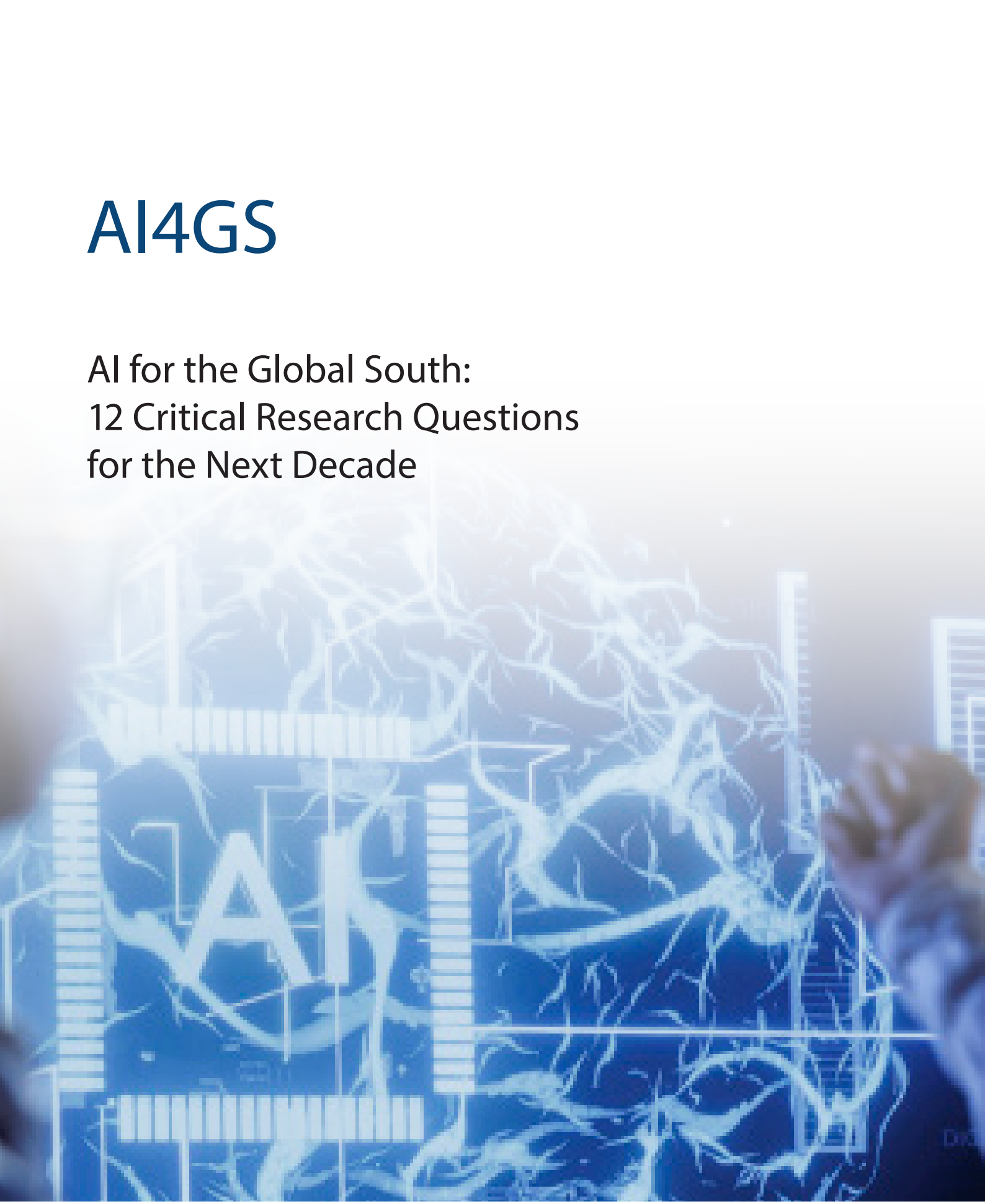
AI4GS

AI for the Global South: 12 Critical Research Questions for the Next Decade



Co-organizers

Mohamed bin Zayed University of Artificial Intelligence

المعهد الهندي للتكنولوجيا دلهي أبوظبي

INDIAN INSTITUTE OF TECHNOLOGY DELHI ABU DHABI

Endorser

AI IMPACT SUMMIT

Sponsor

Microsoft

www.x.com/mbzuai
www.facebook.com/mbzuai
www.linkedin.com/school/mbzuai
www.instagram.com.mbzuai

Contents

Find out more

Introduction

Why?

While advancements in Artificial Intelligence (AI) have made extraordinary progress, its benefits remain unevenly distributed. The research and development of AI are largely influenced by efforts concentrated in the Global North, leaving vast opportunities for the Global South[1] to take a more central role in shaping AI's future. AI for the Global South (AI4GS) was a convening of researchers who work for and with Global South communities to co-define a 10-year holistic and transdisciplinary research agenda for increasing AI's positive impact on the Global South. As an official pre-summit event of the India AI Impact Summit, AI4GS brought together leading voices from across the world, including AI researchers, social scientists, policymakers, ethicists, NGOs, and practitioners, to collaboratively design a decade-long roadmap that ensures AI development is inclusive, equitable, and responsive to real developmental needs.

How?

The AI4GS seminar was organized into thematic sessions on **technology, users, and society**, each featuring a vision talk by a field expert followed by lightning talks. We also ran a series of structured brainstorming sessions. First, participants selected their area of expertise and joined a focused discussion to identify key questions at the intersection of AI and the Global South. Next, we regrouped into interdisciplinary teams to refine and synthesize these ideas. Participants then voted to converge on a final set of twelve questions, which were further developed, articulating their significance, key challenges, potential solutions, and risks.

What?

This report lists the twelve trans-disciplinary research challenges that were identified through this exercise, and are understood to be critical for an equitable and positive impact of AI for everyone on the planet. In conclusion, we discuss the impact of this report on policy making, funding, and research investment strategies.

Who?

This report is an outcome of the AI4GS convening, which was hosted by and at Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, from 11-13th December 2025, as an official pre-summit event of the India AI Impact Summit 2026. 40+ experts (full list of contributors included at the end) attended the event and contributed to the content of this report. AI4GS was co-organized by MBZUAI and IIT Delhi, Abu Dhabi, and sponsored by Microsoft.

[1]Our framing refers to the regions, languages, cultures and contexts traditionally under-represented in mainstream AI development and research

Q1

How can human capability be built and strengthened so that the next generation is resilient and socially responsible in an AI-driven world?

AI is reshaping how people learn, work, communicate, and participate in society, with particularly profound implications for today's youth. The next generation will enter a world in which AI is not an optional tool but a pervasive infrastructure shaping economic opportunity, social interaction, and civic life. Preparing young people only with technical skills is insufficient; they must also develop the resilience to navigate uncertainty, adapt to shifting labor markets, and respond critically to rapid technological change, while remaining grounded in social responsibility and collective well-being.

The relevance of this question is especially strong in the Global South, where structural inequalities in education, infrastructure, and access to technology coexist with youthful populations and high stakes for future development. Unequal access to digital resources, gaps in AI literacy, and limited representation of local values in AI systems risk deepening existing disparities. At the same time, investments in human capacity through education, policy, and community-centered approaches offer an opportunity to enable youth to participate confidently in global AI-enabled futures without sacrificing local cultures, social cohesion, or societal priorities.

Core concepts such as "resilience" and "society-centeredness" are not universally defined or easily measurable. Resilience can refer to individual adaptability, psychological well-being, economic security, or collective capacity to respond to disruption, while society-centeredness encompasses ethical awareness, civic responsibility, and sensitivity to local and communal values. Translating these abstract qualities into educational objectives, institutional strategies, and evaluative metrics is analytically and practically challenging, particularly across diverse cultural and socioeconomic contexts. Additional barriers arise from structural and institutional constraints. Educational systems in many regions struggle with limited resources, outdated curricula, and uneven teacher training, while access to reliable internet, computing infrastructure, and AI tools remains highly unequal. Fear and uncertainty surrounding AI, driven by concerns about job displacement, surveillance, and loss of identity, can further complicate efforts to integrate AI literacy into education in constructive ways. There is also a risk that narrowly defined AI-literacy initiatives may prioritize technical proficiency over broader societal outcomes, inadvertently reinforcing technocentric or exclusionary models of progress.

A first set of research questions concerns definition and assessment: how resilience and society-centeredness should be conceptualized in the context of AI, and how individual, institutional, and societal capacity can be meaningfully evaluated. This includes developing context-sensitive frameworks and instruments to assess AI literacy, adaptive capacity, ethical reasoning, and civic orientation among youth, as well as establishing baseline measures to understand current readiness across regions. A second cluster of questions focuses on intervention and design: what educational, policy, and institutional strategies best foster resilient and society-centered human capacity. This includes examining the role of curriculum design, teacher training, AI-safety education, and tertiary-level computer science and AI programs in embedding social values alongside technical competence. Participatory and community-engaged approaches are particularly important in the Global South, ensuring that capacity-building initiatives reflect local priorities and lived realities rather than imported assumptions. Finally, this line of inquiry explores enabling conditions: the human, infrastructural, and governance arrangements required to support long-term capacity development. This includes policies for equitable access to digital infrastructure, support for educators and researchers, and coordination across education, labor, and technology policy. Success is not defined solely by higher levels of AI adoption or technical skill, but by the extent to which the next generation is equipped to thrive amid uncertainty, contribute to their societies, and shape AI-enabled futures that are socially grounded, inclusive, and resilient.

Q2 | What national, regional, and global institutional structures are required to enable AI systems for the Global South?

AI is rapidly emerging as a foundational infrastructure, embedding itself in welfare delivery, education, health systems, agriculture, and crisis response. For the Global South, this shift is uniquely consequential. The region is poised to become the dominant demographic and economic force of the twenty-first century, yet it simultaneously faces fragile political institutions, uneven digital public infrastructure, and deep socio-economic inequality. In such contexts, AI systems do not merely optimize services; they actively reshape power, access, and agency.

Institutional structures act as the binding mechanism connecting technical capability to social outcomes. Without robust governance, AI risks intensifying digital colonialism, algorithmic discrimination, and the exploitation of informal or legally unprotected populations. Effective institutions must therefore embed plural moral ecologies, rights protections, and diplomatic intelligence, ensuring that AI systems support human agency rather than automate exclusion. As AI deployment accelerates, particularly in public services, the absence of coherent national, regional, and global governance frameworks transforms innovation into systemic risk.

Key concepts such as institutional readiness, diplomatic intelligence, and plural moral ecologies are difficult to define and operationalize across diverse contexts. Institutional capacity spans legal authority, bureaucratic competence, political legitimacy, and cultural grounding, all of which vary widely within the Global South. Resource asymmetries relative to multinational technology firms and Global North standards bodies further constrain governance autonomy, while fragmentation within the Global South weakens collective negotiating power.

Domestically, AI governance may be politicized or centralized in ways that undermine accountability, and under-resourced bureaucracies may lack the expertise to oversee complex systems. There is also a risk of vendor lock-in, militarization, opaque accountability, and the marginalization of humanities-based and ethical expertise, leaving vulnerable populations exposed to harms from systems not designed for their realities.

A first research strand concerns governance architecture: how national governments can be restructured to support AI-literate civil services, participatory evaluation mechanisms, and cross-agency crisis protocols, and how regional or Global South-led governance accords can establish shared negotiating positions in global AI forums.

A second line of inquiry focuses on diplomatic intelligence and capacity-building, including embedding diplomatic intelligence into AI education, civil service training, and diplomatic practice, and benchmarking AI systems for contextual negotiation rather than narrow technical alignment. Training diplomats and policymakers in AI affairs becomes a strategic necessity, not an optional add-on.

A third area addresses rights and societal impact, encompassing AI rights charters for non-citizens and informal workers, due process and grievance mechanisms within digital public infrastructure, and metrics that track reductions in algorithmic exclusion and harm. Success is defined not by AI adoption alone, but by whether institutional trust increases and AI systems measurably advance social equity and human agency within the Global South.

Q3

How can a distributed, collaborative research network be designed to expand participation, technical capacity, and drive innovation in the Global South?

Research communities across the Global South remain structurally fragmented, both in relation to the Global North and within and across Global South regions themselves. This fragmentation limits collaboration, restricts access to resources, and significantly reduces the visibility and impact of research produced in Global South contexts, as reflected in lower citation rates and reduced participation in global scientific discourse. These dynamics are reinforced by asymmetric access to funding, compute, educational materials, and evaluation mechanisms, as well as by research agendas and framing norms that are predominantly shaped outside the Global South. A distributed, collaborative research network offers a potential mechanism to address these constraints by enabling stronger South-South connections, pooling limited resources, and creating pathways for shared identity, ownership, and long-term participation in research. Such a network is not only a means of increasing efficiency or output, but a structural intervention aimed at enabling Global South institutions and communities to define, pursue, and lead research agendas that reflect their own priorities, contexts, and needs.

Designing and sustaining such a research network presents a set of interrelated challenges. Fragmentation persists at multiple levels, driven by divergent national policies, institutional constraints, and uneven infrastructure, which complicates coordination and scaling across regions. Researchers in the Global South often operate under high teaching loads, limited funding, and restricted access to shared resources, leaving little capacity for sustained research collaboration. Language barriers and the predominance of English-language technical and educational content further restrict accessibility and participation, particularly beyond formal academic settings. Increased collaboration also raises governance challenges related to credit assignment, incentive structures, and the equitable recognition of contributions, which, if left unresolved, risk reproducing hierarchies within the Global South itself. Finally, long-term sustainability remains a central concern, as limited researcher bandwidth and reliance on short-term or externally driven funding models can undermine continuity, engagement, and collective ownership of the network over time.

A distributed research network can help address these challenges by providing shared infrastructure and governance mechanisms that lower barriers to participation while strengthening autonomy and visibility for Global South researchers. By facilitating resource pooling across compute, educational materials, and human expertise, such a network can enable research to proceed without requiring individual institutions or communities to first overcome prohibitive resource constraints. Shared platforms for discovering and disseminating research artifacts can improve visibility and reduce the marginalization of Global South scholarship, while translation of technical and educational content into languages beyond English can significantly expand access and participation. Establishing sovereign peer review processes, recognition mechanisms, and fellowships can reduce dependence on Global North validation and support sustainable research careers within the Global South. AI-supported tools may assist by reducing friction in cross-lingual communication and improving the discoverability of research outputs, but their role remains supportive rather than determinative. Ultimately, the effectiveness of such a network depends on governance structures that prioritize shared ownership, equitable recognition, and long-term sustainability, ensuring that research agendas, outputs, and benefits remain anchored in the priorities and contexts of the Global South.

Q4

How can we design evidence-based Impact Estimation methods for the Global South that can match the speed and scale of AI deployments while measuring real-world, context-specific outcomes?

AI development and deployment demand substantial financial, technical, and organizational investment, making it essential to strategically prioritize limited resources. At the same time, AI systems operate in complex sociotechnical environments: interactions between technology, local context, institutions, and human behavior can produce outcomes that are hard to anticipate. Because many AI systems are conceived and built primarily from a Global North perspective, success in Global North settings does not reliably translate to successful or equitable outcomes in the Global South. Yet evidence on AI's real-world impacts in the Global South remains extremely limited, with most studies still concentrated in the United States and Europe. Accurate impact assessment is critical to surface divergences, ensure contextual relevance, and guide responsible development, especially where misfit can amplify harm and inequality.

Evidence-based impact estimation methods such as randomized controlled trials (RCTs) typically require long timelines that don't match AI deployment cycles. Models and products often change substantially before results are available, creating a persistent drift that is compounded by this nature. In the Global South, impact assessment is further constrained by limited and fragmented data, heterogeneous and rapidly changing environments, resource and capacity limitations, and uneven technical infrastructure. Moreover, AI deployments are often controlled by private or centralized actors: who gets access, when, and which version may be opaque or adaptive, which is more consequential in the Global South due to weaker observability, fewer accountability mechanisms, and greater dependence on external platforms or vendors. Methods designed for relatively static, single-shot interventions struggle to capture the evolving and systemic effects of AI deployments, especially when systems are iterated, adopted unevenly, and embedded in broader institutions.

Classical impact evaluations (especially RCTs) tend to be strong on internal validity but weaker on external validity: similar interventions can yield widely different effect sizes across sites, making it difficult to judge transferability, especially across Global South settings with substantial heterogeneity in institutions, governance, infrastructure, and social practices.

When rigorous impact estimation is infeasible or too slow, stakeholders often skip impact estimation entirely or default to quasi-assessment proxies such as benchmarks on non-representative datasets, DAU/MAU growth and engagement, and version-to-version performance or safety improvements. These signals can be useful for product iteration but are frequently mistaken for real-world impact; they often fail to reflect outcomes for the intended target populations, who are excluded, or how benefits and harms are distributed over time. The result is a higher risk of misplaced confidence, missed harms (especially for vulnerable groups), and over-optimistic claims that can misdirect resources. Given that impact claims can shape distributional regulation, labor and welfare decisions, and infrastructure priorities, systematic overclaiming or mis-calibration can have outsized downstream consequences.

Progress requires impact assessment methods that operate on substantially shorter timelines aligned with the iterative pace of AI systems and that reduce the drift between the system evaluated and the system actually deployed. This will likely require innovations in evidence collection (e.g., low-cost digital sensing and adaptive survey methods). There is also potential to use AI itself to support estimation (for example, scenario simulations). Success also means moving beyond proxy metrics toward grounded impact claims about real outcomes for clearly defined target populations, including distributional effects and culturally grounded, relational outcomes (e.g., autonomy, dignity, and shifts in local power) where feasible. Because perfect control and full observability are unrealistic in many Global South deployments, success further includes enabling credible "shadow evaluations" and audits under partial data access, with transparent assumptions and uncertainty. Equally important is improving how evidence is communicated: transparent reporting, explicit acknowledgment of uncertainty, and clear articulation of methodological limitations to avoid overclaiming. Finally, success includes building human capacity by training researchers and practitioners to interpret imperfect evidence critically, communicate constraints clearly, and resist overstating findings, so the impact assessment ecosystem becomes more trustworthy and responsible.

Q5

How can the Global South act as a "smart latecomer" to build sovereign, locally meaningful, resource-frugal AI?

The current trajectory of AI development threatens to entrench, rather than alleviate, global economic and social inequalities. The core challenge is the exponential and prohibitive cost of frontier AI development, an upward trend that creates an "AI oligarchy" and immediately translates into a multiple-fold higher cost of adoption for the GS. This high barrier to entry results in a situation where the AI technology that is adopted is frequently subpar in quality for localized needs. It is often a repurposed, non-contextualized product not designed for the unique infrastructural, linguistic, or social realities of the GS. This systemic misalignment means that for the same investment, the Global South gains disproportionately less benefit from meaningful AI use-cases, compounding the opportunity cost. Ultimately, the prohibitive cost of AI technology for the Global South community leads to overreliance on Global North providers, which means the Global South has no agency or sovereignty over what AI it owns or truly needs, leaving its economies, cultures, and governance structures vulnerable to technologies that may be entirely misaligned or even detrimental to local cultural contexts.

Success in frugal innovation is assessed by the creation of an equitable, self-reliant AI ecosystem of Global South communities. It requires a structural shift where the Global South transitions from a passive consumer of imported intelligence to an active architect of its own digital future. In this reimagined landscape, success is defined by:

- Technological Leapfrogging: We capitalize on the "second-mover advantage" to bypass the capital-intensive trial-and-error phases of the Global North. Instead of replicating inefficient legacy paths, we treat our compute constraints as a catalyst for superior engineering and innovation, forcing the development of hyper-efficient, agile architectures that leapfrog the wasteful "brute-force" scaling era of the GN. By prioritizing the development of efficient technologies, Global South innovators can deliver global impact while securing visibility as frontrunners in the AI landscape.
- Democratized Capability: Global South Global South communities must command access to powerful, democratized AI technologies that drastically lower the barrier to entry. High capability should not require prohibitive resources, enabling widespread adoption and independent development on accessible infrastructure, thereby driving sustainable economic growth..
- Capable AI with Contextual Fidelity: The Global South must no longer be constrained by the anglocentric defaults of global AI technology. Success ensures high-fidelity, culturally relevant performance where local communities exercise absolute control over value alignment, deciding explicitly which behaviors are supported and which are rejected based on local norms.
- Sovereignty and Strategic Autonomy: The ecosystem must transition from dependency to agency. The Global South community retains full ownership of the model lifecycle, from training data to deployment, breaking the cycle of AI colonization and ensuring AI serves as a sovereign asset for local economic resilience.

Key challenges include the risk of perpetual dependency through technological lock-in, as even frugal adoption can still rely on infrastructure, operating systems, or foundation models built by first-movers, alongside a talent and skills bottleneck, since building and maintaining smaller, locally meaningful AI demands specialized expertise (e.g., distillation, efficient architectures, low-resource language processing). There is also a "good enough" quality ceiling, where prioritizing low-spec models may cap performance relative to the largest systems, and an opportunity cost dynamic in which the ease and payoff of adaptation can crowd out long-term investment in foundational AI capabilities tailored to unique Global South needs, such as new hardware-efficient architectures or novel data collection methods.

Q6

How can the Global South design and enforce data governance frameworks that give communities meaningful control across the entire data lifecycle?

Modern AI systems are data-hungry, requiring vast datasets to train frontier models. The Global South has become a key source of data due to low-cost labor for data work, multilingual populations filling gaps in existing AI, and large user bases generating interaction data. Yet the current data lifecycle is marked by deep asymmetries: low privacy protections, minimal compensation, limited control over data once produced, and little transparency into downstream use. This pattern of data colonization sees valuable resources flow outward to build models that are then sold back to originating communities at premium prices. Three crucial issues define this challenge: **agency** (the ability to make informed decisions about data and exercise meaningful consent), **utility** (ensuring AI systems serve Global South needs rather than solely benefiting external actors), and **dignity-respecting governance** (protecting privacy, providing fair compensation, and ensuring humane data work conditions). Data governance must balance the interests of multiple actors: individuals generating data through daily interactions, communities whose cultural and knowledge systems are captured in datasets, countries seeking data sovereignty, and coalitions pursuing collective bargaining power.

Global South workers increasingly train AI systems via annotation and interaction at low wages, while value accrues to Global North firms, creating an extraction dynamic. Existing power structures create enforcement gaps, as corporations and governments often have misaligned incentives. Corporations benefit from weak regulation and free data access, while governments may lack the capacity or will to enforce protections, perpetuating extractive relationships. Data work provides immediate income but at the cost of relinquishing control over cultural knowledge systems and personal information. Once extracted, regaining control becomes nearly impossible, creating path dependencies that entrench inequality. Scaling successful governance models, such as data cooperatives and stewardship structures, remains difficult. These require infrastructure, legal frameworks, and coordination that are challenging to replicate in resource-constrained settings with heterogeneous institutions. Cross-border data flows complicate matters further, as transnational platforms make it difficult for individual nations to regulate data use without coordinated multilateral agreements. Perhaps most challenging is governing existing data: retroactively applying governance to already-extracted data has no straightforward legal or technical solution. Additional challenges include limited awareness about data exploitation, weak intellectual property and labor protections, and the absence of infrastructure for transparent cataloguing and provenance tracking.

Success requires harmonizing competing priorities through participatory governance structures that respect individual consent, protect communal resources, and assert national sovereignty. Individuals and communities need meaningful control at every lifecycle stage: including fair wage negotiation, opt-out rights from exploitative processes, data deletion and revocation capabilities, and transparent tracking of data use, including cross-border deployment. Data governance should actively create value through fair compensation and economic participation in the AI value chain, investment in skills development for higher-value roles, and AI systems addressing real Global South needs rather than solely serving external markets. Success means reduced dependency, increased economic resilience, and tangible community benefits. Concretely, this includes transparent cataloguing systems, usage-based compensation, culturally aware sensitivity assessments, robust provenance tracking, and enforceable consent protocols. Only by centering agency, utility, and dignity can data governance transform extractive relationships into equitable partnerships that empower rather than exploit Global South communities.

Q7 | How can community-engaged approaches enable the identification, design, and evaluation of culturally and contextually aligned AI in the Global South?

Anthropological, social science, and psychological research demonstrates that core concepts such as self-identity, well-being, harm, and human flourishing vary substantially across cultures and communities, shaped by distinct worldviews (e.g., individualistic versus holistic, low-context versus high-context). Consequently, definitions of what is good, desirable, or harmful are inherently subjective and culture-dependent. If the overarching goal of AI is to reduce human suffering and enhance well-being, then understanding community-specific pain points, values, and priorities is essential. Yet many existing AI systems are designed from the perspective of developers rather than end-user communities, particularly in the Global South. This misalignment can result in low adoption, ineffective solutions, or unintended harm. A clear articulation and prioritization of community needs can guide more responsible and impactful AI research, design, and evaluation.

Several challenges complicate efforts toward contextual and cultural alignment. **First**, it is difficult to reach a truly representative set of community members, often leading to reliance on proxy or surrogate users rather than meaningful participation. **Second**, research and solutions may disproportionately focus on more privileged or accessible communities, leaving others excluded. **Third**, culture itself is dynamic and pluralistic; reducing it to static proxies such as geography or language risks oversimplification. Fourth, AI is a broad and evolving concept, making it unclear which methods count as "AI" and how community interaction with AI should be measured. Additional challenges include rapidly evolving community needs, limited incentives or funding to build systems aligned with identified needs, fast-changing model capabilities that affect usefulness estimates, and risks around scaling solutions without losing community agency and care. There are also ethical risks, including potential misuse of AI (e.g., for scams) and the reality that many community goals may not require AI at all. Hence, communities should be able to opt out of AI-based solutions without losing opportunities or cultural identity.

A long-term research agenda should adopt a sustained, community-engaged approach to mapping needs, solutions, and gaps. This raises a set of interrelated research questions that should be systematically answered: (i) How to elicit and understand community needs over time? Answering this requires understanding how communities define well-being, harm, and success. How their priorities shift with social, economic, and technological change, and how community-engaged methods can capture these dynamics at a grassroots level. (ii) When and why are existing technical solutions being adopted or rejected? This requires examining what technical tools communities already use, which available solutions remain unused, and how technical limitations, social norms, institutional constraints, trust, and incentives shape adoption decisions. (iii) How can alignment, safety, and evaluation frameworks reflect cultural plurality and change over time? This requires understanding alignment and safety principles (e.g., helpful, honest, harmless) through local lenses, which enables the design of community-based evaluation frameworks and the correct assessment of impact. (iv) How can communities exercise sustained agency over AI design, deployment, and governance? This requires understanding and addressing the drawbacks of current participatory processes, feedback loops, and institutional structures so that communities can shape AI-based solutions as they need them, not as developers desire. This question also intersects our research question on data sovereignty, which calls for governance frameworks that grant individuals and communities real control over how their data is collected, used, monetized, and withdrawn, including rights to economic participation and revocation. (v) How to meaningfully estimate the impact of AI-based solutions? Intersecting with our question on impact estimation, this question requires constructing success metrics that are oriented towards individual communities, rather than being universal and broad.

Q8

How can AI systems be developed so that access across languages, dialects, registers, and modalities is not a bottleneck?

Even when AI systems are contextually and culturally aligned, they often remain inaccessible due to language barriers. Contemporary AI development is overwhelmingly centered on English and sometimes, a handful of other languages, such as Mandarin for AI built in China, which function as the default interface language for most AI systems, datasets, benchmarks, and evaluation protocols. However, the majority of the world's population does not speak English, and linguistic diversity within the Global South is vast - not only across languages, but also within them. Languages differ by region, class, occupation, education, and medium (spoken vs. written), and these differences materially shape how people seek information, express uncertainty, and act on advice. Moreover, "supporting a language" is frequently conflated with supporting a standard variant of that language. In practice, this creates AI systems that are technically multilingual but functionally unusable.

Language and accessibility face a distinct set of challenges that go beyond cultural alignment. **First**, AI research disproportionately privileges written text over spoken language, despite the fact that many communities rely primarily on oral communication. **Second**, language data is treated as interchangeable across regions and populations, masking deep intra-language variation; a single language label (e.g., "Hindi" or "Bengali") obscures mutually intelligible but socially distinct forms of speech. **Third**, most benchmarks evaluate grammatical correctness or translation fidelity rather than usability, comprehension, or actionability for specific user groups. **Fourth**, language models often encode assumptions about literacy, abstraction, and formal reasoning that do not match how information is exchanged in everyday settings. **Fifth,** economic incentives strongly favor high-resource languages and large markets, leaving dialects, minority languages, and occupational registers underrepresented. **Finally**, there is a growing risk that AI-mediated services become gatekept by language proficiency, further marginalizing those who cannot or choose not to interact in dominant or "standard" linguistic forms.

Addressing language and accessibility requires rethinking language in AI as infrastructure rather than interface, and raises the following core research questions: (i) How do language use and modality vary within and across communities and domains? Answering this would require understanding the diversity of dialects, registers, and modalities across domains such as agriculture, healthcare, or public services. This would, in turn, inform AI systems to enable targeted solutions, without relying on standardized or elite language forms. (ii) How can linguistic access be defined and evaluated in terms of task completion rather than language coverage or fluency scores? Currently, most systems use language coverage as a metric of accessibility. It seldom analyzes the scale of benefit as a success metric. After understanding the diversity, evaluation schemes should couple language coverage and the scale of benefit from using AI-based solutions as a composite success metric for linguistic access. (iii) Can language technology be purely built from speech data without any text-based interfaces? Since language literacy and lack of text-based data are primary bottlenecks for low-resource languages, investigating architectures and training paradigms that can directly leverage speech and other modalities might be beneficial. (iv) How can participatory data stewardship give communities control over how their languages and speech are represented, shared, and reused in AI systems? This intersects our research question on data sovereignty, which focuses on exploring data governance frameworks that give individuals meaningful control, economic participation, and revocation rights over their data. (v) Finally, how to ensure AI-mediated solutions do not become a gatekeeper to essential services? Technological change should be participatory and not enforced. Hence, careful oversight and planning are required to ensure that, on one hand, AI-mediated language access improves comprehension, trust, and sustained use, and on the other, preserves non-AI pathways so that language does not become a gatekeeper to essential services.

Q9

How should AI infrastructure be designed for system-level decision-making across the Food-Water-Energy-Climate-Health (FWECH) nexus in the Global South?

AI is increasingly positioned as a vital tool for decision-making across domains such as agriculture, water management, energy systems, climate adaptation, and public health. In the Global South, where climate risks are intensifying, energy access remains uneven, and structural inequalities persist, the effectiveness and legitimacy of AI systems depend fundamentally on their ability to reflect local environmental conditions, infrastructural realities, and social contexts. The Food-Water-Energy-Climate-Health (FWECH) nexus provides a framework for understanding these challenges, as an imbalance in one domain routinely cascades into others. This produces risks that cannot be addressed through isolated interventions. Shifts in water availability affect food security and health outcomes; energy poverty constrains healthcare delivery and climate resilience; and climate extremes simultaneously disrupt agricultural production, energy infrastructure, and public health systems. In such settings, AI systems not trained for the Global South-specific ecosystem often fail to generalize to tropical, data-sparse regions.

Developing AI systems for the FWECH nexus in the Global South faces significant structural, scientific, and governance-related challenges. Data scarcity remains a fundamental constraint, as observation networks for rainfall, soil moisture, river flows, air quality, health surveillance, and energy usage are often fragmented and undermine model reliability and robustness. Infrastructure limitations further restrict the sustained operation of AI tools. These challenges are compounded by the siloed organization of government agencies and sectoral programs, which inhibits the data integration necessary to model cascading risks across the nexus. Additionally, reliance on non-region-specific AI models, frequently trained on Global North data, brings algorithmic bias and poor generalization. This fails to account for the realities of smallholder agriculture, decentralized energy systems, local disease vectors, and informal socio-economic structures.

Addressing these challenges requires a purpose-built approach to AI development for the FWECH nexus that prioritizes contextual relevance, system integration, and participatory governance. AI systems should be designed to operate effectively under sparse, noisy, and incomplete data conditions, while supporting the integration of information across food, water, energy, climate, and health domains to detect and anticipate cascading risks. Emphasis should be placed on developing interoperable datasets, decision-support tools that function in low-bandwidth and low-infrastructure environments, and models that reflect local environmental dynamics and social realities.

At the system level, success depends on aligning AI development with institutional coordination across sectors, enabling integrated responses to climate extremes, resource stress, and public health emergencies. Governance frameworks that involve stakeholders across the FWECH nexus are essential to ensure that AI tools are trusted, equitable, and responsive to local priorities. In this framing, AI serves not as an autonomous solution, but as an enabling infrastructure that supports adaptive decision-making, resilience, and sustainability across interconnected systems in the Global South.

Q10

What interventions can improve Global South labor market outcomes for informal workers and reduce AI-driven inequality?

Studying the impact of AI on labor markets is critical because labor is the primary asset of the poor in the Global South, making it a first-order concern for living standards, poverty, and income inequality. Labor markets in the Global South are uniquely different from those in the Global North in terms of their informality. Over two-thirds of the labor force in the Global South (and >80% in the poorest countries of the Global South) is employed in the informal sector, generally meaning low productivity, low wages, no job security, and no benefits (such as health or unemployment insurance, pension, or disability protection). AI also risks substituting traditional styles of informal work with new types of informal work, like the gig economy and delivery jobs. On the positive side, AI could create new opportunities for development, facilitate job creation, and enhance the skills and productivity of frontline workers. Conversely, the adoption of AI may also suppress economic growth by substituting existing jobs, intensifying the migration of digital talent to the global north, or leading to the exploitation of digital workers.

A core challenge is that while AI is widely expected to substitute for some human labor, the scale and pattern of displacement will differ sharply by task type (routine vs. non-routine; manual vs. cognitive), making aggregate predictions unreliable. Another challenge is distributional uncertainty: because workers are unevenly concentrated across task types by age, gender, education, and caste/ethnic/tribal status, AI could produce very different job losses across identity groups. A third challenge is cross-country heterogeneity: since informality varies widely across Global South labor markets, AI-driven displacement may look fundamentally different across settings (e.g., Indonesia vs. Uganda; South Africa vs. Bolivia), limiting the transferability of findings. Finally, there is the challenge of inequality dynamics: evidence from the Global North suggests AI can widen wage gaps by rewarding skilled workers faster than others, but it is unclear whether the same pattern will hold in Global South contexts. Accurate measurement of AI-induced job switching and new job creation may be difficult, as new jobs may be created that we are currently unaware of.

Current labor laws are not equipped for a world of "algorithms as employers," where workers face AI-driven income volatility, opaque and potentially unfair rating and pricing systems, and a lack of basic protections like health insurance and pensions, pointing to the need for a new legal category for platform and informal digital workers with enforceable rights and benefits. In parallel, policy must strengthen digital inclusion by expanding access to smartphones and connectivity while investing in locally grounded digital skills, since device access alone does not translate into real capacity.

Effective regulation should also require algorithmic transparency and fairness so workers can understand how pay and ratings are computed and be protected from discriminatory allocation practices. Finally, education systems should be redesigned to preserve and amplify uniquely human skills such as critical thinking, creativity, and judgment to reduce long-run cognitive erosion and future brain drain. If not managed well, AI has the potential to retard economic growth and development in the Global South beyond job substitution, for example, via import dependence, profit repatriation, and premature deindustrialization. A desirable future is therefore one where AI helps the Global South move from "cheap labor" to dignified work with fewer extractive practices and stronger institutions that phase out the least dignified jobs. AI must also strengthen existing jobs and services by boosting the effectiveness and productivity of frontline workers and other service providers, increasing benefits for end users and society. Over time, these shifts could help reduce global inequality, not by slowing Global North growth, but by enabling faster, broader-based gains in the South.

Q11

How does the deep integration of AI systems into emotional, social, political, and economic domains reshape human agency, cultural identity, and meaningful civic participation?

AI systems are no longer confined to instrumental or background functions; they are increasingly positioned as socially embedded actors that mediate knowledge, provide emotional support, influence decision-making, and, in some cases, substitute for human relationships and institutions. As AI becomes entwined with processes of meaning-making, social interaction, and governance, it has the capacity to subtly but profoundly reconfigure foundational aspects of what it means to act, choose, relate, and participate as a human being. The relevance of this inquiry is particularly acute for the Global South, where historical power asymmetries, cultural marginalization, and institutional fragility intersect with rapid technological adoption. AI systems trained predominantly on Global North data and value systems risk misrepresenting lived realities, eroding linguistic and cultural diversity, and reinforcing existing hierarchies of knowledge and authority. At the same time, AI holds significant potential to expand access to information, strengthen civic participation, and compensate for gaps in education, healthcare, and public services. Addressing this question, therefore, enables the development of culturally grounded, human-centered approaches that ensure AI serves as a tool for empowerment rather than exclusion, and that it contributes to social resilience and equity rather than exacerbating vulnerability.

This is a particularly challenging problem because its core constructs—human agency, identity, cultural voice, and civic participation—are inherently multidimensional, context-dependent, and socially constructed. Their meanings vary across cultures, communities, and generations, making it difficult to develop analytical frameworks that are both rigorous and culturally sensitive. Moreover, AI's role in emotional and relational life is evolving at a pace that outstrips theoretical consolidation, resulting in objects of study that are fluid, unstable, and continuously reshaped by emerging technological affordances. Methodological complexity further compounds these difficulties. Many of the most consequential impacts of AI manifest at the level of subjective experience, relational dynamics, and subtle shifts in dependence or trust, which are difficult to observe or measure at scale. Behavioral data related to AI use are often opaque, proprietary, or decontextualized, while conventional computational methods struggle to capture cultural nuance, idiomatic expression, and situated meaning. There is also a nontrivial risk that research itself may reproduce harm: by flattening intra-regional diversity, reinforcing deficit-based narratives about Global South communities, or generating findings that could be appropriated to legitimize surveillance, techno-solutionism, or extractive market expansion.

One core set of research questions concerns representation and voice: how AI systems shape whose perspectives are amplified, distorted, or silenced, and how Global South cultural, linguistic, and epistemic traditions can be meaningfully reflected in AI outputs. Addressing these questions requires participatory and community-led approaches to data collection, evaluation, and benchmarking, alongside interdisciplinary collaboration with cultural studies, anthropology, and indigenous knowledge systems to define contextually appropriate standards of fidelity and respect. A second cluster of questions focuses on agency and autonomy, particularly how reliance on AI for emotional support, education, healthcare guidance, or decision-making influences perceptions of self-determination and free will in contexts with limited regulatory safeguards. Mixed-method and longitudinal research designs—integrating insights from psychology, human–computer interaction, and digital sociology—are essential to disentangling algorithmic influence from broader structural conditions such as economic precarity, institutional absence, and social inequality. A third line of inquiry examines civic participation and democratic processes, including how AI-mediated information ecosystems, surveillance practices, and misinformation dynamics reshape public discourse and political engagement in societies already facing structural inequities. Progress in this area depends on collaboration with political science, education, and civil society actors to develop governance frameworks, AI literacy initiatives, and participatory design models that protect autonomy, cultural integrity, and democratic resilience.

Q12

What science and systems are needed to meaningfully disrupt and reshape AI development?

Despite great efforts toward safe, ethical, responsible, or value-aligned AI, there is no guarantee that future AI systems will, in fact, be aligned to these goals. There are suggestions from history, political economy, and critical theory that the entities that are likely to unleash the most powerful, most impactful AI systems on the world are driven primarily by profit and accumulation of power for themselves, not the benefit of humanity as a whole, and especially not the Global South. Meanwhile, the world's top AI experts have expressed concerns that AI has the potential to end human civilization, cause nuclear-level catastrophes, and increase the economic divide. In contexts such as these – when AI is either imposed top-down in a way that is harmful, or when AI goes "rogue", and causes harm that even its owners/creators did not intend, an ethical response could be to disrupt AI, that is, limit AI harm once a harmful AI system is deployed. Such harms have been shown to affect more vulnerable and marginalized communities. Disruption might involve disseminating viruses that shut down AI systems; interfering with or manipulating AI function; sabotaging AI training data through the insertion of dirty data; mass user action to confuse an AI system; and so on.

Relevant work along these lines is already underway, even if it is not consistently framed as AI disruption. There is, for instance, substantial work on adversarial attacks on AI systems and jailbreaking methodologies. Another adjacent area is data poisoning, where strategically introduced training data alters a model's behavior in subtle or significant ways. Such techniques are often studied defensively but could, in principle, be deliberately employed to undermine AI systems during development or deployment, as in the case of Red-teaming.

In parallel, community-led AI auditing approaches seek to involve marginalized communities directly, equipping them with the capacity to identify and document how AI systems negatively affect their lives and social environments. These practices overlap with traditions of ethical hacking, where digital systems are disrupted not for personal gain but to advance socially constructive or justice-oriented objectives. From this perspective, disruption itself can be understood as one legitimate mode of resisting harmful AI systems.

Although many of these disruption-oriented methods already exist in specific domains, a key challenge lies in extending and future-proofing them for use by marginalized and vulnerable communities at larger scale. Doing so would require not only technical adaptation but also the cultivation of a sustained scientific and civic community focused on AI disruption, contestation, and reshaping in service of social equity and collective agency.

As with any science, disruptive AI benefits from a community of researchers and practitioners moving the science forward. Potential avenues for building such a community could include creating research conferences or journals, teaching university courses, training lay people on AI disruption methods, and informing citizens about the dangers of closed AI systems, all to make disruption methodologies widely known so that people can counteract harm caused by AI for their own purposes. The social impact of AI disruption research would become evident in any context in which a community or population experiences significant AI harm and takes steps to reduce the harm through technologies or techniques based on disruption research.

Efforts to disrupt AI will always be a "cat and mouse" game in which disruption efforts prompt AI system owners to address and protect against those disruptions, at which point new forms of disruption will need to be identified for effective disruption, and so on in an endless cycle. An ongoing challenge is developing disruption methods legally and safely - it would be counterproductive if disruption techniques are used to harm, or if they interfere with the ethical operation of AI. Also, as disruption is unwanted by AI system creators and owners, they may seek to suppress the development of disruption methodologies. Finally, because the vast majority of capital in AI is directed toward increasing commercial ROI, independent disruption research and outreach is left underfunded, with Global South researchers lacking the massive compute power and financial backing necessary to stress-test the very systems they aim to keep in check.

Conclusion

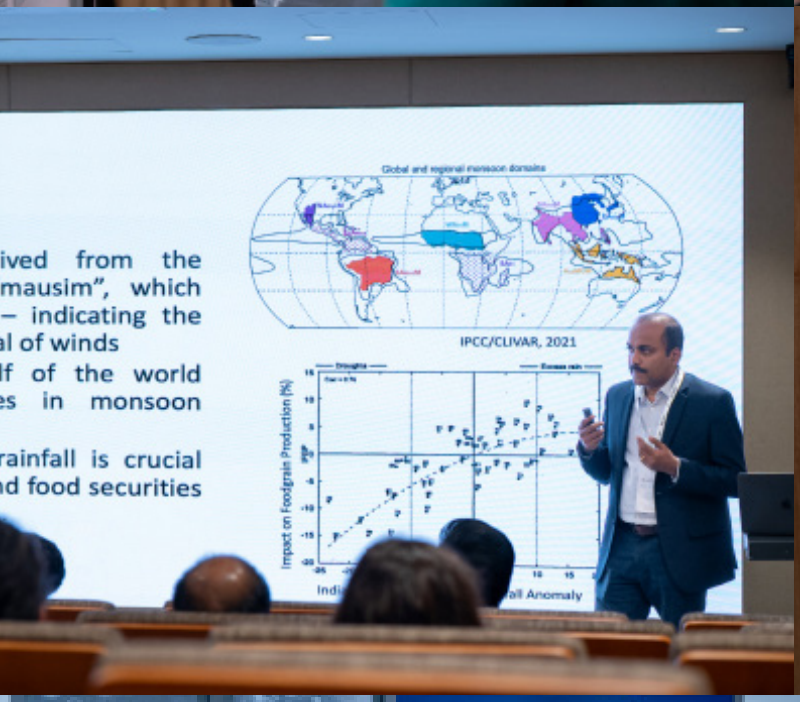
This report synthesizes a 10-year holistic, transdisciplinary research agenda for increasing AI's positive impact on the Global South, developed in collaboration by leading AI researchers, social scientists, policymakers, ethicists, NGOs, and practitioners. Together, the twelve questions encapsulate a roadmap for AI that advances human agency, equity, and locally grounded development in the Global South by addressing the intertwined challenges of technology, institutions, communities, and real-world impact.

For **policymakers**, this document offers a shared vocabulary for institutional readiness, rights protections, language access, and impact estimation, enabling governments to move beyond importing Global North frameworks toward governance that reflects local realities. In practice, it can inform national AI strategies, public-sector procurement standards, digital public infrastructure roadmaps, and accountability mechanisms.

For **funders and science investment leaders**, the report serves as a guide for prioritizing research and infrastructure investments over the next decade. It clarifies which research bets are foundational, such as impact measurement capacity, multilingual and multimodal access, data governance, and institution-building. It strengthens the case for sustained investment in public goods, including datasets, evaluation tools, community-engaged protocols, training programs, and regional research networks, that reduce dependence and enable locally owned innovation.

For **NGOs and civil-society organizations**, the document provides both an advocacy framework and practical guidance for implementation. It treats community-defined needs, participatory evaluation, and culturally grounded definitions of harm and benefit as core requirements rather than optional additions. It also supports coalition-building across sectors so communities can shape what is built, how it is monitored, and what success should mean in their contexts.

For **scientists and academia**, the report offers ideas for a transdisciplinary socio-technical research program. It highlights methodological gaps where academic leadership is essential, including fast yet credible impact estimation aligned with iterative deployments, longitudinal and drift-aware evaluation, frugal and efficient model architectures, and participatory methods that remain rigorous at scale. This agenda can strengthen Global South leadership in defining research questions and standards that evaluate AI not only by capability, but by equitable, measurable impact.


ived from the
mausim", which
– indicating the
l of winds
lf of the world
es in monsoon
rainfall is crucial
d food securities
Global and regional monsoon domains
IPCC/CLIVAR, 2021
Impact on Foodgrain Production (%)


environment
quantify im
mitigate
labour markets
18 - educati

Contributors

Monojit Choudhury | MBZUAI
Sunayana Sitaram | Microsoft Research India
Aditya Vashistha | Cornell University
Elizabeth Churchill | MBZUAI
Fajri Koto | MBZUAI
Sunipa Dev | Google
Hariprasad Kodamana | IIT Delhi Abu Dhabi
Amr Keleg | MBZUAI
Ananya Agrawal | MBZUAI
Aseem Srivastava | MBZUAI
Gayatri Bhat | Karya Inc
Madhur Jindal | MBZUAI
Muhammad Dehan Al Kautsar | MBZUAI
Prashant Kodali | Microsoft Research India
Salsabila Zahirah Pranida | MBZUAI
Sougata Saha | MBZUAI
Sourabrata Mukherjee | Microsoft Research India
Alham Fikri Aji | MBZUAI
Anil B. Deolalikar | UC Riverside
Bashar Alhafni | MBZUAI
Boulbaba Ben Amor | Inception
Chetan Arora | IIT Delhi
Danish Pruthi | Indian Institute of Science
Howard Lakougna | Gates Foundation
Syed Ishtiaque Ahmed | University of Toronto
Joyojeet Pal | University of Michigan
Julia Kreutzer | Cohere
Kalika Bali | Microsoft Research India
Kentaro Toyama | University of Michigan
Luciana Benotti | Universidad Nacional de Córdoba and NGO Fundación Vía Libre
Luis Chiruzzo | Universidad de la República Uruguay
Lyantoniette Chua | Center for AI and Digital Policy
Melissa Densmore | UCT South Africa
Munmun De Choudhury | Georgia Tech
Parag Singla | IIT Delhi
Salman Khan | MBZUAI
Sandeep Sukumaran | IIT Delhi
Talal Rahwan | NYU Abu Dhabi
Tatapong Beyala | UNIVERSITY of Yaounde
Thamar Solorio | MBZUAI
Vukosi Marivate | University of Pretoria and Lelapa AI
Wessam Ezzeddine | AI Safety UAE
Wiebke Denecke | Massachusetts Institute of Technology
Yasir Zaki | NYU Abu Dhabi

AI4GS

Find out more

Co-organizers

Mohamed bin Zayed University of Artificial Intelligence

المعهد الهندي للتكنولوجيا دلهي أبوظبي

INDIAN INSTITUTE OF TECHNOLOGY DELHI ABU DHABI

Endorser

AI IMPACT SUMMIT

Sponsor

Microsoft