection demand
- Land: scarcity of infrastructure-ready corridors
- Construction: deployment speed and power integration
- Professional Services: execution and enterprise rollout complexity

As a result, value is shifting away from pure software and toward the infrastructure required to operate AI at scale.

The economics across these layers are highly attractive because most infrastructure assets benefit from:

- Long-term demand visibility
- High switching costs

- Recurring operational revenue
- Structural supply constraints

At the same time, AI workloads are fundamentally changing infrastructure requirements:

- Rack density increases from 5–15 kW → 30–160 kW
- Cooling transitions from air → liquid and immersion systems
- Energy becomes 30–40% of operating costs
- Edge and interconnection demand scale with real-time AI usage
- Prime digital land corridors command massive premiums

This creates a broad investment cycle across:

- Data centers & EPC infrastructure
- Energy generation, storage, and grid systems
- Semiconductors & AI hardware supply chains
- Cooling & thermal management systems
- Fiber, subsea cables, and edge connectivity
- Infrastructure-ready industrial land
- AI consulting, integration, and managed services

Strategic Positioning

The strongest long-term opportunities lie in companies that control:

- Compute availability
- Reliable power access
- Thermal efficiency
- High-speed connectivity
- Infrastructure-ready deployment capacity
- Enterprise execution layers

The market is moving toward an environment where physical scalability becomes more valuable than software innovation alone.

Bottom Line

AI is no longer just a technology cycle... it is a full-scale industrial infrastructure cycle.

The durable winners will not simply be the companies building AI models, but the ones controlling the infrastructure that allows AI to run continuously, efficiently, and at hyperscale.

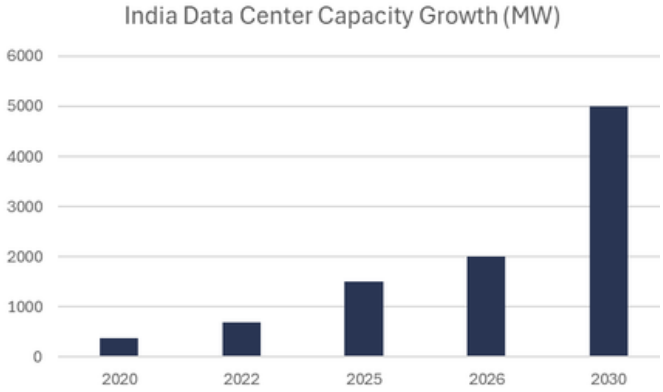
- Whoever controls the constraints... power, chips, cooling, connectivity, land, and deployment- controls the economics of the AI era.

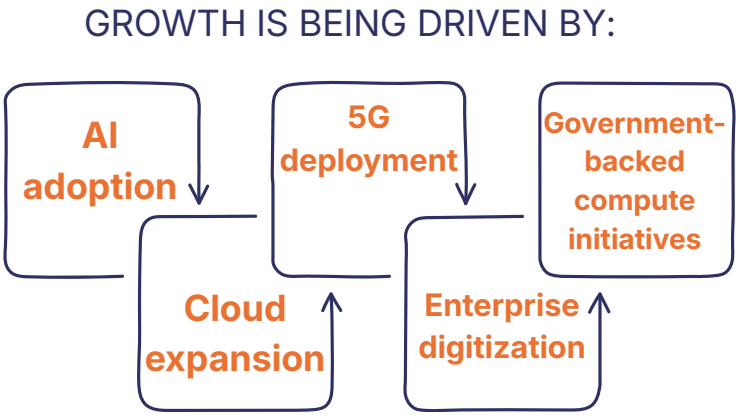
5. INDIA SUPPLY MAP

5.1 Compute Infrastructure

Compute infrastructure forms the backbone of India's AI ecosystem, determining how much AI capacity the country can actually deploy, scale, and operate. This layer includes data center capacity, GPU availability, cloud infrastructure, and compute supply chains, all of which directly influence AI adoption speed, cost, and competitiveness.

India's compute ecosystem is scaling rapidly, but it remains significantly smaller than global leaders and heavily dependent on imported hardware and hyperscaler ecosystems. Total operational data center capacity is currently estimated at approximately ~1.3–1.5 GW (2025), up sharply from just 375 MW in 2020. Capacity is expected to expand further to approximately 4–5 GW by 2030, reflecting one of the fastest infrastructure growth cycles globally.





At the same time, India's compute infrastructure remains highly concentrated geographically. Major hubs include:

- Mumbai: largest hub (~25%+ share) due to subsea connectivity and financial ecosystems
- Chennai: growing rapidly through international cable connectivity
- Delhi NCR: enterprise and government demand
- Hyderabad & Bengaluru: cloud and technology ecosystems

Despite rapid data center expansion, GPU availability remains the largest bottleneck. Most advanced AI hardware... including NVIDIA and AMD GPUs... continues to be imported, creating dependency on global supply chains and hyperscaler partnerships.

India has begun scaling sovereign compute initiatives, with more than 38,000 GPUs onboarded under government-linked frameworks, but the domestic ecosystem remains early-stage. GPU deployment timelines still average ~3 months or longer for large-scale clusters, while long-term constraints continue around HBM memory and advanced semiconductor supply chains.
The broader ecosystem is gradually evolving from:

Importer → Partial assembler → Long-term domestic builder

However, at present, strategic control over high-end compute infrastructure still largely sits outside India.

Metric	Detail
India Data Center Capacity (2025)	**~1.3–1.5 GW**
Capacity Growth (2020 → 2025)	**375 MW → 1500 MW**
Expected Capacity (2030)	**4–5 GW**
Annual Addition (2025)	**+387 MW**
GPUs Onboarded (Gov Frameworks)	**38,000+**
GPU Market Growth	**13–36% CAGR**
GPU Deployment Lead Time	**~3 months+**
Mumbai Market Share	**~25%+**

5.2 Power & Energy Layer

The power and energy layer forms the operational backbone of India's AI, industrial, and digital infrastructure ecosystem. Data centers, AI clusters, manufacturing facilities, and cloud infrastructure all depend on one core requirement: stable, scalable, and cost-efficient electricity supply.
As AI workloads become increasingly power-intensive, India's challenge is

no longer basic electricity generation... it is the efficient management, transmission, pricing, and stabilization of power across the grid.
India's installed power capacity has now reached approximately ~524 GW (2026), while peak demand crossed ~250 GW in 2024, reflecting rapid industrial and digital expansion. The country has also significantly improved transmission infrastructure and reduced Aggregate Technical & Commercial (AT&C) losses from approximately ~22% to ~15%, improving overall grid efficiency.

At the infrastructure level, India now operates approximately:

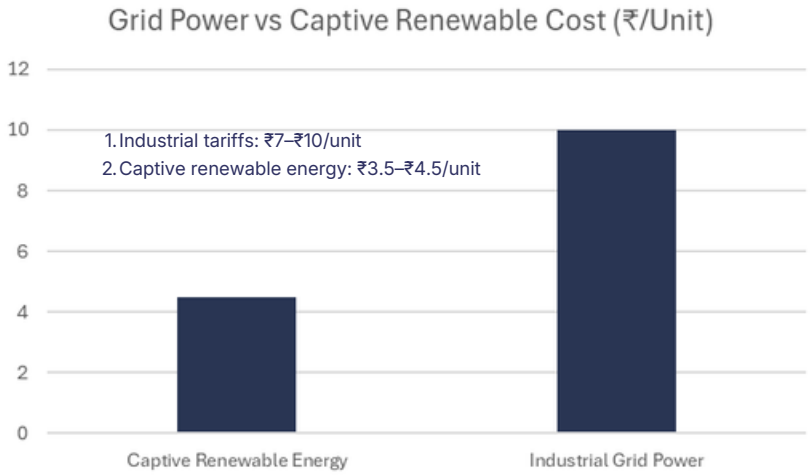
- ~1,147 GVA AC substation capacity
- ~902 GVA distribution transformer capacity

These substations form the critical control layer that determines whether new AI clusters, factories, and hyperscale data centers can actually connect to the grid. In many regions, substation upgrade timelines and land constraints are becoming hidden infrastructure bottlenecks.

At the same time, industrial electricity pricing remains a major competitive factor. India follows a cross-subsidy tariff structure where industrial users effectively subsidize residential and agricultural consumption. As a result:

- Industrial tariffs: ₹7–₹10/unit
- Captive renewable energy: ₹3.5–₹4.5/unit

This large pricing gap is driving rapid growth in captive solar, wind, and hybrid renewable systems, especially among data centers and industrial clusters seeking both cost savings and energy security.



India's renewable energy transition is also accelerating rapidly, with a national target of 500 GW renewable capacity by 2030. However, this introduces new challenges around:

- Renewable intermittency
- Transmission congestion

- Grid balancing
- Long-duration energy storage

The country's energy problem is therefore shifting from:

"Can India generate enough power?" → "Can India deliver stable, affordable power at scale?"

At the same time, hybrid renewable systems and captive generation are gradually transforming industrial consumers into partial energy producers, reducing dependency on traditional DISCOM structures.

5.3 Network Infrastructure

Network infrastructure forms the transport layer of India's digital economy, enabling data movement between users, enterprises, clouds, AI systems, and global internet gateways. As AI workloads become increasingly latency-sensitive and distributed, the quality of network infrastructure directly determines inference speed, cloud performance, and real-time application scalability.

India's network ecosystem consists of four major layers:

- Internet Exchange Points (IXPs)
- Long-haul optical fiber networks
- Subsea cable connectivity
- Intra-data-center interconnection infrastructure

Together, these systems form the connectivity backbone required for AI deployment at national scale.

India currently operates 31 active Internet Exchange Points (IXPs) across 928 member networks, with approximately 60 Tbps installed capacity. However, these exchanges remain concentrated in only 7 major cities, despite India having more than 168 cities with populations above 300,000. This concentration creates structural inefficiencies, particularly for AI inference and low-latency applications. Much of India's domestic internet traffic still routes internationally through hubs such as Singapore before returning to local destinations, increasing latency and reducing network efficiency. Current IXP utilisation remains only 5–6%, compared to 40–60% utilisation in mature Western markets.

At the same time, India's long-haul optical fiber backbone has expanded rapidly. Total optical fiber route length reached approximately 4.24 million route-km by end-2025, up from 1.94 million km in 2019, making India one of the largest fiber markets globally.
Major contributors to this buildout include:

- BSNL backbone infrastructure
- RailTel railway corridor fiber
- PowerGrid transmission-linked fiber
- BharatNet rural connectivity program

BharatNet alone represents one of the world's largest digital infrastructure investments at approximately ₹1.39 lakh crore (~$16.9B), connecting more than 2.14 lakh Gram Panchayats by 2026. This expansion has significantly reduced data costs from approximately ₹269/GB to ₹8–10/GB, while helping India surpass 1 billion broadband subscribers.

International connectivity is also scaling rapidly. India's lit subsea bandwidth capacity increased from 139 Tbps (2022) to 309 Tbps (2025), representing approximately 60% YoY growth in 2025 alone. Several major cable projects... including SEA-ME-WE-6, MIST, 2Africa Pearls, and Jio's IAX/IEX systems... are expected to add substantial additional capacity over the next few years.

At the data center layer, operational IT load capacity reached approximately 1,500 MW by end-2025, with a record 387 MW added in 2025 alone, reflecting a 103% YoY increase over 2024 additions. This acceleration is being driven largely by hyperscaler and AI infrastructure demand.

- However, the long-term challenge is not simply adding capacity... it is ensuring that new infrastructure supports GPU-dense, high-power AI workloads, rather than only general-purpose cloud deployments.

Metric	Detail
Active IXPs	**31**
IXP Member Networks	**928**
IXP Installed Capacity	**~60 Tbps**
IXP Utilisation	**5–6%**
Fiber Route Length (2025)	**4.24M route-km**
Fiber Route Length (2019)	**1.94M route-km**
BharatNet Investment	**₹1.39 lakh crore (~$16.9B)**
Gram Panchayats Connected	**2.14 lakh+**
Data Cost Reduction	**₹269/GB → ₹8–10/GB**
Broadband Subscribers	**1B+**
Subsea Lit Capacity (2025)	**309 Tbps**
Subsea Lit Capacity (2022)	**139 Tbps**
Operational DC Capacity (2025)	**1,500 MW**
New DC Capacity Added (2025)	**387 MW**
Projected DC Capacity (2030)	**4,000 MW**

5.4 Talent Ecosystem

Talent is the human infrastructure layer behind India's AI economy. While compute, power, and networking determine technical capacity, long-term AI competitiveness ultimately depends on the availability of engineers, researchers, operators, and execution talent capable of building and maintaining AI systems at scale.

India's AI talent ecosystem has expanded rapidly, with the broader AI/ML workforce reaching approximately 2.75 million professionals in 2025, reflecting a 55% YoY increase. However, the core frontier-level expertise pool remains significantly smaller, estimated at approximately 600,000–650,000 professionals, and projected to grow to 1.25M+ by 2027 at approximately 15% CAGR.

This creates a structural mismatch:

- AI demand growth: 25–35% CAGR
- Frontier talent growth: ~15% CAGR

As a result, demand for specialized AI talent is now growing materially faster than supply.

Geographically, ***Bengaluru remains India's dominant AI talent hub with approximately 478,620 AI/ML professionals,*** supported by a strong inflow advantage over other cities. Hyderabad and Chennai are also scaling rapidly, while Tier-2 cities such as Coimbatore (+72%) and Ahmedabad (+65%) are emerging as new talent clusters.

At the infrastructure operations layer, Site Reliability Engineering (SRE) has become one of the fastest-growing technical disciplines. India currently has more than 23,000 active SRE job openings, with approximately 1,852 new listings added weekly. Demand is being driven directly by hyperscale cloud expansion, GPU infrastructure deployment, and AI operations complexity.

The most acute shortage now exists at the intersection of:

- AI infrastructure operations
- GPU thermal management
- CUDA-level debugging
- High-density compute reliability

These are highly specialized skill sets with limited institutional training pipelines.

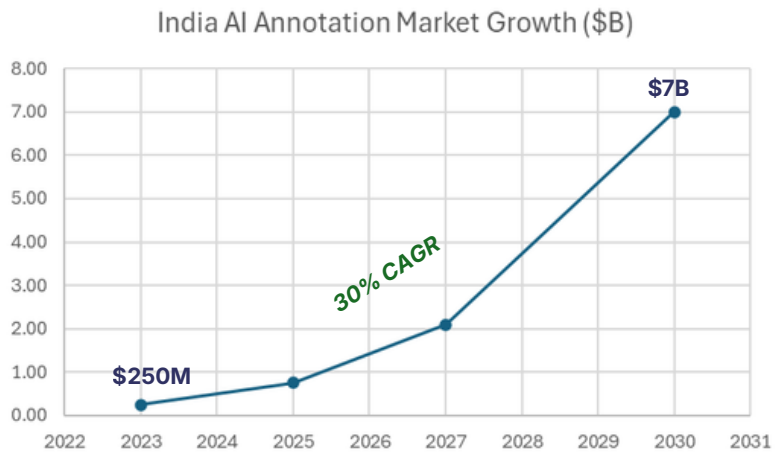
At the physical infrastructure layer, India's data center engineering workforce currently stands at approximately 25,000–35,000 professionals, but will need to nearly double by 2027 to support ongoing capacity expansion. Skill shortages are especially severe among high-voltage power engineers and experienced infrastructure operators.

India's academic pipeline remains globally competitive. In 2024–25 alone, approximately 390,245 students enrolled in Computer Science and Engineering programs across 1,600+ institutions offering AI/ML curricula.

INDIA ALSO RANKS:

- #1 Globally in AI skill penetration
- #1 In AI talent concentration
- #5 In AI scientific publications

At the data layer, India has also emerged as the global leader in AI data annotation and labeling services. ***Approximately 70,000 professionals currently work in annotation-related roles,*** while the market is projected to expand from $250M (2023) to approximately $7B by 2030.



(Vector databases are scaling rapidly (1.66 → 7.34B), becoming a critical layer for AI retrieval and GenAI applications.

Metric	Detail
India AI/ML Workforce (2025)	**~2.75M professionals**
Core Deep-Expertise Talent Pool	**600K–650K**
Projected Frontier Talent (2027)	**1.25M+**
Frontier Talent Growth	**~15% CAGR**
AI Market Growth	**25–35% CAGR**
Bengaluru AI Workforce	**478,620 professionals**
Active SRE Job Openings	**23,000+**
Weekly New SRE Listings	**1,852+**
Data Centre Engineering Workforce	**25K–35K**
CSE Admissions (2024–25)	**390,245 students**
Institutions Offering AI/ML	**1,600+**
India Annotation Market (2023)	**$250M**
Projected Annotation Market (2030)	**$7B**
Global Annotation Tools Market CAGR	**20.71%**

INVESTMENT OUTLOOK

- **Compute & Infrastructure Scaling:** Demand growth is severely outpacing supply readiness, driving high-yield opportunities into data center expansion, hyperscale campuses, cloud-linked platforms, and localized sovereign GPU infrastructure, leasing, and domestic assembly to reduce global import dependence.

- **Behind-the-Meter Power Moats:** Because electricity availability and pricing dictate compute economics, value is shifting to operators securing low-cost, scalable long-term access via captive solar/wind infrastructure, hybrid renewable systems + Battery Energy Storage Systems (BESS), grid modernization, smart substations, and energy management platforms to maintain grid stability.

- **Latency & Edge Connectivity:** High-throughput connectivity matches compute in structural importance, moving capital into high-density ecosystems including Internet Exchange Points, metro/long-haul fiber, subsea cables, and edge interconnection for real-time inference workloads, alongside a major distributed expansion runway in underpenetrated Tier-2 regions.

- **Global Talent Supply Anchor:** Backed by scale, cost efficiency, English proficiency, and a massive STEM pipeline, India is a dominant global human capital hub. A widening frontier supply-demand mismatch grants immense pricing power to platforms driving AI/ML upskilling, university-industry research partnerships, data annotation, enterprise staffing, and specialized SRE infrastructure operations talent.

6. FINAL INVESTMENT ANALYSIS

6.1 Seven-Year Spending Forecast

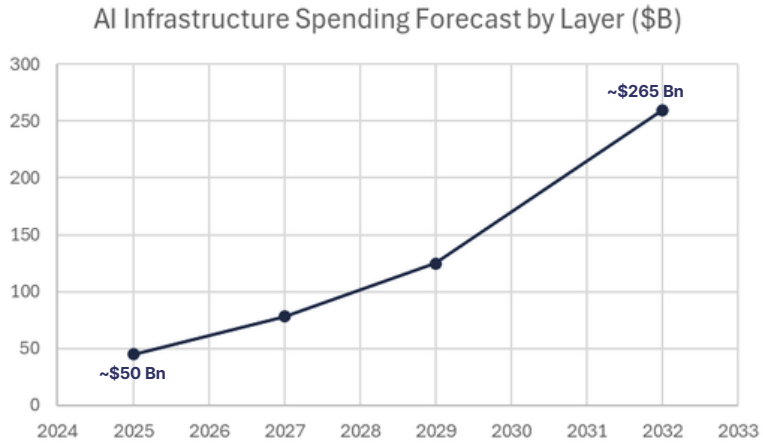
India's AI infrastructure cycle is entering its largest capital deployment phase to date. Over the next seven years, spending is expected to shift from experimentation toward full-scale industrial deployment across compute, energy, networking, cooling, and enterprise AI operations.
The investment cycle will not be concentrated in one layer alone. Instead, AI growth will require synchronized expansion across the entire pipeline:

- Compute infrastructure
- Power and energy systems
- Networking and connectivity
- Cooling and thermal management
- Software deployment layers
- Professional services and operations

The strongest spending acceleration is expected in infrastructure-heavy segments because AI workloads are fundamentally constrained by physical capacity.

7-Year Spending Forecast (India AI Infrastructure Ecosystem)

Layer	2025 Base Estimate	2032 Forecast	Implied CAGR	Rationale
Compute Infrastructure	**$8–10B**	**$45–60B**	**~28–31%**	India DC capacity projected from ~1.5 GW → 4–5 GW+, GPU
Power & Energy	**$12–15B**	**$70–90B**	**~29–32%**	AI power demand + renewable + BESS
Networking & Fiber	**$6–8B**	**$30–40B**	**~25–28%**	5G densification + edge growth + subsea scaling
Cooling & Thermal	**$2–3B**	**$18–25B**	**~33–36%**	Shift from air cooling → liquid & immersion
Industrial Land & Parks	**$4–5B**	**$20–30B**	**~24–29%**	Premium digital corridor expansion
Software Infrastructure	**$5–7B**	**$35–50B**	**~31–34%**	Enterprise AI deployment + MLOps growth
Professional Services	**$8–10B**	**$40–55B**	**~24–28%**	Managed AI services + integration demand



6.2 Current Indian Capacity Estimate

India's current AI infrastructure base is expanding rapidly, but remains relatively early-stage compared to the scale of projected demand.

Infrastructure Layer	Current Capacity (2025–26)
Operational Data Center Capacity	**~1.5 GW**
Installed Power Capacity	**~524 GW**
Subsea Lit Capacity	**309 Tbps**
Fiber Backbone	**4.24M route-km**

GPUs Onboarded	**38,000+**
AI/ML Workforce	**2.75M professionals**
Internet Exchanges	**31 IXPs**
Renewable Capacity	**262+ GW non-fossil**

Despite strong growth, most infrastructure layers remain concentrated in a small number of cities and dependent on imported hardware and external supply chains.

6.3 Future Demand Projection

India's future demand profile is being shaped by simultaneous expansion across:

- AI inference workloads
- Cloud adoption
- Enterprise digitization
- 5G rollout
- Edge infrastructure
- Sovereign AI initiatives

Projected demand by 2030–2032 includes:

Demand Layer	Projected Requirement
Data Center Capacity	**4–5 GW+**
GPU Demand	**Several hundred thousand AI GPUs**
Renewable Integration Need	**500 GW+**
Battery Storage Need	**411 GWh+**
Subsea Bandwidth Demand	**10x current traffic levels**
AI Talent Requirement	**1.25M+ frontier professionals**
Edge Data Center Demand	**19%+ CAGR growth**

AI demand growth is expected to materially outpace infrastructure readiness in nearly every major layer.

6.4 Gap Analysis

The gap between current infrastructure capacity and future AI demand is now becoming structurally significant.

Layer	Current Status	Future Gap
GPU Supply	Import-dependent	Severe shortage risk
Power Grid	Strong generation	Transmission + stability bottlenecks
Cooling	Mostly air-cooled	AI-density readiness gap
Networking	Strong backbone	Weak Tier-2 interconnection
Talent	Large workforce	Frontier skill shortage
Land	Large supply	Infra-ready scarcity
Data Centers	Rapid expansion	AI-grade capacity shortage

The key issue is not lack of demand... it is the speed at which infrastructure can scale to meet that demand.

6.5 Critical Bottlenecks

Across the AI infrastructure stack, several bottlenecks repeatedly appear:

1. GPU & Semiconductor Dependency

- Imported GPUs dominate supply
- HBM memory shortages remain critical
- Advanced packaging capacity constrained globally

2. Power & Grid Stability

- Renewable intermittency
- Transmission congestion
- Industrial tariff burden
- Storage infrastructure deficit

3. Cooling & Density Constraints

- Air cooling inadequate for AI workloads
- CDU and liquid cooling shortages

High retrofit costs

4. Connectivity Concentration

- IXPs concentrated in few cities
- Tier-2 interconnection weakness
- Subsea landing concentration risk

5. Talent Deficit

- Frontier AI talent growth slower than demand
- SRE and AI infra operations shortages
- Limited GPU-level operational expertise

6. Execution Delays

- Land approvals
- Power connectivity timelines
- Infrastructure integration complexity

6.6 High-Conviction Investment Opportunities

The strongest investment opportunities are concentrated in layers where:

- Demand visibility is extremely high
- Supply remains constrained
- Revenue is recurring or infrastructure-linked
- Switching costs are strong

Theme	Why It Matters
AI Data Centers	Core compute infrastructure layer
Captive Renewable + BESS	AI power economics advantage
GPU Cloud & Sovereign Compute	Hardware scarcity opportunity
Liquid Cooling Infrastructure	Mandatory for high-density AI
Edge & Fiber Networks	Real-time AI connectivity demand
Managed AI Services	Sticky recurring enterprise revenue
Industrial Land Corridors	Infrastructure-ready scarcity
AI Operations & SRE Services	Long-term operational dependency

FINAL STRATEGIC CONCLUSION

The core conclusion across the entire report is consistent:
India's AI opportunity is fundamentally an infrastructure opportunity.

The country already possesses:

- Massive digital demand
- A large technical workforce
- Strong services capability
- Rapid infrastructure expansion momentum

However, the long-term winners will not simply be software companies or AI model builders. The durable value will accrue to firms controlling the hard constraints of the ecosystem:

- Compute
- Power
- Cooling
- Connectivity
- Infrastructure-ready land
- AI operations and execution

The AI economy is becoming increasingly physical, capital-intensive, and infrastructure-dependent. Whoever controls the bottlenecks will control the economics of India's AI future.

RESEARCH BY-

Hardik Jain
Project director

Rashi Jain
Research Lead

Yuvraj Singh
Research Lead

Shikhar Omer
Research Lead

Suhani Kharbanda
Research Lead

Devansh Khandelwal
Associate

Jishnu Narayan
Associate

Suhani Singh
Associate

Chirag Arora
Associate

Rashi Jain
Associate

Kartik Bhatia
Associate

Aryan
Associate

Diya Gera
Associate

SOURCES AND CITATIONS

- Synergy Research Group; Gartner, global cloud infrastructure and IaaS market share estimates, 2024–2025.
- Washington Post, report on Amazon, Microsoft, and Google infrastructure investment pledges to India, 26 December 2025.
- Ministry of Electronics and Information Technology (MeitY), Lok Sabha reply on India's data-centre capacity, 18 March 2026.
- Savills India, India Data Centre Capacity More Than Doubled to 387 MW in 2025, Projected to Triple by 2030, January 2026, via ANI News,
- JLL India, India Data Centre Market H1 2025, capacity and forecast to 2027.
- Nomura, India data-centre capacity forecast to 9.2 GW by 2030.
- Institute for Energy Economics and Financial Analysis (IEEFA) and Jefferies, India data-centre capacity build and investment requirement estimates, 2024–2025.
- IndiaAI Compute Portal, Government of India, GPU deployment statistics (38,231 GPUs by December 2025).
- Press Information Bureau, Government of India, statement by Union Minister Ashwini Vaishnaw on IndiaAI GPUs (34,000 GPUs milestone, 18,417 existing + 15,916 new), May 2025.
- NASSCOM and Oliver Wyman, India Cloud Market Report 2022.
- IDC, India Public Cloud Services Tracker, H1 2024 (published July 2025), USD 10.9 billion public cloud market and top-5 CSP share.
- rahulmalodia.com, "AWS vs Azure vs Google Cloud India 2026" (AWS ~40% share of India's cloud services market in 2024)
- Gartner, global IaaS market share of Big-3 hyperscalers (71% share, 2024).
- Mordor Intelligence, India AI Data Centre Market 2025 (cloud providers' 55.82% share).
- *Civo, "2026 Indian Cloud Market Trends", 2026.*
- *NASSCOM, AI Adoption Index 2025 (share of Indian enterprises deploying AI via public cloud).*
- *JLL India, India Data Centre Update H1 2025, BFSI hybrid cloud trend and AWS region timelines.*

- Expert Market Research, citing IDC, corporate bank cloud adoption in India (80% on cloud in 2024).
- Press Information Bureau, Ministry of Electronics and Information Technology, Ayushman Bharat Digital Mission (ABDM) migration to NxtGen Chennai data centre for data localisation.
- IndiaAI Mission, Government of India, guidance on sovereign and government cloud workloads.
- Reserve Bank of India, Cloud Policy for regulated entities, November 2024.
- Ministry of Power, Government of India, “India achieves 50% non-fossil fuel capacity target ahead of COP-26 commitment”, Press Information Bureau release, 29 November 2025, https://pib.gov.in/PressReleasePage.aspx?PRID=2209478.
- Institute for Energy Economics and Financial Analysis (IEEFA), “India’s Power-Hungry Data Centre Sector at a Crossroads”, 2024, https://ieefa.org.
- Petroleum Planning and Analysis Cell (PPAC), Ministry of Petroleum and Natural Gas, India natural gas production and demand statistics 2024.
- Institute for Global Public Policy (IGPP), “The Energy Cost of India’s Data Centre Expansion”, 2025.
- Astute Analytica, India Data Centre Market report, February 2026 (Mumbai >52% share of DC capacity end-2025), via GlobeNewswire.
- Council on Energy, Environment and Water (CEEW), “How Is Data Centre Infrastructure in India Shaping Power and Water Use?”, February 2026.
- Grid-India (National Load Despatch Centre), “Overview of National and Regional Load Despatch Centres”.
- Power Grid Corporation of India Limited (PGCIL), Inter-State Transmission System overview.
- Central Electricity Regulatory Commission (CERC), regulations and framework for GNA and power markets
- IEEFA and JMK Research, “Green Power Transmission Development in India”, 24 September 2025
- Down to Earth, “Green Corridor Projects Failing to Evacuate Renewable Energy”, 24 September 2025
- Central Electricity Authority (CEA), statements on 600 GW renewable integration target by 2032 and Green Energy Corridor phases.

- Down to Earth, coverage on Green Energy Corridor Phases 3 and 4 (Rs 56,000 crore proposal), January 2026
- Union Budget 2026, Government of India, allocation of Rs 600 crore for Green Energy Corridor intra-state transmission lines, Press Information Bureau.
- GlobalPetrolPrices.com, “India Electricity Prices, June 2025”,
- BESCOM, BSES Rajdhani, CESC, KSEB, MSEDCL, TANGEDCO, state DISCOM tariff documents compiled in GlobalPetrolPrices dataset.
- Karnataka Electricity Regulatory Commission (KERC), FY2024–25 Tariff Order, as summarised by Mercom India, May 2024
- Council on Energy, Environment and Water (CEEW), “Tariff Reforms Are Reshaping Power Markets”, January 2026
- TradeBrains, “Data Centers in India: Energy Costs and Policy”, December 2025
- Institute for Energy Economics and Financial Analysis (IEEFA), “India’s record 59 GW renewable energy auction in 2024”, December 2024
- Ministry of New and Renewable Energy (MNRE), Government of India, renewable capacity and auction data
- Council on Energy, Environment and Water (CEEW), open-access and captive renewable energy market analysis, January 2026
- Mercom India, coverage of C&I renewable open-access growth and data-centre PPAs (Nxtra, STT GDC, AdaniConneX), August 2024
- Central Electricity Authority, “National Electricity Plan 2023”, battery energy storage targets (411 GWh / 74 GW)
- Rystad Energy, India battery energy storage opportunity (USD 487 billion, ~2.56 GW installed BESS), November 2025, cited via ESS News
- S&P Global, GPU data-centre rack power density estimates for AI clusters, September 2025
- Next Move Strategy Consulting, GPU data-centre cluster power-density and market outlook, December 2025.
- Government of Andhra Pradesh, “AP Data Centre Policy 4.0”, industrial tariff incentives for AI data centres, as summarised in TradeBrains, December 2025.
- Takshashila Institution, “India’s AI Data Centre Construction Cost Advantage”, October 2025

- Institute for Energy Economics and Financial Analysis (IEEFA), “More than 40 GW of Renewable Projects Stranded without Power Purchase Agreements”, December 2024,
- Credence Research, “India Data Center UPS Market Report”, November 2024
- Global Market Insights, “Data Center Infrastructure Management (DCIM) Market Report 2024”
- Vertiv India, data-centre UPS and infrastructure solutions
- Schneider Electric India, EcoStruxure IT and data-centre infrastructure solutions,
- Eaton India, data-centre UPS and power management solutions
- Legrand / Numeric India, UPS solutions for data centres
- Delta Electronics, data-centre UPS and power solutionsom.
- Global Market Insights, DCIM market share data (Schneider Electric >6% share), 2024.
- Credence Research, India Data Centre Market report (key DC power management players)
- Mordor Intelligence, India Data Centre Power Market, May 2024
- Honeywell India, Honeywell Forge / EBI building management and DC energy management
- Sterlite Technologies, optical fibre and power management infrastructure for data centres
- Ministry of Power, Revamped Distribution Sector Scheme (RDSS) for smart metering, Press Information Bureau
- Internet Society, Pulse IXP Tracker – India, February 2026
- Unnamed author, “IX Research Paper, 2023–2024”, Scribd document on Indian IXPs and peering
- DE-CIX India, “Internet Exchange Point” pages
- PeeringDB, India IXP and network listings
- Government of India, Digital Personal Data Protection Act 2023.
- Press Information Bureau, Ministry of Communications, BharatNet coverage status, article “BharatNet connects 2.15 lakh Gram Panchayats: India’s digital infrastructure”, March 2026, via Blitz India Media
- Press Information Bureau, “BharatNet Project” release
- Universal Service Obligation Fund (USOF), “BharatNet Project”, May 2025

- Communications Today, “Over 2.14 lakh Gram Panchayats connected under BharatNet: Scindia”, February 2026
- Tele.net.in, “India targets investment-friendly regime for submarine cables, says TRAI chairman”, February 2026
- Outlook Business, “India must build 10x more cable landing stations to compete in global data race: TRAI chief”, March 2025
- Submarine Networks, “Cable Landing Stations in India”
- Observer Research Foundation (ORF), “A roadmap for securing India’s undersea cables”
- India Briefing, “India: Three Subsea Cable Projects Launch March 2025”, March 2025
- Cushman & Wakefield, “India Data Centre Update H1 2025”
- PeopleLogic Business Solutions, “India’s AI-ML Talent Market: What the Numbers Actually Tell Us”, 2025
- NASSCOM, “State of Data Science & AI Skills: India – Data and Art of Smart Intelligence”
- Deloitte and NASSCOM, “Bridging the AI talent gap to boost India’s tech and economic impact”, August 2024
- IndiaAI.gov, “India’s AI talent pool to grow to 1.25 million by 2027 – NASSCOM–Deloitte India report”
- LinkedIn, India Site Reliability Engineer job listings
- Coursera, “Site Reliability Engineer Salary India: 2026”
- Vinsys, “SRE and Role of SRE Engineer”,
- Makers Now, “India’s Data Infrastructure Boom Will Create an Entirely New Talent Segment”, March 2026
- Alpha Apex Group, “Top Data Center Roles 2026”
- Lovely Professional University (LPU), “Top 10 CSE & AI Colleges in India for 2026 to Launch Your Career in Computer Science & Artificial Intelligence”
- Communications of the ACM, “AI and Data Science Centers in Top Indian Academic Institutions”
- *iMerit, “AI Data Jobs Attract Highly Skilled Indians”, 2025*
- *IMARC Group, “Data Annotation Tools Market”*
- *Ken Research, “India Data Annotation Tools Market”*
- *Grand View Research, “Data Collection and Labeling Market Size, Share & Trends”, 2024*

- Sify Technologies, Sify Infinit Spaces, data-centre portfolio
- Adani Enterprises / AdaniConneX, corporate disclosures on 1 GW AI data-centre campus and pan-India DC pipeline
- WifiTalents, “Edge Computing and AI Workloads 2026” (enterprise data at the edge and device counts), 2026
- Business Insider, article on AI workload migration to edge infrastructure, 2025–2026.
- Axios, coverage of global edge computing market reaching USD 200+ billion annually, 2025.
- IndustryResearch (industryresearch.biz–style report), global edge data-centre count growth 2022–2026.
- Snowflake Inc., FY2025 Q4 earnings release and investor presentation, product revenue and customer metrics (USD 943.3 million product revenue, 2025)
- Amazon Web Services, Amazon Redshift pricing documentation (provisioned and serverless pricing, 3-year reservations), aws.amazon.com/redshift/pricing
- Academic research on Snowflake micro-partition pruning efficiency (99.4% reduction in scanned data) – unnamed paper cited in the document.
- India AI market forecasts from MRFR-style reports (Market Research Future or similar), 2024–2030, cited for India enterprise AI market size and CAGR.
- National Payments Corporation of India (NPCI), UPI transaction volumes and growth statistics used to illustrate real-time AI workloads in payments.
- Press Information Bureau, Government of India, Press Release on India’s AI compute and data-centre policy, PRID=2209737
- E2E Networks Ltd., independent Indian GPU-cloud provider
- Yotta Data Services, “Yotta Shakti Cloud” brochure, Scribd
- IndiaAI.gov, “SCIKIQ” startup profile