TOWARDS EQUITABLE POST-AGI FUTURES: THE IMPERATIVE FOR MULTIMODAL AI ASSURANCE IN THE GLOBAL SOUTH

Anonymous authors
Paper under double-blind review

ABSTRACT

The rapid advancement of innovations towards Artificial General Intelligence (AGI) and advanced multimodal AI systems marks a transformative shift in computational capabilities, with profound implications for global equity, governance, and societal development. While AGI research has traditionally focused on technical benchmarks, its impending deployment in the Global South introduces complex challenges that necessitate a fundamental rethinking of AI assurance frameworks. The Multimodal AI Assurance framework, introduced in this paper, integrates technical robustness, participatory governance, and adaptive policy mechanisms to ensure that AGI systems are equitable, transparent, and accountable in low-resource and culturally diverse settings. We detail the technical and policy dimensions of MAIA, providing a roadmap for researchers, policymakers, and practitioners to operationalize equity-centered AGI assurance.

1 INTRODUCTION

The emergence of Artificial General Intelligence (AGI) and advanced multimodal AI systems represents a paradigm shift in computational capabilities, with profound implications for global equity and governance (3). While AGI research has historically prioritized technical benchmarks, such as scalability, generalization, and computational efficiency, its deployment in the Global South introduces a complex array of challenges that demand a fundamental rethinking of AI assurance frameworks (2; 11). The Global South, encompassing regions in Africa, Latin America, the Caribbean, Asia (excluding high-income economies), and Oceania, faces structural inequities in data representation, infrastructure disparities, and governance gaps that exacerbate the risks of AGI deployment. For instance, less than 1% of globally labeled AI training data originates from Africa, despite the continent representing 17% of the world's population. This data scarcity leads to algorithmic bias, where AGI systems trained on Western-centric datasets perform poorly or introduce harmful biases when applied to diverse populations (4), an effect that will likely snowball with the emergence of post-AGI agents, which will most certainly exude precursors of bias from pre-AGI societies in the absence of active corrective interventions (12). Additionally, infrastructure constraints, such as limited internet connectivity (40% in Africa vs. 93% in developed nations) and unreliable electricity access (600 million people in sub-Saharan Africa lack electricity), further complicate the AGI-society dynamic. Without deliberate interventions, AGI risks becoming another tool of digital colonialism, entrenching existing inequalities and eroding public trust (8; 13).

The Multimodal AI Assurance (MAIA) framework, introduced in this paper, is designed to address these challenges by integrating technical robustness, participatory governance, and adaptive policy mechanisms to ensure that AGI systems are equitable, transparent, and accountable in low-resource and culturally diverse settings. MAIA is structured around four core pillars: **equity**, ensuring AGI systems are inclusive, representative, and accessible to marginalized communities; **transparency**, providing contextual explainability tailored to local sociocultural norms; **adaptability**, developing modular approaches that evolve with community needs; and **accountability**, establishing community-driven oversight mechanisms to monitor and mitigate harms. This paper expands on the technical and policy dimensions of MAIA, providing a detailed roadmap for researchers, policymakers, and practitioners to operationalize equity-centered AGI assurance in the Global South.

2 TECHNICAL CHALLENGES OF AGI IN THE GLOBAL SOUTH

Representational Bias AGI systems rely on large-scale, diverse datasets to achieve generalization and robustness. However, the Global South is severely underrepresented in global AI training data, leading to systemic biases in AGI applications (10; 7). For example, facial recognition systems exhibit error rates up to 100 times higher for darker-skinned women compared to lighter-skinned men. Similarly, medical imaging tools trained on Western datasets misclassify skin lesions in Black patients at twice the rate of white patients in African clinics. Voice recognition systems also fail to recognize indigenous languages, with over 80% of voice datasets derived from English, Mandarin, or European languages. To mitigate these biases, technical interventions such as federated learning can enable models to be trained across decentralized devices, such as clinics, farms, or community centers, without raw data leaving local control. Synthetic data augmentation techniques can generate culturally and environmentally diverse datasets to supplement scarce real-world data. Additionally, transfer learning with local fine-tuning allows pretrained models, such as large language models, to be fine-tuned on local datasets. All of these are currently available paradigms even in the pre-AGI era, and their role in the continual optimization of post-AGI artefacts would be indispensable.

Infrastructure Constraints AGI systems often require high computational resources, which are unavailable or unaffordable in many Global South contexts. Cloud-based tools are often inaccessible due to low internet penetration (e.g., less than 20% in rural Africa), while edge devices such as smartphones or Raspberry Pi may lack the processing power for real-time inference. Energy instability further disrupts applications requiring continuous power, such as disaster prediction or telemedicine. To address these constraints, model compression and quantization techniques can reduce model sizes and computational requirements. For example, quantized neural networks can run on low-cost devices with minimal performance loss, enabling deployment in resource-limited settings. Edge AI deployment allows models to operate offline or in low-connectivity environments. Hybrid cloud-edge architectures combine cloud-based training with edge deployment to balance performance and accessibility. For example, an AGI system for flood prediction could use cloud-based training but deploy lightweight models to local weather stations, ensuring real-time functionality without relying on continuous connectivity.

Ethical and Societal Risks AGI introductions in the Global South raises profound ethical concerns, including surveillance and human rights violations, exclusion of marginalized groups, and erosion of indigenous knowledge. For example, AGI-powered facial recognition can be used to target political dissidents in some countries with a far greater power imbalance in favor of authoritarian setups. Similar instances of current societal risks are likely to be significantly exacerbated in a post-AGI world. To mitigate these risks, bias audits and fairness constraints should be integrated into present day algorithm development and models should undergo pre-deployment bias audits using metrics such as demographic parity, equalized odds, and counterfactual fairness to ensure equitable outcomes. Participatory design involves end-users, such as farmers, healthcare workers, and indigenous communities in co-creating machine intelligence systems to ensure enduring cultural relevance.

3 THE MAIA FRAMEWORK: TECHNICAL AND POLICY DIMENSIONS

3.1 EQUITY: INCLUSIVE AGI DESIGN

In a post-AGI future, an optimistic view could be the aspiration to achieve equity using AGI systems and their downstream effects. Multilingual and multimodal AGI systems should support indigenous languages and dialects in voice and speech recognition. It may be reasonable to assume that general intelligence model ecosystems would also spawn 'shallow intelligence' models for specific aspects (15). For instance, an AGI-powered educational chatbot in Peru could integrate Quechua and Aymara alongside Spanish, ensuring accessibility for indigenous communities. Low-resource optimization, such as model pruning, quantization, and distillation, allow AGI models to deploy on edge devices. For example, a quantized downstream model for soil analysis could run on a $50 smartphone, making it accessible to smallholder farmers. Governments should mandate diverse training data for AI systems, requiring developers to disclose dataset demographics and ensure representation of marginalized groups. Public funding for local research can establish national AI Safety and Alignment institutes, such as Nigeria's National Centre for AI, to develop culturally relevant

responses to societal questions around emerging AGI. Data sovereignty laws should be extended to AGI contexts to protect local data from extraction by foreign corporations.

3.2 Transparency: Contextual Explainability

Localized XAI interfaces provide explanations in local languages and cultural contexts, similar in intent to current explainability ideas but significantly enhanced with advanced reasoning on world models. Participatory explanation design involves community health workers in designing AGI explanations. For instance, a malaria detection AGI could explain predictions using local symptom descriptions, improving usability and trust. This can pave the way for audit trails for AGI decisions to log reasoning and chains of thought for post-hoc review to enable accountability (5). Right to explanation laws should require AGI developers to provide understandable explanations for high-stakes decisions, such as healthcare or finance. AGI literacy programs can train public officials, healthcare workers, and farmers in interpreting AGI outputs, ensuring informed usage. Independent AGI audits, conducted by third-party bodies should verify transparency and fairness.

3.3 Adaptability: Dynamic AGI Systems

Continual learning allows AGI models to update incrementally with new data. For example, an AGI weather predictor in Bangladesh could adapt to changing climate patterns, improving accuracy over time. Modular AGI architectures enable component-wise updates, such as swapping a crop disease classifier without retraining the entire system, reducing computational costs. Robustness testing evaluates AGI performance under real-world variability, such as dialectal speech or low-light medical images, ensuring reliability in diverse conditions. AGI sandbox regulations allow controlled testing of adaptive general models in real-world settings, such as curated AI sandboxes for healthcare and similarly sensitive applications.

3.4 Accountability: Community-Driven Oversight

Community Oversight Boards (COBs) empower marginalized groups to monitor AGI deployments. For example, a COB in Brazil's favelas could audit police AGI tools for racial bias, ensuring fairness. Harm mitigation protocols enable real-time AGI corrections, such as halting a biased loan AGI. Legal liability frameworks hold AGI developers financially and legally accountable for harms, such as misdiagnoses or financial exclusion. COBs should be formalized as statutory bodies with enforcement powers with legal and constitutional safeguards prioritizing human individual and collective interests. Multilateral harm compensation funds and arbitration mechanisms can compensate victims of AGI failures or downstream catastrophic risks. Cross-border AGI accountability enables Global South countries to sue foreign developers for harms, similar to EU-style GDPR fines.

4 Policy Roadmap for MAIA Implementation

National-Level Reforms At the national level, governments should enact equity-centric AGI laws that mandate bias audits for high-risk AGI applications, such as healthcare and criminal justice. Public AGI infrastructure, such as national AGI labs or AI Safety Institutes, can develop localized tools tailored to regional needs of model monitoring and safeguards provisioning (6). AGI workforce development programs, including university partnerships and vocational training, can build local capacity in AGI ethics and auditing.

Regional and Global Coordination Regional bodies, such as the African Union (AU) and ASEAN, should establish cross-border AGI governance frameworks, borrowing from the last century of multilateral experiments on cross-border regulation of high risk technologies and their interface with the global commons (1). A global AGI Governance Framework could include, for example, a supranational AGI Ethics Board to harmonize standards and a cross-border federated learning initiative for healthcare and agriculture. Simiarly, regional bodies like the ASEAN can leverage localised Digital Economy Frameworks could implement joint AGI audits and a regional AGI sandbox for testing adaptive models. Globally, the UN Global Digital Compact (14) and AI-specific derivatives should create a Global South AGI Fund to support local AGI research and binding AGI human rights standards, such as the right to contest AGI decisions.

Private Sector Accountability The private sector must be held accountable through AGI developer licensing, requiring mandatory equity audits and revoking licenses for repeated bias or harm. Corporate AGI transparency should mandate public disclosure of training data demographics and open-source AGI models for critical sectors, such as healthcare. This should include developers of large foundation models or world models that may exude general intelligence capabilities, but not be limited to them. AGI harm liability laws should enable class-action lawsuits for discriminatory AGI and fines proportional to corporate revenue, similar to GDPR and the EU AI Act penalties.

5 Conclusion: Ways forward

As AGI systems edge closer to surpassing human-level capabilities, their deployment across Africa, Latin America, Asia, and other low-resource regions will not only reshape local economies, healthcare, and governance but will fundamentally redefine the geopolitical and ethical landscape of the 21st century. The decisions made today—whether to prioritize equity, transparency, and community-driven accountability or to default to centralized, extractive models of AGI development—will determine whether this technology becomes a force for emancipation or subjugation, for democratization or digital colonialism (9). Without deliberate intervention, AGI risks replicating and amplifying the very inequities it claims to solve, entrenching surveillance states, deepening economic exclusion, and eroding cultural sovereignty, particularly in regions already grappling with infrastructural fragilities, data asymmetries, and weak governance. MAIA offers a blueprint for resistance and reformation, advocating for a future where AGI is decentralized, participatory, and aligned with the needs of marginalized communities, but this vision demands urgent action. In the medium term, if we fail to institutionalize frameworks like MAIA, we risk locking the Global South into a dependency paradigm, where AGI systems are designed mainly in the Global North, deployed without due consent, and governed without accountability, turning these regions into mere testing grounds for automated decision-making that benefits distant elites. The alternative, a future where AGI is treated as a public good, governed by principles of equity, transparency, and collective stewardship, requires a fundamental shift in how we develop, deploy, and regulate this technology. This means polycentric AGI innovation, with regional hubs in Africa, Latin America, and Asia leading the creation of culturally attuned systems; participatory governance through Community Oversight Boards that empower farmers, healthcare workers, and indigenous leaders to audit and shape AGI; and a restorative justice approach, where AGI actively compensates for historical injustices by prioritizing underserved populations, amplifying indigenous knowledge, and democratizing access to multilingual, context-aware tools. Achieving this future will not be passive, and requires structured South-South collaboration, where countries share data and models without Northern intermediation, reform of global governance bodies to center Global South leadership, and the establishment of a new multilateral framework, a ”Bretton Woods for AGI”, that prioritizes sovereignty over surveillance and collective benefit over profit. The risks of inaction include a dystopian default where AGI systems, trained on biased data and deployed without oversight, lead to widespread misdiagnoses, financial exclusion, and state-sponsored suppression, while democracies erode under algorithmic governance and climate adaptation is hijacked by profit-driven monocultures. Will we allow AGI to deepen inequality and entrench oppression, or will we seize this moment to build a future where AGI serves the many, restores rather than extracts, empowers rather than controls, and unites rather than divides? The future of AGI is not predetermined; it is ours to shape, and the stakes are high.

Table 1: Comparison of MAIA Framework vs. Current AGI Systems

Pillar	**Current AGI Systems**	**MAIA Vision**
Equity	Biased datasets, Western-centric training, exclusion of marginalized groups	Federated learning, local data sovereignty, multilingual/multimodal AGI
Transparency	Black-box models, technical jargon, no recourse for users	Contextual XAI, participatory design, audit trails for decisions
Adaptability	Static models, high resource demand, cloud dependency	Continual learning, modular architectures, robustness testing
Accountability	No accountability, corporate self-regulation, weak enforcement	Community Oversight Boards, harm mitigation protocols, legal liability

REFERENCES

[1] Tom Everitt, Gary Lea, and Marcus Hutter. Agi safety literature review. *arXiv preprint arXiv:1805.01109*, 2018.

[2] Eugenio V Garcia. The technological leap of ai and the global south: Deepening asymmetries and the future of international security. In *Research Handbook on Warfare and Artificial Intelligence*, pages 370–387. Edward Elgar Publishing, 2024.

[3] Jerome C Glenn. Why agi should be the world's top priority. *Horizons: Journal of International Relations and Sustainable Development*, (30):86–95, 2025.

[4] Nima Kordzadeh and Maryam Ghasemaghaei. Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3):388–409, 2022.

[5] Phoebe Koundouri, Fivos Papadimitriou, Georgios Feretzakis, Theodoros Daglis, and Vera Alexandropoulou. Ai policies towards the agi challenge: An international assessment. Technical report, Athens University of Economics and Business, 2025.

[6] Spyros Makridakis and Maria Michailides. The forthcoming agi revolution: Its impact on society and firms. 2025.

[7] Helya Modarresi Tehrani. What would be the impact of agi on society with a focus on un's sustainable and development goals? 2025.

[8] Arthur GO Mutambara. *Artificial Intelligence: A driver of inclusive development and shared prosperity for the Global South*. CRC Press, 2025.

[9] Wim Naudé and Nicola Dimitri. The race for an artificial general intelligence: implications for public policy. *AI & society*, 35(2):367–379, 2020.

[10] Chinasa T Okolo. Ai in the global south: Opportunities and challenges towards more inclusive governance. *The Brookings Institution*, 2023.

[11] Marie-Therese Png. At the tensions of south and north: Critical roles of global south stakeholders in ai governance. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 1434–1445, 2022.

[12] Raghu Raman, Akshay Iyer, and Prema Nedungadi. Forecasting artificial general intelligence for sustainable development goals: a data-driven analysis of research trends. *Sustainability*, 17(16):7347, 2025.

[13] Shahana Rayhan. Ethical implications of creating agi: impact on human society, privacy, and power dynamics. *Artificial Intelligence Review*, 11(2):44–59, 2023.

[14] Allison Wylde. The un global digital compact (gdc), achieving a trusted, free, open, and secure internet: Trust-building. In *European Conference on Cyber Warfare and Security*, pages 544–XXIII. Academic Conferences International Limited, 2023.

[15] Gokul Yenduri, Ramalingam Murugan, Praveen Kumar Reddy Maddikunta, Sweta Bhattacharya, Devulapalli Sudheer, and Bharath Bhushan Savarala. Artificial general intelligence: Advancements, challenges, and future directions in agi research. *IEEE Access*, 2025.