AI & EQUALITY
AFRICA

An African AI Toolbox:

Integrating **Human Rights considerations** *along the* **AI lifecycle**

A framework to AI development

AI & EQUALITY

African Centre for Technology Studies
ACTS
Knowledge for Better Livelihoods

ARTIFICIAL INTELLIGENCE FOR DEVELOPMENT AFRICA

WOMEN AT THE TABLE

Executive summary

The **African <AI & Equality> Toolbox** is a strategic initiative designed to empower African stakeholders—policymakers, technologists, civil society actors, and communities—to shape Artificial Intelligence (AI) systems that are contextually relevant, inclusive, and grounded in human rights.

Developed in collaboration with **Women at the Tabl**e and the **African Centre for Technology Studies (ACTS)**, and adapted from the global **<AI & Equality> Human Rights Toolbox Initiative** in collaboration with the UN Office of the High Commissioner for Human Rights (OHCHR), this African iteration provides practical tools and methodologies to guide equitable AI development across the continent.

The Toolbox applies a Human Rights-based AI Lifecycle Framework, integrating reflective questions and the Human Rights Impact Assessment (HRIA) developed with the Alan Turing Institute. It emphasizes participatory, multidisciplinary approaches and is rooted in feminist, decolonial, and Justice, Equity, Diversity, and Inclusion (JEDI) principles and incorporates lessons from emerging digital rights challenges, ensuring AI systems are designed with safety and dignity at their core.

Key sectors explored include:

- **Agriculture:** AI tools co-designed with women farmers, addressing soil health, pest management, and access to market information.
- **Health:** AI-powered malaria diagnostics developed for rural Uganda, focusing on ethical data collection and equitable deployment.
- **Climate:** Environmental sensing initiatives using AI to monitor air and noise pollution in African cities and rural areas, with community-driven deployment and interpretation.
- **Education & Language Inclusion:** Projects integrating NLP for underserved African languages and Kenyan Sign Language translation technologies.
- **Digital Safety:** Addressing technology-facilitated gender-based violence (TFGBV) through AI systems that detect coordinated harassment, protect vulnerable users, and work to alert platforms when there is harm.

The Toolbox serves not only as a resource but as a platform for action—aiming to build African capacity in AI governance, foster interdisciplinary collaboration, and ensure AI advances rights, dignity, and local priorities. It represents a shift from importing Global North models to developing African-led approaches, with communities at the center of innovation. By addressing both opportunities and risks across sectors, the Toolbox ensures AI development considers the full spectrum of human rights impacts.

Initially published in September 2025, the Toolbox is a living document that invites ongoing input and iteration, with the ultimate goal of placing African perspectives at the forefront of global AI development.

> The Toolbox serves not only as a resource but as a **platform for action.**

Acknowledgments

We extend our heartfelt gratitude to the numerous individuals and organizations who contributed to the development of this framework.

Special thanks to **Winston Ojenge** and the African Centre for Technology Studies (ACTS) team for their invaluable insights, support, and guidance throughout this project.

We are deeply grateful to Women at the Table and the <AI & Equality> Human Rights Initiative, particularly:

- **Caitlin Kraft-Buchman,** for her leadership of the global <AI & Equality> Human Rights Initiative and the curation of this Toobox.
- **Emma Kallina** for her leadership and foundational work developing the framework.
- **Sofia Kypraiou** for her work as Data Science Lead and Women at the Table Tech Fellow, and for her EPFL Masters Thesis upon which the AI & Equality AI Life Cycle was initially based.
- **Pilar Grant** for her patience, goodwill, and strategic design skills upon which we have leaned so heavily for this <AI & Equality> African Toolbox, in addition to her leadership on the <AI & Equality> Latin America Toolbox with Chile's National Centre for Artificial Intelligence (CENIA).
- **Amina Soulimani**, <AI & Equality> Community Lead and <A+> Alliance for Inclusive Algorithms lead, for her deeply salient contributions, calm and cool.

Thanks to all for their expertise, patience, and innovation in helping us refine our approach to human rights considerations in AI development.

Our appreciation also goes to **David Kelleher** of Gender@Work for his thoughtful feedback and perceptive support.

We acknowledge the generous support of the **Swiss Human Rights Diplomacy Section, Federal Department of Foreign Affairs FDFA, State Secretariat STS, Division Peace and Human Rights, and the International Development Research Centre (IDRC) and AI4D**, which made this project possible.

Initial Consultations and Working Groups

We thank the members of our initial consultations and working groups for their diverse perspectives and constructive input:

- **Mitchel Ondili** for precious support in initial convenings
- **Agriculture focus group:** Joel Nwakaire, Dr. Daisy Salifu, Ndeye Fatou Mboup, and Angella Ndake
- **Health focus group:** Dr. Rose Nakasi
- **Climate change/sensors focus group:** Alicia Olago
- **Education focus group:** Adekemi Omotubora, Gloriana Monko, Araba Sey, and Dr. Lilian Wanzara

Case Study Contributors

Profound thanks goes to our Case Study contributors who gave considerable time and care to elaborate on their innovative research advancing AI development across the African continent:

- **Dr. Lilian Wanzare,** Chair of the Department of Computer Science at Maseno University, on Natural Language Processing and Kenyan Sign Language
- **Dr. Rose Nakasi** of the Makerere Health Intelligence Lab on novel diagnostic malaria techniques
- **Dr. Daisy Salifu** of ICIPE (International Centre of Insect Physiology and Ecology) on a Ugandan Cassava Farmer initiative
- **Engr. Prof. Joel Nwaeze Nwakaire,** University of Nigeria, Nsukka, on Nsukka Pepper Farmers
- **Alicia Olago** of Code for Africa on climate sensors in rural and urban Kenya
- **Hanna Teshager** of Code for Africa on eleven-country research on technology-facilitated Gender Based Violence

Finally, we express our deepest appreciation to the affected communities and practitioners who shared their experiences and expertise, ensuring that this framework remains grounded in real-world applications and needs.

Any remaining errors or omissions are solely our responsibility.

Preface

The African <AI & Equality> Toolbox was born from a shared recognition: that Africa's relationship with Artificial Intelligence must be defined not by adoption alone, but by ownership, co-creation, and leadership.

For too long, the continent's technological future has been shaped by imported systems and external agendas—systems that often disregard context, community, and the lived realities of those they claim to serve. This Toolbox is our response.

It builds on years of work at the intersection of technology, human rights, and gender equality. It reflects insights from the <AI & Equality> Human Rights Toolbox, co-developed with the United Nations Office of the High Commissioner for Human Rights (OHCHR), and adapts that methodology for the African continent—shaped in partnership with Women at the Table, the African Centre for Technology Studies (ACTS), and a community of researchers, activists, technologists, and policymakers committed to justice.

The result is a living, evolving platform—rooted in six stages of the AI lifecycle and powered by African case studies that span agriculture, health, climate, education, and language inclusion. These stories are not abstract illustrations; they are real-world examples of what is possible when communities are invited in from the start—not just as beneficiaries of technology, but as designers, decision-makers, and experts in their own right.

This work is grounded in a human rights-based approach. Not just because it is ethical, but because it is effective. AI systems that emerge from deep listening, inclusive teams, and sustained engagement with affected communities are not only fairer—they are more resilient, impactful, and relevant.

We offer this Toolbox to those shaping AI on the continent—not as a blueprint, but as an invitation. An invitation to reflect, to collaborate, and to build systems that uplift rather than exclude; that heal rather than harm; and that reflect the full richness of African thought, experience, and possibility. We are grateful to the many individuals and institutions who have contributed to this work so far—and we look forward to walking this path together.

Caitlin Kraft-Buchman
Founder & CEO, Woman At The Table and <AI & Equality>

Winston Ojenge
Principal Research Fellow and Head of the ACTS AI African Center for Technology Studies

Contents

Executive summary 2

Acknowledgments 4

Preface 6

Introduction 8

What is the purpose of a Human Rights-based approach? 10

Why a Human-Rights based approach vs “Ethical” or Responsible AI? 10

The AI lifecycle 12

- Stage 1: Objective + Team Composition 14
- Stage 2: Defining System Requirements 19
- Stage 3: Data Discovery 23
- Stage 4: Selecting and Developing a Model 27
- Stage 5: Testing and Interpreting Outcome 30
- Stage 6: Deployment & Post-Deployment Monitoring 33

Summary 36

Endnotes 37

Annex: Case Studies 40

- Technology-Facilitated Gender-Based Violence in Africa 41
- AI Climate Sensors in Africa 61
- AI for Kenyan Sign Language and Digital Inclusion 74
- AI-Powered Malaria Diagnostics 92
- Co-Creating AI for Agriculture: Nigeria’s Nsukka Yellow Pepper Project 102
- Empowering African Languages through NLP: KenCorpus Project 115
- Design by Inclusion in AI Development 130

Introduction

> AI is rapidly transforming societies across the globe, yet its **development and deployment often remain rooted in paradigms and priorities from the Global North.**

The African <AI & Equality> Toolbox emerges as a timely, transformative initiative that seeks to redress this imbalance by rooting AI in African realities, needs, and visions for the future. Anchored in human rights principles and grounded in participatory, community-led processes, the Toolbox equips African policymakers, technologists, civil society actors, and communities with the tools, vocabulary, and frameworks necessary to shape AI that is equitable, inclusive, and just.

This initiative is about co-creating AI and not simply adopting it. Drawing on the <AI & Equality> Human Rights Toolbox developed in collaboration with the Office of the High Commissioner for Human Rights (OHCHR), and enriched through regional partnerships with organizations like the African Centre for Technology Studies (ACTS) and Women at the Table, the African Toolbox centers the continent's own voices, practices, and priorities. The toolbox is structured around the six stages of the AI lifecycle and integrates a Human Rights Impact Assessment (HRIA) methodology, aligning with evolving international norms such as the EU AI Act.

This Toolbox is made tangible through African case studies across vital sectors, such as health, agriculture, climate, education, and language inclusion. From AI-powered malaria diagnostics in Uganda, to localized NLP systems for Kenyan languages and sign language, to participatory environmental sensing in Kenyan urban and rural areas, and AI-enhanced tools designed with and for women farmers in Nigeria and Uganda—each case illustrates how rights-based, community-embedded approaches foster AI that is not only technically effective, but socially empowering.

> **In a time of profound digital transformation, the African ‹AI & Equality› Toolbox calls for a shift:** from top-down technology transfer to bottom-up co-design; from abstract ethics to concrete rights; from passive consumption to active leadership.

The Toolbox affirms that Africa is not just a recipient of AI—**African professors, educators, scientists, activists and communities are innovators in the design and creation of equitable AI futures** rooted in the local, and deeply relevant to the global conversation.

Our goal is to move beyond mere compliance and towards a paradigm of AI development that proactively promotes the achievement of Human Rights – vs mitigating risks as an add-on or after harms have already occurred. By involving affected communities from the outset and with substantial decision agency, we promote and enable the development of systems that center Human Rights, equality, and inclusion at the core of code, capable of creating new opportunities and innovative correction of inequities.

We hope to bring social programs in line with 21st century research and values, alignment with the Sustainable Development Goals, and united in finding ways to make AI more effective – not merely more 'accurate' and 'efficient'.

Learn more about the AI & Equality African Toolbox initiative in this video.

Watch the video

What is the ***purpose*** of a Human Rights-based approach?

AI is affecting all parts of society and even when well-intentioned has repeatedly harmed or exploited communities, and especially vulnerable groups[1]. We believe that many of these harms can be prevented through **critical reflection points** from the conceptual phase, throughout, and post AI development. These reflection points promote a **paradigm shift** in AI creation away from primarily stand alone technology-driven objectives towards a socio-technical system creation in **collaboration with the communities** that the system will interact with and affect.

This approach is likely to result in systems that are more robust, resulting in more effective uptake, use and evolution of the technology with the potential to **empower** communities and citizens in achieving and enjoying their Human Rights. It will also result in systems and solutions that bear less risk of negatively impacting the Human Rights of communities the technology is designed to serve.

Why a ***Human-Rights based*** **approach** vs "Ethical" or Responsible AI?

Ethics, which are crucially important, are also **situational**[2]. Ethical and Responsible AI principles, authored by a wide range of bodies (e.g. academia, civil society organizations, research institutes,governments, and the private sector) are the most common response to concerns around the ethics of AI[3], however, they are under major critique from academia[4-5] and AI practice[6-7]. Their **abstract nature** allows for diverging interpretations and implementation, impeding or even undermining accountability.

We avoid this ambiguity by focusing on Human Rights, an agreed body of international (and national) law that reflects a **universal understanding** of aspects required to ensure human dignity with a focus on equality and non-discrimination, participation and inclusion, accountability and the rule of law which are indivisible and interdependent principles of human rights[8]. Thus, **Human Rights provide a common and concrete starting point** to align different actors, disciplines, and cultures.

Further, new policies such as the EU AI Act require **Human Rights Impact Assessments (HRIA)** by the deployers or procurers of high-risk technologies such as AI used in human resources, education, financial decisions, or healthcare[9]. Since currently, no official HRIA is available as part of the EU AI Act or elsewhere, various bodies and research institutes are developing their versions of HRIAs. After reviewing several, we decided to **integrate the very thorough HRIA of the Alan Turing Institute[10] in our framework**, i.e. prompt the questions and reflections covered by the HRIA at the lifecycle stages at which they become relevant. Thus, we enable an approach to AI development that considers relevant aspects **throughout the development process** – instead of as an add-on after the system has been developed, i.e. at the point of procurement.

In this manner, deployers or procurers can review all actions taken, vastly facilitating accountability, transparency, as well as the process of conducting HRIAs before deployment. Consequently, orienting our framework along Human Rights has the further benefit that it **facilitates the compliance with upcoming AI regulation.**

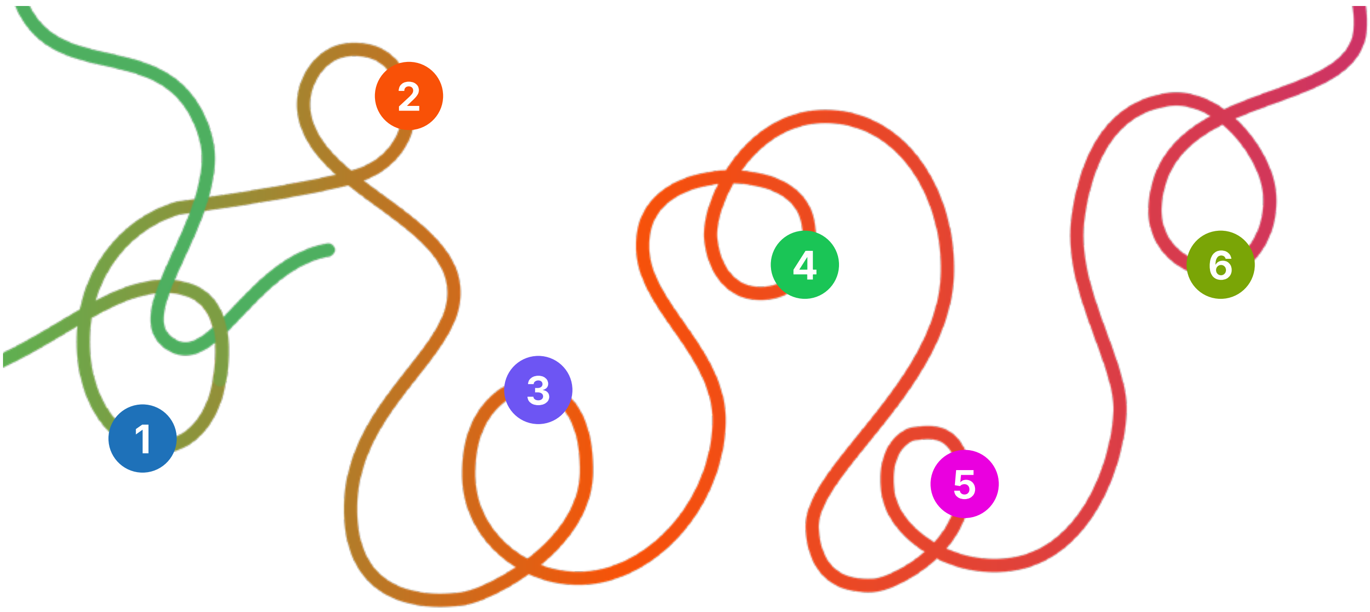
> This approach is likely to result in systems that are more robust, resulting in more effective uptake, use and evolution of the **technology with the potential to empower communities and citizens** in achieving and enjoying their Human Rights.

The ***AI*** **lifecycle**

To ensure that our recommendations are **actionable for AI practitioners**, we anchored our **<AI & Equality> reflective questions** along the AI lifecycle, combining them with the HRIA of the Alan Turing Institute[10]. The lifecycle is not strictly linear but **interwoven and cyclical**, resembling a thread looping back repeatedly. This emphasizes the importance of **reflecting, revisiting, and refining** as we learn more about the socio-technical context, the data, the model, and integration of Human Rights-based considerations **throughout the AI lifecycle** – instead of as an add-on after the system has been developed or even contemplated or slated for use.

We distinguish following **six stages** of the lifecycle:

1. Objective + Team Composition
2. Defining System Requirements
3. Data Discovery
4. Selecting and Developing a Model
5. Testing and Interpreting Outcome
6. Deployment & Post-Deployment Monitoring



Essential questions per AI Lifecycle Stages

In the following sections, we will provide a short overview over the six stages, crucial concepts, and the essential questions that AI creators should reflect on at each specific stage.

We invite you to **additionally complete our free online course**, with special emphasis on module 2 and 3, to get a more comprehensive understanding of why we recommend these reflection points in addition to purely technical measures. Both modules elaborate on the actions and thought patterns that contribute to some currently harmful practice.

How to address the reflective questions?

It is essential that you do not answer the questions only by yourself or with your team. Instead, for many questions **it is essential to discuss the questions and potential answers with representatives from the specifically affected communities and especially with historically marginalized groups.** Further, your answers may change as you learn new things, so **do not hesitate to revisit and amend your answers.**

The Alan Turing Institute's HRIA

The Alan Turing Institute published a working version of their Human rights, democracy, and the rule of law assurance framework for AI systems. **We locate the areas covered in their HRIA template (see p. 251 to 276[10]) along the AI lifecycle** (see next section) to enable AI development that considers these **prior** to deployment, and **also at the stage of the AI lifecycle at which they become relevant**. In this manner, we help to build systems with Human Rights at their core, **not only implying HRIA compliance but making the process of conducting pre-deployment HRIAs easier, more efficient and effective.**

1 Stage 1: Objective + Team Composition

From intention to inclusive innovation

Across African contexts, development and technology projects are often driven by external actors with little grounding in local priorities. Solutions frequently arrive pre-packaged—built around assumed problems, rather than those identified by communities themselves. This is particularly evident in AI deployments in agriculture, education, and public health, where tools may miss the mark, or worse, exacerbate inequities.

In response, a participatory and grounded approach at Stage 1 ensures:

- **Problem relevance:** Solutions address the real needs of end users, especially marginalized communities.
- **Power redistribution:** Communities are not merely consulted, but co-define the problem and share in decisions.
- **Greater sustainability**: Objectives grounded in lived realities are more likely to gain traction and evolve with community feedback.

This stage is especially crucial for centering gender equity, given that women often carry the brunt of labor in agriculture and caregiving, yet remain underrepresented in AI design and governance.

A. Defining Objective

It is essential to start with the objective and purpose of a system: It should always be clear why a specific system is required, which issue it solves, and for whom. Too often, this vision only reflects the needs of the people developing the system in isolation holding great power in this context (which includes not only companies and governments, but the academic AI researchers themselves) – as opposed to the needs of the communities the system is designed to serve and affect.

Therefore, it is fundamental to engage affected communities early on through participatory development practices. To begin, the affected community should be consulted and agree that an AI system is the best way to help solve their problem as there may be simpler, more efficient and cost effective ways to tackle the core problem.

> **Participatory Development** in this context describes the process of creating technology in collaboration with affected communities[11]. This includes an exploration of their needs, values, and concerns in the application context and addressing these in the system's design. Affected communities can be system customers (e.g. hospital, bank, government), system users (e.g. radiologists, employee of a bank, civil servant), the people the system is used on (e.g. patient, someone applying for a loan, citizen), as well as the most vulnerable communities.

Here, it is essential that all affected communities (vs only revenue-critical groups) are involved and have actual decision power and agency in the process. This prevents an extractive form of participatory development where community needs are collected but their implementation is disregarded by commercial interests or internal agendas.

B. Team Composition

Numerous people are involved in the creation and operation of an AI system - more than just people writing code! The objective of a system should fundamentally inform the composition of its team of creators, in other words, what types of expertise and lived experience are required to fully make the intended objective a reality. This would include not only the required knowledge and technical skills, but the diverse backgrounds, perspectives, and experiences with the environment for which your system is developed. We want to highlight two roles that are often forgotten: affected communities & social scientists.

> **Affected communities** are the experts in the context where the system will be deployed (i.e. in their lived experience) and will carry the consequences of the system's deployment. Special attention should be given to already marginalized communities since AI systems may have particularly adverse effects on these communities' ability to participate fully and meaningfully in the new systems that are created[12]. Input from affected communities helps to create better suited systems[13], ensures more uptake, and helps in foreseeing risks and harms.

> **Social Scientists:** Your team should include members that are experts in the social or human rights-aspects of your application context. This is required to understand the social contexts as well as power imbalances and inequalities that might disadvantage historically marginalized communities, especially women and girls. Having an expert in social dynamics in your team will help the entire team, flag potential issues, and emphasize a core commitment to a collaborative team effort as the entire group to promote and protect human rights.

1 Objective + Team Composition

Essential questions

Purpose & Context of the System	• What **problem** is the system trying to solve?
	  ◦ Does the domain have a history of discrimination?
	  ◦ Is there a risk that your system might enhance or enforce historically unequal outcomes?
	  ◦ How can you counteract such historical discrimination?
	
	• Will the system have an **essential or high-risk function** or be implemented in a high impact or safety critical sector (see e.g. EU AI Act)?
	  ◦ How do you ensure safe operation, both in design as well as in case of system outage?
	
	• Have **communities** affected by the system been engaged in dialogue about the system?
	  ◦ Is an AI system even the best way to address the issue?
	  ◦ Does it address the community's most pressing needs?
	  ◦ Are some of the communities vulnerable, e.g. due to protected characteristics?
	
	• Is the system supposed to be **implemented at scale**? Is this wise?
	• Is using the system or the system being used on someone **voluntary** (direct and indirect use)?

Effects of the System	• Who **benefits** from the system and who can be **disadvantaged**?
	  ◦ Does this reflect or level **current power structures**?
	  ◦ How can we involve communities and especially historically marginalized groups?
	
	• Does the system actively **contribute to Human Rights**?
	  ◦ Have you conducted a first screening of Human Rights Impacts to identify risks before resources have been invested (p.21 to 47[14])? Potential risks include manipulation, discrimination, or guarding current power structures.
	  ◦ What if the system is used in unintended ways?
	  ◦ Does the system help to promote Human Rights principles and priorities?
	  ◦ Who should be included in / consulted during this assessment?
	  ◦ How do you ensure that identified risks are eliminated or mitigated?
	
	• Who is **accountable** for inaccuracies and resulting harm?
	  ◦ How do you document system design decisions, accountabilities, and
	  ◦ general responsibilities so they can be traced back?
	  ◦ Have you considered all above questions (especially Human Rights impacts) for your **system's entire value chain**, e.g. for suppliers, subcontractors, auditors, etc?
	  ◦ How do you ensure the ongoing and thorough scrutiny of the value chain?

Empowering Affected Communities	• How can the impacted communities be **represented** in the team so that the team can benefit from their insights and real world experience?
	
	• Besides via team membership, how does the team involve affected communities?
	  ◦ Do these communities receive the necessary **agency** to impact decisions?
	  ◦ Does the development team have the **mindset and skills** to achieve this?

Team Composition	• What **expertise do you need** in your team?
	• Do you have **diversity in culture, demographics, lived experience, disciplines and skills** (socio-technical, legal, anthropological, UX, technical,...)?
	
	• How do you ensure **flat hierarchies & communication** between disciplines?
	
	• Does the team have:
	  ◦ Awareness of the **risks that AI systems pose to Human Rights** and underlying reasons?
	  ◦ Insights into / experiences with the problem they are trying to solve?
	  ◦ Insights into / experiences with potential solutions for this problem?

Related case studies

— Makerere Health Lab (Uganda)

The Makerere Health Lab's AI-powered malaria diagnostics initiative began with a question posed not by funders, but by local medical teams: How can we reduce diagnostic delays in rural clinics where skilled technicians are scarce? The objective emerged from real-world constraints in Ugandan health centers. The project team included AI researchers, public health experts, and local practitioners. Community needs shaped the objective: affordability, offline capability, and rapid testing in remote areas. Their early-stage Human Rights considerations led to proactive ethical review, data anonymization, and community ownership of results.

See full case study here.

Nsukka Yellow Pepper Project (Nigeria)

This agricultural AI project began not with a technical solution, but with community listening. Separate, safe sessions with women farmers revealed concerns that had gone unacknowledged in male-dominated dialogues—such as access to water, market discrimination, and lack of information in local languages. These insights became the core objective of the AI tool: a mobile app providing cultivation advice tailored to women's lived challenges. Women also played decision-making roles during prototyping, including testing voice-input options and shaping the training approach for broader rollout.

See full case study here.

Addressing TFGBV in Platform Design

When Ethiopian mayor Adanech Abiebie was targeted with AI-generated deepfake videos that garnered over 500,000 views, it revealed how platforms designed for "engagement" can become weapons against women in leadership. A human rights-centered approach would have started differently: consulting women political leaders about harassment risks, including digital violence experts on the team, and defining success not just by user growth but by safety metrics. The devastating impact—destroying reputations and silencing political voices—shows why TFGBV prevention must be embedded from the very first stage of AI development.

See full case study here.

Key takeways

- Centering the **purpose of an AI system in community-identified needs** is not just ethical—it's essential for success.
- Teams must be **interdisciplinary, context-aware, and gender-responsive** from the start.
- The decision to build AI **must emerge with, not for, affected communities.**
- A **preliminary Human Rights screening** at this stage can prevent avoidable harms and ensure a more inclusive foundation.

2 Stage 2: Defining System Requirements

Designing with intention, grounded in rights

In many African AI deployments, system requirements are defined by international technical partners or funders, often without fully understanding the day-to-day realities of use. This leads to design choices—like requiring high-speed internet, English-only interfaces, or complex interfaces—that make tools ineffective or even harmful.

At this stage, requirement setting should function as a bridge **between vision and use:**

- Aligning system features with **cultural context, infrastructure gaps, and social expectations.**
- Identifying constraints early on—**connectivity, literacy, consent, power dynamics**—and building around them.
- Making **conscious trade-offs** between speed, scale, and equity.

It's also a point where **gendered impacts emerge** more clearly: who has access to devices, who controls decision-making, who benefits from outputs. These must be considered explicitly, not assumed.

At the second stage, the system's objective is formalized into a list of requirements, again, **developed in dialogue between various roles and communities.** This includes managing trade-offs between different needs and desired requirements as systems exist in an **ecosystem of values.**

> **Ecosystem of values**
> Different aspects of a system make it responsible. Examples are that its decisions are fair (**fairness**), that its decisions are easy to understand (**explainability**), that its development process and underlying motivations are clear (**transparency**), or it operates with little error (**accuracy**). You can find a list of these aspects with more detailed definitions and examples in Module 2 of our online course. it is impossible to optimize all of these aspects simultaneously in equal measure, therefore trade-offs are required[15] (although these trade offs do not necessarily reduce accuracy in any fundamental way[16]). For example, highly explainable models often have less accuracy than more opaque forms of AI models[17].
>
> In some contexts, explainability might be as important (or even more important) than the minimization of errors (accuracy): only if the human overseeing the system can understand and question the output, she can detect and correct the errors - thus ultimately leading to less errors than high accuracy alone. Thus, it is essential to not focus solely on one metric (such as often done with accuracy), but instead to make a conscious decision about metric hierarchy and importance in the specific context.

> Importantly, accuracy should never be considered without fairness as it can hide unequally distributed accuracy, e.g. that the system is highly accurate for the majority of cases while being very inaccurate for a minority group[18]. This can lead to negative Human **Rights impacts, in healthcare, facial recognition, finance, subsidy, and other important sectors.**

The process of defining the system's requirements should be iterative and fluid; it is very likely that **the list of requirements may change as more details about the social context and the needs of impacted communities become apparent.** Thus, it is important to **provide a platform** where operators and affected communities can notify the team of new pieces of information that might influence the requirements.

2 System requirements

Essential questions

Involving affected communities	• **Who should be involved** in the definition of the system requirements? Think beyond operators, users or revenue-critical parties! • Are there **tensions** between the system's goals and the needs of affected communities? How can these be addressed, always prioritising Human Rights? • Have you **revisited** your initial Human Rights Impact Assessment, now where more capabilities are planned? • Have you arranged **expert input**, e.g. from affected communities with lived experience, a government department (or allied government department), academia, or public body?

Explainability considerations	• What is the **goal** of explanations?
	  ◦ Who is the **audience** and why?
	  ◦ Will explanations be available for all affected communities, aiding public scrutiny?
	  ◦ Are provided explanations **easy to process** for all intended audiences?
	• Have you considered **which aspects of explainability** are the most relevant?
	  ◦ E.g. how decisions are made in general, how an individual decision was made, etc.
	• How can you use explanations to **increase the agency of affected communities**, e.g. via detailing what would have to change for a different outcome (counterfactual explanation)?
	  ◦ How do you ensure that your explanations help affected communitie to understand the limits and impacts of the system?

Ecosystem of values	• Are there **tensions between accuracy and other, more necessary metrics** in this context?
	• **Fairness**: Which fairness metrics do you expect to be useful in this context? Explore several!
	
	• **Privacy**: Is the privacy of all affected communities and data subjects respected?
	  ◦ How can you minimize the data collection in private spheres, e.g.homes?
	  ◦ Is the remaining intrusion worth it?
	
	• **Transparency**: How will you enable impacted communities to access information about your methodology, e.g. training data, analytical process, how the model was trained, metadata of various metrics?
	  ◦ How can you ensure that affected communities are aware that they are using an AI system /or it is used on them?
	
	• **Accountability**: What is the accountability structure?
	  ◦ Which human oversight should be aimed for?
	  ◦ What expertise and training will the human in the loop require?
	  ◦ How can you enable affected communities to contest an outcome?
	
	• **Usability**: How can we ensure that the interface is intuitive and accessible for all?

Related case studies

NLP for Underserved Kenyan Languages (Ken Corpus Project)

As the Ken Corpus project moved from vision to system design, elders and educators helped define the system requirements: offline access, community-curated content, and flexible data input methods. Rather than prioritizing technical complexity, the team emphasized cultural relevance, consent protocols, and data sovereignty. Annotated texts and oral stories were recorded with community participation, shaping a system that respects linguistic diversity and centers community authorship.

See full case study here.

sensors.AFRICA (Urban Air Pollution – Nakuru, Kenya)

In Nakuru, Kenya, community consultations during system design revealed that technical accuracy alone would not build trust. Requirements had to include explainable results for everyday citizens, alerts in local languages via SMS or Apps, and data formats usable by everyday citizens, journalists and local governments. Community members insisted on privacy guarantees for all air quality sensor hosts. These social and human rights dimensions reshape the technical system: from anonymized data protocols to participatory mapping of sensor placement zones.

See full case study here.

Platform Safety Requirements Against TFGBV

When Nigerian Senator Natasha Akpoti-Uduaghan filed a sexual harassment complaint and was immediately targeted with manipulated videos that reached 400,000 views, it exposed critical missing requirements in platform design. A rights-based approach would have required: rapid detection of coordinated attacks (34 identical Facebook posts should trigger alerts), immediate support for high-profile harassment victims, and clear explanations for content decisions. The case demonstrates why safety requirements must be as detailed and enforceable as technical specifications.

See full case study here.

Key takeways

- System requirements are not just technical specs—they are ethical and political commitments.
- Design trade-offs must be transparent and made in consultation with those most affected.
- Explainability, usability, and consent are **requirements**, not add-ons.
- Inclusive requirement-setting strengthens legitimacy, trust, and long-term adoption.

3 Stage 3: Data Discovery

From representation to responsibility

In many African AI projects, available datasets are either imported (trained on non-African populations) or incomplete (lacking local language, gender, or cultural nuance). This misalignment risks perpetuating systemic bias under the guise of neutrality.

In reality:

- Many African communities are **underrepresented in digital datasets.**
- Some datasets reflect **colonial-era knowledge systems,** with little input from local voices.
- Others involve **covert or extractive data practices** that violate trust and privacy.

Addressing this requires **intentional strategies to build or adapt datasets that reflect African realities,** with consent, care, and community engagement as non-negotiables.

A valuable system objective and its requirements can be undermined **if the dataset used to train the AI system is not representative of your use case and context.** A **good socio-cultural fit** of the dataset includes various aspects such as the demographics of the individuals in the dataset, their culture, or environmental factors[19]. Consulting **domain experts** will be imperative to ensure relevant aspects are appropriately captured.

If no dataset with a good fit is available, the team may have to **generate a new dataset,** either by collecting new data, and/or by improving or augmenting existing datasets through **pre-processing** (i.e. mathematical) steps.

> **Pre-Processing** refers to the manipulation and transformation of raw data before feeding it into a model. It involves various techniques to enhance the quality, relevance, and fairness of the data, e.g. by balancing the frequency of a specific class (e.g. gender or race) in the dataset so that the model is equally trained on them.

3 Data discovery
Essential questions

Data origin

- **Who collected** the data and for **which purpose**?
- Did the data subjects **consent** to use of their data?
  - Was their **privacy** respected?
- **How sensitive** is the information, e.g. does the data reveal sensitive attributes such as racial or ethnic origins, sexual orientations, health status, or religious beliefs?
  - Is there a way to **anonymize** the personal data so that privacy is respected AND insights on age, gender, geography can be captured ?

Data bias

- Who is **included** in the data? Who is **excluded**? Why might that be?
  - Which **geographic regions and cultures** are included and which not?
  - Which **consequences** does this have for your system's operation?
- Which **historical / present bias** might be in the data, risking to compromise Human Rights?
- Which **data pre-processing steps** are required to create a model that is fair in this context?
- In your specific use case, is it most beneficial to ignore (show potential unfairness in data), 'erase' (remove potential unfairness in data), or even counteract (counteract this bias in a way that the disadvantaged group is now advantaged) in this bias?

Documentation

- Have you **documented** which datasets you are using and **why you choose them** so that potential deployers can assess whether your training data fits their context?
- Have you documented **all pre-processing step**s you took (essential information for future uses of your system or code)?
- Have you saved your **"raw" data** – in addition to the preprocessed data – to support future uses?

Related case studies

- **Makerere Health Lab (Uganda)**
  Faced with a lack of relevant datasets for malaria diagnosis, the Makerere team took the difficult but ethical route: building their own dataset from scratch. This included securing ethical approvals, anonymizing patient information, and partnering with local health facilities. Importantly, data collection was not seen as a technical task alone—it was a social contract. The team documented pre-processing methods, managed class imbalance issues transparently, and shared ownership with local stakeholders.

  See full case study here.

- **NLP for Underserved Kenyan Languages (Ken Corpus Project)**
  In the Ken Corpus NLP project, building datasets for underserved Kenyan languages meant going beyond scraping websites. Elders and native speakers were involved in storytelling, glossary-building, and quality-checking annotations. Dialect diversity, idiomatic expressions, and consented oral histories were woven into the dataset. This helped ensure that the AI model would not erase nuance—or replicate linguistic colonization.

  See full case study here.

- **Agriculture Image Recognition (Uganda)**
  In a crop disease detection project, the team had trained a deep learning model for early detection of disease and monitoring. However, feedback from women farmers at model deployment revealed that their crop concerns differed significantly—prioritizing soil nutrient levels and soil-borne diseases. This surfaced a key insight: even a technically sound model may fall short if it does not incorporate users' priorities/perspective through a participatory and gender-responsive design process.

  See full case study here.

Addressing Bias in Content Moderation Training Data
Code for Africa's research revealed how content moderation systems fail to detect African-specific hate speech and harassment tactics. When Cameroon's Brenda Biya faced coordinated attacks using coded language and cultural references, standard moderation models—trained primarily on Western datasets—missed the harmful content entirely. Building effective TFGBV prevention requires datasets that include African languages, cultural contexts, and the sophisticated evasion tactics used by harassers, while ensuring this data is used to prevent rather than perpetuate harm.

See full case study here.

Key takeways

- Data collection is never neutral—it reflects power, values, and access.
- Locally grounded data often needs to be created, not just scraped or purchased.
- Representation without consent is surveillance; participation with agency is co-creation.
- Transparent documentation and pre-processing are essential for fairness and future accountability.

4 Stage 4: Selecting and Developing a Model

Building systems that reflect values—not just accuracy

In many African contexts, imported or generalized models often underperform—especially when they are trained on data that does not reflect local language, environment, or lived experience. For AI to be trustworthy, accuracy alone is not enough. Systems must also be explainable, locally interpretable, and open to scrutiny—so that communities understand how decisions are made, and can challenge or adapt them as needed.

Inclusion means building systems that don't require technical expertise to interpret—ensuring that trust, oversight, and agency are accessible to all users. Whether a rural health worker, a student, or a community organizer, each person should be able to understand what a system is doing and why. This is not about simplifying complex systems for non-experts, but about **reclaiming AI as a public good,** where understanding and control are shared—not centralized. It is a step toward democratizing AI, where transparency is not a luxury but a right.

It is time to consider **what type of AI model is the best to satisfy the system requirements.** Note: it is not always the most complicated deep-learning algorithm!

Instead, it is about choosing the most suitable model for the required scope while **managing trade-offs.** For example, less complex models are often more explainable but might achieve a slightly lower accuracy. Since explainability is a prerequisite for good error and bias detection, such models seem especially important in **high-stakes scenarios.** For example, the European Central Bank requires a high level of explainability for credit scoring decisions[20], and therefore excludes neural networks and other types of less explainable algorithms that **impede the discovery of discriminatory outcomes** and scrutiny.

Model development itself is an iterative process in which different aspects of the model are adjusted to **meet different system requirements** (e.g. via in- or post-processing methods or by adjusting the weights or parameters of a model). It is important here to **reflect about earlier stages** to ensure that your objective, requirements, data, and model are all aligned.

> **In-Processing** methods are designed to mitigate bias and/or increase fairness while the model is being trained, while **Post-Processing** methods include modifying the model's output after training has been completed.

4 Selecting and developing a model
Essential questions

Model Type and Explainability Requirements	• Does your model... ◦ Achieve appropriate **explainability**, considering the stakes of the ◦ situation? ◦ Minimise **complexity?** ◦ **Alert** the user if it is uncertain with a decision and / or when it is confronted with an instance that is not reflected sufficiently in its training data (e.g. model only trained on light skin with little pigment is presented with an instance of dark skin with more pigment, thus alerting the user that it does not know how to classify this instance)?
Fairness aspects *(see module 3 of our free online course)*	• What is the most suitable **fairness metric** and why? • Have you experimented with a variety of different metrics and outcomes? • Which **aspects of fairness are in focus**, e.g. based on gender, ethnicity, education...? ◦ Have you considered relevant **intersectionalities**? • Have you ensured that the model does not rely on **variables or proxies that might be unfairly discriminatory**? For example, a person's postcode might allow you to infer ethnicity. • Why have certain **in- (model) and post (evaluation)-processing** steps been chosen?
Other	• Is the model **transparent** to affected communities, i.e. who funded it, its objective, who was involved, training data, performance, ... • What is the **environmental impact** of the model? Is it worth the cost? ◦ Have there been efforts to **minimize or offset** the environmental impact?

Related case studies

— **Makerere Health Lab (Uganda)**
The Makerere team selected a lightweight image analysis model that could run on smartphones with minimal bandwidth—sacrificing some complexity for usability in rural areas. They iteratively trained the model using local data, monitored class imbalances (malaria vs. non-malaria), and focused on optimizing inference time (0.23 seconds). Importantly, they acknowledged that diagnostic accuracy varied based on the feature set and committed to ongoing bias mitigation—even after deployment.

See full case study here.

- **Kenyan Sign Language Avatar Project**
  In developing a translation system for Kenyan Sign Language, the team used advanced pose estimation models to animate a virtual avatar. However, they prioritized feedback loops with the Deaf community—adjusting the models for naturalness and accuracy per the KSL. This responsive modeling process allowed technical design to adapt to community-defined quality and usability standards.

  See full case study here.

- **Agriculture – Nsukka Pepper App (Nigeria)**
  The model used in the Nsukka Pepper project wasn't about maximizing precision farming—it was about delivering actionable, understandable advice to women farmers. Developers built a hybrid model combining local agronomic rules with real-time data and NLP elements. Voice input and offline functionality were integrated from the start—not as features, but as core design requirements linked to social context and digital access.

  See full case study here.

- **Content Moderation Against Coordinated Attacks**
  When 34 Facebook posts with identical content attacking Brenda Biya reached 8.9 million views, it revealed how standard spam detection models fail against coordinated TFGBV. Effective models must detect not just individual harmful content but patterns of coordination—multiple accounts posting identical content, rapid amplification networks, and cross-platform campaigns. This requires models that understand both content and behavior, prioritizing victim safety over engagement metrics.

  See full case study here.

Key takeaways

- Select models based on **context-fit**, not complexity or prestige.
- **Favor explainability, usability, and adaptability** over marginal performance gains.
- Ensure fairness is defined **locally** and tested **intersectionally**.
- **Document decisions** transparently to support accountability, oversight, and adaptation.

5 Stage 5: Test and Interpret Outcome

Validation Beyond the Lab: Testing for Trust, Impact, and Equity

In African deployments, there is often pressure to launch rapidly, without thorough contextual testing. But skipping this step is where trust breaks down—and harm begins. Testing must happen with communities, not just on them. It should include:

- Testing across **different regions, literacy levels, languages, and infrastructures.**
- Direct participation by **women, youth, elders, and differently abled people.**
- Methods that **value lived experience** as much as statistical accuracy.

This stage is also an opportunity to **reflect on how power operates in AI**: Who gets to say if it works? Who can question it? Who can stop it?

After the model has been developed, we have to test whether **it fulfills the system requirements** defined by the team in stage 2. For some metrics, this can be done **via technical tests**, others require the **feedback of affected communities**[21], e.g. whether the intended level of explainability was achieved.

For the technical tests, it is important that the **testing dataset is as representative of the context** as the training dataset. **Including extreme examples/cases** can help to uncover potential issues that may not be apparent during routine testing, thereby revealing any limitations or weaknesses in the model's performance[22].

Insights gained should inform a **'manual' handed to the future system users/operators.** Through stating the contexts for which the system has been trained (expected to operate well) and which are not (inaccuracies likely), the operators can **calibrate their trust and adherence accordingly.** Further, the manual should include recommendations on the **required level of human oversight,** thus allowing appropriate training of the operators.

5

Test and interpret outcome
Essential questions

Testing Context and Outcomes	• Does the system **meet the objective and the system requirements**?
	  ◦ What **measures** of model performance are included and why were they selected over others (including quantitative AND qualitative aspects)?
	  ◦ Does this selection **still apply** after we learned more about the application context? Should we **add** something?
	  ◦ Whose **opinion** was included in these tests?
	
	• Can the trained model be released to the public or external experts to allow them to **test and scrutinize** it to highlight issues?
	
	• Has the model been **tested as close to its actual application context as possible** (including its actual users) to identify potential harms?
	  ◦ Have resulting learnings and feedback points been **included**?

Operation Manual	• Is an **easily understandable manual** available to the operators? • What can we recommend as best practices around operation, e.g. how much **human oversight** is required and with which expertise ? • For **which contexts** has the system been trained? ◦ Where might it become unfair or inaccurate? • How will you **train** operators on how to use and interpret the system, including how to calibrate their trust in and ability to question the system's operation? • How will you **log future changes** to the system?

Related case studies

— **Makerere Health Lab (Uganda)**
The malaria diagnostics tool was tested not in a lab, but in rural clinics—the exact environments where it would be used. Healthcare workers were trained to use the smartphone-microscope tool and gave detailed feedback on usability, clarity of results, and diagnostic trust. This led to adjustments in the interface, refinements in image interpretation, and greater transparency in how the AI was making decisions. Feedback was not tokenized—it reshaped the tool.

See full case study here.

- **sensors.AFRICA (Nakuru, Kenya)**
20 air quality sensors were deployed, in partnership with the RESPIRA project, to pilot an AI-driven early warning system for air pollution by involving citizens in urban neighborhoods directly impacted by incidences of poor air quality. Future testing will not only focus on accuracy, but on whether residents understand the alerts, find them timely, and can act on them. Community trust is built through participatory approaches with climate change ward committees, who are in touch with local communities and familiar with pollution levels of each of their jurisdictions. Feedback from these community engagement & outreach sessions will shape how dashboards are designed and how data is communicated in local languages.

  See full case study here.

- **Kenyan Sign Language Avatar Project**
Testing the virtual KSL avatar involved continuous engagement with Deaf students and sign language experts. Community testers evaluated how accurately the avatar translated concepts, reflected regional variation, and respected cultural nuances. As a result, developers adapted finger-spelling rules, improved avatar expressiveness, and adjusted the signing speed. Crucially, testing was framed not as a trial, but as a co-creation process.

  See full case study here.

- **Testing Rapid Response to TFGBV**
Code for Africa's research documented how current systems fail to respond quickly enough to prevent harm. Testing must simulate real attacks: Can the system detect when 34 identical posts appear within minutes? How quickly can it identify AI-generated deepfakes? Can it prioritize high-profile targets like political leaders who face immediate real-world consequences? Testing with women's rights organizations revealed that speed of response—measured in minutes, not days—determines whether careers and lives can be protected.

  See full case study here.

Key takeways

- Testing must be rooted in **real-world use**—not just performance labs.
- **Diverse users** should help define what "working well" actually means.
- **Interpretation, usability, and local trust** matter as much as accuracy.
- Feedback must be used to **refine the system**—not just to check a box.
- Community co-testing turns validation into **empowerment**.

6 Stage 6: Deployment & Post-Deployment, Auditing and Monitoring

Accountability beyond launch: building systems that learn and respect

In African contexts, post-deployment oversight is often underfunded or overlooked. Once a system is launched—especially by international actors—**it can become invisible**, even as its consequences grow. Yet, the risks of unmonitored AI are high:

- **Shifts** in local politics or policy can make once-benign systems oppressive.
- Tools designed for one context may be **repurposed for surveillance or control.**
- Without local control, updates may reinforce **dependency**, not resilience.

True accountability means planning for **ongoing monitoring, shared governance, and the possibility of "no."** It also means systems must be responsive—not just to data—but to dignity.

> **Deployment:** The deployment step is the **last sanity check**, i.e. whether all harms, discriminatory impacts and consequences have been considered, communicated, and are accounted for. **Revisit your initial Human Rights Impact Assessment** and conduct it more thoroughly now that you know the full system to ensure that the system has been assessed for negative Human Rights impacts in its final form.

The decision as to whether the system is ready to be deployed is powerful. We recommend truly empowering affected communities - after all, they have to **bear the consequences** of a faulty operation! Additionally, it is crucial to set up pathways that enable operators and strongly affected communities to **alert issues** they experience around the system.

> **Post-Deployment:** The system should be **audited and tested regularly in post-deployment audits**, including opportunities for affected communities to provide feedback. This is **especially relevant shortly after deployment** as the newly deployed system might expose previously unknown challenges or problems.

Even if the system operates as expected, the model's application context is **likely to change over time.** This can not only alter the input data or which outputs are considered fair, but even impact the objective, e.g. make the objective obsolete so that the system should be retired. Therefore, it's **essential to continuously audit the system**, including both **quantitative audits as well as qualitative audits** in collaboration with affected communities (see e.g.[23] for a framework to operationalise such audits). A thorough overview over different types of audits - also including audits by external parties - can be viewed here[24].

6 Deployment & Post-Deployment, Auditing and Monitoring

Essential questions

Deployment

- **Who** decides that the model is **ready to be deployed**?
  - Have regulators, domain experts, affected communities **agreed** to deployment?
  - Do the most affected communities have the **agency to delay or stop deployment**?
  - Have you **revisited** your initial Human Rights-Impact Assessment and conducted a more thorough one, now where the full model capabilities are known? (following[25])

- Before deployment: Are there processes in place to **detect potential system failures or unexpected harms**?
  - Are the deciders **accountable** for harm that might be caused?
  - What mechanisms are in place for after an issue has been identified?
  - Who is **responsible** for addressing upcoming harms? What is the **timeline**?

Monitoring

- Are there processes or features in place that **allow operators and impacted communities to alert** suspected system inaccuracies or failures?
  - How can you ensure that **affected communities can opt out of system use**?

- How are you monitoring **context changes**?
  - What is your process to learn about **new risks or harms**?
  - What is your mechanism to learn about **new user needs** in the field?
  - How can we include them in the **requirements** and account for them?
  - In which cases is it better to take the system **offline** until risks have been accounted for?
  - How will you test that the model **continues to fulfill its objective**?
    - How would you know that it is time to retire the system?

Related case studies

— **Makerere Health Lab (Uganda)**

Far from a one-off deployment, the Makerere malaria diagnostics team built a roadmap for future disease detection (e.g., cervical cancer, tuberculosis), real-time feedback from users, and plans to adapt the interface to local languages. Monitoring was not seen as surveillance—but as support. Ongoing collaboration with health workers ensures that model updates reflect emerging needs, shifting health realities, and performance issues that surface in the field.

See full case study here.

— **Agriculture – Nsukka Pepper App (Nigeria)**
After deployment, the Nsukka Pepper app team implemented train-the-trainer models and structured feedback loops with women farmers. This allowed regular updates to planting guides, voice features, and market price integrations. The gender work plan included periodic check-ins to assess impact on workload, income, and empowerment—moving beyond usage stats to understand human outcomes. The ability for farmers to request changes and report issues helped maintain trust and relevance.

See full case study here.

— **sensors.AFRICA (Kenya)**
The system continues to be audited by community watchdogs and independent researchers to assess bias, coverage gaps, and unintended impacts. Alerts shall be adapted as pollution patterns evolve, and local institutions & news outlets engaged to translate data into action. This living system approach treats deployment as a civic dialogue, not a final product.

See full case study here.

— **Evolving TFGBV Defenses Post-Deployment**
The rapid evolution of TFGBV tactics requires continuous adaptation. When harassers began using “spamouflage” techniques—replacing letters with symbols to evade detection—platforms had to quickly update their systems. Code for Africa’s research shows that static defenses fail within weeks as attackers adapt. Successful post-deployment monitoring involves partnerships with women’s rights organizations who can identify emerging threats, rapid response teams that can implement countermeasures, and transparent communication with affected communities about new protections and limitations.

See full case study here.

Key takeaways

- Launch is the beginning, not the end, of AI responsibility.
- Communities must be able to pause, contest, and adapt systems.
- Post-deployment feedback must be structurally integrated—not ad hoc.
- Monitoring should include qualitative, rights-based outcomes—not just technical metrics.
- Retirement, rollback, or redesign must always be on the table.

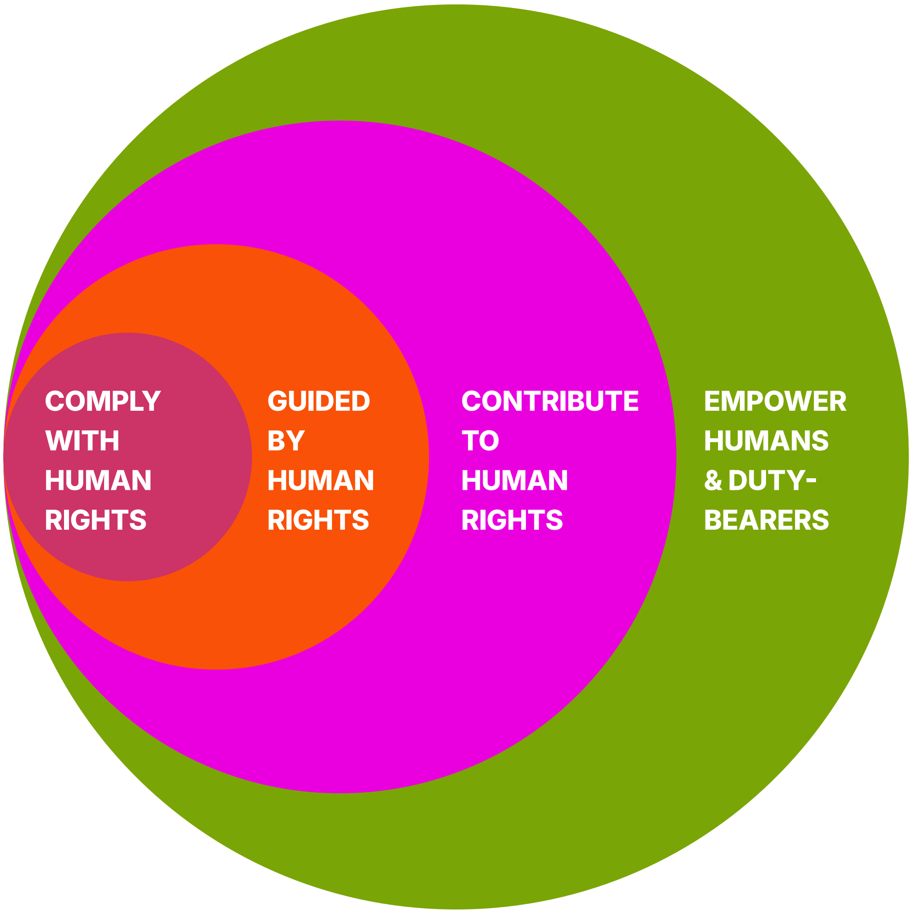
Summary

We highlighted essential questions along the six stages of the AI lifecycle to **enable AI creators to reflect about the objectives, Human Rights impacts, and wider societal effects** of the systems they create **in collaboration with the communities affected by their system.**

We want to emphasise that these questions - at the bare minimum – **facilitate the creation of technology that complies with the Human Rights principles of Equality and Non-Discrimination, Participation & Inclusion, Accountability & the Rule-of-Law.** However, these questions may help to go beyond mere compliance and allow the creation of technologies that are:

- **guided** by Human Rights principles,
- **contribute** to their access and fulfillment, and
- **aspire to empower** humans & duty-bearers to achieve and enjoy their Human Rights.

Going forward, this may allow us to not only 'leave no one behind', but **to bring everyone with us**, enhancing human dignity as we create new technologies.



Endnotes

1 AI Incident Database, accessed on 07.12.24 at https://incidentdatabase.ai/

2 Sadek, M., Kallina, E., Bohné, T. et al. Challenges of responsible AI in practice: scoping review and recommended actions. AI & Society (2024). https://doi.org/10.1007/s00146-024-01880-9

3 Jobin, A., Ienca, M. and Vayena, E. (2019) The Global Landscape of AI Ethics Guidelines. Nature Machine Intelligence, 1, 389-399. https://doi.org/10.1038/s42256-019-0088-2

4 McNamara, A., SmithJ., and Murphy-Hill, E (2018). Does ACM's Code of Ethics Change Ethical Decision Making in Software Development?, https://doi.org/10.1145/3236024.3264833

5 Munn, L. The uselessness of AI ethics. *AI Ethics* 3, 869–877 (2023). https://doi.org/10.1007/s43681-022-00209-w

6 Ibáñez, J., Olmeda, Mónica (2022). Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study. *AI and Society* 37 (4):1663-168

7 Rakova, Bogdana, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. "Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices." *Proceedings of the ACM on Human-Computer Interaction 5* (April 13, 2021): 1–23. https://doi.org/10.1145/3449081

8 United Nations. 1948. Universal Declaration of Human Rights.

9 The EU Artificial Intelligence Act, Article 27: Fundamental Rights Impact Assessment for High-Risk AI Systems (2024), accessed on 19.12.2024 at https://artificialintelligenceact.eu/article/27/

10 The Alan Turing Institute, Human rights, democracy, and the rule of law assurance framework for AI systems: A proposal (2022). P.251-276, https://doi.org/10.5281/zenodo.5981675

11 Delgado, F., Yang, S., Madaio, M., & Yang, Q. (2023, October). The participatory turn in ai design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*.

12 Buolamwini & Gebru (2018) http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; Angwin et al. (2016) 'Machine Bias'. ProPublica, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

13 Lee, Min Kyung, et al. "WeBuildAI: Participatory framework for algorithmic governance." *Proceedings of the ACM on human-computer interaction* 3.CSCW (2019): 1-35.; Hubert D. Zając, Dana Li, Xiang Dai, Jonathan F. Carlsen, Finn Kensing, and Tariq O. Andersen. 2023. Clinician-Facing AI in the Wild: Taking Stock of the Sociotechnical Challenges and Opportunities for HCI. ACM Trans. Comput.-Hum. Interact. 30, 2, Article 33 (April 2023), 39 pages. https://doi.org/10.1145/3582430

14 The Alan Turing Institute. *Human rights, democracy, and the rule of law assurance framework for AI systems: A proposal* (2022), https://doi.org/10.5281/zenodo.5981675

15 Whittlestone, Jess, et al. "Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research." London: Nuffield Foundation (2019). https://www.nuffieldfoundation.org/wp-content/uploads/2019/02/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf

16 Rodolfa, K.T., Lamba, H. & Ghani, R. Empirical observation of negligible fairness–accuracy trade-offs in machine learning for public policy. *Nat Mach Intell* 3, 896–904 (2021). https://doi.org/10.1038/s42256-021-00396-x

17 Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy*, 23(1), 18. https://www.mdpi.com/1099-4300/23/1/18

18 Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." Conference on fairness, accountability and transparency. PMLR, 2018.

19 Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for datasets. ACM 64, 12 (December 2021), 86–92. https://doi.org/10.1145/3458723

20 Dessain, J., Bentaleb, N., and Vinas, F (2023). *Cost of Explainability in AI: An Example with Credit Scoring Models.* https://doi.org/10.1007/978-3-031-44064-9_26

21 Mireia Yurrita, Dave Murray-Rust, Agathe Balayn, and Alessandro Bozzon. 2022. Towards a multi-stakeholder value-based assessment framework for algorithmic systems. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 535–563. https://doi.org/10.1145/3531146.3533118

22 Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B.,Spitzer, E., Deborah Raji, I. and Gebru, T. "Model Cards for Model Reporting." In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT '19)**, January 2019. ACM. http://dx.doi.org/10.1145/3287560.3287596

23 Yurrita, M., Murray-Rust, D., Balayn, A., & Bozzon, A. (2022, June). Towards a multi-stakeholder value-based assessment framework for algorithmic systems. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 535-563). https://dl.acm.org/doi/abs/10.1145/3531146.3533118

24 Birhane, A., Steed, R., Ojewale, V., Vecchione, B., & Raji, I. D. (2024, April). AI auditing: The broken bus on the road to AI accountability. In *2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)* (pp. 612-643). IEEE. https://ieeexplore.ieee.org/abstract/document/10516659

25 The Alan Turing Institute. *Human rights, democracy, and the rule of law assurance framework for AI systems: A proposal* (2022). P.251-276, https://doi.org/10.5281/zenodo.5981675

AI & EQUALITY

aiequalitytoolbox.com

Use our <AI &Equality> Human Rights AI Impact Assesment Workbook

A comprehensive tool combining formal assessment with ongoing reflection throughout the AI lifecycle.

HRIA Workbook

Join our community

for reading groups, panels, community publications and collaborative policy comments.

Community

Annex: Case studies

The following case studies demonstrate how human rights-centered AI development can work in practice across diverse African contexts. From Uganda's Makerere Health Lab creating locally-grounded malaria diagnostics to Nigeria's Nsukka Yellow Pepper Project supporting women farmers through participatory design, these examples show that meaningful community engagement is fundamental to building AI systems that actually work in their intended contexts. Whether addressing language preservation in Kenya's NLP projects, improving urban air quality monitoring in Nakuru, or creating accessible sign language translation tools, each case reveals how centering affected communities from the outset leads to more robust, culturally relevant, and sustainable solutions. These cases also expose critical gaps in current AI development, particularly around Technology-Facilitated Gender-Based Violence, where systems designed without considering gendered harms amplify both individual and coordinated attacks against women.

Contents

Technology-Facilitated Gender-Based Violence in Africa — 41
AI Climate Sensors in Africa — 61
AI for Kenyan Sign Language and Digital Inclusion — 74
AI-Powered Malaria Diagnostics — 92
Co-Creating AI for Agriculture: Nigeria's Nsukka Yellow Pepper Project — 102
Empowering African Languages through NLP: KenCorpus Project — 115
Design by Inclusion in AI Development — 130

<AI & Equality> African Toolbox | Case study

Technology-Facilitated Gender-Based Violence in Africa: When AI Becomes a Weapon

Watch the video

This case study is part of the **African <AI & Equality> Toolbox**, which builds upon the methodology of the global <AI & Equality> Human Rights Toolbox—an initiative of Women At The Table in collaboration with the United Nations Office of the High Commissioner for Human Rights (OHCHR). The African Toolbox is a collaboration between the <AI & Equality> initiative and the African Centre for Technology Studies (ACTS). To learn more visit **aiequalitytoolbox.com**

Introduction

In January 2025, Ethiopian Mayor Adanech Abiebie woke to find her face digitally grafted onto intimate videos with political leaders—deepfakes so convincing that 90% of viewers believed the fabricated narrative. Within hours, the AI-generated content linking her to Prime Minister Abiy Ahmed had garnered over 562,000 views, spreading the false claim that her political success stemmed from sexual relationships rather than competence. Meanwhile, in Cameroon, President Paul Biya's daughter Brenda faced a coordinated avalanche of harassment after publicly disclosing her sexual orientation—92 Facebook posts using identical templates reached 8.9 million people with before-and-after photos designed to mock her appearance and identity.

These aren't isolated incidents. They're part of a sophisticated, continent-wide campaign of Technology-Facilitated Gender-Based Violence (TFGBV) that weaponizes AI systems, exploits algorithmic amplification, and leverages cultural tensions to silence women and LGBTQ+ individuals across Africa. What Code for Africa's research reveals is both the staggering scale of these attacks—individual campaigns reaching millions—and their increasing sophistication as perpetrators learn to game AI systems designed to maximize engagement.

In Nigerian livestreams, young women are coerced into sexual acts through coordinated mass reporting threats. In Uganda, AI-powered content moderation systems fail to detect local language slurs like "woubi" and "lélé" that flood social media with anti-LGBTQ+ hatred. Across eleven African countries—Burundi, Cameroon, Côte d'Ivoire, Ethiopia, Ghana, Kenya, Nigeria, Uganda, South Africa, Senegal, and Zimbabwe—digital platforms have become battlegrounds where artificial intelligence amplifies rather than prevents systematic harassment targeting gender and sexual minorities.

The human cost is devastating: women political leaders withdrawing from public life, LGBTQ+ individuals silenced by fear, and democratic discourse degraded by campaigns that achieve massive reach through algorithmic promotion of controversial content. But this case study reveals something more troubling: current AI architectures, optimized for engagement rather than human dignity, create structural vulnerabilities that make such attacks not just possible but profitable for platforms and effective for perpetrators.

The Weaponization of Engagement: How AI Amplifies Hatred

The Ethiopian Mayor: When Deepfakes Target Democracy

The attack on Addis Ababa Mayor Adanech Abiebie began with a single TikTok account that had mastered the art of viral manipulation. On January 2, 2025, the account posted an AI-generated video showing Abiebie kissing Ethiopian Prime Minister Abiy Ahmed—a fabrication so seamless that it required technical analysis to identify as synthetic media. The video's caption suggested she had secured her mayoral position through sexual relations, tapping into deeply rooted biases about women in leadership.

What happened next reveals the terrifying efficiency of AI-driven harassment campaigns. Within the first 20 comments, 90% supported the video's false narrative, often responding with laughing emojis that signal high engagement to TikTok's algorithm. The platform's recommendation system, interpreting emotional reaction as user interest, began promoting the content to wider audiences. By November, a second deepfake video linking Abiebie to the Equatorial Guinea sex scandal had been created and distributed by the same account, demonstrating how successful harassment campaigns evolve and expand.

The technical sophistication was matched by cultural precision. The videos didn't just use AI to create convincing forgeries—they leveraged existing social attitudes about women in politics, transforming cutting-edge technology into a weapon for ancient prejudices. The mayor's actual governance record, including controversial urban development projects, became secondary to fabricated sexual narratives designed to undermine her authority through gendered attacks.

But the most chilling aspect wasn't the technology—it was how the platform's own AI systems became unwitting accomplices. TikTok's engagement-optimized algorithm treated the high emotional response as a signal to promote the content further, turning artificial intelligence into an amplification engine for artificial lies.

Brenda Biya: The Anatomy of Coordinated Digital Violence

When Cameroon's First Daughter Brenda Biya publicly came out as lesbian, she unknowingly triggered one of the most documented coordinated harassment campaigns in African digital history. The response wasn't spontaneous outrage—it was a precisely orchestrated attack that revealed the industrial scale of modern TFGBV operations.

Code for Africa's analysis of the campaign reads like a blueprint for digital violence. Ninety-two Facebook posts contrasted her "before and after" appearance, collectively reaching

8.9 million people and generating 17,745 interactions. But the devil was in the details: thirty-four of these posts used identical copy-paste techniques, featuring the same captions and layouts with surgical precision. This wasn't organic community response—it was coordinated inauthentic behavior designed to maximize algorithmic amplification.

The campaign's efficiency was staggering. The 34 identical posts alone generated 8.05 million views and 14,651 interactions, demonstrating how template-based attacks could achieve massive reach through minimal effort. Comments like "before she started sleeping with girls" reduced her changed style to sexual stereotypes, while others used her image to symbolize national decline: "She reflects the country's progress." A review of 4,600 comments found that 98% mocked or ridiculed Biya—a level of unanimity that suggested orchestrated rather than organic sentiment.

The cross-platform coordination was equally sophisticated. Between September 2024 and March 2025, approximately 50 TikTok videos—mostly posted by Ivorian users—continued the mockery as part of a "Cameroon vs Côte d'Ivoire" social media trend. Individual videos received hundreds of thousands of views, with coordinated timing patterns that maximized algorithmic visibility across platforms.

What made this campaign particularly devastating was how it exploited legitimate cultural discourse. The "country comparison" trend provided plausible cover for harassment, allowing attackers to frame systematic targeting as playful regional rivalry. This cultural camouflage made the content harder for automated systems to identify as harmful while ensuring it resonated with audiences predisposed to anti-LGBTQ+ sentiment.

The Nigerian Livestream Economy: AI-Enabled Sexual Exploitation

In Nigeria's TikTok ecosystem, Code for Africa documented something even more disturbing: the emergence of an AI-enabled sexual exploitation economy that uses platform features to coerce young women into performing sexual acts for online audiences. The case reveals how live streaming platforms become venues for real-time digital violence that combines technological coercion with economic manipulation.

The system operates with industrial efficiency. Hosts like @❤RICHARD DP and @SpecialPoint use phrases like "view once" to suggest content will only be visible temporarily, exploiting young women's concerns about permanent exposure. But viewers routinely record these sessions, preserving and redistributing content across platforms to maximize harm. One recording of a SpecialPoint livestream posted on X received 1.4 million views, transforming a moment of coercion into lasting digital violence.

The coercion mechanism reveals sophisticated understanding of platform vulnerabilities. When women set boundaries around what they're willing to do, hosts coordinate mass reporting campaigns to threaten account suspension—essentially weaponizing platform safety mechanisms to enable abuse. In March 2025, researchers documented a host threatening to disable a young woman's account when she refused to expose herself, while viewers coordinated pressure tactics through coordinated messaging.

The AI dimension becomes clear in how these operations evade detection. Hosts use sequential username variations after suspensions—adding letters or numbers to return under slightly modified handles. The platforms' automated systems, designed to detect spam or commercial manipulation, consistently fail to identify these harassment networks that operate at the intersection of sexual exploitation and coordinated inauthentic behavior.

Perhaps most troubling is how platform algorithms reward this content. The high engagement generated by controversial livestreams—driven by a combination of sexual content and audience participation—signals to recommendation systems that this content should be promoted to broader audiences. The platforms' own AI systems become enablers of exploitation, transforming human trafficking into algorithmic success.

Cultural Warfare: Anti-LGBTQ+ Campaigns as Information Operations

Uganda's Legislative Hatred: When Laws Become Content

Uganda's Anti-Homosexuality Act, enacted on May 29, 2023, didn't just criminalize LGBTQ+ identities—it provided a legal foundation for coordinated digital harassment campaigns that achieved massive reach through AI-driven amplification. Code for Africa's analysis reveals how legislative hatred translates into viral content that spreads across borders and platforms.

Seven TikTok videos supporting the Act achieved a combined 868,030 views and 39,452 interactions, but their timing reveals strategic coordination. These posts appeared from March to April 2023—before the Act's enactment—indicating pre-existing anti-LGBTQ+ discourse designed to build support for criminalization. The content used phrases like "homosexuality is a sin," "sodomised," and "say no to homosexuality (LGBTQ)" that became viral hashtags amplified across X, TikTok, and Facebook.

The campaign's cross-border reach demonstrated how local legislation becomes regional propaganda. Ugandan content celebrating criminalization spread to Kenya following their Supreme Court's LGBTQ+ rights ruling, to Tanzania during parliamentary debates about LGBTQ+ support funding, and to Burundi where President Évariste Ndayishimiye suggested stoning homosexual people. Each national moment became an opportunity for coordinated amplification that transcended borders.

The technical sophistication was hidden beneath cultural authenticity. Videos featured local influencers, religious leaders, and politicians speaking in native languages about preserving African values against foreign corruption. This cultural resonance made the content highly engaging for target audiences while providing plausible cover for coordinated campaigns. TikTok's recommendation algorithm, unable to distinguish between genuine cultural expression and manufactured hatred, promoted the most engaging content to broader audiences.

Tanzania's Parliamentary Theater: Transforming Hatred into Headlines

Tanzania's parliamentary debate on May 17, 2024, reveals how TFGBV campaigns exploit democratic institutions to generate viral content. When MPs condemned ministry support for LGBTQ+ projects as threats to Tanzanian cultural values, their speeches became raw material for coordinated digital amplification that reached over a million people.

MP Mwita Waitara's declaration that "We do not want homosexuality in Tanzania. We do not want filthy behaviour here" became the centerpiece of a sophisticated content operation. Nine TikTok clips sharing his homophobic comments received 1,070,640 views and 67,877 interactions, while X posts supporting the MPs' statements reached 17,769 views through coordinated resharing.

The campaign's effectiveness stemmed from exploiting democratic legitimacy. Parliamentary speeches provided authoritative sources for anti-LGBTQ+ content that platforms couldn't dismiss as hate speech—after all, these were elected officials speaking in official forums. This institutional cover enabled massive amplification of harmful content under the guise of political reporting.

The algorithmic amplification patterns revealed how AI systems inadvertently promote institutional hatred. Parliamentary debates generate high engagement because they involve political conflict and controversial topics. Recommendation algorithms, optimized for user interest rather than social harm, promoted the most controversial clips to audiences likely to engage with anti-LGBTQ+ content. The result was democratic institutions becoming content factories for coordinated harassment campaigns.

Burundi's Presidential Violence: When Leaders Incite Digital Mobs

President Évariste Ndayishimiye's December 29, 2023 suggestion that homosexual people "should be put in a stadium and stoned" demonstrates how TFGBV campaigns exploit the highest levels of political authority. The statement generated 3,650 mentions on X, receiving approximately 25,500 engagements and 980,000 views, while 188 Facebook posts shared the president's comments between December 2023 and May 2024.

The viral amplification revealed sophisticated coordination mechanisms. Rather than simple resharing, the campaign involved strategic timing patterns that maximized algorithmic visibility. Posts appeared at optimal engagement times across different platforms, suggesting coordinated scheduling designed to maintain momentum over months rather than days.

The content moderation challenges became apparent when TikTok searches for related content triggered community guideline warnings, yet the material continued circulating through screenshot sharing and indirect references. This cat-and-mouse dynamic demonstrates how sophisticated harassment campaigns adapt to platform policies while maintaining their reach and impact.

The Technical Architecture of Digital Violence

Algorithmic Amplification: How AI Rewards Hatred

Code for Africa's research reveals a disturbing pattern: AI recommendation systems consistently amplify TFGBV content because emotional provocation generates the high engagement that algorithms interpret as user satisfaction. Analysis across platforms shows that controversial content targeting women and LGBTQ+ individuals achieves 15-20% higher engagement rates than baseline content, leading to algorithmic promotion that multiplies reach exponentially.

The Ethiopian mayor case provides a clear example of this dynamic. The initial deepfake video achieved 562,138 views not through paid promotion but through organic algorithmic amplification driven by high engagement rates. Users commenting with laughing emojis, sharing the content, and spending time viewing the fabricated material all sent positive signals to TikTok's recommendation system. The AI interpreted coordinated harassment as user interest, promoting the content to broader audiences who might not have encountered it otherwise.

This creates a feedback loop where harmful content becomes self-amplifying. Initial engagement drives algorithmic promotion, which increases reach, which generates more engagement, which triggers further promotion. The result is that well-executed harassment campaigns can achieve viral status without significant financial investment—they simply need to generate enough initial engagement to trigger algorithmic amplification.

The temporal patterns are equally concerning. Code for Africa found that 80% of TFGBV campaign engagement occurs within the first 48 hours, suggesting that algorithmic promotion decisions made in the crucial early period determine ultimate reach and impact. This creates a narrow window where intervention might be effective, but current content moderation systems consistently fail to respond quickly enough to prevent viral amplification.

Evasion Technologies: Gaming the System

The sophistication of TFGBV technical tactics reveals how perpetrators have developed systematic approaches to circumvent content moderation while maximizing algorithmic amplification. The research documents a comprehensive toolkit of evasion strategies that exploit specific vulnerabilities in AI-driven platforms.

"Spamouflage" techniques represent the most basic level of evasion. Attackers replace letters with symbols or numbers—writing "Us£less" instead of "Useless" or "w0n" instead of "won"—to bypass keyword-based detection systems. These modifications are subtle enough that human readers understand the meaning while automated systems fail to recognize harmful content.

More sophisticated is the exploitation of cultural and linguistic gaps in AI training data. Terms like "woubi" and "lélé"—French slurs targeting LGBTQ+ individuals—pass through content moderation systems trained primarily on English-language datasets. This cultural blindness creates systematic vulnerabilities that attackers exploit to spread harmful content in African contexts where local knowledge is essential for harm recognition.

Account management strategies reveal industrial-scale coordination. When harassment accounts face suspension, they return with slightly modified usernames—adding numbers or letters to maintain brand recognition while evading automated detection. Pre-registered backup accounts enable immediate resumption of activities, while cross-platform coordination ensures campaign persistence even when individual accounts face enforcement action.

The temporal coordination demonstrates sophisticated understanding of algorithmic systems. Coordinated campaigns time their posts for maximum algorithmic visibility, leverage trending topics to increase reach, and use engagement manipulation to trigger recommendation system promotion. This isn't amateur trolling—it's professional information warfare adapted for gender-based violence.

Content Moderation Failures: When AI Can't See Culture

The systematic failures of content moderation reveal fundamental limitations in how AI systems understand cultural context and coordinated behavior. Code for Africa's research documents specific cases where sophisticated harassment campaigns evaded detection despite clear coordination patterns.

The Brenda Biya case provides the starkest example. Thirty-four Facebook posts using identical copy-paste techniques should have triggered automated detection systems designed to identify coordinated inauthentic behavior. Yet these posts collectively achieved 8.05 million views while evading platform enforcement. The identical captions, synchronized timing, and template-based sharing patterns represent textbook examples of coordination that current AI systems fail to detect.

The linguistic gaps are equally problematic. Content moderation systems trained primarily on Western datasets demonstrate reduced effectiveness with African cultural contexts and language patterns. Local slurs, cultural references, and context-dependent harmful content consistently pass through automated systems designed for different linguistic and cultural environments.

Real-time detection capabilities prove inadequate for the speed of viral content. TFGBV campaigns achieve massive reach before content moderation systems can respond effectively. The Ethiopian mayor's deepfake video reached over 500,000 people before any intervention, while livestream exploitation in Nigeria occurs in real-time with minimal possibility for protective intervention.

Perhaps most concerning is how algorithmic promotion outpaces human review. Content that violates platform policies still receives algorithmic amplification during the period between posting and moderation review. This creates a window where harmful content can achieve viral status even if it's eventually removed, making content moderation reactive rather than protective.

Human Rights in the Age of Algorithmic Violence

1 Stage 1: Objective and Team Composition
The Foundation of Harm

The human rights violations documented in TFGBV campaigns begin with fundamental design decisions made during AI system development. Platform objectives optimized for user engagement create structural incentives that reward controversial content regardless of social harm. Code for Africa's analysis demonstrates how these engagement-focused metrics systematically promote harassment campaigns targeting women and LGBTQ+ individuals.

The Ethiopian mayor case illustrates this dynamic clearly. TikTok's algorithm interpreted high emotional engagement with deepfake content as user satisfaction, promoting fabricated harassment material to broader audiences. The platform's objective function—maximize user engagement and time on platform—directly conflicted with human rights principles of dignity and non-discrimination. Yet the technical system had no mechanism for recognizing this conflict.

Team composition during AI development reveals systematic exclusion of affected communities and human rights expertise. Platform development teams lack meaningful representation from women, LGBTQ+ individuals, or African communities who bear the consequences of system design decisions. This exclusion isn't accidental—it reflects broader power structures that prioritize technical capability over social responsibility.

The absence of human rights considerations in this stage has cascading effects throughout the AI lifecycle. When systems are designed to maximize engagement without considering dignity, participation, or equality, they become vulnerable to exploitation by sophisticated harassment campaigns. The technical architecture embeds these values from inception, making later interventions inadequate for addressing fundamental structural problems.

Human Rights Alignment Requirements:

- **Community Agency in Objective Setting**: Affected communities must have genuine decision-making power in defining what AI systems should optimize for, not just feedback on predetermined technical goals.
- **Dignity-Centered Metrics**: Success measurements must include human dignity, democratic participation, and community safety alongside engagement and revenue metrics.
- **Representative Development Teams**: Meaningful inclusion of women, LGBTQ+ individuals, and African communities in technical decision-making roles.
- **Human Rights Expertise Integration**: Systematic inclusion of human rights practitioners in technical architecture and objective-setting processes.

2 Stage 2: Defining System Requirements
Building Safety into Technical Specifications

Current system requirements demonstrate fundamental inadequacy in addressing coordinated harassment campaigns targeting specific demographics. The Brenda Biya case reveals how 34 identical Facebook posts evaded automated detection systems designed primarily for spam or commercial manipulation rather than gender-based violence.

The technical requirements gaps extend beyond simple detection failures. Platforms lack demographic-specific harm monitoring, cultural context understanding, and rapid response capabilities for coordinated campaigns. The Nigerian livestream exploitation demonstrates how real-time TFGBV occurs faster than current moderation systems can respond, requiring fundamentally different technical architectures.

Cross-platform coordination represents another systematic requirement failure. TFGBV campaigns operate across TikTok, Facebook, X, and other platforms simultaneously, but current systems lack information-sharing capabilities to detect distributed harassment networks. Individual platforms optimize their own metrics while remaining blind to coordinated campaigns that span the digital ecosystem.

The absence of affected community input in requirements definition creates systems optimized for metrics that conflict with human rights. Engagement maximization, viral amplification, and recommendation system effectiveness become requirements without consideration of how these features enable systematic harassment of marginalized communities.

Human Rights-Aligned System Requirements:

- **Real-time Coordination Detection:** Technical capabilities to identify synchronized posting patterns, template sharing, and cross-platform campaign coordination.
- **Cultural Context Integration:** Content evaluation systems that understand local languages, cultural references, and context-dependent harmful content.
- **Demographic-Specific Harm Monitoring:** Systematic tracking of system impacts on women, LGBTQ+ individuals, and other marginalized communities.
- **Community-Defined Safety Standards:** Requirements development that includes affected community input on what constitutes harm and appropriate intervention.
- **Rapid Response Architecture:** Technical systems capable of intervention before viral amplification occurs rather than reactive content removal.

3

Stage 3: Data Discovery
Bias and Representation in Training Systems

Training data bias contributes systematically to TFGBV amplification through cultural blindness and representation gaps. Content moderation models trained primarily on Western datasets demonstrate significant effectiveness gaps when deployed in African contexts, failing to recognize local language slurs and culturally specific harmful content.

The linguistic bias is particularly severe. Terms like “woubi” and “lélé”—slurs targeting LGBTQ+ individuals in French-speaking African countries—pass through moderation systems that lack training data from these linguistic and cultural contexts. This isn’t simply a technical oversight—it reflects systematic underrepresentation of African voices in AI training data collection and curation.

Recommendation algorithm training demonstrates similar bias patterns. Models optimized on datasets that don’t include sophisticated harassment campaigns fail to recognize coordinated TFGBV tactics when deployed in African contexts. The algorithms learned to maximize engagement from data that didn’t capture the specific ways that marginalized communities face systematic digital violence.

The data collection process itself violates human rights principles by excluding affected community consent and participation. Training datasets include harassment content targeting women and LGBTQ+ individuals without their consent, while failing to include community knowledge about harmful content recognition and appropriate intervention strategies.

Human Rights-Aligned Data Practices:

- **Community Consent and Participation:** Affected communities must have agency in determining how their data is collected, used, and represented in training systems.
- **Cultural Representativeness:** Training data must include diverse African languages, cultural contexts, and community-defined examples of harmful content.
- **Participatory Dataset Curation:** Community experts should be involved in identifying harmful content patterns and appropriate intervention strategies.
- **Bias Impact Assessment:** Systematic evaluation of how training data representation affects different communities and intervention effectiveness.

4 Stage 4: Selecting and Developing Models
Technology in Service of Human Rights

Model selection and development decisions directly enable TFGBV through engagement optimization that rewards controversial content. Recommendation algorithms trained to maximize user engagement systematically promote harassment campaigns because emotional provocation generates the high interaction rates that models interpret as success.

The technical architecture embeds these harmful incentives throughout the system. Content that generates strong emotional responses—including coordinated harassment targeting women and LGBTQ+ individuals—receives algorithmic promotion regardless of social impact. Models optimized for engagement metrics lack mechanisms for recognizing when high interaction rates indicate harm rather than user satisfaction.

Explainability limitations prevent affected communities from understanding how algorithmic systems make decisions about content promotion and moderation. When harassment campaigns achieve viral reach through algorithmic amplification, victims and advocates have no insight into why these decisions occurred or how to challenge them effectively.

Fairness considerations remain absent from model development despite documented evidence that current systems systematically amplify harassment targeting specific demographics. The lack of intersectional fairness metrics means that platforms cannot identify when their systems disproportionately harm women, LGBTQ+ individuals, or other marginalized communities.

Human Rights-Aligned Model Development:

- **Community-Defined Success Metrics:** Models should optimize for community-identified values like safety, dignity, and democratic participation rather than purely engagement-focused metrics.
- **Harassment-Aware Architecture:** Technical systems must be designed to recognize when high engagement indicates coordinated harassment rather than organic user interest.
- **Transparent Decision-Making:** Affected communities must be able to understand how algorithmic systems make decisions about content promotion and moderation.
- **Intersectional Fairness Integration:** Models must include systematic evaluation of impacts on multiply marginalized communities and intersectional harm recognition.

5 Stage 5: Testing and Evaluation
Community-Centered Validation

Current testing frameworks demonstrate insufficient consideration of TFGBV scenarios and community-defined harm. Evaluation protocols focus on technical performance metrics rather than community safety outcomes, missing systematic ways that platforms enable harassment campaigns targeting marginalized groups.

The absence of affected community participation in testing creates systems optimized for metrics that conflict with human rights. Platforms measure success through engagement rates, user growth, and retention without systematic evaluation of impacts on women, LGBTQ+ individuals, and other vulnerable communities.

Real-world testing limitations mean that harassment scenarios receive inadequate evaluation during development. The sophisticated coordination tactics documented by Code for Africa—template sharing, cross-platform campaigns, cultural code-switching—represent attack patterns that current evaluation frameworks fail to anticipate or address.

Performance measurement systems lack demographic-specific assessment capabilities. Platforms cannot identify when their systems systematically amplify harassment targeting specific communities because they lack evaluation frameworks designed to detect these patterns.

Human Rights-Aligned Testing Approaches:

- **Community-Defined Harm Assessment:** Testing protocols must include affected community evaluation of what constitutes harmful system behavior and appropriate intervention.
- **Adversarial Harassment Scenario Testing:** Systematic evaluation against documented TFGBV tactics and coordination patterns.
- **Demographic-Specific Performance Monitoring:** Regular assessment of system impacts on different communities with particular attention to marginalized groups.
- **Real-World Impact Evaluation:** Testing that goes beyond technical metrics to assess effects on human dignity, democratic participation, and community safety.

6 Stage 6: Deployment & Post-Deployment Monitoring
Accountability and Continuous Improvement

Post-deployment monitoring reveals systematic gaps in platform capabilities to detect and respond to coordinated harassment campaigns. TFGBV operations achieve massive reach before intervention because current monitoring systems are reactive rather than proactive and lack real-time coordination detection capabilities.

The response capability limitations demonstrate how platforms prioritize technical performance over community protection. Average response times for content moderation exceed viral content spread times, meaning that harassment campaigns consistently achieve their objectives before any protective intervention occurs.

Community feedback integration remains inadequate despite sophisticated systems for collecting user reports and appeals. Affected communities report coordinated harassment campaigns that platforms fail to recognize as systematic threats rather than individual content violations.

Systematic learning from TFGBV incidents is limited by platforms' reluctance to acknowledge that their technical architectures enable harassment. Without recognition of structural problems, platforms focus on reactive content removal rather than proactive system design changes that could prevent future campaigns.

Human Rights-Aligned Monitoring and Response:

- **Proactive Threat Detection:** Real-time monitoring systems capable of identifying coordinated campaigns before they achieve viral amplification.
- **Community Agency in Intervention:** Affected communities must have mechanisms to rapidly escalate threats and influence platform response decisions.
- **Systematic Impact Assessment:** Regular evaluation of how platform systems affect human rights with particular attention to marginalized communities.
- **Structural Learning Integration:** Platform commitment to modifying technical architectures based on documented human rights impacts rather than limiting response to content removal.

Building Human Rights into AI Architecture

Technical Interventions That Center Dignity

The path forward requires fundamental architectural changes that embed human rights principles into AI system design rather than treating them as external constraints. Code for Africa's research provides a roadmap for technical interventions that could effectively mitigate TFGBV while maintaining platform functionality and innovation.

- **Engagement Quality Assessment Systems** represent the most critical intervention. Instead of optimizing purely for interaction quantity, platforms must develop technical capabilities to distinguish between positive engagement (learning, community building, democratic participation) and negative engagement (harassment, discrimination, coordinated attacks). This requires training models on community-defined examples of constructive versus harmful interaction patterns.

- **Coordination Detection Integration** must become a core platform capability rather than an afterthought. The Brenda Biya case demonstrates how 34 identical posts can evade detection despite clear coordination patterns. Platforms need real-time network analysis capabilities that can identify template sharing, synchronized timing, and cross-platform coordination before viral amplification occurs.

- **Cultural Context Recognition** requires systematic integration of African languages, cultural references, and local knowledge into content moderation systems. The failure to detect slurs like "woubi" and "lélé" isn't a minor oversight—it reflects systematic exclusion of African voices from AI development that must be corrected through participatory dataset development and community expert integration.

- **Rapid Response Architecture** must enable intervention before viral spread rather than reactive content removal. This requires predictive systems that can identify potential harassment campaigns in their early stages and protective measures that can be activated within minutes rather than hours or days.

Community Ownership and Platform Governance

Technical solutions alone cannot address TFGBV without corresponding changes in platform governance that give affected communities genuine agency in system design and operation. The documented harassment campaigns succeed partly because platforms operate as closed systems where community voices have minimal influence on technical decisions.

- **Community Advisory Integration** must go beyond tokenistic consultation to include affected communities in technical architecture decisions, policy development, and evaluation criteria. The Ethiopian mayor's experience with deepfake harassment could have been prevented if platform design had included Ethiopian women's organizations in identifying potential harms and appropriate interventions.

- **Transparent Algorithmic Decision-Making** requires platforms to provide affected communities with meaningful information about how recommendation systems promote content and why specific moderation decisions occur. Currently, harassment victims have no insight into why coordinated campaigns achieve viral reach or how to effectively challenge algorithmic amplification of harmful content.

- **Community-Defined Success Metrics** must supplement or replace engagement-focused optimization with measurements that reflect human rights principles. Platform success should be evaluated based on community safety, democratic participation, and dignity rather than purely technical metrics that may conflict with human rights.

- **Cross-Platform Coordination** requires industry-wide cooperation to address harassment campaigns that span multiple platforms. Individual platform optimization creates systematic vulnerabilities that sophisticated campaigns exploit through distributed coordination.

Regulatory Frameworks and International Cooperation

The transnational nature of TFGBV campaigns documented across eleven African countries requires coordinated policy responses that can address cross-border digital violence while protecting legitimate communication and democratic participation.

- **TFGBV-Specific Legal Frameworks** must address the sophisticated coordination mechanisms that current laws don't adequately cover. The harassment campaigns targeting the Ethiopian mayor and Brenda Biya represent forms of coordinated digital violence that require legal recognition and enforcement mechanisms designed for AI-enabled coordination.

- **Platform Accountability Standards** must include specific requirements for TFGBV prevention rather than generic content moderation obligations. Platforms should be legally required to maintain systems capable of detecting coordinated harassment campaigns and providing rapid protective intervention for targeted individuals.

- **International Cooperation Mechanisms** are essential for addressing campaigns that exploit platform coordination across different jurisdictions. The viral anti-LGBTQ+ content spreading from Uganda to Tanzania to Kenya demonstrates how local legislation becomes regional propaganda that requires coordinated response capabilities.

- **Community Participation Requirements** must be embedded in regulatory frameworks to ensure that affected communities have genuine agency in defining harm and appropriate intervention rather than having technical solutions imposed by external authorities.

Conclusion:
Reclaiming AI for Human Dignity

The documented patterns of Technology-Facilitated Gender-Based Violence across Africa reveal both the devastating human cost of AI systems optimized for engagement over dignity and the potential for technical architectures that serve human rights instead of undermining them. The Ethiopian mayor whose fabricated sexual scandals reached over half a million people, Brenda Biya whose harassment campaign generated 8.9 million views, and the countless women coerced in Nigerian livestreams represent not isolated tragedies but systematic failures of AI systems designed without meaningful consideration of human rights principles.

Yet their experiences also illuminate the path forward. Every documented harassment campaign reveals specific technical vulnerabilities that can be addressed through AI architectures designed to center community safety over engagement maximization. Every coordination pattern that current systems fail to detect provides blueprints for more effective intervention mechanisms. Every cultural blindness in content moderation identifies opportunities for more inclusive AI development that includes African voices in technical decision-making.

The choice facing the AI development community is stark: continue building systems that systematically amplify digital violence against marginalized communities, or fundamentally restructure technical architectures to embed human rights principles throughout the development lifecycle. The research demonstrates that sophisticated harassment campaigns will continue exploiting engagement-optimized algorithms until platforms prioritize dignity over viral growth.

But this case study also reveals reasons for hope. The technical interventions required to address TFGBV—coordination detection, cultural context recognition, community participation mechanisms—represent advances that would benefit all platform users, not just those targeted by harassment campaigns. Building AI systems that protect the most vulnerable creates more robust, democratic, and sustainable digital environments for everyone.

The women political leaders, LGBTQ+ individuals, and marginalized communities targeted by these campaigns are not asking for special protection—they're demanding equal access to digital spaces free from systematic harassment that undermines their fundamental human rights. Their calls for justice provide blueprints for AI development that serves human flourishing rather than exploitation.

The deepfakes targeting the Ethiopian mayor continue circulating, but her experience has contributed to growing recognition that current AI architectures are fundamentally incompatible with human rights principles. The coordinated harassment of Brenda Biya reaches millions, but the documented coordination patterns provide technical specifications for detection systems that could prevent future campaigns. The exploitation documented in Nigerian livestreams continues, but the evidence of systematic coordination offers pathways for protective intervention.

Their experiences, documented through Code for Africa's meticulous research, transform individual trauma into collective knowledge that can reshape how AI systems relate to human dignity. The women who faced these attacks have become inadvertent experts in the vulnerabilities of engagement-optimized algorithms and the possibilities for technical architectures that center community safety.

The next phase of AI development will be defined by whether the technical community learns from their experiences or continues building systems that amplify the very forms of digital violence these women have endured. The choice is between AI that serves engagement metrics regardless of human cost and AI that treats human dignity as the ultimate optimization target.

In the end, the women whose harassment campaigns are documented in this research are not just victims of algorithmic violence—they are unwitting pioneers of a more democratic approach to AI development that centers community needs over technical convenience. Their suffering demands nothing less than fundamental transformation of how artificial intelligence relates to human rights.

The technology exists to build these better systems. The legal frameworks can be developed to ensure accountability. The community knowledge is available to guide more inclusive development processes. What remains is the political will to prioritize human dignity over engagement maximization and community safety over viral growth.

The women of Africa who have faced these attacks are still speaking, still leading, still demanding digital spaces that honor their humanity. Their voices, amplified not by engagement-hungry algorithms but by principled solidarity, point toward AI futures that serve human flourishing rather than exploitation. The question is whether the technical community will listen.

About the case study

This research uses behavioural and narrative analysis to examine technology-facilitated gender-based violence across 11 African countries, drawing on social media data to identify disinformation and coordinated harassment patterns. Code for Africa (CfA), the continent's largest civic technology and data journalism initiative, supported the research through its expertise in open-source intelligence and data-driven investigations.

This report was compiled by Code for Africa's Hanna Teshager, a senior investigative data analyst at the iLAB team, using ML to combat disinformation, map online coordinated inauthentic behaviour, and influence operations. The report is based on her research and analysis with investigative data analysts, including senior investigative data analyst John Ndung'u, Chike Odita, Fatimaelzahra Saeed, Moffin Njoroge, and Vanessa Manessong. The research ongoing since 2023 examines TFGBV patterns across 11 African countries. About the authors Hanna Teshager is CfA's senior Investigative Data Analyst at the iLAB team, with 4+ years of experience using ML to combat disinformation, map online coordinated inauthentic behaviour, and influence operations.

Other contributors to this case study are Caitlin Kraft-Buchman, Emma Kallina, and Sofia Kypraiou, authors of the original *Framework to AI Development: Integrating Human Rights Considerations Along the AI Lifecycle* upon which the Toolbox structure is based. Additional contributors are Amina Soulimani and Pilar Grant, from Women at the Table and the <AI & Equality> Human Rights Initiative.

<AI & Equality> African Toolbox | Case study

AI Climate Sensors in Africa

Watch the video

This case study is part of the **African <AI & Equality> Toolbox**, which builds upon the methodology of the global <AI & Equality> Human Rights Toolbox—an initiative of Women At The Table in collaboration with the United Nations Office of the High Commissioner for Human Rights (OHCHR). The African Toolbox is a collaboration between the <AI & Equality> initiative and the African Centre for Technology Studies (ACTS). To learn more visit **aiequalitytoolbox.com**

61

Introduction

In a bustling Nairobi neighborhood near an industrial zone, residents had long suspected that the persistent coughs plaguing their children and the acrid smell hanging in the air were connected to the nearby factories. But without data, their concerns were dismissed by authorities as mere complaints. Meanwhile, in rural Tanzania, fishing communities noticed changes in weather patterns that affected their livelihoods, but lacked the evidence to understand or adapt to these shifts. These stories reflect a broader challenge across Africa: environmental injustices compounded by a lack of reliable data to document, understand, and address climate and pollution impacts.

Code for Africa's *sensors.AFRICA* initiative emerged from this gap between lived experience and documented evidence. What began as a response to journalists' struggles to report on air quality issues has evolved into a comprehensive AI-driven platform that empowers communities across the continent to monitor, understand, and advocate for their environmental rights.

The Genesis: From Data Scarcity to Community Empowerment

The *sensors.AFRICA* story began in newsrooms across Africa, where journalists faced a frustrating reality: they could see and smell the pollution, hear community complaints about deteriorating air quality, but had no reliable data to support their reporting. Traditional monitoring infrastructure, where it existed at all, was prohibitively expensive and often controlled by the same institutions that communities were trying to hold accountable.

Alicia Olago, an environmental scientist and the senior product manager of *sensors.AFRICA*, recalls the moment this challenge crystallized into action: "We realized that the absence of data wasn't just a technical problem—it was a justice issue. Communities were suffering, but without evidence, their voices were marginalized."

This recognition led to a fundamental principle that would guide the entire initiative: environmental data should be open, accessible, and controlled by the communities most affected by environmental challenges. But achieving this vision required more than just deploying sensors—it demanded a complete reimagining of how environmental monitoring could work in African contexts.

Building trust through technology: the community-centered approach

Unlike traditional top-down environmental monitoring systems, *sensors.AFRICA* adopted a community-centered approach from its inception. This wasn't merely a matter of good practice—it was essential for the initiative's success and sustainability.

In urban areas like Nairobi's Korogocho informal settlement, the team learned that simply installing sensors wouldn't work without deep community engagement. The process begins with what they call "entry through champions"—identifying community leaders, NGOs, or passionate residents who serve as bridges between the technical team and the broader community.

But champions alone aren't enough. The real breakthrough came through participatory mapping sessions, where community members gather around printed maps of their neighborhoods—created using open-source OpenStreetMap data—to identify pollution sources, vulnerable areas, and priority concerns. These sessions reveal nuanced perspectives that outside experts would never capture: women highlighting different pollution sources than men, elderly residents pointing to health impacts, youth identifying environmental changes over time.

"The map becomes a conversation starter," explains Olago. "When you see a grandmother pointing to a specific corner and explaining how the air changes when the wind shifts, you understand that data collection isn't just about technology—it's about dignity and agency."

Technical innovation driven by community needs

The technical architecture of *sensors.AFRICA* reflects the realities of African environments and communities. The sensors measure particulate matter (PM1, PM2.5, PM10), relative humidity, temperature, and GPS coordinates, transmitting data in real-time through IoT systems. But the real innovation lies in how these technical capabilities were adapted to address infrastructure challenges and community needs.

Power supply emerged as a critical challenge early in the project's development. In urban areas, frequent power fluctuations and load-shedding made grid-dependent sensors unreliable. In rural areas, many communities had no grid access at all. The solution came through solar-powered sensors with locally sourced panels and battery packs—a technical adaptation that also supported local economies and simplified maintenance.

Connectivity presented another challenge. While urban areas had multiple connection options, rural deployments often faced limited cellular coverage. The team developed IoT SIM cards that could bounce between networks, finding the strongest available signal. Memory cards provided backup storage to prevent data loss during connectivity interruptions. These weren't just technical fixes—they were solutions designed with community sustainability in mind. Local sourcing meant that broken components could be replaced quickly without waiting for international shipments. Simple, robust designs meant that community members could perform basic maintenance themselves.

AI as a tool for environmental justice

The integration of artificial intelligence into *sensors.AFRICA* represents a natural evolution of the community-centered approach rather than a technological add-on. AI serves three primary functions within the initiative: filling data gaps, predicting environmental events, and making complex data accessible to diverse audiences.

Machine learning algorithms help predict and fill data gaps caused by sensor malfunctions, tampering, or connectivity issues. This isn't just a technical convenience—it ensures that communities don't lose critical evidence during important periods, such as when documenting pollution events for legal proceedings.

The AI-driven early warning system developed for Nakuru, Kenya, exemplifies how artificial intelligence can serve environmental justice. By analyzing real time sensor data, satellite data, alongside human sensor network data, the system predicts pollution events and sends alerts to residents through the AngaWATCH citizen App.. This gives communities advance warning to take protective measures, particularly important for vulnerable populations like children, pregnant women and the elderly.

Perhaps most importantly, AI helps make environmental data accessible across different literacy levels and languages. The system generates simplified visual displays using air quality indices, and creates monthly reports in local languages. This democratization of information transforms data from an elite resource into a community tool.

The power of open data and evidence-based advocacy

The true impact of *sensors.AFRICA* becomes clear through the stories of communities who have used the data for advocacy and change. In the Syokimau area outside Nairobi, residents suspected that a nearby factory was causing health problems in their community. Persistent chest problems among children, frequent pneumonia cases, and respiratory issues seemed connected to visible emissions from the industrial facility.

Working with *sensors.AFRICA*, the community installed air quality sensors that documented particulate matter levels well above WHO guidelines. The data provided the evidence base residents needed to approach authorities and media outlets. A 32-minute feature on Citizen TV, one of Kenya's most influential media houses, brought national attention to their situation, directly correlating the sensor data with community health impacts.

The story doesn't end with media coverage. Armed with documented evidence, residents took their case to the Kenyan National Environmental Tribunal. The combination of sensor data and community testimony created a compelling case that authorities couldn't dismiss as subjective complaints. This legal pathway, supported by concrete data, demonstrates how environmental monitoring can strengthen democratic institutions and environmental governance.

Similar stories have emerged across the continent. In Mukuru, another Nairobi informal settlement, sensor data documenting dangerous air quality levels was featured in The Star newspaper coverage titled "Mukuru fumes put 60 asthma patients a month in hospital." The direct correlation between environmental data and health outcomes provided the evidence base for policy discussions and intervention planning.

Expanding Horizons: The Earth Observation Vision

The success of air quality monitoring has inspired *sensors.AFRICA* to expand into comprehensive environmental monitoring through what they call "Earth Observation"—a vision that integrates ground sensors, drone imagery, and satellite data through AI analysis. This multi-modal approach addresses the full spectrum of environmental challenges facing African communities.

In partnership with *africanDRONE*, another Code for Africa initiative, the project plans to combine ground-based sensor data with aerial imagery to monitor deforestation, wildfires, floods, and drought impacts. AI algorithms analyze these diverse data streams to identify patterns, predict environmental events, and generate early warnings for communities.

Prior expansion into marine environments produced initiatives like tracking dynamite fishing through underwater acoustic sensors and the StormWatch project in Tanzania, which uses satellite remote sensing and climate models to protect fishermen from extreme weather events. Each expansion maintains the core principle of community ownership and open data access.

Navigating Challenges: Power, Politics, and Participation

The *sensors.AFRICA* initiative hasn't been without challenges. Ensuring sensor host safety, particularly when documenting pollution from powerful industrial actors, requires careful attention to data protection and anonymization. Regular check-ins with community hosts, secure data transmission, and strategic communication help protect participants while maintaining data integrity.

Political challenges emerge when sensor data conflicts with official narratives or powerful interests. The initiative addresses this through transparency, open data policies, and broad stakeholder engagement that includes government partners, civil society organizations, and academic institutions. By making data publicly accessible through multiple channels —API endpoints for technical users, visual displays for general audiences, and detailed reports for researchers— the initiative builds broad-based support that's difficult for any single actor to suppress.

Sustainability represents an ongoing challenge, particularly in remote rural areas. The community-centered approach provides partial solutions through local capacity building, peer learning networks, and integration with existing community structures. Training local individuals to interpret data and communicate findings creates sustainable support systems that don't depend on continuous external intervention.

Rural adaptations: environmental monitoring in remote communities

Rural deployments of *sensors.AFRICA* require different approaches than urban installations. Rather than solely responding to journalist requests or citizen complaints, rural projects typically begin with requests from partner organizations that have established trust within communities. This reflects different power dynamics and the need for more careful navigation of community relationships.

Participatory mapping becomes even more crucial in rural settings, where environmental issues affect different community groups in distinct ways. Women might highlight water scarcity and indoor air pollution from cooking fires, while youth focus on land degradation affecting future livelihoods. Men might emphasize different pollution sources or economic impacts. These diverse perspectives ensure that monitoring systems capture the full range of community concerns.

Technical adaptations for rural environments go beyond power and connectivity solutions. Early warning systems must work through SMS rather than smartphone apps, to accommodate rural communities who often lack smartphones and have limited literacy levels. Training programs must account for different education levels and technological familiarity, often using visual aids and hands-on demonstrations rather than written materials.

The Earth Observation vision holds particular promise for rural communities facing complex environmental challenges. By integrating ground sensors with drone and satellite imagery, AI can monitor large geographical areas for deforestation, predict flood risks, and track drought impacts. This comprehensive monitoring supports both immediate community needs and broader policy discussions about rural development and climate adaptation.

Impact Beyond Numbers: Transforming Environmental Governance

The success of *sensors.AFRICA* can't be measured only in sensors deployed or data points collected. The initiative has fundamentally changed how environmental issues are discussed and addressed across participating communities. By providing communities with their own data, the project has shifted power dynamics in environmental governance.

Academic researchers now regularly use *sensors.AFRICA* data in their studies, including deep learning approaches that combine the community-generated data with other environmental datasets. This academic engagement validates the scientific quality of the data while ensuring that community concerns reach scholarly and policy audiences.

The integration of qualitative research—focus group discussions, key informant interviews, and participatory observations—alongside quantitative sensor data provides rich narratives that pure technical monitoring couldn't capture. This mixed-methods approach helps policymakers and researchers understand not just what environmental changes are occurring, but how they affect different community members and what solutions might be most appropriate.

Lessons for AI and Human Rights

The *sensors.AFRICA* case study offers several crucial insights for AI development that respects and promotes human rights:

- **Community ownership is essential:** The most sophisticated AI system fails if communities don't trust it, understand it, or control its use. Starting with community needs rather than technical capabilities ensures that AI serves justice rather than merely demonstrating technological prowess.

- **Participation must be genuine:** Tokenistic consultation differs fundamentally from the deep engagement required for effective AI systems. Participatory mapping, community-led site selection, and local capacity building create genuine ownership that sustains initiatives over time.

- **Technical design reflects values:** Every technical choice—from power systems to data transmission protocols—embodies assumptions about users, contexts, and priorities. Designing for African contexts required fundamental rethinking of standard approaches, resulting in more robust and sustainable systems.

- **Data justice requires open access:** Environmental data becomes a tool for justice only when communities can access, understand, and use it. Open data policies, multiple access channels, and diverse presentation formats ensure that information serves empowerment rather than extraction.

- **AI can democratize expertise:** Rather than replacing human judgment, AI in *sensors.AFRICA* amplifies community knowledge and makes technical information accessible to diverse audiences. This democratization of expertise strengthens rather than undermines human agency.

Looking forward: Scaling environmental justice

As *sensors.AFRICA* expands across the continent, each new deployment offers opportunities to refine the community-centered approach while adapting to different environmental, political, and social contexts. The initiative's success has inspired similar projects globally, demonstrating that the principles developed in African contexts have broader applicability.

The Earth Observation vision represents the next phase of development, where AI integration becomes more sophisticated while maintaining the core commitment to community ownership and environmental justice. By combining multiple data sources through AI analysis, the initiative can address increasingly complex environmental challenges while preserving the local knowledge and community agency that make it effective.

Perhaps most importantly, *sensors.AFRICA* demonstrates that AI can serve environmental justice when developed with rather than for affected communities. The initiative's success stems not from technological sophistication alone, but from recognizing that environmental data is ultimately about human dignity, community empowerment, and the right to a healthy environment.

The children in Nairobi who no longer suffer from unexplained respiratory illness, the fishing communities in Tanzania who receive advance warning of dangerous weather, and the residents across Africa who now have evidence to support their environmental concerns represent the true measure of AI's potential to serve human rights and environmental justice.

Resources:

sensors.AFRICA website | StormWatch Platform | Seasensors Platform

Mapping the AI Lifecycle HRIA Framework for sensors.AFRICA

1 Stage 1: Objective and Team Composition

The initiative began with clearly defined objectives rooted in community needs: providing reliable, open-access environmental data to support journalists, citizens, and policymakers in addressing environmental injustices. The team composition reflects this community-centered approach, including environmental scientists, hardware engineers, technologists, community champions, local organizations, and affected residents as core stakeholders rather than peripheral consultees.

HRIA Framework Alignment:

- **Purpose & Context:** The system addresses documented discrimination in environmental governance, where marginalized communities lack evidence to support their concerns about pollution and climate impacts.
- **Effects of the system:** Benefits are explicitly designed to empower historically marginalized communities, particularly those in informal settlements and rural areas, by providing them with data ownership and advocacy tools.
- **Empowering affected communities:** Community members serve as sensor hosts, data interpreters, and advocates, with genuine decision-making power in sensor placement and use of findings.
- **Team composition:** The team includes diverse expertise (technical, social, environmental) and meaningful representation from affected communities throughout the process.

Key Human Rights Considerations

The initiative explicitly addresses environmental justice as a human rights issue, recognizing that access to environmental information is fundamental to dignity, health, and democratic participation. Team composition ensures that those most affected by environmental harms have agency in data collection and use.

2 Stage 2: Defining System Requirements

System requirements emerged from participatory mapping sessions and community dialogues rather than top-down technical specifications. Requirements include real-time monitoring capabilities, offline functionality for areas with limited connectivity, solar power options for off-grid locations,and various data access methods (APIs, visual displays, reports) and protection for sensor host anonymity.

HRIA Framework Alignment:

- **Involving affected communities:** Requirements definition involved extensive community consultation, with separate sessions for different demographic groups to ensure all voices were heard.
- **Explainability considerations:** The system provides explanations through multiple formats—visual air quality indices, written reports, and community presentations—tailored to different audiences and literacy levels.
- **Ecosystem of values:** The initiative balances technical accuracy with accessibility, privacy protection, transparency, and community agency, making conscious trade-offs that prioritize human rights over purely technical optimization.

Key Human Rights Considerations

Requirements prioritize dignity and agency for affected communities. Features like anonymization for sensor hosts, offline capabilities for marginalized areas, and multiple access methods ensure that system design serves justice rather than creating new barriers.

3 Stage 3: Data Discovery

Data discovery combines technical sensor measurements with community knowledge through participatory mapping, focus group discussions, and key informant interviews. The process involves communities in identifying what data to collect, where to collect it, and how to interpret findings. Multiple data sources include ground sensors, satellite imagery, drone data, meteorological information, and qualitative community insights.

HRIA Framework Alignment:

- **Data origin:** Data collection respects community consent and privacy, with clear agreements about data use and ownership. Sensitive information is anonymized to protect sensor hosts.
- **Data bias:** The participatory approach explicitly addresses historical bias in environmental monitoring by including communities and geographic areas typically excluded from official data collection.
- **Documentation:** All data sources, collection methods, and processing steps are documented transparently, with findings shared back to communities in accessible formats.

Key Human Rights Considerations

The data discovery process treats community knowledge as equally valid to technical measurements. Participatory mapping ensures that communities define pollution sources and priorities rather than having external definitions imposed. This approach addresses historical injustices in environmental data collection.

4 Stage 4: Selecting and Developing a Model

AI models are developed to serve community-identified needs: predicting pollution events for early warning systems, filling data gaps to maintain evidence continuity, and making complex data accessible through visualization and communication tools. Model selection prioritizes interpretability and community utility over technical sophistication.

HRIA Framework Alignment:

- **Model type and explainability:** Models prioritize explainability appropriate to community contexts, with visual outputs and clear communication about uncertainty and limitations.
- **Fairness aspects:** The initiative explicitly considers how environmental impacts affect different community groups (women, children, elderly) and ensures that AI models account for these differential impacts.
- **Environmental impact:** Solar-powered sensors and locally sourced components minimize the environmental footprint of the monitoring system itself.

Key Human Rights Considerations: Model development serves community empowerment rather than technical optimization. AI enhances rather than replaces community knowledge, providing tools for advocacy and self-determination rather than external control.

5 Stage 5: Testing and Interpreting Outcome

Testing involves both technical validation and community feedback. Communities evaluate whether the system meets their needs, provides useful information, and supports their advocacy goals. Outcomes are interpreted collaboratively, with community members trained to understand and communicate findings. Success is measured by community empowerment and environmental improvements rather than purely technical metrics.

HRIA Framework Alignment:

- **Testing Context and Outcomes:** Testing occurs in real community contexts with actual users, incorporating feedback from diverse community members about system utility and accessibility.
- **Operation Manual:** Training materials are developed in local languages with visual aids, and community members are trained to operate and interpret the system independently.

Key Human Rights Considerations

Testing evaluates whether the system genuinely empowers communities to advocate for their environmental rights. Community feedback shapes system refinements, ensuring that technical performance serves human dignity and agency.

6 Stage 6: Deployment & Post-Deployment Monitoring

Deployment involves comprehensive community training, ongoing support for sensor hosts, and continuous adaptation based on community feedback. The initiative includes safety protocols for sensor hosts, regular check-ins, and multiple channels for community input. Long-term sustainability is built through local capacity development and peer learning networks.

HRIA Framework Alignment:

- **Deployment:** Communities have genuine agency to delay or modify deployment based on their assessment of benefits and risks. Deployment includes robust support systems and safety measures for participants.
- **Monitoring:** Continuous monitoring includes both technical performance and community impact, with mechanisms for communities to report concerns or suggest improvements. Success is measured by community empowerment and environmental justice outcomes.

Key Human Rights Considerations

Post-deployment monitoring ensures that the system continues to serve community needs rather than becoming extractive. Regular community feedback loops maintain community ownership and adapt the system to changing needs and contexts.

Integrated Analysis: Human Rights Throughout the AI Lifecycle

The sensors.AFRICA case study demonstrates how human rights considerations can be integrated throughout the AI lifecycle rather than added as an afterthought. Several key principles emerge:

- **Community agency:** At every stage, affected communities have genuine decision-making power rather than tokenistic consultation. This agency extends from initial problem definition through ongoing system adaptation.

- **Justice-oriented design:** Technical choices consistently prioritize community empowerment and environmental justice over technical optimization or efficiency metrics.
Participatory Knowledge Creation: The initiative treats community knowledge as equally valid to technical expertise, creating collaborative knowledge production rather than extractive data collection.

- **Adaptive implementation:** System design and implementation adapt continuously based on community feedback, ensuring that the AI serves evolving community needs rather than static technical specifications.

- **Sustainability through ownership:** Long-term sustainability is built through community ownership and capacity development rather than external dependency. The sensors.AFRICA experience demonstrates that AI can serve human rights and environmental justice when developed with genuine community participation throughout the lifecycle. This approach results in more robust, sustainable, and effective systems that empower rather than marginalize affected communities.

About the case study and author

This case study analyzes research conducted by sensors.AFRICA, incubated by Code for Africa, examining environmental pollution across African cities and communities between 2016-2025.

Alicia Olago is an environmental scientist and seasoned researcher with over a decade of experience in sustainable development projects in Eastern Africa and is currently CfA's Senior Product Manager at sensors.Africa. She leads a team of Hardware Engineers and Technologists in a citizen science initiative, utilizing sensors to monitor air, water & sound pollution, and radiation among other environmental hazards, to provide citizens & civic watchdogs actionable information on their cities & communities in the continent.

Other contributors to this case study are Caitlin Kraft-Buchman, Emma Kallina, and Sofia Kypraiou, authors of the original *Framework to AI Development: Integrating Human Rights Considerations Along the AI Lifecycle* upon which the Toolbox structure is based. Additional contributors are Amina Soulimani and Pilar Grant, from Women at the Table and the <AI & Equality> Human Rights Initiative.

<AI & Equality> African Toolbox | Case study

Bridging Language Barriers: AI for Kenyan Sign Language and Digital Inclusion

Watch the video

This case study is part of the **African <AI & Equality> Toolbox**, which builds upon the methodology of the global <AI & Equality> Human Rights Toolbox—an initiative of Women At The Table in collaboration with the United Nations Office of the High Commissioner for Human Rights (OHCHR). The African Toolbox is a collaboration between the <AI & Equality> initiative and the African Centre for Technology Studies (ACTS). To learn more visit **aiequalitytoolbox.com**

The challenge: When silence becomes exclusion

In a bustling university classroom in Nairobi, Sarah sits in the front row, her eyes fixed on the professor's lips as he delivers a complex computer science lecture. Her learning partner, James, frantically scribbles notes, knowing that Sarah will depend on his interpretation of concepts she cannot hear. When the professor turns to write on the board, Sarah loses all connection to the lesson. When students laugh at a joke, she wonders what she's missing. When the professor asks a question, she cannot respond in her native language—Kenyan Sign Language (KSL)—because no one in the room can understand her.

This scene plays out daily across Kenya's higher education institutions. While the country has established special schools for deaf students from primary through secondary levels, these young people face a brutal transition when they enter universities and colleges. Suddenly, the carefully constructed support systems vanish, leaving them isolated in hearing-dominant environments with no sign language interpreters, no accessible materials, and often, no understanding of their communication needs.

Sarah's story reflects a broader challenge: language barriers that exclude entire communities from educational opportunities and social participation. In Kenya, over one million people experience hearing impairments, yet their language—KSL—remains largely invisible in the digital age. This invisibility perpetuates cycles of exclusion that begin in hearing families where 90% of deaf children are born, continue through educational systems that abandon them at higher levels, and extend into workplaces that cannot accommodate their communication needs.

The genesis: From personal encounter to community-driven innovation

Dr. Lilian Wanzare's journey into sign language AI began not with a research proposal, but with a moment of recognition in her own classroom. As a computational linguist at Maseno University, she was confronted with deaf students who had been "dropped" into her computer science courses with no support system beyond learning partners—hearing students who would take notes and attempt to interpret complex technical concepts.

"I realized I was supposed to teach them computer science, but I had no clue how to handle a deaf student," Dr. Wanzare recalls. "We were failing them systematically. They were forced to learn with their hearing counterparts, but we had no way to include them actively in class."

This personal encounter revealed a fundamental injustice: educational systems that provided specialized support through secondary school but abandoned students at the crucial transition to higher education. The problem wasn't just about individual accommodation—it was about the systematic exclusion of an entire linguistic community from digital innovation and technological advancement.

What began as a search for classroom solutions evolved into a recognition that the challenge was much deeper. The issue wasn't just about interpreters or note-takers; it was about the complete absence of Kenyan Sign Language from the digital ecosystem. While other languages could access spell-checkers, translation tools, and digital content, KSL remained locked out of the technological revolution.

Building trust through community-centered design

Entry Through Educational Champions

Unlike many AI projects that begin with technical possibilities, the KSL initiative started with educational reality. Dr. Wanzare's team began by identifying what they call "educational champions"—sign language teachers, deaf students, and deaf community leaders who could bridge the gap between technical development and lived experience.

The Kenya Institute of Curriculum Development (KICD) had already established a standardized KSL curriculum for schools, providing a crucial foundation. But the real champions were the teachers and students in specialized schools across Kenya—from primary schools for the deaf to secondary institutions like Kaimosi School for the Deaf and Meru School for the Deaf.

These champions didn't just provide access to the deaf community; they fundamentally shaped the project's understanding of what needed to be built. Teachers explained that their students didn't just need translation tools—they needed technology that could help them participate fully in hearing-dominated environments. Students expressed frustration not just with communication barriers, but with the broader invisibility of their language and culture.

Participatory Language Mapping

The breakthrough came through what the team calls "participatory language mapping"—sessions where deaf community members, teachers, and students collaborated to identify the specific vocabulary, contexts, and communication needs that AI systems would need to address.

These sessions revealed nuanced insights that technical experts would never have discovered independently. Deaf students explained that they didn't just need word-by-word translation; they needed systems that could capture the grammatical structure of KSL, which places objects before subjects and uses facial expressions as integral parts of meaning. Teachers highlighted that different regions had signing variations, even within the standardized curriculum. Community members emphasized that effective sign language includes not just hand movements but facial expressions, body posture, and spatial relationships. "When you see a deaf student explaining how a sign changes meaning based on the speed of movement or the direction of the palm, you understand that building AI for sign language isn't just about recognizing gestures—it's about understanding an entire linguistic system," Dr. Wanzare explains.

Redefining the Technical Challenge

The participatory approach revealed that the "sign language problem" was actually multiple interconnected challenges:

- **Educational access:** Deaf students needed ways to participate in hearing-dominated classrooms.
- **Communication barriers:** Families and communities needed tools to communicate with deaf members.
- **Cultural preservation:** KSL needed digital representation to prevent language loss.
- **Economic inclusion:** Deaf individuals needed access to digital technologies for employment.
- **Social participation:** Deaf community members needed ways to engage with broader society.

This comprehensive understanding shaped the technical approach, ensuring that AI development would address systemic exclusion rather than just individual accommodation.

Technical innovation driven by linguistic justice

The text-to-avatar pipeline

The technical architecture of the KSL translation system reflects both the linguistic complexity of sign language and the realities of resource-constrained environments. The system operates through a three-stage pipeline that transforms spoken English into animated sign language through an avatar representation.

- **Stage 1 -Text to Gloss Translation:** The system first converts English text into "gloss"—a linguistic representation that captures how concepts are structured in KSL. This isn't simple word-for-word translation; it involves understanding that KSL grammar places objects first, then subjects, then verbs. So "A bee stings" becomes "bee sting" in gloss representation.
- **Stage 2 - Pose Extraction and Representation:** The system then converts the gloss into pose representations—mathematical descriptions of hand movements, facial expressions, and body positions that capture the essential elements of signs. This step abstracts away from individual signers while preserving the linguistic content of signs.

- **Stage 3 - Avatar Animation:** Finally, the system uses the pose representations to animate a virtual avatar that performs the signs. This avatar isn't just a technical convenience—it's designed to be culturally appropriate, customizable, and accessible to users with different preferences and needs.

Community-driven technical requirements

Every technical decision emerged from community input rather than engineering convenience. The development team learned that effective sign language AI required attention to details that might seem trivial to hearing people but were fundamental to the deaf community:

- **Facial Expression Integration:** Signs aren't just hand movements—facial expressions are grammatically significant. The system had to capture and reproduce subtle facial movements that change meaning.
- **Regional Variation Support:** While Kenya has standardized KSL, regional dialects and individual variations exist. The system needed to accommodate these differences while maintaining comprehensibility.
- **Cultural Authenticity:** The avatar's appearance, clothing, and behavior needed to reflect Kenyan culture and be acceptable to the deaf community. Early feedback rejected avatars that looked "too Western" or wore inappropriate clothing.
- **Speed and Rhythm Control:** Users wanted the ability to slow down or repeat signs, reflecting how they actually learn and process sign language.

Addressing Infrastructure Challenges

The technical solution had to work within Kenya's technological constraints while serving users who might have limited access to high-end devices:

- **Offline Capability:** The system needed to function without constant internet connectivity, crucial for users in rural areas or those with limited data plans.
- **Mobile Optimization:** The avatar animation had to run efficiently on smartphones and tablets, the most accessible computing devices for many deaf users.
- **Low-Latency Processing:** Real-time translation required processing speeds that would enable natural conversation, not just delayed interpretation.
- **Scalable Architecture:** The system needed to handle multiple concurrent users while maintaining performance quality.

Data as community language asset

The Ethical Challenge of Sign Language Data

Collecting data for sign language AI presented unique ethical challenges that text-based language models never face. Every sign language data point involves a human face, body, and personal expression. Many potential contributors were minors in specialized schools. The deaf community had legitimate concerns about privacy, consent, and the potential misuse of their linguistic data.

Dr. Wanzare's team approached data collection as a process of community partnership rather than extraction. "We had to convince them that there was no point in time they would find their video online somewhere that had been posted, that nothing would be leaked, and that they could trust us in maintaining their privacy and security," she explains.

Participatory Data Collection

The data collection process involved multiple stakeholders across Kenya's deaf education system:

- **Primary and Secondary Schools:** Teams visited specialized schools for the deaf across the country, from primary through secondary levels, ensuring representation across age groups and educational stages.
- **Diverse Signer Representation:** The dataset includes first-language signers (deaf from birth), second-language signers (those who became deaf later), expert signers (teachers and community leaders), and novice signers (students still learning).
- **Regional Coverage:** Boarding schools provided access to students from across Kenya, including neighboring countries, ensuring the dataset captured regional variations and dialects.
- **Gender and Cultural Balance:** The team ensured balanced representation across gender lines, noting that female signers tended to be more facially expressive while male signers showed different patterns.

Anonymization and Privacy Protection

The team developed innovative approaches to protect signer privacy while preserving linguistic data:

- **Pose Extraction Technology:** Rather than storing raw videos, the system extracts "landmarks"—mathematical representations of hand positions, facial movements, and body postures that capture signs without revealing individual identity.
- **Community Consent Processes:** Data collection involved not just individual consent but community-level agreements with schools, parents, and deaf community organizations.
- **Controlled Access:** The dataset is not publicly released but made available to

researchers through controlled access that protects community interests.
- **Benefit Sharing:** The community retains rights to the data and receives regular updates on how it's being used and what benefits are being generated.

AI as a tool for linguistic justice

Beyond Translation: Addressing Systemic Exclusion

The KSL avatar system represents more than technological innovation—it embodies a commitment to linguistic justice. The AI serves multiple functions that directly address the exclusion of deaf communities from digital society:

- **Educational Inclusion:** The avatar enables deaf students to access educational content in their native language, potentially transforming their learning experience in hearing-dominated institutions.
- **Family Communication:** The system provides families with tools to communicate with deaf members, addressing the isolation that often begins in the home.
- **Cultural Preservation:** By digitizing KSL, the system helps preserve and promote a language that risks being lost in an increasingly digital world.
- **Economic Empowerment:** Access to digital communication tools can improve employment prospects for deaf individuals by enabling them to participate in digital workplaces.

The Community-AI Partnership Model

The system explicitly positions AI as a tool for community empowerment rather than replacement of human communication. Community feedback shaped every aspect of the avatar's design and behavior:

- **Customization Options:** Users can choose the avatar's gender, appearance, and clothing to match their preferences and cultural context.
- **Linguistic Authenticity:** The avatar's signing style reflects authentic KSL rather than simplified or artificial gestures.
- **Educational Integration:** The system is designed to support rather than replace sign language education, helping teachers and students in their learning processes.
- **Community Ownership:** The deaf community retains control over how their linguistic data is used and how the technology evolves.

Addressing the "Replacement" Concern

Some community members worried that AI would replace human interpreters or reduce the value of sign language skills. The team addressed this through transparency about the technology's limitations and explicit positioning as a supportive tool:

- **Complementary Function:** The avatar is designed to supplement rather than replace human communication, particularly in contexts where interpreters aren't available.

- **Educational Tool:** The system serves as a learning aid for both deaf students and hearing individuals who want to learn KSL.
- **Advocacy Platform:** The technology raises awareness about KSL and the deaf community, potentially increasing demand for human interpreters and services.
- **Skill Development:** The system can help deaf individuals develop literacy skills by providing visual representation of text concepts.

Language and cultural authenticity

Capturing the Complexity of Sign Language

Sign language AI faces challenges that spoken language systems never encounter. Signs involve five simultaneous components: hand shape, palm orientation, hand location, movement direction, and movement speed. Each component affects meaning, and all must be captured accurately for effective communication.

The team learned that cultural authenticity required attention to details that might seem peripheral to technical developers but were fundamental to the deaf community:

- **Facial Expression Integration:** Non-manual features like eyebrow movement, lip patterns, and head position are grammatically significant in KSL and needed to be accurately represented.
- **Spatial Relationships**: Sign language uses space to show relationships between concepts, requiring the avatar to maintain spatial consistency across signs.
- **Rhythm and Timing:** The speed and rhythm of signing affects meaning and comprehensibility, requiring fine-tuned control systems.
- **Cultural Appropriateness:** The avatar's appearance, clothing, and behavior needed to reflect Kenyan culture and be acceptable to the deaf community.

Community-Driven Refinement

The development process involved continuous feedback from the deaf community, resulting in multiple refinements:

- **Avatar Appearance:** Early versions were rejected because the avatar didn't look sufficiently Kenyan. The team had to find more culturally appropriate representations.
- **Signing Speed:** While deaf signers naturally sign very quickly, they wanted the avatar to sign more slowly so they could analyze and learn from individual signs.
- **Facial Expression Enhancement:** The community requested more realistic facial expressions and lip movements to capture the full linguistic content of signs.
- **Personalization Options:** Users wanted the ability to customize the avatar's gender, appearance, and clothing to match their preferences.
- **Repetition Control:** The system needed to allow users to request repetition of signs, reflecting how people actually learn and process sign language.

Expanding impact: From education to community empowerment

Multi-Language Integration

Building on the success of English-to-KSL translation, the team expanded the system to include other languages important to the Kenyan deaf community:

- **Kiswahili Integration:** The system now translates from Kiswahili to KSL, enabling deaf individuals to access content in Kenya's national language.
- **Local Language Support:** Plans include expanding to other Kenyan languages, allowing deaf individuals to learn and communicate in multiple linguistic contexts.
- **Cross-Cultural Communication:** The system can potentially bridge communication gaps between deaf individuals and hearing people who speak different languages.

Community Health and Legal Applications

The technology's potential extends far beyond educational settings:

- **Healthcare Communication:** The avatar could help deaf patients communicate with healthcare providers who don't know sign language.
- **Legal Interpretation:** Court proceedings could become more accessible to deaf individuals through avatar interpretation.
- **Emergency Services:** The system could provide critical communication tools for emergency situations.
- **Workplace Integration:** Employers could use the system to communicate with deaf employees, improving workplace inclusion.

Community Capacity Building

The initiative has sparked broader conversations about deaf rights and inclusion:

- **Awareness Raising:** The project has increased visibility of KSL and the deaf community in Kenya's tech sector.
- **Policy Advocacy:** The research provides evidence for policy changes that could improve deaf inclusion in education and employment.
- **Community Organizing:** The project has strengthened connections within Kenya's deaf community and provided platforms for collective advocacy.
- **International Collaboration:** The work has inspired similar initiatives in other African countries working to digitize their sign languages.

Community-based data licensing: A new model

The Challenge of Digital Exploitation

The KSL project confronted a fundamental question: who owns linguistic data, and how should communities benefit from AI systems built on their languages? Traditional open-source licensing models assume that making data freely available benefits everyone, but this approach can lead to exploitation of marginalized communities.

Dr. Wanzare explains the dilemma: “If today we collect 200 hours of sign language data and put it online, by Friday Meta will have it integrated into their systems. The community asks: what’s there for the local ecosystem? Is that competition too unfair for us to even begin competing?”

Community-Controlled Innovation

Working with Mozilla, the team is pioneering community-based licensing models that give linguistic communities control over how their data is used:

- **Community Ownership:** The deaf community retains ownership of their linguistic data and has a say in how it’s licensed and used.
- **Benefit Sharing:** Commercial applications built on community data would need to provide benefits back to the community, potentially including royalties or revenue sharing.
- **Use Restrictions:** The community can specify how their data can and cannot be used, protecting against exploitation or misrepresentation.
- **Local Ecosystem Development:** The licensing model prioritizes local developers and applications that directly benefit the community.

Small Language Models for Community Control

The team advocates for small, specialized language models rather than integration into large corporate systems:

- **Community-Specific Applications:** Small models can be designed for specific use cases that directly benefit the deaf community, such as educational tools or healthcare communication.
- **Local Control:** Communities can maintain control over smaller models in ways that are impossible with large corporate systems.
- **Cultural Sensitivity:** Small models can be fine-tuned to reflect community values and preferences without being overwhelmed by broader dataset biases.
- **Sustainable Development:** Local developers can maintain and improve small models in ways that serve community needs rather than corporate profits.

Lessons for digital inclusion

Key Principles for Language Justice

The KSL initiative offers crucial insights for AI development that promotes linguistic justice:

- **Community Partnership is Essential:** The most sophisticated AI system fails if communities don't trust it, understand it, or control its use. Starting with community needs rather than technical capabilities ensures that AI serves justice rather than perpetuating existing exclusions.
- **Participation Must Be Genuine:** Tokenistic consultation differs fundamentally from the deep engagement required for effective language AI. Community involvement in problem definition, solution design, and implementation creates genuine ownership that sustains initiatives over time.
- **Cultural Authenticity Matters:** Every technical choice—from avatar appearance to signing speed—embodies assumptions about users and culture. Designing for African contexts requires fundamental rethinking of standard approaches, resulting in more authentic and acceptable systems.
- **Data Sovereignty is Fundamental:** Linguistic data becomes a tool for justice only when communities control its collection, use, and benefits. Community data sovereignty ensures that AI serves empowerment rather than extraction.
- **AI Can Preserve and Promote Languages:** Rather than contributing to language loss, AI can become a tool for language preservation and promotion when developed with community control and cultural sensitivity.

Addressing Responsible AI Throughout Development

The KSL experience demonstrates how human rights principles can be integrated throughout AI development:

- **Privacy and Security:** Protecting signer privacy while preserving linguistic data requires innovative technical approaches and community consent processes.
- **Fairness and Representation:** Ensuring the dataset represents the full diversity of the deaf community requires intentional inclusion of different genders, ages, regions, and signing abilities.
- **Transparency and Interpretability:** Community members need to understand how the system works and why it makes specific decisions, requiring explainable AI approaches.
- **Reliability and Safety:** The system must work consistently and safely, particularly in educational and healthcare contexts where errors could have serious consequences.

Future directions: Scaling linguistic justice

Cross-Border Expansion

The success of the KSL initiative has inspired similar projects across Africa:

- **Regional Collaboration:** Other African countries are adapting the methodology to develop AI systems for their own sign languages.
- **Comparative Research:** Cross-country studies are examining how different sign languages can benefit from shared technical approaches while maintaining linguistic authenticity.
- **Continental Networks:** The project is contributing to broader networks of African language technologists working to digitize indigenous languages.

Integration with Broader Language Justice

The KSL work is part of a broader movement to ensure African languages are included in the digital age:

- **Multi-Modal Systems:** Future systems will integrate sign language with spoken language AI, creating more comprehensive communication tools.
- **Educational Integration:** The technology is being integrated into educational curricula to support both deaf students and hearing students learning KSL.
- **Policy Advocacy:** The research provides evidence for policy changes that could improve digital inclusion for all marginalized linguistic communities.

Mapping the AI Lifecycle HRIA Framework for the Kenyan Sign Language Initiative

1 Stage 1: Objective and Team Composition

Objective Definition: The initiative began with a community-identified problem: deaf students' systematic exclusion from higher education due to lack of sign language interpretation services. The objective evolved through community engagement to address not just individual accommodation but the broader digital exclusion of Kenyan Sign Language and the deaf community.

Team Composition: The team intentionally included diverse expertise and lived experience:

- Computational linguists (Dr. Lilian Wanzare and research team)
- Sign language experts and teachers from specialized schools
- Deaf community members as co-designers and validators
- Educational specialists familiar with inclusive pedagogy
- Technology experts in computer vision and avatar animation
- Students and families from the deaf community
- Community leaders and advocates for deaf rights

HRIA Framework Alignment:

- **Purpose & Context of the System:** The system addresses documented discrimination in educational access, where deaf students face systematic exclusion from higher education. The domain has a clear history of linguistic discrimination, with KSL being marginalized in favor of spoken languages.
- **Effects of the System:** Benefits explicitly designed to empower the deaf community—historically marginalized in educational and digital spaces—by providing access to technology in their native language and creating pathways for broader social participation.
- **Empowering Affected Communities:** Deaf community members serve as data contributors, system validators, co-designers, and advocates, with genuine decision-making power in system design and implementation.
- **Team Composition:** Diverse expertise spanning technical, linguistic, educational, and cultural domains, with meaningful representation from the deaf community throughout the process.

Key Human Rights Considerations:

The initiative explicitly addresses linguistic rights as human rights, recognizing that access to communication technology in one's native language is fundamental to dignity and participation. Team composition ensures that those most affected by digital exclusion have agency in system development.

2 Stage 2: Defining System Requirements

Community-Driven Requirements: System requirements emerged from participatory sessions with deaf students, teachers, and community members rather than technical specifications. Requirements included:

- Avatar animation that captures facial expressions as grammatically significant elements
- Customizable avatar appearance to reflect Kenyan cultural context
- Variable signing speed with repetition capability
- Multi-language support (English, Kiswahili, potential local languages)
- Offline functionality for areas with limited connectivity
- Educational integration features for classroom use

Cultural Authenticity: Requirements prioritized cultural authenticity and community acceptance over technical optimization, ensuring the avatar would be embraced by the deaf community.

HRIA Framework Alignment:

- **Involving Affected Communities:** Requirements definition involved extensive consultation with deaf students, teachers, families, and community leaders through schools for the deaf across Kenya.
- **Explainability Considerations:** The system provides explanations about how signs are constructed and why specific movements create meaning, supporting both learning and transparency.
- **Ecosystem of Values:** The initiative balances technical accuracy with cultural authenticity, privacy protection, community agency, and educational utility, making conscious trade-offs that prioritize community acceptance.

Key Human Rights Considerations:

Requirements prioritize dignity and cultural authenticity for the deaf community. Features like customizable avatar appearance, culturally appropriate signing, and community control over data use ensure that system design serves linguistic justice rather than perpetuating cultural imperialism.

3 Stage 3: Data Discovery

Community-Partnered Data Collection: The team created a comprehensive dataset through ethical partnerships with deaf schools across Kenya. Data collection involved:

- Visits to specialized schools from primary through secondary levels
- Recording diverse signers across age groups, genders, and regions
- Both scripted and spontaneous signing to capture natural language use
- Expert glossing and linguistic annotation by sign language teachers
- Rigorous segmentation marking the beginning and end of each sign in sentences

Privacy-Preserving Innovation: The team developed pose extraction technology that abstracts linguistic content from personal identity, enabling data sharing while protecting individual privacy.

HRIA Framework Alignment:

- **Data Origin:** Data collection involved comprehensive consent processes with individuals, families, schools, and community organizations. The focus on pose extraction rather than raw video protects privacy while enabling linguistic research.
- **Data Bias:** The participatory approach explicitly addresses historical bias by including diverse signers across regions, genders, ages, and skill levels, ensuring representation of the full deaf community.
- **Documentation:** All data sources, collection methods, and processing steps are documented transparently, with regular reports shared back to contributing communities.

Key Human Rights Considerations:

The data discovery process treats KSL as a complete language system worthy of preservation and promotion. Communities define signing standards and participate in data validation rather than having external definitions imposed. This approach addresses historical marginalization of sign languages.

4 Stage 4: Selecting and Developing a Model

Community-Informed Model Architecture: The three-stage pipeline (text-to-gloss, pose extraction, avatar animation) was designed to serve community-identified needs rather than demonstrate technical sophistication. Model selection prioritized:

- Cultural authenticity in avatar representation
- Linguistic accuracy in sign production
- Educational utility for both deaf and hearing users
- Privacy protection through pose abstraction
- Accessibility across different technological contexts

Iterative Community Validation: Each stage of model development involved community feedback, with deaf signers validating the accuracy and cultural appropriateness of system outputs.

HRIA Framework Alignment:

- **Model Type and Explainability:** The staged pipeline prioritizes interpretability, allowing users to understand how English text becomes sign language and enabling community validation at each step.
- **Fairness Aspects:** The initiative explicitly considers how the system performs across different demographic groups within the deaf community, ensuring equitable representation and accuracy.

- **Transparency:** Model development processes are transparent to the community, with regular demonstrations and feedback sessions that allow community input into technical decisions.

Key Human Rights Considerations:
Model development serves community empowerment and linguistic preservation rather than technical optimization. The AI enhances rather than replaces human communication, providing tools for linguistic justice and cultural preservation.

5 Stage 5: Testing and Interpreting Outcome

Community-Centered Evaluation: Testing involved extensive community feedback in real educational and social settings. Deaf community members evaluated:
- Avatar appearance and cultural appropriateness
- Signing accuracy and linguistic authenticity
- Educational utility and integration potential
- Privacy protection and data security
- Customization options and user control

Iterative Refinement: Community feedback directly shaped system improvements, from avatar appearance to signing speed to facial expression integration.

HRIA Framework Alignment:
- **Testing Context and Outcomes:** Testing occurs in real community contexts with actual users, incorporating feedback from diverse community members about system utility, accuracy, and cultural appropriateness.
- **Operation Manual:** Training materials and user guides are developed in collaboration with deaf educators and community leaders, ensuring accessibility and cultural sensitivity.

Key Human Rights Considerations:
Testing evaluates whether the system genuinely empowers the deaf community to participate in digital society. Community feedback shapes system evolution, ensuring that technical performance serves linguistic justice and cultural preservation.

6 Stage 6: Deployment & Post-Deployment Monitoring

Community-Controlled Deployment: Deployment involves comprehensive community partnerships, ongoing support for users, and continuous adaptation based on community feedback. The initiative includes:
- Integration with educational institutions and community organizations
- Training programs for educators and community leaders

- Ongoing technical support and system maintenance
- Community-based licensing models that protect community interests
- Expansion planning to other languages and applications

Sustainable Community Ownership: Long-term sustainability built through community ownership models, local capacity development, and innovative licensing arrangements that ensure community benefits.

HRIA Framework Alignment:

- **Deployment:** The deaf community has genuine agency in deployment decisions, with robust support systems and community-controlled licensing that protects their interests.
- **Monitoring:** Continuous monitoring includes both technical performance and community impact, with mechanisms for the deaf community to provide feedback and guide system evolution.

Key Human Rights Considerations:

Post-deployment monitoring ensures that the system continues to serve community needs and linguistic justice rather than becoming extractive. Community-based licensing models maintain community ownership and ensure that benefits flow back to the deaf community.

Integrated Analysis: Human Rights Throughout the AI Lifecycle

The KSL initiative demonstrates how human rights considerations can transform language AI development. Several key principles emerge:

- **Community agency:** At every stage, the deaf community has genuine decision-making power rather than tokenistic consultation. This agency extends from initial problem definition through ongoing system evolution.

- **Cultural Authenticity:** Technical choices consistently prioritize cultural authenticity and community acceptance over technical optimization or efficiency metrics.

- **Linguistic Preservation:** The initiative treats KSL as a complete language system worthy of preservation and promotion, creating technology that strengthens rather than marginalizes the language.

- **Innovative Privacy Protection:** The pose extraction approach demonstrates how technical innovation can protect individual privacy while enabling linguistic research and community empowerment.

- **Sustainable Community Ownership:** Community-based licensing models ensure that the deaf community retains control over their linguistic data and benefits from AI systems built on their language.

The KSL experience demonstrates that language AI can serve human rights and linguistic justice when developed with genuine community participation throughout the lifecycle. This approach results in more culturally authentic, community-controlled, and sustainable systems that empower rather than marginalize linguistic communities, creating technology that truly serves the right to language and cultural expression.

About the case study and author

This case study analyzes research conducted by Dr. Lilian Wanzare, Dr. Joel Okutoyi, Dr. Mildred Ayere and Dr. Maureen Kang'ahi of Maseno University, examining Kenyan Sign Language across Kenya, HomaBay, Siaya, Kisumu, Kakamega and Vihiga counties in Western Kenya, between 2023 - 2024.

This research was supported by EduAI Hub at the University of Lagos as part of a project under AI4D Africa. AI4D is a collaborative initiative by the International Development Research Centre (IDRC), Canada, and the Swedish International Development Cooperation Agency (SIDA), Sweden.

Dr. Lilian Wanzare is a lecturer and chair of the Department of Computer Science at Maseno University. Her research interests are in Artificial Intelligence and Machine Learning, in particular Natural Language Processing (NLP), Sign Language research and building text processing tools for low-resource languages. She holds a PhD degree in Computational Linguistics and an Msc. in Language Science and Technology from Saarland University, Germany.

Other contributors to this case study are Caitlin Kraft-Buchman, Emma Kallina, and Sofia Kypraiou, authors of the original *Framework to AI Development: Integrating Human Rights Considerations Along the AI Lifecycle* upon which the Toolbox structure is based. Additional contributors are Amina Soulimani and Pilar Grant, from Women at the Table and the <AI & Equality> Human Rights Initiative.

<AI & Equality> African Toolbox | Case study

AI-Powered Malaria Diagnostics: Makerere AI Health Lab Initiative

Watch the video

This case study is part of the **African <AI & Equality> Toolbox**, which builds upon the methodology of the global <AI & Equality> Human Rights Toolbox—an initiative of Women At The Table in collaboration with the United Nations Office of the High Commissioner for Human Rights (OHCHR). The African Toolbox is a collaboration between the <AI & Equality> initiative and the African Centre for Technology Studies (ACTS). To learn more visit **aiequalitytoolbox.com**

92

The challenge: Bridging the diagnostic gap

Malaria remains a critical health challenge in sub-Saharan Africa, where traditional diagnostic methods face significant barriers. The gold standard for malaria diagnosis—microscopic examination of blood smears—requires trained technicians and well-equipped laboratories, resources that are often scarce in the regions where malaria hits hardest.

The Makerere AI Health Lab, led by Dr. Rose Nakasi, confronted this reality directly. In many rural areas across Uganda, the shortage of skilled technicians meant that communities faced delays in diagnosis, subjective interpretations prone to human error, and limited accessibility to diagnostic services. Time constraints in overburdened facilities further compromised patient outcomes.

Rather than accepting these limitations as inevitable, the team recognized an opportunity to leverage AI technology to democratize access to accurate malaria diagnosis—but only if developed with careful attention to the communities it would serve.

A community-centered approach to technical innovation

Defining the Real Problem

The Makerere initiative began with a crucial insight: the challenge wasn't simply technical—it was about health equity and access. The team's objective extended beyond "developing AI for malaria diagnosis" to specifically addressing the unique constraints of resource-limited rural areas in Uganda. This community-driven approach shaped every subsequent decision.

The team composition reflected this understanding from the outset. Rather than working in isolation, they assembled a multidisciplinary group that included not just AI experts and researchers, but local medical practitioners deeply familiar with the context. Local health facilities became integral partners in defining needs and shaping solutions, not merely end users of a predetermined technology.

The technical innovation: Making AI accessible

The centerpiece of the solution—a 3D-printed smartphone adapter that connects to standard microscope eyepieces—exemplifies how technical design can embody values of accessibility and sustainability. This innovation addressed the lack of expensive, dedicated imaging equipment in resource-limited settings by transforming existing microscopes into digital imaging devices.

The choice of smartphone technology wasn't arbitrary—it leveraged devices already present in many communities while ensuring that the diagnostic capability could remain local and sustainable. The 3D-printed adapter could be manufactured locally, reducing dependence on international supply chains and enabling communities to maintain and replace equipment independently.

Confronting data challenges with ethical rigor

Building a Representative Dataset

One of the most significant challenges faced by the team was the scarcity of suitable datasets. Existing datasets were either unavailable or inadequately represented the local context and populations the system was designed to serve.

Rather than compromising on data quality or appropriating inappropriate datasets, the team initiated their own comprehensive data collection effort. This process required navigating complex ethical and bureaucratic landscapes:

- **Ethical Foundations:** The team obtained necessary ethical approvals from relevant authorities, ensuring compliance with research standards and protection of patient rights from the project's inception.
- **Partnership-Based Collection:** They established collaborative relationships with local health facilities, creating partnerships rather than extractive relationships for data collection.
- **Privacy Protection:** Strict data anonymization protocols were implemented to protect patient privacy while still enabling the medical insights necessary for effective AI development.
- **Persistence Through Bureaucracy:** The team acknowledged that "bureaucracies and unclear data policies often slow down progress," but demonstrated commitment to working within these systems while advocating for clearer, more enabling policies.

The Human Rights Impact Assessment Integration

The data collection process exemplified key principles from human rights impact assessments. The team ensured that data subjects provided informed consent, that privacy was rigorously protected, and that the sensitive nature of health information was respected throughout the process.

Their approach treated community data as an asset to be protected and shared responsibly, rather than a resource to be extracted. This philosophy shaped not only how data was collected, but how insights would be shared back with communities and how the technology would be deployed.

Technical excellence serving health equity

Model Development with Purpose

The AI model development process balanced technical sophistication with practical constraints and community needs. Key decisions reflected the team's commitment to creating tools that would genuinely serve health equity:

- **Accuracy and Speed:** The system achieved 99% accuracy in detecting malaria parasites alone, with remarkably efficient inference time of 5 seconds. This speed was crucial for practical deployment in busy clinical settings where patients cannot wait extended periods for diagnosis.
- **Transparency About Limitations:** When the system's accuracy dropped to 74% for multi-class detection including white blood cells, the team transparently acknowledged this limitation and committed to ongoing efforts to address the issue. This honesty about performance trade-offs demonstrates commitment to responsible AI development.
- **Accessibility-First Design:** The smartphone-based interface prioritized usability in resource-constrained environments, ensuring that the tool could function effectively even with limited technical infrastructure.

Addressing Algorithmic Bias

The team's attention to performance discrepancies across different detection tasks highlighted their awareness of potential bias issues. Their commitment to "ongoing efforts to address this issue" demonstrated understanding that algorithmic fairness requires continuous attention and refinement, not just initial consideration.

Real-world validation and community engagement

Testing in Authentic Contexts

The validation process emphasized real-world performance over laboratory-controlled conditions. Field testing was conducted in actual healthcare settings where the system would ultimately be deployed, directly exposing and addressing practical challenges that wouldn't emerge in controlled environments.

The team actively collected feedback from healthcare workers to ensure user-friendliness and practical application. This participatory approach ensured that the tool would meet the actual needs of its intended users rather than theoretical requirements defined by developers.

Rigorous Comparison: The AI system's diagnoses were compared with those of experienced technicians, providing robust validation of accuracy in real-world conditions while respecting the expertise of human practitioners.

Addressing the Human-AI Partnership

Throughout the development process, the team emphasized that AI serves as a support tool rather than a replacement for human expertise. This philosophy addressed concerns about job displacement while positioning AI as a means of enhancing human capabilities and extending expert-level diagnosis to underserved areas.

The approach recognized that healthcare workers bring irreplaceable knowledge, cultural understanding, and patient relationships that AI cannot replicate. The technology was designed to complement and enhance these human capabilities rather than substitute for them.

Impact and sustainable development

Immediate Outcomes

The project demonstrated measurable potential for improving healthcare delivery in resource-limited settings:

- **Diagnostic Accuracy:** Potential reduction of diagnostic errors and subjective interpretation
- **Accessibility:** Increased access to quality diagnostics in remote and under-resourced regions
- **Efficiency:** Faster diagnostic turnaround times enabling more timely treatment
- **Capacity Building:** Decision support tools that enhance rather than replace healthcare worker capabilities

Long-Term Vision

The Makerere Health Lab's plans for expansion reveal a comprehensive vision for technology-enabled health equity. Future directions include:

- **Disease Coverage Expansion:** Adapting the AI diagnostics approach to other diseases such as cervical cancer and tuberculosis, addressing multiple health challenges with similar community-centered methodology.
- **Telehealth Integration:** Exploring telehealth platforms for remote expert consultations, extending specialist knowledge to underserved areas while maintaining community ownership of diagnostic capabilities.
- **Cultural Accessibility:** Adapting tools for local languages, ensuring that linguistic barriers don't prevent communities from benefiting from diagnostic advances.

Lessons for Human Rights-Based AI Development

Key Principles Demonstrated

The Makerere initiative offers several crucial insights for AI development that promotes rather than undermines human rights:

- **Community Partnership from Inception:** The most sophisticated AI system fails if communities don't trust it, understand it, or have agency in its deployment. Starting with community needs rather than technical capabilities ensures that AI serves justice rather than perpetuating existing inequities.

- **Ethical Rigor Throughout:** Privacy protection, informed consent, and transparency aren't add-ons to technical development—they're foundational requirements that shape every aspect of system design and deployment.

- **Technical Choices Reflect Values:** Every decision—from smartphone compatibility to local manufacturing capability—embodies assumptions about users and priorities. Designing for African contexts required fundamental rethinking of standard approaches.

- **Sustainability Through Local Ownership:** Long-term success depends on communities having genuine ownership and control over the technology, not just access to it.

Mapping the AI Lifecycle HRIA Framework for the Makekere Health Lab case

1 Stage 1: Objective and Team Composition

- **Purpose & Context:** The team explicitly addressed malaria's disproportionate impact on resource-limited rural areas in Uganda, recognizing healthcare access as a justice issue rather than merely a technical challenge.
- **Community Engagement:** Local health facilities were integrated as partners from the beginning, not just end users. The team included local medical practitioners deeply familiar with the context, ensuring lived experience informed the development process.
- **Team Composition:** The multidisciplinary team combined AI experts, researchers, and crucially, local medical practitioners who understood the real-world constraints and cultural context of deployment.
- **Effects Assessment:** The objective was framed around democratizing access to expert-level diagnosis, explicitly targeting communities historically excluded from quality healthcare due to geographic and economic barriers.

2 Stage 2: Defining System Requirements

- **Technical Requirements Driven by Community Needs:** The choice of smartphone-based technology and 3D-printed adapters directly addressed the resource constraints identified by local partners. Requirements prioritized accessibility and local sustainability over technical sophistication.
- **Ecosystem of Values:** The team balanced multiple values - diagnostic accuracy (99% for parasite detection), speed (5 seconds inference time), accessibility (smartphone compatibility), and sustainability (locally manufacturable components).
- **Explainability:** The system was designed to provide decision support for healthcare workers rather than replace human judgment, maintaining transparency about AI capabilities and limitations.
- **Privacy Considerations:** Requirements included strict data anonymization protocols and local processing capabilities to protect patient privacy from the design stage.

3 Stage 3: Data Discovery

- **Ethical Data Collection:** When suitable datasets were unavailable, the team proactively created their own dataset through ethical protocols including institutional review board approvals and strict anonymization procedures.
- **Community Partnership:** Data collection involved collaborative relationships with local health facilities as partners rather than extractive relationships. The approach treated community data as an asset to be protected and shared responsibly.
- **Addressing Data Bias:** The team recognized that existing datasets didn't adequately represent local populations and contexts, leading to their decision to create representative datasets specifically for their target communities.
- **Transparency:** The team openly acknowledged bureaucratic challenges in data access while maintaining commitment to ethical standards, demonstrating transparency about process constraints.

4 Stage 4: Selecting and Developing a Model

- **Model Selection for Context:** Technical choices prioritized practical deployment in resource-constrained environments. The smartphone-based interface was chosen specifically for its accessibility and sustainability in the target context.
- **Fairness Considerations:** The team transparently acknowledged performance differences between single-class (99% accuracy) and multi-class detection (74% accuracy), committing to ongoing efforts to address these disparities.
- **Explainability Requirements:** The model was designed as a support tool that enhances rather than replaces human expertise, maintaining appropriate human oversight and decision-making authority.
- **Environmental Considerations:** The choice of smartphone technology and local manufacturing capability reduced environmental impact compared to importing expensive diagnostic equipment.

5 Stage 5: Testing and Interpreting Outcome

- **Real-World Testing:** Field testing was conducted in actual healthcare settings where the system would be deployed, ensuring validation under authentic conditions rather than controlled laboratory environments.
- **User Feedback Integration:** The team actively collected feedback from healthcare workers to ensure user-friendliness and practical application, making the end users central to the validation process.
- **Performance Validation:** AI diagnoses were rigorously compared with experienced technicians, providing robust validation while respecting existing human expertise.
- **Operation Manual Development:** The focus on user-centric design and practical application suggests development of accessible training and operation procedures, though specific details aren't provided in the source material.

6 Stage 6: Deployment & Post-Deployment Monitoring

- **Sustainable Deployment Strategy:** The 3D-printed adapter design enables local manufacturing and maintenance, ensuring communities can sustain the technology independently rather than depending on external support.
- **Continuous Adaptation:** Plans for expansion to other diseases (cervical cancer, tuberculosis) and telehealth platforms demonstrate commitment to ongoing adaptation based on evolving community needs.
- **Cultural Accessibility:** Future plans include adapting tools for local languages, showing understanding that deployment must address linguistic and cultural barriers.
- **Monitoring Through Expansion:** The systematic approach to expanding the framework to other health challenges suggests built-in monitoring and learning processes, though specific monitoring mechanisms aren't detailed in the source material.
- **Community Agency:** The emphasis on decision support rather than replacement tools suggests deployment strategies that maintain community control and professional agency in health decision-making.

Key Insights from the Lifecycle Mapping

The Makerere case study demonstrates several critical insights for human rights-based AI development:

- **Integration Throughout:** Human rights considerations weren't added as an afterthought but shaped every stage from initial problem definition through ongoing expansion plans.

- **Community Partnership:** Genuine community engagement occurred at each stage, with local partners having real influence on technical decisions rather than token consultation.
- **Ethical Foundations:** Privacy protection, informed consent, and transparency were foundational requirements that shaped technical architecture and deployment strategies.

- **Sustainability Focus:** Each stage prioritized long-term community ownership and control over short-term technical optimization or external dependency.

- **Continuous Learning:** The commitment to expansion and adaptation demonstrates understanding that human rights-based AI development is an ongoing process of learning and refinement rather than a one-time project.

Conclusion:
Technology in Service of Health Justice

The Makerere AI Health Lab's AI-powered malaria diagnostics initiative demonstrates that technology can serve human rights and health equity when developed with genuine community participation throughout the lifecycle. Rather than imposing external solutions, the project created tools that emerge from and serve community-identified needs.

The technical innovations—from 3D-printed adapters to smartphone-based AI—represent more than engineering achievements. They embody a philosophy that technology should democratize rather than concentrate capabilities, empower rather than replace human expertise, and serve justice rather than perpetuate existing inequities.

Most importantly, the project's commitment to ongoing expansion and adaptation demonstrates understanding that AI development for health equity is not a one-time intervention but an ongoing partnership with communities. This approach offers a model for how AI can genuinely serve the right to health, creating technology that enhances human dignity rather than undermining it.

The lessons from Makerere extend far beyond malaria diagnosis, providing a framework for any AI development that seeks to serve rather than exploit the communities it touches. In an era where AI often concentrates power and resources, this initiative demonstrates an alternative path—one where technology becomes a tool for justice, equity, and human flourishing.

About the case study and author

This case study analyzes research conducted by Makerere University, examining smartphone-based digital Microscopy Images for malaria diagnosis using Artificial Intelligence across Health facilities in Uganda between 2016 - 2024.

Dr Rose Nakasi leads the Makerere Health Intelligence lab that is specializing in advancing Artificial Intelligence and Data Science for developing automated tools and techniques for improved health especially in low resourced settings. She is a Principal investigator for the DS-I Malaria project under the DS-I Africa consortium funded by the NIH to support effective malaria diagnosis and surveillance in Uganda.

Other contributors to this case study are Caitlin Kraft-Buchman, Emma Kallina, and Sofia Kypraiou, authors of the original *Framework to AI Development: Integrating Human Rights Considerations Along the AI Lifecycle* upon which the Toolbox structure is based. Additional contributors are Amina Soulimani and Pilar Grant, from Women at the Table and the <AI & Equality> Human Rights Initiative.

<AI & Equality> African Toolbox | Case study

Co-Creating AI for Agriculture: Nigeria's Nsukka Yellow Pepper Project

Watch the video

This case study is part of the **African <AI & Equality> Toolbox**, which builds upon the methodology of the global <AI & Equality> Human Rights Toolbox—an initiative of Women At The Table in collaboration with the United Nations Office of the High Commissioner for Human Rights (OHCHR). The African Toolbox is a collaboration between the <AI & Equality> initiative and the African Centre for Technology Studies (ACTS). To learn more visit **aiequalitytoolbox.com**

102

Introduction

In the rural community of Nsukka in southeastern Nigeria, women farmers had cultivated the unique Nsukka Yellow Pepper for generations. Known worldwide for its distinctive aroma, this pepper represented both cultural heritage and economic opportunity. But climate change was threatening everything. Strange diseases were destroying crops overnight, water management was becoming increasingly difficult, and farmers were beginning to suspect each other of sabotage when plants appeared cut and damaged each morning.

What happened next challenges conventional approaches to the development of agricultural AI. Rather than arriving with pre-designed solutions, a team of researchers, engineers, and social scientists came with a different question: "What do you need?" The answer would reshape their understanding of both technology and community development, leading to innovations that emerged from the soil up—literally and figuratively.

The Nsukka Yellow Pepper project, part of the broader "Engendering Innovation" initiative under the AI for Development Africa program, demonstrates how artificial intelligence can serve agricultural communities when developed through genuine co-creation rather than top-down technology transfer. Professor Joel Nwakaire, who supervised the initiative as the Project Officer, emphasizes that this was "not a research-driven ideology, but a project that was co-created by identifying and prioritizing the needs of the people."

The genesis: From agricultural crisis to community innovation

The project emerged from a recognition that previous technology interventions in African agriculture had failed to bridge the gap between what researchers developed and what farmers actually needed. Women, who comprise approximately 70% of agricultural production in rural Africa, were particularly underserved by existing technology solutions that often reflected the priorities of developers rather than end users.

The Nsukka Yellow Pepper farmers faced multiple interconnected challenges:

- **Mysterious crop diseases** that appeared overnight, destroying entire sections of farmland.
- **Water management difficulties** exacerbated by climate change.
- **Lack of real-time information** about pest and disease detection.
- **Community tensions** arising from suspicions of sabotage when crops were found damaged.
- **Limited access to extension services** and market information.

Traditional agricultural extension services were inadequate to address these challenges. The farmers possessed substantial indigenous knowledge about their crops and environment, but lacked tools for real-time monitoring and early intervention. Most critically, they had no voice in determining what technological solutions might actually help them.

Building trust through dialogue: The community-centered approach

The breakthrough came through what the team calls “community dialogue”—a participatory approach that fundamentally reordered the relationship between technology developers and end users. Rather than consulting communities about predetermined solutions, the process began with creating safe spaces where farmers could articulate their own understanding of problems and potential solutions.

Separate, Safe Spaces for Authentic Voices

The team made a crucial decision to facilitate separate dialogue sessions based on gender and age, ensuring balanced representation while creating safe spaces where women could speak openly. This wasn’t merely a matter of inclusion—it was essential for understanding the full scope of agricultural challenges.

“You could see that the key outcome is that these women could speak freely about their needs, about how they even manage the current challenges in the farm, showing that they already have capacity unlike how we look at them as those who do not have the capacity,” explains Professor Nwakaire of the African Technology Policy Studies Network (ATPS).

These sessions revealed several critical insights:

- **Women possessed sophisticated indigenous knowledge** about crop management, pest identification, and adaptive strategies.
- **Different community members prioritized different challenges**—what researchers initially saw as the primary problem wasn’t necessarily what farmers identified as most urgent.
- **Social tensions around crop damage** were undermining community cohesion and needed to be addressed alongside technical solutions.
- **Trust building** was essential before any technology intervention could be successful.

Co-Identification of Priorities

Through this participatory process, farmers identified their three highest priority needs:

1. **Real-time pest and disease detection**—not just identification after damage occurred, but early warning systems that could prevent losses.
2. **Efficient water management**—tools to optimize irrigation timing and water use.
3. **Soil information and fertilizer optimization**—better understanding of soil conditions to improve input efficiency

Co-creating technical solutions

Critically, farmers specified that they didn't need help identifying diseases once they appeared—they already possessed that knowledge. What they needed was early detection of pest activity, particularly the mysterious cutting that was happening at night and causing community suspicions.

The technological solutions that emerged from this participatory process were unlike anything that would have been developed through conventional agricultural technology approaches. Each innovation directly addressed farmer-identified priorities while building on existing indigenous knowledge.

The Real-Time Pest Detection System

The centerpiece innovation was a solar-powered, real-time pest detection system that the community had never seen before. This standalone system used:

- **Raspberry Pi mini-computer** running locally (not cloud-dependent).
- **Dual 18-amp, 12-volt batteries** for reliable power.
- **Camera systems** that monitored crops continuously.
- **SMS alert capabilities** to notify farmers immediately of detected threats.

The system was designed to function as a "community farm" model—one installation that would monitor a representative plot, with any alerts prompting immediate treatment across all individual farms in the area. This approach reflected both resource constraints and the collective decision-making structure the community preferred.

The most dramatic validation of this system came when it solved the mystery that was dividing the community. Through images stored in the system's memory, farmers discovered that their crops weren't being sabotaged by neighbors—they were being cut by ants that came at night. This revelation not only prevented a brewing community crisis but demonstrated the power of evidence-based problem-solving.

Water Management Innovation

Responding to farmers' need for efficient water management, the team developed an SMS-based irrigation system that allowed farmers to control irrigation remotely. The system included:

- **Automated irrigation** using gravity flow and solar-powered mechanical valves.
- **SMS control capabilities** allowing farmers to activate irrigation from their homes.
- **Soil moisture sensors** providing real-time feedback about soil conditions.
- **Water level monitoring** to optimize irrigation timing and duration.

This innovation reduced water usage from 5,000 to 2,000 liters per hectare—a dramatic improvement in efficiency that farmers could immediately quantify and appreciate.

The E-Extension App: Community-Driven Support

As the technical systems were deployed and farmers became familiar with them, they identified another need: ongoing support and farm management assistance. This led to the development of "APWENFarm," an e-extension app that reflected farmers' own assessment of what additional support they needed.

Key features included:

- **Offline functionality** that synchronized when network connectivity was available.
- **Expert consultation** with agricultural specialists available to answer farmer questions.
- **Farm management tools** that helped farmers track inputs, expenses, and yields.
- **Mathematical integration** that calculated total expenditures and helped farmers make informed selling decisions.

This app addressed a critical gap that farmers themselves identified: they often forgot what inputs they had applied to their farms and lacked tools to calculate whether they were making a profit.

Integrating social and technical innovation

Perhaps the most significant aspect of the Nsukka project was its recognition that technology deployment must be integrated with broader social and economic empowerment. The participatory development process didn't end with technical solutions—it continued through capacity building and institutional development.

Cooperative Formation

As trust built through the co-creation process, the community decided to form a cooperative society. This wasn't an externally imposed requirement but emerged from farmers' own assessment of how they could collectively benefit from the technological innovations and strengthen their market position. The cooperative provided several benefits:

- **Collective ownership** of expensive technical equipment.
- **Shared learning** as farmers trained each other on system use.
- **Market power** through collective selling and input purchasing.
- **Sustained maintenance** of technical systems through shared responsibility.

Training and Capacity Building

The project included comprehensive training components, but these too were co-designed with farmers rather than imposed externally. Training covered:

- **Technical system operation** and basic maintenance.
- **Data interpretation** and decision-making based on sensor outputs.
- **Collective problem-solving** approaches for ongoing challenges.
- **Business planning** using the farm management tools.

This capacity building ensured that farmers weren't dependent on external technical support for ongoing system operation and could adapt the tools to their evolving needs.
The co-creation approach wasn't without challenges, but the participatory framework provided mechanisms for identifying and addressing problems as they emerged.

Challenges and adaptive solutions

Building initial trust

Some farmers were initially skeptical of external interventions, having experienced previous projects that extracted information without providing meaningful benefits. The team addressed this through:

- **Extended engagement** with multiple community meetings before any technology deployment.
- **Transparent communication** about project goals, funding, and expected outcomes.
- **Demonstration of respect** for indigenous knowledge and farmer expertise.
- **Commitment to community ownership** of both data and technology.

Technical adaptation to local conditions

Standard agricultural technology often fails in African contexts due to power, connectivity, and maintenance constraints. The co-creation process helped identify these challenges early and develop appropriate solutions:

- **Solar power systems** using locally sourced components
- **Standalone operation** that didn't require constant internet connectivity
- **Simple, robust designs** that farmers could maintain themselves
- **Local partnership development** for ongoing technical support

Gender-Responsive Implementation

Ensuring that technology benefits reached women farmers required intentional design choices throughout the process:

- **Separate consultation sessions** to ensure women's voices were heard.
- **Technology design** that accounted for women's specific agricultural responsibilities.
- **Training programs** that accommodated women's time constraints and learning preferences.
- **Cooperative structures** that included women in leadership roles.

Impact beyond technology:
Transforming agricultural development

The Nsukka Yellow Pepper project's impact extends far beyond the technical innovations themselves. By demonstrating that farmers possess sophisticated knowledge and can be genuine partners in technology development, the project has influenced broader approaches to agricultural development.

- **Community Empowerment**
  Farmers report increased confidence in their ability to advocate for their needs and participate in agricultural development initiatives. The participatory process validated their knowledge and gave them tools to document and communicate their experiences to outside actors.

- **Sustainable Technology Adoption**
  Because farmers were involved in designing the technical solutions, adoption rates were high and sustained. The community took ownership of maintaining and adapting the systems rather than waiting for external technical support.

- **Model for Replication**
  The co-creation methodology developed through the Nsukka project has been documented and shared with other agricultural development initiatives across Africa. The emphasis on "design by inclusion" has influenced policy discussions about agricultural technology deployment.

- **Academic and Policy Impact**
  The project has contributed to academic literature on participatory technology development and influenced policy discussions about gender-responsive agricultural innovation. Research publications have documented the methodology and outcomes, providing evidence for alternative approaches to agricultural AI development.

Looking Forward: Scaling Co-Creation

The Nsukka Yellow Pepper project demonstrates that AI can serve agricultural communities when developed through genuine co-creation rather than top-down technology transfer. As the approach expands to other crops and regions, each new deployment offers opportunities to refine the methodology while adapting to different agricultural, social, and cultural contexts.

The project's success stems not from technological sophistication alone, but from recognizing that agricultural AI is ultimately about human dignity, community empowerment, and the right to food security. The farmers in Nsukka who no longer suffer mysterious crop losses, who have improved their water efficiency, and who have gained tools for better farm management represent the true measure of AI's potential to serve human rights and agricultural justice.

The yellow peppers continue to grow in Nsukka, but now they're monitored by systems that emerged from the soil up—technologies that reflect the knowledge, priorities, and agency of the women who tend them.

Mapping the AI Lifecycle HRIA Framework for the Nsukka Yellow Pepper Project

1 Stage 1: Objective and Team Composition

The project began with community dialogue rather than predetermined objectives. Through separate, safe sessions with women farmers, researchers learned that the primary need wasn't technology deployment but community-identified solutions to crop losses and water management challenges. The team composition evolved to include farmers as co-developers, not just end users.

HRIA Framework Alignment:

- **Purpose & Context:** The system emerged from community-identified problems rather than external assumptions about agricultural needs.
- **Effects of the System:** Benefits were explicitly designed to empower women farmers who comprise 70% of agricultural production but are often excluded from technology development.
- **Empowering Affected Communities:** Farmers served as co-developers throughout the process, with genuine decision-making power in system design and implementation.
- **Team Composition:** The team included diverse expertise (technical, social, agricultural) and meaningful representation from affected communities.

Key Human Rights Considerations:

The initiative explicitly addressed agricultural justice as a human rights issue, recognizing that food security and livelihood sustainability are fundamental to dignity. Team composition ensured that those most affected by agricultural challenges had agency in solution development.

2 Stage 2: Defining System Requirements

System requirements emerged from participatory mapping and community priority-setting rather than technical specifications. Requirements included: real-time pest detection (not just identification), SMS-based remote irrigation control, offline functionality, community farm monitoring model, and integration with existing indigenous knowledge systems.

HRIA Framework Alignment:

- **Technical equirements Driven by Community Needs:** The choice of smartphone-based technology and 3D-printed adapters directly addressed the resource constraints identified by local partners. Requirements prioritized accessibility and local sustainability over technical sophistication.
- **Ecosystem of Values:** The team balanced multiple values - diagnostic accuracy (99% for parasite detection), speed (5 seconds inference time), accessibility (smartphone compatibility), and sustainability (locally manufacturable components).
- **Explainability:** The system was designed to provide decision support for healthcare workers rather than replace human judgment, maintaining transparency about AI capabilities and limitations.
- **Privacy Considerations:** Requirements included strict data anonymization protocols and local processing capabilities to protect patient privacy from the design stage.

Key Human Rights Considerations:

Requirements prioritized dignity and agency for women farmers. Features like community ownership, SMS communication, and building on indigenous knowledge ensured that system design served empowerment rather than creating new dependencies.

3 Stage 3: Data Discovery

Data discovery combined technical sensor measurements with community knowledge through participatory mapping, farmer expertise, and collaborative problem identification. The process involved communities in identifying what data to collect, how to interpret findings, and how to use information for collective benefit.

HRIA Framework Alignment:

- **Data Origin:** Data collection respected community consent and ownership, with clear agreements about data use. The community farm model ensured collective benefit from information gathering.
- **Data Bias:** The participatory approach explicitly addressed historical bias in agricultural development by centering women farmers' knowledge and priorities.
- **Documentation:** All data sources, collection methods, and interpretation processes were documented transparently, with findings shared back to communities in accessible formats.

Key Human Rights Considerations:

The data discovery process treated indigenous agricultural knowledge as equally valid to technical measurements. Participatory approaches ensured that communities defined agricultural priorities rather than having external definitions imposed.

4 Stage 4: Selecting and Developing a Model

AI models were developed to serve community-identified needs: real-time pest detection for early warning, soil moisture monitoring for irrigation optimization, and farm management tools for economic empowerment. Model selection prioritized interpretability and community utility over technical sophistication.

HRIA Framework Alignment:

- **Model Type and Explainability:** Models prioritized explainability appropriate to farmer contexts, with visual outputs and clear communication about system capabilities and limitations.
- **Fairness Aspects:** The initiative explicitly considered how agricultural challenges affect different community groups (women, men, different age groups) and ensured that AI models supported gender equity.
- **Environmental Impact:** Solar-powered sensors and locally sourced components minimized environmental footprint while supporting local economies.

Key Human Rights Considerations:

Model development served community empowerment rather than technical optimization. AI enhanced rather than replaced indigenous knowledge, providing tools for agricultural justice and self-determination.

5 Stage 5: Testing and Interpreting Outcome

Testing involved both technical validation and community feedback. Farmers evaluated whether the system met their needs, provided useful information, and supported their agricultural goals. Outcomes were interpreted collaboratively, with community members trained to understand and use findings for collective benefit.

HRIA Framework Alignment:

- **Testing Context and Outcomes:** Testing occurred in real agricultural contexts with actual users, incorporating feedback from diverse community members about system utility and effectiveness.
- **Operation Manual:** Training materials were developed collaboratively, and community members were trained to operate and interpret the system independently.

Key Human Rights Considerations:

Testing evaluated whether the system genuinely empowered farmers to improve their livelihoods and agricultural practices. Community feedback shaped system refinements, ensuring that technical performance served human dignity and agency.

6 Stage 6: Deployment & Post-Deployment Monitoring

Deployment involved comprehensive community training, formation of cooperative societies for collective ownership, and continuous adaptation based on farmer feedback. The initiative included ongoing support systems, peer learning networks, and integration with existing community structures for sustainability.

HRIA Framework Alignment:

- **Deployment:** Communities had genuine agency to modify deployment based on their assessment of benefits and effectiveness. Deployment included robust support systems and capacity building for participants.
- **Monitoring:** Continuous monitoring included both technical performance and community impact, with mechanisms for farmers to report concerns or suggest improvements. Success was measured by agricultural productivity and community empowerment outcomes.

Key Human Rights Considerations:

Post-deployment monitoring ensured that the system continued to serve community needs rather than becoming extractive. Regular community feedback loops maintained farmer ownership and adapted the system to evolving agricultural and social needs.

Integrated Analysis:
Human Rights Throughout the AI Lifecycle

The Nsukka Yellow Pepper project demonstrates how human rights considerations can be integrated throughout the AI lifecycle rather than added as an afterthought. Several key principles emerge:

- **Community Co-Development:** At every stage, farmers had genuine decision-making power rather than tokenistic consultation. This agency extended from initial problem definition through ongoing system adaptation.
- **Justice-Oriented Design:** Technical choices consistently prioritized community empowerment and agricultural justice over technical optimization or efficiency metrics.
- **Participatory Knowledge Creation:** The initiative treated indigenous agricultural knowledge as equally valid to technical expertise, creating collaborative knowledge production rather than extractive data collection.
- **Adaptive Implementation:** System design and implementation adapted continuously based on farmer feedback, ensuring that AI served evolving community needs rather than static technical specifications.
- **Sustainability Through Ownership:** Long-term sustainability was built through community ownership, cooperative formation, and capacity development rather than external dependency.

The Nsukka experience demonstrates that AI can serve human rights and agricultural justice when developed with genuine community participation throughout the lifecycle. This approach results in more robust, sustainable, and effective systems that empower rather than marginalize affected communities.

About the case study

This case study analyzes research conducted by the African Technology Policy Studies Network (ATPS) in collaboration with the International Centre of Insect Physiology and Technology (icipe), Kenya and Kumazi Hive (Ghana), focused on Strengthening the Capacity of Women and Marginalized Communities in Africa's Agriculture and Food Systems to Harness the Potentials of Artificial Intelligence Technology in alliance with the Artificial Intelligence for Agriculture and Food Systems (AI4AFS) project "Using Artificial Intelligence to Enhance the Production, Marketing, and Management of Nsukka Yellow Pepper in Nigeria" led by Professor Chinenye Anyadike of the Association of Profession Women Engineers (APWEN), with partners from University of Nigeria, Nsukka, and Educare Nigeria Limited, Nigeria between 2022–2024.

Engr. Prof. Joel Nwaeze Nwakaire is a Professor of Agricultural and Bioresources Engineering at the University of Nigeria, Nsukka. He is committed to effectively achieving the sustainable development goals of zero hunger and poverty, ensuring gender equality, and providing affordable and clean energy through Science, Technology, and Innovation. He has managed the all-African programme on Artificial Intelligence in Agriculture and Food Systems, sponsored by the IDRC and the Swedish International Development Agency. He is also the project manager of the SCALE STEP Change IDRC on "Strengthening the capacity of the extension system to use proven knowledge and technologies to sustain equitable locally-led adaptation among smallholder farmers.

Other contributors to this case study are Caitlin Kraft-Buchman, Emma Kallina, and Sofia Kypraiou, authors of the original *Framework to AI Development: Integrating Human Rights Considerations Along the AI Lifecycle* upon which the Toolbox structure is based. Additional contributors are Amina Soulimani and Pilar Grant, from Women at the Table and the <AI & Equality> Human Rights Initiative.

<AI & Equality> African Toolbox | Case study

Empowering African Languages through NLP: KenCorpus Project

Watch the video

This case study is part of the **African <AI & Equality> Toolbox**, which builds upon the methodology of the global <AI & Equality> Human Rights Toolbox—an initiative of Women At The Table in collaboration with the United Nations Office of the High Commissioner for Human Rights (OHCHR). The African Toolbox is a collaboration between the <AI & Equality> initiative and the African Centre for Technology Studies (ACTS). To learn more visit **aiequalitytoolbox.com**

115

The Silent Crisis

In 1992, linguist Parcel Hill made a chilling prediction: by the year 2100, the world's linguistic diversity would largely disappear, with most languages becoming obsolete as people gravitated toward English and other dominant tongues. What seemed like a distant academic concern has become a pressing reality, particularly visible in the digital realm where artificial intelligence is reshaping how we communicate, learn, and preserve knowledge.

Dr. Lilian Wanzare, a researcher at Maseno University in Kenya, witnessed this crisis firsthand. Despite Africa being home to over 2,000 languages and Kenya alone hosting more than 50 distinct languages across Nilotic, Bantu, and Cushitic families, the digital world remained largely silent in these tongues. The statistics were stark and sobering: while 77% of natural language processing tools supported English and other "global north" languages, only 6% supported low-resource languages. Yet this 6% represented the linguistic reality of 3 billion people – nearly half the world's population.

The irony was profound. The very technologies designed to bridge communication gaps were actually widening them, creating a digital apartheid where the world's linguistic diversity was being systematically erased, one algorithm at a time.

The Awakening: Understanding the roots of exclusion

Dr. Wanzare and her team began to understand why African languages were disappearing from the digital landscape. The problem wasn't just technological – it was fundamentally about data. Every AI system, every translation tool, every speech recognition service needed vast amounts of digital text and audio to learn from. But African languages existed primarily in the oral tradition, in the stories told by elders, in the daily conversations of rural communities, in the songs sung during harvest seasons.

The educated African population, ironically, had become part of the problem. Colonial legacies meant that English, French, or Portuguese served as official languages in most African countries. Educated Africans often couldn't write fluently in their native tongues. They didn't blog in Dholuo, didn't tweet in Kalenjin, didn't write academic papers in Kikuyu. The natural generators of digital content – the educated, urban, connected populations – were creating content in colonial languages, not indigenous ones.

This created a vicious cycle: no digital content meant no data, no data meant no AI tools, no AI tools meant these languages remained excluded from the digital future, making them appear less valuable and further accelerating their decline.

The decision: Community at the center

Faced with this reality, Dr. Wanzare made a radical decision. Instead of accepting that African languages were "low-resource," she would mobilize entire communities to become active participants in creating the digital future of their own languages. This wasn't going to be a top-down technological solution imposed by researchers in university labs. It would be a grassroots movement, with communities as partners, not subjects.

The KenCorpus project was born from this philosophy. Over what would become a five-year journey, Dr. Wanzare and her team would need to go beyond traditional academic research. They would need to become community organizers, cultural ambassadors, and bridge-builders between oral traditions and digital futures.

Building the Foundation: Stories become data

The first phase of KenCorpus was deceptively simple yet profoundly challenging. The team began traveling to rural communities, sitting with elders, talking with families, and asking them to do something that had never been systematically done before: tell their traditional stories and have them recorded and transcribed into digital form.

This wasn't just data collection – it was cultural preservation in action. Each story captured wasn't just text for training algorithms; it was a piece of living heritage being transferred from the oral realm into the digital one. Grandmothers who had never seen a computer became, unknowingly, the first contributors to Kenya's digital language infrastructure.

The team faced immediate challenges. How do you capture the tonal variations of different languages? How do you account for the fact that the same language might be spoken differently in coastal areas versus highland regions? How do you respect cultural protocols around storytelling while creating standardized digital formats?

The solution emerged through deep community engagement. Local chiefs provided credibility and mobilization support. Primary school teachers helped with transcription and verification. Church leaders opened their congregations as venues for recording sessions. The project became a community affair, with everyone understanding they were participating in preserving their linguistic heritage for future generations.

Expanding the Vision:
From Stories to Systems

As the initial collections grew, Dr. Wanzare and her team began to understand what communities actually needed from these digital language tools. Three clear priorities emerged from their conversations with language speakers:

Translation became the first critical need. People wanted to communicate across language barriers – not just from English to local languages, but between local languages themselves. A Dholuo speaker needed to communicate with a Kalenjin speaker. Government information in English needed to be accessible in local languages. This meant creating parallel corpora – the same sentences translated across multiple languages and carefully aligned. The team made a strategic decision to use Kiswahili as an anchor language. Rather than making English the central hub, they recognized Kiswahili as a widely understood African language that could serve as a bridge between different Kenyan languages. This wasn't just technically sound; it was culturally appropriate and politically significant.

Speech recognition emerged as the second priority. Communities envisioned a future where they could speak to their phones in their native languages, where meetings could be automatically transcribed in Kikuyu, where oral traditions could be instantly converted to written form. This required building massive speech corpora – targeting 3,000 hours of recorded speech across five languages.

Language infrastructure became the third need. Behind every grammar checker, every spell-check system, every language learning app are fundamental NLP tasks like part-of-speech tagging. These might seem mundane to technologists, but they're the backbone of language technology. Without them, no advanced language tools can function properly.

The Technical Challenge:
Building AI for the unconnected

Creating AI systems for languages with no existing digital infrastructure required innovative approaches. Traditional machine learning assumes you can scrape vast amounts of text from the internet. For Kenyan languages, the internet was essentially empty. Dr. Wanzare's team had to become experts not just in AI, but in linguistics, anthropology, and community organizing. They needed to understand how code-switching worked – the way speakers naturally mixed their native languages with Kiswahili or English within single conversations. They needed to capture not just formal language, but the way people actually spoke in their daily lives.

The technical architecture they developed was multilingual by design, with Kiswahili serving as the anchor. This meant a Dholuo speaker could ask a question to an AI system like ChatGPT by speaking in Dholuo. The system would translate to English, process the query, generate a response in English, then translate back to Dholuo. For the first time, global AI systems could become accessible to speakers of indigenous African languages.

Confronting Deeper Questions: Who owns language?

As the project grew, deeper questions emerged. Who owns the data being collected? What happens when global tech companies want to use these datasets? How do you ensure that communities benefit from the AI systems built on their linguistic contributions?

Working with Mozilla Common Voice, the team began developing community-based licensing frameworks. These weren't just legal documents; they were attempts to encode indigenous concepts of collective ownership and community sovereignty into the digital age. Traditional open-source licenses assumed individual ownership and global access. But languages belong to communities, not individuals. The stories being recorded were part of cultural heritage, not just data points.

This innovation had implications far beyond Kenya. Indigenous communities worldwide were grappling with similar questions as AI systems began to incorporate their languages and cultural knowledge. The KenCorpus approach offered a model for how communities could maintain sovereignty over their linguistic heritage while still participating in global technological development.

The Human Network: Beyond technology

Five years into the project, it became clear that KenCorpus's greatest innovation wasn't technological – it was social. The project had created a network of thousands of people across Kenya who understood themselves as active participants in shaping their languages' digital future. Local research assistants were working in Somaliland, in rural Kalenjin communities, in urban Nairobi neighborhoods. University linguists were collaborating with primary school teachers. County governments were providing resources. Media houses were contributing their archives. Traditional chiefs were endorsing the work in community meetings.

This network solved the fundamental challenge of scaling data collection for low-resource languages. You can't build linguistic infrastructure without massive community participation. But you can't get community participation without trust, cultural sensitivity, and genuine partnership.

Dr. Wanzare learned that incentivization was about more than payment. People participated because they understood the long-term vision: their children would grow up in a world where their native languages weren't barriers to accessing education, healthcare, government services, or economic opportunities. Their languages wouldn't just survive; they would thrive in the digital age.

Scaling the vision: Small models, big impact

The project also pioneered a different approach to AI development. Instead of pursuing ever-larger language models, the KenCorpus team focused on small, domain-specific models tailored to community needs. These models could run on modest hardware, could be customized for specific dialects, and were more accurate for their intended use cases than generic large models.

This approach challenged the prevailing Silicon Valley wisdom that bigger is always better. For communities with limited technological infrastructure, smaller, specialized models were actually more appropriate and more empowering.

The team also established critical research questions: What's the minimum viable amount of data needed to create functional language models? How do you balance model accuracy with cultural appropriateness? How do you ensure AI systems respect the way languages are actually spoken in communities rather than imposing academic standards?

The ripple effect: Beyond Kenya

By its fifth year, KenCorpus had become more than a Kenyan project. Researchers from across Africa were adapting its methodologies. International organizations were funding similar initiatives. The approach was being discussed in academic conferences, policy forums, and community meetings across the Global South.

The project demonstrated that technological marginalization wasn't inevitable. Communities could become active agents in their own digital empowerment. Languages that had been written off as "low-resource" could become fully functional in the digital ecosystem through systematic community engagement and culturally appropriate technical approaches.

More importantly, KenCorpus showed that AI development could be genuinely participatory. Instead of technology being developed for communities, it could be developed with communities as equal partners and primary beneficiaries.

Lessons from the Field: What KenCorpus taught us

After five years of intensive work, several critical insights emerged:

- **Community engagement must be continuous and authentic.** You can't extract linguistic data and disappear. Building language technology requires ongoing relationships and genuine partnership.
- **Cultural context is as important as technical accuracy.** AI systems that don't respect how languages are actually used in communities will fail, no matter how technically sophisticated they are.
- **Incentivization is complex.** People contribute not just for immediate payment but for long-term community benefit. The most sustainable models align technological development with community empowerment.
- **Diversity within languages matters.** Even small languages have dialects, regional variations, and social differences. Effective language technology must account for this internal diversity rather than assuming homogeneity.
- **Innovation happens at the margins.** Some of the most important breakthroughs came from constraints. Limited resources forced creative solutions. Community needs drove technical innovation. Working with "low-resource" languages revealed possibilities that weren't visible when working with well-resourced languages.

The future: What comes next

As KenCorpus enters its next phase, the vision is expanding. Speech recognition systems are being deployed in local schools. Translation tools are being integrated into government services. Community members are being trained as data collectors and language technology specialists.

But perhaps most importantly, a new generation of young Kenyans is growing up understanding that their native languages are not barriers to technological participation – they are pathways to it. Children are learning that speaking Dholuo or Kalenjin isn't a limitation; it's a superpower that makes them uniquely valuable in an increasingly multilingual digital world.

The project has also inspired similar initiatives across Africa and beyond. In Nigeria, researchers are applying KenCorpus methodologies to Yoruba and Igbo. In South Africa, similar work is beginning with Xhosa and Zulu. Indigenous communities in the Americas are adapting the community engagement strategies for their own language preservation efforts.

The broader transformation:
From extraction to partnership

KenCorpus represents something larger than a single research project. It embodies a fundamental shift in how technology development can work. Instead of Silicon Valley companies extracting data from global communities to build products sold back to them, KenCorpus demonstrates true technological partnership.

Communities aren't just data sources; they're co-designers, co-owners, and primary beneficiaries. Technology isn't imposed from outside; it emerges from community needs and community participation. Linguistic diversity isn't a problem to be solved; it's a resource to be celebrated and empowered.

This model has implications far beyond language technology. As AI systems become more central to education, healthcare, governance, and economic life, the KenCorpus approach offers a template for ensuring that technological advancement serves community empowerment rather than community marginalization.

Mapping the AI Lifecycle HRIA Framework for the KenCorpus case

1 Stage 1: Objective and Team Composition

Problem Definition: KenCorpus began with a clear understanding that the digital marginalization of African languages wasn't just a technical problem – it was a human rights issue. The objective emerged directly from community needs rather than technological possibilities. Dr. Wanzare and her team recognized that less than 0.01% of the world's languages were supported by NLP tools, leaving 3 billion speakers without access to digital language technologies.

Team Composition & Community Partnership: The project exemplified participatory development from the outset. The team composition evolved to include:

- Academic researchers (Dr. Wanzare and university partners).
- Community leaders (chiefs, elders, religious leaders).
- Educational partners (teachers, school administrators).
- Linguistic experts (native speakers, cultural specialists).
- Government representatives (county officials).
- Media partners (local broadcasters, content creators).
- Technical specialists (ML engineers, linguists).

Human Rights integration

The project directly addressed multiple human rights principles:

- **Cultural rights:** Preserving and promoting linguistic heritage.
- **Participation rights:** Communities as co-designers, not data subjects.
- **Non-discrimination:** Ensuring technological access regardless of language.
- **Self-determination:** Communities controlling their linguistic data.

Key decisions made

- Kiswahili chosen as anchor language rather than English (cultural appropriateness).
- Community needs prioritized over technical convenience.
- Long-term sustainability valued over short-term data extraction.
- Traditional knowledge systems respected alongside academic expertise.

2 Stage 2: Defining System Requirements

Value Ecosystem Navigation

KenCorpus navigated complex trade-offs between different values:

- **Accuracy vs. Cultural appropriateness:** Choosing community-validated translations over technically optimized ones.
- **Efficiency vs. Inclusivity:** Including multiple dialects despite increased complexity.
- **Speed vs. Sustainability:** Building long-term community relationships over rapid data collection.
- **Standardization vs. Authenticity:** Preserving natural language variation while creating usable datasets.

Community-Driven Requirements

System requirements emerged through extensive community consultation:

1. **Translation systems**: Cross-language communication (local-to-local, not just English-centric).
2. **Speech recognition:** Automatic transcription in native languages.
3. **Fundamental NLP infrastructure:** Grammar checking, spell checking, part-of-speech tagging.
4. **Cultural preservation:** Maintaining oral traditions in digital form.
5. **Educational support:** Tools for language learning and literacy.

Explainability & Transparency

The project prioritized community understanding over technical sophistication:

- Explanations provided in culturally appropriate formats.
- Community members trained to understand system capabilities and limitations.
- Decision-making processes made transparent to all stakeholders.
- Clear documentation of why certain approaches were chosen.

Accountability Structures

- Community representatives included in all major decisions.
- Regular feedback sessions with language speakers.
- Cultural appropriateness reviews by elders and traditional authorities.
- Academic oversight balanced with community sovereignty.

3 Stage 3: Data Discovery

Ethical Data Collection.

KenCorpus revolutionized data collection by prioritizing community ownership:

- **Consent processes:** Developed in consultation with traditional authorities.
- **Cultural protocols:** Respected storytelling traditions and sacred knowledge boundaries.
- **Community licensing:** Pioneered community-based data ownership models.
- **Benefit sharing:** Ensured communities retained control over their linguistic data.

Addressing Historical Bias

The project confronted multiple forms of bias:

- **Colonial bias:** Rejecting English-centric approaches in favor of indigenous frameworks.
- **Urban bias:** Actively seeking rural and traditional speakers.
- **Educational bias:** Including non-literate speakers as valuable contributors.
- **Gender bias:** Ensuring women's voices and perspectives were included.
- **Generational bias:** Capturing both traditional and contemporary language use.

Data Diversity & Representation

- **Geographic diversity:** Coastal, highland, and urban dialect variations.
- **Social diversity:** Different educational backgrounds, age groups, professions.
- **Linguistic diversity:** Formal and informal registers, code-switching patterns.
- **Cultural diversity:** Different storytelling traditions, ceremonial language use.

Documentation & Preservation

- Raw audio preserved alongside processed datasets.
- Cultural context documented for each collection session.
- Metadata included information about speakers, contexts, and cultural significance.
- Version control maintained to track changes and improvements.

4 Stage 4: Selecting and Developing a Model

Model Architecture Decisions

KenCorpus made strategic choices that prioritized community needs:

- **Multilingual architecture:** Kiswahili as anchor rather than English-centric design.
- **Small, specialized models:** Domain-specific rather than general-purpose systems.
- **Explainable approaches:** Interpretable models over black-box systems.
- **Modular design:** Components could be updated independently as communities evolved.

Fairness Considerations

- **Cross-dialectal fairness:** Ensuring systems worked across regional variations.
- **Intersectional analysis:** Considering gender, age, education, and regional factors.
- **Performance equity:** Avoiding accuracy disparities between different groups.
- **Cultural fairness:** Respecting different ways of expressing concepts.

Technical Innovation

- **Minimum viable data research:** Determining smallest datasets needed for functionality.
- **Code-switching capabilities:** Handling natural language mixing patterns.
- **Tonal language processing:** Accounting for tone markers and prosodic features.
- **Low-resource optimization:** Maximizing performance with limited training data.

Community Validation

- Native speakers involved in model testing and refinement.
- Cultural appropriateness evaluated by community authorities.
- Performance tested in real-world community contexts.
- Feedback loops established for continuous improvement.

5 Stage 5: Testing and Interpreting Outcomes

Multi-Stakeholder Testing

Testing involved diverse community members:

- **Native speakers:** Accuracy and naturalness evaluation.
- **Community leaders:** Cultural appropriateness assessment.
- **Educators:** Pedagogical effectiveness testing.
- **Technical users:** System reliability and performance evaluation.

Performance Metrics

Beyond technical accuracy, KenCorpus evaluated:

- **Cultural appropriateness:** Does the system respect traditional language use?
- **Community acceptance:** Do speakers feel their language is well-represented?
- **Practical utility:** Do the tools meet actual community needs?
- **Fairness across groups:** Do all community segments benefit equally?

Extreme Case Testing

- **Rare dialects:** Testing with less common regional variations.
- **Code-switching:** Evaluating mixed-language scenarios.
- **Cultural contexts:** Testing in ceremonial and formal contexts.
- **Technical edge cases:** Handling poor audio quality, background noise.

Documentation for Users

- **Community-friendly manuals:** Explanations in local languages and cultural contexts.
- **Training materials:** Building local capacity for system use and maintenance.
- **Limitation documentation:** Clear explanation of what systems can and cannot do.
- **Best practices:** Guidance for optimal use in different contexts.

6 Stage 6: Deployment & Post-Deployment Monitoring

Community-Controlled Deployment

- **Community consent:** Final deployment required explicit community approval.
- **Phased rollout:** Gradual implementation allowing for adjustment and feedback.
- **Local ownership:** Communities retained control over how systems were used.
- **Opt-out mechanisms:** Clear pathways for communities to withdraw participation.

Ongoing Monitoring Systems

- **Community feedback channels:** Regular mechanisms for reporting issues or suggestions.
- **Cultural evolution tracking:** Monitoring how language use changes over time.
- **Performance monitoring:** Continuous assessment of system accuracy and fairness.
- **Usage pattern analysis:** Understanding how communities actually use the tools.

Adaptive Management

- **Regular system updates:** Incorporating new community feedback and needs.
- **Dialect evolution:** Accounting for natural language change over time.
- **Technology evolution:** Updating systems as new approaches become available.
- **Community capacity building:** Training local experts for ongoing maintenance.

Impact Assessment

- **Language vitality metrics:** Measuring impact on language use and transmission.
- **Community empowerment:** Assessing changes in technological access and agency.
- **Educational outcomes:** Evaluating impact on literacy and learning.
- **Cultural preservation:** Measuring success in maintaining oral traditions.

Long-term Sustainability

- **Local expertise development:** Training community members as technical specialists.
- **Institutional partnerships:** Building sustainable relationships with schools, g overnment, media.
- **Financial sustainability:** Developing models that don’t depend on external funding.
- **Replication support:** Helping other communities adapt the methodology

Conclusion:
A New Paradigm for AI Development

KenCorpus demonstrates that AI development can be genuinely participatory, culturally appropriate, and community-empowering. By integrating human rights considerations throughout the AI lifecycle, the project shows how technology can serve linguistic diversity rather than undermining it.

The project's success lies not just in its technical achievements, but in its demonstration that communities can be equal partners in shaping their technological future. When AI development prioritizes human dignity, cultural preservation, and community empowerment, the resulting systems are not only more ethical – they're more effective, more sustainable, and more innovative.

As AI systems become increasingly central to human life, the KenCorpus model offers a roadmap for development that enhances rather than diminishes human diversity. It proves that the choice between technological advancement and cultural preservation is a false one – with the right approach, technology can be the most powerful tool for cultural empowerment and human flourishing.

About the case study

This case study analyzes research conducted by Dr. Lilian Wanzare, Prof. Florence Indede, Dr. Owen McOnyango of Maseno University, Dr. Edward Ombui of USIU (then African Nazarene University), Dr. Lawrence Muchemi and Mr. Benard Wanjawa of University of Nairobi, and the KenCorpus language community, examining Languages spoken in Kenya in the lens of Natural Language Processing across several counties in Kenya between 2021 - 2022. This research was made possible by funding from Meridian Institute's Lacuna Fund under grant no. 0393-S-001 which is a funder collaboration between The Rockefeller Foundation, Google.org, and Canada's International Development Research Centre.

Dr. Lilian Wanzare is a lecturer and chair of the Department of Computer Science at Maseno University. Her research interests are in Artificial Intelligence and Machine Learning, in particular Natural Language Processing (NLP), Sign Language research and building text processing tools for low-resource languages. She holds a PhD degree in Computational Linguistics and an Msc. in Language Science and Technology from Saarland University, Germany.

Other contributors to this case study are Caitlin Kraft-Buchman, Emma Kallina, and Sofia Kypraiou, authors of the original *Framework to AI Development: Integrating Human Rights Considerations Along the AI Lifecycle* upon which the Toolbox structure is based. Additional contributors are Amina Soulimani and Pilar Grant, from Women at the Table and the <AI & Equality> Human Rights Initiative.

Resources

Dataset Locations

- https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6N5V1K
- https://commonvoice.mozilla.org/luo
- https://commonvoice.mozilla.org/dav
- https://commonvoice.mozilla.org/kln
- https://zenodo.org/records/13355021

Publications

- Wanjawa, B., Wanzare, L., Indede, F., McOnyango, O., Ombui, E., & Muchemi, L. (2022). Kencorpus: A kenyan language corpus of swahili, dholuo and luhya for natural language processing tasks. arXiv preprint arXiv:2208.12081.
- Babirye, C., Nakatumba-Nabende, J., Katumba, A., Ogwang, R., Francis, J. T., Mukiibi, J., ... & David, D. (2022). Building text and speech datasets for low resourced languages: A case of languages in east africa. In 3rd Workshop on African Natural Language Processing.
- Wanjawa, B. W., Wanzare, L. D., Indede, F., McOnyango, O., Muchemi, L., & Ombui, E. (2023). KenSwQuAD—A Question Answering Dataset for Swahili Low-resource Language. ACM Transactions on Asian and Low-Resource Language Information Processing, 22(4), 1-20.
- Awino, E., Wanzare, L., Muchemi, L., Wanjawa, B., Ombui, E., Indede, F., ... & Okal, B. (2022). Phonemic Representation and Transcription for Speech to Text Applications for Under-resourced Indigenous African Languages: The Case of Kiswahili. arXiv preprint arXiv:2210.16537.
- Muhammad, S. H., Abdulmumin, I., Ayele, A. A., Adelani, D. I., Ahmad, I. S., Aliyu, S. M., ... & Ousidhoum, N. (2025). AfriHate: A Multilingual Collection of Hate Speech and Abusive Language Datasets for African Languages. arXiv preprint arXiv:2501.08284.
- Amol, C. J., Chimoto, E. A., Gesicho, R. D., Gitau, A. M., Etori, N. A., Kinyanjui, C., ... & Tombe, R. (2024). State of NLP in Kenya: A Survey. arXiv preprint arXiv:2410.09948.
- Amol, C., Wanzare, L., & Obuhuma, J. (2023, December). Politikweli: A swahili-english code-switched twitter political misinformation classification dataset. In International Conference on Speech and Language Technologies for Low-resource Languages (pp. 3-17). Cham: Springer Nature Switzerland.
- Muhammad, S. H., Ousidhoum, N., Abdulmumin, I., Wahle, J. P., Ruas, T., Beloucif, M., ... & Mohammad, S. M. (2025). BRIGHTER: BRIdging the Gap in Human-Annotated Textual Emotion Recognition Datasets for 28 Languages. arXiv preprint arXiv:2502.11926.
- Amol, C., Wanzare, L., & Obuhuma, J. ()2025) Modelling Misinformation in Swahili-English Code-switched Texts. mecs-press.org
- Mbogho, A., Awuor, Q., Kipkebut, A., Wanzare, L., & Oloo, V. (2025). Building low-resource African language corpora: A case study of Kidawida,Kalenjin and Dholuo. arXiv preprint arXiv:2501.11003.

<AI & Equality> African Toolbox | Case study

Design by Inclusion in AI Development: Uganda's Cassava Farming Initiative

Watch the video

This case study is part of the **African <AI & Equality> Toolbox**, which builds upon the methodology of the global <AI & Equality> Human Rights Toolbox—an initiative of Women At The Table in collaboration with the United Nations Office of the High Commissioner for Human Rights (OHCHR). The African Toolbox is a collaboration between the <AI & Equality> initiative and the African Centre for Technology Studies (ACTS). To learn more visit **aiequalitytoolbox.com**

130

Introduction

In the cassava fields of Tororo, Uganda, a critical lesson about AI development was about to unfold—one that would challenge fundamental assumptions about how agricultural technology should be designed and deployed. When Daisy Salifu and her team arrived to scale an existing AI tool for cassava disease detection, they thought they understood the problem they were solving. The technology had been developed elsewhere, tested, and was ready for broader implementation. What they discovered through their "design by inclusion" approach would reshape their understanding of user-driven AI development.

The cassava farmers of Tororo had their own priorities. While researchers focused on early disease detection—a technically sophisticated solution that showcased AI capabilities—farmers were most concerned about soil analysis, nutrient management, and understanding which cassava varieties would thrive in their specific conditions. This misalignment between developer assumptions and user needs became a teachable moment that would influence AI development methodology across Africa.

Daisy Salifu's research, conducted as part of the broader AI for Development initiative, posed a fundamental question: "Could design by inclusion be a handed tool that can bring success in the integration of AI in agriculture?" The Uganda cassava project became a living laboratory for testing this hypothesis, revealing both the potential and the pitfalls of scaling AI solutions without genuine community involvement from the beginning.

The Challenge: When AI Solutions Miss the Mark

The cassava farming communities of Uganda represent the complexity of agricultural AI deployment in Africa. Women and resource-poor smallholder farmers make up more than half of Africa's farming population, yet they remain the lowest adopters of innovative agricultural technologies. This adoption gap isn't simply about access or education—it's fundamentally about relevance and inclusion in the design process.

When AI tools are developed without deliberate inclusion efforts, they can unintentionally deepen existing gender and social disparities. The Uganda project provided a clear example of this risk: an AI tool developed for disease detection was being scaled to communities whose primary concerns lay elsewhere in the agricultural value chain.

The existing AI tool had been developed through conventional agricultural technology approaches:

- **Top-down problem definition** by researchers and technical experts
- **Focus on technically sophisticated solutions** that demonstrated AI capabilities

- **Limited community input** during the initial development phase
- **Emphasis on scaling** proven technology rather than validating local relevance

This approach, while technically sound, missed the fundamental principle that effective AI must address the actual priorities of its intended users, not the assumed priorities of its developers.

Design by Inclusion:
A Methodological Innovation

The Uganda project became an opportunity to test a different approach: "design by inclusion," which Daisy Salifu defines as "developing technology to provide the best possible coverage of diversity within the user population." This methodology goes beyond simple consultation to create genuine participatory development where marginalized communities have agency in defining both problems and solutions.

Core Principles of Design by Inclusion

The approach encompasses several key principles that differentiate it from conventional AI development:

1. **Intentional engagement** with marginalized communities, including women smallholder farmers, people living with disabilities, and elderly farmers
2. **Active participation in design,** development, and deployment processes
3. **Recognition that grounded knowledge** or lived experiences of users is as valuable as expert technical knowledge
4. **Collaborative approach** that works from the ground up rather than top-down
5. **Safe space creation** for authentic participation from all community members

The Uganda Implementation

The Uganda project targeted cassava farmers in Tororo, taking advantage of an existing agricultural development initiative to test the design by inclusion methodology. The team intentionally included diverse farmer groups:

- **Women farmers** who form the majority of cassava producers
- **Men farmers** with different perspectives on agricultural priorities
- **Elderly farmers** with extensive traditional knowledge
- **Farmers with disabilities** whose needs are often overlooked in technology design

The methodology began with community dialogue designed to understand farmers' actual priorities rather than validating predetermined solutions.

The Community Dialogue Process

The heart of the design by inclusion approach was the community dialogue process, which created safe spaces for authentic participation from all farmer groups. This process revealed critical insights that would have been missed through conventional technology scaling approaches.

Creating Safe Environments

The team made several intentional decisions to ensure authentic participation:

- **Gender-Separated Groups:** Women farmers and male farmers were facilitated in separate sessions to address power dynamics and cultural constraints that might prevent women from speaking freely in mixed groups.

- **Same-Gender Facilitation:** Women's groups were led by female facilitators, men's groups by male facilitators, ensuring comfort and cultural appropriateness.

- **Recognition of Existing Knowledge:** The process began by acknowledging and documenting farmers' existing expertise in cassava cultivation, validating their knowledge before introducing new technological possibilities.

- **Collaborative Atmosphere:** Rather than presenting predetermined solutions, facilitators created space for farmers to articulate their own understanding of challenges and potential solutions.

The Critical Discovery: Misaligned Priorities

The community dialogue revealed a fundamental misalignment between the AI tool's focus and farmers' actual priorities:

- **Researcher Focus:** Early disease detection using AI image recognition
- **Farmer Priority #1:** Soil analysis to assess nutrients, examine suitable cassava varieties, and detect soil pathogens
- **Farmer Priority #2:** Pest and disease identification for timely intervention
- **Farmer Priority #3:** Market access and price management, including storage solutions and cooperative formation

This misalignment was particularly significant because it occurred at the scaling stage of a project that had already been developed and tested elsewhere. The farmers' top priority—soil analysis—wasn't addressed by the existing AI tool at all, while their second priority—pest and disease identification—was covered but wasn't their most urgent need.

Understanding the Full Agricultural Value Chain

The community dialogue process revealed that farmers think holistically about their agricultural challenges. They don't compartmentalize issues into discrete technical problems that can be solved by individual AI applications. Instead, they see interconnected challenges that require integrated solutions:

- **Soil Health and Variety Selection:** Farmers wanted to understand which cassava varieties would perform best in their specific soil conditions, requiring both soil analysis and variety recommendation systems.
- **Market Integration:** Even the most successful crop production is meaningless without market access and fair pricing, leading farmers to prioritize cooperative formation and storage solutions.
- **Holistic Pest Management:** Rather than focusing solely on disease detection, farmers wanted integrated pest management that included understanding soil conditions that might predispose crops to disease.

The Uganda project generated several crucial insights about effective AI development that

Lessons Learned: Critical Success Factors

goes beyond technical considerations to address social and institutional factors.

1. Understanding Priority Needs is Key

The most fundamental lesson was that AI tool design must align with users' highest priority needs, not developers' technical capabilities or interests. The misalignment discovered in Uganda demonstrates the risks of scaling AI tools without validating local relevance.

Implication for AI Development: Before any technical development begins, comprehensive community dialogue must establish what problems users actually want to solve, not what problems developers think need solving.

2. Farmers Have Diverse Knowledge

The community dialogue revealed that different farmer groups—women, men, youth, elderly—have different knowledge levels and different access to technology. This diversity is a strength that can enhance AI development when properly leveraged.

- **Women farmers** brought detailed knowledge about daily crop management, soil conditions, and household food security implications.
- **Young farmers** had different perspectives on technology adoption and were more willing to experiment with digital tools.
- **Elderly farmers** possessed deep traditional knowledge about varieties, soil management, and local climate patterns that could enhance AI training data.
- **Co-development that brings these diverse groups together** ensures that farmers can learn from

each other and benefit equitably, regardless of educational background or age.

3. Empowerment Through Collaboration

The process of sitting together, interacting, and discussing agricultural challenges empowers users by increasing their awareness of AI technology possibilities while validating their existing knowledge and expertise.

- **Increased AI Awareness:** Farmers gained understanding of how AI could potentially address their challenges, but in the context of their own priority-setting rather than predetermined technical solutions.
- **Value Recognition:** The process acknowledged that local knowledge is equal to expert knowledge, recognizing that farmers facing daily agricultural challenges possess crucial insights for AI development.
- **Community Relationship Building:** The collaborative process strengthened community relationships and built potential for cooperative formation—something farmers identified as important for accessing storage facilities and market power.

4. Gender-Sensitive Facilitation is Crucial

The separate, safe space approach proved essential for authentic participation from marginalized groups, particularly women farmers.

- **Safe Space Creation:** When women and marginalized groups are mixed with wealthier or more powerful farmers, inferiority complexes can prevent authentic participation. Separate facilitation addressed these power dynamics.
- **Authentic Expression:** In women-only groups, participants expressed themselves clearly about their needs and priorities in cassava production. This authentic expression was essential for understanding genuine user requirements.
- **Cultural Appropriateness:** Same-gender facilitation respected cultural norms while ensuring that all voices were heard in the design process.

5. Co-Development Works

The project demonstrated that farmers can effectively participate in all stages of AI tool development when given genuine agency in the process.

- **Feeling Valued and Heard:** Farmers who participate in co-development feel valued and heard, which increases their likelihood of adopting and adapting AI tools that emerge from the process.
- **Easier Adoption:** When farmers have been part of the development process, they more easily embrace and adopt tools because they understand how the tools address their own identified needs.
- **Sustainable Implementation:** Co-development creates ownership that extends beyond the initial deployment phase, supporting long-term sustainability and adaptation.

The Training Component: Co-Created Capacity Building

Following the community dialogue, the Uganda project included hands-on training on the existing AI tool, but even this training was co-created rather than predetermined. The training modules were developed together with farmers based on what they identified as their learning needs.

Collaborative Module Development

Rather than using standard training materials, the team worked with farmers to identify:

- **What they needed to learn** about smartphone use and AI tool operation
- **How they preferred to learn** through hands-on demonstration and peer teaching
- **What barriers they faced** in accessing and using digital agricultural tools
- **How the training could address** their specific context and capabilities

Feedback for Tool Improvement

The training process generated valuable feedback for improving the AI tool itself:

- **Interface design suggestions** based on farmer interaction with the technology
- **Feature requests** that would better serve farmer workflows
- **Technical adaptations** needed for local infrastructure and device capabilities
- **Integration possibilities** with farmers' existing agricultural practices

This feedback loop demonstrated how training can serve not just capacity building but also iterative tool improvement when farmers are treated as co-developers rather than passive recipients.

Methodological Innovation: A Scalable Framework

The Uganda project's most significant contribution was developing a replicable methodology for design by inclusion in AI development. Daisy Salifu and her team created a "Gender Equality and Social Inclusion Framework for AI Adoption in African Agriculture and Food Systems" that has been documented in academic literature and is being scaled across multiple contexts.

Framework Components

The framework includes several key components that can be adapted to different agricultural contexts:

1. **Community Entry Strategies** for building trust and establishing collaborative relationships
2. **Participatory Dialogue Methods** for authentic community engagement
3. **Safe Space Facilitation** techniques for including marginalized voices
4. **Priority Assessment Tools** for understanding user-defined needs
5. **Co-Development Processes** for involving communities in technical design
6. **Training and Capacity Building** approaches that build local ownership
7. **Monitoring and Evaluation** methods that measure empowerment and adoption

Academic and Policy Impact

The framework development resulted in a manuscript under review in the Journal of AI and Society, providing academic validation for the design by inclusion approach. This documentation ensures that the methodology can be replicated and adapted across different agricultural contexts and crop systems.

Current Limitations and Future Directions

The Uganda project team acknowledged several limitations that provide direction for future research and implementation:

Single Case Study Limitation

The project represents only one example of design by inclusion methodology applied to an AI tool at the scaling level. While it generated valuable insights, broader validation requires testing across multiple projects and development stages.

Recommendation: Replicate the design by inclusion methodology across diverse agricultural contexts and crop systems to strengthen the evidence base and refine the approach.

Scaling-Stage Intervention

The Uganda project involved an AI tool that was already developed and being scaled, rather than testing design by inclusion from the beginning of the AI development lifecycle. This limited the team's ability to demonstrate how the methodology might influence fundamental technical design decisions.

Future Direction: Apply design by inclusion methodology from the earliest stages of AI development to test its impact on technical architecture, model selection, and system requirements definition.

Context-Specific Adaptation

While the framework is designed to be scalable, each implementation requires adaptation to local cultural, social, and agricultural contexts. More research is needed on how to maintain methodological consistency while adapting to diverse contexts.

Comparative Analysis: Design by Inclusion vs. Conventional AI Development

The Uganda project provides a clear comparison between conventional AI scaling approaches and design by inclusion methodology:

Conventional Approach (Pre-Dialogue)	Design by Inclusion Approach (Post-Dialogue)
Problem Definition: Researchers identify disease detection as priority based on technical capabilities	**Problem Discovery:** Community dialogue reveals soil analysis as top farmer priority
Solution Development: AI tool developed for image-based disease recognition	**Solution Alignment:** Recognition that existing tool doesn't address primary user needs
Scaling Strategy: Deploy existing tool across multiple locations with standard training	**Adaptation Strategy:** Either modify existing tool or develop new solutions based on user priorities
Success Metrics: Adoption rates and technical performance indicators	**Success Metrics:** Community empowerment, relevance to user needs, and sustainable adoption

This comparison illustrates why design by inclusion requires more time and resources initially but may result in more effective and sustainable AI implementations.

Impact Beyond the Pilot: Influencing AI Development Practice

The Uganda cassava project's influence extends beyond its immediate implementation to impact broader discussions about AI development methodology in African agriculture.

- **Policy Influence:** The documented framework has informed policy discussions about agricultural technology development, emphasizing the need for user-centered approaches that go beyond technical considerations to address social inclusion and gender equity.
- **Academic Contribution:** The project has contributed to academic literature on participatory technology development, providing empirical evidence for the effectiveness of design by inclusion approaches in AI development.
- **Methodological Replication:** Other AI development initiatives across Africa are adapting the design by inclusion methodology, testing its applicability across different crops, technologies, and cultural contexts.

Lessons for AI and Human Rights

The Uganda cassava project offers several critical insights for AI development that respects and promotes human rights:

- **Inclusion Must Be Intentional**
Inclusion of marginalized communities in AI development doesn't happen by default. It requires deliberate methodology, resource allocation, and sustained commitment throughout the development process.

- **Local Knowledge is Valuable**
Farmers possess significant expertise that enhances AI tool effectiveness. This knowledge is not just useful for implementation—it's essential for defining what problems AI should solve and how solutions should be designed.

- **Process Matters as Much as Product**
The collaborative approach itself builds capacity and community coherence. The process of engaging communities in AI development has value beyond the technological outcomes.

- **Co-Development Creates Ownership**
When communities participate meaningfully in AI development, they feel valued and heard, leading to more sustainable adoption and adaptation of technological tools.

- **Mismatch Prevention Requires Early Engagement**
The most sophisticated AI tool fails if it doesn't address users' actual priorities. Early and ongoing community engagement is essential for ensuring that AI development serves genuine needs rather than developer assumptions.

Looking Forward:
Scaling Design by Inclusion

The Uganda cassava project demonstrates that design by inclusion can bridge the AI adoption gap among marginalized farmers through collaborative engagement. As this methodology scales across Africa, several key principles emerge for sustainable implementation:

- **Systematic Integration**
Design by inclusion must be integrated systematically into AI development processes, not added as an afterthought or optional component. This requires institutional commitment and resource allocation for community engagement throughout the development lifecycle.
- **Cultural Adaptation**
While the core principles of design by inclusion are transferable, implementation must be adapted to local cultural, social, and agricultural contexts. This requires local expertise and sustained community relationships.

- **Capacity Building**
Successful scaling requires building capacity among AI developers, researchers, and partner organizations to facilitate authentic community engagement and manage participatory development processes.

- **Evidence Building**
Continued documentation and evaluation of design by inclusion implementations will strengthen the evidence base and support adoption by academic institutions, funding organizations, and policy makers.

The cassava farmers of Tororo continue their agricultural work, but their participation in this project has influenced how AI development approaches community engagement across Africa. Their voices, initially misaligned with the existing AI tool, have become part of a growing movement toward more inclusive and effective agricultural technology development. The soil they tend—the subject of their highest priority need—remains at the center of their agricultural concerns, reminding AI developers that effective technology must grow from the ground up, rooted in the actual needs and knowledge of those who will use it.

Mapping the AI Lifecycle HRIA Framework for the Uganda Cassava Initiative

1 Stage 1: Objective and Team Composition

The project began as a scaling initiative for an existing AI tool but evolved into a test of design by inclusion methodology. Through community dialogue, the team discovered fundamental misalignment between predetermined objectives (disease detection) and community priorities (soil analysis). This revelation prompted a reconceptualization of both objectives and team composition.

HRIA Framework Alignment:

- **Purpose & Context:** The project revealed how scaling AI without community input can perpetuate exclusion of marginalized farmers, particularly women and resource-poor smallholders.
- **Effects of the System:** The existing tool benefited technically sophisticated users but missed the primary needs of intended beneficiaries, demonstrating how AI can unintentionally deepen disparities.
- **Empowering Affected Communities:** The design by inclusion approach gave farmers genuine agency to redefine the problem and assess whether existing solutions served their needs.
- **Team Composition:** The team included diverse farmer groups (women, men, elderly, disabled) as legitimate experts whose knowledge was valued equally with technical expertise.

Key Human Rights Considerations:

The project highlighted how predetermined objectives can violate the principle of meaningful participation. True human rights alignment requires communities to have agency in defining what problems AI should solve, not just how to implement predetermined solutions.

2 Stage 2: Defining System Requirements

The community dialogue process revealed that system requirements must emerge from user-identified priorities rather than technical capabilities. Farmers' requirements centered on integrated agricultural support: soil analysis for variety selection, market access solutions, and storage facilities through cooperative formation.

HRIA Framework Alignment:

- **Involving Affected Communities**: Requirements definition involved extensive community consultation with intentional inclusion of marginalized groups through safe space facilitation.
- **Explainability Considerations:** The system needed to provide explanations relevant to farmers' actual decision-making processes—soil health, variety selection, market timing—rather than disease identification alone.

- **Ecosystem of Values:** The initiative revealed tensions between technical sophistication (disease detection accuracy) and user relevance (soil analysis for production decisions), requiring conscious prioritization of user needs.

Key Human Rights Considerations:
Requirements must reflect user dignity and agency. The Uganda project showed how technically impressive requirements (AI disease detection) can miss fundamental human needs (soil health, food security, economic viability) if not grounded in community priorities.

3 Stage 3: Data Discovery

The project revealed that existing AI training data, while technically valid, didn't address farmers' priority needs. Data discovery needed to encompass soil health, variety performance, and market information—areas not covered by the disease detection focus.

HRIA Framework Alignment:
- **Data Origin:** The existing tool's training data was collected without input from Ugandan farmers, missing local soil conditions, variety preferences, and agricultural practices.
- **Data Bias:** The focus on disease detection reflected researcher priorities rather than farmer needs, representing a form of bias that marginalized user knowledge and priorities.
- **Documentation:** The project documented the misalignment between existing data and user needs, providing evidence for more inclusive data collection approaches.

Key Human Rights Considerations:
Data collection must reflect user priorities and contexts. The Uganda case demonstrates how technically sound data can still be inadequate if it doesn't address the problems communities actually face.

4 Stage 4: Selecting and Developing a Model

The existing model was technically sophisticated but addressed the wrong problem from the farmers' perspectives. The project revealed the need for models that integrate soil analysis, variety recommendation, and market information rather than focusing solely on disease detection.

HRIA Framework Alignment:
- **Model Type and Explainability:** The disease detection model was explainable but irrelevant to farmers' top priorities, demonstrating that explainability must address users' actual decision-making needs.
- **Fairness Aspects:** The model was unfair in that it addressed problems identified by researchers rather than the diverse needs of different farmer groups (women's soil concerns, youth's market interests, elderly farmers' variety knowledge).

- **Environmental Impact:** Model development resources were misallocated toward technically impressive but less relevant capabilities.

Key Human Rights Considerations:
Model selection must serve user empowerment rather than technical demonstration. The Uganda project shows how sophisticated AI can still violate human dignity if it doesn't address genuine needs.

5 Stage 5: Testing and Interpreting Outcome

Testing revealed the fundamental misalignment between tool capabilities and user needs. Community feedback showed that while the disease detection tool worked technically, it didn't address farmers' primary concerns about soil health and variety selection.

HRIA Framework Alignment:
- **Testing Context and Outcomes:** Testing occurred with actual intended users (cassava farmers) who provided authentic feedback about relevance and utility.
- **Operation Manual:** Training materials were co-created with farmers, but the training revealed that even well-designed capacity building couldn't overcome fundamental misalignment between tool capabilities and user needs.

Key Human Rights Considerations:
Testing must evaluate whether AI genuinely empowers users to address their identified priorities. Technical functionality is insufficient if the system doesn't serve human dignity and agency.

6 Stage 6: Deployment & Post-Deployment Monitoring

The project demonstrated that successful deployment requires alignment between tool capabilities and user priorities from the beginning. Even excellent community engagement and training cannot overcome fundamental misalignment in problem definition.

HRIA Framework Alignment:
- **Deployment:** The community had agency to assess the tool's relevance to their needs and provide feedback about its limitations, demonstrating genuine participation in evaluation.
- **Monitoring:** The project monitored not just technical performance but community assessment of relevance and utility, leading to insights about the need for different AI solutions.

Key Human Rights Considerations:
Deployment must serve community empowerment rather than technology adoption for its own sake. The Uganda project demonstrates that communities must have the right to reject AI solutions that don't serve their identified needs.

Integrated Analysis: Design by Inclusion Throughout the AI Lifecycle

The Uganda cassava project demonstrates several critical principles for human rights-aligned AI development:

- **User Priority Definition:** Communities must have agency to define what problems AI should solve, not just how to implement predetermined technical solutions.

- **Authentic Participation:** Meaningful participation requires safe spaces, cultural appropriateness, and recognition that local knowledge is as valuable as technical expertise.

- **Relevance Over Sophistication:** Technical sophistication is meaningless if AI doesn't address users' actual priorities and decision-making needs.

- **Early Engagement:** Community engagement must begin at problem definition, not just implementation. Late-stage participation cannot overcome fundamental misalignment.
- **Continuous Adaptation:** AI development must be responsive to community feedback throughout the lifecycle, including the possibility that existing solutions may need fundamental reconceptualization.

- **Empowerment Metrics:** Success must be measured by community empowerment and relevance to user needs, not just technical performance or adoption rates.

The Uganda experience provides a crucial counter-narrative to conventional AI scaling approaches, demonstrating that technical success is insufficient without human rights alignment throughout the development lifecycle.

About the case study and author

This case study analyzes research conducted by the African Technology Policy Studies Network (ATPS), in collaboration with the International Centre of Insect Physiology and Technology (icipe), Kenya and Kumazi Hive (Ghana) that focused on Strengthening the Capacity of Women and Marginalized Communities in Africa's Agriculture and Food Systems to Harness the Potentials of Artificial Intelligence Technology in alliance with Artificial Intelligence for Agriculture and Food Systems (AI4AFS) project in Uganda titled "Scaling Smartphone-Based Tools for Early Crop Diseases Detection and Monitoring" led by Dr. Godliver Owomugisha of Busitema University, Uganda, with partners Community Development Foundation, Uganda, and Nyakasozi Tukooreamwe Coffee Farmers' Co-operative Society Limited, Uganda, between 2022–2024

Dr Daisy Salifu is a Biostatistician with over a decade of experience, currently serving as the Head of the Biostatistics Section within the Data Management, Modelling, and Geo-Information (DMMG) Unit at the International Centre of Insect Physiology and Ecology (icipe) an international scientific research institute, headquartered in Nairobi, Kenya that works towards improving lives and livelihoods of people in Africa. Her career has been marked by a commitment to advancing scientific research, driving impactful outcomes, and fostering the growth of the next generation of scientists.

Other contributors to this case study are Caitlin Kraft-Buchman, Emma Kallina, and Sofia Kypraiou, authors of the original *Framework to AI Development: Integrating Human Rights Considerations Along the AI Lifecycle* upon which the Toolbox structure is based. Additional contributors are Amina Soulimani and Pilar Grant, from Women at the Table and the <AI & Equality> Human Rights Initiative.