



KGID
2026
SEJONG

Small AI to Big AI

World Bank's AI Vision and Korea's Role

Jieun Choi

Senior Economist, World Bank Group

Key Messages

- World Bank's Digital and AI Strategy emphasizes a focus on "Small AI"
- Korea's AI Policy, while centered on "Big AI," offers valuable lessons for LMICs
- The World Bank's Global AI and Digital Center plays a key role in translating Korea's policy lessons for client countries

WB STRATEGY AIMS TO DIRECTLY TACKLE THESE THROUGH FOUR STRATEGIC OBJECTIVES...



WBG Digital mission is to:

Connect all people, governments, enterprises to deliver digitally enabled services, harness safe and secure digital and AI use, and create more economic and job opportunities for development impact.

1

Access

Connect people in hard-to-reach areas and improve quality of connectivity

2

Affordability

Deliver affordable devices, especially for the bottom 40%

3

Ecosystem

Unlock digitally enabled public and private services

4

AI Readiness

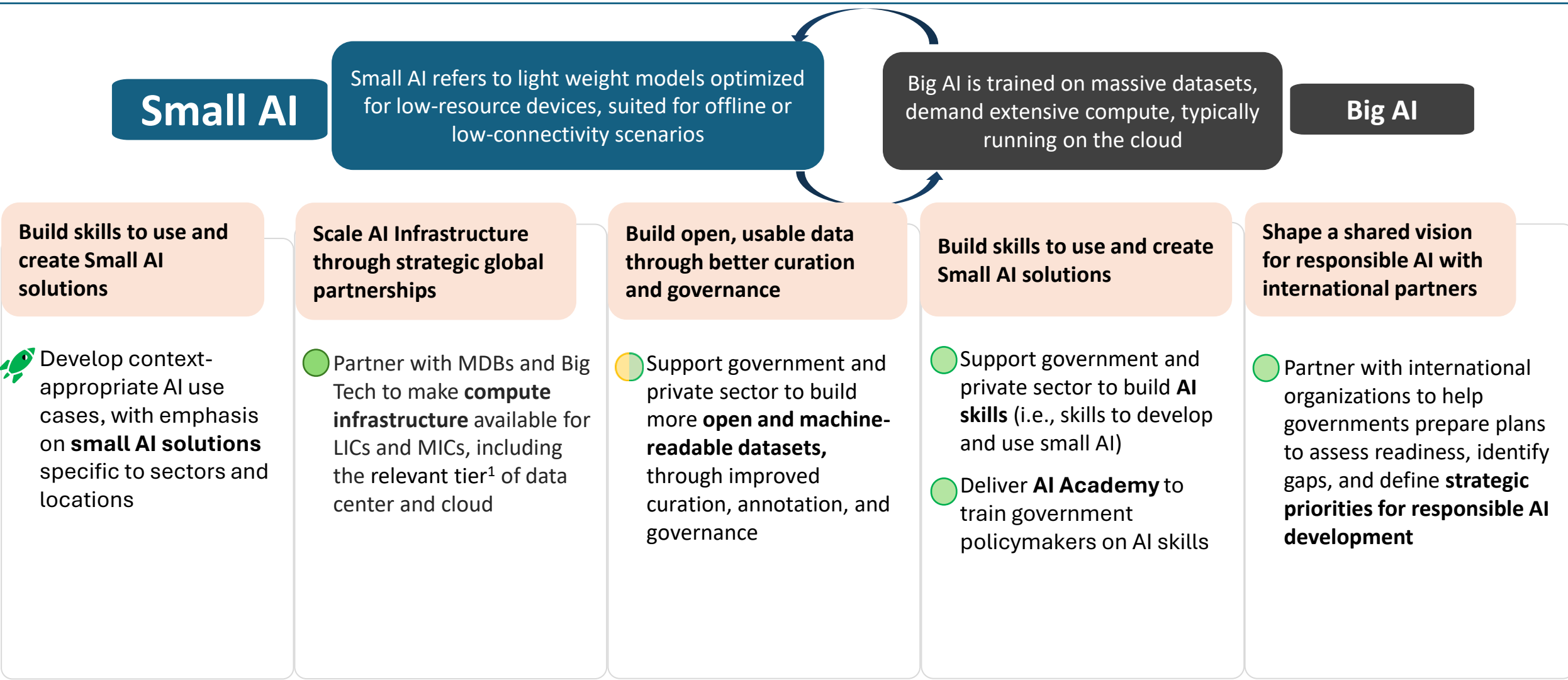
Bridge the widening digital divide with the rise of AI

Digital Safeguards - cybersecurity, data protection, privacy, and online safety

Greater relevance for LICs

Greater relevance for MICs

(4) AI READINESS: HELP UNTANGLE AI IMPLICATIONS AND HOW TO HARNESS ITS POTENTIAL



Small AI models are deliberately designed to be compact and compute-efficient

Small AI refers to simpler artificial intelligence models that are designed to perform narrow, specific tasks and can run on mobile phones, laptops, and other edge devices. Small AI can roll out faster and cheaper, reach people at the last mile, and improve services in farming, health, classrooms, and small businesses.

Key characteristics of Small AI



Task specific intelligence: Small systems are designed to perform a limited number of well-defined tasks rather than generalizing to new scope areas.



Lightweight and resource-efficient: Small AI models are cheaper to train and often optimized to run on edge devices like smartphones, wearables and microcontrollers.



Privacy-preserving: Because small AI often runs locally on devices, it reduces the need to transmit sensitive data to external servers.



Low latency: Models run locally, avoiding delays from communication with cloud servers.



Composable: Due to their smaller size, small AI models can be used to perform narrow tasks as part of a broader team of small models tasked with a broader task.

Implications for LMICs

Enables targeted solutions for local, context-specific challenges. (e.g.- crop disease detection, maternal health).

Lowers capital commitment required for developing or deploying models. Edge deployment enables use cases in remote areas.

Provides citizen and other users with more control over their personal or sensitive data.

Enables real-time support in areas where data centers may be unfeasible or over-utilized.

Enables the development of powerful AI systems without the costs associated with training foundation models.

Small AI solutions are already driving impact in LMICs across multiple sectors

InkubaLM for low-resource African languages

Use case overview

- InkubaLM is a 0.4B parameter model designed specifically for five African languages: Hausa, Yoruba, Swahili, isiZulu, and isiXhosa.
- InkubaLM achieves similar performance to larger models on machine translation, and outperforms many larger models in sentiment analysis.
- Heavily leveraged existing architecture and attention mechanisms to enhance efficiency.

Critical challenges

- **Language inconsistencies:** The model may sometimes use different languages when generating text responses
- **Accuracy and bias:** Similar to other language models, InkubaLM can struggle to provide accurate, logically consistent and bias-free responses

Lessons for Policymakers in LMICs

- **Growing techniques for developing SLMs:** The paper introducing InkubaLM discusses numerous techniques used to build the model, which can be directly leveraged by other model developers
- **Academia as a data source:** The paper also introduces two novel datasets used to train InkubaLM. These data sets could be leveraged to train more expansive models for African languages.

Fostering literacy and preserving culture- RobotsMali in Mali

Use case overview

- RobotsMali is using AI to create educational materials in Bambara, a local language in Mali.
- By generating schoolbooks and content in Bambara, RobotsMali empowers students and fosters literacy, ensuring that educational resources are accessible & culturally relevant.

Critical challenges

- **Data volume:** Limited digital or annotated data in Bambara to train AI models.
- **Vendor dependence:** Lack of local compute resources increased dependence on external platforms.
- **Cultural barriers:** Many users had not interacted with digital content in their native language.

Lessons for Policymakers in LMICs

- **Localization drives value:** AI tools that support local languages can enhance participation and learning outcomes.
- **Human-in-the-loop:** Human oversight of AI outputs helps manage quality and maintain trust in AI generated materials.

Agriculture resilience support- Plantix in India

Use case overview

- Plantix is an image recognition app which processes photos uploaded by users of diseases or pests which are present on their crops. The app also suggests solutions to remedy the issue.
- In India, Plantix partnered with the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) to train small farm holders

Critical challenges

- **Measuring feedback:** Plantix has struggled to comprehensively measure impact of the tool beyond proxy metrics such as number of users
- **Achieving expert-level accuracy:** While the app was still maturing, the app would only provide a list of potential identifications. As the model has matured, the confidence in the predictions has increased.

Lessons for Policymakers in LMICs

- **Training partnerships:** Leveraging ICRISAT's network of research labs and universities, the partnership was able to train over 3,500 students and staff on the Plantix app.
- **Community-based annotation:** Government agricultural workers, students of agricultural universities, and farmers in India helped with identify and annotating training images

Policymakers should assess small AI readiness across numerous dimensions

Dimensions	Diagnostic questions	Example pilot initiatives
1. Skills and human capital readiness	<ul style="list-style-type: none"> • Policymaker literacy- Do policymakers understand the benefits of small AI and strategies for encouraging the adoption of the technology? • Workforce readiness- Are members of the public supportive of AI adoption? Do they have the digital skills to deploy and maintain AI-enabled solutions locally? • Technical capacity- Do we have domestic ML experts who understand current small AI techniques and can build or customize small AI models? 	<p>India's national AI strategy recognized the pressing need to reskill the workforce for AI. This strategy included:</p> <ul style="list-style-type: none"> • Standardization of local informal training institutions • Creation/usage of open learning platforms • Creating financial incentives for reskilling employees • Encouraging higher education institutions to adapt curriculums
2. Data readiness	<ul style="list-style-type: none"> • Domain-specificity- Are there unique data sources we have or could feasible access which would lead to differentiated model performance? • Data quality & useability- Are publicly-available datasets sufficiently relevant for local context? Do we have the necessary resources to generate, label, and make data sources machine-readable? • Data availability- Is the necessary data infrastructure (e.g., data pipelines, warehouses, metadata catalogues) in place to support the use of small AI? 	<ul style="list-style-type: none"> • Conduct a review of publicly available datasets frequently used for training AI models • Compare these datasets to local language patterns and norms required for widespread national adoption • Identify gaps where local-language-specific datasets need to be created
3. Connectivity and infrastructure readiness	<ul style="list-style-type: none"> • Private sector optionality- Is our infrastructure supportive of the different methods that domestic businesses or entrepreneurs may want to use small AI? <ul style="list-style-type: none"> • API access- Is our national internet infrastructure sufficient to support models which rely on continuous API access? • Training & fine-tuning- Do we have ways of accessing sufficient compute to support domestic model training or fine-tuning? 	<p>An infrastructure readiness assessment could assess:</p> <ul style="list-style-type: none"> • Current methods citizens and businesses are accessing AI models • Infrastructure gaps which prevent other model adoption/access methods (e.g., API, local hosting) • Cost-benefit analysis between infrastructure buildout costs and additional model access methods
4. Devices and edge readiness	<ul style="list-style-type: none"> • Understanding device usage- Which devices do our local populations interact with most frequently? • Edge AI experience- Do we have experience procuring IOT sensors or other edge devices? • Incentive mechanisms- Are there any existing funding or incentive mechanisms for IOT or edge devices? • Secure standards- Have we established privacy and security standards for edge devices? 	<p>An assessment of edge device readiness should identify:</p> <ul style="list-style-type: none"> • Edge device penetration in target sectors • Best-in-class edge devices for resource-constrained environments • Compute capabilities of current devices being used versus requirements for SLMs or other AI models • Local manufacturing and import capabilities

Immediate actions for policymakers to promote the development and adoption of small AI

CREATING AN ECOSYSTEM

- a. Negotiate with universities and foreign partners to localize existing models under open licenses.
- b. Create a national open-source repository for small AI models, encourage participation from academia and the private sector.

ESTABLISH INITIAL GOVERNANCE

- a. Develop ethical AI guidelines and provide risk management frameworks tailored to local context.
- b. Establish reporting channels for citizens to report AI abuse or misuse.

GENERATING MOMENTUM

- a. Launch government-sponsored challenge grants for small AI models in priority sectors. Tie the results of the challenge grant to procurement commitments
- b. Establish a use case sandbox for small AI experimentation.

LEADING BY EXAMPLE

- a. Create short-term certification programs for civil servants in areas such as responsible AI use.
- b. Pilot small AI models in government agencies to signal commitment to AI and allow civil servants to learn by doing

REGIONAL COOPERATION

- a. Join or initiate regional AI cooperation framework which governs issues such as cross-border data transfer.
- b. Explore initial appetite for shared regional compute clusters.

FINANCING AND SUSTAINABILITY

- a. Establish public funding sources and mechanisms to support entrepreneurship, upskilling, and adoption.
- b. Identify initial partners for public-private partnerships.

Korean Case

- **Goal:** Leaping to a Top 3 Global AI Powerhouse
- **Principles :** People-centered inclusive AI, Public-Private Collaboration, AI-Ready System, Balanced AI Development



Korean Case: In the context of WB's Strategy

IN THE CONTEXT OF WB'S DIGITAL AND AI STRATEGY (4) AI READINESS –THE 4C+R FRAMEWORK

AI Specific Aspects

Digital Economy Aspects Required for AI

C¹

Competency

Digital skills

- From basic AI literacy to advanced technical training, skills are critical for adoption and responsible use
- Policymakers, institutions, and citizens need capacity to critically evaluate AI outputs and design safe systems
- Embedding AI competencies into STEM education builds the pipeline for long-term readiness

C²

Context

Data & algorithms

- Localized, high-quality, representative data ensures accuracy, fairness, and trust
- Governments must invest in data governance, interoperability, and secure data-sharing frameworks
- Balancing open-source flexibility with proprietary reliability helps protect sovereignty while fostering innovation

C³

Compute

Processing power

- Affordable access to computing (on-premise, cloud, or edge) is essential for AI innovation
- Enables local firms to train and fine-tune models tailored to their needs
- Supports real-time AI inference for services in health, education, agriculture, and beyond

C⁴

Connectivity

Infrastructure

- Broadband connectivity, sustainable power, and affordable devices are the foundation for AI adoption
- Without reliable electricity and internet, neither small-scale AI applications nor advanced data centers can function
- Investment in resilient digital and energy infrastructure is a prerequisite for scaling AI solutions

R

Responsible AI

Governance & trust

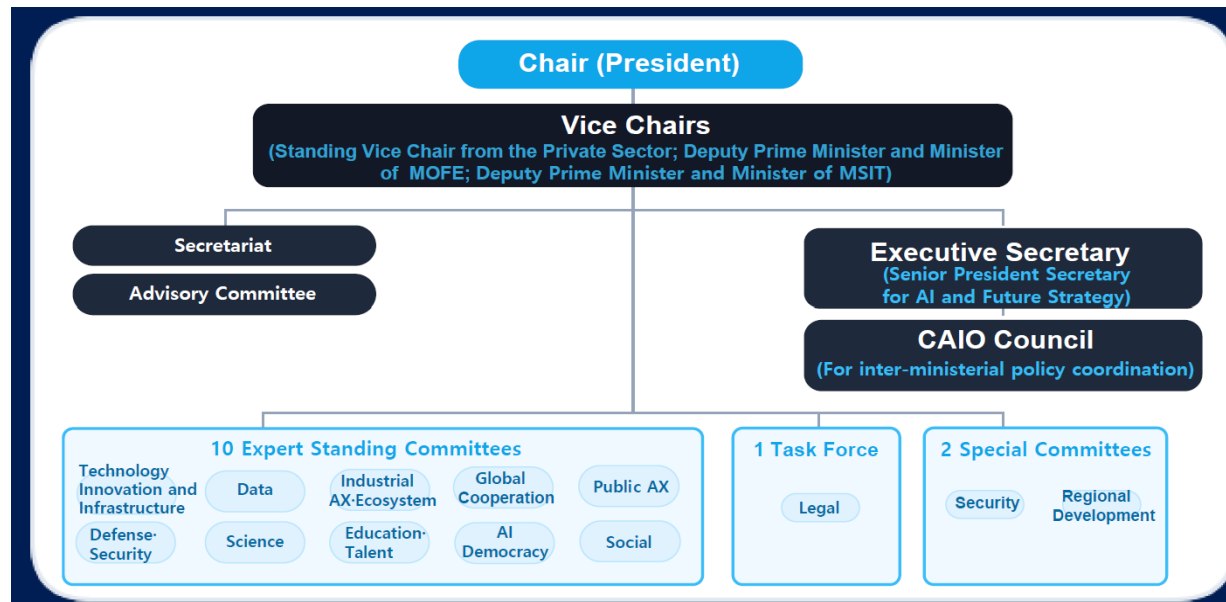
- Ensure frameworks balance innovation with risk mitigation
- Promote fairness, transparency, accountability, and inclusion
- Safeguard privacy, security, and human rights, especially in sensitive sectors
- Build capacity to enforce and update governance approaches

Korean Case: National AI Committee

Roles

- Design the national AI master plan and vision
- Lead planning and macro-level coordination of government-wide AI policies
- Oversee implementation, inspection, and performance management of core tasks
- Strategically allocate the national AI budget
- Advance deregulation and establish safe, ethical governance

Organizational Chart



Korean Case: AI Action Plan

Pillars

- Creation of AI Innovation Ecosystem
- Nation-wide AI Transformation
- Contribution to Global AI Basic Society

The plan presents policy recommendations and legislative tasks (20 laws, 4 enforcement decrees) to be implemented by each ministry

Number of Policy Recommendations



Legislative Tasks by Standing Committee



Global AI and Digital Knowledge Center (GAIDKC)



Purpose

Programmatic Advisory Services and Analytics anchoring global knowledge work, with embedded technical assistance in country programs.

Translate Korea's end-to-end digital experience into scalable, practical solutions for World Bank operations

Leverage the World Bank's global reach to turn proven knowledge into impact



Progress

Since launching in December 2025

Staffing underway (target: 8 staff)

Call for Proposals for Innovation Grants currently underway

Serving as a global knowledge exchange platform. Planned events in 2026 include

- Cybersecurity Academy (with KISA), July
- AI Academy, October
- Global Digital Summit, October

GAIDKC Thematic Focus

Digital Infrastructure, Digital Public Services, Inclusive Digital Economy with Responsible use of AI as cross-cutting enabler.

Areas of Work – with AI at its core

Digital and Data Infrastructure

Strengthening inclusive, affordable, and trusted digital systems

Digital Public Services

Supporting governments to deliver integrated, secure, and user-centered services

Inclusive Digital Economy

Driving innovation, entrepreneurship, and skills development, particularly for women and youth



KNOWLEDGE & ANALYTICS

Policy toolkits, diagnostics, applied research, and guidance for national digital strategies

CAPACITY BUILDING & LEARNING

Leadership programs, peer exchanges, and South-South learning that translate knowledge into practical results

CONVENING & PARTNERSHIPS

Global dialogues connecting governments, academia, private sector, and development partners

G
R
A
N
T
S

How the GAIDKC Delivers

PASA

- Programmatic Advisory Services and Analytics anchoring global knowledge work, with embedded technical assistance in country programs.

Grants

- Innovation and implementation grants linked to ongoing and pipeline lending across GPs and regions

Leveraging Partnerships

- Mobilizing Korean institutions and experts to support delivery and knowledge exchange
- Public Partners: MOEF · MSIT · NIPA · NIA · KISA · PIPC · National AI Committee, etc
- Working closely with Private Sector through strategic partnership and AI solution depository



WORLD BANK GROUP

THE WORLD BANK
IBRD · IDA

IFC
International
Finance Corporation

MIGA
Multilateral Investment
Guarantee Agency

Thank you!



Annex

Understanding differences between SLMs and LLMs helps assess when each model type should be used

Key Model Strengths		SLMs	Choosing between SLMs and LLMs	LLMs
SLMs	LLMs	Limited compute budget	How much compute is available for training and inference?	Significant compute budget
<ul style="list-style-type: none"> Resource-efficiency on specific tasks Small size and low latency contribute to faster response time Lightweight enough to be deployed locally on edge devices Composability into complex systems of multiple SLMs Can be trained to possess very specific expertise 	<ul style="list-style-type: none"> Performance on complex tasks which require nuanced reasoning Ability to generalize to new tasks or data increases model versatility Maintain contextual understanding over large inputs Better multimodal and multilingual support Current technical focus of frontier labs 	Edge devices, low-compute environments	What is the planned deployment environment?	Cloud deployment, complex infrastructures
		Narrow, specialized tasks	How complex is the task performed by the model?	Broad, general tasks
		Depth of knowledge	Does the task require more depth or breadth of knowledge?	Breadth of knowledge
		Limited amounts of very high-quality data	Do we have any major data constraints to consider?	Large amounts of data with varying quality
		Data should be stored locally to preserve privacy	Are there any privacy considerations which may restrict the flow of data?	No special privacy considerations
		Can be frequently fine-tuned or retrained	Will the model require access to frequently updated or real-time data?	Can be solved with RAG, but more complex to implement
		Real-time response needed	Is real-time or near-real-time response critical for the use case?	Tasks are not as time sensitive
		Modular deployment allow for multiple specialized skillsets	How many different skills must the solution possess?	Dependent on generalized model capabilities

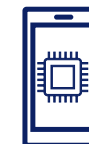
Understanding different deployment options for small AI



Cloud Servers

Processing Localization

Edge Processing



Thin-Client AI	Hybrid AI	On-Device Small AI
<p>Strengths</p> <ul style="list-style-type: none"> • Quick to deploy and runs on basic phones. • Minimal local infrastructure or specialized AI skills needed. • Ideal for rapid results and early value recognition. 	<p>Strengths</p> <ul style="list-style-type: none"> • Increases local control and relevance (e.g., adapting to crops, dialects, or guidelines). • Partially reduces dependence on centralized cloud providers. 	<p>Strengths</p> <ul style="list-style-type: none"> • Fully offline, improving resiliency and preserving privacy. • Lower ongoing operating costs once deployed.
<p>Challenges</p> <ul style="list-style-type: none"> • Dependent on internet connectivity and external providers. • Requires pre-processing and caching to preserve privacy and data security. • Costs tied to API usage and bandwidth. 	<p>Challenges</p> <ul style="list-style-type: none"> • Requires modest local data management and a small technical team. • Partially dependent on cloud infrastructure and APIs. 	<p>Challenges</p> <ul style="list-style-type: none"> • High upfront demands: local data collection, labeling, training, and packaging. • Needs mechanisms for updates, monitoring, and retraining to avoid model drift. • Hardware constraints: limited RAM, compute, and battery on affordable devices.
<p>Necessary Skills</p> <ul style="list-style-type: none"> • Mobile and web app development, API integration, data anonymization. 	<p>Necessary Skills</p> <ul style="list-style-type: none"> • Local data curation, integration of retrieval or rules engines, light server maintenance. 	<p>Necessary Skills</p> <ul style="list-style-type: none"> • ML engineering, MLOps, model optimization, and device-specific deployment.
<p>Best Used When</p> <ul style="list-style-type: none"> • Edge devices are memory-constrained • Internet connection is reliable • Workloads are large or complex 	<p>Best Used When</p> <ul style="list-style-type: none"> • User demands may be variable in their compute requirements • Local teams can optimize workloads to manage cost 	<p>Best Used When</p> <ul style="list-style-type: none"> • Low latency is critical • Edge device battery and compute capabilities are sufficient to support inference

Illustrative sector-specific pre-requisites for deploying small AI solutions in LMICs

Dimensions	Healthcare	Education
1. Skills and human capital readiness	<ul style="list-style-type: none"> Minimum viable teams for deploying a pre-built solution include: <ul style="list-style-type: none"> Software engineer to handle user interfaces, error handling, etc. Product manager for overseeing use case success, output quality, etc. DevOps engineer for hosting infrastructure, configuring model monitoring and logging, etc. Provide practitioners AI safety training using resources such as the Coalition for Health Responsible AI Guide Domain specialists who can validate model outputs for clinical accuracy 	<ul style="list-style-type: none"> Engineers and compliance teams responsible for assessing solutions which may interact with children must have a deep understanding of AI safety benchmarks, content toxicity filters, and other leading guardrails Administrators and teachers understanding of tactics to integrate language models into teaching workflows Curriculum developers recommending AI-enhanced learning pathways Local or regional support staff for troubleshooting and maintenance
2. Data readiness	<ul style="list-style-type: none"> Availability of de-identified datasets, with a clearly defined lawful basis for use in AI models Datasets which reflect country-specific case mix (e.g., co-morbidities, endemic diseases, etc.) Data standards for datasets which will be central for AI training. These standards should consider factors such as demographic data, skin tone, social determinants of health, etc. 	<ul style="list-style-type: none"> Do we have the data infrastructure necessary to deliver the requisite data to geographically disperse schools? Systems for collecting learning outcome data to measure impact Collect datasets for all languages used across the education system to promote AI tools that function at the same level for all languages
3. Connectivity and infrastructure readiness	<ul style="list-style-type: none"> Understanding of power, internet and cellular reliability, as well as the compute and data storage capabilities for hospitals and local health clinics to support informed decisions on model architectures Encryption, firewalls, and other cybersecurity minimum requirements for protecting personal health information Cellular or landline connectivity to support telemedicine integration 	<ul style="list-style-type: none"> Sufficient power supply to sustain energy flow or charge AI-enabled devices in the classroom At minimum, internet connectivity to conduct batch content downloads/uploads during periods of lower internet use (e.g., nightly) Local access networks for connecting devices to local mini-servers
4. Devices and edge readiness	<ul style="list-style-type: none"> Compatibility of AI-enabled edge devices with existing medical devices Secure endpoints for AI-enabled edge devices with role-based access and audit trails Physical security controls to protect laptops, tablets, etc. 	<ul style="list-style-type: none"> Low-cost tablets or refurbished laptops capable of displaying a web browser or basic app Plans for sharing devices when all students/teachers cannot access devices Local mini-server with at least 4GB VRAM for providing compute to edge devices Ventilation and cooling mechanisms for servers

Example - Responsible AI Governance

A shared global foundation adaptable to national contexts and development priorities



Rights & data protection

Safeguard individuals' rights and privacy across the AI lifecycle. Ensure people control how their data is used.



Transparency & explainability

Make AI decisions understandable, traceable, and open to scrutiny to foster public trust.



Accountability

Define clear responsibility for AI outcomes and establish oversight mechanisms throughout design, deployment, and use.



Inclusion & non-discrimination

Promote equitable access to AI benefits and prevent algorithmic bias or systemic exclusion.



Redress & oversight

Provide accessible remedies for harms caused by AI and ensure continuous monitoring and ethical review.

Global frameworks, including:

- the OECD AI Principles (2019)
- the UNESCO's Recommendation on the Ethics of AI (2021)
- the G7 Code of Conduct (2023)

converge around these five pillars.

LMICs can use them as a foundation for designing context-sensitive, rights-based AI policies aligned with international norms

Examples of approaches adopted by countries



Canada

Combines transparency and accountability through its *Algorithmic Impact Assessment and Directive on Automated Decision-Making*. Strong internal oversight makes this model replicable for other governments building AI governance capacity.



Korea

Integrates risk-based self-assessment into its *AI Personal Information Protection Checklist*. Shows how risk classification can operationalize ethical principles without full legislation.



Uruguay

Introduced *Algorithmic Impact Assessments* in the public sector. Demonstrates how LMICs can embed accountability mechanisms even with limited resources.



Singapore

AI Verify framework blends transparency and safety testing, offering a scalable sandbox model adaptable for other developing economies.