

Built Lean, Built to Scale:

An Artificial Intelligence
Compendium from
the Global South





Table of Contents

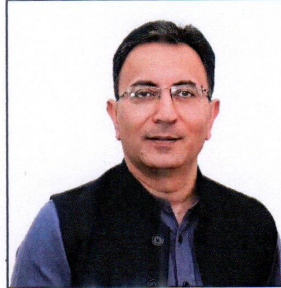
01	Foreword	04
02	Introduction	05
03	Compute	08
04	Data	24
05	Skilling	37
06	Conclusion	53

जितिन प्रसाद
JITIN PRASADA



राज्य मंत्री
वाणिज्य एवं उद्योग,
इलेक्ट्रॉनिकी और सूचना प्रौद्योगिकी
भारत सरकार

Minister of State
Commerce & Industry,
Electronics and Information Technology
Government of India



MESSAGE

Artificial Intelligence is becoming central to how economies grow. Its value should be measured by its impact on everyday life and the increase in the size of economy by either reduction of cost or increase in productivity. For the Global South, AI must expand access, support livelihoods and create prosperity at scale.

India's approach has been shaped by its own development journey. Under the leadership of Hon'ble Prime Minister Shri Narendra Modi ji, India built Digital Public Infrastructure that is open, interoperable and designed for more than a billion people. This experience shaped India's approach to AI: technology creates the greatest value when built as shared infrastructure and designed for access.

The IndiaAI Mission puts this principle into practice by expanding access to high-end computing at subsidised rates for startups, researchers, students and public institutions. It supports indigenous foundation models and wider access to datasets through AIKosh. BHASHINI enables multilingual and voice-based access to public services. The Mission also supports sector-specific AI solutions for agriculture, healthcare and governance. People remain central alongside Data and AI Labs for future skills.

India has also approached digital and AI infrastructure as shared global public goods. During its 2023 G20 Presidency, India launched the Global Digital Public Infrastructure Repository and supported adoption of India Stack, DigiLocker and UPI across 20 countries. At the 2025 AI Action Summit, India backed open-source systems, unbiased datasets and locally relevant solutions. At the 2026 India AI Impact Summit, the New Delhi Declaration on AI Impact was endorsed by 92 countries and international organisations.

Built Lean, Built to Scale shows these principles in practice. Across India and the Global South, innovators demonstrate that AI does not always require the largest models, expensive infrastructure or concentrated capital. Efficient models, representative data, local languages, shared compute and skilled people can produce practical, locally grounded and adaptable solutions.

The next chapter of AI must be a more collaborative effort. India stands ready to share, learn and build with the Global South so that AI strengthens national capabilities, expands opportunity and improves lives.

(Jitin Prasada)



कार्यालय : इलेक्ट्रॉनिक्स निकेतन, 6, सी.जी.ओ. कॉम्प्लेक्स, नई दिल्ली-110003, दूरभाष : 011-24368757-58
Office : Electronics Niketan , 6, C.G.O. Complex, New Delhi-110003, Tel. : 011-24368757-58
E-Mail : mos-eit@gov.in, Website : www.meity.gov.in

Introduction



As AI capabilities expand rapidly, countries worldwide are looking to build locally grounded, contextually relevant solutions rather than simply adopting what is built elsewhere. Currently, the Global North largely leads the AI race, controlling the core infrastructure, compute, chips, and the underlying supply chains, that AI development depends on. Holding limited control over either that infrastructure or the deployment of major AI systems, the Global South is not waiting on the sidelines for that balance to shift. It is charting its own path: one that starts from these dependencies rather than around them and treats them as the starting point for innovation rather than a limitation to work past. This shift reflects broader changes in global AI governance. Recent AI summits, from Bletchley Park in 2023 to Seoul in 2024 and Paris in 2025, initially focused on the existential risks of advanced systems. Each summit expanded its agenda, culminating in the 2026 India AI Impact Summit, the first hosted by a Global South nation and the first to prioritise real-world impact over frontier-model safety. Organised around People, Planet, and Progress, and seven themes including human capital, inclusion, trusted AI, resilience, science, democratisation of AI resources, and AI for economic and social good, the summit received support from over 100 countries. This approach aligns with the UN's Global Digital Compact, adopted in September 2024, which commits member states to govern emerging technologies, including AI, "for humanity," and to close digital divides through inclusive, development-focused cooperation. The global AI conversation is shifting from building the largest models to focusing on who benefits from AI and under what conditions.

This compendium examines the contextual solutions emerging from this shift, highlighting how the Global South is adapting, building, and at times leading in AI solutions, with constraints that the Global North are not equipped to address at present. It argues that the Global South is not competing on the Global North's terms, but instead innovating within existing gaps, limited data, compute, and skills to create equitable and responsible AI solutions by design. The compendium is structured around three key areas: compute, data, and skilling. Each chapter explores how the Global South is developing its own AI narrative, using constraints as drivers of innovation and positioning the region as a hub for adaptable, context-specific AI.

GLOBAL AI SUMMITS: AN EXPANDING AGENDA

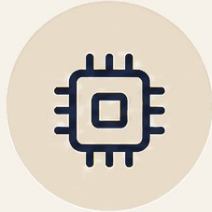


First hosted by a Global South nation, and the first to prioritise real-world impact over frontier-model safety.

To ground this argument beyond the global examples already discussed, each chapter also draws on case studies. These are built from conversations with the people actually doing this work. This includes founders and engineers at startups, bureaucrats who build and run public digital infrastructure, and technologists at established companies and big tech firms. Their insights sit alongside the sector-specific and infrastructure examples in each chapter. Together, they offer a closer, more textured view of how these solutions are actually built, deployed, and adapted on the ground. The aim is not only to show what has been achieved, but how, and against what practical, everyday constraints.

Through this compendium, we aim to surface the initiatives and voices emerging from the Global South that embody this ethos, and to document their lived experience of building AI on their own terms.

THE COMPENDIUM IS STRUCTURED AROUND THREE KEY AREAS



01

Compute



02

Data



03

Skilling



01

Compute



INTRODUCTION

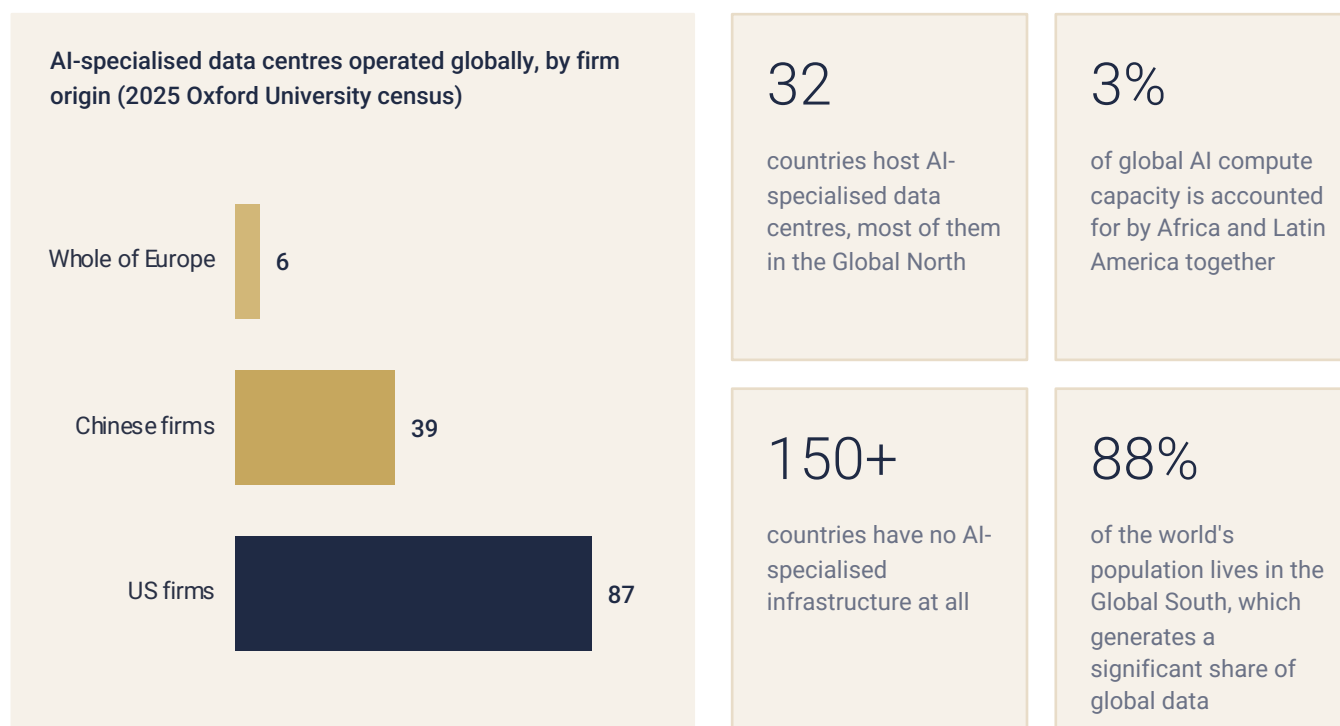
AI development is fundamentally shaped by access to compute, yet the capacity to build and deploy advanced AI systems remains unevenly distributed, with much of the world's high-end compute concentrated in the Global North. As countries across the Global South expand their AI capabilities, this imbalance is increasingly being approached not simply as a constraint, but as an opportunity to develop new models of AI suited to local needs, resources and conditions. South-South collaboration is reinforcing this shift, alongside the emerging idea of prosocial AI: moving beyond questions of whether AI is developed ethically to asking how AI infrastructure, capabilities and benefits can be made more broadly accessible.

This connects closely with the principles of AI for All and AI for Good, where the objective is not only to deploy AI responsibly but to ensure that its benefits reach underserved communities and economies. The chapter examines how countries in the Global South are responding to compute constraints through approaches such as GPU optimisation, frugal AI and Edge AI, which can reduce dependence on large-scale compute infrastructure while enabling practical deployment. The broader argument is that limitations in compute need not simply restrict AI development. They can also drive innovation in how compute is designed, optimised and deployed, creating more affordable and locally relevant AI solutions that can serve wider communities.

Much of the conversation around AI still centres on scale - how large data centres are, or how quickly AI infrastructure is growing. Scale is easy to measure, but it does not tell us where the ability to develop and use AI actually resides. Today, only 32 countries host AI-specialised data centres, most of them in the Global North, while Africa and Latin America together account for just 3% of global AI compute capacity. A 2025 Oxford University census of the world's AI data centres found that US firms alone operate 87 such facilities globally and Chinese firms 39, against just six for the whole of Europe, and that more than 150 countries have no AI-specialised infrastructure at all. Reports from Harvard's Kempner Institute indicate the scale of this disparity. The institute today possesses more computing power than all AI facilities owned by all African entities combined. Meanwhile, while the Global South constitutes 88% of the world's population and generates a significant share of global data, much of the infrastructure needed to process and use it sits outside the region.

This means the global AI divide will not be determined only by who adopts AI fastest. It will increasingly depend on who controls the basic inputs that allow AI to function: computing capacity, reliable and affordable energy, data centres, advanced chips, supply chains, capital and specialised talent. OECD research has similarly stressed the importance of countries building domestic compute capacity as part of a wider AI strategy, particularly for economic development, resilience, and technological sovereignty.

For the Global South, the consequences go beyond slower access to the latest models. Without sufficient domestic infrastructure, countries risk becoming primarily consumers of AI services built elsewhere. At the same time, their data, talent and economic value continue to flow into infrastructure controlled by a small number of countries and companies. This creates a structural dependence not just on technology, but on the infrastructure, costs and terms under which that technology is made available. Closing the AI divide, therefore, requires more than deploying applications - it requires building the underlying capacity to produce, process, and control AI within the countries meant to benefit from it.



Global South approach

India's approach has reflected the central principles of AI for All and AI for Good. As part of the IndiaAI Mission, the compute pillar was launched in 2024 with an initial target of 10,000 GPUs, empanelling cloud and data-centre providers through a public-private tender process. A deliberate choice was made to empanel sovereign cloud providers as part of this shared compute resource, so that the data being processed and trained on remains within India. Over the past two years, total empanelled capacity has grown to more than 38,000 GPUs as well as TPUs, including the first integration of Google's Trillium chips, nearly four times the original planned target.

Today, the compute pillar demonstrates what shared, subsidised compute can do at population scale. The IndiaAI Mission now offers some of the cheapest compute available anywhere, accessible to startups, researchers, students and academic institutions at a subsidised rate of roughly ₹65–83 (about \$1) per GPU-hour - typically 60–70% below commercial cloud pricing. The strategy does not eliminate the underlying capital cost, which is borne largely by the government's outlay of more than ₹10,000 crore. Still, it converts compute from a fixed cost that only well-funded labs can absorb into a variable, pay-as-you-go resource accessible to smaller players.

10,000 → 38,000+

GPUs: initial 2024 target versus total empanelled capacity today, nearly four times the original planned target

₹65-83 (about \$1)

subsidised rate per GPU-hour for startups, researchers, students and academic institutions

60-70%

Access to GPU's from the shared compute cluster is typically, below commercial cloud pricing

₹10,000 crore+

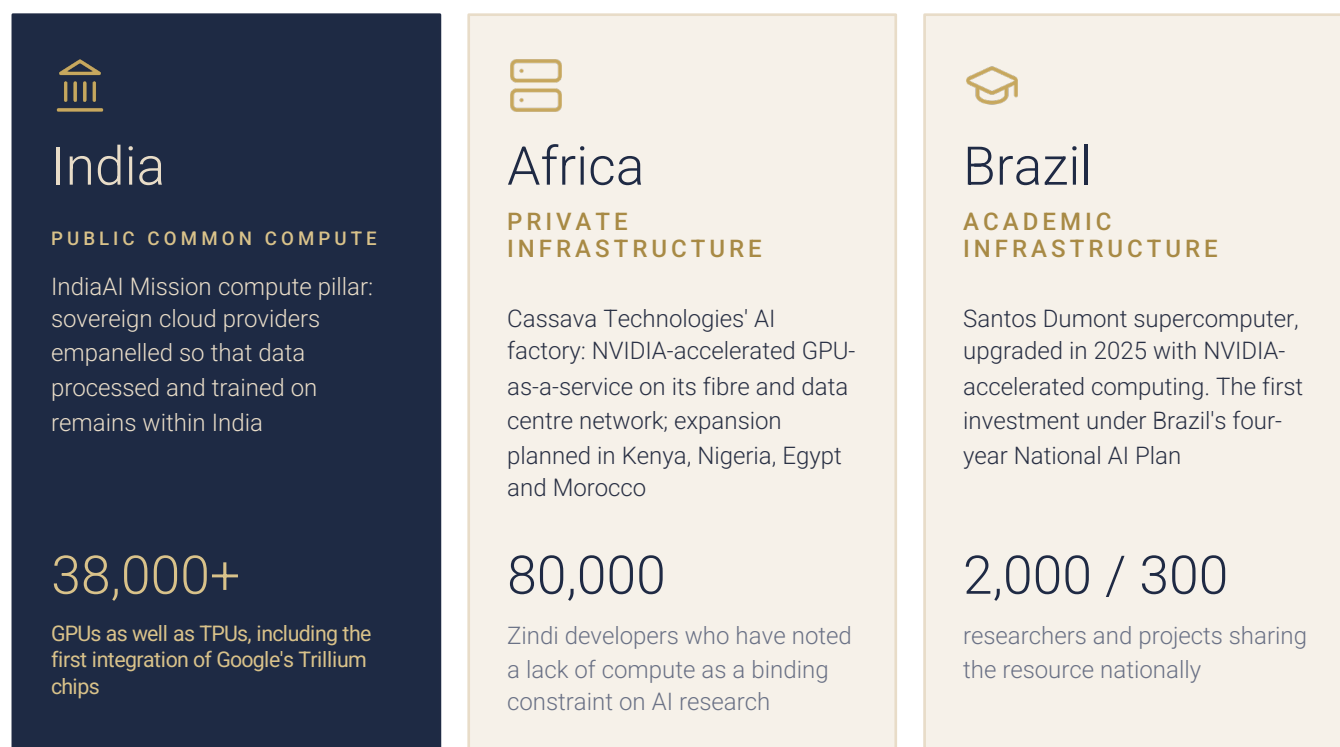
government outlay bearing the underlying capital cost



Africa is following a similar logic of pooled access, but through private rather than state infrastructure. Pan-African telecommunications and technology group Cassava Technologies launched the continent's first "AI factory" in 2025, deploying NVIDIA-accelerated GPU-as-a-service (GPUaaS) on its existing fibre and data centre network. This network expansion is also planned in Kenya, Nigeria, Egypt and Morocco. To further expand access for researchers, they also partnered with Zindi, whose 80,000 developers have noted a lack of compute as a binding constraint on AI research. Through these initiatives, they are committed to training models in African languages such as Swahili, Zulu and Afrikaans. Akin to India's philosophy, keeping the infrastructure within African jurisdiction is explicitly being framed as a sovereignty objective – ensuring that the data and workload are not processed on foreign servers.

While most Global South examples so far reflect a public-private or commercial route to procuring GPU capacity, Brazil offers the clearest example of compute built as academic rather than commercial infrastructure. Brazil's Santos Dumont supercomputer, a shared resource among roughly 2,000 researchers and 300 projects nationally, was upgraded in 2025 with NVIDIA-accelerated computing as the first infrastructure investment under Brazil's four-year National AI Plan.

THREE ROUTES TO POOLED COMPUTE ACCESS



South-South cooperation

A clear pattern is emerging across the Global South. As reliance on compute grows, driven by high entry costs, dependence on foreign infrastructure, and geopolitical risks also increases. With AI becoming central to national economic strategies, this dependence raises not only access challenges but also concerns about contextual fit, long-term resilience, and economic autonomy. While some reliance is unavoidable due to current resource distribution and geopolitical factors, South-South cooperation can help countries maximise the value of their resources by pooling efforts, sharing implementation insights, and developing AI systems tailored to local contexts.

In the lead-up to the India AI Impact Summit 2026, held in New Delhi in February 2026, a Carnegie Endowment analysis argued for a more nimble, targeted model of South-South AI cooperation - anchoring collaboration in contextually fitted national initiatives, creating sustained cooperation "practice spaces," and pooling and aligning financing mechanisms for shared infrastructure and shared outcomes at population scale. Through these emerging patterns, a narrative of Prosocial AI takes place. The ethos points towards moving beyond the discourse of "ethical AI" and "AI for Good" towards an actionable, proactive framework for designing and deploying AI systems with measurable, positive outcomes for humanity. Hence, the Global South, to secure digital sovereignty and collective progress, prioritises shared investment in data trusts, compute infrastructure, and open-source foundational models. Building these systems collectively through South-South collaboration ensures the algorithmic architecture powering AI remains grounded in local realities rather than being externally owned. This strengthens self-reliance while enabling contextually relevant innovation aligned with regional needs and values.

Optimisation, Frugal AI and Edge AI

For most of AI's short history, the assumption has been that better models require more compute. DeepSeek-V3 broke that assumption. Released by the Chinese lab in December 2024, it matched the performance of far larger frontier models while being trained on a fraction of the hardware - roughly 2,000 GPUs- at a reported cost of around \$5.6 million, against the tens of thousands of chips and hundreds of millions of dollars typically associated with models of that class. It achieved this not by brute force but through architectural restraint: activating only a small subset of parameters for each query and training at lower numerical precision. The emphasis was less on DeepSeek itself and more on demonstrating that the compute gap facing the Global South isn't necessarily a performance gap. India took that lesson early. Sarvam AI and Krutrim were built from the outset around efficiency rather than scale for its own sake. The idea reaches its purest form in the Saving Voices Project, which built a speech-AI system for the Soliga, an Indigenous community in Karnataka whose language has no written script and whose villages have no internet access - and ran it entirely offline on a Raspberry Pi. There was no larger model to shrink down to size here; the hardware and the community's needs defined the model from the start, not the other way around.

The same instinct now runs through Indian public infrastructure. The migration was almost incidental compared to the platform's true test: months earlier, at Maha Kumbh 2025, Bhashini's multilingual assistant provided real-time translation and voice support in eleven Indian languages for tens of millions of pilgrims. This was arguably the largest live deployment of an Indian-language AI system to date. Officials noted that shifting to sovereign infrastructure improved performance and reduced costs compared to the hyperscaler setup, not despite going local but because of it. Additionally, since Bhashini's models and datasets are open-source, this infrastructure now supports work beyond the government - adopted by startups, researchers, and public projects that would otherwise have had to develop their own solutions.

DEEPSEEK-V3 (DECEMBER 2024): MATCHED FAR LARGER FRONTIER MODELS

~2,000 GPUs · ~\$5.6 million

Trained on a fraction of the hardware, at a reported cost — through architectural restraint: activating only a small subset of parameters for each query and training at lower numerical precision

Tens of thousands of chips
Hundreds of millions of dollars

Typically associated with models of that class



BHASHINI

CASE STUDY: BHASHINI

1. Introduction

Artificial intelligence is most often discussed through the lens of scale, measured by the size of models, the volume of computing power available, or the quantity of data used in training. For much of the Global South, the more consequential question is different. It is whether advanced artificial intelligence can work across linguistic diversity, uneven connectivity, limited digital literacy and the practical realities of delivering services at the last mile. India is a demanding environment in which to answer that question. It has hundreds of languages and dialects, significant variation in digital access and literacy, and a public service ecosystem operating at enormous scale. BHASHINI, India's National Language Translation Mission, has emerged from this context. Its approach is not confined to developing language models. It is the construction of a shared, open and reusable technology layer through which people, governments, institutions and innovators can interact with digital systems in the languages and modes natural to them.

This design philosophy places accessibility alongside capability. Multilingual models, voice-first interfaces, low-latency inference, model optimisation, shared computing infrastructure and selected offline and edge capabilities are developed not as isolated technical features but as the means of making artificial intelligence usable where bandwidth, computing resources, language data and digital literacy cannot be assumed to be abundant. A citizen should not have to become proficient in English, learn to type, interpret a complex interface or hold uninterrupted high-speed connectivity before reaching a public service. In a village, a government office or a telephone interaction, the interface may need to begin with speech, operate in a familiar language and respond with minimal delay.

The significance of this extends beyond India. Countries across the Global South face the same structural questions: how artificial intelligence can serve populations speaking many languages, how systems can operate when connectivity is intermittent, how limited computing resources can be used efficiently, and how technology can reach people who have remained outside formal digital systems. The emerging answer is not necessarily to replicate the infrastructure model of the most compute-rich economies. It can instead involve building differently, through open ecosystems, efficient models, multilingual architectures, voice-first interaction, edge deployment and infrastructure reusable across applications. The constraints themselves become design requirements and, in turn, drivers of innovation.

2. Language Technology as Digital Public Infrastructure

BHASHINI's operational footprint spans 36 Indian languages for text, 65 Indian languages for voice and 35 international languages, including the official languages of the United Nations, in addition to the 22 constitutionally scheduled languages with which the programme began. This capability is delivered through a catalogue of more than 350 deployed models, built through contributions from multiple organisations and research ecosystems, including foundations such as AI4Bharat. The significance lies not in the count of languages or models but in the creation of a shared language-technology layer reusable across government services, applications and the wider innovation ecosystem.

The underlying model families illustrate how architecture itself can be adapted to linguistic diversity. Translation is built on a Transformer encoder-decoder architecture, with the flagship Large variant at approximately 1.1 billion parameters. Because the architecture is multilingual, a common model works across multiple languages and scripts rather than requiring an independent model for each. This matters directly for low-resource languages: knowledge learned from languages with richer datasets helps the system form useful representations for languages with far less data, making broad coverage practical rather than prohibitive. Training draws on large parallel datasets such as the Bharat Parallel Corpus Collection, in which equivalent content exists across multiple languages, allowing a single investment in data and models to benefit many languages rather than remaining siloed within one.

Speech recognition uses Conformer-based acoustic models, including a Conformer-Large variant of approximately 120 million parameters operating on 16 kHz audio, with hybrid CTC-RNNT decoding. The Conformer architecture combines convolution and self-attention, allowing the model to attend both to the fine acoustic detail of individual sounds and to the wider context of a spoken sentence, a design objective oriented towards working across diverse Indian languages, accents and speaking styles rather than optimising only for languages with abundant speech data. Text-to-speech has evolved from acoustic-model approaches such as FastPitch and HiFi-GAN towards a codec-language-model architecture, in which the system generates a sequence of learned audio tokens from a natural-language description of the desired voice, subsequently reconstructed into a 44.1 kHz waveform. In practical terms this moves speech synthesis closer to the way modern language models generate text, enabling more natural and expressive output than earlier spectrogram-and-vocoder methods.

The common thread is that each architectural choice answers a specific constraint. Multilingual models address the scarcity of data. Efficient architectures and optimisation address the scarcity of computing. Speech, streaming and edge-oriented technologies address the scarcity of connectivity and the difficulty of the last mile.

This is where language technology begins to take on the characteristics of digital public infrastructure. As digital identity and payment infrastructure provide common capabilities on which many applications are built, a shared language layer provides translation, speech recognition and speech synthesis as reusable building blocks. A government department need not build its own language model, and a startup need not assemble an entire language stack from scratch. Both build on a common foundation.

3. Case Study: Optimisation for Scale

What allows the BHASHINI stack to serve at government scale while continuing to function in low-bandwidth and offline settings is not a single technique but a layered approach to inference optimisation. Inference optimisation means making a model perform the same work faster and with fewer computing resources. It is applied across four levels: the model itself, the hardware it runs on, the infrastructure that serves it, and the way requests are handled.

Compilation and hardware tuning.

Models developed in PyTorch are converted into deployment-ready formats such as ONNX, a common representation allowing a model to run efficiently across different environments, and then into hardware-optimised engines such as TensorRT. The effect is comparable to tuning a vehicle for a particular road: the model is prepared to make the fullest use of the specific processor on which it will run. Techniques such as operator fusion, which combines several small operations into one, and CUDA Graphs, which reduce repeated instructions issued to the processor, cut overhead and response time further.

Precision and quantisation.

A model can often be represented using smaller numbers without losing the information that matters, much as a file can be compressed. BHASHINI uses formats including FP16, BF16, FP8 and INT8 according to the workload and the accuracy that can be accepted and applies techniques such as Activation-aware Weight Quantisation to reduce the memory footprint of larger models. Critically, each optimisation is validated against a higher-precision reference model using measures such as Word Error Rate, which records how often a speech system misrecognises words, alongside punctuation and speech-quality measures. This validation step is what separates efficiency from degradation: compression is adopted only where the citizen-facing experience is preserved.

Serving efficiency.

These optimised models operate on shared infrastructure of approximately 200 H100 GPUs, with computing capacity pooled across models and applications rather than dedicated one processor to one model. Requests are handled through dynamic and continuous batching, processing several together rather than sequentially, with PagedAttention used to manage processor memory more effectively when many requests run concurrently. Low-Rank Adaptation allows a small specialist layer to be added to a common base model instead of maintaining a wholly separate model for every application. Repeated computation is avoided through a distributed cache: where an identical or sufficiently similar request has already been processed, the stored result is returned rather than running the full pipeline again, and prefix and key-value caching allow recurring portions of prompts and context to be reused. This is especially consequential in government applications, where interactions contain recurring terminology, service descriptions, application forms and frequently asked queries.

Constrained environments.

For real-time voice, speech is processed in small chunks as it arrives rather than waiting for a complete recording, using bidirectional streaming that allows audio and responses to flow continuously, including over 8 kHz narrowband telephone networks, where available bandwidth is a fraction of modern broadband. Where connectivity is limited or absent, selected models are quantised and compiled into compact edge runtimes for deployment closer to the point of use.

Taken together, these layers convert a finite pool of computing capacity into services operating at national scale. The shared infrastructure currently supports more than 22 million inferences per day. The significance of that figure lies in its denominator: scale of this order is achieved not by continually adding computing infrastructure, but by making every available unit of computing work harder while preserving the quality and responsiveness that citizen-facing services demand. For economies where computing capacity is scarce and expensive, this is the more replicable proposition.

4. BHASHINI's Art of Possibilities

The technical choices described above are not ends in themselves. Efficient inference, shared computing, caching, batching, quantisation, streaming and edge deployment are the mechanisms that make responsive language and voice capability affordable across a large and diverse population. Their clearest return is in public service delivery itself: each deployment below optimises a service that already existed, removing steps, intermediaries and literacy requirements that stood between a citizen and an entitlement. The BHASHINI ecosystem currently supports more than twenty such deployments across governance, agriculture, justice, health, transport, identity and heritage; seven are set out here

eGramSwaraj.

The Ministry of Panchayati Raj's flagship platform for Panchayat-level planning, accounting and monitoring has been integrated with BHASHINI to provide access across 22 scheduled Indian languages, reaching a user base of more than 270,000 Panchayati Raj Institutions across 28 States and 6 Union Territories. The relevant point is what did not happen: no state and no institution had to commission its own language stack. A shared layer resolved the problem once for all of them, which is the difference between an integration cost borne once and one borne several hundred thousand times.

SabhaSaar.

Developed with the Ministry of Panchayati Raj, SabhaSaar converts audio and video recordings of Gram Sabha and Panchayat meetings into structured minutes, transcripts and summaries, with translation support across 23 regional languages, now in use across more than 120,000 Gram Panchayats. Before this, a functionary recorded proceedings by hand, and the administrative burden of documentation competed directly with the work of governing. What changes on the ground is that spoken participation at the village level — the form in which grassroots democracy actually occurs, becomes a structured digital record without anyone having to write it down.

ASHA Telephony.

There is a difference between making artificial intelligence available and making it accessible. A sophisticated digital service remains out of reach if it presupposes a smartphone, an application and a degree of digital familiarity the intended user does not have. Developed in the context of the Department of Food and Public Distribution, ASHA Telephony uses voice and the ordinary telephone as the access point for engagement with beneficiaries of the Public Distribution System, serving around 2 million beneficiaries each month. The beneficiary downloads nothing. The interaction begins with a phone call and a voice, in a language they are comfortable speaking, which, for a service of this scale and linguistic range, is a simpler entry point than any application or website could offer.

UIDAI

Digital identity is a foundational layer through which citizens reach a wide range of services, and at a scale of more than 1.4 billion Aadhaar numbers issued, accessibility cannot be separated from linguistic diversity. Integrating BHASHINI's capabilities into UIDAI demonstrates language being built into an existing public service rather than offered as a separate application. The citizen does not leave the service to find a language solution, and the problem does not need solving again for every platform that depends on digital identity.

Rail Madad.

Indian Railways serves a scale few public services match, with approximately 7.41 billion passenger journeys in 2025-26 across more than 13,000 passenger trains. Speech recognition, translation, speech synthesis and optical character recognition support multilingual and voice-based interaction across railway platforms including Rail Madad, with support for 22 Indian languages. The passenger learns no new interface in order to seek assistance; the language capability operates in the background of a platform they already use.

BhashaDaan.

Every capability described in this paper rests on a prior condition: for a system to understand a language, that language must first exist meaningfully in digital form. For many Indian languages, and for a great many languages across the Global South, it does not. BhashaDaan addresses this at the root by enabling citizens to contribute spoken and written language data in their own languages through four modes, Bolo, Likho, Suno and Dekho, with contributions validated and quality-reviewed to build reliable datasets. Its relevance is greatest precisely where digital resources are thinnest. The Bhili initiative described above is what this looks like in a single district; BhashaDaan is the standing national mechanism for it. The consequence is a different relationship between a language community and an artificial intelligence system: the community is not only served by the model but supplies the data from which it learns, which is also the only durable answer to the problem of low-resource languages at scale. The common structure across these deployments is that a single technology foundation supports very different languages, modalities and service environments, while optimisation allows a shared pool of computing to serve all of them. The value of the infrastructure is measured not by the number of models or languages it supports, but by whether those capabilities can be carried into real environments and reused at scale.

5. Conclusion

The next phase of artificial intelligence innovation need not be defined only by larger models and greater computing capacity. It could be defined by the ability to make advanced technology work efficiently across languages, devices, connectivity conditions and levels of digital literacy. By building an open, reusable ecosystem, optimising a finite pool of computing, and carrying those capabilities into live public service environments, BHASHINI demonstrates infrastructure designed around the people it is intended to serve. The techniques that make this possible, efficient models, quantisation, shared inference, streaming and edge deployment, were developed under constraint, and are for that reason directly transferable to economies facing the same constraints. India's experience suggests that the Global South can contribute to the evolution of artificial intelligence not simply by adopting what is built elsewhere, but by building infrastructure shaped by its own scale, diversity and realities.

AI4Bharat closes the loop by working backwards from the same premise: instead of building the biggest model it can and hoping it adapts to Indian conditions, it starts from India's actual constraints - patchy connectivity, low-cost devices - and designs for those first.

Across all four examples, the pattern holds: efficiency was not a compromise the Global South settled for in the absence of compute. It became, on its own terms, a design philosophy.



11 languages

Bhashini's multilingual assistant provided real-time translation and voice support in eleven Indian languages for tens of millions of pilgrims at Maha Kumbh 2025 — arguably the largest live deployment of an Indian-language AI system to date

Edge AI

A typical assumption about frontier hardware is that it diffuses downward: chipmakers build for markets that can pay full price, and cheaper, lighter, optimised versions trickle down later. Edge and frugal AI have not followed that script in the Global South. In several of the cases already discussed, demand for AI that runs on minimal, often offline hardware showed up first - built out of necessity by people with no access to, and no interest in waiting for, a purpose-built chip. The Saving Voices Project didn't wait for an NPU-equipped device to become affordable; it built its Soliga-language model, trained on just five hours of voice data, around a low-cost, hobbyist-grade Raspberry Pi board, because that was what could actually reach a village with no internet connection.

Karnataka's KEO device makes the same point more starkly, because it dispenses with Western chip supply chains altogether. Launched in November 2025, the device, built by the state's electronics corporation, KEONICS, runs on an open-source RISC-V processor with its own on-device AI core, designed to run AI locally in a state where computer ownership is around 15%. Its distinguishing feature is BUDDH, an AI tutor trained on Karnataka's own school syllabus, built so that classroom AI keeps working even when the connection doesn't. What KEO solved was not "build a scaled-down version of a Western product" but a specific local access problem, using an open-source architecture designed for that problem from the start.



5 hours

of voice data used to train the Saving Voices Project's Soliga-language model, run entirely offline on a low-cost, hobbyist-grade Raspberry Pi board



~15%

computer ownership in Karnataka — the state whose electronics corporation, KEONICS, built the KEO device to run AI locally



RISC-V

open-source processor with its own on-device AI core powering KEO, launched November 2025, with BUDDH, an AI tutor trained on Karnataka's own school syllabus

Global chipmakers have since begun building products that formalise this same use case. Intel's Core Ultra line now ships with a dedicated Neural Processing Unit built for on-device inference. Microsoft's "Copilot+ PC" certification - requiring at least 40 trillion operations per second from that NPU - effectively codifies offline AI capability as a mainstream hardware tier rather than a niche one, a category that didn't exist three years ago.

Notably, Intel's own deployments have followed the Global South's use case rather than led it: in Guatemala, it helped build an offline large language model for students with no stable internet access, a project one of its own engineers linked directly to his childhood in rural southern India; and in India, Intel has been developing hardware-virtualisation approaches that let one set of AI-capable machines serve many students in a single government school.

A guiding principle in these efforts is that affordability and inconsistent connectivity are considered initial constraints to work within, rather than issues to fix solely by increasing infrastructure.



The next chapter of AI will be defined by our ability to bring intelligence closer to where people live, work, and learn. Edge AI can help the Global South overcome constraints of connectivity and infrastructure by enabling more localized, efficient, and accessible AI solutions.

By processing data closer to where it is generated, Edge AI can also enhance security and privacy while making more efficient use of computing, energy, and network resources. This approach helps extend the benefits of AI to more people and communities, including those in underserved or resource-constrained environments. But technology must be matched with skills and ecosystems that empower countries to build and adapt AI for their own needs.

At Intel, we believe AI sovereignty starts with sovereign skills - enabling nations to move from being consumers of AI to becoming creators of solutions grounded in local realities



**Sarah Kemp, Vice President,
International Government
Affairs, Intel Corporation**





LEARNING AT THE EDGE: THE 5G DIGITAL SCHOOL LIBRARY AND PENANG'S ON-PREMISE MODEL FOR DIGITAL EDUCATION POWERED BY INTEL.

The growth of cloud and edge computing, and the parallel expansion of artificial intelligence (AI) and fifth generation (5G) mobile networks, has changed how education can be delivered. In the Asia-Pacific and Japan region, and in Malaysia and Indonesia in particular, a young demographic profile and a receptive economy have made digitally enabled education a route to innovation rather than a supplement to it. The obstacles to that shift, however, are consistent across the region and beyond: limited internet infrastructure coverage, inadequate device access, and a skill gap among educators.

Overcoming those challenges required a digital solution that went beyond the usual server-based digital libraries and broadband connectivity. For the 5G Digital School Library project in Penang, Intel helped its collaborators and the state government create an on-premise edge solution resting on three pillars: the latest 5G and cloud technologies, ease of access, and simplicity of use and management.

The constraint: the limits of cloud-only delivery

Most education technology platforms treat a cloud repository as the central hub from which learning material is distributed over the internet to end-user devices, student devices, teacher laptops and interactive smart boards. The model assumes conditions that do not hold uniformly. Bandwidth is the first assumption: not all locations have access to high-speed internet, which limits content distribution. Network reliability is the second: the possibility of network outages can disrupt access to learning materials. These remain common barriers to delivering digital education in remote and underserved areas

A hybrid architecture built around the Content Access Server

The solution adopted a hybrid platform approach with three design characteristics. Content storage flexibility allows learning resources to be held both at the edge and in the cloud, with the placement decided by content sensitivity and the need for high availability. Pre-caching at the edge ensures material remains available during periods of limited or no connectivity. Simplified content management keeps the platform accessible to users regardless of technical expertise and allows pre-cached content to be accessed both locally and remotely.

The architecture centres on a Content Access Server (CAS), a compact appliance that combines Network Attached Storage (NAS) and file management, a router with an access point creating the local wired and wireless network, a Digital Living Network Alliance (DLNA) and Universal Plug and Play (UPnP) media server, and an open-source e-book management system. The CAS maintains a 5G connection to online educational platforms and internet-based resources while functioning as the local hub through which students and teachers reach pre-cached material. Student terminals are laptops running a web-based e-book reader, DLNA and UPnP client players, and links to external sources such as the Penang Digital Library.

The practical consequence is that, by harnessing 5G, remote and underserved areas can bypass the limitations of traditional fibre optic infrastructure and access educational content from the Content Access Server even without an internet connection. The devices in the setup are computers equipped with high-performance processors, so a range of educational applications can run on the terminal itself rather than depending on a remote service.

A student-centric approach and collaborative ecosystem

Designed for a younger audience, the project brought together partners with expertise in user-centric technology and education. Penang Science Cluster, the implementer, selected Intel Malaysia as technical consultant, while 88 Captains, a non-profit organisation and N50 partner focused on student development, brought together SNS Network Technology, Yes by YTL Communications and OCBC Bank. In partnership with the Penang State Education Department (Jabatan Pendidikan Negeri Pulau Pinang), 88 Captains identified 88 pioneer schools across Penang.

Each school received a start-up kit comprising an enterprise-grade 5G modem, a Content Access Server capable of storing thousands of e-books, and three laptops. Together, these enabled students and educators to access their school's digital library beyond the traditional classroom. Intel Malaysia designed the end-to-end digital library architecture. SNS Network Technology supplied 264 Joi laptops, while Yes by YTL Communications provided connectivity and 88 enterprise-grade 5G modems sponsored by OCBC Bank. Python Resources supported the digital library platform through its e-book storage and access system, using Intel processors, cloud services, fault-tolerant Network Attached Storage (NAS) and an open-source e-book management system.

The Intel processors were selected for their compatibility with the platform and their ability to deliver secure, stable, and reliable performance, which the case study ties directly to an uninterrupted, streamlined learning experience in a collaborative environment. Intel vPro and Intel Core series processors bring purpose-built integrated accelerators and advanced security technologies designed to support demanding and fast-growing workloads in educational settings. Intel Active Management Technology (Intel AMT), a feature of Intel vPro technology, allows users and administrators to better discover, heal, and protect networked computing assets, a capability with direct consequences across an estate of 88 dispersed school sites.

Reported outcomes

Of the 88 schools surveyed, all reported positive experiences and indicated they would recommend the 5G service to other schools, with speed and support rated highly. To date, the '5G Digital School Library Program' successfully enabled ~50% of Penang's rural schools with internet access. In doing so, it has enabled almost 200,000 K-12 students to access updated knowledge and expand their horizons. Above all, this solution isn't just about providing students with a digital mode of learning; instead, it is more about bringing them on par with the rest of the world in terms of digital literacy, digital accessibility and opportunities for upskilling and personal development.

The CEO of Penang Science Cluster, Peng Ee Ooi, characterised the deployment as an infrastructure milestone for Penang's schools, noting the response times and availability achieved through the hybrid model. Building on the results, similar libraries were planned for students in other Malaysian states, and 88 Captains was invited to become a partner organisation of N50.

The programme also fed back into infrastructure policy. In collaboration with Digital Nasional Berhad, an undertaking of the Ministry of Finance, the Penang State Government committed to 80 per cent 5G network coverage across the state by the end of 2023, the target recorded at the time of the case study. These steps sit within the Malaysian government's wider objective of upgrading digital infrastructure and developing domestic digital talent.



Conclusion

This project transfers lessons beyond Penang. Firstly, the design answered the constraint directly. Placing storage and delivery at the edge, and using 5G, allowed a school to be equipped without waiting for fibre optic infrastructure, which is the condition much of the Global South is working within.

Additionally, the deployment unit was defined and repeatable. One server, one modem, and three laptops per school is a specification that can be costed, procured, and replicated without redesign, allowing a pilot in 88 schools to serve as a template for other states.

And finally, the consortium structure distributed a cost no single participant would have carried alone: the state education department supplied site selection and institutional legitimacy, a non-profit convened the partners, technology firms contributed architecture and hardware, and a bank funded the connectivity equipment. Beyond the numbers, the collaboration reflects a shared long-term vision for education, and stands as an example of how solution providers, technology manufacturers and local organisations can work together to bridge the digital divide in education. The result is access and infrastructure, the precondition on which learning gains are built.

This approach is significant. Edge and frugal AI originated in the Global South as necessity-driven engineering, not as scaled-down versions of technology from more resource-rich regions. The broader industry has since formalised and commercialised these proven solutions.

This logic connects to a longer-running idea: appropriate technology, originally developed to meet the needs of underserved communities with cost-effective, environmentally sustainable solutions, is now being reshaped by advances in AI and robotics. In recent years, "Frugal AI" has gained growing recognition in international policy and development circles - recent research, for instance, evaluates grassroots frugal-AI and TinyML projects directly against the principles in UNESCO's 2021 Recommendation on the Ethics of Artificial Intelligence, including human-centred design, sustainability and capability-building. In this context, frugal AI aims to maximise societal benefit within real resource constraints, whether computational, financial, or educational. This paradigm is increasingly important for developing countries, where rapid AI adoption could deepen inequality without inclusive capacity-building and local innovation.



02

Data



INTRO

As AI applications continue to expand, greater attention needs to be paid to how these systems are trained and the quality and relevance of the data they rely on. As generative AI moves into areas such as healthcare, agriculture, education, governance and financial services, access to diverse, representative and high-quality datasets is becoming increasingly important. Globally, frontier AI developers are also confronting the limits of available training data, driving interest in solutions such as synthetic data and advanced data curation. While much of this debate has focused on the needs of the Global North, the challenge for the Global South is also about ensuring that AI reflects its diverse languages, cultures, social contexts and local realities, which remain underrepresented in existing datasets.

For the Global South, building useful AI therefore begins with building the right datasets, including sector-specific and locally relevant datasets that respond to the problems communities actually face. This is creating space for startups and research initiatives that are developing datasets from the ground up and using them to build AI solutions tailored to sectors such as agriculture, healthcare, education and public services. The chapter highlights these emerging case studies and examines how startups across the Global South are using locally relevant data to develop sector-specific AI solutions that address real-world needs and deliver practical benefits to communities.

As the various use cases for Generative AI expand, and foundation models become more widely deployed, the availability and quality of training data are becoming increasingly important determinants of model performance. The first phase of the AI race was shaped largely by access to capital, compute, and advanced chips. Attention is now also shifting towards the availability of high-quality, human-generated data. Research by Epoch AI estimates the effective stock of quality- and repetition-adjusted public human-generated text at around 300 trillion tokens, which, under current scaling trends, could be substantially utilised between 2026 and 2032. While this does not imply that AI development will literally exhaust the world's data, it indicates that access to additional high-quality human data is becoming a potential constraint on the further scaling of language models.

This constraint is not limited to the volume of available data. Frontier models are trained predominantly on publicly available digital content, and the distribution of that content is highly uneven across languages and geographies. English remains disproportionately represented online, while thousands of languages spoken across Africa, Asia and Latin America have much smaller digital footprints. Meta's Llama 3.1, for example, was trained on more than 15 trillion tokens, but its published data mix allocated approximately 8% of tokens to multilingual content, with the large majority devoted to general knowledge, mathematics, reasoning and code. The model subsequently supports only a small subset of the world's languages as officially supported languages.

This creates what can be described as a 'resourcedness gap' in AI. High-resource languages benefit from large volumes of text, speech and labelled datasets; established benchmarks; digital dictionaries; and other language technologies. Low-resource languages often face gaps in both availability and quality. Data may be limited in volume, poorly labelled or concentrated in a narrow set of sources. In addition, many languages have limited representation in scientific, educational and technical content, reducing the extent to which models can acquire domain-specific knowledge. These limitations affect not only translation, but also speech recognition, information retrieval, reasoning and the ability of models to interpret local expressions and cultural context. Recent research on African NLP (Natural Language Processing) estimates that 88% of African languages are severely underrepresented, or effectively absent, from computational linguistics. The African Languages Lab demonstrates the scale of the problem: a quality-controlled dataset spanning 40 African languages was able to generate substantial gains from relatively modest models when they were trained on targeted language data.



In a multilingual democracy, people should not have to leave their mother tongue to enter public life. In Nandurbar, with the help of our incredible partners, we were able to show what community-centered AI development can look like. We are now excited to scale this approach globally.

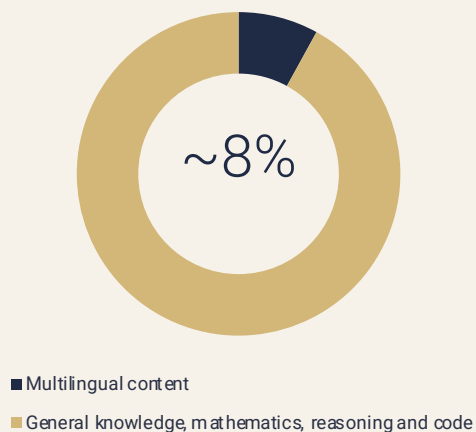
AI is the single most important technology of our lifetime, and today, for 3 billion people, one-third of humanity, AI doesn't understand their mother tongue. Nandurbar shows us the path to change this status quo."



**Manu Chopra, Co-founder and
CEO, Karya**



Meta's Llama 3.1 published data mix: share of training tokens



300 trillion

Tokens - Epoch AI's estimate of the effective stock of public human-generated text, which could be substantially utilised between 2026 & 2032

15 trillion+

tokens Llama 3.1 was trained on, yet it officially supports only a small subset of the world's languages

88%

of African languages are severely underrepresented, or effectively absent, from computational linguistics

40

African languages in the African Languages Lab's quality-controlled dataset, generating substantial gains from relatively modest models



CASE STUDY:

KARYA - BUILDING BHILI LANGUAGE INFRASTRUCTURE IN NANDURBAR

Bhili is spoken by more than 10 million people across parts of western and central India, including Nandurbar in Maharashtra, where nearly 69% of the population belongs to tribal communities. Yet, despite its large speaker base, Bhili had almost no representation in major AI systems. This meant that communities speaking Bhili were largely excluded from AI-enabled services such as translation, speech recognition and access to digital information. The underlying problem was not a lack of speakers, but a lack of usable digital language data needed to train and deploy AI systems.

Karya, a nonprofit that builds AI datasets and models with the communities that speak the language, partnered with the Nandurbar district administration to address this gap. More than 400 native Bhili speakers participated in creating the dataset, recording, translating, reviewing and validating the language data themselves. This community-led process helped ensure that the dataset captured how Bhili is actually spoken and used, rather than treating the community simply as a source of raw data. The approach also demonstrated how languages with limited existing digital resources can be brought into AI systems through targeted, locally grounded data creation.

The resulting Bhili model was integrated into MahaVISTAAR AI, enabling agricultural information to be delivered in Bhili for the first time. The datasets were also made available through AIKosh, India's national AI data platform, which describes it as the country's first tribal-language AI model. On 31 May 2026, Bhashini used the model to livestream the Prime Minister's monthly national address in Bhili, marking the first time the language had entered a national public communication channel at this scale. What began as a community-level effort to create missing language data therefore became part of wider public digital infrastructure.

The Bhili case demonstrates that the central constraint in building AI for underserved languages is often not the absence of technology, but the absence of the data needed to make that technology work. Instead of waiting for general-purpose AI systems to eventually develop adequate Bhili capabilities, the project created the missing data with the community itself and connected it to existing public digital infrastructure. This offers a different approach to AI development in the Global South: build for the context first, rather than expecting a general-purpose system to eventually become relevant to it.

The scale of the challenge extends far beyond Bhili. Karya currently builds datasets across 115 languages, 87 of which started with no usable data of any kind. Across the Global South, more than 200 languages remain barely supported by AI despite being spoken by more than 3 billion people, roughly a third of humanity. The systems being trained between now and 2028 will help determine which languages are represented in the next generation of AI and which remain digitally marginalised. Bhili shows that where conventional AI development sees a data constraint, community participation, targeted data creation and public digital infrastructure can turn that constraint into the foundation for locally relevant AI.

The implications of this imbalance extend beyond language. Most widely used AI benchmarks are designed to measure general capabilities such as reasoning, factual knowledge, mathematics and question answering. These benchmarks are useful for comparing models, but they do not necessarily capture whether an AI system performs effectively within a particular country's institutional, linguistic or socio-economic context. A healthcare model, for example, may perform well on a global benchmark while having limited understanding of local clinical terminology or treatment guidelines. Similarly, a general-purpose model may generate fluent Hindi while still performing poorly on regional dialects, code-mixed language or agriculture-related queries. The challenge for the Global South is therefore increasingly one of relevance and representation: whether the underlying data reflects the populations, languages and sectors in which these systems are expected to operate.

India provides an early example of a coordinated response to this challenge. Rather than relying exclusively on datasets already available on the internet, India is investing in mechanisms to collect, organise and make domestic data available for AI development. AIKosh, the national repository for AI datasets and models under the IndiaAI Mission, brought together 15,799+ datasets and 352 AI models across 22 sectors by September 2026. The platform brings together data from government and non-government sources and is intended to make datasets easier to discover and reuse across applications. Its significance lies not only in the number of datasets, but in the creation of a common layer through which otherwise fragmented information can be made more accessible to developers and researchers.



We believe AI has to be truly inclusive. As our Hon'ble Prime Minister has emphasised, AI should benefit every section of society, including those at the bottom of the pyramid. People should not have to come to AI; AI should go to them. That is why we believe AI needs to be multimodal, multilingual and multichannel, so that people across languages, devices and levels of digital access can benefit from it.

At CoRover, we believe in building hundreds of small, purpose-built models for specific domains, specific purposes and enterprise use cases. These models can be more precise for the information and context they are designed to handle, while enabling lower costs, lower energy consumption, reduced hallucination and better accuracy. One example is PortGPT, which we are building for the port ecosystem. It is designed as a single intelligent operating layer for a port authority combining conversational AI, a live operations command centre, and a report and drafting engine, all built on a common knowledge base and accessible across devices.

The objective is to replace hours of manual information searching and manually compiled reports with instant, source-cited answers, helping transform scattered port data into actionable insights and, ultimately, better decisions.



**Ankush Sabharwal, Founder
and CEO, CoRover.ai**



India is pursuing a similarly deliberate approach to language data. Project Vaani, led by IISc and ARTPARK, is building large-scale speech resources across Indian languages and dialects by collecting data directly from speakers. This approach addresses a gap that cannot be solved through conventional web scraping alone, particularly for languages and dialects with limited digital content. AI4Bharat has followed a complementary approach through IndicLLMSuite, which provides resources across all 22 scheduled Indian languages and comprises 251 billion tokens and 74.8 million instruction-response pairs. The project combines manually verified data with other web, document and video sources, alongside translation, transliteration and synthetic data, and establishes a pipeline for crawling, cleaning, filtering and deduplicating multilingual content.

AIKOSH: NATIONAL REPOSITORY FOR AI DATASETS AND MODELS, FEB 2026

15,799+

datasets

352

AI models

22

sectors

INDICLLMSUITE — AI4BHARAT

22

scheduled Indian languages

251 billion

tokens

74.8 million

instruction-response pairs



This is also changing the role of synthetic data. Frontier developers are increasingly using synthetic examples to supplement human-generated data, particularly for post-training and specialised capabilities. Meta, for example, reports using synthetic data as part of the iterative post-training of Llama 3.1. For low-resource languages and domains, synthetic generation, machine translation and back-translation can similarly help expand limited datasets. However, these methods are most effective when anchored in high-quality local data. Synthetic data can increase the volume of available examples, but it cannot by itself address the absence of local knowledge, linguistic variation or cultural context.

The emerging direction across the Global South is therefore not simply to accumulate larger datasets, but to improve the representation and usability of the data that underpins AI. This includes bringing together fragmented national datasets, creating digital resources for underrepresented languages, involving communities in data collection, improving the quality and structure of existing information, and developing specialised datasets for sectors where local context is critical. India's experience with AIKosh, Project Vaani and IndicLLMSuite, alongside Africa's growing investment in low-resource language datasets, demonstrates that countries can build AI capability by investing in the data layer even where they do not possess the scale of compute or capital available to frontier AI developers. For the Global South, this makes data infrastructure an important component of AI capability in its own right. The strategic objective is not necessarily to reproduce the largest general-purpose models, but to build the datasets, language resources, benchmarks and sectoral knowledge bases required for AI systems to perform reliably in local environments. The countries that are able to convert fragmented information into high-quality, representative and reusable training data will be better positioned to develop AI systems that reflect their own languages, institutions and economic priorities.





The next chapter of AI development should not be defined simply by building ever larger general-purpose models. Today's large general-purpose AI models are a remarkable achievement, and open weights are steadily turning them into something like infrastructure that is increasingly available to everyone. However, recreating and serving them at national scale would drain resources that a large developing country should be spending elsewhere, particularly if their main impact is to improve the efficiency of tasks that people are already capable of performing. For a country with our demographic profile, speed of execution is the wrong yardstick. What matters is the energy required to complete an end task, and whether a deployment leaves more people in purposeful work than it found. This calls for a purposefully developed, frugal and holistic AI that strengthens a workforce rather than simply displacing it.

Two decades of digitisation have left mountains of data locked inside individual institutions: railways, land records, courts, GST, and healthcare. Little of it is the plain text that language models were built on. It is a dense mixture of structured and unstructured records spanning registers, sketches, event logs, and ledgers that no global model reads. Its value never sits in one document waiting to be retrieved. It emerges when patterns are stitched across linked collections, over time, space, and modality.

Extracting goal-driven intelligence from such datasets calls for model designs that could be very different from the large language models that largely define modern AI. Sequence models over national railway event logs joined with natural-language customer feedback could provide better capacity planning. Graph networks over transactional movements could learn what ordinary commerce looks like, so that unusual patterns can be preferentially processed for alerts. Models over court dockets could predict when a case will be ready. Parcel-level models could tell an individual what their land risk actually is. The resources required to train such models are insignificant compared with training an LLM. They can be trained where the data already sits, with the constraints and knowledge of the institution built into the system. Success should be measured not by the size of the model, but by the capability created per unit of compute, energy, and human effort.

These systems will not merely speed up existing work; they will create capabilities that did not exist before—for example, helping an officer make allocation and ranking decisions under real budget constraints, or giving citizens honest projections about the services they depend on. This is the kind of AI that can make a developing country more capable without requiring it to reproduce the enormous computational infrastructure of the most advanced economies.



**Professor Sunita Sarawagi,
Computer Science and
Engineering, IIT-Bombay**



Building AI around local challenges

The second emerging priority is the development of sector-specific datasets and models. For many applications in the Global South, the value of data lies less in its scale than in its relevance to a particular domain. Agriculture is one example. Data on crops, soil, weather, pests, land, markets and government schemes is highly specific to local conditions, and generic internet-scale datasets provide limited coverage of this information. BharatGen's AgriParam illustrates how this can translate into a specialised model. The team assembled an India-centric agricultural corpus from 17,000 open-source articles and knowledge passages, built a taxonomy covering crops, weather, soil, pests, fertilisers, markets and government schemes, and generated approximately 2 million question-answer pairs and 6 million multi-turn conversations. The resulting model is designed around agricultural use cases and Indian users rather than general language generation alone.

A comparable example is also visible in Africa. PlantVillage, which is developed in partnership with Penn State University and African research partners, released Nuru. The application was built on a database of more than 1 lakh annotated images of crops such as cassava, along with maize and potato diseases, which are region-specific to East Africa, serving the needs of over 5,000 to 15,000 smallholder farmers. The application works offline and delivers guidance in the local language, Swahili, as well as English. Similar to the ethos of AgriParam, the underlying issue was that generic internet-scale image datasets are dominated by commercial agriculture in temperate regions and inadequately capture crop varieties, disease presentations or growing conditions relevant to smallholder farming in the Global South. Hence, the value lies in curating region-specific datasets rather than large, generic ones.

AGRIPARAM: BHARATGEN'S INDIA-CENTRIC AGRICULTURAL CORPUS

17,000	~2 million	6 million
open-source articles and knowledge passages	question-answer pairs	multi-turn conversations

NURU: PLANTVILLAGE, PENN STATE UNIVERSITY AND AFRICAN RESEARCH PARTNERS

1 lakh+	5,000+	Offline
annotated images of cassava, maize and potato diseases specific to East Africa	smallholder farmers served	with guidance in Swahili as well as English

The same logic applies to other sectors. In healthcare, for instance, the value of the model depends on access to locally relevant clinical data, terminology and guidelines. In the case of law, it depends on access to domestic statutes, case law, and legal procedures; in public administration, it depends on local governance structures and the processes through which citizens interact with the state. While these datasets might be small in terms of corpus size, their value is greater because they contain information that is often poorly represented in general-purpose training data.

CASE STUDY: COROVER.AI

Ports generate large volumes of operational information every day, covering vessels, cargo, berths, traffic, tenders, revenue and environmental conditions. Much of this information sits across documents, spreadsheets, databases and external portals, making it difficult to access quickly and use for decision-making. V.O. Chidambaranar Port Authority (VOCPA), Thoothukudi, addressed this challenge through **PortGPT**, an AI-powered platform developed with CoRover.ai and powered by BharatGPT. The platform brings these different sources together into a single interface through which port officials can access information, monitor operations and prepare reports.

The primary challenge was making this information usable without requiring officials to search through multiple systems or work directly with databases. PortGPT addresses this through a **conversational AI workspace** that allows users to ask questions in plain language, through text or voice, and receive structured answers with citations to the underlying documents and records. Queries can cover routine operational requirements, such as current tenders, annual traffic trends or vessels waiting at anchorage. This makes the port's existing information accessible through a simpler interface, while retaining a link to the source used to generate the response.

The platform also connects AI with live port operations. Its **Live Operations Dashboard** provides a consolidated view of vessels at anchorage, expected arrivals, berth occupancy, solar generation and air quality. It contains 31 live KPI panels covering port operations, environmental indicators, container traffic, cargo and revenue, and vessel and tonnage data. PortGPT also includes a **Report and Draft Engine**, which can generate 12 types of official reports using live port data and assist with drafting tenders and circulars across six departments. This brings information retrieval, operational monitoring and routine documentation into the same system.

A key part of the implementation is the **Knowledge and Data Fabric** built around VOCPA's own information. More than 562 PDFs, documents and presentations have been indexed, while 20 external domains, including the port authority, the Ministry of Ports, Shipping and Waterways and other ports, are tracked through a web crawler. Historical datasets have also been structured for analytical queries, with 40 years of records made searchable. The system is supported by an India-hosted AI and data architecture with role-based access controls and encrypted data transmission, allowing the platform to be used within a defined security framework.

PortGPT demonstrates how generative AI can be adapted to the specific information and workflows of a public-sector institution. Rather than introducing AI as a standalone chatbot, the deployment connects the technology to existing documents, historical records, live operational data and administrative processes. At VOCPA, this creates a common interface for accessing institutional knowledge, monitoring port activity and preparing routine reports, offering a practical model for applying sovereign and domain-specific AI to other public infrastructure and government institutions.



CASE STUDY:

Digitising Paper to Data by Prof Sunita Sarawagi, IIT-Bombay

The creation of high-quality datasets remains one of the most significant challenges in developing economies. Physical land records hold immense economic value, but much of this information remains locked away in paper archives. Prof. Sunita Sarawagi's research addresses this challenge by exploring how high-quality datasets can be created from fragmented physical records.

*The primary challenge in digitising these land records is that they do not follow neat or predictable templates. A single land archive may contain documents with multiple, uneven layouts. When scanned, conventional **Optical Character Recognition (OCR) software** can turn the text into a garbled string of words, with missing punctuation and misread characters. This is partly because traditional systems often rely on rigid rules to identify specific fields such as the owner's name or plot number.*

*To address this challenge, Prof. Sarawagi approached document digitisation not simply as a scanning exercise, but as a problem of context and probability. In her research on **Information Extraction and Conditional Random Fields (CRFs)**, she developed statistical machine learning models that interpret documents in a way that more closely resembles how humans read them. Rather than examining words in isolation or relying on a fixed document layout, these models use the words and information surrounding a piece of text to determine its meaning and likely role within the document.*

*A key innovation in this work was the **use of Semi-Markov CRFs**, which allow computers to analyse complete phrases or sequences of words rather than treating each word as a separate unit. This is particularly useful for land records, where addresses, owner names and property descriptions can span multiple words. The model can therefore identify a longer block of text as a single property description, even when the document contains variations in formatting or errors introduced during scanning.*

*Prof. Sarawagi also recognised that land records can vary significantly across regions, making it difficult to develop a single model that works equally well for every archive. To address this, her research **incorporated active learning techniques, allowing the system to improve with limited human intervention**. When the algorithm encounters an unfamiliar document layout or an ambiguous property code, it can flag that specific section for review by a human expert. Once the expert provides a correction, the system can use that information to recognise similar patterns across the archive, reducing the amount of manual intervention required as digitisation progresses.*

The broader significance of this approach is that digitisation does not have to mean simply converting paper into digital images. The real value lies in converting fragmented and unstructured records into reliable, usable data. For developing economies, where valuable public records are often spread across decades of physical archives, such methods can help turn existing information assets into datasets that can support better administration, research and public services.

This is clearly evident in healthcare and legal sectors. In dermatology, it highlights how biased outcomes can result from skewed datasets. Large skin image collections such as ISIC and Fitzpatrick17k primarily consist of lighter skin tones. Analysis reveals an imbalance ratio of 7.57 between light and dark skin representation. Additionally, AI-generated dermatological training images tend to carry this bias. This pattern appears beyond dermatology as well: pulse oximeters, primarily calibrated on lighter skin, have been shown in multiple studies to overestimate blood oxygen in darker-skinned patients. One significant retrospective study found that Black patients were nearly three times more likely than white patients to have hypoxemia undetected by these devices.

A parallel effort is underway to bridge this gap with more representative data. India-specific initiatives follow a similar approach: DermaCon-IN, a clinical dataset with over 5,400 expert-annotated images from Indian patients, focuses on infectious, inflammatory, and pigmentary conditions prevalent in South Asian populations. These images are labelled using both the Fitzpatrick and Monk scales, thereby addressing the underrepresentation of darker skin tones and region-specific disease patterns in global datasets that are predominantly based on European and North American patients.

REPRESENTATION IN CLINICAL DATA



7.57

imbalance ratio between light and dark skin representation in large skin image collections such as ISIC and Fitzpatrick17k



Nearly 3x

more likely: Black patients having hypoxemia undetected by pulse oximeters than white patients, in one significant retrospective study



5,400+

expert-annotated images from Indian patients in DermaCon-IN, focused on infectious, inflammatory and pigmentary conditions prevalent in South Asian populations



Fitzpatrick + Monk

the two scales used to label DermaCon-IN images, addressing the underrepresentation of darker skin tones and region-specific disease patterns

India's judiciary highlights a similar dynamic. Adalat AI, a legal-technology non-profit startup, has built a courtroom transcription tool that is now deployed in more than 6,000 courtrooms across 11 states. As per reports, the tool has helped transcribe roughly 1.2 million minutes of proceedings, saving an estimated 60,000 hours of court time. In 2025, the Kerala High Court became the first in India to mandate use of the tool for recording witness depositions, while embedding safeguards requiring judicial officers to verify all AI-generated transcripts.

5,000+

courtrooms across 10
states where Adalat
AI's transcription tool
is deployed

1.2 million

minutes of
proceedings
transcribed, as per
reports

60,000

hours of court time
saved, estimated

2025

Kerala High Court
became the first in
India to mandate the
tool for witness
depositions, with
verification safeguards

In fields like dermatology, agriculture, and the judiciary, the common pattern is that general-use frontier models, trained on globally dominant data, often underperform in applications where success hinges on local skin tones, crops, languages, or institutional procedures. In India, Africa, and other regions, the prevailing response is not to rely on cutting-edge developers to address these deficiencies but to create targeted datasets - though modest in size - that enable AI systems to operate reliably within their specific populations and institutions.



03

Skilling

INTRO

AI development depends not only on computing infrastructure and data, but also on a workforce capable of building, deploying and managing increasingly complex AI systems. The Global South has historically built strong IT and digital talent pools, providing an important foundation for the next phase of AI development. However, the rapid emergence of generative AI, automation and AI-enabled systems is changing the skills required, while many countries continue to operate under constraints around funding, compute, infrastructure and access to advanced training.

The challenge, therefore, is to build an AI workforce that is globally competitive while remaining relevant to local realities. This means developing engineers, researchers and technical professionals who can work with advanced AI systems but also understand the resource constraints, languages, sectors and problems that shape deployment in the Global South. The chapter explores how countries and institutions are approaching this challenge through new models of AI education, industry-led training, university partnerships and practical skilling initiatives, with a focus on building talent that can both compete globally and develop AI for local needs.

The existing approach to AI Capability treats human capital as a downstream variable. Compute is procured, models are trained, and talent is assumed to follow, whether recruited, imported, or created. However, this approach only works for countries that can afford to compete for scarce talent. For most of the Global South, the problem was not a lack of ambition, but a lack of advanced computing, chip manufacturing, strong research universities, and the funding needed to build them.

The distributional evidence is clear. UNCTAD projects the global AI market to reach USD 4.8 trillion by 2033. But this growth is concentrated in about 100 firms, mainly in the United States and China, which make up about 40 per cent of global corporate AI research and development spending, while 118 countries, mostly in the Global South, could remain absent or largely peripheral in major global AI developments. Fewer than one-third of developing countries have a national AI strategy; among least developed countries, the figure is around 12 per cent, with roughly 65 per cent of their populations remaining offline.

Multilateral and development forums have sought mechanisms to make the AI race more inclusive; but projections continue to indicate that the economic gains from artificial intelligence will be heavily concentrated in advanced economies, with developing countries at risk of being left behind. So the question arises: what can the Global South do to ensure the benefits of AI reach its entire population?

THE DISTRIBUTIONAL EVIDENCE (UNCTAD)

USD 4.8 trillion

projected size of the global AI market by 2033

~100 firms

mainly in the United States and China, concentrate this growth

~40%

of global corporate AI research and development spending is made up by those firms

118 countries

mostly in the Global South, could remain absent or largely peripheral in major global AI developments

< one-third

of developing countries have a national AI strategy

~12%

of least developed countries have a national AI strategy

~65%

of least developed countries' populations remain offline

What followed from these constraints was not a smaller version of the Northern model. It was a different architecture. With limited scope for buying capability or infrastructure, a set of middle-income economies chose to manufacture it at population scale, through institutions that already existed, and at the point where AI is used rather than where it is invented. This made the application layer become the means to empower a generation of people, not just catching up in the AI race but creating cutting-edge solutions for their own communities in their own languages.

This idea fundamentally reshapes the “AI race”, which is expected to run on a single track with a common finish line. However, what the ‘Global South’ is now forging is an alternate track oriented toward diffusion rather than frontier capability: the depth and speed at which AI competence reaches a district administrator, small manufacturer, and the median worker.

For example, Brazil’s “AIED Unplugged” framework, supported by UNESCO, demonstrates an offline-first approach to AI in education. It is being implemented in public schools to deliver AI-enabled literacy without core infrastructure, reaching hundreds of thousands of students. This shows how low-connectivity, locally grounded AI solutions can build capacity and drive large-scale impact. In higher education and professional development, industrial and vocational training mechanisms function as key units for AI capacity and deployment.

However, it bears mentioning that the very instruments and ‘global’ standards used to measure AI capability are largely Northern artefacts. Conventional measures of AI capability can understate activity in the Global South. Indicators based on professional networks, conferences and publications often favour researchers and institutions with greater access to funding, international travel and digital platforms. The United Nations Educational, Scientific and Cultural Organisation (UNESCO) also shows that many countries have limited formal AI programmes, and much activity remains informal or locally led. Global indices should therefore be complemented by bottom-up, country-level mapping to capture AI activity that standard measures may miss.

RESHAPING THE “AI RACE”



A single track

Expected to run with a common finish line, oriented toward frontier capability.



An alternate track

Oriented toward diffusion: the depth and speed at which AI competence reaches a district administrator, small manufacturer, and the median worker.

1. Training at Population Scale: The Vocational Rail

India's strategy presents this logic well because it made a clear institutional choice: it used existing delivery infrastructure rather than building a new tier of elite AI institutions.

Under the FutureSkills pillar of the IndiaAI Mission, the Government has planned a network of 570 Data and Artificial Intelligence Labs across Industrial Training Institutes (ITIs) and polytechnics in Tier-2 and Tier-3 cities, with the first tranche launched in partnership with the National Institute of Electronics and Information Technology (NIELIT) and industry. These labs teach AI fundamentals, data annotation, data curation and applied data science.

Additionally, the mission's mandates include 58 AI Centres of Excellence across States and Union Territories, and 686 fellowships awarded across 178 institutions, with sanctioned support for 500 doctoral, 5,000 postgraduate, and 8,000 undergraduate students. Compute access, the constraint most often cited as disqualifying, was addressed not only by building sovereign frontier capacity but also by aggregating and subsidising it through a national common compute pool, as set out in chapter one of this document.

The scale of the persistent gap should not be disregarded. On figures cited in a joint report by IndiaAI and IBM, roughly 200 million of India's 600 million workers are AI-literate, which is around 30 per cent, with the stated ambition of doubling that share by 2030. This identifies data readiness and governance capacity, not model access, as the reason most Indian enterprises have failed to move AI from pilot to scale.



570

Data and AI Labs planned across ITIs and polytechnics in Tier-2 and Tier-3 cities



58

AI Centres of Excellence across States and Union Territories



686

fellowships awarded across 178 institutions



8,000 · 5,000 · 500

undergraduate, postgraduate and doctoral students with sanctioned support



200M of 600M

Indian workers are AI-literate — around 30 per cent (IndiaAI-IBM joint report)



2× by 2030

stated ambition of doubling the AI-literate share

SKILLING ON EXISTING DELIVERY RAILS

150 million+

people trained by India's Skill India Mission across more than forty industrial sectors

PBIA 2024–2028

Brazil anchors AI training and diffusion in the SENAI and SENAC technical education networks

3MTT

Nigeria's 3-Million Technical Talent programme, run alongside an AI Skills Initiative for Developers and the Government

The three design decisions made under this mission are worth isolating, because they are portable:

01

First, the delivery rail is vocational, not academic: An ITI in a Tier-3 district reaches a cohort that a research university may not, structurally. It is also more fiscally viable to convert existing infrastructure, avoiding investments often seen as sunk costs such as building, establishing the enrolment pipeline and building the instructor cadre from the ground up. India's Skill India Mission has trained over 150 million people across more than forty industrial sectors on this logic, emphasising employability over formal credentials.

02

Second, the curriculum is honest about where students begin their educational journey. By combining annotation and curation with applied data science coursework, it understands that for fresh graduates, their first job will be in data science.

03

Third, from the outset, compute is envisioned not as an asset to be owned but as a shared utility: Subsidised hourly access converts a capital barrier into an operating cost and enables more equitable diffusion.

Structurally, the same model has emerged across the Global South, adapted to local institutions. Brazil anchors the training and diffusion axis of the Plano Brasileiro de Inteligência Artificial (Brazilian Artificial Intelligence Plan, PBIA) 2024–2028, which hosts a specific vertical for diffusion and training in the SENAI (National Service for Industrial Training) and SENAC (National Service for Commercial Training) technical education networks rather than in the federal universities alone. Nigeria runs the 3-Million Technical Talent (3MTT) programme, an AI Skills Initiative for Developers and the Government in parallel, pairing the developer pipeline with senior public-sector AI literacy.



SKILLING AT THE LAST MILE: NIELIT AND THE MAKING OF APPLIED ARTIFICIAL INTELLIGENCE IN MANIPUR

The National Institute of Electronics and Information Technology (NIELIT) is an autonomous scientific society under the Ministry of Electronics and Information Technology (MeitY), Government of India, mandated to develop human resources in Information, Electronics and Communications Technology. Its distinguishing feature is not scale alone but distribution.

Rather than concentrating capability in a flagship campus, NIELIT operates 56 centres of its own alongside more than 9,000 partner institutes, structured to deliver training even in rural and remote parts of the country, and 22 of those 56 centres are located in the Northeastern Region. An additional 750+ accredited institutes and more than 9,000 facilitation centres extend the same certification standard through third-party delivery. This architecture produces measurable volume at the periphery. NIELIT has trained more than 31 lakh candidates across its programme categories over the last three years, including in emerging technologies, of whom more than 3.05 lakh are from the Northeastern Region.

NIELIT has been granted Deemed to be University status under the distinct category by the Ministry of Education, with its main campus in Ropar, Punjab, and 11 constituent campuses across regions including the Northeast, Jammu and Kashmir, and other underserved areas. Under the IndiaAI Mission's FutureSkills pillar, 27 IndiaAI Data and AI Labs have also been established through NIELIT centres, including 10 in the Northeast and five in the Himalayan region. This builds on FutureSkills PRIME, under which 22 NIELIT centres and the Centre for Development of Advanced Computing (C-DAC) serve as Lead and Co-Lead Resource Centres across 10 emerging technologies, having trained over 32,700 government officials, 2,367 trainers and more than 40,400 candidates.

Permeation of Technology: The Imphal centre

NIELIT Imphal is one of the eleven constituent campuses and the source of two projects discussed here. Founded on 1 October 1988, it became operational in July 1989 with a three-year diploma in electronics engineering and occupies a campus in Akampat, Imphal East district. The campus spans more than twenty acres and lies about five kilometres from the centre of Imphal city.

Its academic portfolio covers both sides of NIELIT's mandate. Its formal programmes include All India Council for Technical Education (AICTE) recognised diplomas in Electronics and Communication Engineering and Computer Science and Engineering, as well as postgraduate diploma, bachelor's and master's programmes in Computer Applications and Information Technology, affiliated with Manipur University. Its non-formal programmes include NIELIT's government-recognised 'O' Level courses in Information Technology, Hardware and Multimedia.

Since inception, the institute reports having trained over 4,000 students through formal university-affiliated programmes, more than 6,500 through non-formal NIELIT courses and over 50,000 candidates through short-term courses. Delivery extends beyond the valley. Two extension centres in Churachandpur and Senapati became operational in the 2013-14 academic session, with capacity to train 600 students a year each. Both are hill districts. The institute's statement of intent is to deliver digital literacy for all and to the last mile, generate employment, develop entrepreneurship through training, and pursue research and development projects as its four focus areas.

Project one: sign language detection and audio generation

The first project addresses a problem at the edge of most technology agendas and solves it with equipment already on the desk: a webcam watches a hand, and a laptop speaks the letter it sees. A team from NIELIT Imphal, including Dr Yumnam Jayanta Singh, Executive Director of NIELIT Imphal, and Dr Angom Buboo Singh, Academic Head, developed an approach to convert signed letters captured on a webcam into spoken audio in real time.

The pipeline is deliberately plain. A live webcam feed is cropped to a fixed 200×200-pixel region of interest, resized to 64×64 and normalised, then classified by a convolutional neural network trained on the American Sign Language Alphabet dataset, approximately 87,000 images across 29 classes, balanced at roughly 3,000 images per class, with a 90:10 train-validation split. The predicted letter is rendered on screen and passed to a text-to-speech engine.

The substantive engineering lies downstream of the model, and this is what makes the project a frugal-innovation exhibit rather than a classroom exercise. Speech synthesis runs on a separate daemon thread, so audio output doesn't freeze the video feed. A stability filter requires ten consecutive frames to agree before a letter is spoken, and a 1.5-second cooldown separates speech events, eliminating the repetition that would otherwise render the output unusable. The team reports end-to-end latency of about 80 to 120 milliseconds between the hand entering the region of interest and the on-screen overlay updating, below the 200-millisecond threshold conventionally cited in human-computer interaction research for an interaction to feel instantaneous.

The entire system runs on a webcam, a laptop and open-source libraries. There is no accelerator cluster, no inference endpoint, and no recurring per-user cost; this is the economic profile several underserved regions need to engage with technology. The project uses existing, limited infrastructure to address community needs. The team's own framing also identifies the outstanding gap. The model is trained on American Sign Language while the population the institute serves uses Indian Sign Language, highlighting the need to create large-scale indigenous models.

Project two: the Virtual Problem Solver

The second project addresses a pedagogical constraint rather than an accessibility one. Students lose substantial time transcribing mathematical expressions and diagrams into software before they can receive assistance, and existing tools cannot interpret handwritten visual problems.

The system removes the transcription step. Hand-landmark detection converts finger gestures into the input mechanism: index fingers to draw, a selection panel for pen and eraser, thumb up to clear the canvas; the completed drawing is transmitted to a large language model through an application programming interface, and the reasoning and answer are returned to a Streamlit interface. The reported test cases for pen, eraser, clear-canvas and output generation each returned successful results.

The cost logic is the same as the first projects. Removing the stylus, touchscreen and interactive whiteboard leaves a camera and a machine capable of running a Python-supported development environment, which the team notes is any ordinary system. The team then extends the design beyond its original brief, noting that the same gesture-selection layer would allow people with physical disabilities to operate virtual controls on industrial automation lines they cannot reach physically.

Conclusion

That both teams arrived independently at accessibility applications is perhaps the most instructive detail in the pair. Neither was directed to do so. Problem selection tracked proximity to the problem. Three conclusions follow, each transferable beyond India.

First, skilling infrastructure becomes innovation infrastructure once it reaches sufficient depth. NIELIT Imphal is a training institution assessed on employability, not a research laboratory, yet its faculty, laboratories, and student cohorts produced two working applied artificial-intelligence systems addressing locally visible problems. For learning ecosystems that cannot finance a research institute in every region, building the skilling network first produces both the trained workforce and, as a by-product, the applications.

Second, location determines the problem set. Both projects respond to constraints observable from Imphal: assistive communication, and classrooms where hardware is scarce and instructional time expensive. Centralised institutions may not generate these problem statements or device intentionally lean models that can be deployed on the ground in geographically limited connectivity regions.

Third, frugality is a teachable engineering discipline, not a compromise. What made the first system usable was not the model but the threading, the stability filter and the cooldown timer, the craft of making a system run on hardware already in the room. That craft is exactly what a skilling institution exists to transmit, and it is the capability that scales across low-resource settings.

The African Union's Continental Artificial Intelligence Strategy provides a broad umbrella; the Global AI Summit on Africa in Kigali (April 2025) produced the Africa Declaration on AI, endorsed by ministers from 49 countries. Multilateral scaffolding has followed: the International Telecommunication Union's (ITU's) AI Skills Coalition now works with over 70 organisations to deliver close to 200 online AI courses in 13 languages, spanning schoolchildren to diplomats. Kenya has moved AI into its national curriculum and hosts the Africa Centre of Competence for Digital and Artificial Intelligence Skilling, established with the United Nations Development Programme and Microsoft, whose first cohort draws participants from Kenya, Uganda, Malawi, Cabo Verde, Rwanda, Tanzania and Nigeria. Ghana runs a One Million Coders Programme; Egypt begins AI literacy from grade four; Senegal, Tunisia, Mauritius and Togo run population or cadre-specific programmes.

Fundamentally, on paper these programs target a population base that still struggles to ensure basic access to digital infrastructure. Thirty-eight of Africa's fifty-four countries still have no national AI strategy, and four countries, namely, Nigeria, Kenya, South Africa and Egypt, account for an estimated 83 per cent of AI investment on the continent. Meanwhile, most developing economies have limited absorptive capacity. These problems are compounded by the ambitious task of training a large population to enter a market with limited placement capacity, and the distributional burdens often left unaccounted for, particularly the impact on women and marginalised social groups. Analysis by the International Labour Organisation (ILO) indicates that in developing countries, women's jobs may be disproportionately exposed to AI-driven automation. A skilling architecture that scales without being gender-conscious could scale an existing inequality with unusual efficiency.

AI SKILLING ACROSS AFRICA

49

countries whose ministers endorsed the Africa Declaration on AI at the Global AI Summit on Africa, Kigali (April 2025)

~200 courses

online AI courses in 13 languages delivered by the ITU's AI Skills Coalition with over 70 organisations, spanning schoolchildren to diplomats

1 million

coders targeted by Ghana's One Million Coders Programme; Egypt begins AI literacy from grade four

38 of 54

African countries still have no national AI strategy

83%

of AI investment on the continent is accounted for by four countries: Nigeria, Kenya, South Africa and Egypt

7 countries

in the first cohort of Kenya's Africa Centre of Competence for Digital and AI Skilling (with UNDP and Microsoft)

2. Design and Application Capability: Entering Capital-Intensive Chains at the Capital-Light Point

Capability read at the frontier consistently misreads the Global South. Its true value shows up in design and application. This last-mile capacity is being built and defined by the constraints most Global South economies face.

Compute and financial constraints have created a prevailing capital-light logic for model development. Studies have found that fine-tuned Small Language Models (SLMs) can achieve approximately 80-90% of Large Language Models (LLMs) performance at 10-25% of the computational cost, with estimated energy reductions of over 70%, while offering more curated, culturally specific performance in languages that frontier developers have no commercial reason to serve well. The resulting body of work is a design portfolio, not a set of derivative applications.

Language and voice created the first application class, addressing the principal challenge that large user bases may not be able to engage with the script global LLM interfaces use, by speaking a language most frontier models may not serve well. Giving rise to models operating in local languages, and voice-first models. In India, Sarvam AI has released models that support 22 Indian languages, alongside speech recognition and dialogue-based systems, with edge deployments aimed at on-device operation in local languages. The Digital Public Infrastructure revolution was driven by the National Payments Corporation of India's voice-enabled Unified Payments Interface, which extends transaction capability to users who cannot use a text interface. In Africa, Lelapa AI's InkubaLM covers Swahili, Yoruba, IsiXhosa, Hausa and IsiZulu, serving over 300 million speakers.

FINE-TUNED SMALL LANGUAGE MODELS VERSUS LARGE LANGUAGE MODELS

80–90%

of LLM performance achieved

10–25%

of the computational cost

> 70%

estimated energy reductions

LANGUAGE AND VOICE: THE FIRST APPLICATION CLASS



22 Indian languages

supported by Sarvam AI's models, alongside speech recognition and dialogue-based systems, with edge deployments aimed at on-device operation



300 million+ speakers

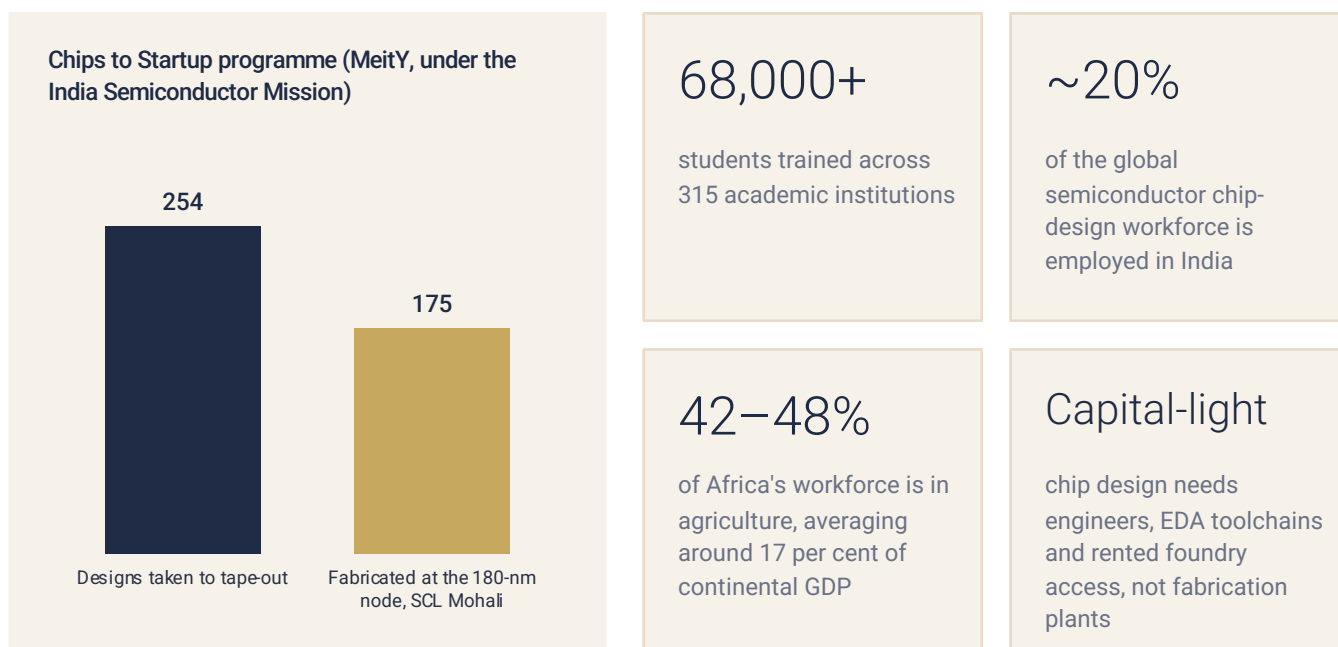
served by Lelapa AI's InkubaLM, covering Swahili, Yoruba, IsiXhosa, Hausa and IsiZulu

Scale is another perceived challenge that the Global South has harnessed into a strength. An African model built for agriculture can immediately have access to a user base that serves between 42 and 48 per cent of Africa's workforce and averages around 17 per cent of continental Gross Domestic Product (GDP), giving any working intervention an immediate reach that no comparable Northern deployment can match. Design approaches have matched the opportunity by building open-source computer vision and geospatial architectures, available at no pretraining cost, to deliver crop monitoring, pest identification, yield prediction and weather forecasting directly to smallholders.

The same capital-light logic extends into the hardware layer that makes these applications possible. Chip design is capital-light relative to fabrication: it requires engineers, Electronic Design Automation (EDA) toolchains and rented foundry access. India has leveraged this opportunity to become key to the physical infrastructure enabling AI, employing nearly 20 per cent of the global semiconductor chip-design workforce and hosting a significant share of global semiconductor design centres.

The Chips to Startup programme, run by the Ministry of Electronics and Information Technology (MeitY) under the India Semiconductor Mission (ISM), reports over 68,000 students trained and 254 designs taken to tape-out, 175 of them fabricated at the mature 180-nanometre node at the Semiconductor Laboratory, Mohali. Participation spans 315 academic institutions. Fabricating student work at a mature node is not a compromise: it teaches the full design-to-silicon cycle at a fraction of the cost, and mature-node capability is where a significant segment of global automotive, industrial and power-electronics demand actually sits.

However, instruments across the world do not currently address retention capacity. Several Global South skilling initiatives lack retention instruments such as competitive public-sector pay scales, tenure-linked incentives, or domestic research funding at credible intensity, in effect putting them at risk of becoming subsidised talent export hubs.



3. Data Work: The Grey Economy and the Terms of Participation

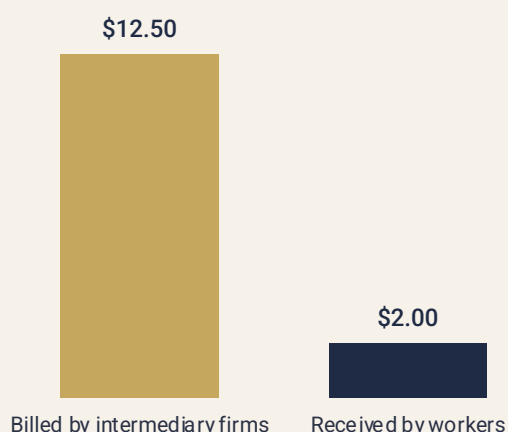
Every frontier model in commercial use in part rests on human judgement, data that has been annotated, enriched, verified and evaluated by people. That work is performed disproportionately in Kenya, India, the Philippines, Venezuela and Colombia. No account of Global South AI capability is complete without it, and no honest account can begin anywhere other than with the terms on which it has been performed.

Reporting on outsourced safety-data work for large model developers found intermediary firms billing on the order of USD 12.50 per hour while the workers themselves received approximately USD 2, for sustained exposure to violent and abusive material. Wage estimates for Kenyan data labelling have been reported between USD 1.50 and 2.00 per hour. The result is a grey economy in the precise sense of the term: legal, economically significant, and almost entirely outside the labour classification, social protection and skills-recognition systems of the states that host it.

But this argument isn't limited to wages; it circles back to the fundamental approach for integrating and assessing the role Global South economies play in enabling the global AI ecosystem. Ethnographic research on AI data work in India documents a regime of counting, assuming that everything from tasks and deliverables to workers, hours, quality, and performance can be managed through enumeration. Clients determine what is counted and therefore what is accountable, while the consequences of what goes uncounted are absorbed entirely at the worker's end.

This category of work is also not permanent. UNCTAD's assessment is that AI-powered automation could affect up to 40 per cent of jobs worldwide, and that it specifically threatens to erode the low-cost labour advantage many developing economies have relied on for growth and poverty reduction. Applied to annotation, this is not an abstract risk: workers labelling content to train moderation systems are, in a direct sense, training the systems that will absorb the task.

Outsourced safety-data work for large model developers: USD per hour



USD 1.50–2.00

per hour – reported wage estimates for Kenyan data labelling



Up to 40%

of jobs worldwide could be affected by AI-powered automation (UNCTAD), threatening to erode the low-cost labour advantage of many developing economies

This adversity has given rise to responses originating in the Global South rather than being imposed on it.

01

The first is reclassification by the state. Under India's IndiaAI Mission, skilling initiatives and the AI Labs teach annotation and curation as formal, certified curriculum, giving an informal task an institutional home, a syllabus, and a credential. A credentialed skill can be priced, benchmarked, and progressed from. Creating a base rung to build on, consequential at scale.

02

The second is a restructuring of the ownership claim itself. Karya, an India-based data organisation, under its Public License structuring, workers retain a claim on the datasets they help create and receive royalties on resale. The organisation reports reaching over 100,000 workers and building more than seventy language datasets, with operations extended to Kenya and Ethiopia. This approach is architectural rather than numerical, establishing that data work can be structured as an asset-generating activity rather than a piece-rate transaction, supporting advanced local model development and protecting social capital.

03

The third is the emergence of standards backed by purchasing power. The Fairwork programme, developed by the Oxford Internet Institute with the WZB Berlin Social Science Centre, has extended its principles to AI supply chains and operates an audit-and-certification pathway covering annotation and data enrichment. Its methodology has been applied across 41 countries. Its ratings have been referenced in state-level studies of platform work in India and in official reports on the platform economy in Brazil.

Beneath all three lies the fact that changes the negotiating position entirely. High-quality data within low-resource languages, regional dialects and culturally specific contexts cannot be manufactured from outside the communities that hold it. Its absence is not only an equity problem for the South but a validity problem for the North, essential for resolving concerns regarding biased inference and degraded performance. The strategic task, therefore, is to shift position within the value chain rather than secure a better price at the bottom.

RECLASSIFICATION, OWNERSHIP, STANDARDS



100,000+

workers reached by Karya, with more than seventy language datasets built and operations extended to Kenya and Ethiopia



Royalties

under Karya's Public License structuring, workers retain a claim on the datasets they help create and receive royalties on resale



41 countries

where the Fairwork programme's methodology (Oxford Internet Institute with the WZB Berlin Social Science Centre) has been applied

CASE STUDY: KATHA



Katha is an NGO based in India that works extensively to integrate literacy into grassroots initiatives in education and urban transformation. The organisation works to bring children living in poverty into reading and quality education. Its flagship initiative, the **Katha Adivasi-Dalit Learning Lab (KADL)**, ran its core educational programme across government schools, Katha Magic Labs, Mobile Katha Magic Labs and community reading centres in the Minjur and Poondi blocks of Thiruvallur district from January to March 2026, under a CSR partnership with a global MNC. The programme reached approximately **1,800 children in Grades 1 to 5**, and was based on the premise that AI readiness does not need to begin with a device or a line of code.

The sessions were conducted entirely without devices and built into Katha's existing StoryPedagogy curriculum, known as **Vidyalogam**, using structured worksheets, observation, discussion and reflection rather than screens or software. Learning progressed by grade level. Children in Grades 1 and 2 learned that machines follow instructions and that humans remain responsible for what those instructions produce. By Grade 3, children were introduced to the idea of prediction as a helpful guess rather than a certainty. Grade 4 introduced concepts of decision-making and fairness, with children learning to identify biased or unfair outcomes. By Grade 5, children were able to articulate basic principles of AI ethics and recognise the risks associated with misuse of the technology, all without ever using a computer.

Alongside the children's sessions, KADL conducted three community sessions for women, girls and youth, using the **Maa Mandal** community structure already established in these villages. One session focused on health, hygiene and the importance of continuing girls' education beyond the primary years, with participating mothers pledging to prioritise keeping their daughters in school. A second session, aimed at youth aged 16 to 25, introduced AI and data annotation in simple terms as potential livelihood opportunities. Several participants who owned smartphones expressed interest in learning more about available training pathways. A third session connected the AI-readiness work in schools with broader economic opportunities for women, leading three mothers to volunteer as **StorySeeds narrators** for the community oral history component.

The programme placed a deliberate emphasis on **thinking, dialogue and reflection**, making it suitable for an equitable and safe learning environment in a low-resource setting. The initiative demonstrates that many of the cognitive foundations that underpin AI, including **observation, pattern recognition, sequencing and an understanding of fairness**, can be developed at an early age without specialised equipment. It also shows that children from some of the country's most underserved communities, including first-generation learners, can engage meaningfully with these concepts when they are introduced through appropriate, locally grounded teaching methods.

4. What the Global South is Building

Read together, three principles define what is being built: skilling as the core diffusion infrastructure, where the measure of success is not the number of frontier researchers produced but the depth of AI competence in the application layers supporting operations on the ground; capability fortified at the point of use, placed where the work is and at the specification the work requires rather than centralised at the highest available specification; and a labour-to-ownership ladder, insisting that participation in the AI supply chain generate retained value on terms that drive inclusion across all domestic user bases.

The New Delhi Declaration on AI Impact, adopted on 19 February 2026 at the first global AI summit hosted in the Global South, names human capital as one of its cornerstones and carries an AI Workforce Development Playbook and a set of voluntary AI skilling principles among its deliverables, alongside a Charter for the Democratic Diffusion of AI and a Global AI Impact Commons. Its true value will depend on how these instruments carry forward shared agendas for the Global South.

THREE PRINCIPLES DEFINE WHAT IS BEING BUILT



Skilling as the core diffusion infrastructure

The measure of success is not the number of frontier researchers produced but the depth of AI competence in the application layers supporting operations on the ground.



Capability fortified at the point of use

Placed where the work is and at the specification the work requires rather than centralised at the highest available specification.



A labour-to-ownership ladder

Insisting that participation in the AI supply chain generate retained value on terms that drive inclusion across all domestic user bases.

19 Feb
2026

NEW DELHI DECLARATION ON AI IMPACT: DELIVERABLES

- AI Workforce Development Playbook
- A set of voluntary AI skilling principles
- Charter for the Democratic Diffusion of AI
- Global AI Impact Commons

Human capital, named as one of its cornerstones, adopted at the first global AI summit hosted in the Global South.

Primarily, mutual recognition of credentials. Certifications and design qualifications earned in one participating economy should have mechanisms to make them portable across others. Ensure national skilling programmes do not become isolated labour pools competing against each other and enable cross-border innovation. Portability is also the most direct retention instrument available, as qualifications that travel across the Global South can anchor talent in the local economy.

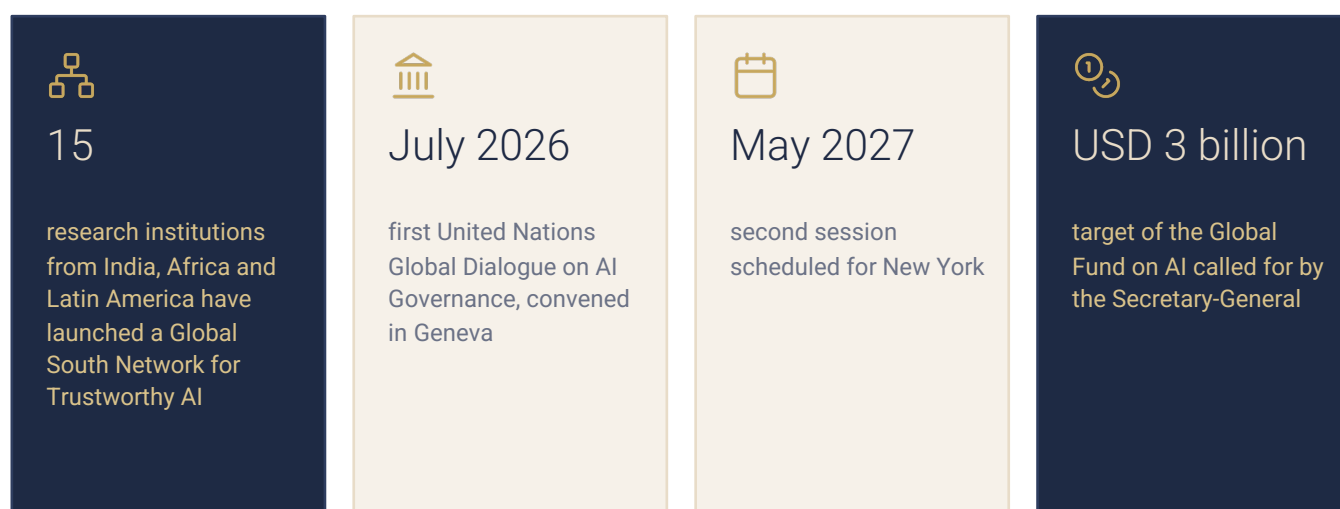
Second, enable pooled, open-access tooling and compute to extend South-South cooperation. India's Chips to Startups demonstrates that free institutional access to commercial toolchains produces measurable design output at national scale. The same architecture, negotiated as a bloc rather than individually, enhances access to EDA licences, subsidised compute and small-model training infrastructure across participating economies.

Third, common frameworks should measure AI use, not just participation. New commitments from leading AI developers to measure real-world use and improve multilingual and use-case testing create an opportunity to do this. Global South governments could collectively push for these measures to be broken down by region, language and gender. This would ensure that AI activity missed by standard measures is properly captured. What is not measured is unlikely to be funded.

The institutional vehicles are assembling. Fifteen research institutions from India, Africa and Latin America have launched a Global South Network for Trustworthy AI. The first United Nations Global Dialogue on AI Governance convened in Geneva in July 2026, with a second session scheduled for New York in May 2027. The Secretary-General has called for a Global Fund on AI, targeting USD 3 billion. Each provides a platform which can drive this shared vision forward in its own capacity.

The constraint was real. What it produced was a set of institutional answers that the Global North, having never faced the constraint, did not need to invent but now has several reasons to study.

THE INSTITUTIONAL VEHICLES ARE ASSEMBLING



Conclusion

Today humanity stands in front of an opportunity to define the future of AI and as members representing one of world's largest population basis, we can choose what the future of AI looks like not just as consumers but as co-creators of inclusive systems. For the Global South, this is not simply a question of adopting technologies developed elsewhere, it is about seizing an opportunity to shape what AI becomes and who it serves. Making the most of this opportunity means building AI systems and architecture that actually reflect the environment they're developed and deployed in. That means local, contextualized data; compute that's affordable and accessible; a skilled workforce; solid infrastructure; and models built to work within real-world constraints, not around them. It also means letting go of the assumption that AI progress has to follow a straight line defined by ever-bigger compute or ever-larger models. The examples in this compendium tell a different story: constraints can actually spark innovation in how AI systems get designed, trained, and deployed - giving rise to approaches like frugal AI, edge AI, and specialized datasets that squeeze far more out of limited computing resources.

This compendium brings together a range of initiatives, startups, and case studies that show these approaches in action. They come from different countries, sectors, and stages of the AI ecosystem, but they share a common thread: solving problems rooted in local realities. Together, they make the case that resource constraints don't have to mean falling behind technologically. In fact, they can push innovators toward solutions that are leaner, more context-aware, and more accessible - especially for communities that conventional AI development often overlooks. The bigger takeaway is that AI's future doesn't have to be written solely by those with the deepest compute reserves or the largest tech companies. The Global South has a real role to play in shaping how AI is built, where it's deployed, and whose needs it actually serves. The work now is to build on the innovation already underway, deepen collaboration across countries and institutions, and create the right conditions for locally grounded AI to grow and scale. Ultimately, the future of AI shouldn't be shaped only by what the technology can do: but by what societies choose to build with it, and who they choose to build it for.



Contributors



Contributors



Nilaya Varma
Group CEO & Co-Founder



Vinay Vijay Singh
Managing Director



Vidhi Narang
Assistant Vice President



Sharad Venkatraman
Manager



Indranuj Pathak
Manager



Mayuri Dutta
Manager



Ishita Garg
Manager



Sakshi Naphade
Senior Consultant



Parul Kataria
Consultant

Contact Us

Nilaya Varma, Group CEO & Co-Founder

Phone: +1 (346) 503-1835 | Email: nilaya@primuspartners.in

Vinay Vijay Singh, Managing Director

Phone: +1 (267) 243-3184 | Email: vinayv@primuspartners.in

PRIMUS

PASSION

for providing solutions to help clients achieve their goals

RESPECT

for all and alternate viewpoints

INTEGRITY

of thoughts and actions

MASTERY

of our chosen subject to drive innovative and insightful solutions

US

representing the Primus collective, where each individual matters

STEWARDSHIP

for building a better tomorrow



PRIMUS PARTNERS®

Solutions for Tomorrow

Primus Partners has been set up to partner with clients in 'navigating' India, by experts with decades of experience in doing so for large global firms. Set up on the principle of 'Idea Realization', it brings to bear 'experience in action'. 'Idea Realization'— a unique approach to examine futuristic ideas required for the growth of an organization or a sector or geography, from the perspective of assured on ground implementability.

Our core strength comes from our founding partners, who are goal-oriented, with extensive hands-on experience and subject-matter expertise, which is well recognized in the industry. Established by seasoned industry leaders with extensive experience in global organizations, Primus Partners boasts a team of over 250 consultants and additional advisors, showcasing some of the finest talent in the nation.

The firm has a presence across multiple cities in India, as well as Dubai, UAE. In addition, the firm has successfully executed projects across Africa, Asia Pacific and the Americas.

India Offices



Bengaluru

91 Springboard
Business Hub 175, 176
Bannerghatta Rd,
Dollars Colony,
Bengaluru - 560076



Chandigarh

2nd Floor, Netsmartz,
Plot No. 10, Rajiv
Gandhi Chandigarh
Technology Park,
Chandigarh - 160019



Chennai

147, Pathari Rd, Door #3,
WorkEz Hansa Building,
RK Swamy Centre,
Thousand Lights,
Chennai, TN - 600006



Delhi

1 to 7, UG Floor,
Tolstoy House,
Tolstoy Road,
Connaught Place
New Delhi - 110001



Kolkata

Siddhartha Apartments
4th Floor, 188/2,
Block J,
New Alipore,
Kolkata - 700053



Mumbai

156/157, 15th Floor,
Nariman Bhavan,
NCPA Road,
Nariman Point,
Mumbai - 400021

International Offices



Dubai

United Arab Emirates
(UAE)



Dammam

Kingdom of Saudi Arabia
(KSA)



Washington D.C

United States of America
(USA)

www.primuspartners.in

info@primuspartners.in

[in](#) Primus Partners India

[X](#) @partners_primus

[@primuspartners7128](#)