

BharatCompute for AI ascendancy

business-standard.com/opinion/columns/bharatcompute-for-ai-ascendancy-124052801943_1.html

May 28, 2024

India needs a strategy that goes beyond setting up graphical processing units and focuses on leveraging its abundant talent



[Ajay Kumar -](#)

5 min read Last Updated : May 28 2024 | 9:50 PM IST

Artificial Intelligence (AI) capability rests on four key pillars: Algorithms, data, talent, and AI compute, with AI compute often considered the most crucial. Graphical processing units (GPUs) have become the standard measure of AI compute, spurring a global race to build massive GPU capacities.

In this context, the government's March decision to deploy over 10,000 GPUs as part of the India AI Mission is a significant step forward. This initiative, to be achieved through public-private partnership (PPP), reflects India's willingness to leverage the strengths of the private sector, a departure from past practices of confining such capital-intensive national capacities to the public sector.

[Click here to connect with us on WhatsApp](#)

India's AI compute capacities are modest compared to leaders like the US and China. The global GPU market was valued at \$23 billion in 2022, with an estimated 1 to 2 billion GPUs worldwide, predominantly in the US and China. Both countries have aggressively acquired GPUs over the past decade and have bigger plans to advance their AI compute capabilities. The US launched the National AI Research Resource programme in January,

while China aims to increase its aggregate compute power by over 50 per cent, including 10 exascale systems, by 2025. India cannot and does not need to mimic their strategy. Instead, it should adopt a smart fast-follower's approach, which I term BharatCompute.

Why is AI compute so important? In AI, "compute" refers to the computational power needed to perform complex operations, like training and running AI models. It is crucial for optimising deep learning models, which often have millions or billions of parameters involving large datasets requiring extensive computation. Additionally, compute power is needed for real-time inference, enabling predictions from new data.

The global demand for AI compute surged over the past decade, driven by the computational needs of the large-language-models (LLMs). However, as LLMs become widely available, this demand is stabilising, shifting focus to models for image processing, gaming, multi-modal AI and such other models, which require less computational power. BharatCompute could prioritise developing foundational models in economic, social and strategic domains while leveraging existing open-source LLMs already developed. This approach promises significant dividends even with lower compute resources.

BharatCompute should also aim to leverage India's existing CPU or Central processing unit-based compute capacities for AI advancements, with India's robust information technology (IT) industry and large user base providing substantial CPU-based compute power, compensating for its lack of GPUs. Despite CPUs being slower for AI tasks, leveraging this resource can enhance accessibility and reduce costs.

The BharatCompute initiative advocates for government-established AI compute infrastructure to be accessible to those unable to afford market rates. Currently, AI compute costs range from \$1 to \$4 per GPU hour, alongside additional expenses for memory and IT infrastructure. Government subsidies could encourage more startups to enter the AI field with the growth of AI startups serving as a metric for the initiative's success. The government could also prioritise developing foundational models in domains like agriculture, education, healthcare, and water, among others, tailored to Indian requirements.

BharatCompute should support India's sovereign AI needs, especially in strategic and security domains. Currently, the AI compute market relies heavily on cloud infrastructure, controlled by US and Chinese hyperscalers. A study shows that US-based Amazon Web Services, Microsoft Azure, and Google Cloud command about 70 per cent of the global public cloud market, while Chinese tech giants Alibaba, Huawei, and Tencent control much of the rest. Public clouds pose privacy and security concerns. To ensure security and sovereignty, BharatCompute should require that the compute infrastructure be physically located in India and utilise an Indian cloud platform. The government could dedicate a portion of the compute infrastructure in a secure, non-cloud environment for sensitive or classified applications.

Setting up AI compute in a PPP model is a welcome step. This can be operationalised by empanelling a set of Indian vendors who contribute to a pool of 10,000 GPUs, which the government underwrites. Such a process would be much faster than setting up a greenfield facility and ensure a competitive ecosystem providing cost-efficient, high-quality services. The private sector will bring in resources, skills and innovation. However, the PPP framework must address potential pitfalls, including continuity in case of vendor change, avoiding vendor lock-in, data security and privacy issues, and ensuring access to micro, small and medium enterprises and startups. Therefore, it is felt that the management of BharatCompute should be with an autonomous body, led by industry leaders with professional expertise and with participation from government representatives.

The AI compute market in India is still nascent. While the Indian industry is ramping up AI compute infrastructure, with some companies developing it for internal use and others offering it as a service, BharatCompute should do more than just set up 10,000 GPUs. It should catalyse the larger AI compute market and encourage industry to scale-up capacities significantly. Granting preferential access to domestic players for the 10,000 GPUs can give industry a much-needed kick-start. Proposing tax breaks to incentivise private investment in AI compute infrastructure and supporting customised AI chip-development for foundational models is recommended. The booming market for customised AI chips, highlighted by Nvidia's growth, OpenAI's Sam Altman's \$ 7 trillion chipmaking venture, as well as Microsoft and Amazon's entry into AI chip design suggests a significant emerging opportunity. India, with its abundant talent-pool in chip design, is well-positioned to capitalise on it. It has an abundance of talent and data. BharatCompute could be its strategy to leverage these strengths for AI ascendancy.

The writer is distinguished visiting professor, IIT Kanpur, and a former defence secretary

Disclaimer: These are personal views of the writer. They do not necessarily reflect the opinion of www.business-standard.com or the Business Standard newspaper

Topics : [Artificial intelligence](#) [BS Opinion](#) [Indian companies](#)

[Don't miss the most important news and views of the day. Get them on our !\[\]\(a03a7eb2f4046e1d3c76772003e549ea_img.jpg\) Telegram channel](#)

First Published: May 28 2024 | 9:34 PM IST