Caribou Digital

IDRC · CRDI
Canada

UK International Development
Partnership | Progress | Prosperity

ARTIFICIAL INTELLIGENCE FOR DEVELOPMENT AFRICA

Case Studies in the Practice of Responsible AI for Development

Examples from agriculture, education, health, and human rights

Published September 2024

For videos and more details about the report please visit https://bit.ly/responsibleAI4D.

Authors

This report was written by Gratiana Fu with contributions by Jonathan Donner.

Acknowledgments

We are grateful to the International Development Research Centre (IDRC) for supporting this work, to Laurent Elder and Matthew Smith for suggestions and engagement thought the research process, to Mia Foster and Abubakar Nuuman at Story × Design for producing the videos, to Shelby Brewster for editing, and particularly to each of the participants for sharing their experiences and perspectives with the community of practice. Thank you.

Recommended Citation

Fu, Gratiana. "Case Studies in the Practice of Responsible AI for Development". Farnham, Surrey, UK: Caribou Digital Publishing, September 2024. https://www.cariboudigital.net/publication/case-studies-in-responsible-ai-for-development.

Caribou Digital delivers fund management, learning partnerships, and research, advisory, and evaluation services, supporting organizations worldwide to build more inclusive and ethical digital economies.

www.cariboudigital.net

This work was carried out with the aid of a grant from the International Development Research Centre, Ottawa, Canada. The views expressed herein do not necessarily represent those of IDRC or its Board of Governors. Quotes have been edited only for clarity and brevity.

Acessibility

This report has been designed using Adobe's PDF accessibility evaluation tool, a checklist of considerations to align to Web Content Accessibility Guidelines (WCAG) 2.1 Level AA.

Contents

List of abbreviations 4

Executive summary 5

Introduction 11

askNivi: A retrieval-based chatbot improving access to local healthcare 17

HURIDOCS: An AI-optimized database lowering barriers to information for human rights defenders 28

PROMPTS: A messaging platform strengthening primary healthcare in Kenya 36

EIDU: An AI-powered learning platform using content personalization to maximize learning in Kenya 43

Farmer.Chat: An AI-powered digital assistant supporting agricultural extension workers around the world 52

RobotsMali: An education project using LLMs to encourage national language literacy in Mali 60

ACADIC: Predictive AI tools implementing new ways to track disease spread on the African continent 68

Plantix: A mobile app employing image recognition to improve farmers' productivity in India 76

Discussion: From responsible practice, impact 86

References 99

Appendix 105

List of abbreviations

ACADIC	Africa-Canada AI & Data Innovation Consortium AI artificial intelligence
AI4D	AI for development
CSOs	civil society organizations
DNN	deep neural network
DPG	digital public good
DPI	digital public infrastructure
HURIDOCS	Human Rights Information and Documentation Systems
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
IVR	interactive voice response
LLM	large language model
LSTM	long short-term memory
ML	machine learning
NHRIs	national human rights institutions
OHCHR	Office of the United Nations High Commissioner for Human Rights
RAG	Retrieval-Augmented Generation
RCT	randomized controlled trial
REL-AI4GS	Responsible, Explainable, and Local Artificial Intelligence for Clinical Public and Global Health in the Global South
RNN	recurrent neural network
SBMM	skill-based matchmaking
UPR	Universal Periodic Review

Executive summary

Introduction

This project, supported by Canada's International Development Research Centre (IDRC), documents promising use cases of responsible artificial intelligence (AI) in international development. Drawing on interviews with leaders across eight different organizations in health, education, agriculture, and human rights, the study explores distinct implementation of AI for development: 1) askNivi, a chatbot for health; 2) HURIDOCS/UPR Info, an AI-optimized human rights database; 3) Jacaranda Health's PROMPTS, an AI message triage tool; 4) EIDU's message personalization for Kenyan students; 5) Digital Green's Farmer Chat; 6) RobotsMali's project using AI to create local language school books; 7) ACADIC, using machine learning to track COVID-19 in South Africa; and 8) ICRISAT's Plantix, an image recognition app for agriculture.

Our goal was to engage with projects that have implemented discrete AI use cases resulting in measurable impact and to learn from their practices, successes, and challenges, particularly around the responsible use of AI. These cases involve both predictive and generative AI applications.

There is enthusiasm about the potential impacts of AI on global development practice, though evidence remains relatively scarce and fragmented. Meanwhile, voices from policymakers, academics, researchers, civil society activists, technologists, and industry have been raising urgent concerns about the potential harms and risks associated with the spread of AI.

Both the impact and responsibility discussions around AI often deal with abstractions and aggregations. However, it is helpful to examine emerging practices within organizations from a bottom-up perspective. The work of using AI responsibly is done by individuals within organizations; their choices drive progress toward sustainable development goals.

We frame this contribution to the growing conversation on responsible AI and development through two primary questions:

1. How are organizations approaching AI, *responsibly*?
2. How do development organizations evaluate the impact of AI?

To conduct the case studies, we recruited a heterogeneous set of organizations, each using AI to help them pursue their development goals. Speaking with at least two individuals in each organization, we conducted interviews that focused on:

1. The problems the organizations are trying to solve with AI,
2. The intended impacts,
3. How the organizations measure impact, and
4. How they are engaging with ideas of responsibility in AI.

Several of the organizations agreed to participate in a video documentary, using self-shot video to describe the practices around responsible AI use for development of practice.

Crosscutting findings: From responsible practice, impact

Our conversations with the organizations illustrate a variety of ways in which they have created guiding practices to ensure they are engaging with AI responsibly, including:

1. **Do no harm (go slowly).**
   Many organizations are taking a slow approach to using AI and have strict guidelines for integrating AI into their products.
2. **Keep humans in the loop.**
   Keeping humans in the loop means not letting machines make decisions without oversight and guidance. Several of the organizations described how important this practice was to them. Some referenced it as the first precondition to approach responsible AI.
3. **Pursue fairness for all.**
   Inclusivity in AI requires careful attention to increasing access, reducing bias, and treating users with respect.
4. **Share with the community of practice.**
   Many of our interviewees described their engagements with the community of practice, sharing best practices, code, and datasets to promote transparency and accountability.

Creating responsible practices

These case studies also reveal three ways in which the organizations derived their practices of responsible AI.

1 **Approaches to responsible AI should be bespoke.**
Every organization interviewed had a unique set of practices. None explicitly connected their responsible AI practices to the broader principles of digital development, nor to any of the principles of responsible AI currently circulating in the policy community, though there is overlap. There is work still to be done to increase the salience and actionability of emerging global principles on AI.

2 **Responsible AI practices depend on an organization's position in the stack.**
Within institutions, various decision-makers engaged with responsible AI practices. Some technical and sector experts had previous AI experience and a clear vision of their goals and the necessary practices for responsible implementation. Others were deploying AI for the first time. Some groups were close to users with open communication channels, while others worked with frontline intermediaries deploying the AI.

3 **Practitioners can share practices with the broader community.**
These eight cases offer encouraging evidence of sharing of practices beyond the confines of each organization, promoting transparency across the sectors and the value chains that structure how end users and organizations interact with these powerful technologies.

On impact

The eight organizations report delivering impact across at least a dozen of the UN SDGs, using a variety of approaches. The cases demonstrated three kinds of impact:

1 **Increasing scale and efficiency**
Most organizations described how AI was being deployed, responsibly and carefully, to help the people in the organizations do their jobs more quickly and effectively, whether by increasing the number of queries a staff member could process or reducing team members' response times to emergency messages. The trends in AI's contribution to productivity are encouraging but not yet crystal clear, nor are they irrefutable or well-documented within the information-intensive parts of the development sector, as represented by the organizations in this study.

2 **Improving outcomes**
Some organizations described improved outcomes for individuals. These come via improvements to interfaces, driving engagement and accessibility across language and literacy divides, or via building AI directly into the delivery of an intervention, such as better educational materials in the right languages, and identification of plant diseases that could not be made via mobile phone without the use of AI.

3 **Strengthening community**
Each organization can make a choice: to go it alone, quietly, or participate in a broader community of practice. As evidenced by (at a minimum) participation in this study, and more broadly, in practices reported by many of the organizations. The community is getting stronger all the time.

Strengthening impact practices

The cases also highlight three challenges to strengthening impact measurement practices around AI for development.

1 **It is still early (for generative AI).**
The organizations using generative AI that we have talked to are still in the early stages and have not yet had the time to evaluate the full potential of AI-generated content for their organizations or for their users.

2 **Impact costs money to measure.**
We did not specifically ask case study participants about the resources they were dedicating to measuring AI impacts, but it is clear that documented, quantifiable impact evidence is not inexpensive.

3 **Organizations with direct connections to users or communities can measure impact more easily.**
Interventions leading to AI impact involve multiple layers—from LLM providers and open-source ML algorithm publishers, through AI-specific companies and startups, to implementing partners and frontline organizations. Achieving AI impact may require partnerships among these originators and their customers or implementing organizations, as those with frontline access are often well-positioned but under-resourced to carefully track the impacts of new AI interventions.

For all the reasons above, rigorous, quantitative evidence on the real-time impact of AI interventions will remain precious and rare. While reliance on A/B testing and frequent adjustments is understandable, it may not isolate impact evidence as desired by the development community. This, we argue, is a key moment for the community to support evidence gathering. We are at a transition point where many organizations are integrating AI or modifying workflows to include AI, affecting longstanding processes and theories of change.

Conclusion

In summary, these case studies—representing organizations across various sectors—reflect change underway, not change completed.

The work on responsible AI in practice reveals emerging strategies to help organizations navigate the power and vulnerabilities of opaque, probabilistic AI technologies. A key opportunity is to strengthen the dialogue and exchange between implementing organizations and the global bodies publishing general principles of responsible AI.

The work on impact in practice underscores how most organizations are confident they are driving impact, using A/B testing and other forms of measurement to refine their processes. This unique moment allows them to recall and contrast their processes before using AI, although a key challenge remains in translating confidence and qualitative data into quantitative, generalizable evidence.

Within these challenges lie opportunities. Resolution requires greater community and connection across the value chain and within the broader AI development community. The actions of implementing organizations—contributing to DPIs, DPGs, and best practices—shape the emerging political economy of AI, especially as it intersects with the development sector. We're grateful for the eight organizations here and their contributions to this community.

Summary of findings and case studies

Case	Project	Impact of AI	Impact measurement practices	Responsible AI practices	SDGs
askNivi (Nivi)	Retrieval-based chatbot, improving access to local healthcare	Increased user engagement, contributing to increased access and use of healthcare services	Gender audits	Human Evaluations Equitable outcomes	3.7 3.8 5.6
HURIDOCS/UPR Info	An AI-optimized database lowering barriers to information for human rights defenders	Increased team efficiency, from 2 months to 2 weeks	Tracking tool's impact from both technical and social perspectives	Human checks Redundancy	16.3 16.6 16.7
PROMPTS (Jacaranda Health)	An AI-powered message triage tool strengthening primary healthcare in Kenya	Increased help desk team efficiency, improved access to critical care	Qualitative feedback from staff	6 things/ principles Humans in the loop	3.7 3.8
EIDU	Using content personalization to maximize learning in Kenya	Significant improvement in learning outcomes (SD=0.43) determined from RCT	A/B testing	Test for fairness/bias Beta testing	4.1 4.2
Farmer.Chat (Digital Green)	A generative AI chatbot supporting agricultural extension workers around the world	More effective, informative interfaces available to 13,000 users and extension agents	Focus on the farmer	3 principles: transparency, privacy, inclusivity	2.3 2.4
RobotsMali	Using ChatGPT to create locally relevant school books, encouraging local language literacy in Mali	Increased engagement with learning materials created with AI	Training sessions Qualitative feedback	Humans in the loop and extensive discussion of how to use AI for equitable outcomes	4.1 4.2
ACADIC	Machine learning to track COVID-19 in South Africa	Speed of product creation, reach of products	Engagement numbers	Published REL4AI framework	3.b 3.d
Plantix/ICRISAT	Improving farmers' productivity in India with image recognition	Over 10 million users Small-scale evidence of improvement in agronomic knowledge	Engagement and retention numbers Qualitative in-app feedback One-off studies on user impact	CGIAR's Research Code of Ethics	2.3 2.4

Introduction

Technology never stands still. For as long as there has been a development community, there have been imperatives within it to adapt to each major wave of technological innovation. Telephony, rural radio, personal computers, satellites, the internet, the Web, mobile phones, platforms, big data, etc. Each wave enabled new ways of organizing and acting. Each wave restructured social and economic relationships. And each wave prompted changes in the way development can be done.[1]

The technologies and systems collectively labelled "artificial intelligence" or "AI" are another such wave. As such, there is an urgent, exciting need to grasp the opportunities—and better understand the risks—associated with their use in development practice.

Thanks to improvements in the scale, accuracy, and interface accessibility of large language models (LLMs) via tools like OpenAI's ChatGPT and Google's Gemini, the awareness and use of AI tools in professional workflows has emerged from many years of relatively invisible development in research labs and specialized settings into the professional mainstream.

1 Heeks, "ICT4D 3.0?"

AI impacts in development

Like their colleagues in journalism, the creative industries, the sciences, engineering, education, software development, and so on, professionals in development organizations are focusing more intently on how AI tools and systems will change what they can accomplish.[2] Yet just how rapid, widespread, and disruptive these changes will be for the development community remains to be determined.

As attention to and usage of AI has increased, so has documentation of applications of AI in development. There are communities of practice and portfolios of support, such as the AfricAI conference organized by IDRC, GIZ, and Niyel.[3] IDRC is already a leading voice in this discussion, both through advocacy and thought leadership.[4] Strathmore's AI project includes both sectoral analyses and a map of over 280 distinct applications of AI active in Africa in 2023.[5] Key reports from the Schwab Foundation and the GSMA each detail 90+ cases studies in AI for social innovation and development, respectively.[6] Others take more theoretical approaches, digging into productivity and labour rates to contemplate AI's eventual contribution to or disruption of economic development.[7]

Responsible AI in development

Yet in some critical ways, this wave, this moment of exploration, is unfolding differently from the waves that have preceded it. A range of voices—policymakers, academics, researchers, civil society activists, technologists, and actors within industry—are raising broad and urgent concerns about the potential harms and risks associated with the spread of artificial intelligence. These risks come in a variety of forms and time horizons, from concrete cases of error and bias, to longer-range concerns about disruptions to economies and societies, to concerns about the undermining of truth and a spread of misinformation, to even more existential concerns about whether some forms of artificial intelligence could inflict national or planetary harm on humanity.

2 Björkegren, "Artificial Intelligence for the Poor."
3 AfricAI, "AfricAI Conference 2023."
4 Smith, "Ensuring Generative AI Is Responsible AI."
5 Centre for Intellectual and Property Law, "Mapping of AI Applications in Africa."
6 GSMA, "AI for Africa: Use Cases Delivering Impact."
7 Korinek and Stiglitz, "Artificial Intelligence, Globalization, and Strategies for Economic Development"; Aly, "Digital Transformation, Development and Productivity in Developing Countries."

This awareness of risks invokes hotly contested conversations about how to balance those risks while utilizing these technologies. Perspectives range between optimism and concern, between progress and responsibility. And there are "big picture" prescriptions under development: "global principles" for responsible AI, such as the OECD AI Principles, G20 AI Guidelines, UNESCO's Recommendation on the Ethics of Artificial Intelligence, the UN Resolution on Artificial Intelligence, the UK's Responsible AI, and EU AI Act 22.[8]

There are also regional diagnoses that frame AI as presenting a new set of opportunities and challenges confronting regions and countries throughout the majority world. Papers here focus on critical gaps in skills, training, regulatory safeguards, and infrastructure,[9] as well as the strength of the data ecosystem.[10] The Global Index on Responsible AI works in this vein, highlighting progress in AI capacity, human rights elements, and governance across more than 140 countries.[11]

Case studies connecting responsibility and impact

Both the impact and responsibility conversations often work with abstractions and aggregations. However, a "bottom-up" approach that examines individual organizations engaged in making choices about how to deploy AI in development contexts is helpful. For example, Cooper/Smith highlights specific techniques for using AI in the health sector, including guardrails, UX iteration, and "red-teaming" testing in "the real world."[12]

This study takes a bottom-up approach. By engaging closely with organizations working across a wide array of development sectors, including education, agriculture, and health, we can generate insights and crosscutting lessons to strengthen the sector as a whole, and inform its continued adoption of AI. **We look at emerging practices in organizations. The hard work of coming to use AI responsibly is done by people, working in organizations, making choices and seeking to drive progress towards development goals.** We frame this contribution to the growing conversation on responsible AI and development via two primary questions:

8 Thank you to Justine Gluck at the Fletcher School for the idea to list these kinds of principles as a block. See OECD.AI Policy Observatory, "AI Principles Overview"; UNESCO, "Recommendation on the Ethics of Artificial Intelligence"; UN General Assembly, "Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development"; RAi UK, "Guiding Principles – Responsible AI"; EU Artificial Intelligence Act, "Article 22."

9 Ade-Ibijola and Okonkwo, "Artificial Intelligence in Africa."

10 Okolo, Aruleba, and Obaido, "Responsible AI in Africa—Challenges and Opportunities."

11 Adams, R. et al., "Global Index on Responsible AI 2024 (1st Edition)."

12 Cooper/Smith, "Deploying GenAI in the Real World."

How are organizations approaching AI, *responsibly*?

All technological advancements, from clocks to printing presses to telegraphs to cell phone cameras, come with a new set of dynamics around power. They require their users, in various organizational and cultural contexts, to balance new possibilities with new challenges.

This is not the first time the world has been asked to confront a technology responsibly. But it is a particularly important moment; many believe that AI will usher in a step change in the relationship between people, organizations, and the generation of knowledge and meaning. The ways AI can go wrong are as terrifying as the ways that it can go right are encouraging.

The broad community of practice emerging around AI ethics has a lot to do, and one of the places to look for whether these principles and processes will be up to the challenge is in the ways organizations like those in this study approach the idea of responsible AI use.

How do development organizations approach impact?

A related inquiry involves the impacts that are emerging. The straightforward question is one of presence/absence: How and to what extent is the use of AI in development initiatives potentially having any impact, and for whom? Such programmatic impact questions have been relatively scarce in the literature to date.

We are interested not only in pointing to impact, but to the culture and practices of *impact measurement*. It is a perennial challenge in digital development interventions to develop theories of change, create the culture to test and measure against those theories of change, and to change organizational practices in response to that data. Throughout these case studies we engage with people at every level of these organizations to understand the hard work of propagating and reinforcing this culture of change.

Methods

To conduct the case studies, we recruited a heterogeneous set of organizations, each using AI to help them pursue their development goals. These represent a range of *sectors* (health, climate, education, agriculture, etc.), *regions*, and implicit or explicit *development outcomes*.

Fully populating this matrix of domains × regions × outcomes with case studies of applications of AI would require dozens of cases and stretch beyond the resources of this project. Instead, we engaged with IDRC in interactive, careful discussions to arrive at an illustrative sample of AI applications across several domains and regions, and with as inclusive an approach to implied development outcomes as possible.

Active definitions

We used a wide definition of "AI" to include both predictive AI (applications of machine learning and automation, some with longer implementation track records) and generative AI (the LLMs and generative/creative/conversational advances that only hit the mainstream in 2023).

We did not predetermine definitions of "impact" or "responsibility." Those frames are left to participants to define and share with us. Indeed, this is not a study in which it made sense to arrive with a predetermined set of criteria or principles of "responsible," evaluating each organization against them. We instead left the question open to participants to share with us. The gaps between how the organizations describe responsible AI and how the various global principles might describe them is important to document.

Impact, too, was left open for discussion. Impact is too often under-articulated by programs who are trying to do the right thing, changing frequently to find fit. In other cases, the intended impact is clear, but processes have not been put in place to measure that impact. One resolution for this dilemma might have been to only select for case studies those "AI for Development" (AI4D) programs that have a clear theory of change and a solid plan to measure those impacts in place. But there is considerable value in talking about theories of impact and assessments of impact, even if only indirectly, for the topic must be kept at the core of the development community's engagement with AI.

Interviews and background research

We conducted interviews and discussions that focused on the following questions. The rough discussion guide is contained in the appendix.

1. What problems are the organizations are trying to solve with AI?
2. What are the intended impacts?
3. How do the organizations measure/know they are having impact?
4. How are they engaging with ideas of responsibility in AI?

For each case, at least two individuals involved in the program were interviewed. These were not adversarial, journalistic, or even evaluative interviews; they were, rather, consultations and inquiries. Each respondent was informed of the general idea for the study (centring on the themes of impact and responsibility). Where possible, discussions were supplemented with secondary research.

Video

Several of the organizations agreed to participate in a video documentary using self-shot video to engage with and describe the practices around responsible AI4D. Participants have reviewed their direct quotes in text and the images portrayed from them on video.[13] They should be seen as members of this emerging community of practice rather than interview subjects.

We are grateful for their time and for their willingness to speak frankly and share widely vis à vis these important issues.

13 Videos are available at https://bit.ly/responsibleAI4D.

askNivi: A retrieval-based chatbot improving access to local healthcare

Digital health company Nivi's chatbot, askNivi, provides relevant healthcare information to users and connects them with nearby trusted healthcare providers for continued care. askNivi is a retrieval-based bot that uses the interpretive and categorization power of LLMs to provide users with tailored content. Operating from a state of caution about using generative AI for healthcare, the Nivi team has leveraged generative AI in a way that doesn't expose users directly to generated content, and the risks associated with that. They have deployed this feature responsibly, keeping humans in the loop and proceeding slowly.

Background

Nivi is a digital health company dedicated to creating a safe environment for sharing and accessing healthcare information and services.[14] Now run by an almost twenty-person team, the organization started as a research project by Population Council, a research group dedicated to improving the health and well-being of underserved communities globally.[15] The project brought together technologist Siddhartha Goyal, behaviour science psychologist Dr. Eric Green, and public health and development specialist Dr. Ben Bellows.

Years ago, Nivi's team saw a need for a scalable and trusted method to disseminate health information, a digital mechanism for connecting individuals to healthcare information and safe healthcare providers in environments where seeking that information publicly was not easy or safe.

14 "Nivi | AI Powered Chatbot for Healthcare."
15 Population Council, "Home."

They originally developed an automated interactive voice response (IVR) system that people could use to get answers to routine healthcare questions. However, Nivi soon recognized that IVR could not handle the information that people were asking about. So they switched to a text-based platform, now called askNivi.

"[We] started with an IVR platform. We realised that the amount of information that [users] were trying to give and specific types of information like care referrals don't actually translate that well in audio. [Users] want to be able to revisit what they've learned to be able to act on it."

David Tresner-Kirsch, Nivi CTO

Today, askNivi is an end-to-end chatbot that connects patients, providers, and healthcare institutions with health information, starting on SMS and now primarily available on WhatsApp. The tool is operating in India, Kenya, and Nigeria and has reached over 1.5 million registered users. The majority of Nivi's target communities are already actively using WhatsApp, so capitalizing on an existing tool maximizes the number of people that can interact with askNivi. The Nivi team also keeps a record of historical interactions, enabling users to return to previous askNivi conversations and continue where they left off, a feature that original users of the IVR system were missing. askNivi can also provide referrals to trusted healthcare providers and product offerings. Users who decide to act on the information askNivi provides are matched with up to three nearby physicians and up to three digital resources where they can receive consultations, purchase health products, and more. By offering both in-person and online services, askNivi gives users the autonomy to choose a path forward that they are most comfortable with.

Currently, askNivi is a retrieval-based chatbot that uses LLMs to interpret and categorize messages from users and return curated content. It provides users with healthcare content created by experts on a small set of healthcare topics. While this method limits the number of topics that users can get information about, it meets users where they are. A large majority of askNivi users are multilingual and code switch between languages; those who are not multilingual often speak languages for which there are few available digital language resources. By offering a set number of targeted health topics in select languages, Nivi is able to concentrate its efforts on providing high-quality, reliable, and appropriate information to its users. They currently offer content about select health topics, like reproductive health, maternal health, HIV and AIDS, and nutrition, in a variety of languages, including English, Swahili, and Hindi.

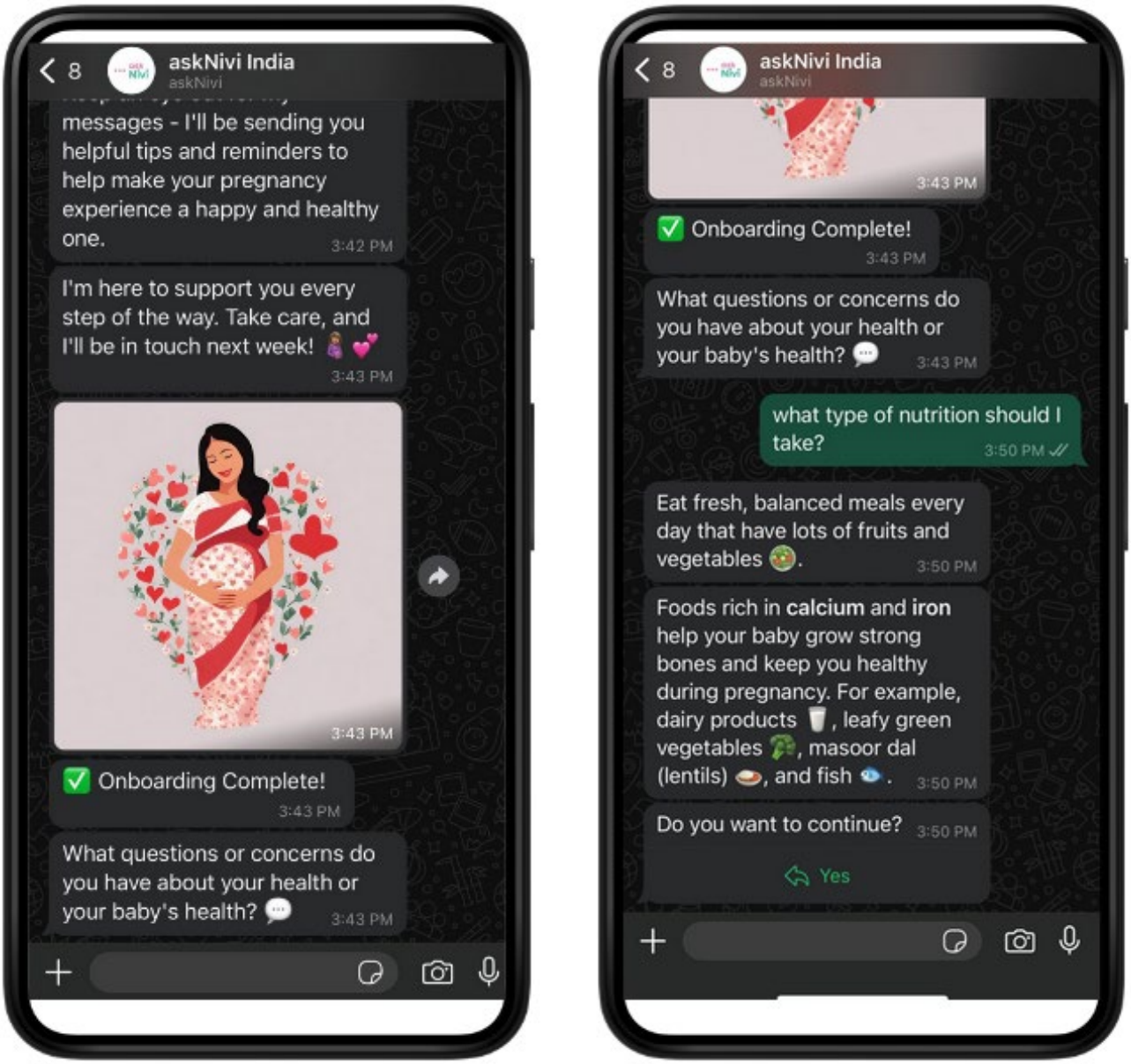


Figure 1 ▲
A WhatsApp exchange with askNivi
Source: askNivi

Integrating AI to improve askNivi

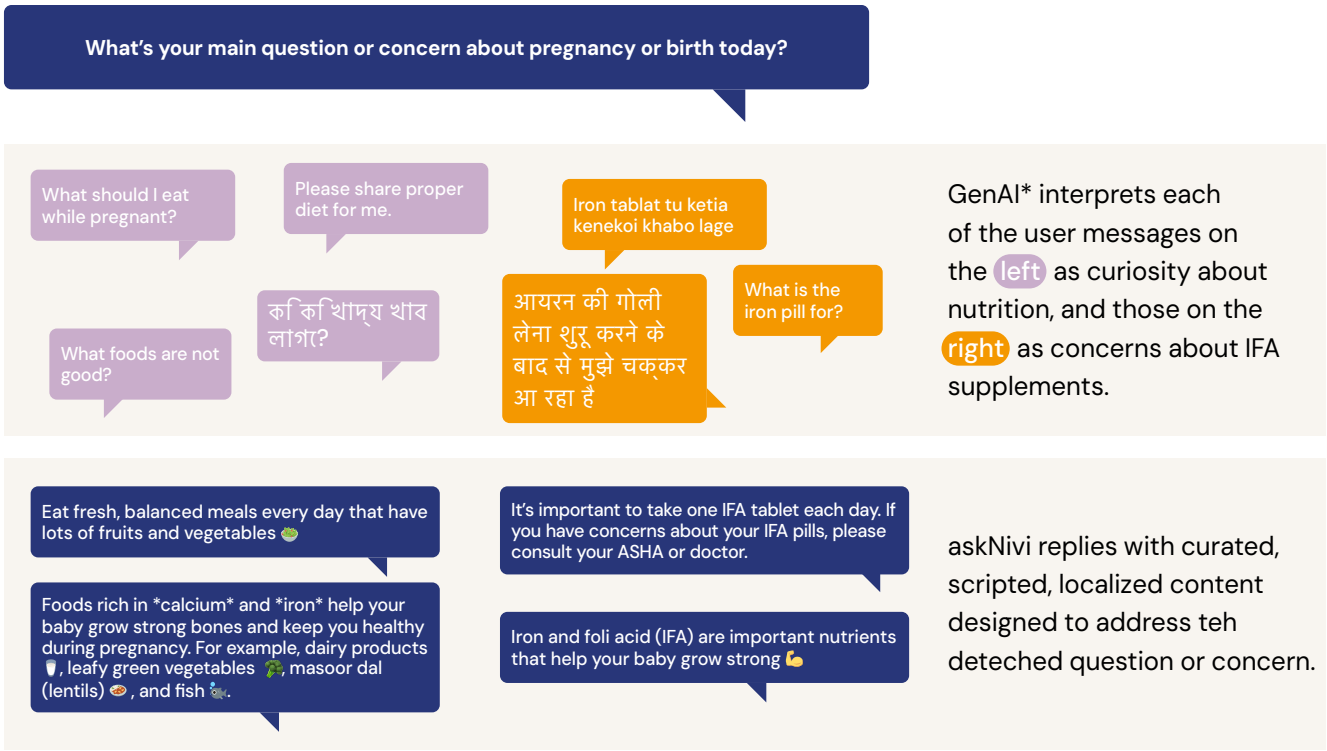
The askNivi tool receives hundreds of unprompted natural language messages each day from users who want to interact with it in a different way. Through in-depth analysis of conversations carried out in askNivi, Nivi found that a large percentage of users disengaged from the platform when they tried to send messages that were not associated with a preset menu option. For example, users would send greeting messages, like "Hello!" or "Good morning," or free-text health questions. The bot would, in turn, respond with error messages (i.e., "I'm only a bot. Please talk to me in numbers or button presses."). The tool's inability to handle natural language was contributing to lower engagement and became a barrier to connecting users with appropriate healthcare.

"The primary way we leverage AI is for intent detection and mindset detection. Helping the Nivi bot determine what a user's state of mind is with respect to their healthcare goals, so that responses can be tailored in effective ways that are relevant to their current needs."

David Tresner-Kirsch, Nivi CTO

Nivi's approach to answering questions combines two AI paradigms: zero-shot learning and retrieval-based chat. askNivi leverages PaLM 2 to interpret various user questions, categorize the questions accordingly, and return the appropriate responses from its library of localized, expert-generated content.

In line with its goal of improving healthcare and quality of life, Nivi is trying to encourage health behaviour change amongst its users. Their approach is rooted in behavioural science and an understanding that making decisions that support users' health is shaped, in large part, by their beliefs. The Nivi team uses a similar method to its question-answering component to address different healthcare barriers encountered, employing few-shot learning (instead of zero-shot) with its retrieval-based bot. These newer question/answer and barrier consultation components offer more flexibility and improved user experience, encouraging more interaction with the tool.



*through an engineered prompt, zero-shot learning, and a reference library of conversation metadata

Figure 2 ▲
askNivi's question-answering feature
Source: askNivi

*through an engineered prompt, few-shot tuning, and a reference library of conversation metadata

Figure 3 ▲
askNivi's barrier consultation feature
Source: askNivi

Nivi in Kenya

One area that askNivi is being used in is for reproductive health in Kenya. In Kenya, Nivi is partnering with Ipas, a nonprofit organization dedicated to expanding access to reproductive healthcare.[16] Nivi provides askNivi to Ipas's networks of pharmacists in the country. One key issue in the country's healthcare system is the disjointed communication between pharmacists and patients regarding post-abortion care. Pharmacists are important providers in medical abortions, especially in rural and underserved areas where other healthcare providers and resources are limited. They are able to dispense take-home abortion medication and often are responsible for follow-up care for people who have abortions. However, until now, there have been no reliable methods for pharmacists to track the status and outcomes of patients after they take abortion pills. Conversely, patients have limited methods for connecting with pharmacists and getting answers to follow-up questions and concerns regarding the medication. Nivi and Ipas seek to fill this communication gap with askNivi.

The tool gives those seeking abortion care accurate information and connects them with pharmacists who can treat them appropriately. It also creates a channel for pharmacists and clients to stay connected after medication is dispensed to ensure proper follow-up care. Results from the partnership are promising. Nivi compared askNivi data and insurance claims data from Ipas to identify the overlap in chatbot users and patients seeking medication abortions. They found that Nivi referrals were responsible for a 4% increase in the total number of medication abortions provided. These results show that the askNivi platform was able to reach a significant number of people and get them to appropriate healthcare in a timely fashion.

16 Ipas, "Home Page."

Impact

Nivi wants to improve healthcare outcomes and encourage healthcare service utilization among its users. Their goal is not solely to maximize the number of people who interact with the askNivi platform, but to help users make more informed decisions and take action about their health. One of the main methods that Nivi uses to measure the impact of their technology is tracking user journeys from the platform to the point of healthcare service delivery.

> *"We've been trying to standardise this [process] by creating a closed loop interoperation at the point of service delivery."*
>
> **David Tresner-Kirsch, Nivi CTO**

Many other technologies determine value by the number of active users and amount of time that users spend communicating with the platform. While these metrics are useful for tracking direct engagement, they are not indicative of any value beyond the platform. Evidence of users acting on information given by askNivi and seeking care from a provider is one major way to prove the effectiveness of the platform.

Nivi collects information about user behaviour outside of the platform in a variety of ways. In some cases, Nivi asks users directly whether they acted on the referral information provided, where they went, and what services they received. Though this is the simplest way to gather the information, it produces noisy, less valid data. The company has also piloted a provider-first approach, rather than a user-first approach, to obtain information at the point of care. Nivi asks participating care providers to post QR codes in their offices and businesses. Patients scan the code and fill out the relevant information during their visit. The company found that data from this approach was higher quality, especially when providers were fully engaged in the process and encouraged patients to use askNivi.

Ultimately, Nivi aims to keep track of users as they progress through their healthcare journey—from users' first interactions with the chatbot to provider visits to (hopefully) changes in health outcomes. The AI components add a new layer to the company's process for measuring askNivi's value. Nivi is in the early stages of partnering with a Kenyan health insurance company and will soon be able to use askNivi data and insurance claims data to identify whether users are following through on referrals provided by the chatbot.

Responsible AI

Nivi is a healthcare technology company, creating useful tools that healthcare organizations can customize and deploy in a variety of settings. Nivi's choices about how and when to deploy powerful AI technologies reflect an attention to detail and to protecting both end users and the organizations servicing them. Conversations with them revealed five themes.

Do no harm.

"The first thing that comes to mind in any discussions around equitable, responsible AI and healthcare is the Hippocratic oath. It's first, do no harm. I think that we have seen a lot of things spun up very quickly but there's an inevitable, unhappy path where people are getting misinformation that is actionable misinformation and could put people in danger. And I think there are a lot of products that are rolling out right now."

David Tresner-Kirsch, Nivi CTO

Pursue fairness for all.

The Nivi team thinks that "*AI should only be added into a tool if it can make the aggregate outcomes for all demographic categories (i.e., gender, race/ethnicity, age, etc.) better.*" This requires the ability to track healthcare outcomes and attribute the outcomes to changes made in the technology. The team believes that, in addition to good model performance for all demographic groups, there must also be a degree of oversight of user interactions that occur on the platform to follow up with the tool in practice.

Keep humans in the loop.

One mechanism that Nivi uses to combat inequitable outcomes is including people in the loop who are representative of the population to review interactions on the platform. People who are fluent, first-language speakers review content in their respective languages (i.e., fluent Swahili speakers evaluate interactions conducted in Swahili in Kenya). Nivi also asks English-speaking reviewers from the countries in which the platform operates to evaluate interactions from those countries. English speakers in Nigeria, instead of English speakers from the United States, for example, review askNivi conversations that were conducted in English to avoid introducing culturally inappropriate biases to the auditing process.

Go slowly.

Following the release of ChatGPT and other LLMs, Nivi has considered integrating these new AI resources into its chatbot to create a fully interactive tool. Rather than completely relying on generative AI, they have decided to incorporate small genAI pieces into askNivi. They use the exceptional capabilities of LLMs to interpret and categorize free text in different languages to identify the types of answers users are asking about and return human-generated content. Nivi determined that the types of health information that users inquire about are too specific and the potential for harm too great to foist responsibility onto a general AI-based tool. Misinformation about medical abortion, for example, can cause additional fear and harm to users who may already be cautious about seeking abortion services. People who are given incorrect information about abortions could turn to more dangerous methods instead of seeking medical care that is safe and medically sound.

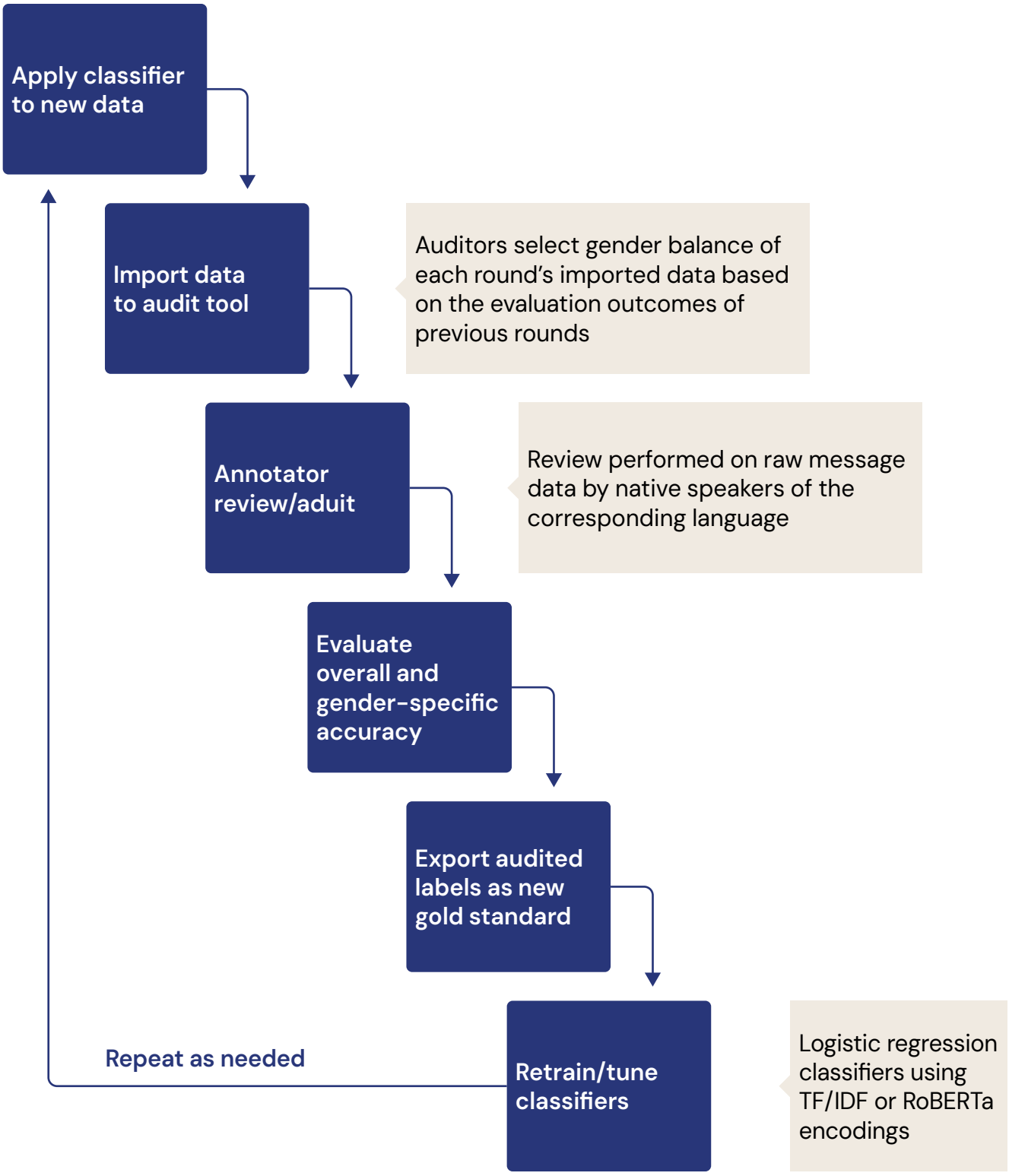


Figure 4 ▲

Development Pipeline for the Gender Auditing Tool
Source: Nivi, "Optimizing for Equity in Natural Language Processing."

Employ gender-aware auditing.

While Nivi has decided that the disadvantages of relying entirely on AI are too great, the team is still acutely aware of the risk that even small AI components bring to its tool and users. They are conscious that the introduction of generative AI can bring unintended bias into the system and Nivi is taking steps to limit the negative effects of this bias. Through USAID's Equitable AI Challenge, Nivi and its partner, UNILAG, have created a gender-aware auditing tool to evaluate user interactions in askNivi and the platform's subsequent responses by gender.[17] The goal of the auditing tool is to "*optimise for equitable outcomes during incremental development of natural language processing systems*." This auditing tool gives the Nivi team a new way to integrate responsible AI into the chatbot as it evolves. This differs from many existing auditing methods that evaluate AI performance after development and deployment has already taken place.

By building an AI-based system that is attuned to the needs of each local population and sensitive to user demographics, askNivi will ideally reach more women, help them make better-informed health decisions, and reduce overall healthcare costs.

What's next

Nivi is regularly testing new technologies to integrate into askNivi and improve the experience for all its users. The team is regularly maintaining its existing AI features and looking into other ways that AI, and genAI specifically, can benefit its users. Additionally, Nivi is a member of a consortium of implementing partners through the Frontiers Health Market Engage project.[18] Nivi will work with partners to implement askNivi in five states in India—Karnataka, Assam, Meghalaya, Maharashtra, and NCR-Delhi—to connect users with sexual and reproductive health information and provide timely referrals to local trusted providers.

17 The challenge was implemented by DAI's Digital Frontiers project. USAID, "Equitable AI Challenge and Community of Practice."

18 Nivi, "Nivi Joins the USAID Funded FHM Engage Consortium as the Technology Partner to Drive Demand and Consumer Insights on Sexual and Reproductive Health."

HURIDOCS: An AI-optimized database lowering barriers to information for human rights defenders

Human rights partners HURIDOCS and UPR Info developed a supervised machine learning tool that processes and automatically tags key human rights documents with specific categories. The feature, built on top of HURIDOCS's open-source database software Uwazi, has reduced workload for UPR Info staff and allowed the team to reallocate scarce resources to supporting UPR's advocacy programs.

Background

Documentation and information are essential for human rights defenders in their work to preserve civic space. Evidence of wrongdoings is necessary in the international justice system to hold perpetrators accountable. Documentation and recorded data provide a foundation for policy changes and, ultimately, widespread institutional change. However, much of this documentation is hard to find and in formats that are difficult to digitally process.

HURIDOCS (Human Rights Information and Documentation Systems) is an international NGO with a mission to make information and evidence more accessible to human rights defenders.[19] Since 1982, the team of archivists and software engineers has developed tools and strategies for information retrieval and data collection to support the use and preservation of human rights documents.

19 HURIDOCS, "Home."

HURIDOCS is guided by three main goals:

1. To explore new technologies and methods with human rights defenders to organize and share information;
2. To support human rights organizations by addressing their unique information needs through bespoke strategic and technological solutions; and
3. To build and maintain infrastructure for sustainable storage of human rights information over time.

HURIDOCS has worked with human rights organizations of all sizes to overcome barriers to information storage, retrieval, and use. For example, with the Institute for Human Rights and Development in Africa, HURIDOCS built the African Human Rights Case Law Analyser to improve access to case law and legislation resources in the African human rights system.[20] The resulting multilingual database not only provides access to African human rights law and case law but also identifies linkages to other potentially related texts and documents that could be useful in users' search for information.

Currently, HURIDOCS is concentrating much of its efforts on Uwazi, a document management system for the human rights community that has evolved into a database application with greater functionality and use. It is a flexible, open-source web-based application that provides users with a secure, private, and organized way to store and share their documents. In August 2022, Uwazi was recognized by the Digital Public Good Alliance as a digital public good (DPG).[21]

The application's features include the ability to manage multilingual documents, user access controls, and tailored search functions for easy information retrieval. Since its inception, Uwazi has been used in a variety of sectors, including to track attacks against environmental activists in Bolivia and to catalogue cases of arbitrary detention, abduction, and enforced disappearance in North Korea.[22]

HURIDOCS has also been trying out different AI tools to better analyze key topics in documents. Over the past eight years, the team has built expertise in and continues to research the latest advances in machine learning (ML) and natural language processing, including LLMs like GPT-4 and BERT. They ventured into the world of ML for text extraction and analysis on an early project involving a collection of ~9,000 court documents; HURIDOCS and their partner organization were able to categorize the documents with supervised ML and used them to identify potential gender bias in court rulings. This project was HURIDOCS's initial foray into AI and ML for improving information access and provided the foundation for the organization's continued AI work.

20 IHRDA, "African Human Rights CLA 2.0."

21 HURIDOCS, "Uwazi Receives Accolade of Being Recognised as a Digital Public Good by the Digital Public Goods Alliance."

22 CEDIB and CONTIOCAP, "Mapa de Ataques a Defensoras Ambientales"; Transitional Justice Working Group, "Footprints: Database of Those Taken by North Korea."

UPR Info and Uwazi

In 2019, HURIDOCS partnered with UPR Info to further develop and expand Uwazi. UPR Info is a nonprofit organization that supports and promotes the Universal Periodic Review (UPR), a process conducted by the UN Human Rights Council to evaluate the human rights performance of each UN Member State. The partnership applied ML to extract information from UPR documents.[23]

UPR Info operates select programs to promote the UPR:

1. The Policy and Advocacy Programme (PAP) helps ensure that the UPR remains an inclusive process by sharing good practices to enhance its effectiveness and creates opportunities for civil society organizations (CSOs) and national human rights institutions (NHRIs) to advocate effectively with UN Member States.
2. The Stakeholders Engagement Programme (SEP) prepares key national and international stakeholders to engage in the UPR, including human rights defenders, CSOs, NHRIs, journalists, universities, and more. The program includes capacity-building workshops and training sessions to build a network of partners who have the knowledge and skills to engage with Member States during the UPR.
3. The In-Country Programme (ICP) facilitates national and regional dialogue between key stakeholders to monitor progress on human rights commitments made during the UPR. The ICP's goal is "*to utilise the UPR to effectively implement human rights obligations and commitments on the ground by providing technical support and cooperation nationally and regionally.*"

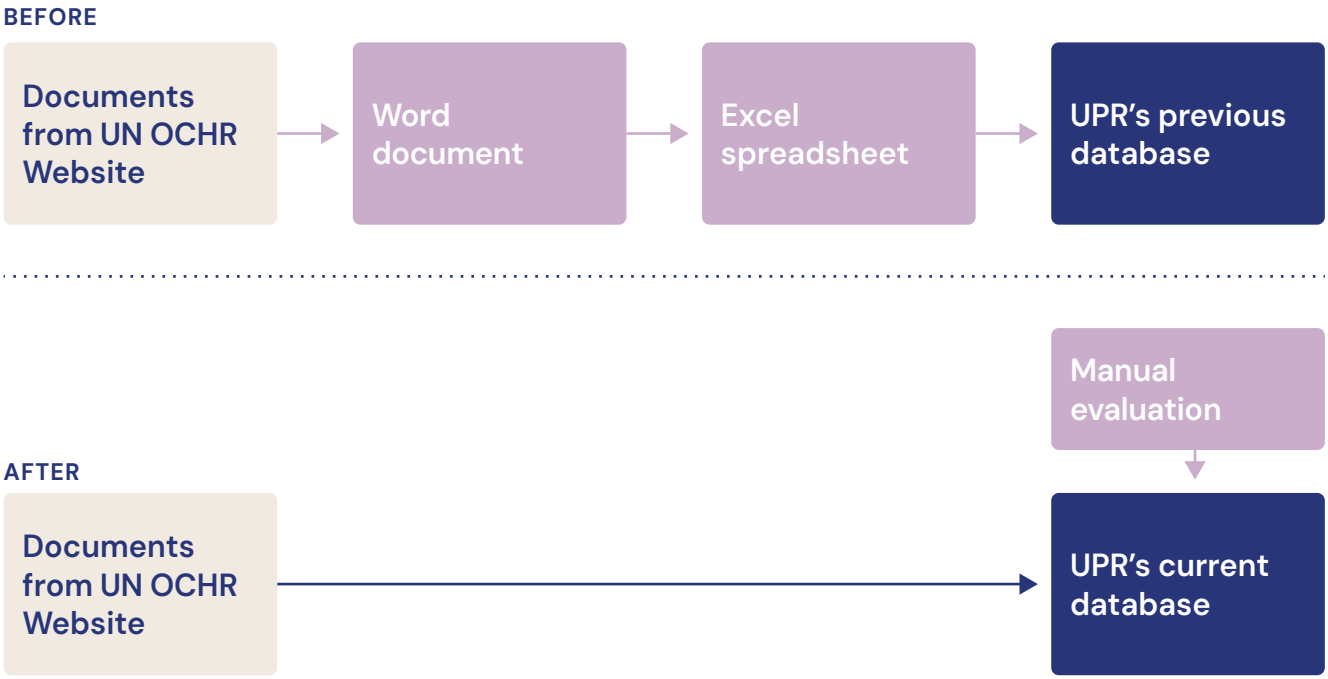
Additionally, UPR Info maintains tools to enable human rights stakeholders to better use the information created by the UPR through its Digital Innovation and Knowledge Management Programme. During each UPR session, Member States receive recommendations from the HRC Working Group and can make pledges to address those recommendations. UPR Info tracks these recommendations and pledges in its Database of UPR.[24]

Until recently, maintenance of this database was a manual process. The recommendations and pledges for each Member State are periodically published in summary documents on the Office of the United Nations High Commissioner for Human Rights (OHCHR) website. UPR Info staff members would visit the website, download and read any newly published materials, extract text on recommendations and pledges, and update the database with the new information. This two-month process was repeated three times per year.

23 The project was supported by a US$1-million grant from the Google AI Impact Challenge. Weinberg, "Geneva-Based HURIDOCS Is a Google AI Impact Grantee."

24 UPR Info, "Database of UPR."

Figure 5 ▼
UPR Info workflow before and after implementing AI
Source: Nivi



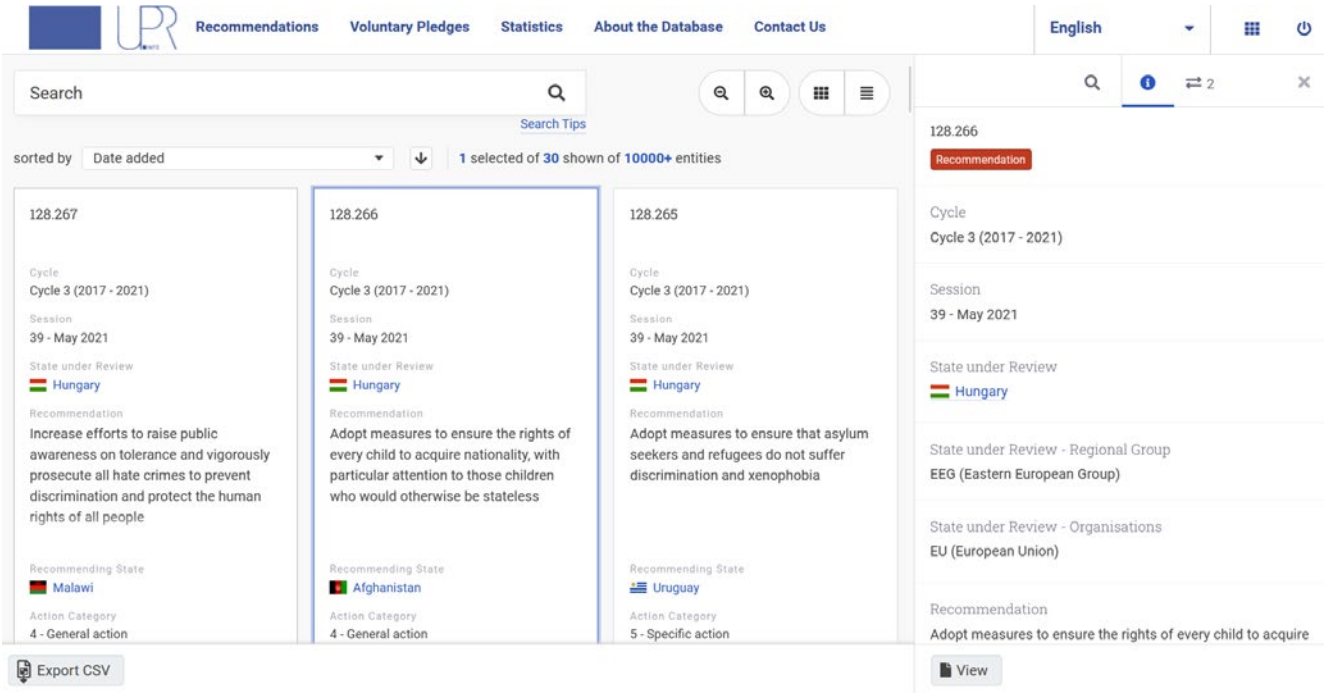


Figure 6 ▲
The Database of UPR
Source: Nivi

To support the speed of updating the database, HURIDOCS, a Google.org Fellow, and UPR Info created an automated tool that downloads new UPR documents from the OHCHR website, processes and extracts recommendations from documents, and suggests tags by topic and action type. The teams first held a series of design meetings to discuss the problems and needs of the Database of UPR and maintenance staff. From there, they determined that using Uwazi and automating parts of the updating process would be the best way to reduce the challenge of maintaining the database.

Developers migrated all data from the previous Database of UPR into a new database built with HURIDOCS's Uwazi software. They then created an automated crawler to periodically check the OHCHR website for new documents and upload them to the database's backend for additional processing. For the processing and tagging component of the tool, the team tested a variety of multilingual LLMs and trained the best-performing model on the existing Database of UPR. Proposed tags are manually reviewed for accuracy by UPR Info staff before being published to the database. In addition to metadata and topic tags, the tool tags documents by "Issues" (i.e., disability rights, women's rights, etc.) and "Actions" (i.e., specific action, minimal action, etc.).

Impact

As the HURIDOCS/UPR Info partnership was grant-funded, measuring impact was an integral part of the project. Driven by a human rights–based approach focused on the needs of their users, HURIDOCS and UPR Info tracked their tool's impact from both technical and social perspectives.

Increased informational accuracy

At the beginning of the project, HURIDOCS estimated that the tool would be able to achieve 83% accuracy in tagging documents with the correct topics. However, the tool has surpassed that estimate and, when originally released in 2020, was achieving between 87% and 94% accuracy on newly tagged documents. After a year and a half of use, the tool's accuracy has increased to about 93%.

Increased efficiency and ability to redirect staff labour

The new functionalities have reduced database maintenance time for UPR Info staff from over two months to two weeks per cycle. Before, maintaining and updating the database was tedious and time-consuming. The ever-present backlog of documents consumed limited team resources and prevented the organization from focusing on other important projects.

"[Uwazi has] freed up time for staff to conduct trainings, create events, [and] develop other initiatives."

Nicoletta Zappile, UPR Info Deputy Director

The team hopes that users will also experience an improvement in their work. Greater efficiency processing documents through the Database of UPR will ideally enable users to access and use more up-to-date information.

Sharing new tools with users

HURIDOCS has harnessed the progress made through its partnership with UPR Info to further develop Uwazi, its ML components, and its user interfaces. The successful integration of AI into the Database of UPR has led to additional funding that will enable HURIDOCS to further improve the system and provide all Uwazi users with new functionalities and improved user experience.

Responsible AI

When asked about their approach to responsible and ethical AI, UPR Info staff said that the organization does not have specific principles that it follows. Given that this is one of their first projects using AI, the technology and its applications are relatively new. However, they acknowledge the importance of having a set of guidelines for the ethical use of AI.

"I believe that it is important to develop a common ethical code on the use of AI. It is important that there is a code that can hold accountable persons and institutions that violate the basic human rights and fundamental freedoms when using AI technology."

Nicoletta Zappile, UPR Info Deputy Director

The absence of official guidelines, however, has not stopped the organizations from putting responsible AI practices in action. Both HURIDOCS and UPR Info approach their work from a human rights perspective and are cautious about how to build and integrate AI into UPR Info's workflow.

Integrate manual validation.

The UPR Info team said that "*keeping staff involved [in updating the database] is important and responsible*." Even though the tool achieves high classification accuracy on new documents, UPR Info does not anticipate removing manual validation in the near future.

Experiment with AI on low-risk issues.

For this tool, AI components do not generate any new content, only analyze existing content. Both HURIDOCS and UPR Info teams acknowledge that, while they are cautious about using AI, this particular use has a very low risk of harming users. An incorrect tag may cause a small inconvenience to a user who is looking for a specific document. However, in comparison to an AI tool used to categorize a person's health status, for example, the negative impact is minimal. The Database of UPR also provides all documents and information—not just the classifications—to its users. As a result, users can, and do, contact UPR Info to flag errors in the database that are then manually resolved.

Build an approach based in human rights.

As organizations dedicated to human rights, both HURIDOCS and UPR Info integrate human rights–based approaches in all their activities.

"We are very concerned with ensuring an inclusive approach, where every stakeholder, regardless of their background and their ability, can benefit from the data and analyze it in a safe manner. I believe that the use of AI requires, first of all, an understanding of the use and the impact of the use of AI technologies."

Nicoletta Zappile, UPR Info Deputy Director

"There's a lot of questions about what responsible AI actually means practically in organizations. For us at HURIDOCS, it simply means keeping humans and human rights as a cornerstone of our practice. For any organization looking to start using machine learning and artificial intelligence in their work, or even start building it themselves, the biggest aspect is not following the hype, not just thinking that this is going to solve all of the problems that we haven't been able to solve so far. It's simply not true. These tools help us work more efficiently, and so systems need to be in place for that work to happen."

Danna Ingleton, HURIDOCS Executive Director

What's next

UPR Info and HURIDOCS updated the Database of UPR through a grant-based partnership with the Google Foundation. HURIDOCS continues to provide maintenance to the database to ensure that the tool continues to work as intended. Now that the new AI-powered functionalities are in place, UPR Info is directing resources toward users. Much of the Database of UPR upgrade occurred in the backend—the ways documents are integrated and processed. UPR Info is now interested in how to improve the user experience of the database. The team hopes to conduct user feedback activities to learn more about the changes and improvements users want to see.

PROMPTS:
A messaging platform strengthening primary healthcare in Kenya

PROMPTS is Jacaranda Health's digital health tool that has empowered three million Kenyan mothers with information via SMS to seek and connect with care at the right time and place. Since launching the tool in 2017, Jacaranda has built and customized the AI that underpins PROMPTS to improve its efficiency, speed, and personalization for mothers, at scale. The AI component of the platform has evolved from a natural language processing model to a sophisticated Swahili-speaking LLM that provides tailored information to mothers based on user profiles and rapid triaging to a human help desk agent if a risk is identified during the exchange.

Background

Jacaranda Health is a Kenya-based nonprofit working to improve quality of care and life for mothers and newborns.[25] Established as a social enterprise in 2011, the organization originally focused on direct healthcare service delivery. It built and operated mobile and fixed health clinics to provide women in peri-urban areas with accessible and affordable healthcare services. Over the past decade, however, Jacaranda has expanded its work to address two major challenges facing mothers and babies seeking care in government health systems: delays in seeking care and gaps in the quality of facility-based care.

25 Jacaranda Health, "Home."

Jacaranda focuses on three groups of healthcare stakeholders and addresses each with tailored solutions:

1. MENTORS is a training-of-trainers program that provides frontline obstetric workers with skills to prevent and address life-threatening pregnancy and postpartum complications. Trainees also have access to DELTA, a WhatsApp-based learning tool that enables users to continue their training remotely with self-paced learning modules and tests.
2. PULSE supports meaningful data analysis and visualization to enable healthcare facility managers and government stakeholders to identify gaps in service delivery and highlight target areas to direct resources. Dashboards and analysis outputs provide real-time feedback from clients and workers in the healthcare system, giving them the opportunity to improve the system that they are a part of.
3. PROMPTS is a two-way SMS tool that encourages women to seek care through personalized messaging via a communication interface. The Jacaranda tech team has recently integrated an AI component to improve the tool's response rate and overall user experience.

Jacaranda's tools and initiatives have proven to have benefitted the lives of mothers and children across Kenya. In the MENTORS program, 87% of nurse participants have improved competence and confidence in standard obstetrics care. More than 20 Kenyan county governments currently access and use data visualizations and dashboards developed through PULSE. PROMPTS has provided information to more than 2.4 million enrolled expecting and new mothers. PROMPTS has already reached 2.9 million women across Kenya and has government-partnered pilot operations in Ghana and Eswatini.

Photo by Jacaranda Health

PROMPTS in action

PROMPTS began as an SMS tool to encourage health-seeking behaviours by sending mothers reminders and information about regular checkups, risk factors, and other key messages related to pregnancy and birth. While PROMPTS was intended to be a one-way method of communication, the technology allowed mothers to send messages back. The Jacaranda team received questions and messages from expectant mothers but were not equipped to respond. Seeing an opportunity to engage with users, Jacaranda began to develop its tool to handle and respond to user inputs, including integrating automated natural language processing modules to lower response time.[26]

Photo by Jacaranda Health

This updated PROMPTS platform was built on RapidPro, an open-source software platform and DPG that uses SMS and other communication channels for data collection and mass communication.[27] Mothers sign up for the tool during visits to public health facilities. PROMPTS then sends unprompted SMS messages in Swahili and English with information on healthy behaviours and necessary healthcare tailored to mothers' stage of pregnancy. It also asks mothers for feedback during their health journey, like whether mothers attended checkups at health facilities and the quality of care they received. This data is relayed back to facility management and government stakeholders to help them improve service delivery and direct resources to priority gaps.

Users can also send free-text messages at any time via PROMPTS. Prior to the integration of AI, a team of help desk agents manually assessed message content, identified messages' urgency, and responded appropriately. As PROMPTS grew, however, the number of messages increased to thousands per day, and the help desk was having difficulty keeping up. The Jacaranda tech team then developed AI models trained on historical messages labelled by help desk agents to categorize and triage messages. PROMPTS now automatically organizes the deluge of messages by type and priority, enabling help desk teams to respond quickly and effectively.

26 Jacaranda Health, "Jacaranda Launches First-in-Kind Swahili Large Language Model."
27 "RapidPro."

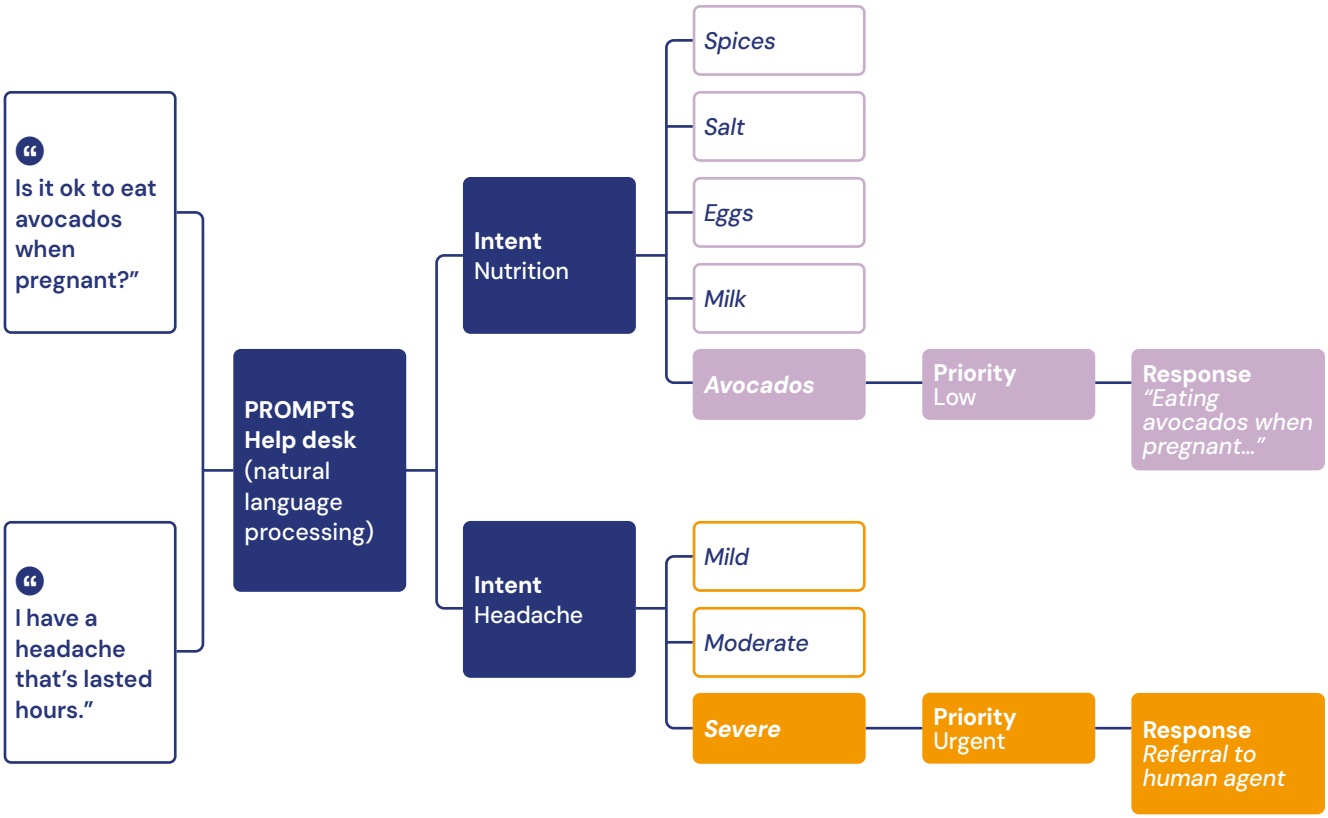


Figure 7 ▲
PROMPTS in action
Source: PROMPTS

Jacaranda Director of Technology Jay Patel noted that the development and maintenance of the tool has been a partnership between the tech team and the help desk. Help desk agents are the primary users of the AI module and have been key to its success. For example, when the AI tool was first introduced, PROMPTS only had one category for nutrition and one generic answer for nutrition questions. It was unable to handle the nuance of nutrition-based questions about drinks, types of food, quantities of food, etc. Due to feedback from the help desk, PROMPTS can now categorize nutrition-related messages into over 30 specific categories. Similarly, the help desk originally noticed that 30% to 40% of the questions they received were categorized into catch-all "general baby" and "pregnancy" categories. Based on their experience reading and triaging messages, the help desk suggested new categories to make the triaging of questions more nuanced.

"When I joined, we were only getting 100 questions a day. At the time, we had one help desk agent and she was able to look at the questions as they came in and decide which ones she needed to answer first … But as we've grown, our question volume has increased. We're getting up to 10,000 questions today and manually reviewing those is no longer possible. So we use AI to flag the urgent questions and make sure that these moms who are experiencing these potential danger signs get the help they need, much more efficiently and faster."

Jay Patel, Jacaranda Health Head of Technology

Photo by Jacaranda Health

Additionally, integrating AI into PROMPTS has made it faster for Jacaranda help desk teams to support mothers who are experiencing signs of a medical emergency. With PROMPTS receiving over 5,000 messages daily, it is difficult for the help desk team to quickly identify which messages need immediate attention (about 5% of messages). The PROMPTS automated classification tool organizes messages accordingly and enables the Jacaranda team to respond more quickly to high-risk cases. At the moment, PROMPTS is able to correctly detect and respond to 86% of user questions.[29]

"AI comes in handy to ensure that questions are triaged according to urgency. It categorizes them into different priorities … so it's efficient for us to reach those moms who are experiencing dangerous things in pregnancy and whose babies are experiencing danger signs and ensure that they are getting to the health facility as quickly as possible. We are also able to map their locations and refer them to the nearest health facilities where they can get help as fast as possible."

Pauline Nafula, Jacaranda Health Senior Help Desk Coordinator

Qualitative data from the help desk shows that the help desk teams significantly value the tool. They actively engage in its improvement by offering suggestions for new categories and responses and consistently providing feedback. Thanks to the tools and the help desk intervention, now 90% of high-risk cases flagged via PROMPTS receive care at a hospital.

28 Jones et al., "A Short Message Service (SMS) Increases Postpartum Care-Seeking Behavior and Uptake of Family Planning of Mothers in Peri-Urban Public Facilities in Kenya."

29 Jacaranda Health, "PROMPTS: Promoting Mothers in Pregnancy and Postpartum Through SMS."

Responsible AI

"[Jacaranda has a] very high risk threshold. Especially [since we're] dealing with health and with underserved populations, we want to make sure we're getting it right."

Sathy Rajasekharan, Jacaranda Health Co-founder

Jacaranda has established a list of values that shape its use of AI: Jacaranda's Principles for Responsible, Person-Centered AI, a codified list of values that is carried out by the behaviours and choices the organization makes to ensure the safe and responsible use of AI for healthcare. The principles are:

First, do no harm.

Jacaranda's mission centres around the improvement of maternal and child healthcare and its work puts mothers and babies first.

Respect user data.

The only personal identifiable information that Jacaranda collects are phone numbers from users and all data is stored on secure cloud servers with strict access controls. Jacaranda also has transparent consent mechanisms in place to ensure that mothers who use the PROMPTS platform are aware of what data will be collected and how it will be used.

Design for equity and local context.

The AI models used with PROMPTS are tested and trained with data that matches the context in which the tool is used. Jacaranda uses data that reflect the communities they work in and works to fill gaps where data from these communities is lacking. For example, after noticing the poor performance of Swahili amongst the language models they tested, the Jacaranda tech team fine-tuned Meta's Llama 2 to better manage Swahili input.

Keep humans in the loop.

Experts are always a part of Jacaranda's AI process. The choice to use AI was guided by the help desk agents, the ones who use it. They also continue to play a part in training and maintaining the tool. The use of AI in PROMPTS was never to replace the help desk and there is no foreseeable future where the task of responding to PROMPTS messages will be fully automatic.

Share with the community of practice.

Jacaranda is a proponent of open-source tools and DPGs, using and adapting existing technologies for its specific purposes. PROMPTS is built on RapidPro, an open-source software created by a partnership between TextIt and UNICEF. The tech team also frequently tests different open-source language resources to improve upon the AI message classification tasks. Additionally, Jacaranda shares its learning and technology with the public. Their Swahili language model is available to all on Hugging Face, an open-source library of AI resources.

Focus on sustainability.

In low-resource communities, the high cost of technology and AI tools can be prohibitive to development and use. Jacaranda capitalizes on existing open-source, high-resourced technologies like Llama 2 and XLM-RoBERTa and adapts them to the Kenyan context. Jacaranda is able to allocate resources to making established AI technologies even better, instead of building them from scratch, a task that requires significantly more human and computational resources.

What's next.

Jacaranda's goal is for all mothers have access to PROMPTS. They are expanding their services in sub-Saharan Africa, specifically Eswatini and Ghana, and developing the tool to offer it in additional languages. Jacaranda is also looking into improving the support that it provides to mothers. The team knows that mothers at risk need different information and kinds of support than other mothers and are exploring ways to change their technology to match mothers' different needs. Finally, Jacaranda is hoping to expand their knowledge base beyond maternal and newborn health. The team wants PROMPTS to be able to handle more health issues, including those that mothers often experience after pregnancy.

EIDU:
An AI-powered learning platform using content personalization to maximize learning in Kenya

EIDU, a social digital learning technology company, integrated an AI personalization algorithm into its digital learning platform that suggests learning content for individual students. The tool, used to support traditional classroom learning, helps teachers identify learning content based on the probability that a student will be able to complete them. It has enabled individual learners to learn at their own pace and in formats that fit their style of learning, ultimately maximizing students' learning outcomes.

Background

Initiated in 2015, EIDU is a social digital learning technology company using AI to improve learning outcomes for schoolchildren.[30] Founder Bernd Roggendorf created the organization and its eponymous digital platform after identifying a need to deliver evidence-based educational content in a cost-effective manner and at scale. In low-resourced education environments, there was a clear need for different tools and methods to support teachers and students in the classroom. The team researched various platforms to deliver content in the classroom but quickly found that no existing software tools fit the digital context and needs of teachers and schoolchildren in low- and middle-income countries, where skills in basic literacy and numeracy are often below high-income countries.

30 "EIDU – Our Classroom Is a Billion Strong."

"The idea [of EIDU] was: can you design something which works in this [very difficult] environment? So from the start, bringing the teachers on board to see what they need [and building a system] that works in their context. This was the starting motivation of our work."

Aidan Friedberg, EIDU Director of Learning and AI

Originally, the organization developed both the educational content and the delivery platform but has redirected its efforts exclusively to building and maintaining the platform. Through partnerships with Kenyan county and national governments EIDU has implemented its platform in over 7,000 schools across the country and has access to government-confirmed content.

Through constant, iterative development cycles with teachers, software engineers, and government stakeholders, the EIDU team built a modular digital platform that is easily customizable to the specific needs of a given educational system. The system supports various education stakeholders to accomplish their respective education goals:

1. Learners receive educational content that is aligned with local curricula and complete digital assessments.
2. Teachers are trained and provided with lesson plans and learner development overviews to enable structured pedagogy.
3. Trainers of teachers are provided with tools that help prioritize support planning and structured feedback for lesson observations.
4. Government ministries are provided with dashboards enabling a real-time aggregated overview of their education systems.

EIDU provides teachers and students with appropriate educational content and information in a manner that fits their local context to maximize students' learning outcomes. The application is built to run on low-cost Android devices, and all content is available offline. Content is delivered via a child-friendly interface and in a gamified way to engage students and improve their learning more efficiently. EIDU captures and analyzes anonymized learning data from students via their app usage, enabling teachers, school administrators, and government stakeholders to track students' progress over time, providing previously unavailable insight into learning conditions.

AI for EIDU

One key feature of EIDU is its ability to suggest personalized learning content for individual learners. Teachers and tutors can manually determine the best learning material for individual students, but this task is time-consuming in a classroom with many students at different stages of learning.

"What's powerful about a tutor is [that] a tutor is able to identify the learning state of a learner … and can select the right content for that learner … What parts of this tutor do we think is powerful? Could we try and replicate [it]? And this is where this individualised learning came in."

Aidan Friedberg, EIDU Director of Learning and AI

The EIDU team decided to use AI to suggest personalized learning content, emulating the role of a tutor, based on Bloom's 2 sigma problem, an educational phenomenon that students taught through one-on-one instruction perform two standard deviations better than students taught in a standard classroom.[31]

The EIDU team tested two models to solve this problem. The first used skill-based matchmaking (SBMM).[32] Used frequently in gaming, SBMM matches players of similar skills to create fair games. Skill is determined by a variety of metrics, including win-loss ratios. EIDU applied SBMM by treating learners as players in a game. Learners gain skill points by completing an exercise and lose skill points by failing to complete an exercise. Concurrently, the exercise loses skill points when solved and vice versa when students do not complete it. The SBMM model aimed to match learners with exercises that they have an 85% chance of solving.

Photo by EIDU

31 Bloom, "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring."

32 Khare, "Exploring Skill-Based Matchmaking Systems in Online Games."

The second method used a long short-term memory neural network (LSTM) to model and predict learners' mastery of skills over time.[33] The model processes learners' performance on exercises—completed or not completed—and outputs a set of probabilities for completing different future exercises. This method selects exercises with a higher average completion probability set, which, theoretically, will help the learner improve the most overall. The EIDU team found that the LSTM method led to significant improvement in learning outcomes.

Impact

EIDU's personalization models are one small component of their larger digital learning platform. The team measures the impact of their AI products accordingly and evaluates the tools as part of the platform's performance. EIDU's team regularly tests their models against standard model evaluation metrics, but they are ultimately focused on maximizing the impact the platform has on learners and learning outcomes.

"Successful impact in our environment can mean different things. Is the user happy with what we've provided them and are they excited to use our platform? These are the classical measures of engagement which we also sometimes rely on. Ultimately though our North Star is the learning outcomes of our students. Measuring this continuously outside of randomised control trials can be complex but we have a number of proxy metrics we use for learning development."

Aidan Friedberg, EIDU Director of Learning and AI

Improved learning outcomes

While the team acknowledges that there are many different metrics of success, EIDU is ultimately driven by the desire to improve learning outcomes. Preliminary data from a one-year randomized controlled trial (RCT) conducted by EdTech Hub examining the personalization tool shows significant improvement in learning outcomes among students using EIDU (SD=0.43).[34] The RCT is one stage of a four-part study that explores different components of personalized learning and its integration into classrooms. The final results of the study are expected to be published in late 2024.

33 Shipra Saxena, "What Is LSTM? Introduction to Long Short-Term Memory."

34 Daltry et al., "Digital Personalised Learning to Improve Literacy and Numeracy Outcomes in Kenyan Classrooms."

Photo by EIDU

Learners typically engage with EIDU for multiple terms, giving the team more time to test and evaluate the impact of the application on learners. New content, recommendations, and models are tested using A/B testing and give the team valuable feedback on the effectiveness of their changes. A/B testing allowed the team to test multiple variants of new material (i.e., content, functionality, interface, models, etc.) by randomly showing users one of the variants and collecting and comparing student performance data on each variant. EIDU's A/B tests run, on average, between full weeks and a full school term.

Increased digital literacy

Anecdotal qualitative data shows learners' and teachers' excitement and happiness about the platform. Teachers who often lack digital literacy skills are enthusiastic about learning to use EIDU on digital devices.

"The benefits and the learning gains that were coming with EIDU gave [teachers] the motivation and encouragement to continue learning how to use the digital devices that we gave them."

Olivia Awuor, EIDU Field Officer

Photo by Story × Design

Similarly, Awuor saw learners consistently show excitement and curiosity over the platform and its educational content. She recounted several cases where students asked their parents for their digital devices to continue learning with EIDU at home. Teachers' and students' responses to EIDU not only indicate interest in the platform and its content but also seem to be bolstering their digital literacy knowledge. The skills that they are developing are transferable across many digital platforms and key to engaging in an increasingly digital world.

"When we introduced the use of AI with the counties that we are working with, we did not experience any hesitation or pushback from the community. What we experienced is actually an eagerness for the communities to learn more about AI. It's like we awakened them to what's happening today and also offered solutions to some of the challenges that they have been facing as a whole."

Olivia Awuor, EIDU Field Officer

Scaling learning successes

EIDU's goal is not only to broadly improve learning outcomes but to improve learning outcomes at scale. EIDU's AI-powered platform is currently operating in 21 Kenyan counties and supporting over 350,000 active learners.

"There have been numerous interesting educational projects that showed a lot of impact in a randomised control trial or in a pilot and then when they scaled they failed. The complexity of scaling is often underestimated in our environments. Our ability to scale our model sustainability is almost as important as ensuring we're creating a large impact on learning."

Aidan Friedberg, EIDU Director of Learning and AI

Responsible AI

Like many AI practitioners in the development world, EIDU has thought extensively about how to ensure that their AI tools are appropriate and effective for their chosen problem. Their approach to responsible AI is based on the foundation that AI should only be used when it is absolutely necessary.

"It has to start with the question, 'Do I need AI at all?' A lot of the time, the solutions that you're trying to build may not actually need AI and can be solved just as effectively with existing tools or simpler methods. Where possible, these other tools should be prioritized."

Aidan Friedberg, EIDU Director of Learning and AI

Continuously evaluate uses of AI.

In line with this foundational value is the need to regularly revisit whether EIDU's AI products are functioning as intended. The team uses a variety of tools to determine whether their personalization algorithms are helping students maximize their learning outcomes including A/B tests and RCTs.

Recognize the bias in AI.

EIDU integrates bias evaluation into their personalization models. The team collects different demographic data, including gender, location, and age, and tests each demographic category against any model that they are testing or implementing. EIDU is continuously training their models and concurrently evaluating them against standard evaluation metrics.

"Is the model still accurate? Is it still as accurate as our original estimates, and secondly, is this affecting any demographic or any category in a skewed way? Are we no longer predicting in a fair and in a fair and balanced way as much as possible?"

Aidan Friedberg, EIDU Director of Learning and AI

Establish reliability for users.

In addition to quantitative model evaluation, EIDU incorporates responsible AI practices into its testing and release of new model updates to schools and learners. The team first releases updates to a small group of beta users, typically a group of ten to twenty schools, who provide qualitative feedback on the performance of the tool. After beta testing, EIDU then pushes out updates in an A/B test for a set period of time and determines whether the new version provides higher learning outcomes or faster learning development than the previous version. The team only releases versions of their model that successfully undergo beta testing and have proven improvements to previous versions in learning outcomes.

What's next?

Through observations and discussions in the classroom, EIDU has identified a need for educational content specifically tailored to students with special needs. Learners with special needs do not have specific content tailored to them but have shown interest when other learners in their class engage with EIDU. Many learners with special needs repeat the sounds and words they hear from other students' devices. Though EIDU no longer creates content for its platform, the organization is working with partners like onebillion, an education nonprofit, to create more content for students with special needs to address this request from students and teachers alike.[35] In addition to content updates, EIDU intends to continue supporting school administrators and government officials in taking ownership of the platform in line with its vision for sustainability. They are currently preparing to implement their tailored digital learning program to all ten counties of the Frontier Counties Development Council, which has the potential to benefit 150,000 pre-primary learners.[36] This extensive rollout underscores EIDU's commitment to making quality education accessible to every learner, regardless of location.

Finally, in terms of the future of AI at EIDU, the team is working to make their tools open source so that any implementer or government can replicate the impact EIDU has achieved in Kenya. EIDU hopes to contribute to a growing community of AI developers and stakeholders who are creating AI products specifically for low-resourced environments.

35 "Onebillion: The Goal - One Billion Children Reading."
36 "Frontier Counties Development Council."

Farmer.Chat: An AI-powered digital assistant supporting agricultural extension workers around the world

Farmer.Chat is a generative AI assistant developed by nonprofit Digital Green that provides agricultural extension workers with timely, locally tailored agronomic information. The assistant, operational in India, Kenya, and Nigeria, harnesses the capabilities of large language models to fill in agronomic knowledge gaps amongst extension workers. Farmer.Chat has reduced the significant burden that underpaid and overworked extension workers face in their efforts to support smallholder farmers increase productivity and income.

Background

Digital Green is a nonprofit organization dedicated to supporting farmers to increase productivity and income through access to technology and data resources.[37] Formally established in 2008 in San Francisco, California, USA, Digital Green was created as a project in Microsoft Research India's Technology for Emerging Markets group.[38] At the time, one of the major challenges facing farmers in India was an overall lack of appropriate knowledge about farming, exacerbated by an overwhelmed public extension system. Farmers were often making uninformed decisions based on intuition and hearsay. Digital Green sought to create new methods for disseminating valuable agricultural information to Indian farmers via digital videos, called a video-based extension approach (VBE).

37 "Digital Green."
38 Gandhi et al., "Digital Green: Participatory Video for Agricultural Extension."

Since then, Digital Green has partnered with communities in India, Nigeria, Kenya, and beyond to create videos containing localized agronomic content and advice. The videos feature members of the communities speaking local languages and demonstrating localized agronomic best practices. Digital Green then trained a network of established agricultural extension agents on how to disseminate the videos to groups of farmers. Extension workers are trained individuals who provide technical agronomic advice and support to local farmers. Digital Green has supported the creation and dissemination of over 8,000 videos that have garnered about 80 million views on YouTube. The organization has reached about 100,000 extension agents through the established dissemination network and about 6.3 million farmers.

Digital Green has developed a variety of other digital tools to support farmers from production to delivery. Mobile app Loop reduces the time burden and financial cost smallholder farmers face on market days, thus increasing their revenue.[39] The application created a way for farmers to sell their produce through knowledgeable extension workers and manage transactions. Connect Offline Connect Online (COCO), a management information system, captures detailed farmer-level data on the individuals who attend video dissemination sessions, the agricultural practices that are screened, and any behavioural changes identified.[40] The tool is still used in Ethiopia but has been phased out in Kenya and India.

After 16 years of operation, Digital Green has seen new challenges arise in knowledge communication for farmers and extension agents due to rapid changes in technology, climate, markets, and gender dynamics. Thanks in part to Digital Green's efforts, government bodies have largely institutionalized the creation and dissemination of agronomic videos. Technology has evolved and farmers have adopted new technologies quickly. Additionally, agricultural stakeholders who have witnessed the success of extension workers have put more work and responsibility onto extension workers. Extension agents are overworked and underpaid and needing more support to accomplish all the tasks that have been assigned to them. Farmers are also needing localized information—specific to their geographical context, crops, and soil types—to find the advice useful and actionable.

39 Digital Green, "Loop Mobile App Makes Farm to Market Linkages Easy."
40 Shah and Joshi, "COCO."

Farmer.Chat

In response to these new challenges, Digital Green created Farmer.Chat, an AI-powered assistant intended to help extension workers provide better agronomic information to farmers, faster.[41] Farmer.Chat is a multi-modal natural language chatbot that is accessible in various local languages and dialects. Extension agents use Farmer.Chat to ask questions that farmers have, triage these questions, and deliver information to farmers. The chatbot offers a new method to deliver timely, localized information to farmers and provides a demand-driven forum for farmers to ask their own questions. It also supports extension agents by filling in their knowledge gaps.

"People deploy AI products for the sake of deploying AI products. And for us, it's never been about the tool. Whether it's video-based or something else, how can it be something that really adds value, and builds trust within extension agents and farmers."

Jona Repishti, Digital Green Head of Global Gender Programs

Farmer.Chat was initially developed as a pilot to answer common questions from Indian chili farmers. The team loaded a collection of documents, FAQs, and video transcripts from the Indian Ministry of Culture and Digital Green's existing videos into a vector database. The model then searched the vector database, added the results into a GPT script, and outputted the answer in a conversational response format. After successfully beta testing Farmer.Chat in a few select regions in India with agricultural extension workers, Digital Green decided to move forward with building out the bot with more content and geographic spread.

Photo by Digital Green

41 Digital Green, "Farmer.Chat."

Currently, Farmer.Chat is a Retrieval-Augmented Generation (RAG) application that combines a corpus of agricultural-specific information with the capability of an LLM, specifically GPT 3.5.[42] At the time of development, GPT was the best-in-class, significantly reducing hallucinations,[43] and tech developer OpenAI was providing technical and financial support to Digital Green. However, Digital Green has been testing various open-source and proprietary models with other tech partners like Meta.

In addition to technology partnerships, Digital Green has developed a network of other partners to create new training content for the application, assist in translation services, and conduct model evaluation. In Kenya, the Kenya Agricultural & Livestock Research Organization (KALRO) brought the most up-to-date and local agronomic content for the region to Farmer.Chat.[44] Another partner, the Climate Action for Smallholders Coalition (CASH Coalition), provided Farmer.Chat with a wealth of content on climate-smart agronomic best practices and knowledge specifically for smallholder farmers.[45] Digital Green has also hired local agronomists to source additional information for the model.

Farmer.Chat is multilingual and can interact with users in five different languages, including Telugu, Hindi, and Swahili. To assist in translation, the tool uses Bhashini, an AI-based translation tool specifically for Indian languages, and Google Translate.[46] Digital Green has formed additional partnerships with organizations like Karya.[47] Karya built datasets containing thousands of sample questions and answers from farmers in local languages that were used to further train Farmer.Chat.

> *"We realise that these tools are far from where we would need them to be in terms of the languages that we want to have capabilities in, particularly when it comes to the language that extension agents and farmers at the language that they use in their day to day."*
>
> **Jona Repishti, Digital Green Head of Global Gender Programs**

Neuron, an India-based company, helps to evaluate the performance of the RAG model. While the internal technical team conducts automated model evaluations, Neuron supports Farmer.Chat by manually comparing question-and-answer pairings. They help to determine whether the answers match the evaluation scores generated by the automated evaluation process and whether the Farmer.Chat answers appropriately match the farmers' and extension agents' questions. Repishti highlighted the importance of Digital Green's partners in continuing to refine Farmer.Chat, noting that "*there are quite a lot of pieces that still need to be pushed forward because of who we are delivering the tool to, and the fact that there just isn't that much content that is up to date and reliable,*" but that the support of many partners has been essential in the work so far.

42 Selvaraj, "What Is Retrieval Augmented Generation (RAG)?"
43 OpenAI, "GPT-4."
44 "Kenya Agricultural and Livestock Research Organization."
45 "CASH Coalition – Climate Action for Smallholders."
46 "Bhashini."
47 "Karya | We Solve Data Needs."

Impact

With the introduction of new tools and technologies, Digital Green's focus on small-scale farmers hasn't changed. Ultimately, the organization is still trying to improve how smallholder farmers are conducting agricultural activities and how successful they are in terms of productivity. The organization is trying to increase farmers' income, support their resilience to climate change, and ensure that impacts are equitable in terms of gender. With that in mind, the MEL team is still working on how to define success metrics for Farmer.Chat beyond product analytics.

"We … look for impact at several different layers. So at the top, it's around these frontline workers and extension agents. These are our current users. Then we look at the farmers that they serve. Is it increasing their knowledge, their adoption, and income? And then we also look at impact at the system level, because we want to make the delivery of information a lot more cost effective."

Jona Repishti, Digital Green Head of Global Gender Programs

Photo by Digital Green

"If there is a certain level of user engagement, then what is the actual value? What is the impact at different levels? We're quite in the early stage in terms of measuring impact."

Tetyana Zelenska, Digital Green Head of Monitoring, Evaluation and Learning

As of January 2024, over 4,500 extension agents are using Farmer.Chat. Number of monthly active users and number of messages sent are simple metrics for the team to collect and analyze, but they give little insight into the value of the tool. Change in agricultural knowledge and improved confidence amongst extension workers are two additional areas of interest for the team. Anecdotal data collected from small samples of users has demonstrated increased self-efficacy and greater intrinsic motivation to engage more with farmers and the tool.

"We understand that perhaps there's going to be intermediate effects of chatbot usage because we're not giving it directly to farmers. But we are learning that there are these behavioral changes that come forward with frontline workers with this new tool in how they serve and how they work with farmers and how the whole system is evolving with this work."

Jona Repishti, Digital Green Head of Global Gender Programs

The Monitoring, Evaluation and Learning team is pursuing other methods for collecting this type of information at a larger scale to determine the true impact of Farmer.Chat. They use A/B tests to determine the effect of new updates and features for Farmer.Chat. The team has also been conducting semantic and quality analysis to understand question typologies: What are users asking about? What are the topics that come up most often? Finally, they conduct user testing and deploy evaluation studies to determine the role of Farmer.Chat on improving agricultural knowledge and adopting best practices.

Responsible AI

"[Our users are] vulnerable populations that have wide-ranging skills and don't fully understand how they are sharing data."

Jona Repishti, Digital Green Head of Global Gender Programs

Create foundational values for the use of AI.

The three pillars of responsible AI and digital technology that Digital Green adheres to are transparency, privacy, and inclusivity.

1. **Transparency**: Whenever possible, Digital Green uses open-source technology and engages in the practice of sharing their technology. The Farmer.Chat platform is open source; other organizations and developers are encouraged to use it. In India, Kenya, and Ethiopia, the technology is being institutionalized and advanced by country governments. In the name of transparency and building trust with users and the larger AI community, organizations are invited in to see how the technology works and improve upon it.
2. **Privacy**: Digital Green is keen on hyper-privacy for its users and their data. To ensure the highest privacy for their users, Digital Green works with policy organizations and lawyers to build comprehensive and understandable data collection, storage, and use policies. Users are informed of these policies in onboarding sessions and consent mechanisms are in place to guarantee that data is being used as users intend.
3. **Inclusivity**: Digital Green integrates inclusivity into Farmer.Chat in a variety of ways. The user interface of Farmer.Chat is designed to be simple to make it easy to use by users with limited digital experience. The tool works offline and with low-bandwidth phones. The organization also conducts in-person workshops to onboard and train users to ensure that they fully understand the tool's capabilities and are comfortable interacting with it. The team also designed a robust quality control and feedback mechanism for Farmer.Chat to maximize inclusivity. Users provide feedback (thumbs up or thumbs down) on messages from the bot and are able to immediately report biased or wrong responses. Digital Green also regularly speaks to users to collect more detailed feedback that they then incorporate into upgrades and new features for Farmer.Chat.

Keep humans in the loop.

In addition to these larger values, Digital Green maintains the human-in-the-loop method to ensure that the information that is communicated to farmers is appropriate and effective. The application is designed for extension workers to be the users of the chatbot, giving extension workers the autonomy to evaluate whether the output message is valid before passing on the information to individual farmers. This workflow design minimizes the spread of incorrect information by incorporating human input and oversight.

What's next

As a nonprofit organization, Digital Green operates mainly through grants. The team has submitted various proposals to expand Farmer.Chat, with the hope of building a foundational product that can integrate with a variety of third-party service providers. Through their extensive work with farmers and extension agents, Digital Green has noticed a desire for non-advisory agricultural content like current market prices and weather information. The team is looking to partner with both public and private sector organizations to develop additional functionalities and support that farmers need to make better decisions. Digital Green sees Farmer.Chat as one tool in a larger long-term strategy of building up agricultural extension networks. Their ultimate goal is for the tool to be fully institutionalized by country governments and serve as a DPG through continued partnerships and technical support.

RobotsMali: An education project using LLMs to encourage national language literacy in Mali

RobotsMali, a Malian nonprofit, used AI tools to create locally relevant school books in Bambara, one of Mali's local languages. Though an estimated 80% of Mali's population speak Bambara, few Malians can read or write the language. RobotsMali created books, complete with AI-generated images, to help students learn to read and write the language. The team was able to improve Bambara literacy amongst primary and secondary students through in-person lessons with these materials.

Background

RobotsMali is a nonprofit organization with a mission to support science, technology, engineering, and mathematics (STEM) education and culture in Mali. Headquartered in Bamako and founded by technologist and educator Michael Leventhal and youth empowerment activist Seydou Katikon in 2017, RobotsMali organizes trainings, conducts research projects, and develops curriculum to further STEM innovation and education for Malian youth. RobotsMali's activities and practices are rooted in evidence-based research and contextual understanding of the educational needs of Malian students. Their activities include:

1 Training: RobotsMali conducts trainings for children and youth through many mediums. They have partnered with numerous schools around the country to teach students about robotics and encourage them to learn more about the field. The school courses operate between one week and one full term, depending on the school. The team also runs a four-week intensive STEM & Robotics Camp every summer. Students from the third year of elementary school to the twelfth year of high school learn about a

wide range of technical topics, including robotics, AI, and programming. Additionally, the organization hosts a teaching certification program for STEM and robotics teachers, supported by the Malian government through the Ministry of Higher Education and Scientific Research. Participants in the program learn to teach STEM and robotics and leave with a wealth of knowledge and a teaching certificate.

2 Competition: Robotics competitions are a significant priority for the organization. They are a way for students to practise what they have learned, engage with international youth, and showcase their skills on an international stage. RobotsMali has supported national robotics teams to compete and place in select robotics competitions, including FIRST Global, Robotex, the Panafrican Robotics Competition (PARC), and the International Festival of Engineering, Science and Technology (IFEST). Malian teams have won 39 medals in these competitions, many of them gold.

3 Research: The RobotsMali team collaborates with Malian organizations on discrete research projects. For example, the YELENKOURA project, initiated and incubated at RobotsMali by inventors Drissa Diarra and Malick Traoré, is a tool to support people with visual disabilities to move safely in physical spaces. Designed in partnership with l'Union Malienne des Aveugles (Malian Union of the Blind), the technology identifies physical objects and obstacles via AI and sensor technology and notifies users via audio and tactile alerts. RobotsMali is also a research partner for Bayɛlɛmabaga, the first machine translator tool between national language Bambara and French. Due to the lack of machine translation tools on African languages, this team decided to take on the task of creating Bambara-specific tools. The team is collecting text and voice data for training AI models, developing data crowdsourcing tools and applications using AI models for education and other critical needs.

AI for Education Project

In 2023, the Bill & Melinda Gates Foundation released a Grand Challenge to identify "*innovative and safe approaches to the use of Large Language Models (ChatGPT-4, or other credible sources with equivalent capability)*" for low- and middle-income countries.[48] The RobotsMali team identified a variety of areas where they could use AI, specifically LLMs, to support Malian literacy and increase the availability of content available in Bambara, one of Mali's most widely spoken national languages.

48 Bill & Melinda Gates Foundation, "Catalyzing Equitable Artificial Intelligence (AI) Use."

"[Although Bambara is] spoken by an overwhelming majority of the Malian population, many Malians do not necessarily know how to read or write in Bambara."

Allahsera Auguste Tapo, RobotsMali Research and Development Lead

Given RobotsMali's existing work with schoolchildren, the team submitted a proposal to refine existing LLMs to create children's storybooks in Bambara. In July 2023 they were selected as a grantee for a three-month grant term. In their initial attempt to generate content, the team contacted several tech companies about accessing their base models, believing that proprietary models would outperform open-source models on Bambara. However, RobotsMali received no response. So the team shifted methods to use readily available resources.

First, they used ChatGPT to generate stories in Bambara, but the responses were nonsensical and unreadable. Their second attempt used ChatGPT to generate content in English and French and then translate that content into Bambara. The team also used the tool , in addition to other image generators like Stable Diffusion, MidJourney, DALL-E 2, and DALL-E 3, to generate images to match the text.

"When we asked the model to generate stories, [the stories were] heavily Eurocentric."

Allahsera Auguste Tapo, RobotsMali Research and Development Lead

This second method proved to be more successful but still needed extensive human resources to ensure that the content was culturally attuned and relevant. Many stories drew from stories and values found in societies in Europe and North America. The machine-translated Bambara content contained significant errors that required human editing. Many of the generated images also skewed Eurocentric. Tapo said that when the team tested image generators on random items, the results "*don't look anything like what we would find in Africa. [For] example, when you ask [the tools] to generate a kitchen … you will see a wonderful, beautiful kitchen that has no relevance whatsoever to Mali's context*." The images were often also slightly off (i.e., an animal with one eye) and had to be manually regenerated and altered.

"It's not a tool that we created but a pipeline of processes on how to generate stories in Bambara and make them context relevant and culturally attuned for Malians."

Allahsera Auguste Tapo, RobotsMali Research and Development Lead

"[Bambara] is an official language but it is not a language that people would really respect. They don't have any role models that have succeeded by learning or reading Bambara."

Allahsera Auguste Tapo, RobotsMali Research and Development Lead

Figure 8 ▾
RobotsMali books for various grade levels

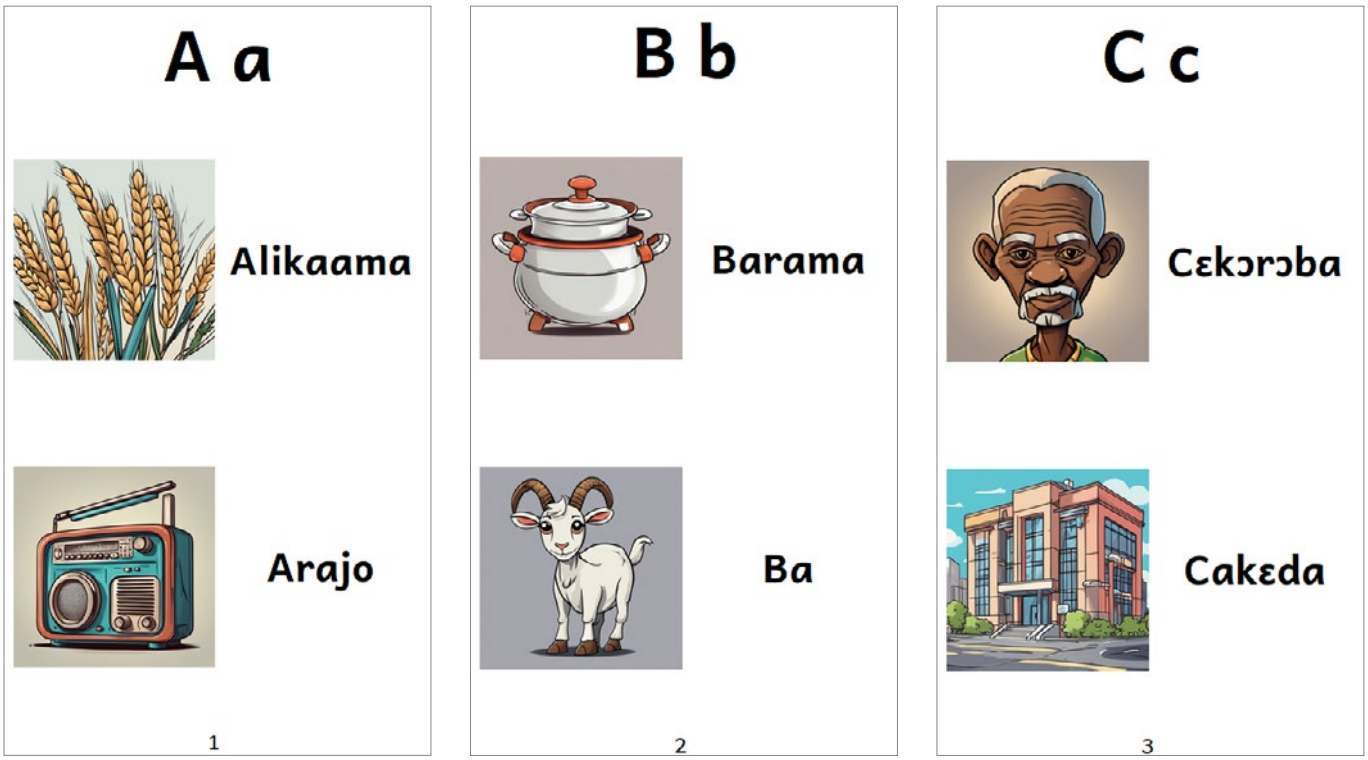


Figure 9 ▴
Screenshot of an early education RobotsMali book for learning the alphabet

Impact

Acceptance of Bambara

RobotsMali tested the edited storybooks with groups of primary and secondary school students, as well as and students who had never attended or had dropped out of school, who did not know how to read Bambara. The team encountered some pushback from parents and teachers, especially those who also did not know how to read Bambara and thus did not see the value for students to learn the language. This made RobotsMali's task more difficult. To address this barrier, the RobotsMali team met with parents before training sessions to introduce parents to the materials and discuss the importance of learning Bambara. The team also followed up with parents after the training sessions.

Anecdotal evidence shows that some students taught their parents to read Bambara using the RobotsMali stories they brought home from school. RobotsMali instructor Nouhoum Coulibaly, who facilitated learning sessions and supported image generation for the storybooks, also noted excitement from school administrators, parents, and students alike for the materials. Students and teachers stayed engaged during training sessions, displaying the desire for additional materials and staying attentive and present during learning sessions.

"They ask so many questions all the time, like 'Are you going to come back so that we can continue learning together?'"

Nouhoum Coulibaly, RobotsMali Instructor

New open-source educational materials

Though its period of performance was relatively short, the project was successful in two major ways. The first was the method the RobotsMali team used to create content. While limited resources prevented them generating content in Bambara, the team was able to use LLMs to create new content in a low-resource language. As Tapo said, the tool the team built was a process, not a product. Given the lack of text and information in Bambara, any new accessible texts and resources were a success for the project. The team created over 180 storybooks that are now publicly available on Bloom, an open-source application for hosting and making digital books.[49]

49 "BloomLibrary."

The storybooks have also had a wide-reaching impact on student learning. Over the course of the project, the RobotsMali team conducted learning sessions with students between the ages of 4 and 15 in homes, community centres, and classrooms. Sessions were conducted across the country, in urban and rural settings, and amongst students from different backgrounds. Some students came from households where parents have no formal education, and others came from households with higher education. In each session, trained facilitators introduced the materials, discussing the purpose of the stories and the goal of the sessions.

Improved staff digital literacy

Finally, the team points to the experience and growth amongst RobotsMali team members. In addition to conducting training sessions, Coulibaly also used DALL-E 3 and other text-to-image models to generate images to accompany the storybook text. He learned to use it across the length of the process and he described the process as "*a great experience, to work with a tool that really facilitates the task and production of literary documents*." Coulibaly expressed excitement over learning a new tool and being able to experience children's positive interactions with both the images and text. RobotsMali was able to influence the knowledge of both students and staff.

Improved efficiency

RobotsMali's method for creating storybooks and educational content is improving the speed and efficiency at which schoolbooks are developed. The use of generative AI has made it easier to create a lot of content over a short amount of time. Founder Leventhal estimates that RobotsMali's technique offers 50 times more savings compared to other techniques for creating schoolbooks.

Responsible AI

Promote AI for Africa.

RobotsMali's approach to using AI responsibly stems from the desire to see African people and languages as equal to their European and North American counterparts. They believe that the goal of the AI for Education project is, at its core, a practice in using AI for good and in a responsible manner. The AI for Education project uses the latest generative AI technology to fill in the gaps and make access to information more equitable.

"In the world, there are about 7,000 languages. On the African continent there are more than 2,000 languages. Google currently supports about a hundred languages, so we have a lot of work to do."

Allahsera Auguste Tapo, RobotsMali Research and Development Lead

Amplify generative AI with human ability.

"AI is a great tool, but it shouldn't be left alone to do whatever it wants. I can't just relax and hope that it comes up with the best solution or the best idea for you."

Allahsera Auguste Tapo, RobotsMali Research and Development Lead

The team also points to its chosen method for creating content as a key practice towards responsible AI use. Through initial testing, they recognized the pitfall of using ChatGPT for a low-resource language. ChatGPT's poor performance on generating content in Bambara proved the necessity of human intervention and support. The team recognized the need to keep the human in the loop to create stories that children would be excited to read and learn from. The team's manual edits and translation evaluations were crucial to create legitimate content.

AI ethics are for everyone.

The team also believes that responsible AI technologies need to be the interest of funders and governments. It is not enough for only researchers to concern themselves with responsible AI; governments also must step in. As barriers to entry for users of AI continue to diminish, RobotsMali believes that all stakeholders need to be fully invested and cognizant of the role that AI is playing in society.

"[Funders] should understand that they shouldn't help people to do only what funders want to do. Funders should take their [grantees] needs into account, what they are comfortable doing, and what they are not comfortable doing. … [Governments also] should be paying heed and waking up to the fact that AI is … going to come to them and either they take it seriously or they don't. If they take it seriously, it will come on their terms. If not, [AI] will just come, sweep governments aside, and do whatever it is that it wants to do."

Allahsera Auguste Tapo, RobotsMali Research and Development Lead

What's next

While the grant period for AI for Education has ended, RobotsMali is continuing to push for Malian innovation and AI resources. Their work providing students with robots and STEM training continues with the support of Malian government partners and external funders. The storybook project illustrated a well-known gap in machine translation tools that RobotsMali intends to fill. The team plans to work on improving the translation quality for Bambara, developing machine translation for the other 12 Malian national languages, building systems of speech recognition for Malian languages, and creating useful applications in education and other critical needs using AI models.

"As a community of researchers and funders, we have the opportunity to speak for the voiceless … In Mali, there are a lot of people who have great potential who are just lacking the ability to access the right information. What is lacking is them being able to access information in a language that they understand."

Allahsera Auguste Tapo, RobotsMali Research and Development Lead

ACADIC: Predictive AI tools implementing new ways to track disease spread on the African continent

Global interdisciplinary consortium ACADIC researches and develops new methods to apply existing machine learning techniques to new public health problems. The team came up with various tools, including a COVID-19 hotspot detection tool, an early warning system to notify governments of potential surges, and public dashboards, all powered by AI. Their work was used by country governments to inform COVID-19 policy and seen by millions of people globally.

Background

When COVID-19 hit the globe in 2020, communities and countries were rendered vulnerable to this relatively unknown disease. Governments, public health organizations, and medical staff had little real-time information about the spread of the virus and limited time and resources available to keep it under control. In response to the wide-spreading pandemic, the Africa-Canada AI & Data Innovation Consortium (ACADIC) at York University in Toronto, Canada, partnered with IDRC and SIDA to develop new technologies to track, analyze, and predict the spread of COVID-19.[50]

ACADIC, formed in January 2020 and led by York University professor Jude Kong, is an interdisciplinary consortium of public health professionals and software and data experts dedicated to reducing the burden of disease in African communities. The consortium includes individuals from Botswana, Cameroon, Canada, Eswatini, Mozambique, Namibia, Nigeria, Rwanda, South Africa, Zambia, and Zimbabwe. ACADIC's mission is to "*mobilize artificial intelligence and big data techniques in an ethical way to build equitable, resilient governance strategies and increase societal preparedness for future global pandemics and climate disasters*."

50 "Swedish International Development Cooperation Agency."

The team initially conducted standard exploratory data analyses of COVID-19 case data and used traditional epidemiological modelling methods like the SIRD model to track the spread of the disease.[51] The SIRD model categorizes a population into four groups: 1) people who are susceptible to the disease; 2) people who have been infected with the disease and can infect others; 3) people who have been infected with the disease and have healed from it; and 4) people who have succumbed to the disease.

This model can help epidemiologists and researchers estimate infection and recovery rates. But in the case of COVID-19, the needs of governments and public health partners quickly grew beyond what the SIRD model could provide. ACADIC turned to more advanced tools of prediction like ML to fill the knowledge gap amongst government stakeholders. ACADIC has created several products to track and predict the spread of the pandemic and its impact on healthcare systems using the latest data science and AI technologies.

"... through engagements with the government, [we realized that] what they needed kind of shifted a little bit. It was valuable for them to get case data predictions [and predictions of] how many hospital beds were going to be required in specific areas within the province."

Finn Stevenson, ACADIC Research Fellow

ACADIC AI technologies

ACADIC developed an AI tool to identify local COVID-19 hotspots. The team tested various models for detecting hotspots in Gauteng Province, South Africa, in order to help public health decision-makers make more informed and timely policy decisions. Hot spots are "*cluster[s] of cases within an area whose spreading dynamics do not conform to the general growth of the pandemic*" and have a higher risk of disease transmission, so identifying them is important. Working with the Gauteng Department of Health, ACADIC created a ML tool to identify geographic clusters and label them as hotspot/non-hotspot. The team trained a Gaussian mixture model on geo-coded COVID-19 data to automatically detect similar clusters of cases.[52]

51 Rapatski, "Mathematical Models of Diseases."
52 Lieberman et al., "Big Data- and Artificial Intelligence-Based Hot-Spot Analysis of COVID-19."

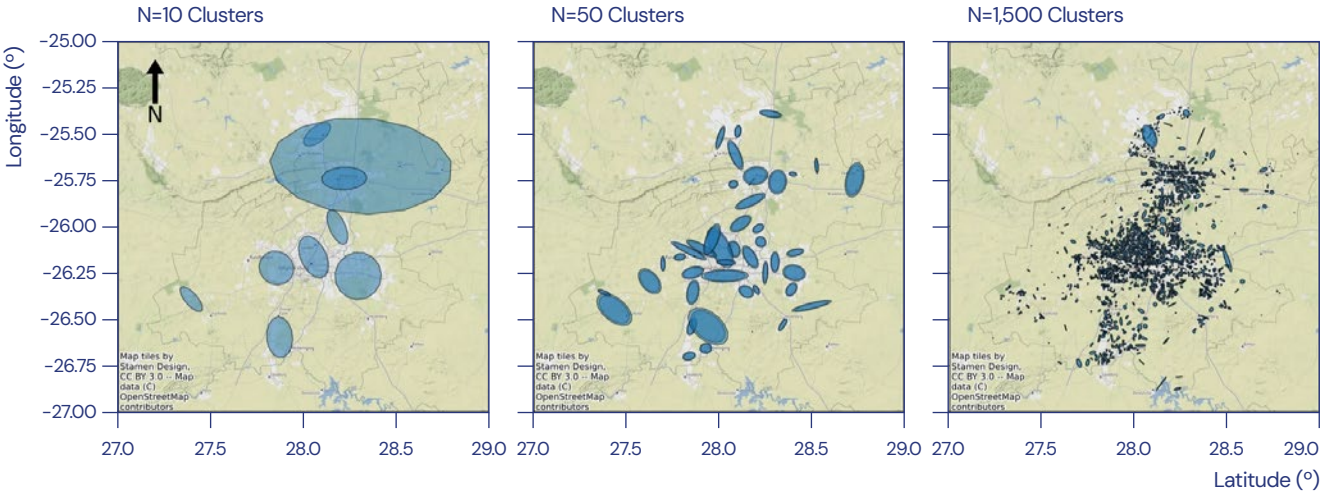


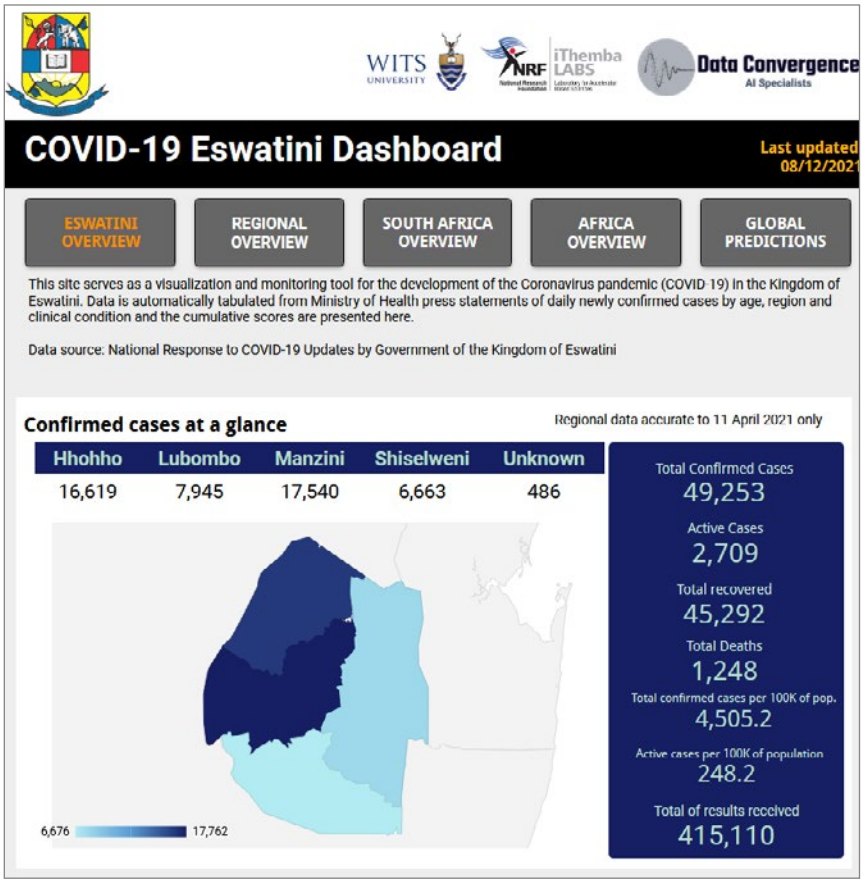
Figure 10 ▲
Maps of COVID-19 hot spots

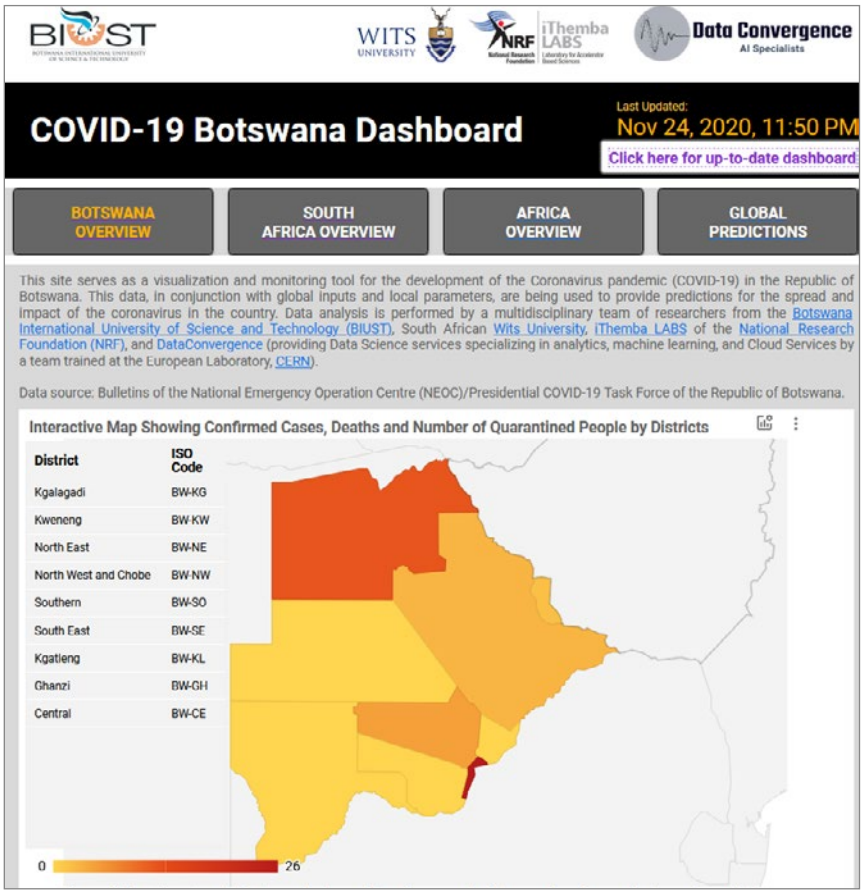
The results from their research offered a new method for real-time hotspot detection to help public health stakeholders reallocate resources toward hotspots and better inform the public on the course of the disease.

ACADIC also built an early warning system, a predictive AI tool, to alert governments, healthcare stakeholders, and the public of potential surges of COVID-19 over time with data from South Africa.[53] Advanced knowledge of surges can help public health decision-makers make more informed decisions to keep the public safe. The team identified a list of factors and correlated data sources that affect the spread of disease:

- Mobility: Google and Facebook mobility data provided a way to track movement of people over time.
- Stringency: Oxford University developed the Oxford COVID-19 Government Response Tracker to track the level or strictness of COVID-19 policies in a country.
- Epidemiology: The South African government collected and published data on new daily COVID-19 cases.

53 Stevenson et al., "Development of an Early Alert System for an Additional Wave of COVID-19 Cases Using a Recurrent Neural Network with Long Short-Term Memory."





◂ Figure 11
COVID-19 dashboards for Eswatini and Botswana

ACADIC trained a recurrent neural network (RNN) with LSTM with indicator data collected between the first and second COVID-19 waves in all provinces in South Africa. The model accurately predicted the start of the second wave, verified by real-time data, and was used by decision-makers for national surveillance of the third wave.

In addition to creating country-specific research products, ACADIC's interdisciplinary team harnessed the power of its international team to analyze COVID-19 vaccine uptake and determine inequities in vaccine adoption amongst low- and high-income countries. They first created a variable, the Vaccine Roll-Out Index (VRI), to estimate vaccination uptake rate in all 142 study countries. The team then identified 36 variables likely associated with vaccine adoption (e.g., age, income level, literacy rate, life expectancy). They used a random forest algorithm to model the relationships, and severity of relationships, between selected variables and countries' VRI.[54] Results from this study highlight inequality in global vaccine adoption and the specific disparity in vaccine adoption by income level. These results give policymakers quantitative evidence of inequitable vaccine distribution and serve as an important data point for working towards more equitable vaccine adoption.

ACADIC also built national-level dashboards to share data and information with the public. At the beginning of the pandemic, little was known about COVID-19 and the public was faced with conflicting information on COVID-19 and precautionary measures. The dashboards were a simple way to share data that would be useful and build trust with the public. Because of the rapidity of the work, the team was not able to conduct user tests or focus groups to improve the user experience of the dashboards.

Instead, ACADIC set up an email address to collect feedback directly from active users, many of whom viewed the dashboards on a regular basis and became very involved in providing feedback. Most of the data analyses and model outputs were made publicly available. While there were select data analysis outputs that were only available to the government, like availability of hospital beds, the team did their best to share as much information as they could with the public. The national dashboards were made available for South Africa, Botswana, Eswatini, Mozambique, Cameroon, and Nigeria.[55]

"Transparency was something that was strived for in the dashboards."

Finn Stevenson, ACADIC Research Fellow

54 Kazemi, Bragazzi, and Kong, "Assessing Inequities in COVID-19 Vaccine Roll-Out Strategy Programs."

55 "COVID-19 Botswana"; "COVID19 Eswatini"; "Official COVID-19 Mozambique"; "COVID 19 Cameroon Dashboard"; Africa-Canada AI & Data Innovation Consortium, "Nigeria."

Impact

ACADIC's main objective was "*to harness the power of gender responsive, responsible artificial intelligence and big data to provide locally relevant COVID-19 information in real time to specific urban and peri-urban communities in nine African countries: Botswana, Eswatini, Cameroon, Mozambique, Namibia, Nigeria, Rwanda, South Africa and Zimbabwe*."

The impact of ACADIC is apparent at multiple levels. At the height of the pandemic, ACADIC's South African dashboard was viewed by over 1 million people daily. Their other national dashboards were also viewed by over 50,000 people daily. While user numbers do not necessarily indicate use, they do indicate reach. In a time of great uncertainty about the spread of COVID 19, ACADIC was able to create a trusted tool that garnered widespread attention.

Trust

ACADIC also developed immense trust between the team, their work, and country governments. This is evident in national and local COVID-19 policies, ACADIC representation on national COVID-19 task forces, and reference to ACADIC tools in national COVID-19 policies. Their work influenced Canadian national and provincial COVID-19 policies regarding the Omicron variant. With national government partners and the Ontario COVID-19 Science Advisory Table, ACADIC used methods tested and validated in South Africa and on the African continent to estimate the disease burden in Canada. Their work with government stakeholders signifies the strength of the partnerships and confidence in ACADIC's analysis outputs.

Fast development

Given the rapidly evolving pandemic, the team had little time to monitor and evaluate the best tool or method for achieving their objective. ACADIC's suite of research products and tools were created to provide the government, other public health decision-makers, and the public with better information. Ultimately, the team believes that they accomplished their goal, despite the time constraints.

"... decisions were made during a very unpredictable time ... every decision was made in quite a lot of uncertainty and with a certain reasoning behind it that was driven by the demand for some answers and clarity."

Finn Stevenson, ACADIC Research Fellow

Responsible AI

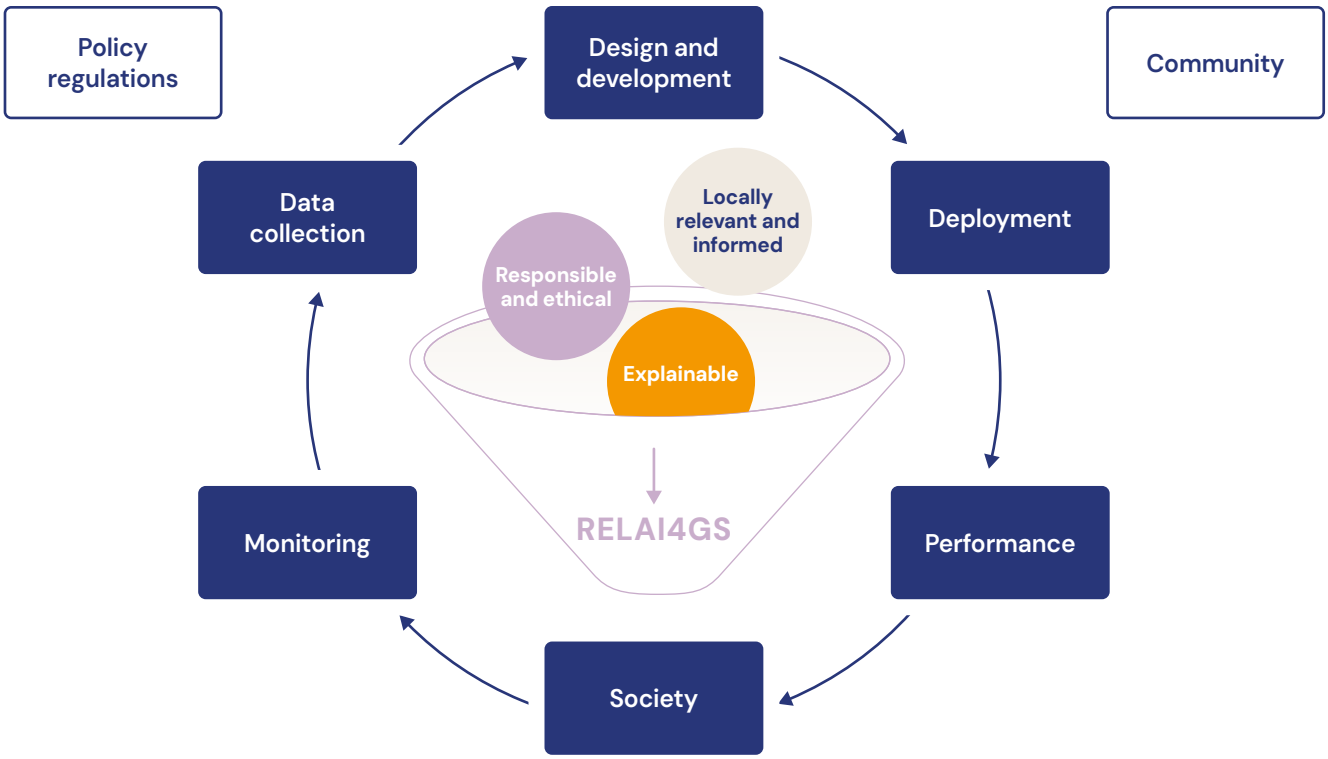
Create guiding values.

The ACADIC team created and published a framework for using AI that reflects the clinical public and global health needs in the Global South, called "Responsible, Explainable, and Local Artificial Intelligence for Clinical Public and Global Health in the Global South" (REL-AI4GS).[56] According to Kong and team, REL-AI4GS addresses a massive gap in ethical governance for AI development and implementation in the Global South. The framework relies on three main values:

1. Responsible: AI should be accountable, auditable, compliant, ethical, respectful, safe, and secure;
2. Explainable: AI should be equitable, fair, impactful, interpretable, and transparent; and
3. Local: AI should be decolonised, human- and community-centred, intersection, and inclusive.

According to ACADIC, these values should provide the foundation for the processes that AI stakeholders undergo to develop AI tools. The framework in Figure 12 provides a comprehensive illustration of REL-AI4GS in practice.

Figure 12 ▾
Responsible, Explainable, and Local Artificial Intelligence for Clinical Public and Global Health in the Global South framework



56 Kong et al., "Leveraging Responsible, Explainable, and Local Artificial Intelligence Solutions for Clinical Public Health in the Global South."

The three values—responsible, explainable, and local—shape the various processes involved in creating and maintaining AI, including the design and development process, data collection, and monitoring. These processes and values must be upheld by various AI stakeholders, including the community and society at large.

Combat inherent biases to prevent misuse.

"... the problem with AI is ... actually the problem with any new technology that comes to the market. It has to be used properly. If you don't use your technology properly, you can have misuse, or in the case of AI, biases."

Bruce Mellado, ACADIC Co-Executive Director

ACADIC worked diligently with partners to help them understand how to use their AI tools and reduce the bias that can often enter the system. They conducted many presentations and training sessions with government officials and public health decision-makers to prevent misuse.

What's next

ACADIC carried out a number of AI research projects during the height of the COVID-19 pandemic. Their work is available via publications in various research journals and applies existing AI technologies to new problems, offering original methods for quickly tracking, analyzing, and predicting the spread of disease. The predictive AI tools that ACADIC has developed can be used and reused during future public health emergencies.

Though the project has ended, members of ACADIC still regularly communicate and collaborate on other projects. ACADIC brought together an international and interdisciplinary team and built a platform for researchers, technologists, public health specialists, and government stakeholders to share information and skills that continues to exist long after the project's period of performance.

Plantix: A mobile app employing image recognition to improve farmers' productivity in India

Plantix is an image recognition mobile application, built by its eponymous technology company, that helps smallholder farmers improve crop yields and, ultimately, increase income and improve their livelihoods. The tool, intended to be used directly by farmers, processes user-uploaded pictures of crop diseases and pests and suggests solutions to remedy the issues. The tool has global reach, thanks in part to Plantix local research partner ICRISAT. ICRISAT provides training on the app to its network of smallholder farmers and develops new training sets to improve the accuracy and scope of the tool.

Background

Plantix is a mobile application that helps farmers make more informed decisions about their farming practices. Smallholder farmers often lack appropriate information and knowledge about crop diseases and uncommon pests. Agricultural extension workers and systems are under-resourced and over-burdened and unable to provide timely information that could help farmers save their crops. When their crops deteriorate, farmers cannot address the root problem and can experience crop loss and, in turn, income loss.

Plantix was created to help smallholder farmers quickly access information about crop damage and advise them on effective solutions. Its goal is to support farmers at all stages of crop problems, from prevention to treatment. Plantix was developed in 2015 by PEAT, a Berlin-based AI tech startup. Founders Simon Strey, Robert Strey, Alexander Kennepohl, and Pierre Munzel came up with the idea for Plantix after research trips during their graduate studies. On visits to rural areas, they interacted with farmers who constantly asked for their help identifying diseases and crop issues, sparking the idea

for a tool that automatically detected such information. The founding team initially developed Plantix supported by university funding; they successfully tested an image recognition approach on tomato plants grown in nutrient-deficient soil in greenhouses. They also enlisted friends to take pictures of plants around Berlin to supplement their training dataset. Eventually, the team created their own organization to fully build out the technology.

Photo by Plantix

Plantix in action

The Android-only app is powered by deep neural networks (DNNs) and ML algorithms that have been trained on annotated images of crops. These trained models are then used to detect user-submitted images. Users upload a picture of their diseased crops. The application then analyzes the image, identifies the issue (i.e., a pest or nutrient deficiency), and provides treatment options. Other features included in the application are:

- A real-time alert function that uses geolocation data collected from Plantix image uploads to warn other users in the area of potential disease outbreaks.
- An online community for farmers, scientists, and other agricultural experts to connect.
- A digital library containing over 700 types of plant damage for users to search and learn more.
- A search tool for users to find nearby retailers offering products that are suitable for solving their unique crop disease problem.

After successfully testing Plantix on lab-grown plants, the team sought to expand their approach for use in field conditions, ultimately deciding to continue their efforts in India. The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) became a key in-country agricultural partner to reach more users and obtain relevant content.

"Smartphone penetration was pretty good in India and mobile connectivity was among the best in … areas where you would have small-scale farmers. There [are] not many crops which aren't being grown in India [and there is] a lot of different diversity as well regarding languages. So we [thought that] if we can make it in India, we might also have a good concept for the rest of the world."

Phil Knake, Plantix Product Manager

Photo by Plantix

ICRISAT is an international nonprofit CGIAR research organization focused on agricultural development in the drylands of Africa and Asia. Founded in 1972 by a consortium assembled by the Ford and Rockefeller Foundations, ICRISAT conducts research in three areas:

1. Accelerated crop improvement: ICRISAT tests new technologies and methods to develop and improve crop varieties that are more nutritious, fruitful, and resilient. Under this domain, ICRISAT produced chickpea, millet, and pigeon pea varieties that have higher drought tolerance and nutrition.
2. Resilient farm and food systems: The team develops new strategies for more efficient and productive farming systems, like the use of satellite imagery and ground sensors to model crop yields.
3. Enabling systems transformation: The research and activities under this pillar address the environment for the adoption of more sustainable and resilient agricultural practices. Activities carried out involve new policy recommendations that benefit smallholder farmers and capacity-building workshops and resources for women and youth.

ICRISAT has developed a suite of digital tools to help farmers, extension workers, and other agricultural professionals increase their resilience to climate change and other environmental challenges and improve their livelihoods. For example, the Intelligent Agricultural Systems Advisory Tool (ISAT) uses historical climate data, weather conditions, soil conditions, and other factors to advise farmers on precise and effective agricultural practices.[57] The advisories are delivered directly to farmers via SMS and in their local languages. Originally built in partnership with Microsoft and the India Meteorological Department, ISAT was piloted with seven villages in India before expanding. The development team has recently incorporated new AI modelling capabilities to improve the accuracy and effectiveness of the advice ISAT provides to farmers.

Photo by Plantix

ICRISAT utilized their extensive network of research labs and universities to train over 3,500 students and staff on the Plantix app. Together, the team built an image database of crops and crop damage. Government agricultural workers, students of agricultural universities, and farmers in India were, and continue to be, involved in the development and improvement of the tool. Their expertise has helped with identifying training images, annotating images with relevant information, and improving the app's user experience. ICRISAT also provides key feedback from direct users to the Plantix development team. This ensures that the tool remains useful and adapts to the changing needs of its users and enabling digital environment.

Additionally, ICRISAT's existing relationships with extension workers, smallholder farmers, and government agencies helped to institutionalize and expand the reach of the application in the Indian agricultural sector. ICRISAT conducts training sessions with agricultural extension workers and smallholder farmers to introduce Plantix and guide them through use of the tool. Trainers demonstrate the application in farmers' local languages (mainly Hindi and Telugu) and show off the different functionalities and features.

57 ICRISAT, "Intelligent Agricultural Systems Advisory Tool (ISAT)."

Impact

The primary focus for the Plantix and ICRISAT organizations is bettering the livelihoods of smallholder farmers. However, the nature of the application prevents the two teams from comprehensively measuring the impact of the tool on smallholder farmers. Farmers typically engage with the tool for short periods of time with limited opportunities for user feedback. As a result, the teams use proxy metrics (like number of users) collected via the application and supplement this data with occasional, resource-intensive impact studies.

Reach

The impact of Plantix is apparent in its popularity. Plantix has been downloaded more than 30 million times and hosts about 10 million users annually. Since its release, Plantix reports it has used AI to analyze more than 100 million crop images from users and is among the most downloaded agriculture technology apps in the world.[58]

Technical accuracy

According to the team, Plantix has an accuracy of 90% across all crops and plants, with performance on some crops achieving up to 95% accuracy. Due to the submission of new crop images from partners like ICRISAT and millions of Plantix users from over 100 different countries, the development team is able to improve the accuracy of the tool through continuous training and testing.

Improvements in farmers' knowledge and productivity

Plantix has conducted in-depth user studies via partner 60 Decibels. In 2021, the study team conducted phone interviews with more than 400 smallholder farmers who use Plantix (and by extension, its AI-powered image recognition) in Maharashtra, Telangana, Madhya Pradesh, and Andhra Pradesh. The study aimed to determine the impact of Plantix on its users and found that:

- Farmers experienced increased crop production, higher income, and improved agronomic knowledge thanks to Plantix.
  - Nearly 90% of study respondents reported an increase in crop production because of Plantix.
  - Over 75% of study respondents reported higher earnings from crops because of Plantix, mainly due to larger crop volumes. Crop sales are a key source of household income for many farmers.

58 "Plantix."

- Farmers also reported an improvement in farming information and knowledge of methods due to use of Plantix.
  - 87% of farmers reported that using Plantix has led to an improvement in their way of farming.
  - Over half of respondents said that they have better knowledge of farming practices, like watering methods, crop rotation, and plant-to-plant distance.

A second study, also conducted by 60 Decibels, was recently done to evaluate Plantix in relation to many other agriculture technology applications; results are slated to be released late 2024.

Responsible AI

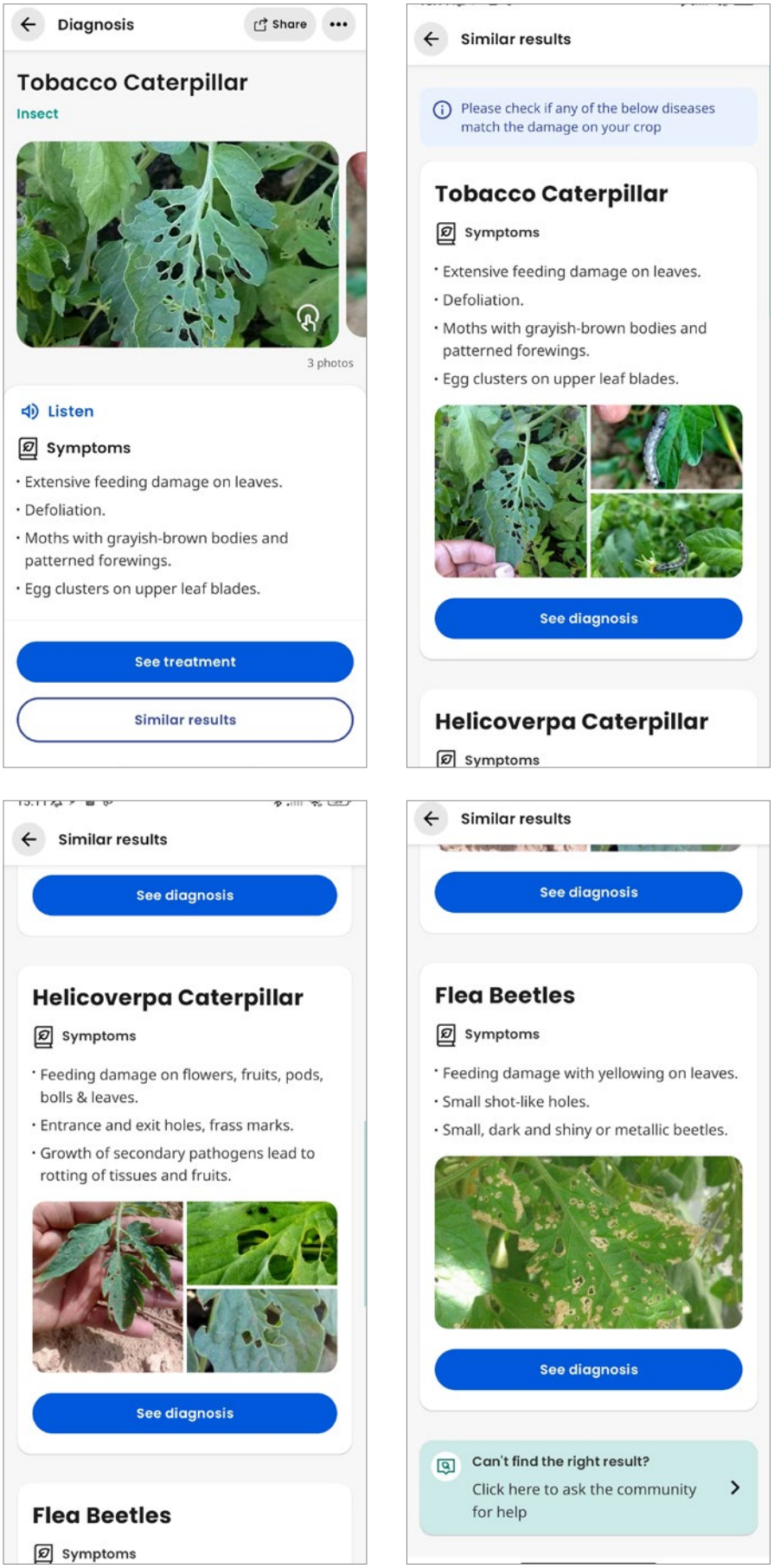
Though neither Plantix nor ICRISAT has codified principles specific to their AI work, both organizations follow responsible AI practices in their respective roles building and maintaining Plantix.

Create guiding values.

As a CGIAR research organization, ICRISAT abides by CGIAR's various policies, including their Research Ethics Code, in its work and AI implementation.[59] The Research Ethics Code, which is regularly updated and approved by the CGIAR System Board, provides guidance on best, ethical research practices for all 15 CGIAR international centres. The code rests on five values:

1. **Integrity:** We are honest, tell the truth, keep promises, pursue objective scientific research, admit mistakes, earn trust, and always act professionally by being accountable and transparent.
2. **Dignity and Respect:** We value and embrace diversity and inclusion, treat all stakeholders with respect and dignity, promote equity, avoid all forms of discrimination, and promote human rights .
3. **Sustainability:** We plan responsibly for the long term, and are committed to environmental, social and economic food security, safety and global prosperity.
4. **Excellence and Innovation:** We strive for excellence by maintaining high standards of scientific rigor, actively encouraging innovation and creativity, and pursuing our passion for learning and discovery.
5. **Partnership:** We value the diverse voices of our internal and external stakeholders, and seek all forms of engagement, collaboration and teamwork.

59 CGIAR System Management Office, "CGIAR Research Ethics Code."



◂ Figure 13
Plantix identification flow for a high-performing crop disease

In its trainings and interactions with smallholder farmers, ICRISAT follows these principles.

Plantix does not have an established framework for using AI responsibly, though some early app decisions and functions illustrate responsible AI in practice. The developers incorporate choice into its plant identifications.

"Early on, we were really hesitant to directly detect disease … We were mostly providing likely options."

Phil Knake, Plantix Product Manager

In the early days of Plantix, the application offered options for each user submission. Users would submit photos of crops and Plantix provided a list of potential identifications with detailed descriptions for users to choose from on one page. Though the Plantix team tried its best to match human expert–level accuracy, they were aware that the application still had (and continues to have) potential inaccuracies and can make misidentifications.

Over time, the team has made small changes to this workflow. Through continuous training and evaluation, the team identified specific crop diseases that Plantix was performing better than human experts on. For these diseases, the application now suggests one main option and offers other potential options that users can access on a separate page. The team's confidence in its AI identification has grown, and changes made to the application reflect that.

Build community among users.

"[We] always wanted to make the farmer an active decision-maker."

Phil Knake, Plantix Product Manager

The application encourages users to consult with other farmers and agricultural experts via a community forum. Integrating user decision-making into the process requires that farmers play a part in identifying crop diseases, maximizing the likelihood of correct identification.

What's next

"Accuracy is the challenge as you keep adding more pests and diseases."

Srikanth Rupavatharam, ICRISAT

The team is testing new AI technologies to forecast disease spread and severity. The development team noticed drops in accuracy with the addition of new crop problems and asked ICRISAT for fresh training datasets to remedy the issue. In response, ICRISAT is conducting greenhouse growing experiments to generate crops with select diseases to create the datasets. These accuracy problems will likely be solved in the next year or so, after the crop cycles are completed.

Discussion: From responsible practice, impact

The case studies shared in this report were not specifically comparative; each stands on its own as a representation of a unique pathway to responsible impact with AI. The sample is an intentionally heterogeneous selection of organizations, rather than a representative one.[60] The eight cases are drawn from a variety of geographies and sectors, and they are using different AI technologies in different ways.

For these reasons, we need to be clear-eyed about how crosscutting findings can be taken from the studies. It is tempting, but ultimately not useful, to frame findings as "Five of eight organizations did X while only two did Y." Instead, this reflections section identifies a few patterns and contrasts only insofar as they trigger ideas for further study and make new connections across disparate parts of this emerging space.

The two themes for discussion are responsible practice and the impact that the organizations are having with AI. We close with some general reflections and an identification of next steps.

60 Palinkas et al., "Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research."

On responsible AI practices

Common responsible practices

Conversations with these organizations illustrate a variety of ways in which they have created guiding practices to ensure they are engaging with AI responsibly. We discuss practices rather than principles. Principles are abstract and universal, proclaimed and inscribed. In contrast, practices are people-specific, organization-specific, invented, and discovered. These case studies revealed several common practices of responsible AI.

Do no harm (go slowly).

Many organizations are taking a slow approach to using AI and have strict guidelines for integrating AI into their products. EIDU, for example, conducts extensive testing of their personalization models for months before releasing the product to its full user base. The team wants to ensure that new updates improve upon its existing product and have no negative impacts on any learner or groups of learners.

Nivi is similarly cautious about integrating AI into their chatbot, evidenced by their choice to refrain from sending AI-generated healthcare content to users. Though LLMs have recently grown in both popularity and accuracy, Nivi is skeptical of their ability to provide relevant and appropriate information to its users, especially healthcare information. The risk of giving users wrong or potentially harmful information is too high for Nivi, and the team is unwilling to put the health and livelihoods of its users in danger.

Keep humans in the loop.

Keeping humans in the loop means not letting machines make decisions without oversight and guidance. Several organizations emphasized the importance of this practice—some even referenced it as the first (and sometimes only) way to approach responsible AI. Nearly all organizations keep humans in the loop. While AI auto-tags documents behind the scenes, UPR Info staff review and approve the tags before they are published to the public Database of UPR. Similarly, the Jacaranda help desk team reviews and edits generative AI content before replying to mothers seeking healthcare via the PROMPTS platform.

Since it will be a while (if ever) before these systems can be trusted to work autonomously, particularly when dealing with sensitive data or risky outcomes, the slow, careful approaches to deploying AI described by many of these eight organizations should be commended and replicated.

Pursue fairness for all.

All organizations mentioned the importance of fairness for their users. Many organizations highlighted the different techniques they use to ensure that their AI tools treated their intended users fairly. For EIDU, that means regularly auditing their personalization model to ensure that it performs well for all demographics of students. Similarly, Nivi's research into gender bias in its AI tools is another key example of evaluating and pursuing fairness for its users. Nivi intentionally biases their AI tool to ensure fairness for all users, regardless of their gender identity.

Other organizations, like RobotsMali, have fairness baked into the tasks they perform; RobotsMali is using AI to make language education and literacy writ large fairer and more accessible.

Share with the community of practice.

Interviewees consistently referenced the need to share best practices and tools with the wider development community. For many, that looks like engaging in the open-source community to share and use proven, successful tools. Jacaranda, for example, fine-tuned Meta's Llama 2 language model on Swahili and pushed the model to Hugging Face, an open-source library of AI resources. Many organizations started from the same place, building upon existing open-source models like Llama 2 and GPT-2 to build their AI tools.

Creating responsible practices

These case studies reveal as much about how the organizations derive their practices of responsible AI as they do about the practices themselves. Here are three findings about the emergence of responsible AI practices in the development sector.

Approaches to responsible AI should be bespoke, but also connected.

Interviews revealed differences in organizations' approach to responsible AI. Some, like Jacaranda Health and Digital Green, have established frameworks and values as foundations, while others appeared to be more reactive about how they approached AI.

Importantly, every organization has landed on their *own* set of practices. None explicitly connected their responsible AI practices to broader principles of digital development, nor to any of the principles of responsible AI currently circulating in the policy community, though there is overlap. We did not ask and cannot discern whether interviewees have read these principles, nor whether they might be informing the individuals responsible for enacting/inventing organizational actions. But the opportunities to increase the connections between global principles and the implementation of local practices is worth noting and exploring further.

The bottom-up practices listed above—"go slowly," "keep humans in the loop," "pursue fairness for all," and "share with the community of practice"—are very helpful. But even as a set, they are likely incomplete, and few of the organizations we spoke to mentioned all four. Thus, even when organizational practices do align with broader principles of responsible AI, they may not be comprehensive or universally applied.

There is a great deal of work still to be done to evangelize global principles on responsible AI, increasing their salience and actionability within the frontline organizations that can do the most to enact them.

Responsible AI practices depend on an organization's position in the stack.

Within institutions, different kinds of decision-makers engage with practices of responsible AI. In some cases, technical and/or sector experts with previous experiences with AI had a very clear idea of both what they wanted to accomplish with AI and the practices that needed to be to responsibly accomplish that goal. In other cases, organizations were deploying AI for the first time and did not have strong prior experiences to guide them. As AI becomes more accessible, many organizations will need to create their practices in real time.

Different institutions showed different strategies in how responsible practices were scoped and propagated between tool creators and tool implementers. Though all organizations interact with users in some shape or form, some have a more direct line and engage in constant interaction.

This is to say nothing of the broader AI stack—the tools behind the tools, including underlying LLM models and ML algorithms, that are even further away from user experience, but are nevertheless important elements of the overall experience and functionality. They, too, have a role to play in promoting the responsible use of AI, if not of AI in development specifically.

Practitioners can share practices with the broader community.

Because AI for development applications stretch across multiple organizations, frequently relying on underlying tech, developers, and implementers, it is clear that responsible practice has to work at two levels. Within organizations, teams need to do all they can to make sure that the ways they use the tech align with the principles of responsible practice. At the same time, organizations need to coordinate and share learnings across organizations, with other parts of the "value chain" or "the stack" that structure how end users and organizations interact with these powerful technologies, and with the sectoral communities of practice to which they belong. Indeed, participating in a study like this one is a clear example of sharing practices.

It is encouraging that among the eight cases profiled here, there is evidence of sharing of practices beyond the confines of each organization, promoting transparency.

On impact

The second focus of this study is the impact that can be attributed to AI. Discussing this impact can be challenging for three reasons. First, few of these organizations are "AI native." AI is a part of their intervention strategy rather than the entirety of it, it can be difficult to discern the impacts that are specific to AI. Second, the technologies (or at least their implementation) are relatively new. This is particularly the case with interventions that draw on LLMs and generative AI, which only began to enter mainstream application in 2023.

Finally, the organizations are not always configured to measure and document impact at the same time they are optimizing for it. Most approaches to impact don't (yet) involve the certainty—and commensurate investment in resources—afforded by a randomized control led trial. Nevertheless, there are key signals and patterns to discuss. We do so below.

Eight organizations and a dozen SDGs

The eight organizations report delivering impact across at least a dozen of the UN SDGs. Table 1 provides an overview of reported impact and the SDGs they address, illustrating a wide range of engagements using AI.

Table 1 ▾

Summary of findings and case studies

Case	Project	Impact of AI	Impact measurement practices	Responsible AI practices	SDGs
askNivi (Nivi)	Retrieval-based chatbot, improving access to local healthcare	Increased user engagement, contributing to increased access and use of healthcare services	Gender audits	Human Evaluations Equitable outcomes	3.7 3.8 5.6
HURIDOCS/UPR Info	An AI-optimized database lowering barriers to information for human rights defenders	Increased team efficiency, from 2 months to 2 weeks	Tracking tool's impact from both technical and social perspectives	Human checks Redundancy	16.3 16.6 16.7
PROMPTS (Jacaranda Health)	An AI-powered message triage tool strengthening primary healthcare in Kenya	Increased help desk team efficiency, improved access to critical care	Qualitative feedback from staff	6 things/ principles Humans in the loop	3.7 3.8
EIDU	Using content personalization to maximize learning in Kenya	Significant improvement in learning outcomes (SD=0.43) determined from RCT	A/B testing	Test for fairness/bias Beta testing	4.1 4.2
Farmer.Chat (Digital Green)	A generative AI chatbot supporting agricultural extension workers around the world	More effective, informative interfaces available to 13,000 users and extension agents	Focus on the farmer	3 principles: transparency, privacy, inclusivity	2.3 2.4
RobotsMali	Using ChatGPT to create locally relevant school books, encouraging local language literacy in Mali	Increased engagement with learning materials created with AI	Training sessions Qualitative feedback	Humans in the loop and extensive discussion of how to use AI for equitable outcomes	4.1 4.2
ACADIC	Machine learning to track COVID-19 in South Africa	Speed of product creation, reach of products	Engagement numbers	Published REL4AI framework	3.b 3.d
Plantix/ICRISAT	Improving farmers' productivity in India with image recognition	Over 10 million users Small-scale evidence of improvement in agronomic knowledge	Engagement and retention numbers Qualitative in-app feedback One-off studies on user impact	CGIAR's Research Code of Ethics	2.3 2.4

Three kinds of impact

The table illustrates a variety of approaches to impact amongst the development organizations profiled in this study. These approaches can be grouped into three types: 1) efforts to increase scale and efficiency; 2) improvements to outcomes and experiences for end users; and 3) expansions of AI practices for the broader development community.

Increasing scale and efficiency

The most commonly reported impacts of AI were perhaps the most tangible and easy to observe, because they were internal to the organizations. Among the applications profiled in this study, some of the most discernible impacts were found in the increased efficiency reported by teams at HURIDOCS and Jacaranda Health. Organizations described how AI was being deployed, responsibly and carefully, to help those within the organizations to do their jobs more quickly and effectively, whether by increasing the number of queries a staff member could process, or reducing team members' response times to emergency messages.

"AI allows us to provide responsive support when addressing pregnancy related concerns and delivers swift personalized assistance during critical solutions. We have been able to be very efficient in responding to these questions with the growing volumes of mums enrolled on the platform."

Pauline Nafula, Jacaranda Senior Help Desk Coordinator

"Thanks to the machine learning features that automatically extract data from the documents and categorize data, UPR Info staff members were able to dedicate their time to implement other types of activities. In particular, in the last years we have been focusing a lot to strengthen the capacities of civil society actors by empowering, for example, human rights defenders through training sessions and also through the development of new tools or good practices that can support their engagement in the UPR."

Nicoletta Zappile, UPR Info Deputy Director

These firsthand observations echo global findings and trends; organizations of many types are reporting productivity gains, particularly among knowledge workers.[61] But these gains might not be enough; others are expressing concerns that the productivity gains due to AI are not yet commensurate with the investments organizations are making in the technologies.[62] The trends in AI's overall contribution to productivity

61 Maslej et al., "Artificial Intelligence Index Report 2023."

62 *The Economist*, "What Happened to the Artificial-Intelligence Revolution?"; Goldman Sachs, "Gen AI: Too Much Spend, Too Little Benefit?"

are encouraging but not yet crystal clear. Indeed, across the eight cases profiled in this study, the impacts attributed to older "predictive" AI were more discernible and documented than changes associated with the emerging impacts of newer generative AI approaches. But for predictive and generative AI alike, these contributions are neither irrefutable nor sufficiently documented within the information-intensive parts of the development sector, as shown in the interviews conducted for this study.

Improving user outcomes and experiences

There are, however, some indications in the cases studies of how AI can improve outcomes for individual end users. One pathway is via improvements to interfaces, such as askNivi and Digital Green, both of which show how AI can drive engagement and accessibility across language and literacy divides. Another pathway is via building AI directly into the delivery of an intervention: identification of plant diseases made with Plantix that could not be identified via mobile phone without the use of AI.

But assessing and quantifying these these impact pathways for end users is a different matter from simply describing them. This measurement task is urgent, since it is ultimately the outcomes on individuals and communities (rather than on implementing organizations) that animate the development community. This measurement task is necessary because AI impact is not always positive, and there is greater risk of harms to individuals interacting with AI systems. This measurement task is complex, since it can be difficult to isolate the impacts of AI from other factors around an intervention. Yet unfortunately, this measurement remains rare.

Expanding AI for the development community

The third kind of impact is best described as participation in the collective effort to improve the application of AI in development. Each organization can make a choice: to go it alone quietly, or to participate in a broader community of practice. As evidenced by (at a minimum) participation in this study and, more broadly, in practices reported by many of the organizations, the community is getting stronger all the time.

> *"We recognize the global learning crisis is far too big for any one organization to tackle. Instead, we are aiming to build the open-source tools that any implementer or any government can use to replicate the impact we've already achieved in Kenya. If we invest sufficient resources to build AI technology that's tailored for lower income environments, then there is a chance for this technology to benefit millions of people around the world. If instead, we hope that solutions that are being developed in high income environments or high-tech environments can easily just be imported or replicated in different environments, we risk missing out on this opportunity of AI being a net benefit for the international development field."*
>
> **Aidan Friedberg, EIDU Director of Learning and AI**

"Ninety percent of what we use is open source. We have a few things that are our own but it's small compared to the rest of the development."

Tomàs Andreu, HURIDOCS CTO

"Making sure that humans are in the loop and then also that whatever we learn has to be shared back with the community. And what I mean by that is any, like, for example, the language models that we create have been open sourced and shared openly, freely with the community for use not only in maternal and newborn health use cases, but in any case so they can be fine-tuned, for example, for agriculture or education or finance. We've developed the models. There's no reason for us to hold on to them and we want to release them openly to be able to benefit other organizations."

Jay Patel, Jacaranda Health Director of Technology

Another way development organizations using AI tools can strengthen the community of practice is by engagement with DPIs and DPGs—both as users of those systems and as contributions to them. Multiple organizations talked about the importance of DPGs. HURIDOCS's Uwazi software is a certified DPG and the team creates tools with the intention of sharing them with the wider community. Jacaranda uses RapidPro, a DPG, for PROMPTS's underlying messaging framework. Notably, while some organizations specifically called out DPGs in their work, almost all organizations noted the importance of open-source tools. The use and creation of open-source tools and infrastructure help shape the political economy of AI in ways that promote transparency and sustainability.

Strengthening impact practices

In tandem with the discussion of creating responsible AI practices, this section reflects on the impact (measurement) practices described by the organizations. These practices are distinct from the impacts themselves—but they are linked because, as mentioned above, it is difficult to ascertain what impacts may be without plans to measure them along the way. It's not novel to point out that many organizations in development, like those in this study, often work without quantifiable evidence that their efforts are changing outcomes, let alone with guidance from the RCTs that have become the "gold standard" of the development community. These eight case studies suggest three challenges to strengthening impact measurement practices around AI for development.

First, **it is still early**, particularly for applications using generative AI. Documenting the impacts of AI, either on organizational productivity or directly for "end users" and communities, will take time.[63] In contrast to some organizations that have been working with predictive AI for several years, all of the organizations in this study using generative AI are still in the early stages and have not yet had the time to evaluate the full potential of AI-generated content on their users.

"We're really early in the implementation of generative AI in the agricultural space, in general. But for us, AI just has a potential to be a net benefit. How can we calibrate this techno optimism by keeping our eye closely on the experience of small-scale farmers? This is what guides us, our North Star, and we do want to meet their needs. We just see a lot of promise there."

Jona Repishti, Digital Green

Irrespective of the time it takes for AI to spread within the development community, **impact costs money to measure**. We did not specifically ask case study participants about the resources they dedicate to measuring AI impacts, but documented, quantifiable impact evidence is an investment. Even "low-cost" RCTs using existing data in policy and development sectors can run US$50,000 or more,[64] a large sum for many organizations. Other, lower-cost approaches, like causal process tracing,[65] might be more feasible while also yielding key insights about how AI's impacts are embedded in broader, more complex interventions.[66]

Finally, **it is easier for organizations with direct connections to users/communities to measure impact**. Yet not all of the organizations interviewed had direct engagement with end users. As described in the section above on responsible AI practices, interventions that ultimately result in an AI intervention that can 'impact' a community involve several layers, from LLM providers and open-source ML algorithm publishers, through AI-specific companies and startups, to implementing partners and frontline organizations. AI impact may require partnerships among these originators and their customers/implementing organizations, as those with frontline access may be well positioned but under-resourced to carefully track the impacts of new AI interventions.

63 Smith, "Ensuring Generative AI Is Responsible AI."
64 Haskins and Feldman, "Low-Cost Randomized Controlled Trials."
65 Collier, "Understanding Process Tracing."
66 Partnership for Finance in a Digital Africa, "Approaches to Determining the Impact of Digital Finance Programs."

Helping organizations assess the impact of AI

For all the reasons above, rigorous quantitative evidence on the impact of AI interventions unfolding in real time is likely to remain precious and rare. A reliance on A/B testing and rapid adjustments is understandable. Without a control condition, A/B testing struggles as an impact assessment methodology, but it does provide key feedback to implementing organizations that enables them to quickly (re)design and improve their offerings.

That said, this is a key moment for the community as a whole to support evidence gathering. Many of the organizations we spoke to for these case studies predate the AI boom; they have added AI to existing workflows rather than emerging as "AI native" organizations. With this long view are the roots of, if not RCTs, then at least, pre-test/post-test ways of learning. Put another way, several of the case studies described here at least have the benefit of having worked on development interventions before the advent of AI. We are at a moment where some of the best learning about how to use AI effectively to advance the SDGs comes from organizations whose engagement predates the adoption of AI tools. Further explorations from more actors about pre- and post-AI ways of working can further inform the AI impact conversation, even if RCTs and more statistically valid approaches to impact measurement remain out of reach for smaller initiatives

All told, this reflection on impact is not intended to be an admonition, but rather a call to help more organizations get access to the measurement and impact tools they need to iterate. Indeed, at the risk of being recursive or turning the mirror around, AI technologies promise to assist in the design and delivery of lower-cost impact evaluations, just as they are important topics for evaluation themselves.[67]

67 Mason, "Finding a Safe Zone in the Highlands."

Conclusion

In summary, these case studies, representing organizations working across several parts of the sector, reflect change underway, rather than change completed.

The work on **responsible AI in practice** shows practices emerging, in place to help organizations navigate the power (and new vulnerabilities) that come with working with opaque and probabilistic AI technologies. A key challenge is how unconnected to the broader focal principles many of these emerging practices of responsible AI seem to be.

The work on **impact in practice** underscores how most organizations are confident that they are driving impact and are using A/B testing and other forms of engagement to refine what they are doing, adjusting and sharing all at the same time. It's a unique moment in that they can recall and contrast their processes before using AI, even if in most cases we surface a key challenge in translating that confidence and qualitative data into more quantitative tests.

Case studies like these often identify key questions that can be addressed by other research designs. In this case, at least three avenues for further inquiry present themselves:

First, a closer look at the costs involved when development organizations implement AI would be a strong contribution to the evidence base. It is important to further quantify both the investments and resulting savings associated with AI in the sector, as well as any implications for the skill sets and work patterns of those closest to the implementation.

Second, there is more to understand about the interplay between AI adoption and the growing role of digital public goods (DPGs) and digital public infrastructures (DPI). Of course this interplay is not limited to the development sector, but as the case studies here have indicated, the sector can be a leader in forging and strengthening these linkages, both via the tools and underlying technologies they select, and via their development and refinement of tools and technologies others will use.

Finally, while this study focused on documenting emergent practices rather than testing awareness of global principles of AI, the circuits between practices and principles, between local and global, and between implementation and amplification, is worth exploring in more detail. This study suggests that the promotion of shared meaning, shared perspectives, and shared language amongst implementing organizations, research, and policy communities may be a key to ensuring greater uptake and replication of responsible AI.

In the meantime, both of the challenges of impact and responsible practice should be understood as opportunities, in which the resolution lies in

greater community and connection—across parts of the value chain, and among the broader community of practice that is AI in development. Organizational choices, though contributing to DPIs and DPGs, and contributing best practices, help shape the emerging political economy of AI, particularly as it intersects with the development sector. The case studies in this report show eight organizations working responsibly to use AI to advance important development outcomes. We're grateful for their time, their insights, and their contributions to this community of practice.

References

Adams, R., F. Adeleke, A. Florido, L. G. de Magalhães Santos, N. Grossman, L. Junck, and K. Stone. "Global Index on Responsible AI 2024 (1st Edition)." South Africa: Global Center on AI Governance, 2024. https://coral-trista-52.tiiny.site/.

Ade-Ibijola, Abejide, and Chinedu Okonkwo. "Artificial Intelligence in Africa: Emerging Challenges." In *Responsible AI in Africa: Challenges and Opportunities*, edited by Damian Okaibedi Eke, Kutoma Wakunuma, and Simisola Akintoye, 101–17. Social and Cultural Studies of Robots and AI. Cham: Springer International Publishing, 2023. https://doi.org/10.1007/978-3-031-08215-3_5.

Africa-Canada AI & Data Innovation Consortium. "Nigeria." https://www.yorku.ca/science/mathstats/acadic/nigeria/.

AfricAI. "AfricAI Conference 2023." https://conference.ai4d.ai/.

Aly, Heidi. "Digital Transformation, Development and Productivity in Developing Countries: Is Artificial Intelligence a Curse or a Blessing?" *Review of Economics and Political Science* 7, no. 4 (January 1, 2020): 238–56. https://doi.org/10.1108/REPS-11-2019-0145.

"Bhashini." https://bhashini.gov.in/en/.

Bill & Melinda Gates Foundation. "Catalyzing Equitable Artificial Intelligence (AI) Use." https://gcgh.grandchallenges.org/challenge/catalyzing-equitable-artificial-intelligence-ai-use.

Björkegren, Daniel. "Artificial Intelligence for the Poor." *Foreign Affairs*, August 9, 2023. https://www.foreignaffairs.com/world/artificial-intelligence-poor.

Bloom, Benjamin S. "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring." *Educational Researcher* 13, no. 6 (1984): 4–16. https://doi.org/10.2307/1175554.

"BloomLibrary." https://bloomlibrary.org/.

"CASH Coalition – Climate Action for Smallholders." https://cashcoalition.earth/.

CEDIB [Centro de Documentación e Información Bolivia] and CONTIOCAP [National Coordinator for the Defense of Native Indigenous Territories and Protected Areas]. "Mapa de Ataques a Defensoras Ambientales." https://basedefensoras.cedib.org/.

Centre for Intellectual and Property Law. "Mapping of AI Applications in Africa." Strathmore University, May 21, 2020. https://cipit.strathmore.edu/artificial-intelligence/.

CGIAR System Management Office. "CGIAR Research Ethics Code." CGIAR, November 2020. https://cgspace.cgiar.org/items/b47956c2-cb0f-4b45-aa7b-7de16abfaf8e.

Collier, David. "Understanding Process Tracing." *PS: Political Science & Politics* 44, no. 4 (October 2011): 823–30. https://doi.org/10.1017/S1049096511001429.

Cooper/Smith. "Deploying GenAI in the Real World." *Cooper/Smith* (blog), April 29, 2024. https://medium.com/cooper-smith/deploying-genai-in-the-real-world-3df5346e3994.

Daltry, Rebecca, Louis Major, Chen Sun, Mary Otieno, Kevin Otieno, and Rebecca Daltry Otieno Louis Major, Chen Sun, Mary Otieno and Kevin. "Digital Personalised Learning to Improve Literacy and Numeracy Outcomes in Kenyan Classrooms." *EdTech Hub* (blog), August 15, 2023. https://edtechhub.org/2023/08/15/emerging-findings-from-a-randomised-controlled-trial-involving-pre-primary-learners-and-eidu/.

"Digital Green." https://digitalgreen.org/.

Digital Green. "Farmer.Chat." https://farmerchat.digitalgreen.org/.

Digital Green. "Loop Mobile App Makes Farm to Market Linkages Easy," January 15, 2017. https://digitalgreen.org/loop-mobile-app-makes-farm-to-market-linkages-easy/.

The Economist. "What Happened to the Artificial-Intelligence Revolution?" *The Economist*, July 2, 2024. https://www.economist.com/finance-and-economics/2024/07/02/what-happened-to-the-artificial-intelligence-revolution.

"EIDU – Our Classroom Is a Billion Strong." https://eidu.com/.

EU Artificial Intelligence Act. "Article 22: Authorised Representatives of Providers of High-Risk AI Systems." https://artificialintelligenceact.eu/article/22/.

'Frontier Counties Development Council." https://fcdc.or.ke/.

Gandhi, Rikin, Rajesh Veeraraghavan, Kentaro Toyama, and Vanaja Ramprasad. "Digital Green: Participatory Video for Agricultural Extension." Bangalore, India, 2007. https://doi.org/doi.org/10.1109/ICTD.2007.4937388.

Goldman Sachs. "Gen AI: Too Much Spend, Too Little Benefit?" June 25, 2024. https://www.goldmansachs.com/intelligence/pages/gs-research/gen-ai-too-much-spend-too-little-benefit/report.pdf.

"Guiding Principles – Responsible AI." https://rai.ac.uk/guiding-principles/.

GSMA. "AI for Africa: Use Cases Delivering Impact."; London, UK: GSMA, July 2024. https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/wp-content/uploads/2024/07/AI-for-Africa-Use-cases-delivering-impact.pdf.

Haskins, Ron, and Andrew Feldman. "Low-Cost Randomized Controlled Trials." Washington DC: Evidence-Based Policymaking Collaborative, November 2016. https://govinnovator.com/wp-content/uploads/2020/05/low-cost_rcts.pdf.

Heeks, Richard. "ICT4D 3.0? Part 1—The Components of an Emerging 'Digital-for-Development' Paradigm." *EJISDC* 86, no. 3 (2020): e12124. https://doi.org/10.1002/isd2.12124.

HURIDOCS. "Home." https://huridocs.org/.

HURIDOCS. "Uwazi Receives Accolade of Being Recognised as a Digital Public Good by the Digital Public Goods Alliance." HURIDOCS, August 26, 2022. https://huridocs.org/2022/08/uwazi-receives-accolade-of-being-recognised-as-a-digital-public-good-by-the-digital-public-goods-alliance/.

ICRISAT [International Crops Research Institute for the Semi-Arid Tropics]. "Intelligent Agricultural Systems Advisory Tool (ISAT)," February 2024. https://www.icrisat.org/assets/Flyers/Thematic-Flyers/ISAT-flyer.pdf.

IHRDA [Institute for Human Rights and Development in Africa]. "African Human Rights CLA 2.0." https://caselaw.ihrda.org/en/.

Ipas. "Home Page." https://www.ipas.org/.

Jacaranda Health. "Home," September 8, 2023. https://jacarandahealth.org/.

Jacaranda Health. "Jacaranda Launches First-in-Kind Swahili Large Language Model." October 31, 2023. https://jacarandahealth.org/jacaranda-launches-first-in-kind-swahili-large-language-model/.

Jacaranda Health. "PROMPTS: Promoting Mothers in Pregnancy and Postpartum Through SMS." 2023. https://jacarandahealth.org/ypoagriw/2023/11/PROMPTS-BROCHURE-V4.pdf.

Jones, Rachel M., Grace Kimenju, Shalini Subbiah, Amy Styles, Nicholas Pearson, and Sathyanath Rajasekharan. "A Short Message Service (SMS) Increases Postpartum Care-Seeking Behavior and Uptake of Family Planning of Mothers in Peri-Urban Public Facilities in Kenya." *PLOS ONE* 15, no. 9 (September 30, 2020): article e0239213. https://doi.org/10.1371/journal.pone.0239213.

"Karya | We Solve Data Needs." https://www.karya.in/.

Kazemi, Mehrdad, Nicola Luigi Bragazzi, and Jude Dzevela Kong. "Assessing Inequities in COVID-19 Vaccine Roll-Out Strategy Programs: A Cross-Country Study Using a Machine Learning Approach." *Vaccines* 10, no. 2 (February 2022): article 194. https://doi.org/10.3390/vaccines10020194.

"Kenya Agricultural and Livestock Research Organization." https://www.kalro.org/.

Khare, Pratyush. "Exploring Skill-Based Matchmaking Systems in Online Games." *Artificial Intelligence in Plain English* (blog), December 25, 2023. https://ai.plainenglish.io/exploring-skill-based-matchmaking-systems-in-online-games-96491816d9e7.

Kong, Jude Dzevela, Ugochukwu Ejike Akpudo, Jake Okechukwu Effoduh, and Nicola Luigi Bragazzi. "Leveraging Responsible, Explainable, and Local Artificial Intelligence Solutions for Clinical Public Health in the Global South." *Healthcare* 11, no. 4 (February 4, 2023): article 457. https://doi.org/10.3390/healthcare11040457.

Korinek, Anton, and Joseph E. Stiglitz. "Artificial Intelligence, Globalization, and Strategies for Economic Development." Working Paper, National Bureau of Economic Research, February 2021. https://doi.org/10.3386/w28453.

Lieberman, Benjamin, Jude Dzevela Kong, Roy Gusinow, Ali Asgary, Nicola Luigi Bragazzi, Joshua Choma, Salah-Eddine Dahbi, et al. "Big Data- and Artificial Intelligence-Based Hot-Spot Analysis of COVID-19: Gauteng, South Africa, as a Case Study." *BMC Medical Informatics and Decision Making* 23, no. 1 (January 26, 2023): 19. https://doi.org/10.1186/s12911-023-02098-3.

Looker Studio. "COVID 19 Cameroon Dashboard." http://lookerstudio.google.com/reporting/941ae4a2-b3da-497c-9195-d52a8ab5659a/page/p_ui7hdp8zqc?feature=opengraph.

Looker Studio. "COVID-19 Botswana." http://lookerstudio.google.com/reporting/0f498d3b-2d44-4893-92cd-dd373f0a3f73/page/wI9JB?feature=opengraph.

Looker Studio. "COVID19 Eswatini." http://lookerstudio.google.com/reporting/b847a713-0793-40ce-8196-e37d1cc9d720/page/2a0LB?feature=opengraph.

Looker Studio. "Official COVID-19 Mozambique." http://lookerstudio.google.com/reporting/b521f3d0-83dd-4430-90ac-d6bff7120ac5/page/wI9JB?feature=opengraph.

Maslej, Nestor, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, et al. "Artificial Intelligence Index Report 2023," October 5, 2023. https://doi.org/10.48550/arXiv.2310.03715.

Mason, Sarah. "Finding a Safe Zone in the Highlands: Exploring Evaluator Competencies in the World of AI." *New Directions for Evaluation* 2023, no. 178–179 (2023): 11–22. https://doi.org/10.1002/ev.20561.

Nivi. "Nivi | AI Powered Chatbot for Healthcare." https://www.nivi.io.

Nivi. "Nivi Joins the USAID Funded FHM Engage Consortium as the Technology Partner to Drive Demand and Consumer Insights on Sexual and Reproductive Health." February 20, 2024. https://www.nivi.io/post/nivi-joins-the-usaid-funded-fhm-engage-consortium-as-the-technology-partner-to-drive-demand-and-cons.

OECD.AI Policy Observatory. "AI Principles Overview." OECD. https://oecd.ai/en/principles.

Okolo, Chinasa T., Kehinde Aruleba, and George Obaido. "Responsible AI in Africa—Challenges and Opportunities." In *Responsible AI in Africa: Challenges and Opportunities*, edited by Damian Okaibedi Eke, Kutoma Wakunuma, and Simisola Akintoye, 35–64. Cham: Springer International Publishing, 2023. https://doi.org/10.1007/978-3-031-08215-3_3.

"Onebillion: The Goal - One Billion Children Reading." https://onebillion.org/about/.

Palinkas, Lawrence A., Sarah M. Horwitz, Carla A. Green, Jennifer P. Wisdom, Naihua Duan, and Kimberly Hoagwood. "Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research." *Administration and Policy in Mental Health* 42, no. 5 (September 2015): 533–44. https://doi.org/10.1007/s10488-013-0528-y.

Partnership for Finance in a Digital Africa. "Approaches to Determining the Impact of Digital Finance Programs." Evidence Snapshots. Farnham, Surrey, UK: Caribou Digital Publishing, April 23, 2018. https://www.cariboudigital.net/wp-content/uploads/2024/03/FiDA-Snapshot-15-Approaches-to-determining-the-impact-of-digital-finance-programs.pdf.

Population Council. "Home." https://popcouncil.org/.

Rapatski, Brandy. "Mathematical Models of Diseases." Essential Elements, Fall 2021. https://stockton.edu/sciences-math/ezine/fall2021/mathematical-models.html.

"RapidPro." https://app.rapidpro.io/.

Schwab Foundation for Social Entrepreneurship. "AI for Impact: The Role of Artificial Intelligence in Social Innovation." Geneva, 2024. https://www3.weforum.org/docs/WEF_AI_for_Impact_Social_Innovation_2024.pdf.

Selvaraj, Natassha. "What Is Retrieval Augmented Generation (RAG)? A Guide to the Basics." Data Camp, January 2024. https://www.datacamp.com/blog/what-is-retrieval-augmented-generation-rag.

Shah, Saureen, and Apurva Joshi. "COCO: A Web-Based Data Tracking Architecture for Challenged Network Environments." *Proceedings of the First ACM Symposium on Computing for Development*, December 17, 2010, 1–7. https://doi.org/10.1145/1926180.1926189.

Shipra Saxena. "What Is LSTM? Introduction to Long Short-Term Memory." Analytics Vidhya, August 6, 2024. https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/.

Sida. "Swedish International Development Cooperation Agency." https://www.sida.se/en.

Smith, Matthew. "Ensuring Generative AI Is Responsible AI: The Case for Allowing Time to Find the Evidence." *IDRC Perspectives* (blog), May 11, 2023. https://idrc-crdi.ca/en/perspectives/ensuring-generative-ai-responsible-ai-case-allowing-time-find-evidence.

Stevenson, Finn, Kentaro Hayasi, Nicola Luigi Bragazzi, Jude Dzevela Kong, Ali Asgary, Benjamin Lieberman, Xifeng Ruan, et al. "Development of an Early Alert System for an Additional Wave of COVID-19 Cases Using a Recurrent Neural Network with Long Short-Term Memory." *International Journal of Environmental Research and Public Health* 18, no. 14 (January 2021): article 7376. https://doi.org/10.3390/ijerph18147376.

Transitional Justice Working Group. "Footprints: Database of Those Taken by North Korea." https://nkfootprints.info/.

UN General Assembly. "Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development," March 11, 2024. https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf.

UNESCO. "Recommendation on the Ethics of Artificial Intelligence." https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

UPR Info. "Database of UPR." https://upr-info-database.uwazi.io/.

USAID [US Agency for International Development]. "Equitable AI Challenge and Community of Practice: Sourcing Approaches to Advance Gender Equity in AI," June 25, 2024. https://www.usaid.gov/innovation/competitions/equitable-ai-challenge.

Weinberg, Friedhelm. "Geneva-Based HURIDOCS Is a Google AI Impact Grantee." *HURIDOCS* (blog), May 7, 2019. https://huridocs.org/2019/05/geneva-based-huridocs-is-a-google-ai-impact-grantee/.

Appendix

Invitation script

This project, in partnership with IDRC, is aiming to highlight successful use cases of responsible artificial intelligence in international development. Though there is a new rush to deploy AI technologies across the world in the wake of ChatGPT, international development projects have been taking advantage of these technologies for the past decade with the intention to do good. Our goal with this work is to feature these projects who have implemented discrete AI use cases resulting in measurable impact and learn from their successes and failures.

What we are asking for:

- Project team participation in at least two hour-long recorded interviews, ideally one with project leadership and one with technical staff.
- Funding partner participation in one hour-long recorded interview.

The projects and learnings will be published in a written report and select case studies will be featured in a video docuseries. Both will be featured on IDRC's website and offer the opportunity to showcase your work to the entire development community. There will be opportunity for projects to review the draft report and offer specific comments for discussion on their respective case study before publication. However, we ask that, in the name of open development and information sharing, projects and participants offer up as much as they can to help the development community move forward in its understanding and implementation of AI technologies.

Use case interview guide draft

Subtopic A: Project scope and narrative

- Can you give a brief overview of your project and its current status?
- What problem(s) are you trying to solve with this project?
- What value does AI bring to solving the problem(s) that could not be achieved through other solutions? What were the other solutions that you considered implementing? How was this problem being addressed before your project came along?
- Who makes up your project team? (Number of team members, roles, partners, etc. – probe for diversity in gender, geographic representation, different disciplinary backgrounds etc.)
- Since you are providing much of the data science expertise, what other subject matter experts did you have to work with to make sure this project went well?
- Who did you see as key stakeholders in this project and why? What was the process for identifying them? How did you engage those key stakeholders? When do you engage those stakeholders throughout the project lifecycle?
- How did you make sure that the data were collected and stored securely?
- Can you give an insight into the conversations that you had around data collection, data storage, and data standards throughout the course of the project? (probe for data representativeness for the context, probe for data readiness for the project, ask about ongoing challenges with maintaining and updating datasets as the project continues)

Subtopic B: Impact

- What is your definition of success for your project/technology/product?
- Has that changed over time and how has it changed?

Subtopic C: Process for tracking impact

- How do they know what impact they are having?
- Do they have a theory of change?

Subtopic D: Lessons learned and challenges

- What are the top 3 lessons that you learned throughout the course of the project?
- What are the top 3 challenges that you encountered throughout the course of the project?
- What was something that surprised you or was counterintuitive that you could not have learned about before digging into the project?
- If you could restart the work knowing what you know now, what would you do differently?
- Knowing what you know now, was AI a suitable approach for this project? Does it offer improvement over the status quo?
- Do you think incorporating AI is expensive? How are you thinking about longer term sustainability of the project?
- What do you wish you had known at the beginning of this project? Ex: partnerships, knowledge gaps, etc.
- How can the AI enabling environment/ecosystem in the context of your project (policies, infrastructure) be strengthened to support responsible AI uptake?
- What are you most proud of/greatest accomplishment from this project?
- What is next for the project?

Caribou Digital

IDRC · CRDI

Canada

UK International Development

Partnership | Progress | Prosperity

ARTIFICIAL INTELLIGENCE FOR DEVELOPMENT AFRICA

September 2024