INDIA NARRATIVE

Beyond the Chip War

India's Blueprint for Inclusive AI

TABLE OF CONTENTS

Executive Summary

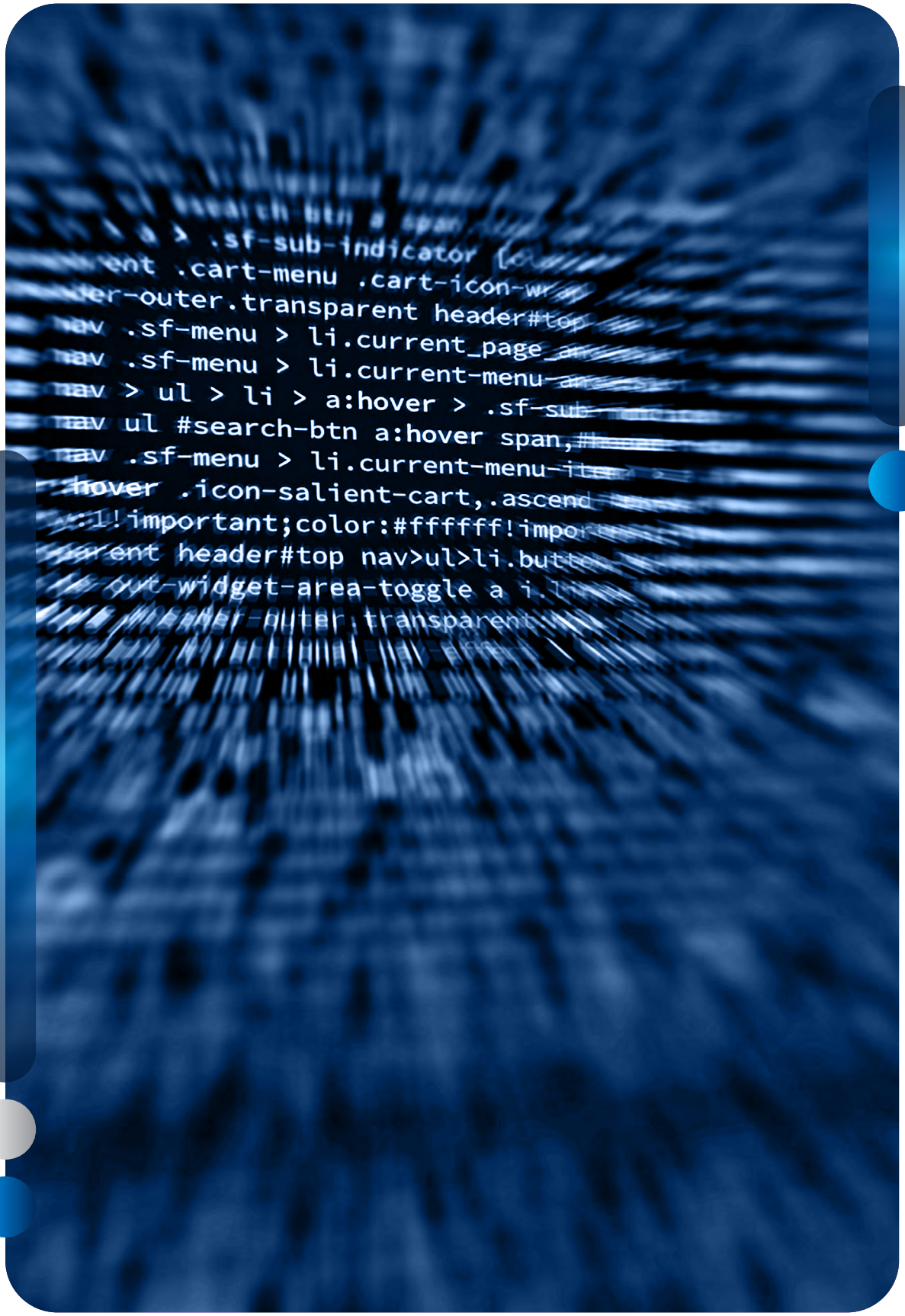
The global artificial intelligence race has definitively shifted from algorithmic research to physical infrastructure and compute capacity. As global superpowers engage in "GPU Nationalism" and funnel billions into hardware, India is charting a distinct, highly effective course. Currently ranking 3rd globally in AI vibrancy, India is transitioning from a traditional IT services hub into the world's foremost AI application engine.

While Western technology giants invest hundreds of millions of dollars into monolithic, proprietary foundation models, these systems often struggle to process complex Indic languages and code-switched dialects efficiently. In response, India is fostering a robust Sovereign AI ecosystem. Driven by indigenous platforms like Krutrim, Sarvam AI, and BharatGPT, the nation is building specialized, highly efficient models tailored to serve over a billion local consumers while adhering to strict data sovereignty requirements.

India's unparalleled advantage lies in its "Application-First" strategy. By layering AI directly onto its existing, world-class Digital Public Infrastructure (DPI)—such as Aadhaar, UPI, and ONDC—India is democratizing access to the digital economy. Initiatives like Bhashini leverage multilingual AI translation to bring non-literate and rural populations online via voice-first, scalable applications.

To fuel this ecosystem, the Indian government is acting as the primary catalyst through the $1.25 billion IndiaAI Mission. By dedicating over $540 million specifically to compute infrastructure—scaling up to deploy tens of thousands of subsidized GPUs at roughly $0.78 per hour—and allocating over $240 million in risk capital for deep-tech startups, the state is actively eliminating the hardware bottlenecks that traditionally stifle innovation.

India is effectively bypassing the trillion-dollar hardware arms race to secure its AI future. By prioritizing frugal innovation, open-source fine-tuning, and unparalleled application scaling, India has designed an inclusive, globally competitive AI paradigm that serves as the ultimate blueprint for the Global South.


.sf-sub-indicator
.cart-menu .cart-icon-wr
er-outer.transparent header#to
.sf-menu > li.current_page
.sf-menu > li.current-menu-
> ul > li > a:hover > .sf-sub
ul #search-btn a:hover span,
.sf-menu > li.current-menu-i
hover .icon-salient-cart,.ascend
important;color:#ffffff!impo
header#top nav>ul>li.but
-widget-area-toggle a i
transparent

01

The Global AI Frontier

Trajectory & Geopolitics

1.1
The Infrastructure Era of AI

For most of the last decade, the AI race was won and lost in code — better architectures, cleverer training tricks, larger datasets. That era is ending. The global AI market has nearly tripled in four years, expanding from $103.6 billion in 2020 to an estimated $288.8 billion in 2024, and this growth is increasingly underwritten not by algorithmic breakthroughs but by physical infrastructure: data centers, power grids, and silicon.

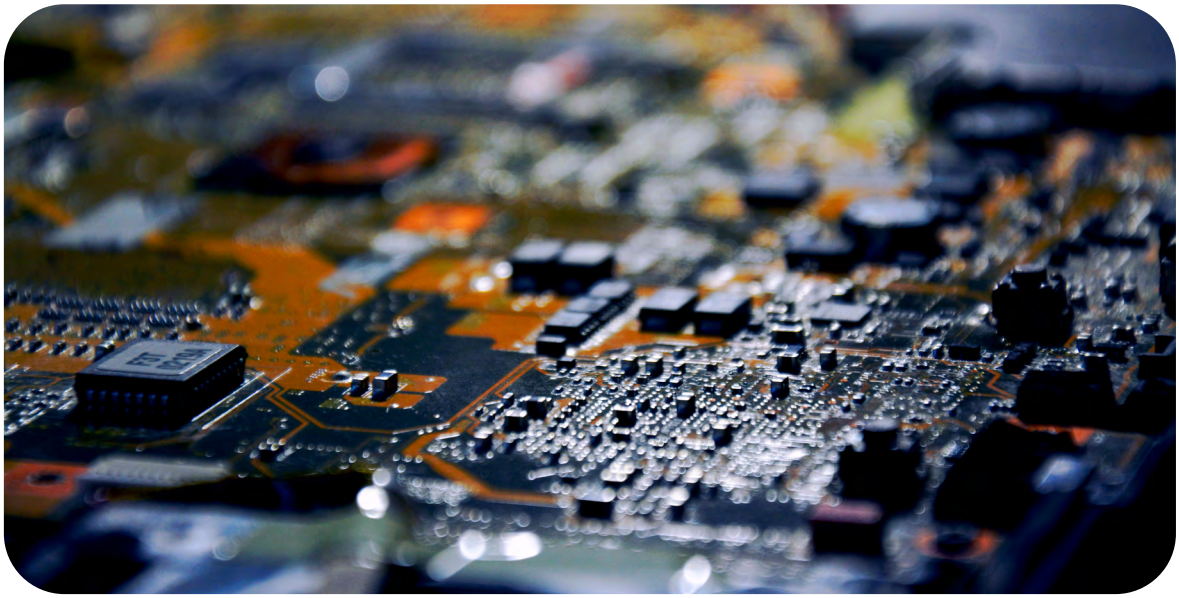
The clearest signal of this shift is cost. Training a single frontier model now runs anywhere from $100 million to $1 billion, a figure that has quietly rewritten the rules of competition. When the dominant cost of building a leading model is compute rather than talent, the barrier to entry stops being a research problem and becomes a balance-sheet problem. Nations and corporations that can marshal capital, energy, and chip supply at scale now hold a structural advantage that clever engineering alone cannot offset. AI has become a capital-intensive industrial project, closer in character to semiconductor fabrication or energy infrastructure than to software development — and that reframing is what drives everything else in this chapter.

1.2
GPU Nationalism & Supply Chain Geopolitics

Because advanced AI chips are manufactured at scale by a small handful of firms, compute has become a chokepoint that governments can control the way they once controlled oil or uranium. The United States has spent the past several years tightening export restrictions on advanced GPUs bound for China, closing loopholes with each successive rule — from the initial 2022 ban on flagship chips, through restrictions on downgraded variants built specifically to skirt earlier thresholds, and more recently through 2026 measures including a tariff on advanced AI chips and a shift toward case-by-case licensing rather than blanket approval.

The result has been a fracturing of the global compute map into tiers of access. Some countries, including India, sit in the top tier with largely unrestricted access to US chip technology, giving them room to build out sovereign compute capacity without the workarounds other nations must adopt. China, facing sustained restriction, has pushed hard into domestic alternatives — most notably training large models on Huawei's Ascend chips rather than Nvidia hardware — while simultaneously continuing to seek access to restricted chips through both legal carve-outs and, per multiple 2026 enforcement cases, illicit smuggling networks. Even with restricted access, analysts estimate China trails US AI compute capacity by a wide margin, though the gap is narrowing as domestic chip production scales.

This dynamic has pushed a growing list of governments — the EU, Saudi Arabia, the UAE, Japan, the UK, and India among them — to treat compute as a matter of national resilience, building or subsidizing domestic GPU reserves so that frontier AI development doesn't remain hostage to another country's export policy. "Sovereign AI" has become the operative phrase: the idea that a nation's capacity to train and run its own models is now a strategic asset on par with energy independence.



1.3
The Global Governance Divergence

Access to chips is only half the story; how nations choose to govern what gets built with them is the other. Three broad models have emerged. The European Union has taken a precautionary, rules-first approach, embedding risk classifications and compliance obligations into law through its AI Act before large-scale harms materialize. The United States has largely favored a market-led model, leaving development to industry with targeted intervention concentrated on export control and national security rather than broad domestic regulation. China's model is state-directed, with AI development folded into national industrial strategy and increasingly shaped by military-civil fusion priorities.

India has charted a distinct fourth path. Rather than adopting a standalone AI law, India has layered a "techno-legal," innovation-first framework atop its existing statutes — data protection, IT law, sector-specific rules — while introducing dedicated AI governance guidelines that favor light-touch, risk-based oversight over preemptive restriction. Regulatory intervention is meant to be triggered by evidence of harm rather than imposed in advance, a philosophy reinforced by initiatives like the IndiaAI Mission's push toward sovereign compute and by a 2026 bilateral agreement with the US explicitly committing both countries to pro-innovation AI regulation. It is a bet that speed and openness, backed by a safety net rather than a wall, is the more durable route to competitiveness — a bet that positions India less as a rule-taker in the global AI order and more as an architect of a third regulatory model between Brussels' precaution and Beijing's control.

02

Global AI Trends & India's Comparative Position

2.1
The Edge AI & Small Language Model (SLM) Shift

While the US and China chase ever-larger frontier models, 2026 has brought a quieter but equally consequential trend: the industry's pivot toward small, task-specific models that run efficiently on private servers and edge devices rather than in massive cloud clusters. Gartner's projection that organizations will use small, task-specific AI models three times more than general-purpose LLMs by 2027 captures a broader recalibration — enterprises are discovering that a compact, fine-tuned model often delivers most of a frontier model's practical value at a fraction of the compute cost.

This shift plays directly to India's structural realities. In a country defined by linguistic diversity, uneven digital infrastructure, and cost-sensitive enterprise buyers, SLMs offer a more viable path to scale than trillion-parameter cloud models. EY's 2026 analysis of India's AI landscape frames this explicitly: the country is positioning small language models — typically ranging from one to a few billion parameters — as the practical backbone for multilingual, regulated, and compliance-heavy sectors like banking and healthcare, while reserving large frontier models for the harder reasoning tasks that genuinely require them. Homegrown players such as Sarvam AI have emerged as part of a small global cohort of non-US, non-China SLM developers building for exactly this niche. The emerging Indian model isn't SLM-versus-LLM; it's hybrid — local models handling the high-volume, everyday workload, with cloud-based frontier models invoked selectively for complex cases. This architecture lowers the capital bar for AI adoption considerably, which matters enormously for a market where the constraint has rarely been ambition and often been infrastructure.

2.2
Benchmarking the Big Three
(US, China, India)

Stanford HAI's 2025 Global AI Vibrancy Index — which scores 36 countries across R&D, talent, infrastructure, policy, economy, responsible AI, and public engagement — now places India third globally, a jump from seventh place just two years earlier, with a score of 21.59 comfortably ahead of South Korea (17.24) and the UK (16.64), whom it has now overtaken. The ranking is unambiguous about the shape of global AI power: the United States leads by a wide margin, China follows, and India now sits alone in third, ahead of every other economy including traditional tech powers like the UK, Japan, and Singapore.

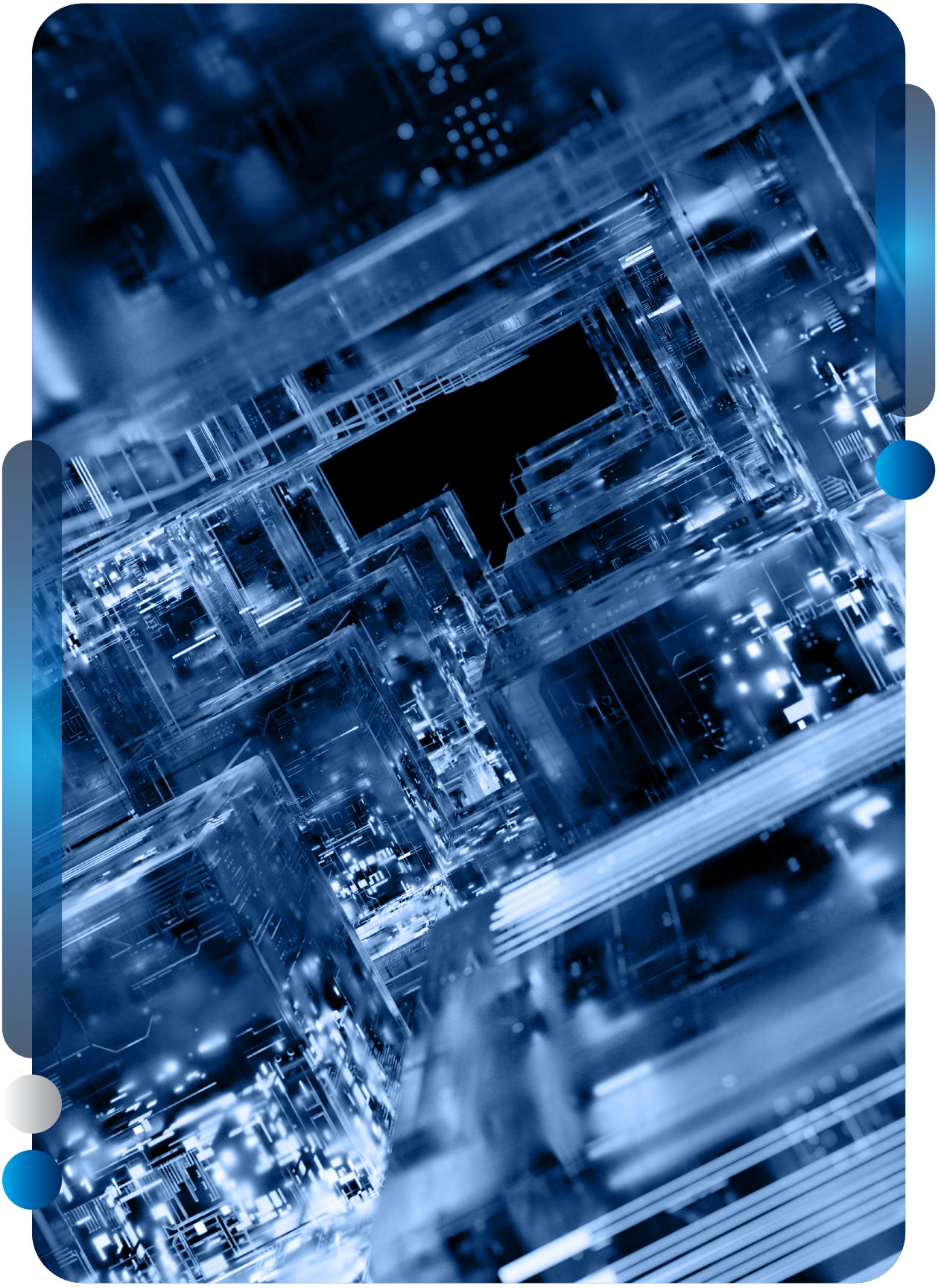
The three countries, however, are not competing on the same terrain. The US retains dominance in capital and frontier compute — training runs costing hundreds of millions of dollars remain concentrated among a handful of American labs with access to the newest chip generations. China compensates for restricted chip access with state-directed scale, domestic silicon substitution (training frontier-class models on Huawei's Ascend chips), and an industrial policy that treats AI as a matter of national strategy. India's comparative advantage lies elsewhere: it leads globally in AI conference citations and ranks near the top in total AI journal publications, and its officials have pointed out that in raw talent metrics India already ranks second worldwide — a reflection of workforce depth rather than capital depth. Where the US and China are racing on infrastructure and models, India is racing on people and deployment velocity — the question being whether the world's largest AI talent pool can be converted into disproportionate application-layer leadership even without matching either rival on compute.

2.3
The Talent Moat:
India's Demographic Dividend in Deep-Tech

The clearest, most quantifiable place India leads outright is open-source software contribution. Indian developers accounted for nearly one-fifth of all global AI projects on GitHub in 2024 — a striking figure given that the US and China together dominate proprietary frontier model development. This open-source weight matters strategically: it signals a developer base building tools, fine-tunes, and applications on top of existing models at a volume no other country outside the US and China can match, and it is precisely the kind of distributed, bottom-up innovation that compounds faster than any single national lab's output.

That developer base sits atop genuine market momentum. NASSCOM, in partnership with BCG, projects India's AI market will reach $17 billion by 2027, growing at a CAGR of 25–35% — a pace that outstrips most mature tech markets and is being driven by rising enterprise tech spending, an expanding AI-skilled workforce, and a surge in AI investment. Demand for AI talent in India is itself expected to grow at roughly 15% CAGR through 2027, with specialized roles like AI engineering already growing at 67% year-on-year, and India sits among the top five nations globally for growth in AI-skilled individuals over the past seven years.

This is the substance behind the "back-office to engine room" reframing. India's edge was never going to be matched compute — that game belongs to Washington and Beijing's balance sheets. India's edge is a demographic one: a vast, English-fluent, increasingly AI-literate engineering workforce feeding both global capability centers and a fast-maturing domestic startup ecosystem, converting talent depth into applied AI adoption at a rate few other economies can replicate. The R&D-versus-application-scaling divide that defines this chapter's comparative frame is, in effect, India's core strategic bet: while frontier research remains concentrated in the US and China, India is positioning itself as the place where AI actually gets deployed, localized, and scaled into products used by over a billion people.



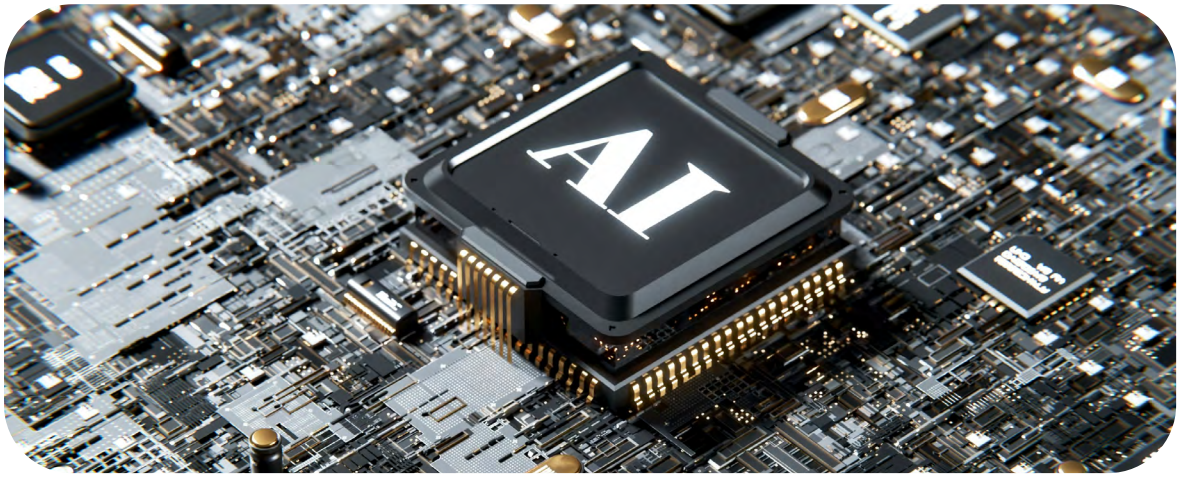
03

The Proprietary Foundation Model Bottleneck & Sovereign AI

3.1
The Economics of the Token Penalty

Western foundation models were built on an English-dominated internet, and that origin shows up first in the mechanics of tokenization — the process by which a model breaks text into billable units before it can read or generate anything. Because tokenizers like OpenAI's are trained on corpora overwhelmingly weighted toward English, common English words compress into single tokens while equivalent words in Indic scripts routinely fragment into many. One recent analysis found Hindi text costing roughly 63% more per million words than English on GPT-4o pricing, while a peer-reviewed study of ten Indian languages found a mean "tokenizer tax" that leaves Indian-language users with a fraction of the effective context window an English speaker gets for the same content. Academic work on Sanskrit-derived text found translation into English or Hindi could inflate token counts by roughly twentyfold relative to the source. For low-resource scripts and complex code-mixed dialects like Hinglish, the penalty compounds further, since Western tokenizers frequently have no clean vocabulary coverage for either the script or the blend.

This is not a cosmetic inefficiency — it is an economic one. An Indian enterprise processing the same conversational volume as a US counterpart pays measurably more per query, gets a smaller effective context window for the same content, and absorbs higher latency because more tokens must be generated and transmitted for identical meaning. At the scale of call centers, banking chatbots, or government service platforms fielding hundreds of millions of interactions a month, this "language tax" becomes a structural cost disadvantage baked into the infrastructure itself — one that newer tokenizers have partially narrowed but not eliminated. For a country whose AI ambitions depend on serving over a billion users across 22 official languages and hundreds of dialects, renting Western infrastructure priced against an English-centric cost curve is simply not viable at scale.



3.2
Cultural Hallucinations & Data Sovereignty

The token penalty is a cost problem; the deeper problem is comprehension. Models trained predominantly on English and Western-web data don't just process Indic languages inefficiently — they often misunderstand them. Idiom, code-mixing, regional context, and culturally specific reasoning frequently get flattened or misread, producing outputs that are grammatically plausible but contextually wrong — the linguistic equivalent of a hallucination rooted not in factual error but in cultural blindness. A model that has barely seen naturally occurring Hinglish, or that lacks deep training exposure to a specific state's dialectal patterns, cannot reliably power a customer service bot or a citizen-facing government platform without embarrassing or costly failures.

Layered on top of this is a harder strategic concern: dependency. Routing citizen data, financial transactions, and government services through foreign-hosted APIs means critical national infrastructure runs on decisions made by companies and governments outside India's jurisdiction — subject to pricing changes, service disruptions, export-control shifts, or policy decisions entirely outside India's control. This is precisely the vulnerability that "Sovereign AI" is designed to close: the principle that a nation's capacity to train, host, and govern its own models — on its own data, under its own legal jurisdiction — is now a matter of strategic resilience, not merely commercial preference. For regulated sectors like banking, healthcare, and public services, strict data localization isn't a nice-to-have; it is often the only way to guarantee that sensitive citizen data never leaves Indian soil.

3.3
The Rise of Indic Foundation Models

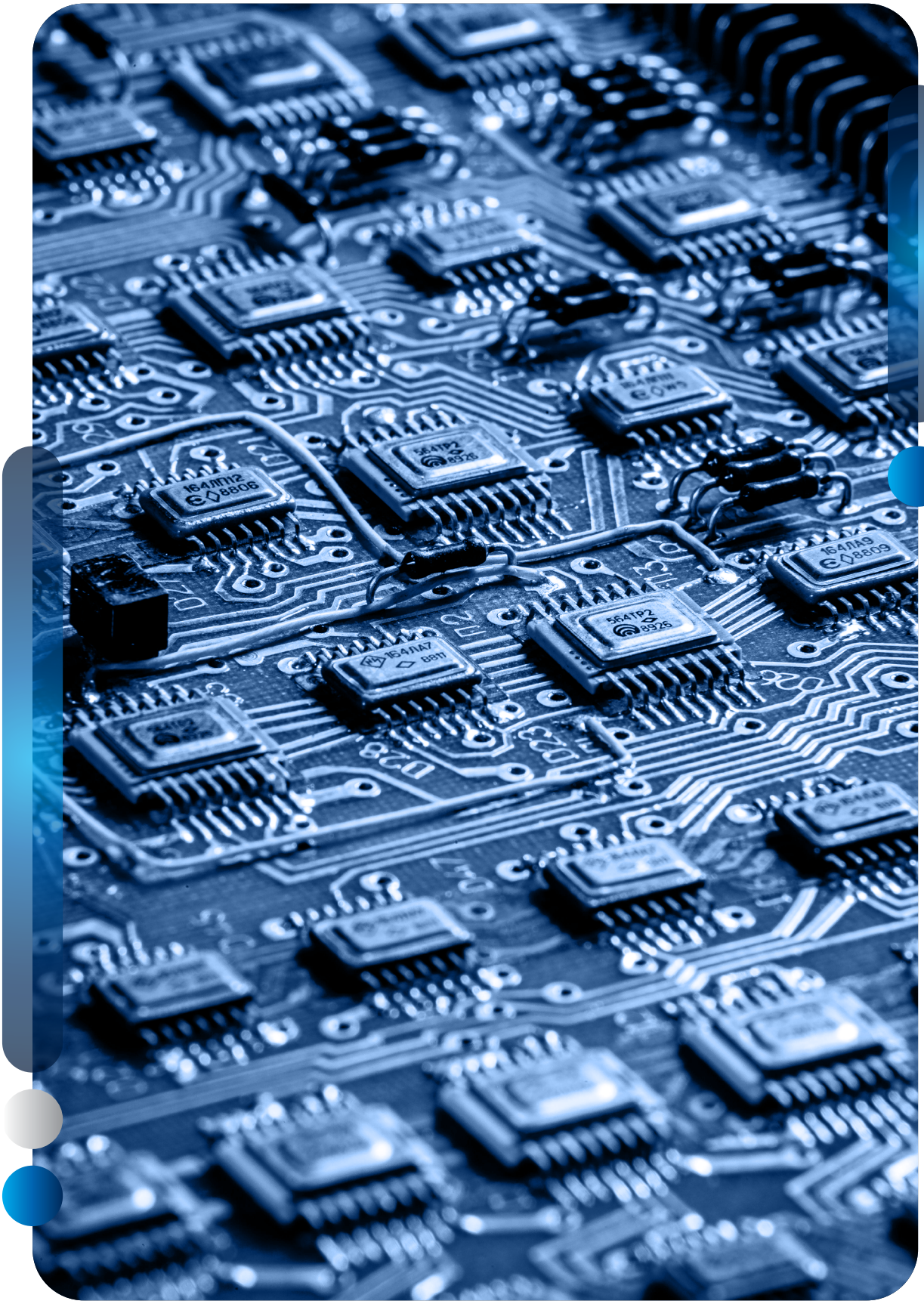
India's answer has been to build rather than rent, and three companies illustrate three distinct strategic bets.

Sarvam AI, backed by Peak XV and Lightspeed and now also supported under the government's IndiaAI Mission, has focused on solving the audio-latency problem that generic Western models struggle with. Its unified audio-to-audio pipeline pushes voice response latency under 500 milliseconds — the threshold below which a phone-based conversation with an AI agent starts to feel natural rather than robotic — making it viable for real-time banking, collections, and customer service bots operating in Hindi and regional dialects. At the India AI Impact Summit 2026, Sarvam extended this further, unveiling a large vision model reporting strong accuracy on Indic-language document understanding and demonstrating on-device translation across ten Indian languages at sub-200ms latency, including a feature-phone assistant built to work without an internet connection — a direct bet that India's next hundred million AI users won't be reached through smartphones and cloud APIs at all.

Krutrim, built by Ola and holding the distinction of being India's first AI unicorn, has taken the opposite approach: scale and breadth. Trained on more than two trillion tokens, Krutrim is designed as a general-purpose consumer and enterprise Indic model, built to power everything from Ola's electric vehicle platform to consumer-facing apps natively in Indian languages rather than through English-first translation layers.

Hanooman, developed by the BharatGPT consortium in collaboration with IIT-Bombay and backed by Reliance's infrastructure muscle, and hosted on Google Cloud specifically to guarantee data localization, targets the regulated-sector and government use case most directly — integrating voice capabilities across 14-plus Indian languages for deployments in healthcare, banking, and public services where data residency isn't optional.

Together, these three models mark a shift in posture: rather than treating Indic-language capability as a fine-tuning afterthought bolted onto a Western base model, Indian AI companies are building the foundation layer itself — tokenizer, training corpus, and voice pipeline — natively for the market it's meant to serve.



04

Future Prospects

India's Application-First Strategy

4.1
The 'Application Layer' Advantage

The preceding chapters have made the case for why India cannot, and likely should not, try to out-compete the US and China on frontier model training. The capital required — hundreds of millions to a billion dollars per training run, backed by chip stockpiles most nations don't have — makes that race a two-horse contest. But India's strategists have arrived at a different, and arguably more pragmatic, conclusion: the greatest economic value in AI doesn't sit at the foundation-model layer at all. It sits at the application layer, where fine-tuned, often open-source models get wired into the systems people actually use.

This is not a consolation strategy — it's a return-on-investment calculation. Training a frontier model from scratch requires enormous fixed capital before it produces a single dollar of value. Fine-tuning an existing open-source model for a specific language, sector, or user population costs a fraction of that, and the resulting deployment reaches revenue-generating or citizen-serving scale almost immediately. For a country with 1.4 billion people, a large share of them rural, non-English-speaking, or new to digital services, the marginal return on "make an existing model work well for this specific population" vastly outstrips the marginal return on "build a marginally better general-purpose model." India's real asymmetric advantage isn't compute — it's an existing, extraordinarily cheap distribution backbone to build AI applications on top of.

That backbone is India's Digital Public Infrastructure.

4.2
AI & Digital Public Infrastructure (DPI):
Supercharging UPI, Aadhaar, and ONDC

Over the past decade, India has built three foundational digital rails at national scale: Aadhaar, the biometric identity system covering over a billion residents; UPI, the real-time payments network now handling the majority of the world's real-time transaction volume; and ONDC, the Open Network for Digital Commerce, an open-protocol effort to unbundle e-commerce from single dominant platforms. Each of these systems was built as a public good — open, interoperable infrastructure that any developer or business can build on top of, rather than infrastructure locked inside a single corporate platform.

This is precisely why AI adoption in India looks structurally different from AI adoption in the US or China. Where American and Chinese AI value tends to accrue inside closed platform ecosystems, India is layering generative AI directly onto public rails that already reach the entire population. NPCI's collaboration with Bhashini to launch "Hello! UPI," a voice-enabled payments feature, is an early example: it lets a UPI transaction be initiated by voice in a local language rather than requiring literacy in English or comfort navigating a text-based app interface. The same logic extends to ONDC, where AI-driven translation and voice tools are being used to let small merchants — who would never have the resources to build a multilingual app themselves — plug directly into a national commerce network.

The strategic bet here is that AI's biggest economic multiplier in India isn't a smarter chatbot for the already-connected, English-speaking, urban middle class. It's unlocking DPI access for the several hundred million Indians who are digitally included in principle (they have Aadhaar, a bank account, a UPI-linked phone) but functionally excluded because the interface assumes literacy and English fluency they don't have.

4.3
Bridging the Digital Language Divide:
The Bhashini Ecosystem

The platform doing the most structural work in this strategy is Bhashini — the National Language Translation Mission, launched by the Prime Minister in 2022 and run by the Digital India Bhashini Division under the Ministry of Electronics and Information Technology. Bhashini is not a single product; it is a public, open-source language infrastructure layer, offering translation, transliteration, speech-to-text, and text-to-speech APIs covering 22 Scheduled Languages and an expanding set of tribal and regional dialects. Crucially, these APIs are made available free or at minimal cost to individual developers, startups, and government-empanelled applications — turning what would otherwise be an expensive, proprietary capability into a shared public utility that any builder can plug into.

This design choice is the crux of why Bhashini matters more than any single consumer-facing AI product. Because it is open infrastructure rather than a closed platform, Bhashini lets government agencies, startups, and NGOs each solve their own last-mile problem without each having to separately solve the underlying language problem. It powers real-time translation for citizen-facing government services, parliamentary proceedings through initiatives like Sansad Bhashini, and educational content — while remaining, in its own framing, a voice-first system explicitly designed to bridge literacy, language, and digital divides simultaneously, not just the language gap alone.

The clearest example of Bhashini's application-layer leverage in action is Saarthi, a reference application launched by ONDC in partnership with Bhashini. Saarthi gives small-town and rural merchants — sellers who could never afford to build a custom multilingual e-commerce app — a ready-made framework for creating their own personalized, multilingual buyer-facing applications on the ONDC network. Rather than forcing every seller and buyer onto a single English-first interface, Saarthi lets a merchant's storefront speak the customer's language, initially covering Hindi, English, Marathi, Bangla, and Tamil, with an explicit roadmap to scale across all 22 Bhashini-supported languages. Features like real-time translation, transliteration, and voice recognition mean a small merchant in a Tier 3 town can transact with a buyer in another state without either party needing a shared language — effectively using AI to erase a barrier that has excluded non-English speakers and non-literate users from digital commerce since its inception in India.

Taken together, Bhashini and Saarthi illustrate the thesis of India's application-first strategy in miniature: India does not need to train the world's most capable foundation model to extract enormous economic and social value from AI. It needs to take existing AI capability, wrap it around infrastructure that already reaches a billion people, and remove the last barrier — language — standing between hundreds of millions of citizens and the digital economy. That is a strategy no amount of raw compute, on its own, can replicate.

05

Policy Frameworks & Infrastructure Execution

5.1
The $1.25 Billion IndiaAI Mission:
Unpacking the Government's Sovereign Deep-Tech Investment

If Chapters 1 through 4 established the shape of India's AI strategy — application-first, talent-heavy, infrastructure-layered — this chapter asks who is actually paying for it. The answer, unusually for a market economy, is the state itself, acting less like a regulator and more like a sovereign venture capitalist.

The IndiaAI Mission, approved by the Union Cabinet in March 2024 with a five-year outlay of ₹10,371.92 crore (roughly $1.25 billion), is the vehicle for this role. Implemented by the IndiaAI Independent Business Division under the Ministry of Electronics and Information Technology, the mission is structured around seven pillars: subsidized compute capacity, an open datasets platform (AIKosh), application development for India-specific problems, indigenous foundation models, future-skills training, startup risk financing, and safe and trusted AI governance. The explicit framing behind all seven — "Making AI in India and Making AI Work for India" — signals that this isn't a research grant program bolted onto existing bureaucracy; it is a coordinated industrial strategy treating AI capability the way earlier decades treated highways, telecom, or power generation: as infrastructure too foundational to leave entirely to private capital.

What makes this approach distinctive is the explicit goal stated at the mission's founding — to democratize access to computing, data, and capital simultaneously — rather than picking winners through R&D grants alone. That triad matters because each on its own is insufficient: capital without compute access still leaves startups priced out of training runs; compute without startup financing still leaves early-stage teams unable to get to a fundable prototype. The mission was designed to remove all three bottlenecks at once, funded through a public-private partnership model rather than pure state spending.

5.2
Building the National Compute Cloud:
Democratizing GPUs for Academia and Enterprise

The most tangible, and most rapidly scaling, piece of this strategy is the national GPU cloud. The mission's original target was modest — 10,000 GPUs. By early 2026 it had overshot that target nearly fourfold, with more than 38,000 GPUs onboarded through the IndiaAI Compute Portal. Then, at the India AI Impact Summit in New Delhi on February 17, 2026, IT Minister Ashwini Vaishnaw announced "AI Mission 2.0" — a further expansion adding 20,000 GPUs, with orders placed within a week and deployment expected over the following six months, bringing the national pool toward roughly 58,000 units and part of a stated ambition to scale further by the end of the year.

The pricing is the real headline. These GPUs are made available to eligible startups, researchers, MSMEs, and academic institutions at a subsidized rate of approximately ₹65 per hour — about $0.78 — against commercial cloud rates in India that typically run several times higher. For a solo founder or a university lab, this isn't a marginal discount; it is the difference between being able to run a training experiment at all and being priced out of AI development entirely before a single model gets built. Framed against a Parliamentary record from March 2026, in which the government confirmed the more-than-38,000-GPU figure directly to the Lok Sabha, this compute layer has moved from pilot to standing national infrastructure — allocated through an application process weighted by eligibility, requested hours, and subsidy need, with larger requests subject to additional review to keep the pool broadly distributed rather than captured by a few large players.

The strategic logic is straightforward: compute has become the single largest barrier to entry in AI, as Chapter 1 established. A country that wants thousands of startups and research groups experimenting with model training — not just a handful of well-capitalized firms — has to either accept that most of that experimentation won't happen, or subsidize the barrier away. India chose the latter, effectively treating GPU-hours the way it once treated electrification: a public utility problem, not a market-clearing price problem.

5.3
Viability Gap Funding & Innovation Hubs:
Centers of Excellence and Startup Financing

Compute access alone doesn't solve the funding gap that kills most deep-tech ventures before they reach commercial viability — the stretch between a working prototype and a product a private investor is willing to underwrite. The IndiaAI Mission's Startup Financing pillar is designed explicitly to bridge that gap, providing risk capital across a startup's lifecycle from early seed-stage prototyping through to commercialization, rather than only funding mature, de-risked companies the way most private venture capital does.

Complementing the financing pillar is a rapidly expanding network of AI Centres of Excellence (CoEs), built in partnership with state governments, academic institutions, and industry. As of mid-2026, 58 AI CoEs have been approved across states and union territories, alongside more than 500 Data and AI Labs — many explicitly targeted at Tier 2 and Tier 3 cities rather than concentrated in existing metro tech hubs. These centers are not abstract research outposts: Kerala's Bio-AI CoE, for instance, is targeted to incubate 48 deep-tech startups across four annual cohorts, create 600 high-value jobs, and train 1,200 professionals in bioinformatics, while a CoE in Visakhapatnam has already channeled over ₹131 crore in funding to eight AI and IoT startups working on real-world manufacturing applications.

Together, subsidized compute, startup risk capital, and a distributed network of regional CoEs form a deliberately layered support structure — compute removes the technical barrier to building, financing removes the capital barrier to surviving the pre-revenue stage, and CoEs remove the geographic barrier that has historically concentrated Indian tech opportunity in a handful of metro cities. It is, in effect, the state constructing the entire early-stage AI venture stack itself, on the premise that if it waits for private capital alone to build this infrastructure, India's vast talent pool — the subject of Chapter 2 — will simply go build elsewhere.

Strategic Roadmap & Conclusion

Across the previous five chapters, a single thesis has taken shape: India is not trying to win the AI race the US and China are running. It has instead defined a different race entirely — one measured not in parameters trained or chips stockpiled, but in people reached. Where the global frontier is contested through capital-intensive compute and export-controlled silicon, India has built what might be called an **Application-Led, Inclusion-Focused AI paradigm** — layering fine-tuned, often open-source intelligence over world-class Digital Public Infrastructure, and routing that intelligence through voice and vernacular language rather than English-first text. Bhashini's translation layer, Sarvam's sub-500ms voice latency, Krutrim's Indic-native scale, and Saarthi's merchant-facing multilingual commerce are not isolated products; they are expressions of one coherent strategy: bring AI to the literacy-constrained, rural, linguistically diverse majority rather than waiting for that majority to adapt to AI built for someone else.

This matters beyond India's borders. The paradigm the US and China have exported — proprietary, English-centric, capital-gated — doesn't map cleanly onto most of the Global South, where linguistic diversity, patchy connectivity, and constrained public budgets closely resemble India's own starting conditions. A model built to serve a billion people across 22 languages on subsidized public compute is, almost by construction, a model built for export.

Three actions would convert this positional advantage into durable leadership:

1. Accelerate GPU-as-a-Service (GPUaaS) beyond the metros.

The IndiaAI Mission's 58,000-plus GPUs must be deployed and made reliably accessible in Tier 2 and Tier 3 cities, not concentrated in Bengaluru, Delhi, and Mumbai. Decentralizing compute is what turns a national resource into genuinely distributed innovation capacity, rather than reproducing the same metro-centric access patterns AI was supposed to disrupt.

2. Fast-track open government data.

India's healthcare, agricultural, and civic datasets — anonymized and released at scale through the IndiaAI Datasets Platform — are a training resource no foreign lab can replicate, because no foreign lab has access to Indian-context data at this scale. Releasing it faster directly fuels better-grounded local models.

3. Export the DPI-AI stack diplomatically.

India should package its combination — UPI-grade digital infrastructure plus Bhashini-grade voice AI — as a technology diplomacy offering to Africa and Southeast Asia, positioning itself not as a follower in the AI order but as the architect of the model most of the world will actually be able to use.

India's path forward, then, is not to catch up to the compute race being run in Washington and Beijing, but to render that race less relevant to how most of the world will actually experience AI. By proving that inclusion, not scale, can be the organizing principle of a national AI strategy, India offers something neither superpower has built: a replicable blueprint for nations that will never own the frontier lab, but can still own the way AI reaches their people.

References

01. Allen Institute for AI (AI2). (2024). *The economics of frontier AI training: Cost estimates for billion-parameter models and tokenization penalties*. AI2 Research.

02. Analytics India Magazine. (2025, October 28). *India surpasses US as world's largest base of open source contributors: GitHub Octoverse 2025*. AIM Publications.

03. Boston Consulting Group (BCG). (2025). *Where's the value in AI? The state of enterprise AI adoption in India*. BCG Global.

04. CoRover.ai. (2024). *BharatGPT and Hanooman: Revolutionizing generative AI for Indic languages*. CoRover Technical Publications.

05. Department of Science and Technology. (2025). *National Supercomputing Mission and the AIRAWAT expansion*. Government of India.

06. European Parliament. (2024, March 13). *Artificial Intelligence Act: MEPs adopt landmark law for comprehensive AI regulation*. European Parliament News.

07. GitHub. (2025). *Octoverse 2025: A new developer joins GitHub every second as AI leads*. GitHub Blog.

08. Global Partnership on Artificial Intelligence (GPAI). (2023). *New Delhi Declaration of the Global Partnership on Artificial Intelligence*. GPAI Secretariat.

09. Government of Rajasthan. (2024). *Integration of voice AI in the Pehchan portal for civil registration*. Department of Economics and Statistics, Rajasthan.

10. Grand View Research. (2024). *Artificial intelligence market size, share & trends analysis report by solution, by technology, by end-use, by region, and segment forecasts, 2024 - 2030*.

11. IndiaAI. (2024). *Empowering deep-tech: The ₹2,000 crore startup funding allocation under the IndiaAI Mission*. National AI Portal of India.

12. International Data Corporation (IDC). (2024). *Worldwide artificial intelligence spending guide and regional shift analysis*.

13. International Telecommunication Union (ITU). (2024). *AI for good: The Global South and the rise of Digital Public Infrastructure (DPI)*. United Nations.

14. McKinsey & Company. (2024). *The economic potential of generative AI: The next productivity frontier and emerging market growth*.

15. Miller, C. (2022). *Chip war: The fight for the world's most critical technology. Scribner*.

16. Ministry of Electronics and Information Technology (MeitY). (2022). *National Language Translation Mission (Bhashini): Breaking the language barrier*. Government of India.

17. Ministry of Finance. (2025). *Union Budget 2025-26: Allocation for Centres of Excellence in Artificial Intelligence*. Government of India.

18. Ministry of Health and Family Welfare. (2023). *Integration of AI in eSanjeevani national telemedicine service for primary health centers*. Press Information Bureau.

19. Ministry of Law and Justice. (2023). *The Digital Personal Data Protection Act, 2023*. The Gazette of India.

20. NASSCOM. (2024). *India's AI market to reach $17 billion by 2027: Deep tech insights*. National Association of Software and Service Companies.

21. NASSCOM. (2025). *AI adoption index: Empowering India's digital transformation*. National Association of Software and Service Companies.

22. NITI Aayog. (2018). *National strategy for artificial intelligence #AIforAll*. Government of India.

23. Open Network for Digital Commerce (ONDC). (2024). *Introduction to Saarthi: Multilingual digital commerce for rural merchants*. ONDC Network.

24. Press Information Bureau (PIB). (2024, March 7). *Cabinet approves IndiaAI Mission with an outlay of Rs. 10,371.92 crore*. Ministry of Electronics & IT, Government of India.

25. Press Information Bureau (PIB). (2026, February 12). *AI@Work: Driving productivity, jobs, and innovation*. Government of India.

26. Sarvam AI. (2023, December). *Introducing OpenHathi: A foundational model for Hindi and Indic dialects*. Sarvam AI Official Blog.

27. Stanford University. (2025). *Artificial Intelligence Index Report 2025: Global AI Vibrancy Ranking*. Stanford Institute for Human-Centered Artificial Intelligence (HAI).

28. TechCrunch. (2024, January 26). *Ola founder's Krutrim becomes India's first AI unicorn*. TechCrunch Media.