As AI Inference Demand Rises, MeitY Plans National Compute Expansion

News reference: 27 November 2026

Published: 28 November 2026

Mains: GS-III, AI infrastructure and science and technology policy



Original symbolic illustration: pcsmantra.com

▼ **In this article**

1. Why in the news
2. Key facts of the plan
3. Why inference capacity matters
4. The IndiaAI Mission: how far along
5. Exam relevance
6. Way forward
7. Conclusion
8. Mains practice question
9. Frequently asked questions (FAQ)
10. Sources

Why in the news

Building an AI model involves two distinct computing needs: training and inference. The compute support extended so far to companies selected under the IndiaAI Mission was primarily for training. According to the MeitY Secretary, inference requirements are far broader. The government also needs inference capacity for its internal work. The ministry is therefore examining an expansion of national compute capacity. The statement comes as AI models built in the country begin commercial use.

Basics

- **AI training:** the computation of teaching a model on large data. Very heavy, largely a one-time process.
- **AI inference:** the computation of drawing answers from a trained model. It happens on every use, again and again, millions of times; the more a model is used, the more inference is needed.
- **GPU/AI accelerator:** processors built for AI computation.
- **Petaflop:** the ability to perform a quadrillion (10^15) calculations a second. The scale of a supercomputer.
- **Empanelment:** the government's listing of qualified providers from whom compute can be bought at subsidised rates.

Key facts of the plan

The main points come from the MeitY Secretary's public remarks. Figures from the oral statement are attributed to him; no independent official document confirming those figures was found.

Point	Detail
Government's own capacity (per the Secretary)	A facility of about 3,300 GPUs is coming up at C-DAC Pune; a high-performance computing and AI facility in Bengaluru, which the Secretary described as a "30 petaflop machine"
G42 arrangement	G42 and the Government of India formalised the commercial framework for the "Condor Galaxy India" AI supercomputer on 15 May 2026: 8 exaflops, 64 Cerebras CS-3 systems, with C-DAC as partner. According to the Secretary, this arrangement will be used primarily for inference
C-DAC empanelment tender	C-DAC has issued an RFE for the empanelment of AI service providers for provisioning of AI compute services (C-DAC/GPU/01/2026); bid closing date 30 November 2026
Order of priority	First the government's needs, then research and academic institutions; surplus capacity to innovators for inference
Status of subsidies	Existing support is for training; domestic capacity will be assessed first, then additional assistance considered; direct or co-investment in companies also possible
New model selection	None for now; the priority is helping first-round companies make their models commercially available

Why inference capacity matters

Training is a one-time cost; inference is an everyday one. When a government service, say language translation, farm advice or grievance redress, runs on an AI model, computation happens for every user's every question. With users in the crores, total inference demand outgrows training many times over. In the Secretary's words, "broadly there needs to be an expansion of the overall compute capacity which comes up in the country, so that more inferencing is created." This is why policy is moving from 'build the model' to 'run the model'.

The IndiaAI Mission: how far along

- **The mission:** approved by the Cabinet in March 2024 with an outlay of ₹10,371.92 crore (about ₹10,372 crore). One of its seven pillars is IndiaAI compute capacity.
- **Empanelled GPUs:** as per the official list on the IndiaAI compute portal, a total of 44,077 GPUs of providers across the country are empanelled (the largest share is Yotta's, 17,408), from which selected teams draw compute at subsidised rates.

PCSMANTRA.COM

Exam relevance

Mains angle: 'compute sovereignty' for AI, meaning public capacity versus private investment; the policy understanding of training versus inference; links with data-centre policy and energy needs.

Facts to remember: the IndiaAI Mission (March 2024, ₹10,371.92 crore); empanelled GPUs (44,077); G42's Condor Galaxy India arrangement (15 May 2026, 8 exaflops); per the Secretary, about 3,300 GPUs at C-DAC Pune and a 30-petaflop facility in Bengaluru.

Way forward

- **Worth tracking:** the shape of the inference-capacity plan (how much subsidy, for whom, how), progress on C-DAC's empanelment tender, and the Bengaluru facility's timeline.
- **Upcoming guidelines:** the Secretary said MeitY's research-project guidelines (industry-oriented, application-driven) will be issued shortly.
- **Assessment:** once the assessment of domestic capacity is complete, additional assistance or investment options may emerge.

Conclusion

The one-line summary of the policy shift: India's next stage in AI is not building models but running them. For the exam, remember what inference is, why it is rising, and which capacities (C-DAC Pune, Bengaluru, the G42 arrangement) the government is adding.

Mains practice question

Question 1. "The next frontier of self-reliance in AI is not model-building but inference capacity." Examine in the light of India's national compute-expansion plan. (250 words, 15 marks)

▼ **Answer outline**

- **Introduction:** the MeitY Secretary's statement of 27 November 2026; the one-line training-versus-inference distinction.
- **Body, supporting the statement:** at the service layer inference is the real consumption; government services (language, agriculture, citizen services) need domestic, secure compute; steps such as, per the Secretary, C-DAC Pune (about 3,300 GPUs) and Bengaluru (30 PF), the G42 Condor Galaxy India arrangement, and 44,077 empanelled GPUs.
- **Balance:** model-building must continue too (the IndiaAI-selected teams); the cost of public capacity, its energy demand and utilisation need monitoring; private investment and co-investment options.
- **Challenges:** chip supply, power and water for data centres, skilled talent.
- **Conclusion:** end on defining AI self-reliance as the dual capacity to 'build and run'.

Frequently asked questions (FAQ)

Short answer: MeitY is planning to expand national compute capacity for the rising demand for AI inference. According to the Secretary, a 3,300-GPU facility at C-DAC Pune and a 30-petaflop facility in Bengaluru are coming up, with priority to the government, then research institutions and innovators.

PCSMANTRA.COM

Q1. What is AI inference?

The computation of drawing an answer or prediction from a trained AI model on a new input. It repeats on every use, so its total demand in large services becomes very large.

Q2. What new capacities is the government adding?

According to the Secretary, a government facility of about 3,300 GPUs at C-DAC Pune and a 30-petaflop HPC-AI facility in Bengaluru. There is also the "Condor Galaxy India" arrangement with G42 (8 exaflops, 64 Cerebras CS-3 systems), which the Secretary said will be used primarily for inference. C-DAC has also floated a tender for empanelment of AI compute service providers.

Q3. Are new models being selected under the IndiaAI Mission?

Not for now. As per the Secretary, the priority is to help the companies selected in the first round make their models commercially available.

Q4. How many GPUs are empanelled under the IndiaAI Mission?

A total of 44,077 GPUs, as per the official list on the IndiaAI compute portal.

Sources

1. MeitY Secretary's remarks at Tec-Verse 2026, The Hindu (27 November 2026): https://www.thehindu.com/business/Industry/meity-plans-to-expand-national-computing-capacity-for-ai-inference/article71750812.ece
2. The Hindu BusinessLine, independent report of the Secretary's remarks: https://www.thehindubusinessline.com/news/evaluating-expanding-compute-support-to-ai-inferencing-under-indiaai-mission-says-meity-secretary/article71751432.ece
3. IndiaAI compute portal, empanelled GPU list: https://compute.indiaai.gov.in/files/empanelled-gpus.pdf
4. Cabinet approval of the IndiaAI Mission, PIB: https://www.pib.gov.in/PressReleasePage.aspx?PRID=2012355
5. G42 and the Government of India, commercial framework for Condor Galaxy India (15 May 2026): https://www.prnewswire.com/news-releases/g42-and-government-of-india-formalize-commercial-framework-for-condor-galaxy-india-ai-supercomputer-302773650.html
6. C-DAC, RFE for empanelment of AI service providers (tender notice): https://www.cdac.in/index.aspx?id=tenders_DateChangeCorrigendum_13Nov26-1

PCSMANTRA.COM