

Subsidise, Own or Rent?

Designing the next phase of India's AI compute programme

Published

September 2026

Author

Rahul Rai

Series

AI as Public Infrastructure, No. 3

Key findings

India rented its way to one of the cheapest public AI compute pools in the world, and outbuilt its own target more than fourfold. The next constraint is not GPUs. It is allocation, co-payment design and fixed prices in a rising market. As the government moves to five-year, ring-fenced contracts, those design choices will decide whether the next phase works.

4.5×

Committed GPUs (45,000+) against the original 10,000 target, June 2026

22%

Share of committed GPUs actually in use by end users, March 2026

5%

Share of the ₹10,372 crore five-year outlay spent by March 2026

- **Supply is not the binding constraint.** Of 33,099 GPUs committed by providers in March 2026, 38% had been assigned workloads and 22% were in use. Officials have since put accessible capacity at about 30,000 GPUs against 45,000 committed, a figure that has not been officially confirmed.
- **Most of the low price came from competition, not subsidy.** On the benchmark H100 line, the winning bidder cut ₹310 per hour from its list price; the state's 40% subsidy was worth ₹68.
- **Fixed prices in a rising market create an incentive to divert.** One-year H100 contract prices rose about 38% between October 2025 and March 2026. One provider said of unallocated capacity: “we gave it outside.”
- **The new five-year contracts magnify the pricing risk.** C-DAC's October 2026 empanelment seeks up to 10,000 ring-fenced compute units for five years. Over that horizon, fixed rates will be wrong in one direction or the other: too low and providers default, too high and the state overpays.
- **Recommendation:** keep the rental model; index Stage-2 prices to a basket of published compute benchmarks within a floor and a cap; set allocation deadlines; fix the co-payment; size the sovereign tranche by demonstrated demand; and publish a monthly utilisation dashboard.

Contents

Executive summary	4
SECTION 1 · How the IndiaAI compute model works.....	6
A state that rents rather than builds	6
What the design got right	7
Who actually pays for cheap compute	8
SECTION 2 · Where capacity is lost.....	9
The utilisation funnel.....	9
The spending gap	9
Three causes.....	10
SECTION 3 · The pricing problem	12
A market that turned	12
An unhedged position	12
Compute is becoming a priced commodity	13
SECTION 4 · The policy shift now under way	14
The October 2026 empanelment.....	14
A government-controlled pool.....	14
Letting providers sell abroad	15
SECTION 5 · Four ways to provide public compute.....	16
Matching option to job	16
SECTION 6 · Recommendations for Stage 2.....	17
1. Index prices to a basket of benchmarks, within a floor and a cap.....	17
2. Set allocation deadlines and use-it-or-lose-it rules.....	18
3. Redesign the co-payment.....	18
4. Size the sovereign tranche by demonstrated demand	18
5. Price the export option.....	18
6. Publish a monthly utilisation dashboard.....	18
SECTION 7 · Objections and responses.....	20
SECTION 8 · Conclusion.....	21
About this report	21
Sources	22

Executive summary

In March 2024 the Union Cabinet approved the IndiaAI Mission with a five-year outlay of ₹10,371.92 crore and a target of at least 10,000 GPUs. Rather than buy hardware, the Mission empanelled private cloud providers, ran competitive price discovery and subsidised usage. By June 2026 providers had committed more than 45,000 GPUs at a headline rate of about ₹65 per GPU-hour.

That design was right. It avoided the capital cost and rapid obsolescence of state-owned hardware, it paid only for use, and it produced prices well below global benchmarks. It also built capacity faster than any comparable public programme. **The difficulty is that capacity has outrun the system that turns it into work.** Government data reported in March 2026 showed that only about a fifth of committed GPUs were being used. Spending tells the same story: by March 2026 the Mission had spent about ₹519 crore, roughly 5% of its outlay.

Three frictions explain most of the gap: slow allocation, a 60% co-payment that is hard for startups to fund and slow for government users to approve, and fixed contract rates in a rising global market, which give providers a reason to sell capacity elsewhere.

The government is now changing the model. In October 2026 C-DAC opened an empanelment for up to 10,000 dedicated, ring-fenced compute units on five-year terms, with a further tranche of government-controlled capacity under discussion. **This report argues that the shift towards long-term contracts is sound, but that its success depends on design choices still open in Stage 2,** especially how prices adjust over five years, how quickly capacity is allocated, and how much the state should own outright.

Recommendations at a glance

1. **Index Stage-2 prices.** Pay providers a rate linked to a basket of published compute price indices, less an agreed discount, within a floor and a cap, reset quarterly.
2. **Set allocation deadlines.** Decide on requests within a fixed number of working days; return unused allocations to the pool after a set window; publish median allocation time.
3. **Fix the co-payment.** Allow milestone-linked or instalment payment for startups; give government users annual pre-approved budgets instead of per-request purchase orders.
4. **Size the sovereign tranche by demand.** Reserve owned or ring-fenced capacity for workloads where control matters, and size it by the queue of such workloads, not by headline targets.
5. **Price the export option.** If providers may sell committed capacity abroad, require a recall right and a revenue share on overseas sales of IndiaAI-committed capacity.
6. **Publish a monthly dashboard** of committed, delivered, allocated and used GPU-hours by provider, GPU type and user category.

Why this matters now

Stage-1 applications under the C-DAC empanelment close on 29 October 2026, and the Stage-2 technical and commercial terms have not yet been issued. Pricing, allocation and export terms can still be set at low cost. Once five-year contracts are signed, changing them will be slow and expensive.

The rest of this report sets out how the current model works (Section 1), where capacity is lost (Section 2), why fixed pricing has become a problem (Section 3), what the government is now changing (Section 4), the options available (Section 5), and specific recommendations for Stage 2 (Section 6). Section 7 addresses the strongest objections.

How the IndiaAI compute model works

A state that rents rather than builds

The IndiaAI Mission is organised in seven pillars. Compute is by far the largest, at ₹4,563.36 crore or 44% of the outlay. Foundation models (₹1,971.37 crore) and startup financing (₹1,942.5 crore) follow. Datasets receive ₹199.55 crore and Safe and Trusted AI ₹20.46 crore.¹

Pillar	Allocation (₹ crore)	Share
Compute capacity	4,563.36	44.0%
Foundation models (Innovation Centre)	1,971.37	19.0%
Startup financing	1,942.50	18.7%
FutureSkills	882.94	8.5%
Application development	689.05	6.6%
Datasets platform (AIKosh)	199.55	1.9%
Safe and Trusted AI	20.46	0.2%
Overheads and contingency	102.69	1.0%
Total	10,371.92	100%

Table 1. IndiaAI Mission allocation by pillar. Source: PIB; MeitY reply in the Lok Sabha, 4 December 2024, as tabulated by Vayam.

The compute pillar works through empanelment. Private providers own and operate GPUs in their own data centres. They bid hourly rates for each GPU type through a request for empanelment, and the lowest bid (L1) sets the reference price that other empanelled providers may match. Eligible users, including startups, academic researchers, government bodies and foundation-model teams, apply through the IndiaAI Compute Portal, which publishes a rate card by GPU type and commitment length.² The Mission then pays up to 40% of the bill; the user pays the remainder. Payment is for hours used, not capacity reserved.

1. Press Information Bureau, IndiaAI Mission component-wise allocation, 2026. <https://www.pib.gov.in/PressReleaseDetail.aspx?PRID=2245233®=3&lang=1>
2. IndiaAI Compute Portal, price list. <https://compute.indiaai.gov.in/pricelist>

What the design got right

Speed. The pool grew from 18,693 GPUs after the first round in January 2025³ to 34,333 by May 2025, 38,231 from 14 providers by March 2026⁴ and more than 45,000 by June 2026.⁵ No state-owned build could have moved this quickly; power, cooling, procurement and staffing alone would have taken longer.

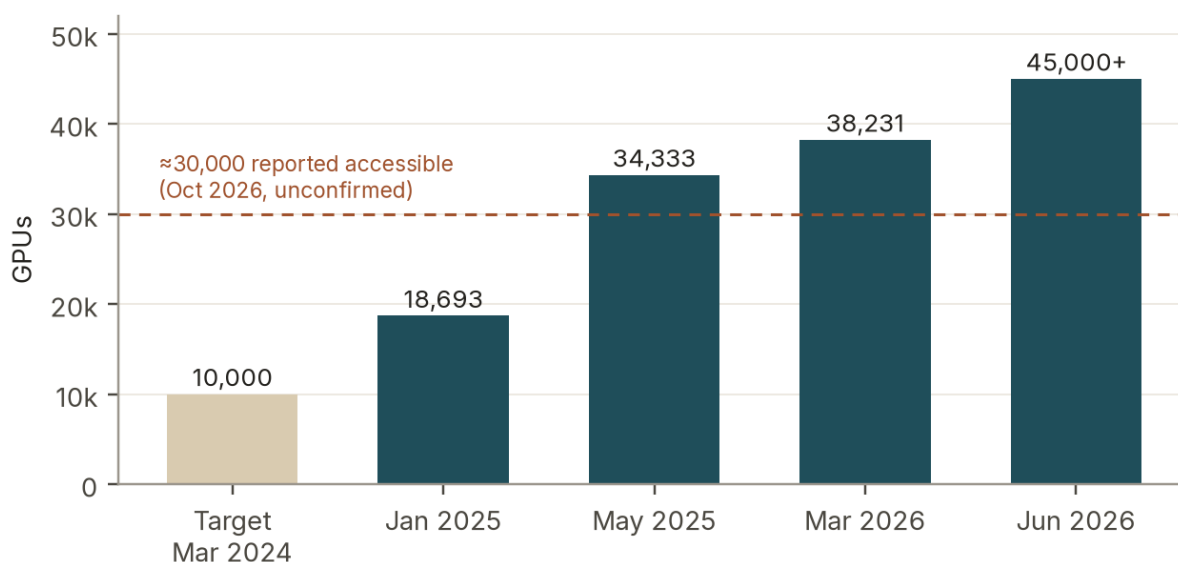


Figure 1: Committed GPUs in the IndiaAI pool against the original target

Bars show provider commitments as reported by PIB. The dashed line marks the approximately 30,000 GPUs an unnamed official told BusinessLine were accessible in October 2026; this figure has not been officially confirmed.

Price. First-round bids averaged ₹115.85 per GPU-hour, against global rates then commonly quoted at \$2.5–3.⁶ After subsidy, the headline user price has been about ₹65 per hour.⁷ Officials and providers have described the effective cost as below \$1 per GPU-hour, among the lowest anywhere.⁸

Fiscal discipline. Because the state pays only for hours used, idle capacity costs the treasury almost nothing. The risk of under-use sits with providers, not taxpayers. This is a genuine advantage over owning hardware, where an idle GPU depreciates whether or not anyone runs a job on it.

3. Press Information Bureau, "India all set to launch its own AI model at affordable cost," 30 January 2025. <https://www.pib.gov.in/PressReleaseDetailm.aspx?PRID=2097709>

4. Press Information Bureau, "Government supporting organisations and consortia to develop sovereign foundational model," March 2026. <https://www.pib.gov.in/PressReleaseDetail.aspx?PRID=2245063®=1&lang=1>

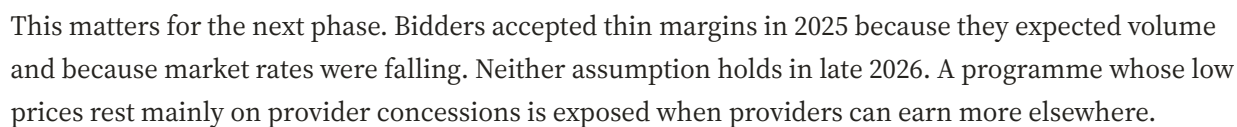
5. Press Information Bureau, "Powering the Future" factsheet, August 2026. <https://www.pib.gov.in/FactsheetDetails.aspx?Id=150860®=20&lang=8>

6. Press Information Bureau, 30 January 2025. <https://www.pib.gov.in/PressReleaseDetailm.aspx?PRID=2097709>

7. Press Information Bureau, IndiaAI compute press note. <https://www.pib.gov.in/PressNoteDetails.aspx?NotelD=156786&ModuleId=3®=3&lang=2>

8. Economic Times, "IndiaAI GPU tender: lowest bidders receive AI workloads," 2025. <https://economictimes.indiatimes.com/tech/artificial-intelligence/indiaai-gpu-tender-lowest-bidders-receive-ai-workloads/articleshow/120627587.cms>

A closer look at the bid tables changes how the low price should be understood. On the H100 SXM on-demand line in the first round, the listed market price was ₹480 per hour and the winning bid was ₹170, a discount of 64.6%. At a 40% subsidy, the user pays ₹102 and the state ₹68. The provider's concession from its own list price, ₹310, is about 4.6 times the public subsidy.⁹



8

SECTION 2

Where capacity is lost

The utilisation funnel

The clearest public evidence comes from a government communication to providers, reported in March 2026. It showed that of 33,099 GPUs committed, 12,638 had been assigned workloads and 7,418 were in use by end users.¹⁰ The pattern held for the largest provider: of 8,192 GPUs Yotta had committed, 4,864 were allocated and 2,024 used.

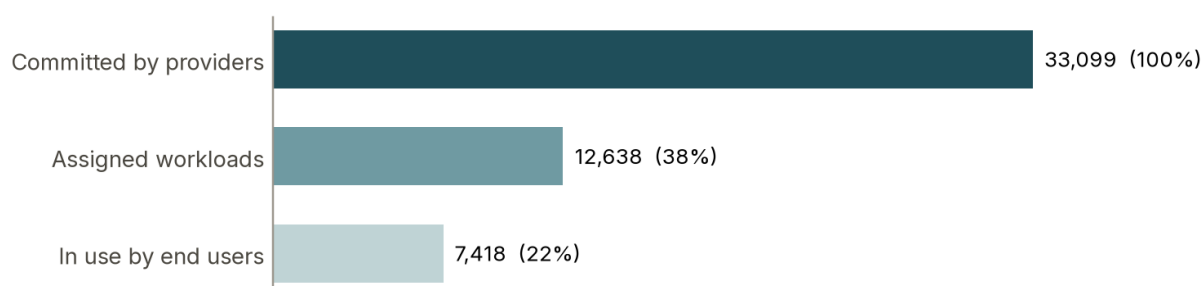


Figure 3: From commitment to use, March 2026

Fewer than one in four committed GPUs were running user workloads. Source: Economic Times, reporting an IndiaAI communication to providers.

A second measure points the same way. By August 2026 the Mission reported 93.18 lakh GPU-hours across 237 projects. Spread across a pool of 45,000 GPUs, that is about 207 hours per GPU, or eight and a half days of work, in a year that has 8,760 hours.¹¹ The figure is cumulative and the pool grew over time, so this understates utilisation of the GPUs available at any moment. It is still a small number. Parliamentary answers have described these hours as “sanctioned” and PIB as “accessed”; neither is the same as hours actually used.

The spending gap

Spending confirms that money, not just hardware, is moving slowly. Budget allocations for the first three years totalled ₹1,973 crore, about 19% of the five-year outlay. Actual expenditure was ₹21.79 crore in 2024–25 and ₹497.08 crore in 2025–26 as of 13 March 2026.¹²

10. Economic Times, March 2026. <https://economictimes.indiatimes.com/tech/artificial-intelligence/underused-gpus-raise-questions-about-indiaai-capacity-build-out/articleshow/128981840.cms>

11. Vayam, citing MeitY reply to Lok Sabha Unstarred Question 4079, 12 August 2026. <https://vayam.press/stories/indiaai-the-state-as-compute-wholesaler>

12. Lok Sabha Unstarred Question 4311, March 2026. https://www.sansad.in/getFile/loksabhaquestions/annex/187/AU4311_SVy5u3.pdf?source=pqals

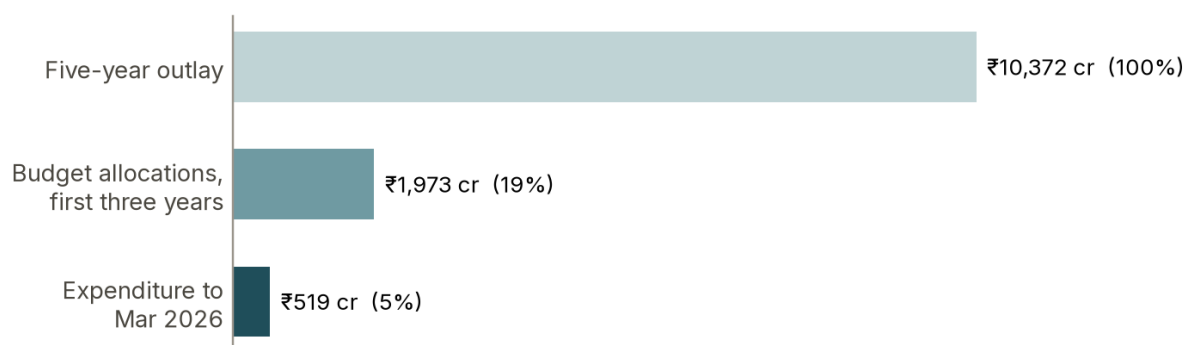


Figure 4: IndiaAI Mission outlay, allocations and spending

Expenditure to March 2026 combines 2024–25 and 2025–26 to 13 March 2026. Sources: Lok Sabha Unstarred Questions 4311 and 4079.

Three causes

Slow allocation

Providers describe IndiaAI's allocation process as slow. A committed GPU earns nothing until it is assigned, so every week of delay is a week in which a provider carries an idle asset. In a rising market, that idle asset has an outside buyer. One provider executive explained the result directly: “As they are not allocating and we are getting better prices outside, we gave it outside.”¹³

The 60% co-payment

The co-payment exists for good reasons: it screens out frivolous requests and keeps users accountable. But it creates two kinds of friction. Early-stage startups often cannot fund 60% of a large training run in cash before revenue arrives. Government users can fund it but must raise a purchase order for each request. One provider described waiting on a government customer: “since it is a government customer, they are taking their own time to process it.”

Uneven drawdown

Users frequently receive an allocation, use part of it and plan to use the rest later. This is rational behaviour for a team that expects demand to grow, but it locks up capacity that another user could run today. Without use-it-or-lose-it rules, the pool fills with reserved but idle hours.

13. Economic Times, March 2026. <https://economictimes.indiatimes.com/tech/artificial-intelligence/underused-gpus-raise-questions-about-indiaai-capacity-build-out/articleshow/128981840.cms>

What the data cannot yet tell us

The Mission does not publish hours actually used by GPU type or provider, a split of GPU-hours by user category, how many committed GPUs sat idle while users waited, or whether foundation-model grants draw on the same pool. In February 2026, a Rajya Sabha question asked directly for utilisation levels; the answer listed user counts but no utilisation figure.¹⁴ Several conclusions in this report would be stronger, or might change, with that data. Recommendation 6 is designed to produce it.

14. Vayam, citing Rajya Sabha Unstarred Question 1668, 13 February 2026. <https://vayam.press/stories/indiaai-the-state-as-compute-wholesaler>

The pricing problem

A market that turned

H100 rental prices fell steeply through much of 2025 as supply caught up with demand. They then reversed. SemiAnalysis's one-year H100 contract index rose from \$1.70 to \$2.35 per GPU-hour between October 2025 and March 2026, an increase of about 38%.¹⁵ The Ornn Compute Price Index, built from executed trades, settled at \$3.15 per H100-hour on 22 October 2026, up 10.5% over the month.¹⁶ Newer Blackwell-generation pricing has also risen in 2026.¹⁷

Input costs are rising too. Providers report sharply higher memory prices and memory suppliers booked well into 2028. Moneycontrol reports that providers' costs have risen sharply in six months and that providers are now looking to overseas customers.¹⁸ An official told BusinessLine that some empanelled providers had been unable to deliver committed capacity as equipment costs rose.¹⁹

An unhedged position

A fixed contract rate in a market that moves is a financial position, whether or not anyone calls it one. When market prices rise above the contract rate, the provider is effectively short: every hour sold to IndiaAI is an hour not sold at the higher market price. Rational providers respond by delivering slowly, prioritising other customers, or failing to bring promised capacity online. None of this requires bad faith. It is the predictable result of combining a fixed price, a volatile underlying market and slow allocation.

Under the current pay-per-use model, the cost of this position falls mainly on users, who find capacity unavailable. Under a five-year ring-fenced contract, the same dynamic becomes larger and harder to unwind. If prices rise, providers face a long period of selling below market and may default or under-deliver. If prices fall, as they did through 2025, the state locks in rates above market for years. **Over five years, a fixed price is almost certain to be wrong in one direction.**

15. GPUsmith, "H100 rental price history and trends," citing SemiAnalysis, 2026. <https://gpsmith.com/articles/en/pdfs/h100-rental-price-history-trends.pdf>

16. Ornn Compute Price Index, accessed 24 October 2026. <https://data.ornn.com/>

17. Communications Today, "IndiaAI Mission faces gap between capacity and deployed GPUs," 2026. <https://www.communicationstoday.co.in/indiaai-mission-faces-gap-between-capacity-and-deployed-gpus/>

18. Moneycontrol, "India's AI compute paradox: GPU providers look abroad as researchers, startups struggle at home," October 2026. <https://www.moneycontrol.com/artificial-intelligence/india-s-ai-compute-paradox-gpu-providers-look-abroad-as-researchers-startups-struggle-at-home-article-14054291.html>

19. BusinessLine, 2 October 2026. <https://www.thehindubusinessline.com/info-tech/centre-fortifies-indias-ai-mission-with-sovereign-compute-pool-amid-rising-gpu-costs/article71537526.ece>

Compute is becoming a priced commodity

Two developments make an alternative practical. First, independent compute price indices now exist and are published regularly. Ornn's H100 index carries its own Bloomberg ticker. Others are built from advertised rather than executed prices. They differ in method and can diverge, sometimes significantly,²⁰ which argues for using a basket rather than a single benchmark. Second, the government itself is moving to longer contracts, which is exactly where indexation matters most.

20. TokenPost, on divergence between GPU rental price indices, 2026. <https://www.tokenpost.com/news/technology/33728>

The policy shift now under way

The October 2026 empanelment

On 7 October 2026, C-DAC, as the Mission's implementing agency, issued a request for empanelment of AI service providers (Ref. C-DAC/GPU/01/2026), closing on 29 October.²¹ Its main features differ from the earlier rounds:

- **Scale and term:** up to 10,000 AI compute units, a category that includes GPUs, TPUs and other accelerators, supplied as end-to-end services for five years, with the option to scale beyond 10,000.
- **Ring-fencing:** capacity committed to C-DAC must be specifically identified, reserved and made available for a Stage-2 procurement.
- **Configuration:** clusters of 4,000, 1,000 or 512 units, each with at least 512 units at one data centre in India.
- **Sovereignty:** all infrastructure, data, logs, backups and disaster recovery must remain in India, with physical and logical isolation from other customers.²²
- **Stage 2:** exact technical requirements, cluster sizes and, by implication, pricing terms will be specified later.

Reports indicate that about 3,000 units would serve government use and about 7,000 would be available to other eligible users.²³ The Minister for Electronics and IT said the arrangement is meant to give industry certainty of demand: "Our demand estimate is for 10,000 GPUs, and the first tender that we are coming out with is for 5,000."

A government-controlled pool

Alongside the empanelment, officials have described a government-controlled pool. It would include capacity through C-DAC and an 8-exaflop system built with G42 and Cerebras, reportedly a gift from the UAE and roughly equivalent to 10,000 GPUs. Pricing for this pool had not been decided as of early October.²⁴ MeitY has also issued a purchase order for a roughly 1.1-exaflop AI system at the NIC data centre in Delhi. No contract for the G42 system has been published, and analysts have noted a lack of clarity on the chip mix.

21. C-DAC, Request for Empanelment of AI Service Providers for Provisioning of AI Compute Capacity, 7 October 2026. https://www.cdac.in/index.aspx?id=tenders_ASP-RFE

22. MediaNama, "MeitY invites AI compute providers to supply 10,000 GPUs under IndiaAI Mission," 13 October 2026. <https://www.medianama.com/2026/10/223-meity-ai-compute-providers-gpus-indiaai-mission/>

23. ET Datacenters, "India to issue AI regulation paper, expands GPU capacity," October 2026. <https://datacenters.economictimes.indiatimes.com/news/ai-compute-infrastructure/india-to-issue-ai-regulation-paper-expands-gpu-capacity/134853002>

24. BusinessLine, 2 October 2026. <https://www.thehindubusinessline.com/info-tech/centre-fortifies-indias-ai-mission-with-sovereign-compute-pool-amid-rising-gpu-costs/article71537526.ece>

Letting providers sell abroad

The government is also reported to be considering allowing successful bidders to offer capacity to overseas customers at commercial rates.²⁵ This would improve provider economics and could keep Indian data centres full. It also formalises the diversion that has already been happening informally. Without clear priority rules, the export option and ring-fencing pull in opposite directions.

Assessment

The direction is right. Long-term contracts give providers the revenue visibility they need to finance hardware in a tight market, and a small government-controlled pool addresses workloads where control matters. But a five-year contract written with a fixed price, slow allocation and an unpriced export option would reproduce the current problems at a larger scale and over a longer period. Stage 2 is where this is decided.

25. Business Standard, "Successful GPU bidders likely to offer compute capacity to global firms," October 2026. https://www.business-standard.com/amp/technology/tech-news/successful-gpu-bidders-likely-to-offer-compute-capacity-to-global-firms-126101100603_1.html

Four ways to provide public compute

Public compute can be supplied in four basic ways. They are not mutually exclusive, and the right programme combines them.

Option	How it works	Cost profile	Speed	Control	Main risk
A. Subsidise use (current)	State pays a share of usage on empanelled private clouds	Pay per use	Fast	Low–medium	Low utilisation; diversion when market rates rise
B. Own capacity	State buys and operates its own clusters	High upfront; fast depreciation	Slow	High	Idle or obsolete hardware; thin operating capacity
C. Fund the user	Vouchers or credits spendable with any qualified provider	Pay per use	Fast	Low	Leakage to low-priority uses; little leverage over providers
D. Contract forward, indexed	Multi-year ring-fenced capacity at prices tied to published benchmarks within bounds	Predictable, bounded	Medium	Medium–high	Index quality; contract complexity

Table 2. Options for supplying public AI compute.

Matching option to job

Broad startup and research access is best served by option A, improved with the allocation and co-payment fixes below. It keeps the programme demand-led and fiscally light.

Large, predictable workloads, such as foundation-model training, government inference at scale and shared research clusters, suit option D. These users need guaranteed capacity for long periods, and providers need revenue certainty to finance hardware. This is the logic of the C-DAC empanelment.

Sensitive workloads involving defence, critical infrastructure or classified government data justify option B, but only at the scale of demonstrated demand. Ownership does not guarantee sovereignty when chips, networking and cooling are imported, and hardware loses value quickly as new generations arrive.

Cash-constrained startups benefit from option C as a way to finance the co-payment, for example through milestone-linked credits.

Recommendations for Stage 2

1. Index prices to a basket of benchmarks, within a floor and a cap

Stage-2 contracts should pay providers a rate that adjusts with the market rather than a single fixed rate for five years. A simple structure would work as follows:

- **Reference price:** the average of two or three published compute price indices for the relevant GPU class, at least one built from executed transactions.
- **Contract rate:** the reference price minus an agreed discount, set competitively in the bid. Bidders compete on the discount, not on an absolute price.
- **Floor and cap:** the rate cannot fall below a floor, protecting providers' financing, or rise above a cap, protecting the budget.
- **Reset:** quarterly, with a published fallback rule if an index is discontinued or diverges sharply from the others.
- **Technology step-down:** rates for a given GPU generation step down on a published schedule as newer generations arrive, unless the provider upgrades.

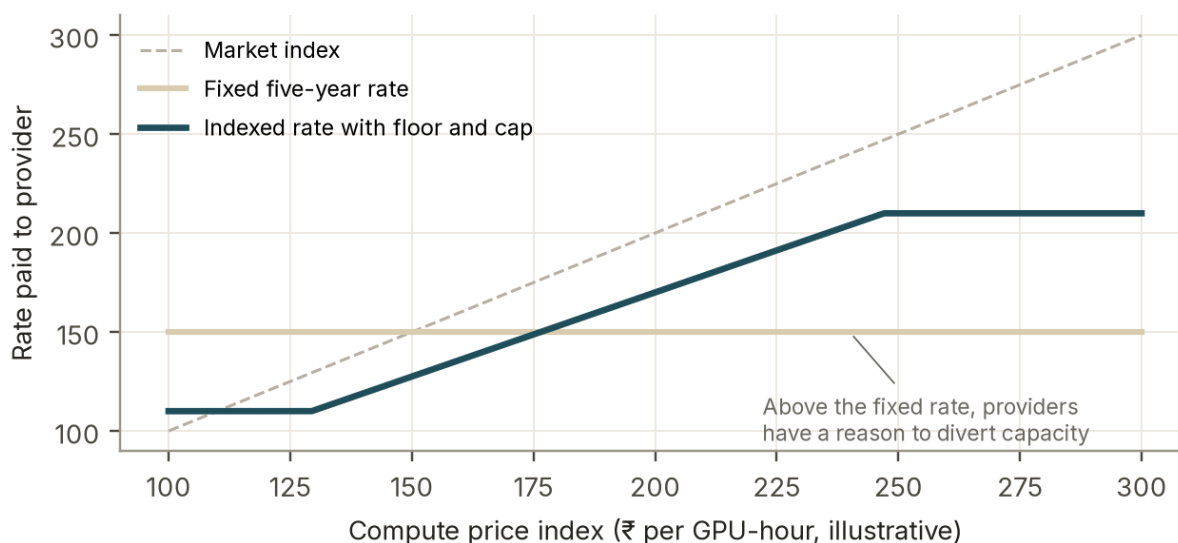


Figure 5: Fixed versus indexed contract rates (illustrative)

A fixed rate (sand) is below market when the index rises, creating an incentive to divert, and above market when it falls. An indexed rate with a floor and cap (teal) tracks the market within a bounded range. Values are illustrative.

This removes most of the incentive to divert capacity when markets rise, because the contract rate rises with them up to the cap. It stops the state from locking in above-market rates if prices fall. And it makes the state's maximum exposure known in advance, which is what finance departments and auditors need. The state does not need to trade derivatives to do this. It only needs contracts that reference a credible benchmark, as fuel and electricity contracts routinely do.

2. Set allocation deadlines and use-it-or-lose-it rules

The Mission should commit to deciding on compute requests within a fixed number of working days and publish its median decision time. Allocations not drawn down within an agreed window, for example 30 days for on-demand and 60 days for reserved capacity, should return to the pool unless the user shows cause. For ring-fenced capacity under the new empanelment, unallocated capacity should be released to the provider on a short notice period at a pre-agreed rate, rather than sitting idle.

3. Redesign the co-payment

The co-payment should stay, because it screens demand. Its form should change:

- **Startups:** allow the 60% share to be paid in instalments over the life of a project, or deferred against milestones, with the provider paid in full by the Mission and the user's share recovered over time.
- **Government users:** issue annual pre-approved compute budgets to ministries and state departments, so individual requests do not each need a new purchase order.
- **Academic users:** consider a higher subsidy share for public universities and research institutions with grant-funded projects, where the co-payment mainly shifts money between public budgets.

4. Size the sovereign tranche by demonstrated demand

Government-owned or exclusively controlled capacity should be reserved for workloads where control is the real requirement. Its size should follow the queue of such workloads, not headline targets. The government should publish the criteria for access to the government-controlled pool, the pricing approach for the G42 system and the NIC cluster, and how these will interact with the empanelled pool. Without clear pricing, a free or near-free government pool would draw demand away from the empanelled providers and weaken the programme the Mission has built.

5. Price the export option

If providers are allowed to sell capacity overseas, the terms should be explicit. IndiaAI-committed capacity should carry a recall right on short notice. Overseas sales of capacity that was committed to the Mission, as opposed to providers' other capacity, could carry a revenue share or a lower Mission rate. Domestic eligible users should keep priority whenever the pool is below a published utilisation threshold.

6. Publish a monthly utilisation dashboard

The March 2026 utilisation figures became public only through press reporting. A monthly dashboard should show, by provider, GPU type and user category: GPUs committed, delivered, allocated and used; GPU-hours sanctioned and actually consumed; median allocation time; and spending against budget. Routine transparency would let the programme correct itself faster, give Parliament the figures it has asked for, and settle the question of whether low utilisation reflects weak demand or friction.

Recommendation	Owner	Instrument	Timing
Indexed pricing with floor and cap	C-DAC, IndiaAI	Stage-2 bid terms	Before Stage-2 issue
Allocation deadlines	IndiaAI	Portal rules, SLA	Immediate
Co-payment redesign	IndiaAI, MeitY	Scheme guidelines	Next guideline revision
Sovereign tranche sizing and pricing	MeitY	Published access policy	Before pool goes live
Export option terms	MeitY, C-DAC	Empanelment conditions	Before any approval
Utilisation dashboard	IndiaAI	Public portal	Within one quarter

Table 3. Implementation summary.

Objections and responses

"India needs to own its compute for sovereignty."

For some workloads, yes, which is why recommendation 4 exists. But owning hardware does not deliver sovereignty when the chips, networking and cooling are imported and the software stack is foreign. Owned hardware also loses value quickly as new GPU generations arrive. For most uses, assured access on fair terms, with data localisation and ring-fencing already written into the new empanelment, gives more strategic value per rupee than ownership.

"Index-linked contracts expose the state to price spikes."

The cap bounds that exposure and makes it known in advance. Today, the cost of a price spike already falls on someone: on users, who lose access when capacity is diverted, and on the programme, when providers fail to deliver. Indexation with a cap makes that cost visible and limited rather than hidden and open-ended.

"Compute indices are new and thin."

They are. That is why the reference should be a basket of indices, with at least one built from executed trades, a published fallback rule and an annual review of the benchmark. The alternative, a fixed five-year price set by a single bid in a volatile market, carries more risk than an imperfect index.

"Low utilisation just means demand is weak."

Perhaps partly. But providers report selling diverted capacity at higher prices elsewhere, and users report delays in allocation and payment. Both suggest demand is being held back by friction rather than absent. The utilisation dashboard would settle the question within a few months.

"Indexation is too complex for government procurement."

Indian public bodies already use price-variation clauses in construction, fuel and power contracts. A compute index clause is no more complex. The harder task is building in-house understanding of compute markets at C-DAC and IndiaAI, which the programme needs regardless.

SECTION 8

Conclusion

India made a sensible bet in renting rather than owning, and it outbuilt its own target. The next phase is less about adding GPUs than about the systems that connect capacity to users and keep prices fair as markets move: faster allocation, simpler co-payments, transparent data and contracts that adjust with the market.

The move to five-year ring-fenced contracts is the right response to a tightening market. It is also the moment when design errors become most expensive, because they will be locked in for five years. Writing indexation, allocation deadlines and clear export terms into Stage 2 now would raise effective capacity faster, and more cheaply, than another round of procurement.

About this report

India 2035 Research is an independent research initiative on AI as public infrastructure in India. This report draws on official releases, parliamentary answers, procurement documents and press reporting available as of 24 October 2026. Where figures come from unnamed officials or provider executives, the text says so. Corrections and comments are welcome at rahul@india2035.ai.

Sources

1. Press Information Bureau, IndiaAI Mission component-wise allocation, 2026. <https://www.pib.gov.in/PressReleaseDetail.aspx?PRID=2245233®=3&lang=1>
2. IndiaAI Compute Portal, price list. <https://compute.indiaai.gov.in/pricelist>
3. Press Information Bureau, "India all set to launch its own AI model at affordable cost," 30 January 2025. <https://www.pib.gov.in/PressReleaseDetail.aspx?PRID=2097709>
4. Press Information Bureau, "Government supporting organisations and consortia to develop sovereign foundational model," March 2026. <https://www.pib.gov.in/PressReleaseDetail.aspx?PRID=2245063®=1&lang=1>
5. Press Information Bureau, "Powering the Future" factsheet, August 2026. <https://www.pib.gov.in/FactsheetDetails.aspx?Id=150860®=20&lang=8>
6. Press Information Bureau, IndiaAI compute press note. <https://www.pib.gov.in/PressNoteDetails.aspx?NotelId=156786&ModuleId=3®=3&lang=2>
7. Economic Times, "IndiaAI GPU tender: lowest bidders receive AI workloads," 2025. <https://economictimes.indiatimes.com/tech/artificial-intelligence/indiaai-gpu-tender-lowest-bidders-receive-ai-workloads/articleshow/120627587.cms>
8. Vayam, "IndiaAI's GPU pool: who really pays for ₹65 compute," 23 October 2026. <https://vayam.press/stories/indiaai-the-state-as-compute-wholesaler>
9. Economic Times, March 2026. <https://economictimes.indiatimes.com/tech/artificial-intelligence/underused-gpus-raise-questions-about-indiaai-capacity-build-out/articleshow/128981840.cms>
10. Lok Sabha Unstarred Question 4311, March 2026. https://www.sansad.in/getFile/loksabhaquestions/annex/187/AU4311_SVy5u3.pdf?source=pqals
11. GPUsmith, "H100 rental price history and trends," citing SemiAnalysis, 2026. <https://gpusmith.com/articles/en/pdfs/h100-rental-price-history-trends.pdf>
12. Ornn Compute Price Index, accessed 24 October 2026. <https://data.ornn.com/>
13. Communications Today, "IndiaAI Mission faces gap between capacity and deployed GPUs," 2026. <https://www.communicationstoday.co.in/indiaai-mission-faces-gap-between-capacity-and-deployed-gpus/>
14. Moneycontrol, "India's AI compute paradox: GPU providers look abroad as researchers, startups struggle at home," October 2026. <https://www.moneycontrol.com/artificial-intelligence/india-s-ai-compute-paradox-gpu-providers-look-abroad-as-researchers-startups-struggle-at-home-article-14054291.html>
15. BusinessLine, 2 October 2026. <https://www.thehindubusinessline.com/info-tech/centre-fortifies-indias-ai-mission-with-sovereign-compute-pool-amid-rising-gpu-costs/article71537526.ece>
16. TokenPost, on divergence between GPU rental price indices, 2026. <https://www.tokenpost.com/news/technology/33728>
17. C-DAC, Request for Empanelment of AI Service Providers for Provisioning of AI Compute Capacity, 7 October 2026. https://www.cdac.in/index.aspx?id=tenders_ASP-RFE
18. MediaNama, "MeitY invites AI compute providers to supply 10,000 GPUs under IndiaAI Mission," 13 October 2026. <https://www.medianama.com/2026/10/223-meity-ai-compute-providers-gpus-indiaai-mission/>
19. ET Datacenters, "India to issue AI regulation paper, expands GPU capacity," October 2026. <https://datacenters.economictimes.indiatimes.com/news/ai-compute-infrastructure/india-to-issue-ai-regulation-paper-expands-gpu-capacity/134853002>
20. Business Standard, "Successful GPU bidders likely to offer compute capacity to global firms," October 2026. https://www.business-standard.com/amp/technology/tech-news/successful-gpu-bidders-likely-to-offer-compute-capacity-to-global-firms-126101100603_1.html