HyperNext Data Centers

RESEARCH PAPER

HN-RP-007

The India AI Compute Gap

Capacity, sovereignty, and the decade ahead

Where the GW are coming from, who needs them, and what will not exist.

This is what is Next.

Series	HyperNext Research
Paper	HN-RP-007
Issued	13 May 2026
Version	1.0
Classification	Public release
Citation	HyperNext Research, HN-RP-007

The India AI Compute Gap

This paper is part of the HyperNext Research series. Methodology, assumptions, and source data are stated openly so other operators can reproduce the analysis on their own facilities. Citation as "HyperNext Research, HN-RP-007" is welcome.

Contents

1. Preview

> PREVIEW, FORTHCOMING Q3 2026
>
> This paper, scheduled for release in the third quarter of 2026, will examine the gap between India installed AI compute capacity in 2026 and the capacity the country will need by 2030 on the projected adoption curve. The working numbers, subject to revision as the analysis matures, suggest a gap of between 6 and 23 gigawatts of equivalent capacity that has to be financed, built, powered, and operated within the four-year horizon. The paper will work through the demand side honestly (enterprise inference, sovereign government workloads, BFSI, the consumer AI surface that an Indian population of 1.4 billion is starting to use at scale) and the supply side (what is being built, by whom, with what financing model, and at what pace). It will examine what international hyperscalers will and will not fund inside India under the current regulatory environment, and identify the capacity that must come from Indian operators with Indian capital. The argument: the next five years are when the structure of India AI supply is fixed for the following twenty.

This is a preview release. The full paper will be issued in Q3 2026. Subscribe via the research page to be notified.

What the paper will cover

The full paper is in active draft and will cover the following structure.

- The current installed AI compute capacity in India in 2026: approximately 2 GW across all operators, broken down by region, by tenant type, and by GPU generation.
- The demand forecast for 2030 across four scenarios. A base case of 12 GW. A constrained case of 8 GW (if adoption stalls). An expansion case of 18 GW (if AI products embed rapidly across the economy). An extreme case of 25 GW (if generative AI becomes the dominant consumer interaction modality).
- The supply side build-out: a survey of announced and committed capacity from HyperNext, Yotta, AdaniConneX, Sify, NTT, ST Telemedia GDC, CtrlS, and the international hyperscalers operating Indian regions.
- Financing analysis. Where the capital for this build-out can come from. What financial structures are working in the Indian market. Where the bottlenecks are.
- The regulatory analysis. What the current sovereignty regime permits the international hyperscalers to build, what it does not, and how this constrains the share of demand they can serve.
- The argument. That the supply structure being established between 2026 and 2030 will determine who serves Indian AI demand for the following two decades, and that this is a structural moment that does not recur.

Methodology note

The paper will be based on public data (regulatory filings, company announcements, industry research from Anarock, IDC, and 451 Research) supplemented by HyperNext market intelligence on capacity ramps across the industry. Where projections are made, the methodology will be open. The assumptions will be stated explicitly so that other operators can re-run the analysis with their own inputs. Forecast intervals will be reported as ranges, not point estimates.

The paper will not be a marketing document. It will state findings that may not favour the HyperNext commercial position, and it will identify areas where the industry will need to coordinate rather than compete. This is consistent with the editorial stance of the HyperNext Research series. The work matters more than the positioning.

TO RECEIVE HN-RP-007 ON RELEASE

> Subscribe on the research page at www.hypernxt.com/research.

> The paper will be released in Q3 2026 and emailed to subscribers on the day of release.

> Updates between now and release will be posted to the LinkedIn page of HyperNext Data Center Limited.

2. Methodology preview

The full paper will work through demand and supply analysis on the methodology summarised below. We publish the methodology preview now so that other operators, researchers, and policy stakeholders can comment before the analysis is finalised.

Demand-side methodology

The Indian inference demand forecast for 2030 is built bottom-up across ten verticals (BFSI, government, telecom, healthcare, retail, software/SaaS, consumer AI, education, manufacturing/logistics, and emerging/other). For each vertical, the demand is modelled as:

```
INFERENCE DEMAND BY VERTICAL
  Demand (MW) = N_users
              × tokens_per_user_per_year
              × joules_per_token
              / (8760 hours × 3.6 × 10^9 joules/MWh)
              × (1 + safety margin)
              × (1 / utilisation factor)

Where:
  N_users                = vertical-specific user count, 2030 forecast
  tokens_per_user_per_year = average inference consumption per user
  joules_per_token       = energy per token at expected 2030 hardware generation
  safety margin          = 20% capacity headroom for peak handling
  utilisation factor     = expected average utilisation, 0.65 to 0.85 depending on vertica
```

The model produces a range for each vertical. The vertical ranges are summed (with explicit covariance handling) to produce the four scenarios mentioned in Section 1.

Supply-side methodology

The supply forecast is built from publicly announced capacity by operator, adjusted for execution risk based on historical industry delivery against announcement. We score each operator on three factors:

- Land secured (yes/no with documentation reference)
- Power allocation in place (MW allocated, source, expected energisation date)
- Hardware order placed (vendor confirmation of order book position)

An operator with all three secured for a given capacity tranche is treated as 90 percent likely to deliver on the stated timeline. An operator with two of three is treated as 60 percent likely. An operator with one or none is treated as 30 percent likely. The probability-weighted aggregate gives the expected 2030 supply.

Sovereignty filter

The supply analysis is then run through the sovereignty filter from HN-RP-003. Capacity that satisfies all three layers (residency, jurisdiction, operational control) is counted toward the sovereign supply. Capacity that satisfies layers 1 and 2 but not 3 is counted toward the partial-sovereign supply. Capacity that satisfies only layer 1 is treated as non-sovereign for the purposes of regulated workload demand.

The mismatch between sovereign-compliant supply and the share of 2030 demand that requires sovereignty is the headline finding the full paper will quantify. The working number, subject to revision, is that approximately 60 percent of 2030 inference demand will require some level of sovereignty compliance and approximately 35 percent will require all three layers. The currently-announced supply that satisfies all three layers is less than half of that 35 percent share.

3. Demand by vertical

The demand build-up across ten verticals is below. The numbers are the result of bottom-up modelling described in Section 2, with each vertical reviewed separately by domain analysts. Where uncertainty bands are wide, both the central forecast and the band are reported.

Vertical	2030 demand low (MW)	2030 mid (MW)	2030 high (MW)	Workload signature
BFSI	1,400	2,300	3,400	Chat-based servicing, fraud, document AI, KYC
Government and PSU	800	1,400	2,200	Citizen services, language platforms, internal analytics
Telecom	600	1,000	1,600	Customer service, network operations, voice translation
Healthcare	400	800	1,500	Radiology, diagnostics, drug discovery
Retail and e-commerce	500	900	1,500	Recommendation, search, customer assist
Software and SaaS	1,200	2,400	4,500	Indian-headquartered SaaS serving global customers
Indian consumer AI	1,000	2,500	5,500	Local-language assistants, search, content
Education	200	500	1,200	Tutoring, content generation, assessment
Manufacturing and logistics	300	700	1,400	Vision, process optimisation, predictive maintenance
Other and emerging	800	1,500	3,200	Including verticals not yet identified
Total	**7,200**	**14,000**	**26,000**	Equivalent inference capacity

BFSI: the lead vertical

Banking, financial services, and insurance is the first vertical to move at scale because it has the combination of three forces. Regulatory pressure pushing toward domestic inference for personal financial data. Customer service volume that lends itself to AI assist at unit economics that work. Risk management workloads (fraud, AML, credit scoring) where the lift from current rules-based systems to AI-assisted systems is large.

Six Indian banks have publicly committed to deploying production AI inference workloads at scale by 2027 or 2028. The combined capacity demand from those six is approximately 600 MW IT load by 2028, rising to 1,200 MW by 2030 as the workload matures. Insurance majors add roughly the same again. NBFCs and fintech add another 300 to 600 MW. The aggregate 2030 demand falls in the 1.4 to 3.4 GW band depending on how aggressively the sector adopts and how much of the inference can be served by smaller dedicated models versus the larger general-purpose ones.

Government and consumer AI

The two largest single-vertical bands by 2030 are software/SaaS (Indian-headquartered software companies serving global customers) and Indian consumer AI (local-language platforms and assistants). Both have upper bounds above 4 GW under aggressive adoption scenarios, and the consumer AI vertical specifically has the widest uncertainty because it depends on the rate at which Indian consumers adopt AI as part of everyday transactions in commerce, education, and entertainment.

Government and PSU demand is more predictable. Citizen-services platforms (passport, taxation, healthcare, welfare) are migrating to AI-assisted versions on programme timelines that are public. Language platforms are well advanced for several Indian languages. Total government demand by 2030 sits between 800 MW low case and 2.2 GW high case.

What the variables are

The variance between low, mid, and high cases is driven by three forces. The pace of model capability improvement, which determines how much inference is needed per user transaction. The unit economics of inference, which determine which use cases pencil out. The regulatory environment, which determines what fraction of workload must be served from Indian capacity (the topic of Section 5).

The low case assumes slow capability improvement, unit economics that stay tight, and regulatory pressure that levels off. The high case assumes rapid capability improvement (with consequent demand expansion), unit economics that improve materially, and continued regulatory pressure to domesticate workload. The mid case is the linear interpolation. We weight the three scenarios roughly equally for the headline 14 GW figure.

4. Supply assessment

The supply side is the bottom-up commitment summary across the eight operators currently building serious capacity for the Indian AI market. The numbers are based on public announcements supplemented by the execution-risk scoring described in Section 2. Operators are listed in alphabetical order, not by capacity ranking.

Operator-by-operator commitment summary

Operator	Announced 2026-2030 capacity (MW IT)	Land secured	Power allocated	Hardware ordered	Execution-risk score
AdaniConneX	2,000+	Yes (3 sites)	Partial (2 of 3)	Partial	Medium
CtrlS	800	Yes	Yes	Yes (first phase)	Low
HyperNext	1,560 (Hyd 250 + Kak 1,200 + NR 100, plus 110 South Africa, plus 5 BTS)	Yes (all India sites)	Yes (Hyd + Kak)	Yes (first phases)	Low
NTT Global Data Centers India	600	Yes	Yes	Yes (first phase)	Low
Sify Infinit	500	Yes	Partial	Partial	Medium
ST Telemedia GDC India	700	Yes	Yes	Yes (first phase)	Low
Yotta Data Services	1,500+	Yes (3 sites)	Yes (Maharashtra)	Yes (first phase)	Low
International hyperscaler regions in India	1,200 (combined estimate)	Mixed	Mixed	Yes	Medium
Total committed	**8,800+**	—	—	—	—

The headline number is approximately 8.8 GW of committed Indian AI compute capacity by 2030 across all operators, with execution-risk weighting applied to bring the expected delivered capacity to roughly 6.5 to 7.2 GW.

The execution-risk discount

The 8.8 GW announced figure is not what will get delivered. Indian data centre delivery against announcement historically runs 70 to 85 percent over five-year horizons, with the gap coming from land delays, power transmission delays, supply chain timing, and project economics changing during the build-out. The execution-risk-weighted figure for 2030 delivered capacity is approximately 6.8 GW central, with a 6.0 to 7.5 GW band.

The discount falls disproportionately on operators with one of three risk markers. Land that is not yet under registered title. Power allocation that depends on transmission infrastructure not yet under construction. Hardware orders that have not been confirmed by the vendor (which means the operator is not in the priority queue and may face 18 to 36 month additional lead times for the rack-scale platforms).

The geographic distribution

Indian AI compute capacity by 2030 will be concentrated in five regions. Hyderabad and southern Andhra Pradesh, where land, power, and water are all available at the scale needed and where state government engagement is most advanced. Mumbai and the western Maharashtra corridor, where the proximity to BFSI customers justifies the higher land cost. Chennai and the Tamil Nadu corridor, where the international cable landings and the existing IT services concentration support demand. The Delhi NCR corridor, primarily for government and BFSI workloads. Bengaluru, but at modest scale because land economics and water availability constrain build-out.

Capacity outside these five regions is small in 2026 and will remain a small fraction of the total through 2030. The build-out is concentrated, not distributed.

What 6.8 GW supply against 14 GW demand means

The headline finding is that demand at the mid case exceeds risk-weighted supply by roughly a factor of two by 2030. The high case puts demand at almost four times supply. Even the low case, where demand is conservatively 7.2 GW, exceeds risk-weighted supply.

The gap will close in one of three ways. Indian inference workloads run on capacity outside India, with the data transit, latency, and sovereignty implications that creates. Indian capacity build-out accelerates beyond the current trajectory, which requires capital and execution discipline that has not yet been demonstrated at this scale. Demand is suppressed by infrastructure shortage, which means Indian AI products underperform their international peers in capability and availability.

None of those three outcomes is desirable for India. The argument of this paper is that the third outcome is the one that policy intervention can prevent, and that the policy window is between 2026 and 2028.

5. The sovereignty mismatch

The capacity gap analyses in Sections 3 and 4 assume any compute capacity can serve any workload. The reality is more constrained. Indian AI demand by 2030 includes a large share that must satisfy the sovereignty framework described in HN-RP-003 (residency, jurisdiction, operational control). Capacity that does not satisfy all three layers cannot serve that share. The mismatch between sovereign-compliant supply and sovereignty-requiring demand is the central finding of this paper.

How much demand needs sovereignty

Estimating the sovereignty-bound share of 2030 demand requires mapping each vertical against the regulatory framework.

Vertical	Sovereignty requirement	Share of 2030 demand (mid case)
BFSI (banking, NBFC, payments)	All three layers (DPDP + RBI + CII)	2,300 MW (100% sovereign)
Government and PSU	All three layers (CII designation)	1,400 MW (100% sovereign)
Insurance	Layers 1 and 2 (DPDP), 3 for designated entities	800 MW (85% sovereign)
Healthcare (regulated)	Layers 1 and 2, 3 for designated platforms	600 MW (70% sovereign)
Telecom (regulated)	Layers 1 and 2	700 MW (70% sovereign)
Retail and e-commerce	Layer 1 (DPDP)	270 MW (30% sovereign)
Software/SaaS (Indian-headquartered)	Variable by customer profile	600 MW (25% sovereign)
Indian consumer AI	Layer 1 for Indian users	500 MW (20% sovereign)
Education	Layer 1 (DPDP) for children data	200 MW (40% sovereign)
Manufacturing and logistics	Generally not sovereign-bound	140 MW (20% sovereign)
Other and emerging	Mixed	450 MW (30% sovereign)
Sovereign-bound demand (mid case)	—	**7,960 MW**

The mid case shows approximately 8.0 GW of Indian AI demand by 2030 that requires some level of sovereignty compliance. Of that, approximately 4.5 GW requires all three layers (full operational control inside India).

How much supply provides sovereignty

On the supply side, the same operator-by-operator analysis can be filtered for sovereignty compliance. An operator that is structurally subject to Indian jurisdiction, with operations entirely staffed inside India, can offer layer-three sovereignty. An operator that is a foreign hyperscaler with an Indian region cannot, regardless of how the residency commitment is structured.

Operator	2030 capacity (MW)	Layer 1 (residency)	Layer 2 (jurisdiction)	Layer 3 (operational)
HyperNext	1,560	Yes	Yes (Indian-incorporated, Indian-domiciled)	Yes (all-Indian operations team)
Yotta	1,500	Yes	Yes	Yes
AdaniConneX	2,000	Yes	Yes	Yes (with EdgeConneX JV nuances)
CtrlS	800	Yes	Yes	Yes
NTT GDC India	600	Yes	Indian-incorporated	Partial (parent-company access protocols)
ST Telemedia GDC	700	Yes	Indian-incorporated	Partial (Singapore parent)
Sify	500	Yes	Yes	Yes
International hyperscaler India regions	1,200	Yes (regional)	Partial (US parent)	No (operations from outside India)

Risk-weighted sovereign-compliant supply by 2030 (capacity that satisfies all three layers): approximately 5.5 GW central case.

Sovereign-bound demand (all three layers required): approximately 4.5 GW central case.

That looks balanced. It is not. The supply numbers are gross announcements. After execution-risk discount, the all-three-layer supply drops to approximately 3.8 GW. The demand for all three layers is firmer because it is set by regulatory floor, not by execution. The all-three-layers mismatch is approximately 700 MW by 2030 central case.

The broader sovereignty bound

Layer-one and layer-two demand (which is broader, including most BFSI and government and a significant share of healthcare and telecom) is approximately 8 GW by 2030. The corresponding supply, after risk discount, is approximately 6.5 GW. The mismatch is roughly 1.5 GW.

The implications are that approximately 700 MW of all-three-layers Indian AI demand cannot be served from sovereign-compliant supply by 2030 at central case, and approximately 1.5 GW of layer-one-or-two demand cannot be served either. Either the demand is suppressed (Indian AI products serve their users at degraded capability), or it migrates to foreign hyperscaler capacity (with the sovereignty implications that creates), or the supply build-out accelerates.

6. Capital allocation implications

The supply-demand mismatch is a market signal. For investors and capital allocators, the implications fall into three categories.

Infrastructure capital

Indian AI data centre capital requirement through 2030 is approximately INR 4.5 to 5.5 lakh crore (USD 50 to 60 billion at USD/INR 96). That is the all-in cost of the supply build-out described in Section 4. The capital is split roughly 60 percent for the data centre infrastructure itself (building, power, cooling, network) and 40 percent for the AI compute hardware (GPUs, server platforms, software stack).

The funding sources for this capital are emerging but not yet at the scale required. Domestic infrastructure debt and equity have raised approximately INR 80,000 crore for data centre purposes between 2022 and 2025. Foreign direct investment in Indian data centres has added another INR 60,000 crore in the same period. Sovereign wealth fund and large pension fund commitments to Indian data centre platforms total roughly INR 30,000 crore from announcements through 2025.

The aggregate is approximately INR 1.7 lakh crore of identified capital against a requirement of INR 4.5 lakh crore. The funding gap is large.

Project finance instruments

The capital gap will not close through equity alone. Project finance instruments specific to data centre development need to mature in India faster than they have. The recommendations the financial community is converging on include:

- Long-tenor infrastructure debt at 15 to 20 year horizons, matched against the asset life of the data centre rather than the typical 7 to 10 year tenor common today.
- Take-or-pay tenant contracts at investment grade or better, which become security for the underlying debt.
- Anchor-tenant pre-leases at construction phase, which de-risk the project economics and allow tighter debt pricing.
- Co-investment by sovereign wealth funds and large pension funds, which have the duration profile to match the asset class.

Implications for the smaller operators

The capital intensity of the AI rack-scale platforms (USD 5 million per rack at NVL576 generation) means that operators below approximately 200 MW IT total capacity face economics that may not be viable in the long run. Either they scale to a level where they can amortise the platform supply chain costs (network, vendor support, hardware refresh), or they specialise into a workload niche where the platform commitment is less central, or they exit the AI segment.

The Indian market has perhaps 25 operators today with some form of data centre footprint. The number with capacity meaningful at the AI scale described in this paper is closer to 8. Industry consolidation through 2030 is likely.

7. Policy implications

The policy actions that would move the supply-demand balance in the direction of meeting Indian demand from Indian capacity are mostly understood. They are not yet uniformly executed across the states where capacity is being built. The recommendations below are sequenced by urgency and by the level of government (central or state) at which the action sits.

Power transmission and allocation

Power is the binding constraint for data centre delivery in most Indian states. Land is available where states have prioritised data centre development. Water is manageable through the architectural choices described in HN-RP-001. Power allocation timelines and transmission infrastructure are not yet matched to the scale and pace required.

The specific actions that would help:

- Dedicated data centre power allocation queues at state distribution companies, with timelines of 12 months from application to allocation rather than the 18 to 36 month industry norm.
- Transmission infrastructure pre-build for designated data centre clusters, anticipating capacity ramps rather than reacting to them.
- Direct grid connection for facilities above 50 MW load, bypassing the distribution company layer and the cross-subsidy structure.
- Open access enablement for inter-state renewable PPAs, which is the mechanism most large data centre operators are using to procure renewable supply.

Sovereignty and data residency

The sovereignty framework described in HN-RP-003 is broadly working. The actions that would make it work better:

- Clarity on the timeline for DPDP Section 16 transfer restrictions notifications. The current uncertainty makes long-term workload placement decisions harder than they need to be.
- A defined CII designation process with predictable timelines and clear operational implications, so operators and customers know in advance what the layer-three compliance bar is.
- Recognition of the three-layer compliance framework (or an equivalent) in procurement guidance for central and state government purchasing, so government workloads are placed on infrastructure that actually satisfies the requirements rather than infrastructure that nominally satisfies layer one only.

Indian capital deployment

The capital gap identified in Section 6 will not be closed by foreign investment alone. Domestic capital deployment is the medium-term lever. The actions that would help:

- Insurance regulator (IRDAI) clarity on data centre infrastructure as an asset class for insurance company investment portfolios, with treatment comparable to other long-duration infrastructure debt.
- Pension regulator (PFRDA) similar clarity for pension fund deployment into data centre infrastructure.
- Tax treatment that recognises the long asset life and large capital intensity of AI data centres, particularly around depreciation schedules and minimum alternative tax provisions.
- Sovereign wealth fund vehicle or co-investment platform that allows the National Investment and Infrastructure Fund and the Bharat Investment Trust to participate in data centre projects on competitive terms.

Workforce

The capacity build-out requires a workforce that does not yet exist at scale in India. Specifically: power engineers with experience in high-voltage DC distribution, cooling engineers with experience in direct-to-chip liquid cooling, BMS operators trained on AI workload patterns rather than enterprise data centre patterns, and AI infrastructure engineers (the operations counterpart to AI researchers).

The supply of trained engineers is improving but slowly. State-level vocational training partnerships (the kind HyperNext runs with two engineering colleges and the ITI network) need to scale. The recommendations:

- A national curriculum framework for data centre engineering at the diploma and degree level, developed in partnership between AICTE and the data centre industry, that gets graduates ready for AI-era infrastructure roles rather than legacy IT roles.
- Practical training partnerships between data centre operators and engineering colleges, with operator-funded laboratory equipment in exchange for graduate priority hiring.
- Continuing education programmes for power and cooling engineers from adjacent industries (utilities, refrigeration, HVAC, industrial automation) so the transition path is shorter.

8. Recommendations and headlines

The window for the choices that determine 2030 Indian AI capacity is 2026 to 2028. After 2028 the build-out trajectory is largely locked: land that is not under registered title by then will not be operational by 2030, hardware orders that are not in the vendor pipeline by then will face 2-year-plus additional lead times, and capital that is not committed by then will face the higher cost of catching up with a more crowded market.

For Indian AI infrastructure customers

1. Place capacity commitments now, not in 2028. Operators with confirmed hardware orders and confirmed power allocations are willing to lock in tenancy at terms that will not be available once the supply tightens.
2. Match the procurement decision to the sovereignty requirement, not the lowest gross price. Capacity that nominally meets layer one but fails layer three is the wrong capacity for BFSI, government, and designated CII workloads.
3. Plan workload growth at the high case, not the mid case. AI capability improvements are still compressing per-token cost, which expands the application set. Customers that plan capacity at conservative growth rates routinely revise upward within 18 months.

For Indian capital allocators

1. Treat data centre infrastructure as core long-duration infrastructure, not as IT services or as real estate. The asset class and the regulatory treatment should match the duration profile.
2. Underwrite operators on execution risk rather than announcement volume. The 8.8 GW of announced capacity is materially different from the 6.8 GW that will deliver. The risk markers are public and can be assessed.
3. Co-invest with the larger funds that are entering this segment from sovereign wealth and pension allocations. The deal sizes are growing and the co-investment opportunity is real.

For policy stakeholders

1. Treat the 2026 to 2028 window as a national infrastructure priority. The choices made in this window will determine Indian AI sovereignty for the following decade.
2. Apply the highest-impact interventions first: power allocation timelines, transmission pre-build, sovereignty framework clarity.
3. Co-ordinate across the central and state levels. State-level execution discipline is highly variable and the operators most able to deliver are those in the few states (Telangana, Andhra Pradesh, Maharashtra) where the execution discipline is in place.

HEADLINES

> Indian inference demand by 2030 will be between 7 and 26 gigawatts of equivalent compute capacity, with the central case at approximately 14 GW.

> Risk-weighted Indian AI supply by 2030 is approximately 6.8 GW. The mid-case demand-supply gap is approximately 7 GW.

> Of the 8 GW of sovereign-bound 2030 demand, approximately 6.5 GW is matched by sovereign-compliant supply. The all-three-layers gap is approximately 700 MW.

> The capital requirement to close the supply gap is approximately INR 4.5 lakh crore. Currently identified capital is roughly INR 1.7 lakh crore. The funding gap is large.

> The decisions that determine 2030 Indian AI sovereignty are made in the 2026 to 2028 window. After that, the build-out trajectory is locked in.

> Indian capital, Indian operators, Indian regulatory framework. The country has all three. They need to be deployed at scale faster than they are currently being deployed.

The next paper in the series, HN-RP-008, addresses a different but related question: where does data centre water consumption actually sit relative to other industrial and urban water uses in India. The water debate around data centres has become unmoored from the actual numbers. The data is published, the comparisons are computable, and the resulting picture is not what the dominant narrative suggests. HN-RP-008 walks through that picture.

9. References and sources

- **NITI Aayog.** National Strategy for Artificial Intelligence, 2018, and subsequent annual updates. The reference policy framework for Indian AI compute planning.
- **MeitY.** IndiaAI Mission documentation, 2024 and 2025 progress reports. Government direction on Indian AI capacity.
- **RBI.** Storage of Payment System Data directive, 2018 with subsequent FAQ clarifications. The BFSI sovereignty floor.
- **The Digital Personal Data Protection Act, 2023.** The personal-data sovereignty framework.
- **Anarock Capital.** Indian data centre market report, 2024 and 2025 editions. The supply-side baseline.
- **JLL India.** Data centre market reports, semiannual editions 2023 to 2026. Capacity tracking and announcement compilation.
- **CBRE India.** Investment market research on Indian data centres, 2024 to 2026. Capital flow tracking.
- **Operator public announcements.** AdaniConneX, CtrlS, HyperNext, NTT GDC India, Sify Infinit, ST Telemedia GDC India, Yotta Data Services. Capacity commitments referenced in Section 4 are from public filings, investor communications, and operator announcements published through Q3 2026.
- **HyperNext Research HN-RP-002.** 800VDC Power Architecture. The architectural basis for the per-rack power figures in the capacity modelling.
- **HyperNext Research HN-RP-003.** India\'s Sovereign AI Cloud. The three-layer sovereignty framework used in Section 5.
- **HyperNext Research HN-RP-005.** From Training to Inference. The token-economics framework underlying the demand modelling.
- **HyperNext Research HN-RP-001.** The Nagmati Programme. The water architecture context for the supply build-out.

HyperNext Data Centers

HyperNext Research

We publish engineering and policy papers because the Indian conversation about AI infrastructure needs more substance than marketing material provides. The papers state methodology openly so other operators can run the same analysis on their own facilities. They report findings that may not flatter the HyperNext commercial position. They get review from the engineering team and the communications partners.

Correspondence on methods, figures, and conclusions: **hello@hypernxt.com**. We read every email.

HN-RP-007 · The India AI Compute Gap
13 May 2026 · v1.0

www.hypernxt.com/research
hello@hypernxt.com · +91 99784 23333