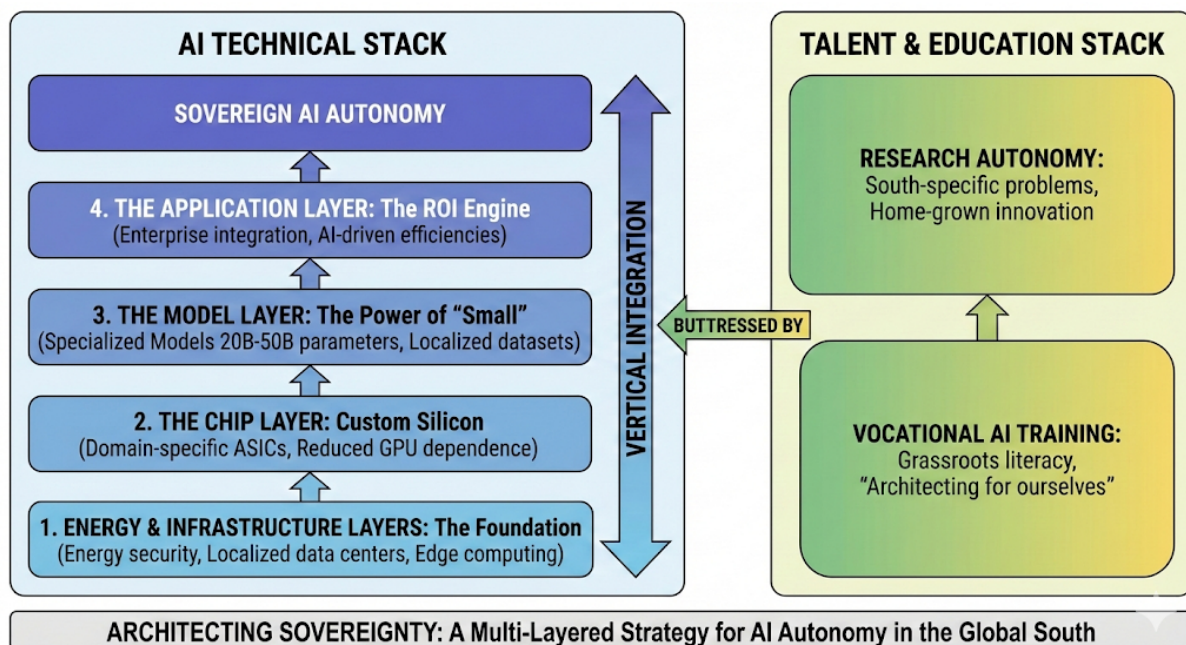# Architecting Sovereignty: A Multi-Layered Strategy for AI Autonomy in the Global South

Soumya Banerjee
University of Cambridge

Abstract

As Artificial Intelligence (AI) becomes the primary engine of global economic growth, the risk of "digital colonialism" looms over the Global South. This paper proposes a framework for AI Sovereignty, arguing that true autonomy requires a vertical integration of five technical layers (Application, Model, Chip, Infrastructure, and Energy) buttressed by a robust Talent and Education stack. Using India as a primary case study, we demonstrate that return on investment driven application development and medium-scale specialized models (20 billion to 50 billion parameters) provide a more sustainable path to sovereignty than the pursuit of massive, general-purpose large language models.

ARCHITECTING SOVEREIGNTY: A Multi-Layered Strategy for AI Autonomy in the Global South

I. The 5-Layer Technical Architecture

To achieve sovereignty, a nation cannot rely on black-box APIs from external providers. It must build and control its own stack.

1. The Energy & Infrastructure Layers

- **The Foundation:** AI is fundamentally a transformation of electricity into intelligence. Sovereignty begins with energy security and the localization of data centers.
- **Edge Computing:** For the Global South, infrastructure must be optimized for intermittent connectivity and localized cooling solutions to suit regional climates.

2. The Chip Layer

- **Custom Silicon:** Reducing dependence on global GPU monopolies by investing in domain-specific accelerators (Application Specific Integrated Circuits: ASICs) tailored for the types of models the Global South actually needs.

3. The Model Layer: The Power of Small

- **The 95% Rule:** Indian minister Ashwini Vaishnaw recently argued that while the world focuses on trillion-parameter models, 95% of enterprise utility may be handled by models in the 20 billion to 50 billion parameter range [1,2].
- **A "Bouquet of Models":** Instead of one giant model, India is developing a specialized fleet of models tuned for local languages, agricultural data, and healthcare specifics [1]. This increases efficiency and reduces the cost of inference. These are smaller, domain-specific models fine-tuned for local context (Indian languages) and local problems (healthcare diagnostics, precision agriculture, and financial inclusion).

4. The Application Layer: The ROI Engine

- **The Services Evolution:** India's historical strength in IT services is its greatest asset. The path to AI success is not just building the model; it is the *integration*.
- **Enterprise Integration:** Success comes from understanding the "business of the enterprise": translating complex workflows into AI-driven efficiencies. This is where the Return on Investment (ROI) is realized. India may need to prioritize the application layer because this is where the greatest economic value may lie [1]. This would involve deeply understanding an enterprise's business and wrapping specialized AI models around their specific workflows.

5. The Education & Talent Stack: The Sixth Layer

Technical infrastructure is hollow without the human capital to steer it. The Global South must pivot its education system from "coding for others" to "architecting for ourselves."

- **Vocational AI Training:** Moving beyond elite universities to provide AI literacy at the grassroots level.
- **Research Autonomy:** Encouraging home-grown research that solves South-specific problems (e.g., AI for dry-land farming or linguistic diversity) rather than chasing Western benchmarks.

II. The Mineral Layer: Securing the Physical Foundations of AI

This section addresses the foundational "Layer 0": the *Mineral and Material Layer*. For a nation like India to achieve true AI sovereignty, it must secure the physical elements that make "Compute" possible. Sovereignty in the application and model layers is precarious if the underlying hardware depends on a fragile, concentrated global supply chain of critical minerals. AI hardware, from GPUs to high-speed fibre optics and data centre cooling systems, relies on a specific set of "AI Minerals."

The table below identifies the minerals essential for the AI stack that India must prioritize:

Mineral	Application in the AI Stack	Criticality
High-Purity Silicon	The base substrate for all semiconductor wafers.	**High**: Basis of all chip manufacturing.
Gallium & Germanium	Used in high-performance AI chips (GaN) and fiber-optic data transmission.	**Extreme**: China controls ~90% of global supply.

Mineral	Application in the AI Stack	Criticality
Rare Earths (Nd, Dy, Tb)	Neodymium, Dysprosium, and Terbium for high-strength magnets in HDD storage and cooling fans.	**Extreme**: Essential for data center efficiency.
Copper & Aluminium	Massive quantities required for data center wiring, power grids, and liquid cooling plates.	**High**: Global supply is tightening as AI demand scales.
Lithium & Cobalt	Essential for UPS systems and backup power for edge computing nodes.	**High**: Required for 24/7 uptime.

Currently, India faces a significant **Strategic Risk** due to:
- **Import Vulnerability:** India imports approximately 85%–95% of its critical mineral requirements.
- **Geopolitical Chokepoints:** Export restrictions on Gallium and Germanium by dominant players have previously led to 30% price spikes within months, threatening the timelines of India's semiconductor "Fabs."
- **Refining Bottlenecks:** Even if raw ore is found domestically, India currently lacks the high-purity refining infrastructure required for semiconductor-grade silicon (99.9999999% purity).

To mitigate these risks, India has launched a multi-pronged **National Critical Minerals Mission (2025)** with a ₹34,300 crore ($4.1 billion) outlay. The strategy includes:

A. Overseas Acquisition (CMOAA)
India has established the **Critical Minerals Overseas Acquisition Authority (CMOAA)**. This sovereign-backed vehicle uses a Public-Private Partnership model to acquire stakes in mines abroad (specifically in Australia, Argentina, and Africa) ensuring a "Mine-to-India" dedicated pipeline.

B. The "Silicon-to-Chip" Vertical
Under the **India Semiconductor Mission (ISM)**, the government is incentivizing the domestic production of **polysilicon and ingots**. By moving from importing wafers to manufacturing them locally from high-purity domestic quartz, India reduces the "Silicon Risk."

C. International Resource Diplomacy
- **Mineral Security Partnership (MSP):** India's active role in this partnership allows for shared investment in refining technologies and mutual defense against supply shocks.
- **Bilateral Pacts:** Recent $100 billion agreements with Saudi Arabia and the UAE (2025) include dedicated "Mineral Corridors" that exchange Indian tech expertise for Gulf-led mineral investments.

III. The Strategy for India and the Global South

India is uniquely positioned to lead this movement due to its massive data footprint and its Service-to-Product transition. AI Sovereignty is not about isolationism; it is about having the right to choose. By building the full stack, Global South nations ensure that their economic future is not dependent on the pricing or political whims of a few global tech giants.

Comparative Advantages

Layer	Strategy for the Global South
Model Size	Focus on 20 billion to 50 billion parameter efficiency.
Data Strategy	Use unique, local datasets (e.g., Indian languages) to create "sovereign moats."
Application	Focus on high-ROI enterprise services rather than pure consumer play.

IV. Conclusion

Building AI sovereignty is a generational project. By mastering the five layers of the technical stack and anchoring them in a revamped education system, the Global South can transition from a back-office service provider to a global AI powerhouse. The goal for countries in the Global South should be: *Intelligence that is locally owned, locally operated, and globally competitive*.

Declarations

AI declaration

The initial ideas for this paper were generated using a generative AI tool (Gemini). The author accepts full responsibility for the content of this manuscript.

References

[1] Ashwini Vaishnaw, World Economic Forum Annual Meeting, Davos, Switzerland 2026
[2] https://www.pib.gov.in/PressReleasePage.aspx?PRID=2168319&reg=3&lang=2, URL accessed Jan 2026