WHITE PAPER

Beyond the Hype: **Building Equitable and Sustainable AI for Social Impact**

April 2025

Bibliothèques Sans Frontières
Libraries Without Borders
kajou
pleias

About us

Bibliothèques Sans Frontières (Libraries Without Borders) is an NGO that since 2017, has been working to make knowledge accessible to all. Through its solidarity-driven initiatives, the organization turns culture, education, and information into powerful tools for resistance, resilience, and empowerment. With innovative solutions, a vast library of content available in over 27 languages, and recognized expertise in community engagement, BSF reaches people in crisis zones, development settings, and areas marked by deep inequality across some thirty countries, including France. ***www.librarieswithoutborders.org***

Kajou is a technology company specializing in mobile learning and digital solutions for social impact. They partner with organizations to develop and deploy innovative applications and platforms that address critical challenges in education, health, and economic development. Kajou's expertise lies in creating user-friendly, accessible technologies that can be effectively implemented in low-resource environments, often leveraging mobile devices and offline capabilities to reach underserved communities. Their focus is on building scalable and sustainable solutions that empower individuals and drive positive social change. ***www.kajou.io***

Pleias specializes in the development of large language models (LLMs) with advanced reasoning capabilities for information-intensive and highly regulated industries. Their approach emphasizes AI compliance, open scientific models and multilingual support to meet diverse industry needs. They focus on the creation of high-quality synthetic data, the development of corpus mining pipelines for untapped data sources, and the integration of semantic data for enhanced AI applications. The founding team is made up of experts affiliated with the Sorbonne's Artificial Intelligence Center and Sciences Po's Médialab, known for their pioneering LLM assistantships and contributions to open access research. ***www.kajou.io***

List of abbreviations

- **AI:** Artificial Intelligence
- **CHWs:** Community Health Workers
- **CRSV:** Conflict-Related Sexual Violence
- **GenAI:** Generative Artificial Intelligence
- **LLMs:** Large Language Models
- **LMICs:** Low- and Middle-Income Countries
- **NLP:** Natural Language Processing
- **POC:** Proof of Concept
- **RAG:** Retrieval Augmented Generation
- **RLHF:** Reinforcement Learning from Human Feedback
- **SSA:** Sub-Saharan African
- **TOPS:** Trillion Operations Per Second

Keywords
Artificial Intelligence, Social Impact, Global South, Equity, Linguistic Diversity, Infrastructure, Governance, Open Source, Education, Responsible AI

Authors
Anastasia Stasenko, Co-Founder, Pleias
Cormac O'Keeffe, Director of Education, Impact and Research, Bibliothèques Sans Frontières
Pierre -Carl Langlais, Co-Founder, Pleias

Coordination & Contribution
Ambre Cerny, Director of Communications, Bibliothèques Sans Frontières
Irène Girard, Data Analyst, Pleias
Jérémy Lachal, General Director & Co-Founder, Bibliothèques Sans Frontières

CONTENTS

INTRODUCTION

Generative AI offers unprecedented potential for social good, particularly in education, healthcare, and democratizing access to information across diverse cultures and languages. However, the promise of AI is threatened by ethical risks, the potential for human marginalization, and the reinforcement of existing inequalities, especially in the Global South. This white paper addresses these opportunities and risks, advocating for responsible AI development and deployment that prioritizes equitable impact through open-source solutions, linguistic diversity, and sustainable infrastructure.

Why This White Paper Now?

Generative AI has unlocked unprecedented possibilities in how we access and create information. It offers tremendous potential for improving education and healthcare, democratizing access to information, and embracing cultural and linguistic diversity. Moreover, as digital transformation accelerates in Sub-Saharan Africa, AI is coming to the fore as a potential source of prosperity, with the United Nations forecasting a contribution of up to $1.5 trillion to the region's economy by 2030. However, this transformative technology also presents risks, including ethical concerns, potential for human marginalization, and the spread of misinformation. There is a risk of closed, black-boxed solutions controlled by a few, becoming out of reach for many, and ultimately perpetuating existing biases and inequalities. **Recognizing this potential and these risks, Bibliothèques Sans Frontières and Kajou have been actively experimenting with AI solutions for over a year, yielding valuable insights and perspectives that inform this paper**

[1] GSMA. (2024). The Mobile Economy Sub-Saharan Africa 2024. GSM Association. https://www.gsmaintelligence.com/research/research-file-download?reportId=50121&assetId=12080

KEY TAKEAWAYS AND TRENDS

LINGUISTIC UNDER-REPRESENTATION: *A CORE CHALLENGE*

The under-representation of diverse languages in AI training data and the skewed distribution of training data toward English and other European languages hinder the development of fair and effective AI for many regions. The scarcity of multilingual data hinders the development of fair, useful applications of AI and ultimately reduces the performance of non-linguistically or culturally relevant models.
(See: Linguistic (under) representation in data and LLMs & The Data Paradigm Problem)

INNOVATION FOR SOCIAL GOOD IS ALREADY HAPPENING

Concrete AI projects like Karibu, kSANTÉ, AI4CRSV, and AI4Teachers showcase the potential of AI to address pressing social challenges in education, health, and access to justice, showing field-based and user-centred applications of the IDEAS AI initiative.
(See: Practical Implementations in the Field)

SUSTAINABLE AND FRUGAL AI: *A RESPONSIBLE PATH*

AI, especially large models, can have a significant impact on elicit and water consumption. Prioritizing sustainable, resource-efficient AI models and technologies is essential to mitigate the environmental impact of AI and enable deployment in resource-constrained environments. **(See: Sustainable and frugal offline AI)**

EMERGING SOLUTIONS: *ALTERNATIVE AI ARCHITECTURES*

Edge AI, offline, mobile-first deployment, and decentralized training offer promising alternatives for enabling sovereign AI development despite infrastructure limitations.
(See: Emerging Solutions: Alternative Approaches to AI Infrastructure)

KEY TAKEAWAYS AND TRENDS

2|2

ADOPTION AND GOVERNANCE CHALLENGES: *A NEED FOR CONTEXTUAL*

Understanding: Successfully adopting and governing AI in social impact sectors requires addressing domain-specific challenges, building trust, developing appropriate evaluation frameworks, and fostering domain-appropriate governance models.
(See: Adoption challenges & Governance challenges in complex implementation environments)

BEYOND TECHNICAL METRICS: *CULTURAL RELEVANCE IN EVALUATION*

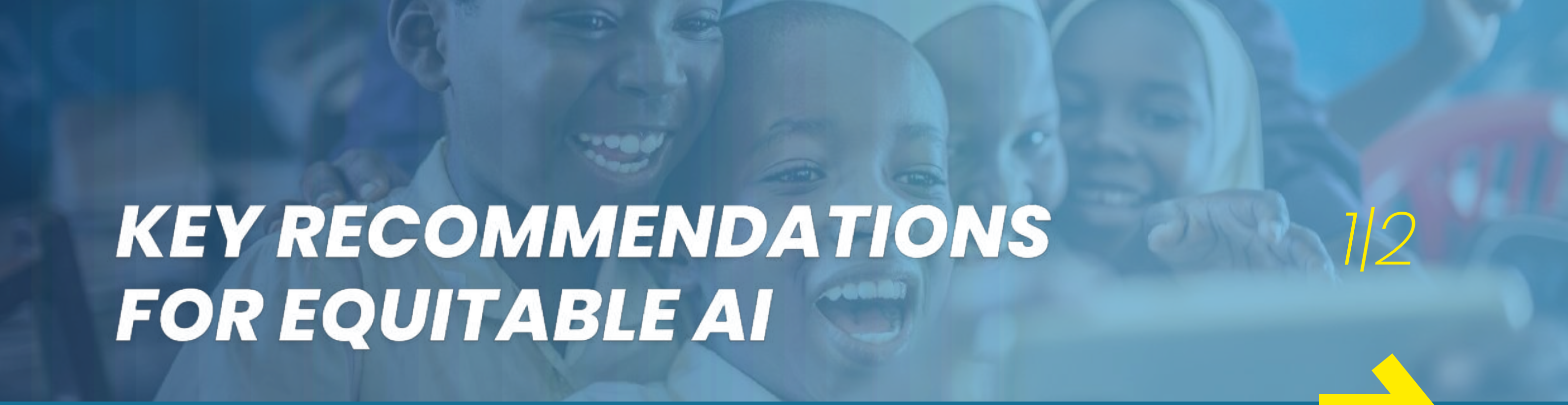
The need to prioritize cultural appropriateness alongside technical performance is becoming increasingly apparent. Evaluation frameworks must assess for localisation, accessibility, and cultural responsiveness to ensure AI improves quality and outcomes in diverse contexts.
(See: Evaluation Crisis Across Implementation Contexts & Establish Benchmarks for Ethical and Inclusive AI)

COMMUNITY-DRIVEN DATA CURATION

Equitable AI requires a shift from data extraction to community data stewardship. Empowering local communities to curate, contribute, and govern linguistic data ensures AI systems reflect local needs and knowledge, fostering local ownership and agency.
(See: Empower Linguistic Communities to Curate AI Data)

DECENTRALIZED TRAINING

A Path to AI Sovereignty: Federated learning and other decentralized training approaches offer promising pathways for emerging markets to develop AI capabilities despite infrastructure limitations. By distributing the computational burden, these techniques enable collective capabilities that exceed individual resources.
(See: Decentralized Training Approaches & Emerging Solutions: Alternative Approaches to AI Infrastructure)



KEY RECOMMENDATIONS FOR EQUITABLE AI

1/2

1

HUMAN-CENTERED AND ETHICAL AI GOVERNANCE: *PRIORITIZE MEANINGFUL OVERSIGHT*

AI systems require human oversight and ethical safeguards, especially in sensitive domains, to respect community values and rights.

Recommendation: Mandate Human-in-the-Loop (HITL) frameworks in all AI implementations to preserve human agency and prevent unintended harms.

2

FRUGAL, SUSTAINABLE, AND OFFLINE AI: *A PATH TO EQUITY AND SUSTAINABILITY*

To promote equity and reduce environmental impact, prioritize resource-efficient models that can operate offline on affordable hardware.

Recommendation: Invest in offline, resource-efficient AI solutions deployable on low-cost or edge devices and support community-driven innovation.

3

LINGUISTIC AND CULTURAL DIVERSITY: *ESSENTIAL FOR FAIR AI SYSTEMS*

Fair AI requires linguistic and cultural inclusion, so supporting the integration of underrepresented languages ensures systems reflect human knowledge and preserves linguistic diversity.

Recommendation: Fund open-source NLP tools and datasets to empower communities to integrate local languages into AI systems.

KEY RECOMMENDATIONS FOR EQUITABLE AI

2|2

4

OPEN SOURCE AND TRANSPARENCY: *FOUNDATIONAL FOR TRUST AND ACCOUNTABILITY*

Promote the adoption of open-source licenses for both AI models and training datasets, ensuring transparency and enabling community-driven auditing, adaptation, and improvement.

Recommendation: Prioritize funding for AI projects that embrace open-source principles, making code, data, and methodologies publicly accessible.

5

INCLUSIVE METRICS: *BEYOND TECHNICAL PERFORMANCE*

Evaluation frameworks need culturally relevant benchmarks to ensure AI is inclusive, unbiased, and accessible, especially in key social sectors.

Recommendation: Develop and adopt evaluation frameworks prioritizing cultural appropriateness, accessibility, and fairness alongside accuracy.

6

LOCAL CAPACITY BUILDING: *SHAPING SUSTAINABLE AI FUTURES*

Sustainable AI development needs local talent, leadership, and infrastructure, which means supporting education, research, and entrepreneurship in LMICs.

Recommendation: Fund AI education and training initiatives, strengthening local ecosystems through support for startups and global partnerships.

AI FOR SOCIAL IMPACT

One year of research and field experience by Bibliothèques Sans Frontières

The rapid advancement of generative artificial intelligence technologies presents unprecedented opportunities for transformative social impact across the globe. However, the promise of these sophisticated systems remains unevenly distributed, particularly in regions where linguistic diversity is highest and financial resources are at their most scarce.

Recognizing the critical barriers to AI equity posed by multilingual data scarcity, infrastructure deficits, and reliance on centralized AI models whose commitment to open science and transparency can be varied, Bibliothèques Sans Frontières (BSF) along with researchers who later went on to found PleIAS, published an op-ed in Le Monde in December 2023 arguing for:

1. **SYSTEMATICALLY ADAPTING MODELS TO LOCAL CONTEXTS**, in particular to reflect and promote linguistic and cultural diversity;
2. **BUILDING MODELS IN "OPEN SCIENCE"** (e.g. whose source code, weights and training corpora are openly accessible and reusable by others), able to operate in frugal tech environments, especially without an Internet connection or with a limited one - which is the norm for half of the world's population.
3. **CREATING MASSIVE CITIZEN TRAINING PLANS** – particularly aimed at the most vulnerable – to the opportunities and challenges brought about by the irruption of AI into our lives, and supporting them to fully master and reach appropriation of these technologies.

This op-ed and the support within the community led to the launch of what was to be the **Ideas AI initiative**[2]. This initiative aims to harness the potential of artificial intelligence by sharing knowledge to developing countries during humanitarian emergencies in the education and health sectors.

---

[2] Libraries Without Borders. (2025). *IDEAS AI AI for sharing knowledge*. Libraries Without Borders. https://www.librarieswithoutborders.org/ideas-ai-en/

Ideas AI focuses specifically on developing and deploying AI solutions that are frugal, open-source, and operational even offline—critical factors for success in regions lacking robust digital infrastructure. These approaches directly address the infrastructural and technological sovereignty challenges highlighted in this paper. Through projects that are firmly grounded in local realities, the initiative builds AI-driven solutions that are contextually relevant, linguistically inclusive, and culturally adapted to local needs.

The initiative comprises three strategic pillars:

1. Impactful Projects
2. Research & Advocacy
3. Targeted Training

Since launching IDEAS AI in early 2024, the initiative has already begun piloting several projects across Senegal, Ivory Coast, Nigeria, Democratic Republic of Congo, and Ukraine. These projects leverage generative AI for educational improvement, healthcare support, and support for survivors of conflict-related sexual violence, reflecting the initiative's comprehensive approach to AI for social good:

- **AI FOR LANGUAGE LEARNING IN WEST AFRICA**

  An Ai-Driven French Language Learning App To Support Senegalese School Teachers In Learning French And Delivering Effective Lessons To Their Students.

- **AI FOR TEACHER PROFESSIONAL DEVELOPMENT & LIFELONG LEARNING**

  A Frugal, Offline Llm For Ultraportable Microservers And Mobile Phones To Support Teacher Education In Rural Senegal In Foundational Learning, Digital Skills, Green Skills, And Gender-Transformative Pedagogy.

- **AI FOR HEALTH LITERACY IN WESTERN AFRICA**

  An Online And Offline Conversational Agent For Community Health Workers In Ivory Coast And Senegal. Validated By Health Professionals, It Will Support Outreach Workers In Daily Practice And Training And Provide Content In The National Languages Of Dioula And Wolof.

- **AI TO END CONFLICT RELATED SEXUAL VIOLENCE**

  An Ai Initiative On Conflict-Related Sexual Violence Adapted To The Languages Spoken In Nigeria, Ukraine, And Democratic Republic Of Congo. The Tool Based On The Foundation’s Legal Database, Will Assist Legal Workers Supporting Survivors.

  -> Refer to Part 2, BSF’s social impact projects’, for more information.

STATE OF AI FOR SOCIAL IMPACT IN 2025

While the case studies and pilot projects within the IDEAS AI initiative highlight the transformative potential of AI in addressing social challenges, they also expose the significant hurdles that must be overcome to scale these efforts. As we move forward, we will explore the systemic barriers—ranging from linguistic representation to infrastructure deficits—that need to be addressed for AI to realize its full potential in driving sustainable, equitable change. Understanding these challenges is critical to advancing the next generation of inclusive AI solutions.

This white paper explores the current landscape of AI development, highlighting the tension between initiatives driving open, equitable, and transformative social impacts, and those trending toward closed, opaque, and restrictive practices. The former push for transparency, inclusivity, and equitable access, fostering meaningful societal transformation, whereas the latter reinforce barriers and deepen the divide between digital "haves" and "have-nots". While some progress has been made, state of the art closed (and open) sourced models still show error rates 1.5 times higher (see Figure 1) for countries from Sub-Saharan Africa compared to North American countries[3].

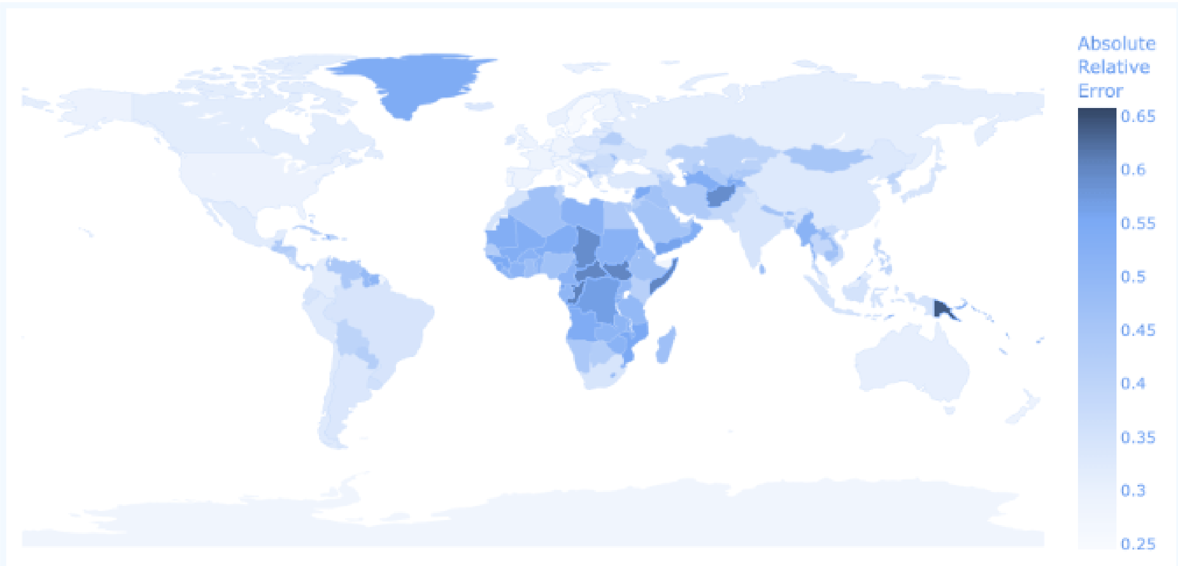


Figure 1: Error rates in the ability of LLMs to recall factual information about specific countries.
Source: WorldBench: Quantifying Geographic Disparities in LLM Factual Recall

[3] Moayeri, M., Tabassi, E., & Feizi, S. (2024). WorldBench: Quantifying Geographic Disparities in LLM Factual Recall. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1211–1228. https://doi.org/10.1145/3630106.3658967

We will examine the key challenges currently faced, highlight innovative solutions being developed by PleIAS, Bibliothèques Sans Frontières, and other actors, and explore opportunities to effectively address pressing issues in health, education, and employment.

This section begins by asking the question, that since December 2024, what has happened to address this call for action?

LINGUISTIC (UNDER) REPRESENTATION IN DATA AND LLMS

The lack of local-language data is not just a technical issue; it has a profound human impact. As we explore the technical, cultural, and infrastructural dimensions of this challenge, it becomes clear that addressing multilingual data scarcity represents perhaps the most significant barrier to equitable AI development globally. Many AI tools fail to work effectively—or not at all—for communities in low-resource regions, deepening existing digital inequalities and limiting opportunities for education, healthcare, and economic empowerment.

The scale of linguistic diversity—with over 7000 languages in the world of which 2,000 languages are spoken across Africa alone—demands coordinated efforts that extend beyond individual corporate initiatives or academic projects[4]. This diversity represents not merely a technical challenge but a rich opportunity to develop more robust, inclusive AI systems that respect and enhance the full spectrum of human knowledge and experience.

The consequences of inaction are significant. Without addressing multilingual data challenges:

1. AI deployment risks reinforcing existing social and economic inequalities rather than mitigating them
2. Linguistic communities may face accelerated language loss as technologies favor dominant languages with one language lost every 3 months[5]
3. Indigenous knowledge systems may be systematically excluded from AI-mediated information spaces
4. Users of AI systems with low cultural context will frequently underperform and provide inaccurate and/or culturally biased information
5. Economic opportunities arising from AI will disproportionately benefit already-advantaged communities

[4] Research center for language intelligence. (2024). How many languages are there in the world? *Ethnologue*.
[5] Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, *6*, 163–173. https://doi.org/10.1038/s41559-021-01604-y

While improving data diversity is essential, it is not enough on its own. Even if we had abundant multilingual data, communities in the Global South face another major obstacle: the lack of physical infrastructure needed to train, deploy, and scale AI systems.

The Data Paradigm Problem: Current state of training data distribution

Data serves as the foundation upon which all AI systems are built, functioning as both the primary knowledge source and the basis for developing an AI system's reasoning capabilities. Current frontier AI models have been trained on massive datasets comprising hundreds of trillions of tokens, yet the linguistic distribution within these datasets remains overwhelmingly skewed toward English and other European languages. Historical policies that favoured European languages have meant that many non European datasets are sparse and marginalised. Not only does this undermine the capacity of LLMs to capture linguistic richness and diversity but it also means that they underperform[6].

This imbalance is not merely a matter of volume but reflects deeper systemic priorities in the technology development ecosystem. The majority of commercially available and open-source large language models (LLMs) dedicate upwards of 90% of their training data to English content[7], with another 10% distributed among high-resource European languages such as Spanish, French, German, and Italian. This leaves less than 1% of training data to represent the thousands of languages spoken by billions of people globally, particularly across Africa, Asia, and indigenous communities worldwide.

The challenge extends far beyond the ability to process queries and to chat in different languages. While these surface-level capabilities are important, they represent only the most basic requirements for truly multilingual AI systems. The deeper issue involves the embedded reasoning frameworks, cultural references, contextual knowledge, and nuanced understanding that form the foundation of truly useful AI applications. This is particularly true in African contexts where socio-cultural realities are underrepresented in

[6] Adebara, I., Toyin, H. O., Ghebremichael, N. T., Elmadany, A., & Abdul-Mageed, M. (2025). *Where Are We? Evaluating LLM Performance on African Languages* (No. arXiv:2502.19582). arXiv. https://doi.org/10.48550/arXiv.2502.19582

[7] Zhang, B., Williams, P., Titov, I., & Sennrich, R. (2020). Improving Massively Multilingual Neural Machine Translation and Zero-Shot Translation. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1628–1639. https://doi.org/10.18653/v1/2020.acl-main.148

the models thus impacting a model's performance and its ability to be used to positively impact social change for good[8].

When models are disproportionately trained on Western cultural contexts and linguistic patterns, they inherently develop reasoning structures that reflect these contexts. The majority of models available today consistently align better with high-income countries when it comes to information retrieval and text generation[9]. This imbalance is further accentuated by the fact that in an attempt to create models in languages other than English, many datasets are simply translations from English. In addition to containing cultural biases, they LLMs based translated data struggle with queries that complex social and cultural contexts[10] and often deliver worse in hate speech tasks[11].

This creates a subtle but profound form of cognitive bias within AI systems that affects their ability to:

1. **Model diverse social norms and expectations**
2. **Generate contextually appropriate responses for different cultural settings**
3. **Recognize and respect different epistemological frameworks**
4. **Incorporate indigenous knowledge systems**

The result is a form of technological colonization where Western cognitive frameworks become embedded in systems deployed globally, reinforcing existing power imbalances, existing inequalities, and contribute to discriminatory practices[12] under the guise of universal technological solutions.

---

[8] AUDA-NEPAD. (2024). *AI and the Future of Work in Africa.* https://www.nepad.org/publication/ai-and-future-of-work-africa-white-paper

[9] Yadav, S., Zhang, Z., Hershcovich, D., & Shutova, E. (n.d.). *Beyond Words: Exploring Cultural Value Sensitivity in Multimodal Models*.

[10] Seto, S., Hoeve, M. ter, Bai, R. H., Schluter, N., & Grangier, D. (2025). *Training Bilingual LLMs with Data Constraints in the Targeted Language* (No. arXiv:2411.12986). arXiv. https://doi.org/10.48550/arXiv.2411.12986

[11] Hasan, M. A., Tarannum, P., Dey, K., Razzak, I., & Naseem, U. (2024). *Do Large Language Models Speak All Languages Equally? A Comparative Study in Low-Resource Settings* (No. arXiv:2408.02237). arXiv. https://doi.org/10.48550/arXiv.2408.02237

[12] Lim, S., & Pérez-Ortiz, M. (2024). *The African Woman is Rhythmic and Soulful: An Investigation of Implicit Biases in LLM Open-ended Text Generation*. In Review. https://doi.org/10.21203/rs.3.rs-5283007/v1

Awarri's LangEasy AI - Building Nigeria's Multilingual AI Foundation

Awarri's LangEasy AI initiative is tackling the critical challenge of linguistic underrepresentation in AI by empowering Nigerians to contribute to a comprehensive, multilingual AI model. LangEasy invites individuals to record audio samples translating English sentences into Nigerian languages, including Yoruba, Hausa, Igbo, Pidgin, and Ibibio. Through this crowdsourced approach, Awarri aims to create a high-quality dataset that reflects the unique linguistic nuances and cultural heritage of Nigeria, enabling AI models to better understand and represent the nation's diverse linguistic landscape. This initiative directly addresses the issue of data scarcity for African languages, enabling a more inclusive and equitable AI ecosystem.

By contributing to LangEasy, individuals play an active role in shaping the future of AI, ensuring that it serves all Nigerians, regardless of the language they speak. Awarri wants to ensure the preservation and advancement of local culture through open source solutions. This has direct application to one of the central goals of sustainable AI.

Cross-Lingual Transfer: promises and limitations

→ Recent advances in zero-shot language transfer

Imagine teaching a child to read. Once they understand the basic principles, they can often apply those skills to new books and even new languages. Cross-lingual transfer learning attempts to do something similar with AI. Recent advances in large language model architectures have demonstrated surprising capabilities for cross-lingual transfer. Models like DeepSeek, Claude, GPT-4, and others have shown the ability to engage in conversations in languages like Swahili, Yoruba, and Amharic with relatively coherent outputs despite having minimal exposure to these languages during training. This capability stems from several technical advancements:

1. The scale effect: as models grow larger, they develop more robust internal representations of linguistic patterns

2. Shared vocabulary spaces through tokenization that sometimes capture cross-lingual morphological similarities
3. Multilingual embedding spaces that map semantically similar concepts across languages
4. Improved parameter efficiency that allows for better generalization from limited examples

These capabilities initially appear promising as they suggest that models might overcome data scarcity through efficient transfer learning. If we can train a model primarily on English and then "transfer" its abilities to hundreds of other languages, we can potentially create multilingual AI systems quickly and cheaply. However, a closer examination reveals significant limitations to this approach, particularly for communities in the Global South.

→ The limitations of transfer learning

While zero-shot and few-shot language transfer can produce superficially coherent outputs, detailed analysis reveals several critical shortcomings:

Surface-level competence: Models may generate grammatically plausible sentences but frequently demonstrate shallow semantic understanding. They struggle with:

- Ambiguous terms with culturally specific meanings
- Handling complex logical reasoning in non-dominant languages
- Maintaining cultural appropriateness across extended interactions
- Accurate representation of specialized knowledge domains

> *Example: A model might be able to translate a sentence about healthcare, but completely miss the cultural nuances related to traditional healing practices*

Persistent Western reasoning frameworks: Even when producing text in diverse languages, these models continue to apply reasoning frameworks derived from their predominantly Western training data. This manifests as:

- Application of Western conceptual hierarchies to non-Western contexts
- Importation of inappropriate cultural assumptions
- Misinterpretation of local context based on dominant-culture defaults
- Reinforcement of Western epistemological frameworks as universal

> *Example: An AI system might provide advice on farming techniques that are completely inappropriate for the local climate or soil conditions.*

The localization gap: True localization goes beyond language to encompass cultural contexts, local knowledge, and community values. Current cross-lingual transfer approaches fail to address:

- Locally relevant examples and reference points
- Community-specific ethical frameworks
- Historical and cultural contexts specific to regions
- Local expertise and traditional knowledge systems

Example: A model trained on European legal documents might be unable to understand or apply traditional dispute resolution mechanisms in an African community.

This gap between linguistic production and cultural understanding reinforces the essential need for genuine representation in training data rather than relying solely on transfer capabilities. To create truly inclusive AI, we must prioritize the creation of high-quality, culturally relevant training data for diverse languages. The following section will explore some of the key challenges related to creating this data and the solutions that are being developed to address them.

The Nature Language Processing Infrastructure Deficit

The promise of AI for social good hinges on more than just sophisticated models; it requires a robust foundation of basic tools and resources for natural language processing (NLP). Africa, home to over 3000 languages spread across 54 countries[13] still has only basic NLP tools that are often underdeveloped - or in many cases, entirely absent. This infrastructural deficit includes:

Language identification and classification: Even fundamental tasks like reliable language detection remain challenging for many African languages. Current language identification systems frequently:

- Misclassify related languages due to insufficient training examples
- Fail to distinguish between regional variants of the same language

[13] Tchindjang, M., Athanase, B., & Ngamgne, L. A.. (2008). *Languages and cultural identities in Africa—UNESCO Bibliothèque Numérique*. https://unesdoc.unesco.org/ark:/48223/pf0000162983

- Perform poorly on code-switched text (mixing multiple languages)
- Misidentify less-documented languages as more common ones

For example, even Swahili, spoken by approximately 200 million people across East Africa, lacks robust language detection tools as well as significant gaps in data resources, language models, and the adaptation of existing NLP technologies to the unique linguistic contexts of African nations[14]. This basic deficiency creates cascading problems throughout the AI development pipeline.

Absence of basic processing tools: Many languages lack even the most basic preprocessing tools that are taken for granted in English NLP:

- Stemmers and lemmatizers to reduce words to their base forms
- Part-of-speech taggers to identify grammatical structures
- Named entity recognizers to identify people, organizations, and locations
- Dependency parsers to analyze grammatical relationships

Without these foundational tools, developing more advanced applications becomes extraordinarily difficult, creating a technological debt that compounds with each new development cycle.

Tokenization challenges: Tokenization—the process of breaking text into manageable units for processing—presents particular challenges for morphologically rich languages common across Africa. Current tokenizers often:

- Split words inappropriately, breaking meaningful morphological units
- Create unnecessarily large vocabulary spaces for agglutinative languages
- Handle diacritics and special characters inconsistently
- Struggle with languages that do not use space-delimited words

These technical challenges have significant implications for model performance, as poor tokenization directly impacts semantic understanding. These are just a few examples of the NLP infrastructure deficit that prevents many communities from fully participating in the AI revolution. This isn't just a technical problem; it's a human problem with significant consequences for education, healthcare, economic opportunity, and cultural preservation.

[14] Amol, C. J., Chimoto, E. A., Gesicho, R. D., Gitau, A. M., Etori, N. A., Kinyanjui, C., Ndung'u, S., Moruye, L., Ooko, S. O., Kitonga, K., Muhia, B., Gitau, C., Ndolo, A., Wanzare, L. D. A., Kahira, A. N., & Tombe, R. (2024). *State of NLP in Kenya: A Survey* (No. arXiv:2410.09948). arXiv. https://doi.org/10.48550/arXiv.2410.09948

To address these challenges, we need a concerted effort to prioritize the development of basic NLP tools for diverse languages. The following section will explore some of the downstream consequences of this deficit and will highlight potential solutions.

The Common Crawl Challenge

Organizations like the Common Crawl Foundation provide invaluable web-crawled datasets used for training many contemporary AI models. However, these resources reflect and often amplify existing data imbalances due to methodological limitations:

Language identification bottlenecks

Common Crawl's pipeline relies on language identification tools to classify its massive corpus. Given the limitations of these tools for less-resourced languages, content in these languages is frequently:

- Misclassified as more dominant languages
- Filtered out due to low confidence scores
- Incorrectly grouped with related languages

These identification failures create a self-reinforcing cycle where languages with insufficient initial representation remain underrepresented in derived datasets.

The "low-resource" paradox

From the perspective of organizations like Common Crawl, virtually all languages except English qualify as "low-resource" due to the extreme imbalance in online content. This creates a situation where even languages spoken by tens of millions of people face severe data scarcity in digital contexts. The statistics are stark:

- English represents approximately 60-70% of web content despite being spoken by only about 15% of the global population
- The top 10 languages online account for over 90% of content
- Thousands of languages with millions of speakers have negligible online presence

This digital representation gap translates directly into training data scarcity for AI systems.

Quality and representativeness issues

Beyond simple volume, the quality and representativeness of available data present additional challenges:

- Available content often skews toward certain domains (tourism, news, education)
- User-generated content may be limited to specific demographic groups
- Religious texts often constitute a disproportionate percentage of available content
- Translated content rather than original material predominates

These qualitative issues further compromise the utility of the limited data that does exist.

The Common Crawl Foundation is a nonprofit organization that has been operating since 2007. Its mission is to preserve and freely share samples of the public Internet. Common Crawl is a key partner to multiple research communities ranging from Web search, AI and security, to archiving, political science and linguistics. Our crawling has always been polite and ethical, strictly obeying robots.txt. However our crawls have been historically biased towards English and North American content, in part because the foundation is based in the US, but also because the seeds of our initial seed crawls were mostly comprised of URLs from North American websites.

Throughout the years, we have seen an ever increasing interest in our dataset across the globe, as smaller linguistic and cultural communities use our dataset to develop technologies and conduct research that was previously only accessible to mostly anglo-centric communities. We thus believe that improving Common Crawl's language and cultural diversity would be of great benefit. However, in trying to expand Common Crawl's coverage we have identified two major roadblocks, the first being the lack of availability of a good Language Identification Algorithm (LangID), where we found that even though LangID solutions exists, they only cover high-resource languages and tend *to be biased towards cleaner and formal text, being unable to recognize more informal language registers even for those high-resource languages. The second roadblock we found is that being a fairly small team, we only speak a handful of languages, making it very challenging for us to manually collect and inspect URLs that could potentially serve as seeds for our seed crawl.*

In order to overcome these two challenges, we have created two initiatives, the Web Languages Project where contributors contribute seeds (URLs of websites) in their languages to us, and the LangID project, where we ask contributors to do simple LangID annotations on Common Crawl data with languages tags so that we can develop our own LangID algorithm. But even though we expect to improve Common Crawl's language and cultural coverage with these projects, we will need significant input and contributions from community, not only because it is impossible for our team to carry out these two projects out by ourselves, but also because we believe that the languages and the content written in them belong in the end to their respective linguistic communities.

Pedro Ortiz Suarez, Senior Research Scientist, Common Crawl Foundation

Downstream Consequences for Applied AI

The lack of basic NLP tools is not just an abstract technical problem. Its absence creates significant barriers to implementing practical AI applications, effectively preventing many communities from leveraging AI for local needs. This section will explore what that absence means in practice

→ Challenges in Retrieval Augmented Generation

Retrieval Augmented Generation (RAG) systems, which enhance generation with retrieved information, depend heavily on language-specific preprocessing capabilities. Without these tools, RAG implementations face severe limitations:

- Inability to effectively index content in local languages
- Poor retrieval performance due to inadequate semantic matching
- Failure to identify relevant passages when queries contain cultural nuances
- Reduced accuracy in generating responses that incorporate local knowledge

These limitations are particularly problematic as RAG represents one of the most promising approaches for creating AI systems grounded in accurate, up-to-date information.

➜ Document processing limitations

Organizations across Africa and other linguistically diverse regions deal with massive volumes of documents in local languages. The absence of appropriate NLP tools severely limits the potential for automation in these contexts:

- PDF extraction tools perform poorly on documents in non-Latin scripts
- Optical character recognition systems have high error rates for many languages
- Form parsing and information extraction remain largely manual processes
- Document classification systems cannot effectively categorize local-language content

These limitations force continued reliance on manual processing, preventing efficiency gains that AI could otherwise enable.

➜ Impact on specialized domain applications

The data and tooling deficit is especially pronounced in specialized domains critical for social impact applications:

Healthcare

Medical AI applications face particular challenges due to:

- Lack of medical terminology datasets in local languages
- Poor performance in translating medical concepts across linguistic contexts

Agricultural applications

AI systems supporting agricultural development struggle with:

- Limited datasets describing local crop varieties and farming techniques
- Inadequate representation of traditional agricultural knowledge
- Poor performance in identifying region-specific plant diseases
- Inability to effectively process local terminology for soil conditions and weather patterns

Legal and administrative systems

Efforts to improve access to legal and administrative services through AI face:

- Scarcity of legal corpus data in local languages
- Limited precedent documentation for legal reasoning models
- Poor performance in understanding regionally specific administrative terminology
- Challenges in aligning with local legal frameworks and processes

These domain-specific limitations highlight how the general data scarcity problem manifests in concrete barriers to developing socially beneficial applications.

The Instruction and Alignment Data Crisis

While pre-training data (the raw material AI learns from) presents significant challenges, the situation becomes even more acute when considering instruction tuning and alignment data—specialized datasets that teach models to follow instructions and align with human preferences. Without good data, AI models will struggle to understand different cultural viewpoints

The instruction tuning gap: Instruction tuning transforms raw language models into assistive systems that can follow user directions. This process requires high-quality datasets of instruction-response pairs that:

1. Cover diverse task types and domains
2. Demonstrate appropriate reasoning patterns
3. Exhibit desired assistant behaviors
4. Represent culturally appropriate interactions

For most languages beyond English, such datasets are virtually nonexistent in open-source repositories. Even for French—a global language with significant resources—open-source instruction tuning datasets remain scarce. This creates a fundamental barrier to developing assistive AI in diverse linguistic contexts.

The alignment challenge: Beyond basic instruction following, alignment techniques like Reinforcement Learning from Human Feedback (RLHF) require:

1. Human preference data comparing multiple responses
2. Detailed evaluative feedback on model outputs

3. Nuanced assessment of factual accuracy, helpfulness, and safety
4. Cultural appropriateness judgments from relevant perspectives

These specialized datasets are almost exclusively available in English, with minimal resources for other languages. This creates a situation where alignment itself becomes linguistically and culturally biased, reinforcing dominant perspectives in how models evaluate "good" responses.

Commercial dependencies and barriers: The absence of open-source alignment resources obliges AI labs working in non-English languages to contract with specialized data creation companies. This approach introduces:

1. Prohibitive costs for community-driven projects
2. Dependencies on external providers who may lack cultural context
3. Intellectual property constraints on resulting datasets
4. Barriers to collaborative improvement and adaptation

The resulting economic barrier effectively prevents grassroots AI development in linguistically diverse communities, creating a dependency on major technology companies and external providers that inherently limits local agency and innovation potential.

Quality control challenges: Even when commercial partnerships are viable, quality control for instruction and alignment data presents particular challenges in diverse linguistic contexts:

1. Scarcity of qualified annotators with both linguistic expertise and technical understanding
2. Difficulties in establishing consistent evaluation criteria across cultural contexts
3. Challenges in capturing diverse dialectal variations within languages
4. Inconsistent terminology for technical concepts across languages

These quality control issues further complicate efforts to develop appropriately aligned systems for diverse linguistic communities. However, developing and sharing data publicly and creating common data spaces, could help ensure that quality control for instruction and alignment data will adhere to high cultural standards.

Current Initiatives

Despite these significant challenges, several initiatives are working to address multilingual representation in AI, offering promising pathways toward more inclusive systems.

Cohere For AI's Aya Initiative

The Aya project represents one of the most significant efforts to expand instruction-following capabilities across diverse languages. Specifically, it:

1. Creates high-quality instruction-following datasets for 101 languages : Aya collaborates with native speakers, linguists, and local researchers worldwide to curate datasets that reflect the linguistic nuances of each language
2. Employs a collaborative, community-driven approach with local stakeholders : There are over 3.000 independent researchers and about 250 “language ambassadors” across 119 countries contributing to the dataset development, ensuring direct representation and local expertise.

   ...
3. Develops open benchmarks for evaluating multilingual performance : Aya works with regional tech communities to design benchmarks that assess model performance fairly across different languages, including underrepresented ones.
4. Releases models fine-tuned on these diverse datasets : The models are openly shared with researchers and developers, fostering transparency and enabling further improvement through collective contributions.

Google’s Gemma

Google's Gemma initiative represents a significant advancement in addressing the challenges of instruction tuning and alignment data for languages beyond English. Specifically, it:

1. Develops open-source language models: Gemma provides 2-billion and 7-billion parameter models that are lightweight and versatile—designed to integrate easily across various frameworks and devices.

2. Offers tools for multilingual instruction tuning: With resources such as comprehensive tutorials and frameworks like Unsloth, Gemma enables efficient fine-tuning for specific tasks and languages, reducing computational overhead while expanding multilingual capabilities.
3. Engages the community through open-source collaboration: Initiatives like the “Unlocking Global Communication with Gemma” competition actively invite developers from around the world to fine-tune these models for their native languages, fostering a collaborative environment that promotes the development of truly multilingual AI systems.
4. Utilizes curated, diverse multilingual training data: Recognizing that AI systems have historically been hampered by training data imbalances—where over 90% of data is dominated by English and a few high-resource European languages—Gemma takes a decisive step toward inclusivity.

Stoney Nakoda’s HarleyCoops

The HarleyCoops initiative is a project aiming at preserving and revitalizing endangered Indigenous languages using AI. The project, developed in collaboration with the community, seeks to create an effective language model for Stoney Nakoda that can be fine-tuned iteratively with community-driven feedback. This ensures that the model not only learns the language but also integrates cultural context and community-specific nuances :

1. AI-Powered Language Revitalization: The HarleyCoops initiative leverages state-of-the-art AI techniques to preserve and revitalize endangered Indigenous languages, creating a robust language model for Stoney Nakoda through data-centric approaches.
2. Community-Centric Cultural Preservation: Developed in close collaboration with native speakers and cultural custodians, the project integrates indigenous linguistic data, narrative traditions, and cultural context to ensure the model authentically represents the language.
3. Scalable and Iterative Model Development: Utilizing a modular pipeline for data ingestion, processing, and fine-tuning, HarleyCoops demonstrates how low-resource languages can be effectively supported through iterative, community-in-the-loop refinements.
4. Commitment to Open-Source Transparency: By openly sharing methodologies and model architectures, the project empowers researchers and communities alike to continuously enhance the model while safeguarding the cultural heritage for future generations.

- Named entity recognizers trained on locally relevant entities (so AI can identify important people, organizations, and places in a community).

As AI continues to rapidly evolve, the gap between English-speaking and non-English-speaking regions in AI adoption and advancement grows wider.

At Mundo AI, we believe that unlocking AI's full potential requires equitable access and development across all languages, cultures, and geographies. Current solutions, while adequate but not perfect for English, fail to address non-English languages effectively due to significant quality data shortages and uneven distributions of accents and dialects.

Mundo AI tackles this challenge through a unique approach focused on generating high-quality human data by building the largest multilingual dataset platform, and empowering organizations worldwide to develop inclusive models that benefit the world."

Garreth Lee, Founder, Mundo AI (YC W25)

Toward Truly Inclusive AI Infrastructure

At present, AI struggles to understand diverse languages and cultures, and we lack the tools to properly train and evaluate AI in these contexts. But how do we fix this? Building truly inclusive AI requires a multifaceted approach that combines technical innovation, institutional support, and, most importantly, active community engagement. It's about creating an AI infrastructure that reflects the world's rich diversity, not just a narrow slice of it.

Here are key priorities for infrastructure development:

1. **Basic NLP tool development: creating foundational processing tools for diverse languages, including:**
    - Language identification tools with improved accuracy for related languages (so AI can tell the difference between similar dialects)
    - Part-of-speech taggers and syntactic parsers for grammatically diverse languages (so AI can understand the structure of sentences, even in languages that don't follow Western grammatical rules)

- Named entity recognizers trained on locally relevant entities (so AI can identify important people, organizations, and places in a community).

2. **Data collection methodologies:**
    - Data is the fuel that drives AI, but for many languages, that fuel is scarce. We need innovative data collection methodologies that leverage existing textual resources (like government documents and educational materials), incorporate oral traditions through speech-to-text technologies, create synthetic data to supplement scarce resources, and design efficient annotation and synthetic data protocols for maximizing value from limited examples.

3. **Evaluation framework development:** We need evaluation frameworks that go beyond simple accuracy scores and assess the real-world impact of AI systems in diverse communities. This means:
    - Assessing performance across diverse task types in local languages not just translating English text
    - Incorporating cultural appropriateness metrics so AI doesn't reinforce harmful stereotypes or biases
    - Evaluating factual accuracy within local contexts so AI reflects local knowledge and realities.

These are the steps required to create robust building blocks for AI in diverse languages. By creating this robust infrastructure, the AI that people are using will be high-quality and useful for the people it serves. However, building inclusive AI infrastructure requires tackling fundamental infrastructure challenges which will be discussed in the next section.

INFRASTRUCTURE CHALLENGES FOR AI DEPLOYMENT IN THE GLOBAL SOUTH

While the previous section explored the critical challenges related to multilingual data availability for AI development, this section examines the equally fundamental infrastructure barriers that threaten to widen the digital divide as AI technologies advance. The physical and technical infrastructure required to develop, deploy, and access AI systems presents a complex set of challenges for emerging markets, particularly across Africa and parts of Asia and Latin America. This infrastructure gap encompasses computing resources, data centers, connectivity, and energy systems—all essential components for meaningful participation in the AI revolution. *Figure 2* shows how the distribution of robust digital infrastructures, advanced AI strategies, and significant human capital development is highly correlated with the ability to develop and maintain AI-based technologies.

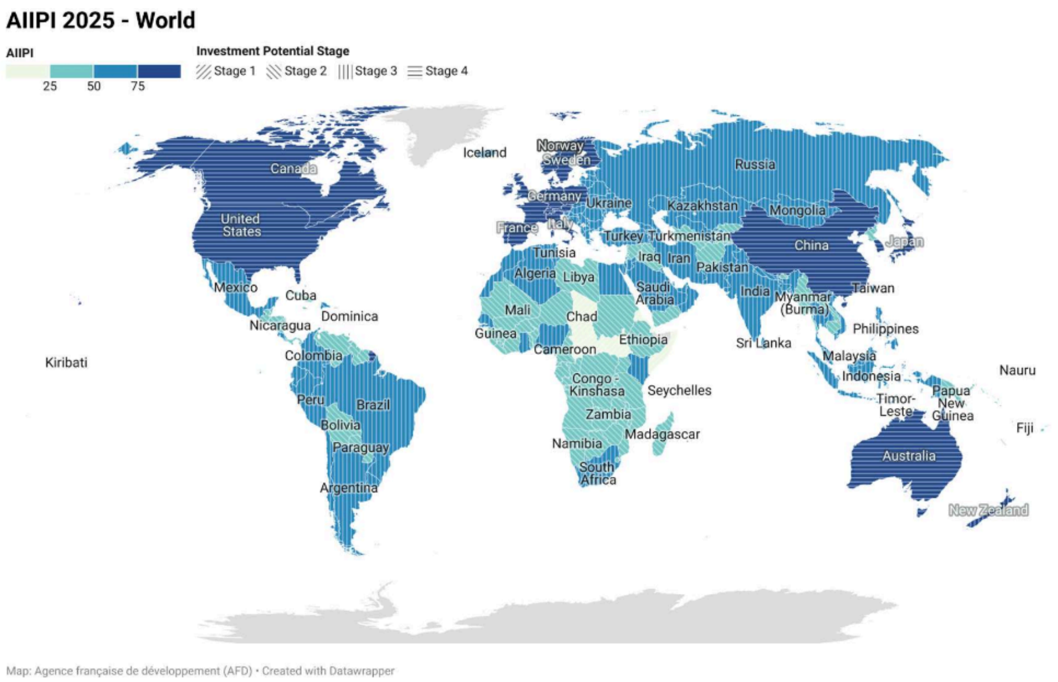


Figure 2: AI Investment Potential Index (AIIPI) showing Stage 1 countries (low) and Stage 4 countries (high)[15]

[15] Addo, P. M., Melonio, T., Taieb, A., & Landrien, L. (2025). *AI Investment Potential Index 2025 Unlocking Equitable Opportunities for Global AI Growth* (No. 342). Agence française de développement.

Computational resources divide

The global distribution of data center infrastructure reveals profound disparities that directly impact AI development and deployment capabilities. Current statistics paint a concerning picture:

- North America hosts approximately 40% of the world's data centers despite representing just 4.7% of the global population
- Europe accounts for roughly 30% of global data center capacity
- The entire African continent, home to 17% of the world's population, hosts less than 1% of global data center capacity[16]
- Within Africa, over 85% of existing data center capacity is concentrated in just three countries: South Africa (approximately 50%), Nigeria (21%), and Kenya (14%)

These disparities translate directly into computational resource availability. As of early 2024, the total data center capacity across Africa stands at approximately 250 MW—less than a single large-scale data center campus in Northern Virginia (which alone can exceed 300 MW).

→ AI training infrastructure

The specialized infrastructure required for AI model development presents an even more pronounced divide:

- Over 95% of the GPU clusters used for training frontier AI models are located in North America, Western Europe, and East Asia
- The top four cloud providers (AWS, Microsoft Azure, Google Cloud, and Alibaba Cloud) control approximately 80% of the available AI training infrastructure globally
- The estimated computing power used to train frontier models has increased by a factor of 1,000,000 in just seven years (2012-2019), with continued exponential growth since
- A single training run for a state-of-the-art model like GPT-4 or Claude 3 Opus requires an estimated 10,000-25,000 high-end GPUs operating for several months, representing tens of millions of dollars in computing resources

[16] US International Development Finance Corporation. (2022). Tackling a critical need for data center infrastructure across Africa. *DFC Blog*. https://www.dfc.gov/investment-story/tackling-critical-need-data-center-infrastructure-across-africa#:~:text=Information%20and%20communications%20technology%20,available%20global%20data%20center%20capacity

For emerging economies, accessing this scale of computing resources remains prohibitively expensive and logistically challenging, creating a fundamental barrier to sovereign AI development.

→ The hardware access challenge

Beyond data centers, the specialized hardware required for AI development faces significant supply constraints:

- High-end AI accelerators like NVIDIA H100 GPUs face global supply shortages, with delivery times often exceeding 6-9 months even for priority customers
- A single NVIDIA H100 GPU costs between $25,000-$40,000, putting even small clusters beyond the reach of most research institutions in emerging markets
- Import restrictions, tariffs, and complex procurement processes further complicate hardware acquisition in many regions
- Maintenance and operational expertise for specialized AI hardware remains concentrated in developed markets

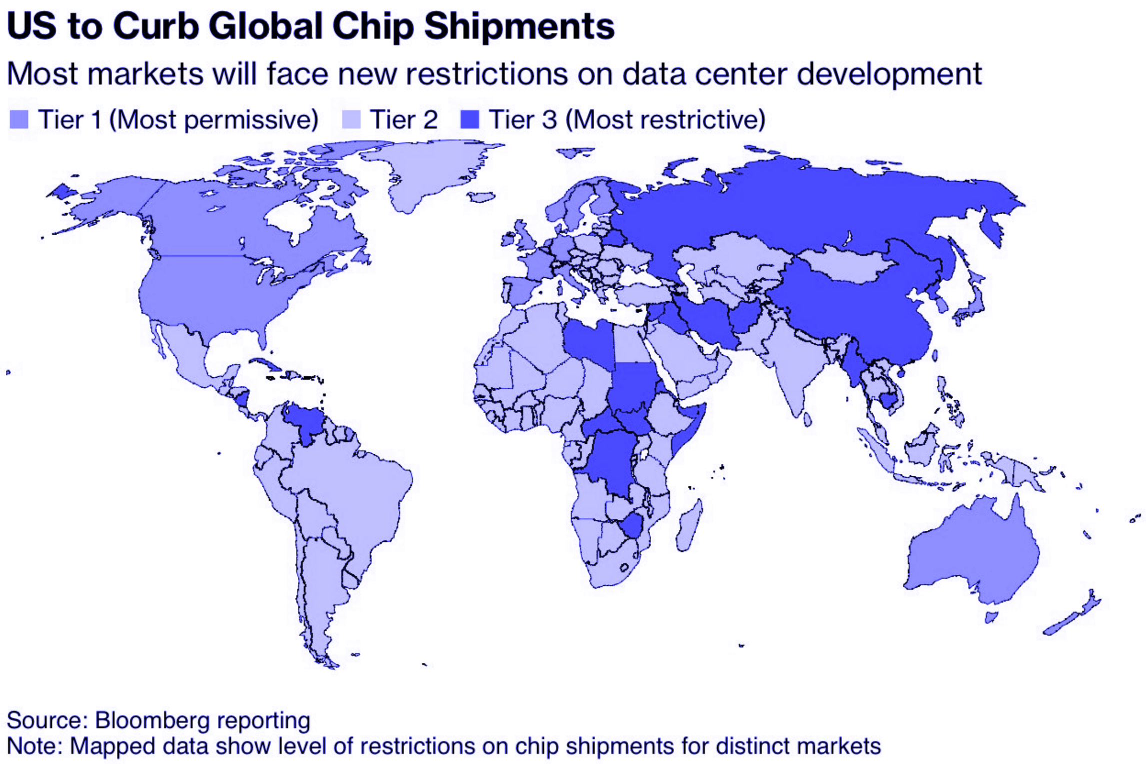


Fig x: Global limitations on the distribution of semiconductors specializing in processing AI applications (Tier 1 few restrictions, Tier 3: severe restrictions)[17]

[17] Nardelli, A. (2025). EU to Raise Biden's AI Chip Curbs with Trump Administration. *Bloomberg*.

These hardware constraints create cascading effects throughout the AI development ecosystem, limiting opportunities for local innovation and research.

For emerging economies, accessing this scale of computing resources remains prohibitively expensive and logistically challenging, creating a fundamental barrier to sovereign AI development.

As Michael S. Mollel, Co-Founder of Sartify LLC and Creator of PAWA, Swahili LLMs, explains

> *Our Swahili LLM performs well in specific domains such as agriculture, health, and education. However, our biggest challenge lies in compute power, particularly for fine-tuning the model when new data becomes available. To enhance this, we require at least one H100 GPU, which would significantly improve our ability to fine-tune the model efficiently. For inference, we currently host on AWS to serve high-demand users, while lower-demand services run locally on a machine with only 48GB of VRAM. To ensure smooth experimentation and seamless deployment for new business requests and demo applications, we are seeking an additional A6000-96GB GPU.*
>
> *The second major challenge is the slow adoption of AI across Africa. Additionally, African startups like Sartify face tough competition from well-funded foreign companies that have significantly more resources and computational power. Overcoming these barriers is essential for fostering a more competitive and locally-driven AI ecosystem.*
>
> **Michael S. Mollel, Co-Founder Sartify LLC, Creator of PAWA, Swahili LLMs— AI tools 4 African**

Connectivity challenges

While data center infrastructure represents a critical bottleneck, internet connectivity forms the foundation upon which all digital services—including AI—must be built.

---

https://www.bloomberg.com/news/articles/2025-01-21/eu-to-raise-biden-s-ai-chip-curbs-with-trump-administration

Current connectivity metrics reveal persistent challenges:

- Internet use across Africa averages 37% compared to 90%+ in North America and Western Europe[18]
- Fixed broadband connections reach less than 5% of African households compared to 80%+ in developed markets
- Average connection speeds vary dramatically:
    - North America: 150+ Mbps average fixed broadband
    - Western Europe: 120+ Mbps average fixed broadband
    - Sub-Saharan Africa: 28 Mbps average fixed broadband (with significant urban-rural disparities)
    - Rural Africa: Often below 10 Mbps where available

These speed disparities directly impact the usability of AI services, particularly as models grow more sophisticated and bandwidth-intensive.

➜ Mobile Connectivity

Mobile networks have expanded significantly across emerging markets, but substantial challenges remain:

- 4G coverage reaches approximately 77% of Africa's population, but actual adoption stands at only 28% due to affordability constraints[19]
- Device limitations further constrain access, with 42% of Africans covered by mobile mobile internet having no device to use it
- 5G deployment remains in early stages, covering less than 3% of Africa's population compared to 80%+ in leading markets[20]
- Data costs average 8-10% of monthly income across Africa compared to less than 1% in North America

For AI applications that require real-time interaction or substantial data transfer, these connectivity limitations create significant barriers to adoption and usability.

[18] ITU. (2023). Global offline population steadily declines to 2.6 billion people in 2023. *Measuring Digital Development Facts and Figures 2023 International Telecommunication Union Telecommunication Development Sector*. https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-use/

[19] GSMA. (2024). *The Mobile Economy Sub-Saharan Africa 2024*. GSM Association. https://www.gsmaintelligence.com/research/research-file-download?reportId=50121&assetId=12080

[20] GSMA. (2024)

➜ Latency: the hidden barrier

Beyond simple connectivity metrics, latency presents a particular challenge for AI services:

- Average round-trip latency from Africa to major cloud regions in Europe or North America ranges from 180-300ms
- Latency-sensitive AI applications like real-time language processing or computer vision typically require <100ms latency for optimal user experience
- Limited submarine cable connectivity and sparse regional interconnection points contribute to these latency issues
- Existing submarine cables connecting Africa have a combined capacity of approximately 300 TeraBits Per Second (Tbps), compared to over 1,500 Tbps connecting North America and Europe

These latency challenges are particularly problematic for real-time AI applications in sectors like healthcare, finance, and education, where response time significantly impacts utility.

SUSTAINABLE AND FRUGAL OFFLINE AI

Although LLMs offer unprecedented AI capabilities in natural language and automated content tasks, their rapid growth comes at an environmental cost, presenting considerable energy consumption challenges.

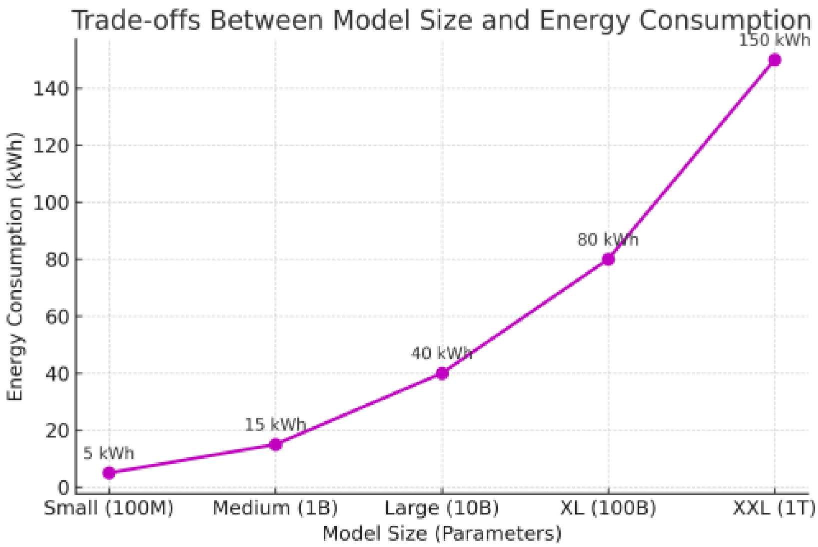


Figure 3: Trade-offs Between Model Size and Energy Consumption[21]

LLMs with large model sizes and with heavy reliance on GPUs (such as Open AI's GPTs) require more computation and memory bandwidth and consume more power (Figure 3) . Training these models demands extensive computational resources, leading to substantial energy consumption. This energy use results in carbon emissions from fossil fuel combustion and significant water usage through evaporation or pollution during power generation, data center operations, and hardware production process. Even the environmental impact of developing relatively small models with 13 billion parameters (the largest models are up to 140 billion parameters) is equivalent to the amount of water consumed by one average person in the United States in about 7.5 years and models up to 20 billion parameters producing the equivalent 24.5 years of water consumption[22].

[21] Chowdhury, M. N. U. R., Haque, A., & Soliman, H. (2025). *The Hidden Cost of AI: Unraveling the Power-Hungry Nature of Large Language Models*. Computer Science and Mathematics. https://doi.org/10.20944/preprints202502.1676.v1

[22] Morrison, J., Na, C., Fernandez, J., Dettmers, T., Strubell, E., & Dodge, J. (2025). *Holistically Evaluating the Environmental Impact of Creating Language Models* (No. arXiv:2503.05804). arXiv. https://doi.org/10.48550/arXiv.2503.05804

This not only makes them unsuitable for use in regions with connectivity issues but also costly to run and use[23]. The cost of running these models, combined with the scarcity and high price of GPU resources, prevents local communities from building and maintaining their own AI solutions. This is especially true where electricity, connectivity, and financial resources are limited, creating a significant hurdle for equitable AI deployment This is particularly true since most current memory-augmented frameworks in use today have built-in significant energy inefficiencies. As a result, actors working in resource-constrained environments, as is the case with much of the global south face disproportionate efficiency penalties[24] meaning that although they have access to LLMs, these will be less capable and effective in meeting local needs.

In resource constrained environments, offline AI using frugal, affordable technology offers several advantages.

- Resource-constrained edge devices such as the Raspberry Pi used in Bibliothèques Sans Frontières Ideas Cube can, with the use of optimized, customised models, offer high accuracy and response times but with lower amounts of energy consumption and financial investment inexpensive hardware[25].

The timing gap: infrastructure development vs. AI advancement

A critical dimension of the infrastructure challenge involves timing—the rate of AI advancement compared to infrastructure development timelines creates a fundamental mismatch that threatens to widen rather than narrow the digital divide.

→ AI Development acceleration

The pace of AI advancement has accelerated dramatically in recent years, creating rapidly evolving capabilities and dependencies. Computing power used in frontier AI

[23] Wu, L., Liu, X., Shi, T., Ye, Z., & Song, D. (2025). *DeServe: Towards Affordable Offline LLM Inference via Decentralization* (No. arXiv:2501.14784). arXiv. https://doi.org/10.48550/arXiv.2501.14784

[24] Wu, H., Wang, X., & Fan, Z. (2025). *Addressing the sustainable AI trilemma: A case study on LLM agents and RAG* (No. arXiv:2501.08262). arXiv. https://doi.org/10.48550/arXiv.2501.08262

[25] Velaga, K. S., & Guo, Y. (2024). Optimizing Large Language Models Assisted Smart Home Assistant Systems at the Edge: An Empirical Study. *OpenReview.Net*.

models has doubled approximately every 3.4 months since 2012—a rate far exceeding Moore's Law. This exponential growth has enabled rapid capability advancement with new breakthrough applications emerging regularly.

New capabilities emerge quickly, with significant advancements often occurring in 6-12 month cycles. The progression from GPT-3 to GPT-4, for example, introduced substantial new capabilities within approximately 18 months, changing expectations and applications across the industry. This rapid evolution creates moving targets for development planning and policy responses.

Commercial applications based on these capabilities follow quickly, creating new economic opportunities and dependencies. Industries from healthcare to financial services are rapidly incorporating AI capabilities into core operations, establishing new competitive baselines that affect global markets. Those without access to these capabilities face growing competitive disadvantages across multiple sectors.

Technical debt accumulates rapidly for those unable to keep pace with these developments. Organizations and regions that delay AI adoption find themselves facing not only the current implementation gap but also accumulated adaptation challenges that grow more complex over time. This accumulating disadvantage creates a compound effect that becomes increasingly difficult to overcome.

➜ Infrastructure development timelines

In contrast, physical infrastructure development follows much longer timelines that cannot match the pace of technological advancement. New submarine cable projects typically require 3-5 years from planning to operation, with complex international projects often facing additional delays due to regulatory approvals, financing challenges, and construction complications.

Major data center developments average 18-36 months from inception to completion under ideal circumstances. In emerging markets with less developed supply chains, regulatory frameworks, and technical expertise, these timelines often extend significantly. Even relatively straightforward deployments face delays related to equipment importation, site preparation, and workforce development.

Power generation projects generally require 3-7 years from planning to commercial operation, with renewable energy projects often at the longer end of this spectrum due to land acquisition, grid integration, and financing complexity. In regions with limited existing capacity, these energy projects become prerequisites for data center development, creating sequential dependencies that further extend timelines.

Last-mile connectivity expansion programs often extend over 5-10 year timeframes, particularly in rural and low-density areas where commercial incentives for infrastructure development are limited. These programs face challenges related to rights-of-way, maintenance capabilities, and sustainable business models, all of which extend implementation timelines.

This timing mismatch creates a situation where infrastructure gaps are likely to widen rather than narrow in the near term, as AI capabilities and applications advance more rapidly than the supporting infrastructure in emerging markets. This divergence threatens to create a self-reinforcing cycle where regions with existing infrastructure advantages can adopt and benefit from AI more quickly, generating resources for further infrastructure investment, while those without such advantages fall further behind.

Sovereignty and Dependency Concerns

The infrastructure gap creates profound implications for technological sovereignty and economic independence, raising questions about control, value capture, and strategic autonomy in an AI-driven future.

→ Technological dependency

The infrastructure gap creates profound implications for technological sovereignty and economic independence, raising questions about control, value capture, and strategic autonomy in an AI-driven future.

Without local infrastructure, emerging economies face increasing technological dependency with far-reaching implications. Critical data must be processed and stored in foreign jurisdictions, creating concerns about privacy, security, and compliance with local regulations. This extraterritorial data storage creates legal ambiguities and sovereignty challenges that many nations find increasingly problematic.

AI capabilities become accessible only through APIs controlled by foreign entities, creating dependency relationships that extend beyond simple service provision. These relationships often include unilateral terms of service, pricing structures, and feature limitations that prioritize the provider's interests over local needs. Access can be restricted or modified based on business decisions, political considerations, or regulatory changes entirely outside local control.

Pricing, feature availability, and terms of service remain outside local control, creating unpredictable operating environments for businesses and public services trying to leverage AI capabilities. This uncertainty complicates planning and investment, further hampering local adoption and development. Abrupt changes in pricing or availability can disrupt critical services and business models with little recourse.

Data governance follows foreign regulatory frameworks rather than local priorities, creating potential conflicts with cultural norms, social values, and development objectives. Content policies developed for Western markets may restrict legitimate local content, while safety mechanisms may fail to address locally relevant concerns. These misalignments can have significant implications for how AI systems serve—or fail to serve—local needs.

→ Economic value capture

The infrastructure gap directly impacts where economic value from AI is captured, with significant implications for economic development pathways. Cloud computing and AI services represent a global market exceeding $500 billion annually, growing at 25-30%. Without local infrastructure, 70-80% of this value typically flows to infrastructure providers in developed markets, creating a substantial economic outflow.

Local economies are relegated to lower-value implementation roles rather than core technology development. This position in the value chain limits economic returns, knowledge development, and competitive advantage. The long-term economic impact of this positioning can reinforce existing development disparities rather than enabling convergence.

Intellectual property development and retention becomes challenging without adequate local infrastructure for research and development. The resulting IP imbalance affects not only economic returns but also the direction of technology development itself, which increasingly reflects the priorities and perspectives of regions with strong R&D infrastructure rather than global needs.

The data generated by local activity—increasingly valuable as training inputs for AI systems—flows primarily to external entities that can capture and monetize this value. This data outflow represents a significant missed opportunity for local value creation and poses questions about equitable compensation for data contributions to global AI advancement.

→ Strategic and security implications

Dependency on foreign infrastructure raises significant strategic concerns with implications beyond economics. Critical national systems becoming vulnerable to foreign service disruptions creates resilience challenges for essential services like healthcare, finance, and government operations. This vulnerability may limit policy autonomy in international relations due to fear of service disruptions.

Limited ability to enforce local data sovereignty requirements compromises regulatory authority and may conflict with cultural or legal norms around privacy and data protection. This regulatory challenge extends to content moderation, algorithmic transparency, and other governance aspects of AI systems that affect social cohesion and democratic processes.

Potential for extraterritorial application of foreign regulations creates complex compliance environments where local entities must navigate potentially conflicting legal requirements. This regulatory uncertainty can chill innovation and adoption, particularly in sensitive domains like healthcare, finance, and government services.

Restricted ability to develop specialized applications for sensitive government functions creates security concerns and capability gaps in areas like defense, law enforcement, and critical infrastructure protection. These limitations may affect not only digital capabilities but national security more broadly as AI becomes increasingly integrated into security operations.

Emerging Solutions: Alternative Approaches to AI Infrastructure

Despite these significant challenges, several promising approaches are emerging to address infrastructure limitations in ways that may enable emerging markets to participate meaningfully in the AI revolution without requiring full replication of traditional infrastructure models.

→ Edge AI and local decentralized models

The development of lightweight AI models capable of running on edge devices presents one of the most promising paths forward for regions with limited centralized infrastructure.

Recent advancements in model efficiency have dramatically reduced the computational requirements for effective AI capabilities.

Small language models like Phi-3, Gemma, and Llama 3 are dramatically reducing computational requirements while maintaining impressive capabilities. The 8B parameter Phi-3 model can run on devices with as little as 16GB RAM—a specification achievable on mid-range laptops rather than specialized infrastructure. This represents a democratization of access that was unimaginable with earlier generations of models requiring hundreds of billions of parameters.

Techniques such as pruning, knowledge distillation, and early-exit inference create more efficient execution paths that maintain performance while reducing resource requirements. Pruning removes redundant or less-important connections in neural networks, reducing computation without significantly affecting accuracy. Knowledge distillation transfers the capabilities of larger models to smaller ones through targeted training approaches. Early-exit inference allows simpler queries to be resolved with partial network traversal, saving computation for straightforward tasks.

Hardware-aware training can optimize models specifically for available computing resources, creating specialized versions that perform optimally on particular devices or architectures. This approach allows for adaptation to locally available hardware rather than requiring standardized infrastructure deployments. By embracing heterogeneity rather than fighting it, these approaches may enable more sustainable local AI ecosystems.

Recent distillation techniques have enabled models with <1B parameters to achieve performance comparable to much larger models in specific domains. These specialized models can run effectively on commodity hardware without GPU acceleration, opening possibilities for deployment in resource-constrained environments. By focusing on specific high-value use cases rather than general-purpose capabilities, these approaches maximize utility within existing infrastructure constraints.

Pleias: small is beautiful

Pleias developed two small frugal language models : a 350 million parameters (Pleias-Pico) and a 1.2 billion parameters (Pleias-Nano).

- Both Nano and Pico were developed with a focus on computational efficiency and RAG capabilities.
- Pleias-Pico achieves notable efficiency through continuous pretraining on a 45-billion-token dataset. This process yields a lightweight architecture that is adept at structured query handling, verifiable source analysis, and grounded response generation, making it a promising option for settings with limited computational resources.
- Pleias-Nano builds on similar principles of in-built RAG capabilities while incorporating enhanced multilingual support and a higher capacity for managing more complex retrieval tasks. Its design represents a balance between scale and efficiency, with preliminary evaluations suggesting competitive performance relative to larger models, despite lower computational overhead.

Both models employ specialized tokens for queries, source identification, and citation parsing, which helps ensure that outputs remain closely linked to verified sources. Their frugal architectures not only help to minimize inference costs and environmental impact but also provide scalable solutions for a range of applications from academic research to practical deployment.

HuggingFace developed three small, frugal language models: one with 135 million parameters (SmolLM-135M), another with 360 million parameters (SmolLM-360M), and a larger model with 1.7 billion parameters (SmolLM-1.7B).

- All three models were developed with a focus on computational efficiency and high-performance reasoning capabilities. These smaller models present important ethical tradeoffs, potentially democratizing AI access while reducing environmental impact.

- SmolLM-135M achieves notable efficiency through training on a high-quality 600-billion-token dataset. This results in a lightweight architecture that excels at common sense reasoning, world knowledge tasks, and educational content generation, making it ideal for environments with limited computational resources.

- SmolLM-360M builds on similar principles while offering enhanced capabilities for handling more complex queries and knowledge retrieval tasks. Its design strikes a balance between model size and performance, with early evaluations suggesting competitive results compared to larger models, despite its smaller computational footprint.
- SmolLM-1.7B offers a higher capacity for more sophisticated tasks, including advanced reasoning and coding tasks, while maintaining efficiency in both training and inference. It represents a balanced approach to scaling, providing high performance without significant resource demands. This family of models demonstrates how AI development can prioritize environmental sustainability without sacrificing functionality. Additionally, their smaller scale promotes scientific reproducibility, as researchers with modest computational resources can verify results and build upon these models.

Creating tiny but mighty AI models matters because they bring powerful technology to everyone while being kinder to our environment.

Giada Pistilli, Principal Ethicist, HuggingFace

Quantization approaches reducing precision from FP16 to INT4 or INT2 can decrease memory requirements by 4-8×, further expanding deployment possibilities. These techniques, while introducing some performance tradeoffs, enable functional AI capabilities on devices previously considered insufficient. The rapid advancement in these efficiency techniques suggests continuing improvements that will further reduce infrastructure requirements.

Mozilla's Llama Files for local solutions

With the increasing reliance on cloud-based LLM applications, concerns about privacy, data usage, and control are growing. Mozilla created llamafile to provide an alternative that allows users to run LLMs locally on their own hardware, ensuring privacy and control over their data. This initiative also helps avoid the monopolization of AI by centralized tech companies, promoting a more transparent, open-source future.

Llama Files is an open-source project that simplifies running large language models (LLMs) on personal computers. It combines llama.cpp (by Georgi Gerganov) and Cosmopolitan Libc (by Justine Tunney) into a single executable file compatible with six operating systems (Windows, macOS, Linux, FreeBSD, OpenBSD, and NetBSD) and multiple hardware architectures (AMD64, ARM64). This eliminates the need for complex installation, making open-source AI accessible to everyone. The project was developed by Georgi Gerganov and Justine Tunney, with support from Mozilla Internet Ecosystem program.

Llama Files enables local AI, meaning AI runs on users' devices, not in the cloud. This ensures privacy, offline accessibility, and control over data. It represents a shift toward reducing corporate dominance in the AI space and providing powerful AI tools to all users.

These advancements open new possibilities for local AI deployment without the need for massive data center infrastructure. By bringing computation closer to users, Llama Files addresses connectivity limitations while maintaining performance, particularly in regions with limited centralized infrastructure.

➜ Mobile-first AI deployment

Leveraging the relatively high mobile penetration in emerging markets offers another pathway to AI deployment that works with existing infrastructure rather than requiring massive new investments. On-device inference capabilities have advanced significantly, with modern smartphone SoCs capable of executing 5-15 TOPS (trillion operations per second)—sufficient for many practical AI applications.

Specialized mobile AI chips like Google's Tensor and Apple's Neural Engine enable efficient local processing for common AI tasks. While these advanced chips appear primarily in higher-end devices, their capabilities are gradually moving downstream to more affordable smartphones. This democratization of on-device AI processing creates opportunities for applications that work effectively despite connectivity limitations.

Hybrid approaches using both cloud and device capabilities can optimize for local constraints, performing critical processing locally while leveraging cloud resources when connectivity allows. These adaptive systems can provide degraded but functional experiences during connectivity interruptions, improving overall reliability in challenging environments. Such resilient design approaches specifically address the infrastructure realities of emerging markets rather than assuming ideal conditions.

Mobile-based AI applications can reach users despite fixed infrastructure limitations, providing valuable services in healthcare, agriculture, education, and financial inclusion. Healthcare diagnostic tools operating primarily offline can provide valuable screening and triage capabilities in rural areas. Agricultural advisory systems with minimal connectivity requirements can improve productivity and resilience for smallholder farmers. Educational applications designed for intermittent connectivity can supplement formal education systems with personalized learning support.

These mobile-centric approaches align with existing technology adoption patterns in emerging markets, building on the smartphone as the primary computing device rather than requiring new infrastructure deployment. By designing AI systems specifically for these constraints rather than importing models from developed markets, more appropriate and sustainable solutions can emerge.

→ Decentralized training approaches

Novel approaches to decentralized model training offer potential pathways to sovereign AI development despite infrastructure limitations. These approaches distribute the computational burden across multiple smaller systems rather than requiring massive centralized infrastructure, potentially enabling collective capabilities that exceed individual resources.

Federated learning techniques enable model training across distributed devices without centralizing data. Prime Intellect's approach combines federated learning with differential privacy to enable privacy-preserving distributed training across heterogeneous devices. This methodology allows organizations to leverage existing computing resources for model development rather than requiring dedicated infrastructure, significantly reducing capital requirements.

Swarm learning methodologies using blockchain-based coordination to manage distributed training offer another approach to decentralized AI development. These systems establish consensus mechanisms for model updates while providing transparency and accountability in the training process. By distributing both computation and governance, these approaches may enable more collaborative and equitable AI development ecosystems.

Hybrid approaches that combine limited centralized infrastructure with edge device capabilities offer pragmatic middle paths for regions with partial infrastructure development. These approaches can leverage whatever centralized capabilities exist while supplementing them with distributed resources, creating flexible systems that can evolve as infrastructure develops. This adaptability may be particularly valuable in dynamic infrastructure environments.

These decentralized approaches could enable communities to develop locally relevant models despite limited centralized infrastructure. By reimagining the AI development process to align with available resources rather than assuming infrastructure abundance, these methodologies offer potential pathways to technological sovereignty despite infrastructure constraints.

Way forward

Addressing the infrastructure challenges for AI deployment in emerging markets requires a balanced approach that combines multiple strategies aligned with local priorities and constraints. No single approach can fully address the complex challenges involved, but complementary strategies can create viable pathways to meaningful AI participation despite infrastructure limitations.

The most promising approaches leverage existing resources and infrastructure through optimization and innovation, recognizing the reality of resource constraints rather than assuming their absence. By designing specifically for these constraints—through edge AI, mobile-first deployment, CPU optimization, and other efficiency techniques—meaningful capabilities can be delivered within existing infrastructure limitations while more comprehensive development proceeds.

Building human capacity alongside physical infrastructure development ensures that investments translate into sustainable capabilities rather than stranded assets. By investing in education, training, and knowledge transfer, emerging markets can develop the expertise necessary to maintain, operate, and adapt AI infrastructure to local needs. This human capacity development may ultimately prove more important than physical infrastructure in determining long-term outcomes.

By pursuing these parallel paths, the Global South can develop models for AI infrastructure that are appropriately scaled, environmentally sustainable, and aligned with local economic and social priorities. These approaches may differ significantly from infrastructure models in developed markets but may ultimately prove more appropriate and sustainable for their contexts.

ADOPTION CHALLENGES

The implementation of large language models in social impact sectors presents significant challenges that extend beyond technical considerations. While infrastructure limitations affect deployment globally, the challenges are particularly acute in the Global South, where contextual disparities, resource constraints, and governance gaps create complex implementation environments.

➜ The mismatch between generalist technology and specialized domains

The gap between general-purpose AI systems and the specialized requirements of regulated domains represents perhaps the most significant implementation barrier. Khan Academy's experience with Khanmigo illustrates this challenge in educational contexts. Despite extensive fine-tuning and domain-specific guardrails, early classroom implementations revealed substantial misalignment between the system's responses and the structured, standards-aligned content required for effective teaching. Teachers in pilot programs reported spending significant time modifying or contextualizing AI outputs to meet curricular requirements – an unsustainable model for widespread adoption, particularly in schools with limited staff resources.

This mismatch becomes even more pronounced in Global South contexts, where educational frameworks, healthcare protocols, and agricultural practices may differ substantially from the Western contexts that dominate AI training data. When educational AI systems trained primarily on U.S. and European curricula are deployed in countries with different educational traditions and objectives, the gap between system capabilities and local requirements widens considerably. A 2023 implementation of AI-assisted learning tools in rural schools across three states in India found that approximately 70% of teacher queries required substantial modification of AI responses to align with local curriculum requirements and cultural contexts[26].

In healthcare settings, this specialization gap creates significant implementation challenges with potential consequences for patient safety. While ChatGPT, was capable of handling general patient education and administrative tasks, it produced concerning inaccuracies in specialty areas like oncology and cardiology[27]. The study found instances where the system generated plausible-sounding but clinically inappropriate

[26] Équipe du Rapport mondial de suivi sur l'éducation, *Global education monitoring report, 2023*: *technology in education*: *a tool on whose terms?*

[27] Cascella, M., Montomoli, J., Bellini, V., & Bignami, E. (2023). Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *Journal of medical systems*, *47*(1), 33.

recommendations, highlighting risks particularly acute in settings where specialist verification may be limited. For healthcare implementations in regions with distinctive epidemiological profiles or treatment protocols that differ from dominant medical literature, these risks multiply significantly.

Agricultural implementations face parallel domain specialization challenges. The AgroLLM project found that effective agricultural advice requires understanding of highly localized factors including soil conditions, weather patterns, available resources, and traditional practices[28]. When agricultural AI systems trained predominantly on industrial farming data from temperate regions are deployed in tropical or semi-arid contexts with smallholder farming practices, the mismatch between system capabilities and local requirements creates implementation barriers that extend far beyond technical performance metrics.

→ System opacity and implementation trust

The "black box" nature of large language models creates significant operational challenges across implementation contexts. In educational settings, when AI systems generate instructional content or student feedback, educators cannot easily determine the basis for these outputs or verify their alignment with pedagogical objectives. This opacity undermines teacher confidence and creates implementation resistance, particularly in contexts where educators bear ultimate responsibility for learning outcomes.

This transparency deficit becomes especially problematic in Global South implementations, where AI systems may operate at greater cultural and contextual distances from their deployment environments. A 2023 study of AI-assisted healthcare implementations across community clinics in Southeast Asia found that clinician trust represented the most significant barrier to effective adoption. Healthcare providers consistently reported discomfort with incorporating AI recommendations they couldn't verify against local clinical [29]knowledge or reconcile with distinctive patient populations and resource constraints.

[28] Samuel, D. J., Skarga-Bandurova, I., Sikolia, D., & Awais, M. (2025). AgroLLM: Connecting Farmers and Agricultural Practices through Large Language Models for Enhanced Knowledge Transfer and Practical Application. *arXiv preprint arXiv:2503.04788*.

[29] Kee Mun Wong B., Vengusamy S., Bastrygina T., (2024). "Healthcare digital transformation through the adoption of artificial intelligence", in *Information Technologies in Healthcare Industry*. https://www.sciencedirect.com/science/article/abs/pii/B9780443215988000142

The implementation consequences of this opacity extend beyond initial resistance to encompass broader governance challenges. Without clear visibility into how AI systems generate outputs, developing appropriate oversight mechanisms, intervention protocols, and accountability frameworks becomes significantly more difficult. This governance challenge is particularly acute in contexts with limited regulatory infrastructure or technical capacity for system evaluation, creating potential risks of either uncritical acceptance or wholesale rejection of potentially beneficial technologies.

→ The autonomy paradox in implementation contexts

AI implementations must navigate complex tensions between system independence and human control across social impact domains. In educational settings, teachers express concerns about maintaining appropriate oversight while incorporating AI capabilities into instructional practices. The NYC Department of Education's withdrawal of a $1.9 million AI reading tutor proposal stemmed partially from teacher unions raising concerns about appropriate human oversight and potential impacts on student-teacher relationships[30].

This autonomy tension manifests differently across implementation contexts. In well-resourced environments, concerns often focus on maintaining professional judgment alongside AI capabilities. In more resource-constrained settings, particularly in the Global South, autonomy concerns frequently revolve around dependency risks and control over technological trajectories. When agricultural AI systems enter rural farming communities with limited technical infrastructure, for example, implementation creates potential dependencies that farmers may have limited capacity to manage or modify as conditions change.

The autonomy challenge extends beyond individual professional control to encompass broader questions of community and institutional self-determination in technological adoption. When AI systems designed according to external priorities and optimization criteria enter diverse implementation contexts, they may displace local decision-making frameworks and evaluation criteria that reflect different priorities and values. This tension emerges clearly in educational implementations, where AI systems optimized for standardized assessment performance may conflict with local educational traditions emphasizing different learning objectives or pedagogical approaches.

[30] Zimmerman, A. (2024, December 18). NYC withdraws $1.9 million proposal for AI reading tutor after criticism from finance watchdog. *Chalkbeat New York*. https://www.chalkbeat.org/newyork/2024/12/18/nyc-pulls-pep-contract-for-ai-reading-tutor-eps-learning/

Governance challenges in complex implementation environments

→ Regulatory frameworks struggling with technological reality

Current governance structures across social impact domains were established before the emergence of large language models, creating profound misalignments between regulatory frameworks and technological capabilities. Existing regulations focus on traditional concerns like data privacy rather than the novel issues of AI-generated content, algorithmic bias, and human-AI collaboration models.

This regulatory gap creates significant operational uncertainty across implementation contexts. Educational governance frameworks, for example, typically divide oversight responsibilities across federal departments, state agencies, local school boards, and professional organizations, creating ambiguity about appropriate approval processes for AI educational tools. Organizations struggle to determine whether these systems should be evaluated as instructional materials, assessment systems, or educational technologies – classifications with substantially different regulatory implications.

The regulatory challenge becomes particularly acute in Global South contexts, where governance frameworks may have even less capacity to address AI-specific considerations. The digital governance landscape in many regions already struggles with basic data protection and privacy oversight; adding the complexities of AI-generated content and algorithmic decision-making creates governance challenges that existing regulatory bodies have limited capacity to address.

Healthcare governance faces parallel challenges, as existing frameworks for medical devices and clinical decision support systems were not designed for adaptive AI technologies. Provider organizations hesitate to implement systems when liability implications remain unclear, particularly when AI recommendations cannot be transparently evaluated against clinical guidelines. This uncertainty creates particular challenges in regions where medical regulatory infrastructure may be limited, potentially creating situations where AI healthcare applications operate with insufficient oversight or face implementation barriers that prevent potentially beneficial applications.

➜ The evaluation crisis across implementation contexts

Perhaps the most acute governance challenge involves the lack of appropriate evaluation frameworks for AI applications in specialized domains. No standardized methods exist for evaluating the quality, accuracy, and domain-appropriateness of outputs in educational, healthcare, or agricultural contexts. Traditional AI evaluation metrics like accuracy percentages or response consistency provide limited insight into domain-specific performance, creating disconnects between technical benchmarks and practical utility.

AI and education benchmarks

This necessitates the development and adoption of robust benchmarks to ensure that these technologies are not only effective but also equitable and aligned with the diverse needs of learners worldwide[31]. Traditional metrics focused solely on technical performance are insufficient to capture the complex realities of educational contexts, particularly in the Global South. Therefore, benchmarks must be designed to evaluate AI systems across a range of critical dimensions:

- **Equity and Inclusion:** AI-powered educational tools should be rigorously assessed for potential biases that could disadvantage particular student populations, including those from marginalized communities, students with disabilities, or learners from diverse linguistic backgrounds. Benchmarks must prioritize fairness in algorithmic design, accessibility for all learners, and culturally responsive content that reflects local values and knowledge systems.
- **Contextual Relevance:** The effectiveness of AI in education is heavily dependent on its ability to adapt to local curricula, pedagogical practices, and linguistic nuances. Benchmarks should evaluate the capacity of AI systems to align with national or regional educational standards, support diverse teaching methodologies, and accommodate the unique learning styles of students in specific cultural contexts.
- **Data Privacy and Security:** Protecting student data is paramount. Benchmarks must evaluate AI systems' adherence to data protection standards, including compliance with local regulations and ethical guidelines. Special attention should be given to the safeguarding of sensitive student information, particularly in vulnerable communities where data breaches could have severe consequences.
- **Transparency and Explainability:** Educational stakeholders, including teachers, students, and parents, need to understand how AI systems arrive at their recommendations and decisions. Benchmarks should prioritize AI models that

[31] AI-for-Education.org. (2025). *AI Benchmarks for Education.* https://ai-for-education.org/ai-benchmarks-for-education/

provide clear explanations of their reasoning processes, enabling educators to critically evaluate and trust AI-driven insights.

- **Replicability**: A significant challenge with existing benchmarks is the frequent failure to report the statistical significance of their findings[32]. Moreover, the methodologies used are often not sufficiently transparent or well-documented to allow for easy replication, making it difficult to validate results and hindering the widespread adoption of these benchmarks.
- **Pedagogical Soundness:** AI systems should enhance, not replace, effective teaching practices. Benchmarks must assess the pedagogical appropriateness of AI-driven interventions, ensuring that they align with established learning theories, promote active student engagement, and foster critical thinking skills.
- **Impact on Learning Outcomes:** Ultimately, AI in education should lead to measurable improvements in student learning outcomes. Benchmarks should evaluate the impact of AI systems on key indicators of academic achievement, such as test scores, graduation rates, and college enrollment, while also considering broader measures of student success, such as social-emotional development and civic engagement.

Agricultural applications face similar evaluation challenges. The effectiveness of agricultural recommendations cannot be measured solely through technical metrics but must consider practical applicability, regional appropriateness, resource requirements, and sustainability implications. As a recent study on LLMs and agricultural services recommended, "evaluation methodologies must move beyond technical performance to assess real-world agricultural outcomes" [33]– a significant shift requiring domain expertise beyond traditional AI evaluation approaches.

The absence of contextually appropriate evaluation frameworks creates significant risks of implementing systems with unknown effects or undetected failure modes, particularly concerning high-stakes domains where errors can have serious consequences for student learning, patient outcomes, or agricultural sustainability. This risk becomes particularly acute in resource-constrained environments where implementation oversight may be limited and consequences of system failures may be more severe due to fewer fallback options.

[32] Reuel, A., Hardy, A., Smith, C., Lamparth, M., Hardy, M., & Kochenderfer, M. J. (2024). *BetterBench: Assessing AI Benchmarks, Uncovering Issues, and Establishing Best Practices* (No. arXiv:2411.12990). arXiv. https://doi.org/10.48550/arXiv.2411.12990

[33] Tzachor, A., Devare, M., Richards, C., Pypers, P., Ghosh, A., Koo, J., ... & King, B. (2023). Large language models and agricultural extension services. *Nature food*, *4*(11), 941-948.

→ Domain-appropriate governance development

Breaking current implementation barriers requires governance frameworks specifically designed for AI applications in specialized domains. Rather than applying generic AI principles or technical performance metrics, effective governance must address the specific requirements, risks, and evaluation criteria relevant to each domain and implementation context.

Educational governance frameworks, for example, must address not just technical performance but pedagogical appropriateness, developmental alignment, and integration with existing educational objectives and approaches. Healthcare governance must ensure clinical accuracy, evidence-based recommendations, and appropriate integration with existing care processes and professional judgment. Agricultural governance must evaluate contextual relevance, resource requirements, and alignment with sustainable and culturally appropriate farming practices.

The development of these domain-specific frameworks requires collaborative processes that bring together diverse expertise: domain professionals who understand implementation requirements, technical experts who comprehend system capabilities and limitations, governance specialists who can develop appropriate oversight mechanisms, and community members who understand local priorities and contextual factors. This collaborative approach recognizes that effective governance cannot emerge from any single perspective but requires integration of diverse knowledge systems.

The U.S. Department of Education's (2023) report on *Artificial Intelligence and the Future of Teaching and Learning*[34] exemplifies this approach by highlighting "centering people" – including educators, parents, students, and policymakers – as a foundational principle for AI governance in education. However, moving from general principles to practical governance frameworks requires sustained collaborative efforts and institutional support beyond what current structures typically provide.

→ Implementation learning networks across diverse contexts

Organizations facing similar implementation challenges can accelerate effective adoption through structured knowledge exchange and collaborative problem-solving. Rather than approaching AI implementation as isolated organizational projects, network approaches enable sharing of adaptation strategies, evaluation methods, and governance frameworks across similar contexts.

---

[34] Cardona, M. A., Rodríguez, R. J., & Ishmael, K. (2023). Artificial intelligence and the future of teaching and learning: Insights and recommendations.

Healthcare implementations demonstrate the value of this networked approach. Hospital systems implementing similar AI applications have established collaborative learning communities that share implementation experiences, adaptation strategies, and governance approaches. These networks enable organizations to identify common challenges, develop shared solutions, and establish evaluation frameworks that address domain-specific requirements beyond technical performance metrics.

Cross-regional learning networks can be particularly valuable for addressing Global South implementation challenges. By connecting organizations facing similar contextual challenges across different regions, these networks can develop adaptation strategies and governance approaches specifically designed for resource-constrained environments or culturally distinctive implementation contexts, creating alternatives to simply modifying approaches designed for different operational realities.

These implementation learning networks become particularly powerful when they connect not just similar organizations but diverse stakeholders across implementation ecosystems: the practitioners who use AI systems, the administrators who manage implementations, the developers who design and modify systems, and the governance professionals who establish oversight frameworks. This multi-perspective approach enables comprehensive understanding of implementation challenges and more effective adaptation strategies than any single stakeholder group could develop independently.

→ The necessity of truly open source AI

The challenges of AI implementation across diverse contexts underscore the fundamental importance of open source approaches that encompass not just model weights but crucially the training data itself. While proprietary, closed systems dominate the current landscape, open source AI represents the clearest path toward addressing many implementation barriers discussed throughout this analysis.

Most current "open source" AI efforts focus primarily on releasing model weights while leaving training data opaque, incomplete, or inaccessible. This half-measure fundamentally limits the potential for meaningful community governance and contextual adaptation. True open source AI must include open training data - making visible and modifiable the actual content that shapes model capabilities, biases, and limitations.

Open training data enables communities to understand what knowledge systems are represented, identify gaps or biases relevant to implementation contexts, and contribute materials that reflect diverse perspectives and priorities. Without this transparency, even technically "open" models remain partially black-boxed, limiting meaningful adaptation particularly for Global South contexts underrepresented in original training materials.

Common Corpus

Even if they are rare, there are truly open source AI initiatives today.

For instance, the release of Common Corpus by Pleias last year - a comprehensive, multilingual public domain dataset comprising over 1 trillion words from diverse open data sources (cultural heritage, governmental open data, open science...) - demonstrates that training fully open and reproducible large language models on copyright-free content is both feasible and scalable.

Another example, FineWeb[35], developed by HuggingFace, offers a 15-trillion-token web-scale dataset meticulously curated from 96 CommonCrawl snapshots.

Both projects enable unrestricted access and use by researchers and developers worldwide.

Open training data creates the foundation for community contributions that serve as both technical resources and governance mechanisms. When communities contribute domain knowledge, contextual understanding, and evaluation criteria to training datasets, they gain partial authority over system capabilities and alignment. This participatory approach transforms AI development from a unidirectional process to a collaborative endeavor that incorporates diverse knowledge systems and priorities.

For implementation in social impact domains, this community participation becomes particularly crucial. Educational AI requires curricula and pedagogical approaches that reflect diverse educational traditions. Healthcare applications need medical knowledge representing various healthcare systems and approaches. Agricultural systems must incorporate farming practices adapted to different environmental contexts and cultural traditions.

[35] FineWeb: decanting the web for the finest text data at scale
https://huggingface.co/spaces/HuggingFaceFW/blogpost-fineweb-v1

→ Resource challenges and structural support

Meaningful open source AI with open training data faces significant resource challenges, particularly for communities with limited technical infrastructure. Addressing these challenges requires structural support that transfers not just access but capacity for meaningful participation. This includes computational resources for local adaptation, technical expertise development in diverse communities, and economic models that sustain community participation without exploitation.

The path forward requires investment in public digital infrastructure that enables diverse communities to participate in open source AI development according to their own priorities and contexts. Without this support, open source approaches risk becoming peripheral alternatives rather than transformative frameworks for more equitable technological development.

Beyond practical implementation benefits, open source AI with open training data represents an ethical imperative in technological development. As AI systems increasingly influence critical social domains, the ability to inspect, understand, modify, and govern these systems should not remain concentrated in a small number of commercial entities or technical communities.

Open approaches create the foundation for technological self-determination - enabling communities to shape AI systems according to their own priorities, knowledge systems, and governance structures rather than adapting to technologies designed for different contexts. This self-determination becomes particularly crucial for Global South implementations, where the distance between development contexts and implementation realities is often greatest.

CONCLUSION

From Promise to Action - Charting a Course for Equitable AI

This white paper has explored both the extraordinary potential and the significant challenges surrounding the deployment of AI for social good, particularly in the Global South. While AI offers unprecedented opportunities to transform education, healthcare, and access to justice, the promise of these technologies remains unevenly distributed. We've seen how linguistic underrepresentation, infrastructure deficits, and biased algorithms threaten to exacerbate existing inequalities, leaving marginalized communities behind. Are we there yet, with AI for Good in 2025? The honest answer is: not quite. But we can be!

We can and must proactively shape AI so that it serves everyone, not just the global north. First, it is imperative to create a more human-centered approach to AI, such as maintaining human oversight. We can harness the transformative potential of AI only if we prioritize ethical AI which is why we need robust benchmarks. The path forward requires a concerted effort from researchers, practitioners, policymakers, and funders to prioritize:

- **Investing in Open Data and Inclusive Infrastructure:** Prioritizing funding for foundational NLP tools (language identification and core processing pipelines) alongside open source platforms for recording, transcription, and annotation of local languages.
- **Empowering Local Communities:** Developing partnerships with grass roots movements and with those experiencing the problems the AI attempts to solve
- **Prioritizing the Development of Offline AI Technologies** By embracing offline AI, localized content, and community empowerment, we can create a more equitable and environmentally sustainable AI future that delivers benefits without further burdening our planet or perpetuating technological dependencies.

This is the way forward. To ensure these efforts are successful, key indicators to monitor will include an increase in the number of LLMs that can handle African languages at parity with English, greater local ownership of AI projects (measured by the percentage of projects led by local institutions), and policy adoption of AI ethics frameworks in Global

South countries. Support from development banks, impact investors, and philanthropic tech funds will be critical to these efforts

In conclusion, the challenges are substantial, but the opportunities are even greater. By embracing a holistic, context-aware approach, and working collaboratively across sectors, we can harness the transformative potential of AI to create more equitable and effective opportunities for all, regardless of their background or location. Let us seize this moment to build an AI-powered future that truly leaves no one behind.

BSF’s social impact projects

1/2

zeljkosantrac

Practical Implementations in the Field

The IDEAS AI initiative by Bibliothèques Sans Frontières demonstrates the transformative potential of artificial intelligence through concrete, scalable projects deployed across diverse regions. By focusing on open-source, frugal technologies that operate effectively offline and prioritizing systematic localization and community engagement, IDEAS AI is addressing critical challenges related to linguistic diversity, digital infrastructure gaps, and technological sovereignty.

The following use cases exemplify how practical AI interventions are making significant contributions toward educational equity, health improvements, and economic empowerment in diverse and challenging environments.

BSF's social impact projects

USE CASE 1 | **Supporting teacher language training**

Teachers are the cornerstone of a successful education system, cultivating learning-conducive environments, fostering critical-thinking skills, and promoting students' social and emotional well-being. However, in Senegal, teaching quality faces significant challenges. Despite French being the official language of instruction, a certain proportion of teachers struggle with its use as a professional language, with only 4.4% speaking it as a first language and a mere 20% using it as their primary teaching language. Although bilingual education is an excellent way to ensure learning, in schools where students and teachers may not share a common language a lingua franca can be essential.

PleIAS, Kajou, and BSF have collaborated to develop the Karibu application, originally designed by BSF to support refugee integration in France and Belgium. The app helps accelerate the acquisition of essential French language skills, fostering autonomy in daily life—particularly for administrative tasks, healthcare access, and professional integration.

To address this challenge, we designed, in collaboration with the Senegalese Education Ministry, a Proof of Concept (POC) in Senegal, tailoring Karibu to local needs. The offline version includes generative AI on microSD cards, ensuring access to language learning resources even without internet connectivity. The online version features Karibot, an AI tutor designed specifically for teachers. It corrects written work and provides instant, personalized feedback on grammar, syntax, and vocabulary, creating customized learning paths that build on each teacher's strengths and address areas for improvement, fostering genuine progress.

BSF's social impact projects

USE CASE 1 | *Supporting teacher language training*

Figure 1: Reading and writing exercises in the Karibu language learning application

Key Innovations and Solutions:

- **Offline and Mobile-first Approach:** Accessible AI solutions embedded on microSD cards to overcome internet connectivity limitations common in remote regions.
- **Localized and Contextualized Content:** Exercises and feedback tailored to local cultural, linguistic, and professional contexts.
- **Personalized Learning Experience:** Adaptive exercises generated based on learner proficiency, aligned with international standards (CEFR), with instant AI-driven feedback enhancing teachers' linguistic competencies.

Expected Outcomes:

- Enhanced confidence and competence among teachers in their professional language capabilities.
- Improved quality and effectiveness of teaching in French.
- Creation of positive classroom environments that foster better student outcomes.

By adapting Karibu to local needs, this application directly provides a cost-effective, scalable solution for linguistic and educational empowerment in the Global South.

USE CASE 2 | **kSANTÉ (My Daily Health) - AI for Community Health Workers in West Africa**

kSANTÉ is an innovative online learning solution and intelligent chatbot meticulously designed to empower Community Health Workers (CHWs) operating in the often-challenging environments of Senegal and Côte d'Ivoire. The core objective of kSANTÉ is to equip these essential frontline healthcare providers with the knowledge and tools necessary to significantly enhance the quality, reach, and effectiveness of healthcare services they deliver within their respective communities.

The healthcare systems of West Africa, often characterized by resource constraints and limited infrastructure, heavily rely on CHWs as the critical first point of contact for a vast segment of the population, particularly in remote and underserved areas. However, despite their pivotal role, CHWs frequently encounter substantial hurdles in gaining access to comprehensive and up-to-date training resources and essential logistical support required to perform their duties effectively.

Unfortunately, CHWs are frequently under-trained, under-resourced, and subject to high turnover rates. This combination of factors results in compromised healthcare service delivery at the community level, with negative repercussions felt most acutely in critical areas such as preventative care, epidemic management, and the promotion of maternal and child health. The lack of readily accessible training modules and real-time support systems for CHWs contributes to significant disparities in healthcare outcomes across the region.

BSF's social impact projects

USE CASE 2 | **kSANTÉ (My Daily Health) - AI for Community Health Workers in West Africa**

Our approach is to build an AI-driven assistant to support CHWs with:

- A carefully fine-tuned, state-of-the-art open-source 7B Large Language Model (LLM) tailored to the specific healthcare context of West Africa. This LLM will serve as the core intelligence driving the kSANTÉ chatbot, ensuring accurate and relevant responses.
- A user-friendly, intuitive, and accessible customized interface seamlessly integrated into the existing Kajou mobile phone application. This strategic integration maximizes reach and leverages an established platform familiar to many CHWs.
- A curated collection of 19 comprehensive offline training modules thoughtfully developed in close collaboration with the Ministries of Health in Senegal and Côte d'Ivoire. These modules will provide CHWs with access to a wealth of evidence-based information on a wide range of healthcare topics, accessible even in areas with limited or no internet connectivity.
- An intelligent chatbot powered by the fine-tuned LLM designed to furnish CHWs with accurate, contextualized, and up-to-date information on diseases, treatment protocols, and patient care guidelines. The chatbot will enable real-time interaction through text-based inquiries and provide personalized feedback to enhance the learning experience.
- Adaptation to national languages: Wolof in Senegal and Dioula in the Ivory Coast.

Figure 2: Questions and answers with audio replies along with the interactive AI assistant

Expected Outcomes:

- **Empowering CHWs:** Equips community health workers with expanded skills and in-depth knowledge across critical healthcare domains, boosting their confidence and competence in providing essential services.
- **Improving Service Quality:** Enhances the quality, speed, and cultural relevance of healthcare services delivered by CHWs, leading to better patient outcomes and increased community trust.
- **Boosting Motivation:** Fosters increased CHW motivation through continuous learning opportunities, personalized support, and streamlined access to accurate and timely information.

By providing CHWs with an AI-driven assistant accessible through their mobile phones, kSANTÉ seeks to significantly enhance healthcare service delivery within communities across West Africa.

*USE CASE 3 | **Mobilizing Artificial Intelligence to Prevent and Fight Against Conflict-Related Sexual Violence***

Conflict-related sexual violence (CRSV) remains a significant issue in conflict zones. States are often unaware of their obligations under international law, which affects their ability to prevent and respond to such violence effectively.

To address this problem, we are piloting a fine-tuned 350M parameter Large Language Model (LLM) with a custom interface to enhance the Red Line Guidebook. This AI-driven tool will provide intelligent search functionalities, personalized recommendations, automated content categorization, customized reports, summarization tools, and document generation to improve access to and management of legal resources for CRSV actors working to support survivors and prevent CRSV-based crimes.

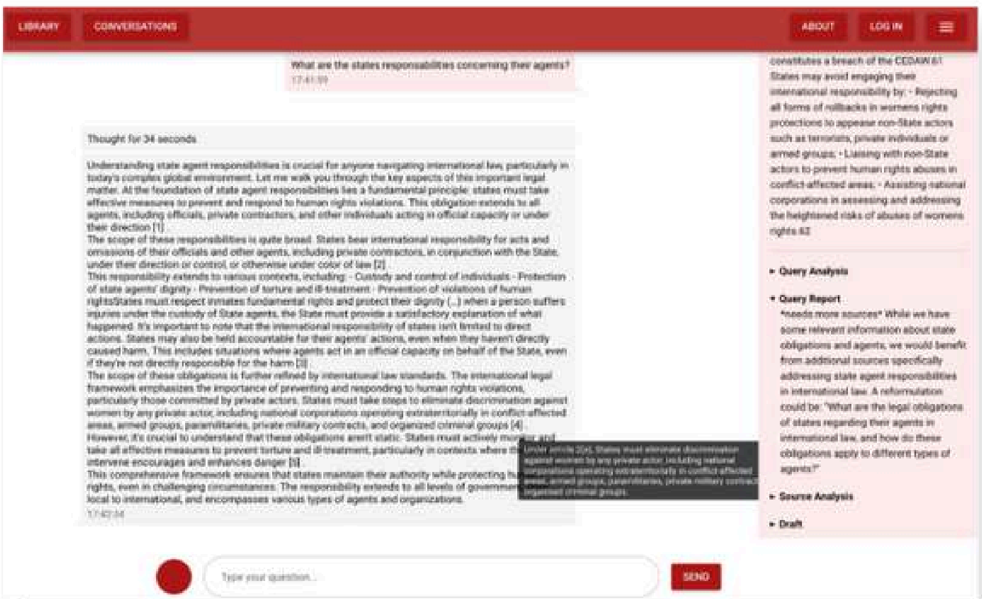


Figure 2: Questions and answers with audio replies along with the interactive AI assistant

BSF's social impact projects

*USE CASE 3 | **kSANTÉ (My Daily Health) - AI for Community Health Workers in West Africa***

Key Innovations and Solutions:

- **Multilingual:** Text and document retrieval and generation in English, French, and Ukraine for users in Nigeria, Ukraine, Democratic Republic of Congo.
- **Intelligent Legal Resource Access:** AI-powered search and recommendation functionalities that streamline access to relevant legal information for CRSV prevention and response.
- **Personalized Legal Guidance:** Customized reports and document generation capabilities tailored to the specific needs of legal professionals and advocacy groups.
- **Automated Content Management:** AI-driven content categorization and summarization tools that improve the efficiency of managing and utilizing legal resources.

Key Innovations and Solutions:

- Improved awareness and understanding of international legal obligations related to CRSV among relevant actors.
- Enhanced capacity of legal professionals and advocacy groups to effectively prevent and respond to CRSV.
- Increased efficiency in accessing, managing, and utilizing legal resources for CRSV prevention and response.

By mobilizing artificial intelligence to enhance the Red Line Guidebook, this project aims to empower actors in conflict zones to more effectively prevent and combat conflict-related sexual violence.

USE CASE 4 | *Mobilizing Artificial Intelligence for Teacher Professional Development*

Teachers often lack access to content and support for ongoing professional training, particularly in regions with limited or no internet access, hindering their ability to effectively deliver quality education.

To address this problem, we are creating an AI-driven professional development assistant for teachers. This assistant will be deployed on the 70 Ideas Cubes Raspberry Pi servers and teacher mobile devices in the Kédougou region.

Expected Outcomes:

- Improved Lesson Planning: Teachers will be better equipped to develop engaging and effective lesson plans.
- Enhanced Digital Skills: Teachers will gain greater proficiency in using digital tools to support their teaching.
- Increased Green Skills: Teachers will be empowered to integrate environmental sustainability concepts into their curriculum.
- Greater Gender Equality Skills: Teachers will be better prepared to promote gender equality in their classrooms.
- Improved Pupil Access: Students will have enhanced access to the curriculum, leading to better learning outcomes.

Key Innovations and Solutions:

- **Offline Accessibility:** The AI assistant functions entirely offline, ensuring continuous access to professional development resources regardless of internet connectivity.
- **Comprehensive Skill Enhancement:** Teachers will improve their lesson planning, digital skills, green skills, gender-equality skills, and improve their pupils' ability to access the curriculum.
- **Ideas Cube Integration:** Taking advantage of a the current deployment of Ideas Cube microsevers across the region ensuring local offline Internet access with built in AI-based resources and tools

By deploying this AI-driven assistant on existing Ideas Cube infrastructure and mobile devices, we aim to empower teachers in the Kédougou region with the skills and resources they need to provide high-quality education to their students.

Project	Description	Impact of AI	Impact measurement practices	Responsible AI practices	SDGs
Karibu	AI-driven French language learning app with offline access, tailored for Senegalese teachers, featuring AI-powered tutor.	Enhanced teacher confidence/competence, improved teaching quality, and positive classroom environments.	Surveys, pre- and post-testing of teachers' language skills, classroom observation data	Localized content, offline access, personalized learning paths, collaboration with the Senegalese Education Ministry.	4.1, 4.5
My Daily Health	AI-driven assistant for CHWs in West Africa, providing offline training modules, a chatbot with healthcare information in local languages.	Empowered CHWs with expanded skills/knowledge, improved service quality and cultural relevance, increased CHW motivation.	Healthcare outcome data, CHW performance metrics, qualitative feedback from CHWs and patients	Open-source LLM, fine-tuned for the local healthcare context, offline accessibility, adaptation to local languages, collaboration with Ministries of Health.	3.4, 3.8
AI for CRSV Preventi on	LLM with a custom interface to enhance the Red Line Guidebook, providing intelligent search, personalized recommendations, content categorization, and document generation for actors involved in CRSV prevention and response.	Improved awareness of international legal obligations, enhanced capacity to prevent/respond to CRSV, increased efficiency in accessing/managing legal resources.	Usage metrics of the tool, qualitative feedback from legal professionals and advocacy groups	Multilingual support, intelligent legal resource access, personalized legal guidance, automated content management.	5.2, 16.3
AI4Teac hers	AI-driven professional development assistant for teachers in the Kédougou region, deployed on Ideas Cubes, offering offline access to training resources for lesson planning, digital skills, green skills, and gender equality.	Improved lesson planning, enhanced digital skills, increased green skills, greater gender equality skills, improved pupil access to the curriculum.	Measurement of teacher skills development through surveys and observation.	Offline accessibility, localized content, comprehensive skill enhancement, Ideas Cube integration.	4.1, 4.5, 4.7

Table 1: The above table summarizes AI-driven projects that showcase impactful interventions in education, health, and access to justice, highlighting their key features, impact measurements, and alignment with sustainable development goals.

Bibliothèques Sans Frontières - April 2025

Bibliothèques Sans Frontières
Libraries Without Borders