Explainable AI in Practice: Practitioner Perspectives on AI for Social Good and User Engagement in the Global South

Chinasa T. Okolo
Center for Technology Innovation, The Brookings Institution
United States of America
cokolo@brookings.edu

Hongjin Lin
Computer Science, Harvard University
United States of America
hongjin_lin@g.harvard.edu

Abstract

AI for Social Good (AI4SG) has been advocated as a way to address social impact problems using emerging technologies, but little research has examined practitioner motivations behind building these tools and how practitioners make such tools understandable to stakeholders and end users, e.g., through leveraging techniques such as explainable AI (XAI). In this study, we interviewed 12 AI4SG practitioners to understand their experiences developing social impact technologies and their perceptions of XAI, focusing on projects in the Global South. While most of our participants were aware of XAI, many did not incorporate these techniques due to a lack of domain expertise, difficulty incorporating XAI into their existing workflows, and perceiving XAI as less valuable for end users with low levels of AI and digital literacy. We conclude by reflecting on the shortcomings of XAI for real-world use and envisioning a future agenda for explainability research.

CCS Concepts

• Human-centered computing → User studies; **• Computing methodologies → Artificial intelligence**; **• Applied computing**;

Keywords

AI for Social Good, Explainable AI, Human-Centered Design, Artificial Intelligence, Social Impact

ACM Reference Format:
Chinasa T. Okolo and Hongjin Lin. 2024. Explainable AI in Practice: Practitioner Perspectives on AI for Social Good and User Engagement in the Global South. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '24), October 29–31, 2024, San Luis Potosi, Mexico.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3689904.3694707

1 Introduction

The "AI for Social Good" (AI4SG) movement has become popular amongst AI researchers and practitioners interested in addressing social issues in low-resource domains [2, 92]. While significant progress has been made in this field, there is a shortage of work critically examining how AI4SG researchers identify and engage with end users during the design and deployment process. Given existing concerns around the negative impacts of AI in low-resource contexts [11, 14, 99], there is also an urgent need to study AI4SG practices and understand efforts enabling the responsible development of AI4SG tools. Explainable AI (XAI) is particularly significant in this context because it aims to bring transparency and accountability into AI systems, enabling users to investigate key qualities (e.g., accuracy, feature importance) of outputs from AI systems [29, 90]. However, existing challenges with the usability and reliability of XAI may inhibit how practical these methods could be for practitioners and end users situated in the Global South [70]. Given that the concepts of AI "responsibility", "transparency," and "accountability" widely encompass various aspects and how XAI is a key facet of prominent responsible AI frameworks from the U.S. National Institute of Standards & Technology [67], Google [4], Accenture [3], and others, our work examining practitioner experiences with XAI provides insights into broader "responsible" AI efforts, particularly in revealing the limitations of using XAI in real-world contexts.

Note that we use the term "Global South", which incorporates Africa, South(east) Asia, Latin America, the Caribbean, and Oceania, throughout the paper for lack of better terms. We acknowledge that any classification of countries risks creating a false hierarchy among nations and ascribing a higher value to some lives. For the purpose of this work, we use the term to describe regions of the world where the uses and perceptions of AI tools are different from wealthy and industry-focused settings due to relatively low economic and industrial development levels and a larger proportion of marginalized and vulnerable populations.

An emerging area of work focuses on evaluating practitioners' current practices and needs when engaging in responsible AI practices, providing valuable knowledge to shape AI development [27, 53, 56, 61, 95, 98, 104]. However, most of this work centers on AI practitioners working in industry and situated in Western contexts. The challenges associated with deploying AI technologies in the Global South, including digital infrastructure deficiencies [71], a lack of access to relevant datasets [1], and a lack of local AI developers [74], underscore the necessity for researchers to examine current AI practices to help shape the future development of inclusive AI tools. Our work builds upon existing research, particularly from Okolo et al. [75], and contributes to the field of human-centered AI by *focusing on AI researchers and practitioners explicitly working on social impact problems and engaging with users in the Global South.* Given the limited amount of studies on perceptions of XAI in the Global South [73, 75], it is important to examine

EAAMO '24, October 29–31, 2024, San Luis Potosi, Mexico

ACM ISBN 979-8-4007-1222-7/24/10
https://doi.org/10.1145/3689904.3694707

explainability practices to better understand ethical considerations around the fairness, trustworthiness, and potential misuse of AI systems deployed in this region.

To address this gap, we conducted an interview-based study to answer the following research questions:

RQ1: How do AI for Social Good practitioners identify problems and engage with end users when designing and building AI systems for use in the Global South?

RQ2: How do AI for Social Good practitioners working in Global South contexts perceive the usefulness of model explainability for themselves and their end users?

RQ3: How do AI for Social Good practitioners consider and implement model explainability in designing, developing, and deploying their AI systems in the Global South?

To answer these questions, we conducted an interview-based study surveying 12 AI4SG practitioners on their experiences developing tools for social impact in the Global South and their perceptions of explainable AI. Our findings show that most AI4SG practitioners identified problems and solutions with end users and stakeholders in mind, aligning with standard practices in human-computer interaction for development (HCI4D) and information and communication technologies for development (ICTD) research. While most of the AI practitioners we interviewed were aware of the concept of explainability, many of them did not incorporate specific explainability techniques in their work due to several challenges: 1) a lack of specific and practical experiences in XAI limiting their ability to use explainable AI methods effectively and accurately, 2) the limited amount of existing tools that allow them to incorporate these techniques into their workflows easily, and 3) explainable AI not being considered as a primary objective of their projects.

All of the AI practitioners we interviewed believed that model explainability would be helpful for their work, especially because it would help them develop more accurate models and communicate their research findings more clearly in publications. However, not all AI practitioners believed explainable AI would be useful for their target end users due to the challenges they faced when communicating AI model outputs to users with relatively low levels of data and digital literacy. This implies that XAI methods alone are insufficient in addressing the challenge of communicating with end users with different domains of expertise and varying levels of technical fluency. Addressing the root causes of the digital divide between the Global North and Global South is crucial for bridging the gap of AI fluency, such as investing in local AI developers and inclusive AI development [74]. Taken together, the contributions of our work include the following:

- We examine how AI4SG practitioners identify social good problems and select AI techniques to address them.
- We provide novel insights detailing how AI4SG practitioners consider model explainability when developing and deploying their technologies.
- We present actionable considerations on reshaping existing notions of explainability to meet the needs of AI4SG practitioners and their end users.

2 Background

2.1 AI for Social Good (AI4SG)

Over the past decade, AI4SG has emerged as a field to address social inequity by leveraging emerging technologies such as AI, ML, and data science. Many of these projects are motivated by the Sustainable Development Goals [66], an initiative by the United Nations to improve global development through 17 goals. With thousands of projects around the globe, AI4SG work has focused on domains such as agriculture [58, 64, 89, 96], conservation [57, 68, 77, 102] education [28, 44, 65], healthcare [33, 35, 105], and humanitarian aid [7, 26, 48]. Our work examining AI4SG practitioners' perspectives on explainable AI is also situated in existing literature in the field of ICTD, which focuses on designing and building technology-based solutions for users in resource-challenged environments [93]. Given the very similar objectives of AI4SG and ICTD, findings from decades of work within ICTD, especially as it pertains to developing and deploying technologies in responsible ways [18, 23, 94], provide valuable lessons for AI4SG practitioners to learn from. Overall, such work holds promise for reshaping current approaches to AI4SG development and further enabling user-centered practices. Our work builds upon this research by interacting directly with AI4SG practitioners who work on projects within the Global South to understand their experiences designing and developing AI tools for social impact.

2.2 Responsible AI in the Global South

Issues of algorithmic bias within primarily Western contexts have motivated initiatives towards "Responsible AI", which aims to develop better sociotechnical solutions to AI by critically examining existing technologies and developing guidelines [6], toolkits [34, 62, 91], and frameworks [16, 78] to shape the development of algorithmic systems. While these methodologies have been instrumental in affecting the responsible development of AI, impacting conference guidelines [19], industry practices [37, 40, 41], and even regulation [22, 69], the majority of these benefits have been reserved for users in Western contexts. As the development of AI systems continues to expand beyond Western settings, perspectives from marginalized backgrounds and low-resource contexts must be examined and prioritized within the design and implementation of AI technologies, as evidenced by an emerging area of research focused on examining how AI systems impact users in non-Western contexts [32, 46, 75, 81, 86, 97]. Influential work in this field by Sambasivan et al. [86] finds that algorithmic bias is often under-analyzed in Indian contexts and that factors such as caste, gender, and religion should be re-contextualized when examining issues of algorithmic fairness in India. Work by Okolo et al. [75] reviews the current state of XAI research in the Global South, finding sixteen papers applying XAI to social-impact problems in domains such as healthcare, education, and finance, and highlights the importance of developing XAI that meets the needs of diverse users in the Global South. Our work contributes to this emerging subfield of AI by focusing specifically on XAI to understand how these techniques are used within AI4SG work.

2.3 Practitioners' Perspectives on Responsible AI

A growing body of work in HCI has examined the practices and needs of AI practitioners as they address Responsible AI issues through the design and development process, focusing on industry settings [42, 61, 82]. Various guidelines [6, 76] and toolkits [12, 101] have been made available to support practitioners in integrating Responsible AI concepts into their project pipelines [63]. However, many challenges remain for practitioners to effectively apply available tools given time and resource constraints on top of their existing workloads [25, 42, 61, 82, 103]. This previous work reveals important directions for Responsible AI design to meet practitioners' needs in industry settings and Western contexts. However, this focus on Western contexts may not be relevant in low-resource settings in the Global South, where additional challenges like the lack of AI practitioners and computing resources pose a challenge to implementing AI. Our study focuses on the context of AI4SG in the Global South, which has been shown to present unique challenges and opportunities in contrast to industry settings in the Global North [73, 86]. We thus build on prior work engaging with AI practitioners to understand how AI4SG practitioners conducting work in low-resource regions perceive XAI and how they implement XAI in practice.

2.4 Explainable AI

XAI generally aims to enable developers and end users to understand how ML models operate and how these models produce specific predictions. For example, feature importance, a widely used XAI approach, indicates how much each feature (a measurable attribute of a particular model) contributes to the model outputs after they become available [60, 84]. Popular methods that leverage feature importance include LIME, which trains surrogate models to estimate predictions from the primary model [84], and SHAP, which calculates the contribution of each feature within a model to a specific prediction [60]. Locally explainable methods like LIME and SHAP can help developers understand what features a model relies on to make a specific prediction. Globally explainable models like decision trees [59] can provide developers with a broader understanding of which features contribute most to all predictions a model produces, allowing them to debug and improve model performance. However, the benefits of XAI have been shown to lean towards those with the technical knowledge to implement and interpret these methods, posing a disadvantage to less technical users. XAI methods have been shown to encode trust in incorrect decisions [9, 50] and provide conflicting explanations [51]. Given these limitations, it is crucial to examine how XAI is used in real-world deployments of AI systems and the associated challenges of using XAI in low-resource contexts as elucidated by Okolo et al. [75].

Since an explanation is only meaningful if a human decision-maker can understand it, a growing body of work recognizes the importance of a user-centered approach to XAI [30]. Many works interviewed practitioners (e.g., UX designers, data scientists), users, and stakeholders to understand gaps in existing XAI tools [13, 27, 29, 47, 53, 56]. Despite an emerging body of research, most research within XAI and human-centered XAI continue to focus on users and stakeholders within Western contexts. Thus, it remains an open question as to how practitioners working on AI4SG projects in the Global South perceive XAI and its usefulness for their end users, collaborators, and AI researchers like themselves. Given the strong interest in using AI to address social problems, the limited research focusing on AI in the Global South omits valuable perspectives and introduces the possibility for bias to compound. Our study aims to contribute further to this domain by engaging AI4SG practitioners who work in the Global South to understand their experiences developing AI solutions and perspectives on XAI.

3 Methodology

To answer the research questions we developed for this study, we conducted interviews from mid-May to June 2023. This section details our participant recruitment strategy, interview procedure, and data analysis.

3.1 Participants

Our study recruited AI4SG practitioners in academia, industry, non-profit, and non-governmental organizations. We classify an "AI4SG practitioner" as an AI developer, researcher, or designer who creates and implements AI solutions that aim to address social impact problems. While academic researchers are not traditionally considered to be "practitioners", work conducted in ICTD is often conducted by people who consider themselves both researchers and practitioners [39]. We also acknowledge that many of the communities targeted by AI interventions in the Global South, like community health workers, farmers, and educators, are practitioners themselves [80], and hope that our classification for this paper provides clarity on the choice of phrasing. Our inclusion criteria required participants to be situated in or conduct research focusing on the Global South (Africa, South/east Asia, Latin America, etc.), be at least 18 years of age, and speak English. We recruited participants through social media (Twitter and LinkedIn), email lists, and directly emailing shortlisted researchers. Interested participants were required to fill out a recruitment form detailing their demographic information (name, occupation, affiliated organization) if their work involved responsible AI or end-user engagement, the domain and regional affiliation of their research, and to confirm their interest in participating. To ensure that participants met the inclusion criteria for our study, we leveraged purposive sampling [55] to aid our selection methods for participants to invite for an interview.

After we invited participants for an interview, they were asked to fill out a pre-interview form to simplify the interviewing process. This form asked for more detailed demographic information (name, gender, age, and home country), the type of institution they are based in (industry, academia, government, nonprofit, NGO, etc.), their occupation, and domain of work. The form also asked participants to detail the number of projects where they have engaged with end users, how long they have been working with AI, how many years they have been working on AI4SG topics, what kinds of populations their research targets, and what kinds of machine learning (ML) methods are used in their work.

Participant Demographics In total, we interviewed 12 AI4SG practitioners (Gender: Female (4), Male (8); Age Range: 18-54). They

Table 1: Participant Demographics

ID	Gender	Age	Home Country	Institution	Domain
P01	Female	25-34	United States	Academia	Environment & Sustainability, Healthcare
P02	Female	45-54	Colombia	Government	Government & Policy
P03	Female	25-34	United States	Academia	Government & Policy
P04	Male	25-34	Brazil	Industry, Academia	Agriculture, Government & Policy, Strategy
P05	Male	35-44	Nigeria	Academia	Agriculture, Environment & Sustainability, Infrastructure
P06	Male	35-44	Botswana	Non-governmental organization	Agriculture, Healthcare
P07	Male	18-24	Kenya	Industry, Academia	Environment & Sustainability, Government & Policy
P08	Male	25-34	Nigeria	Academia	Education, Healthcare
P09	Male	25-34	Uganda	Academia, Nonprofit	Agriculture, Education, Healthcare, Languages
P10	Male	18-24	Uganda	Academia	Agriculture, Environment & Sustainability
P11	Male	25-34	Uganda	Industry, Academia	Education, Government & Policy
P12	Female	25-34	United States	Academia	Agriculture, Finance, Social Media, Government & Policy

also worked across a range of domain areas and countries, including Nigeria, India, Uganda, Brazil, Togo, Bangladesh, Afghanistan, Haiti, and Botswana. The categories that we designated for the occupations of our participants were Administration, Designer, Engineer, Management, Programmer, and Researcher. Most of our participants (10/12) are situated in academia, with 9/12 participants primarily identifying as researchers, two as engineers, and one as a programmer. All participants have experience working with AI and engaging end users situated in the Global South. Our participants have a range of experience working with AI (Range: 2-10 years; Average; 5.4), varied experience working with AI4SG topics (Range: 1-7 years; Average: 4.1 years), and also engaged with end users through numerous projects (Range: 1-20; Average: 4.5; Median 2). We provide detailed demographic information in Table 1.

ML Usage and AI4SG Domains Our participants used a variety of ML paradigms within their work (Supervised learning: n=12; Semi-supervised learning: n=3; Unsupervised learning: n=5). Specific ML techniques used by the AI4SG practitioners we interviewed included deep learning (n=9), computer vision (n=7), natural language processing (n=7), reinforcement learning (n=3), and automated speech recognition (n=2). To understand what domains of AI4SG our participants worked in, we created eight categories to classify their work: Agriculture, Education, Environment & Sustainability, Finance, Healthcare, Infrastructure, Social Media, and Government & Policy. These categories were used in the pre-interview study distributed to shortlisted participants. When we gained more perspective about the work of our participants after interviewing them, Languages was added as a separate domain.

3.2 Interview Procedure

We conducted semi-structured interviews over Zoom. Each interview involved one or two authors and was solely led by one author. Before the interview started, the author leading the interview requested informed consent, familiarized participants with the study objective, and mentioned the voluntary nature of the research study. We then checked if the participants had completed the pre-study interview survey form. If not, we asked the questions from the form in the interview. We then asked for permission to record and indicated the use of automated captioning.

Our interviews were split into two parts: (1) understanding the social impact problems addressed by practitioners, the types of end users impacted in their work, methodologies to engage stakeholders, and challenges deploying AI4SG projects, and (2) understanding how interviewees use XAI in their work, factors that impeded them from using XAI, and how they perceive XAI. Our research questions are available in Appendix B. Before moving to the second part of the interview, we defined XAI as *"methods or techniques that help users understand outputs from AI models or explain model reasoning for a single prediction or set of predictions."* While XAI has traditionally focused on technical methods, our definition of XAI and related questions accounted for nontechnical aspects of explainability, such as explaining model outputs to end users in lay terms. We also introduced the phrase "understandability" to account for such nuances in explaining AI systems. For example, we specifically asked our participants if they used other approaches (aside from XAI) to ensure that AI tools are "understandable" to their end users.

3.3 Analysis

We collected 7.8 hours of audio recordings from our interviews (Range: 21-62 minutes; Average: 39 minutes). After transcribing the interviews, we used inductive thematic analysis [17] to produce key themes from our interview data by repeatedly examining and comparing our qualitative analysis between reviewers. Both authors led the qualitative coding process. We started this process by coding two interviews, each separately, and convened to reconcile our codes by merging similar themes and constructing a codebook. We continued to code the transcripts individually and met throughout this process to continue reconciling codes, resulting in a final, stabilized codebook. This codebook was then used to code the rest of the interviews, which the authors evenly split. We met regularly to examine our progress and further iteratively refine our codes by discussing additions, scrutinizing ambiguities, and reconciling differences. After coding all of the transcripts through multiple passes, we ended up with 226 codes grouped into seven themes: XAI Usage and Methods, XAI Understanding, Challenges, Demographics, Problem Targeting, User Engagement, and Deployment. For example, the "End User Engagement" theme categorized how our interviewees engage with end users and collaborators, whether they use any frameworks or methodologies to guide this engagement, and how they understand the needs of their users when building AI4SG technologies. Examples of codes from this theme include: "End User Engagement: interviews", "End User Engagement: surveys", and "End User Needs: participatory design".

Limitations. This project is subject to several potential limitations. The primary constraint of this project lies in our small participant sample size, highlighting the difficulty of recruiting participants that met our inclusion criteria. Despite this, the AI4SG practitioners we interviewed are from diverse geographic locations, work on a broad spectrum of topics, and work with a wide range of end users (as shown in Table 1 and Appendix C). With this in mind, the perspectives represented in this work might not generalize across the Global South or to Western contexts. However, similar nuances exist in general AI development, to which our findings could prove useful. From our interviews, we provide recommendations that could generally apply to the work of AI practitioners, regardless of where their work is situated. Our study primarily focused on the perspectives of AI practitioners, but considering the perspectives of other key stakeholders, including the direct and indirect end users identified in our findings, is an essential area for future work.

4 Findings

In Section 4.1, we found that end users, different stakeholders' perspectives, and AI4SG practitioners' personal experiences influence the problems they choose to work on. Section 4.2 explores the end user communities our participants engage with, highlighting the importance of regular communication with end users and emphasizing the challenge of communicating across different expertise and levels of data and digital literacy. Section 4.3 explores how practitioners perceive XAI usefulness in their work and end user engagement, finding that all of them agreed that XAI would be helpful for themselves. However, some of our participants believed XAI would **not** be helpful for end users, given their challenges explaining model outputs to end users with relatively low levels of data and digital literacy in Global South contexts. In Section 4.4, we find that AI4SG practitioners encounter many challenges in incorporating XAI techniques into their existing workflows, including 1) a lack of expertise in XAI, limiting their ability to use these methods effectively and accurately, 2) the lack of tools allowing them to easily incorporate these techniques, and 3) XAI being considered as a lower priority within their projects.

4.1 Problem domains and contexts in AI4SG projects

Our participants worked on a diverse range of projects in Government & Policy, Agriculture, and Environment & Sustainability that included humanitarian aid targeting programs, distributing social services to address homelessness, precision farming in Nigeria, cassava disease detection in Uganda, early warning prediction systems for poultry disease in Tanzania, and conservation systems to predict poaching in Southeast Asia and East Africa. Within healthcare, education, and languages, projects included detecting lung disease from mining, automated translation of medical jargon into local Ugandan languages, maternal health in India, using AI to identify issues in student success and improve student success rates, and addressing gender bias in the machine translation of Ugandan languages.

Our participants described several considerations for selecting projects. Some were inspired by their personal experiences of being impacted by problems (P05, P07, P08, P09, P10). For example, in P10's case:

> *"On the occupational health hazard issue, I'm working in an area where we do a lot of sun blasting, so I accidentally sometimes consume a lot of dust, which is not so good. ... so I went on to find ways how can we solve this problem."* (P10)

Some decided to work on the projects after being contacted by funding organizations with specific directives (P09). Others were motivated by pressing global issues (P06, P08, P11) and collaborations with stakeholders (P01, P03, P12), or were directly commissioned by clients (P04). P01 states their approach to problem identification by being engaged with partner organizations who are their project stakeholders:

> *"The way that we identify the solution approaches is by first engaging in long-term conversations with these partner organizations. We try to understand what data is available and what actions they are currently taking."* (P01)

Throughout the interviews, we found that most of our participants (8/12) engaged with stakeholders or end users to help identify what AI4SG topics they wanted to address. This falls in line with practices in HCI and ICTD [24] that advocate for stakeholder engagement when designing and developing technical solutions in low-resource domains.

Overall, we find that AI4SG practitioners engage in a diverse range of methods to identify problems to address and select AI methods to implement their AI4SG solutions. These methods also result in engagement with a diverse range of stakeholders and end users, which we detail next.

4.2 Needs and challenges of end-user engagement and system deployment in practice

Given that most AI4SG projects aim to tackle real-world problems, AI4SG practitioners generally need to engage with the intended end users of their products as part of a development pipeline, primarily to understand and address end users' needs and leverage their respective domain expertise. All of our participants expressed the importance of designing solutions with pre-determined end users in mind, and most of them directly interacted with end users to various degrees throughout their projects, whether their works had been deployed in the real world or not. In this section, we provide insights into *who* and *how* our interviewees engaged with their end users, as well as the challenges that arose in their engagement.

Direct and Indirect End Users. The AI practitioners we interviewed engaged with a wide variety of end users. Some of our interviewees pointed out that there are generally two types of end users - direct and indirect. The *direct end users* typically make decisions assisted by the AI systems and interface directly with the AI practitioners throughout the development pipeline. On the other hand, *indirect end users* are impacted by the decisions made further down the pipeline but are generally not involved in the design pipeline. For example, in the case of humanitarian aid applications, the direct end users are policymakers and government aid program decision-makers who use an AI system to determine what citizens to prioritize when distributing poverty aid in resource-constrained settings. In contrast, the indirect end users are the eligible citizens who receive or do not receive the aid resources. Several participants noted the challenges of involving indirect end users, even though they acknowledged the importance of it. One common challenge, as P03 pointed out, has to do with "low levels of education and digital literacy, importantly related to explainability" in the regions where they work.

The roles of end users depend mainly on the specific projects. For example, for our participants who work in Government & Policy, direct end users include policymakers (P03), and indirect users include social media users (P11) and humanitarian aid recipients (P03). In Agriculture, our interviewees worked with banks as direct users (P04), extension agents as direct users (P05), and farmers as both direct (P12) and indirect users (P04 and P05). Some examples of direct end users in Environment & Sustainability include park rangers (P01). In Healthcare, construction workers (P10) and medical practitioners (P01, P08) are mentioned as direct end users. Direct end users in other domains include nonprofits (P12), translators (P07), media companies (P07), and military personnel (P05).

A "hodgepodge" of engagement methods. While most AI practitioners didn't mention specific frameworks, they stressed the importance of engaging with end users throughout the design pipeline. As P12 pointed out:

> *"I know generally that stakeholder involvement is super important. So having these regular meetings with them is kind of the main way that I've been able to engage with the end users... But I don't have a guide. I kinda just try my best."* (P12)

P01 agreed, stating, "it's this kind of hodgepodge of things that tries to be sensitive to their severe time constraints." This "hodgepodge of things" involves a variety of practical strategies, specifically through regular online meetings (P01, P04, P05, P08, P12), in-depth in-person interviews (P03, P07, P10), informal conversations (P04, P11), questionnaires or surveys (P05, P07, P10, P11), frequent email exchanges (P03), delivery workshops (P04, P08, P12), and written reports (P04). Besides structured engagement, 2/12 participants also mentioned other communication channels like WhatsApp (P04, P09), where AI developers and technicians are available for more ad-hoc inquiries from end users. Aligned with Varanasi and Goyal [95], we found that practitioners tend to adopt a mix of approaches tailored to their specific context without clear and usable guidance on how to apply human-centered design frameworks. While the flexibility might allow practitioners to adapt general guidelines to local contexts, it might also leave practitioners feeling unsupported and needing to take on additional self-guided initiatives on top of existing workloads and tight deadlines [25, 95, 104].

AI4SG system deployment and hand-off challenges. To better understand user engagement in the context of product deployment, we investigated whether the systems our participants developed were meant for deployment on the ground and whether they were intended to be handed off entirely to end users or collaborators. More than half of our interviewees (8/12) described some level of deployment of their work at the time of interviews, while 2/12 participants planned to deploy their solutions in the near future. The rest mentioned that their projects were only meant to be proofs of concepts and academic publications, not for real-world deployment. We find that the deployment of AI tools in real-world settings is often hindered by resource constraints, including limited funding (P05, P12), inconsistent electricity availability (P09), lack of access to computing resources such as GPUs (P05), inadequate availability of data (P07), and a deficiency of in-house technical staff (P01).

A common challenge was accommodating users with varying levels of data and AI literacy (P03, P04, P07) and communicating with end users who were not technical experts themselves (P12). P12 points out the labor involved in such communication, noting that *"there's a lot of translation work [between different fields and expertise] that needs to happen"*. Given the expressed preference for many of our participants to explain their algorithms to various stakeholders and end users, expanding the development of "sociotechnical XAI" [31], XAI that incorporates social and intercultural aspects [52, 72] of explaining predictions from AI systems to a diverse range of users, could enable AI4SG practitioners to more easily communicate model reasoning. Other difficulties our participants faced when deploying their systems included coordination and consensus among multiple technical teams (P11), a lack of project scalability (P12), unstable political environments (P05), the fast turnaround of AI4SG projects and publication timelines (P01, P04, P12), and AI systems impacting end users' sense of agency (P03). Overall, many of the challenges discussed in this section are unique to practitioners working in the Global South, where AI literacy and optimism differ significantly from users in Western settings [75, 83, 86].

Out of the 8 participants that deployed their work, only three handed off their systems to end users, while the rest of the interviewees remained active maintainers and kept in correspondence about the systems. Even though not all systems were meant to be handed off completely to the end users, we asked every participant to describe what their end users might need to operate AI tools effectively to understand whether model comprehension is an important consideration in the deployment process. We learned that a good understanding of the *data*, not *models*, was commonly mentioned as a key thing for end users to know because, as P07 stated, *"if you give [the algorithm] wrong data to train, then it will give you wrong results."*. P03 agreed by emphasizing that AI practitioners should communicate to end users *"first definitely what data is being used."* Participants also noted that it is often unnecessary for the end users to know the inner workings of the models (P04, P08) as long as the procedures and results make sense at a high level. Driven by our specific findings, future efforts toward designing sociotechnical explainability methods should prioritize approaches that effectively explain various facets of AI systems to non-technical audiences, including data, models, and predictions.

Our findings on how AI4SG practitioners currently engage with end users reveal the importance of regular engagement and communication with end users and the challenges that arise. In the next section, we will discuss how practitioners perceive the usefulness of XAI in their work and in engaging with end users.

4.3 Varying perceptions of XAI effectiveness and usefulness

Effectiveness of XAI is context-dependent. Our participants had varying experiences using technical and human-centered XAI approaches. P03 and P10 had positive experiences with the effectiveness of the technical XAI methods they employed in their work. P03 mentioned that XAI was useful in helping them understand how their AI models work and what relationships are being learned from the training data. P03 also mentioned that XAI helped their end users, in this case policymakers, understand more generally how the algorithms worked. P10 mentions how XAI is effective in helping them understand their models and how datasets impact their respective efficacy. P07 mentions that technical XAI methods such as LIME and SHAP were ineffective, highlighting that in their work understanding gender bias in Ugandan languages, these methods could not fully capture the nuances of these languages to validate the correctness of the AI-produced translations, noting that *"it misses a lot of things that are not sufficient enough to show me that this is right in translating gender-neutral sentences into a gender-specific sentence"* and often *"bends towards a specific [gender] stereotype"*. Additionally, our findings from this question illustrated that the effectiveness of XAI is highly dependent on the needs of end users and the objectives of practitioners. P03 describes what "effective" means in their context and how explanations help ensure that stakeholders are confident in the outputs from AI4SG tools:

> *"...they're pretty effective at doing what we want them to do. We want people to have a general sense of how these algorithms are working. We don't expect them to be able to understand how individual eligibility decisions are being determined, but we do want them to have a general sense of how the algorithm is working, and we want them to feel comfortable with it."* (P03)

Usefulness of XAI is dependent on user needs and abilities. While the usefulness and impact of XAI in practice remains an open question to study, all of our participants who actively used XAI methods agreed that XAI is helpful for their work. Technical XAI methods, such as feature importance and marginal graphs, were primarily beneficial for practitioners to diagnose the accuracy of the models and correct the models as needed, as 6 of our interviewees pointed out. They emphasized the importance of understanding the models to ensure their confidence in the accuracy and scientific rigor of their solutions (P01, P04, P05, P06, P07, P11). *You can't build what you can't understand,"* as P09 stated. P08 and P09 also highlighted that being able to explain model outputs is a *"responsibility"* (P08) of the AI developers because they are the *"first people of contact"* (P09) and they need to be able to answer how the models work. Adding to our findings on sociotechnical XAI, participants such as P12 verbally translated technical XAI methods into more user-accessible explanations to aid communication with stakeholders and include nontechnical experts in the decision-making process.

When it comes to their end users, however, some of our interviewees expressed hesitation about the usefulness of XAI. Specifically, P02 and P04 illustrated the difficulty of explaining model outputs to their end users (policymakers and bankers) with relatively low levels of AI literacy, which was also noted in previous work examining AI4SG projects in the Global South [73, 75]. P02 also notes the importance of deploying AI tools responsibly, underlining a need for more accessible explainability methods geared towards nontechnical end users:

> *"We need to guarantee that any system is respecting [end users'] rights and it is not causing any damage. But on the practical side, it is difficult for people even with engineering backgrounds to understand the models, so I have doubts about if an end user should be able to use those tools."* (P02)

However, our participants stressed the importance of users being able to understand how the algorithm functions due to such comprehension, helping end users gain awareness of how the algorithms make recommendations. P05 stressed that *"[XAI] is important because if [end users] don't understand it, then they won't use the solution correctly."* P12 also mentions the importance of XAI in correcting model errors:

> *"It helps end users to potentially dispute the model if they disagree with it and try to incorporate their own expertise and intuition more."* (P12)

When describing the usefulness of XAI for end users, several participants also emphasized its importance in ensuring other responsible AI concepts, such as transparency (P08, P11) and fairness (P07, P08), especially in high-stake decision-making applications like determining education resources, hiring eligibility, and humanitarian aid distribution.

Our interviews demonstrate the nuance of XAI methods, whose utility and effectiveness vary for different groups of users. While all of our participants agreed that explainable AI would be helpful for AI practitioners like themselves, not all believed it would benefit their intended end users. Despite hesitation around the effectiveness for end users, AI4SG practitioners emphasized the importance of such users being able to comprehend outputs from the models they develop. This motivates a careful consideration of designing XAI methods with different end user groups (e.g., practitioners vs. policymakers vs. humanitarian aid recipients) and needs in mind. We next explore how our participants use XAI methods in practice and the challenges that arise.

4.4 Implementing XAI approaches in practice

AI4SG practitioners' varying comprehension of algorithm function. To gauge how well AI4SG practitioners understand their systems with or without using XAI, we began by asking if our participants had a good sense of how their algorithms make predictions or provide results. 9/12 of our participants responded that they understand how the algorithms they build for their AI4SG projects work, with participants (P03, P07) specifically mentioning that technical XAI methods help them better understand how their algorithms function. Three participants responded that they didn't understand exactly how their algorithms function. For example, P11 notes their focus on efficacy over model comprehension:

> *"I'll just look at the documentation of the algorithm and then ensure that it works the way it's supposed to work. But about the background of how the algorithm works, I don't know."* (P11)

While our interviews focused on explainability, we also found that some of our participants were involved in work that aimed to make their algorithmic systems more interpretable. For the sake of the interviews, we considered "model explainability" to be more focused on explaining model reasoning behind a single prediction or set of predictions and "model interpretability" to be how well humans understand how the model operates as a whole. Multiple participants mentioned using simpler models with fewer features to increase the interpretability of their ML models. For example, P12 states using methods such as *"a linear program where you can see your objective"* and P03 mentions using *"a more parsimonious model like a linear lasso regression or something that uses only 20 features which works pretty well."*. In addition to using simpler models, P03 often stated throughout their interview that accuracy was the metric they optimized for the most due to research objectives and stakeholder constraints. While the AI research community often advocates for more interpretable models, such models are shown to be less accurate [36], a metric often seen as more important to AI researchers [15].

While all of our participants agreed that XAI would be helpful for their work (Section 4.3), only 3/12 of our participants used XAI in practice. In some cases, participants responded that they used XAI methods in their research but could not elaborate on the specific methods they used or backtracked on their answers. To avoid leading participants in follow-up questions about their particular usage of XAI, we reiterated our definition of XAI (from Section 3) but did not provide specific examples of XAI tools or methods. If participants were still unable to name XAI methods, we considered it as an indication that participants did not use XAI. For those that used technical XAI in their AI4SG work, such methods included LIME (P07), SHAP (P03, P07), decision trees (P10), feature importance (P03), and heatmaps (P10).

Nontechnical notions of XAI. We also found interesting perspectives on explainability that expanded beyond the typical notion of XAI as solely a technical concept. Our participants commonly mentioned "explainability" as also being focused on providing layman's instructions for end users to either understand decisions produced from AI tools (P03) or instructions on how to operate AI tools (P07). For example, P07 mentions:

> *"Sometimes, we provide a set of instructions during the deployment that is available in web or mobile format. We made simple instructions to ensure the user understands what's happening."* (P07)

Given that our findings highlight multiple dimensions of explainability outside of technical methods, we find that further development of "sociotechnical XAI" [31], XAI that incorporates social and intercultural aspects [72] of explaining predictions from AI systems to a diverse range of users, will be necessary to ensure that low-literate and novice AI communities understand the implications of outcomes from algorithmic decision-making. By developing and incorporating sociotechnical XAI, these methods can democratize access to relevant information, empowering users with low levels of traditional and AI/digital literacy to dispute the decisions made by AI systems and seek recourse for algorithmic harms. Additionally, our findings in Section 4.2 show that future efforts toward designing sociotechnical explainability methods will have to prioritize approaches that effectively explain various facets of AI systems to non-technical audiences, including data, models, and predictions.

Challenges of implementing XAI in AI4SG work. Our participants who implemented XAI methods in their work faced various challenges. These challenges included method selection (P07), ease of using XAI (P10), accuracy tradeoffs (P03), and users' lack of digital and traditional literacy (P07). P10 found issues in the usability of XAI, specifically noting challenges with interoperability when trying to implement XAI on different platforms and when using different programming packages. While P03 didn't report any specific issues that occurred when using XAI in practice, they stated that the practical benefits of using XAI in their work are more centered on conveying results in publications, and due to their high research load, implementing XAI is of low priority unless approaches are easy to implement.

For our participants who indicated that they did not use XAI in their research, we skipped over questions inquiring about the specific XAI methods they used and the challenges associated with integrating and using XAI in the field. However, we asked what expressly prohibited them from pursuing XAI and their interest in using XAI in the future. We found a range of factors impeding the use of XAI in AI4SG work, including the effort needed to integrate these methods (P11), project constraints (P02), a lack of scalability of XAI methods (P01), a limited need to use XAI methods (P09, P12), the extra domain knowledge needed to interpret and use XAI (P02, P06, P11), and computing constraints (P01, P05). More specifically, computing constraints impede the successful implementation of AI4SG projects in the field and impact how XAI can be incorporated, given the added computational complexity XAI often contributes to ML models [21]. Our findings, combined with insights from prior human-centered XAI research, show that much work is still needed to make explainability useful in practice for practitioners and end users alike.

5 Discussion

Within the past few years, interest in using AI to solve pressing social problems has substantially increased with researchers, non-profits, governments, and industry organizations focusing on developing solutions for challenges in agriculture [58, 64, 89, 96], healthcare [33, 35, 105], education [28, 44, 65], and poverty alleviation [5, 48, 54]. Despite burgeoning interest in using AI to solve pressing social problems, little work has focused on understanding practitioner motivations behind developing AI4SG tools and how practitioners ensure that stakeholders and populations affected by AI usage understand outputs from these systems. Our work provides the first analysis of AI4SG in this manner, outlining existing challenges in how XAI is approached within social impact-oriented AI work in the Global South and highlighting opportunities for a paradigm shift. In this section, we reflect on the shortcomings of using XAI in practice and a future agenda for explainability research.

5.1 The Shortcomings of XAI in Practice

A small amount of research has focused on evaluating how XAI is used in practice, showing that the benefits of these methods do not serve the needs of end users [13, 45]. Our study shows similar findings, detailing the challenges that affect how AI4SG practitioners choose to use XAI when developing their models and communicating model outputs to users. Many participants who used XAI methods in practice stated that their primary motivation for using XAI was to aid in publishing their papers rather than improving user comprehension. While it is important that XAI is used to convey the results of AI systems and aid reviewers in understanding the validity of such systems, it is also important for end users, especially those impacted by these systems, to understand the outputs from AI systems. We find that current approaches to using XAI in AI4SG create a severe incentive misalignment, devaluing the needs of users in favor of publication practices in machine learning that often prioritize model performance over user interpretability [15]. Facing the pressure to publish research papers, AI4SG practitioners–especially those situated in academia–might not have the individual agency to prioritize user-centered XAI and other responsible AI approaches. Additionally, given the added complexity and additional computing resources that XAI methods consume [43], employing these methods may not be feasible for researchers in low-resource settings in the Global South and those who lack access to computing clusters.

Given the issues that many of our participants detailed regarding their ability or motivation to incorporate XAI, there is a case for building inherently interpretable models that produce algorithmic decisions understandable by both practitioners and end users rather than relying on post-hoc explanations from "black-box" models. Researchers have strongly advocated for interpretable models to aid in using ML for high-stakes domains such as criminal justice and healthcare due to the ability of explanations to be unreliable and mislead users [85]. Moving away from "black-box" models is also necessary for resisting the exclusionary nature and opaque design of AI systems, which are often rooted in Western contexts and embed the values of large tech corporations. To improve the feasibility of XAI in practice, we find that existing approaches and methodologies to explaining predictions produced from AI/ML models have to be reenvisioned. Considering that wide-scale adoption of inherently interpretable models is still emerging, AI4SG practitioners employing XAI in their work will also have to mitigate existing tensions around the explainability-accuracy tradeoff [79]. This will be needed to understand in which cases improving the accuracy of an AI system is more important than focusing primarily on user interpretability. With this in mind, future research will be needed to understand how AI and ML researchers can develop user-interpretable explainability approaches, the implications surrounding the feasibility of using such methods in low-resource domains, and how these methods should be introduced to low-literate and novice AI users.

5.2 A Future Agenda for Explainability Research

Systematic shifts and publication incentives towards XAI implementation, as seen in efforts by premier ML conferences to have reviewers check whether papers "appropriately reflect" explainability as a necessary characteristic of Trustworthy AI [20], may help encourage XAI as a broader part of responsible AI. However, given the existing challenges with XAI and its lack of usefulness for non-technical end users, such efforts may impose Western notions of "responsible AI" on AI4SG practitioners whose work is primarily situated in the Global South, inadvertently impacting their work and potentially harming users. Thus, ample opportunities exist to re-envision how to bring accountable and responsible AI innovation to AI4SG. A recent collaboration between Anthropic, an AI safety and research company, and the Collective Intelligence Project to draft a constitution based on public input from 1000 Americans to guide the development of Anthropic's LLM [8] could serve as a model to leverage democratic processes in expanding notions of explainability. To illustrate this in practice, AI4SG practitioners could engage community members in a similar democratic process, inquiring what outcomes from AI tools would be "fair" for them and how such decision-making should be explained. Practitioners would initiate this process by explaining their project and outlining goals, intended outcomes, and target end users. Practitioners would then engage in a democratic collective input process where end users identify aspects of the project they agree and disagree with, eventually reaching a consensus on a final set of project goals. These goals could then be integrated into a "constitution" to guide the development and evaluation of AI4SG projects while simultaneously informing novel explanation mechanisms for end users.

While one of the aims of our study was to understand how AI4SG practitioners incorporate model explainability into their work, some of our participants mentioned the concept of explainability being based on how to relay decisions to users affected by algorithmic decision-making. We understand that notions of explainability are quite complex, and techniques that help end users understand outputs from AI systems tend to be distinct from those that aid AI developers in building explainable models. With this in mind, our findings also demonstrate a need for more research within XAI to understand the needs of both practitioners and end users when using explainability in practice. As users become increasingly exposed to AI through the usage of large language models (LLMs) and chatbots powered by these technologies, it will become essential for new research directions to focus on examining the distinct needs of these users, especially those who are low-literate, have less exposure to AI, and are situated in non-Western contexts. Such research can then inform the development of novel user-centric explainability methods that encapsulate technical and sociotechnical aspects of explanations while being effective for a broad spectrum of users and use cases.

To move towards more equitable practices in XAI research and development, we present the following recommendations, summarized from our findings and discussion:

- Promote the use of inherently interpretable models that produce outputs understandable by practitioners and end users.
- Invest in mixed-methods research, leveraging participatory design methods, to understand practitioner and end user needs when using explainability in practice.
- Understand sociotechnical nuances behind explainability to prioritize approaches that effectively explain various facets

of AI systems (data, models, and predictions) to non-technical stakeholders.

6 Conclusion

This paper details an interview-based study conducted with 12 AI4SG practitioners working with a range of end users across a variety of domain contexts in the Global South. Our interviews investigate how these practitioners identify problems to solve and select AI methodologies to address these problems. We also reveal novel findings demonstrating practitioners' motivation behind using XAI and their perceptions regarding the utility and efficacy of these methods. Our findings provide an opportunity to reshape sociotechnical notions of explainability in AI development to recenter value alignment and ensure that the needs of low-literate, novice technology users are prioritized. The use of XAI in domains outside of AI4SG potentially suggests that our findings could be relevant to a broader set of AI practitioners. However, future studies will be required to understand how our work can generalize to other domains.

7 Ethical Considerations

IRB Review. Our institutional IRB office determined our work did not require IRB review or exemption. Despite this, we diligently conducted our project ethically and professionally in a way that respected the autonomy of all participants involved. Before signing up for the study, participants were given an information and consent form to review further study details and how their data would be used. Before each interview, participants were reminded of the confidentiality of their data, that their participation was voluntary, and that they could stop participation at any time. Participants were not compensated for their participation in this study.

Positionality. All authors are researchers based in the United States who have conducted fieldwork with underserved communities in low-resource regions within the Global South (India, Malawi, and China). Both authors identify as female. One author has 3+ years of experience studying healthcare workers in India and AI development in the Global South more widely. Collectively, both authors have 6+ years of experience as AI4SG practitioners. One of the authors has experience working with an international development agency focusing on issues of AI governance within the African continent. One author also has experience working on data for development and SDGs in an international development organization. As AI4SG practitioners whose work centers on marginalized communities, we believe in elevating the voices of local communities and actively including them in AI development. We approach our research through an equity-driven [38] and emancipatory action [10, 49] mindset, where we aim to identify the opportunities and challenges of integrating XAI in AI4SG work while underscoring the needs of local populations who will interact with these technologies.

Adverse Impacts. Our work, along with recent studies in conservation [87, 100] and humanitarian aid distribution [99], highlights some of the unanticipated harms of AI4SG projects aiming to serve vulnerable populations in low-resource domains. While we advocate for further development of "sociotechnical XAI" methods to meet the explanation needs of low literate and novice AI users in the Global South, employing these methods could potentially be used as justification to continue using AI4SG tools that inflict harm (intentionally or unintentionally) on subjects of AI systems.

Given that large-scale deployments of AI-enabled tools are still emerging within the Global South, there remain opportunities to leverage explainability in ways that can attenuate harms from AI systems. To mitigate such harms, thoughtful considerations of cultural and economic contexts, along with the end users that will be directly and indirectly impacted by AI systems, are necessary throughout the design, development, and implementation process [16, 88]. More importantly, researchers and practitioners must consider cases where AI may exacerbate bias (e.g., using facial recognition or other AI-enabled biometric tools to maintain refugee settlements and handle aid distribution) and refrain from using AI as a "panacea" for social issues. In these cases, transparency measures such as XAI should not be used to validate AI outputs or rationalize the continued use of AI tools for sensitive use cases.

Acknowledgments

We are thankful to all of the participants who contributed valuable insights to our study.

References

[1] Rediet Abebe, Kehinde Aruleba, Abeba Birhane, Sara Kingsley, George Obaido, Sekou L Remy, and Swathi Sadagopan. 2021. Narratives and counternarratives on data sharing in Africa. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 329–341.

[2] Rediet Abebe, Solon Barocas, Jon Kleinberg, Karen Levy, Manish Raghavan, and David G Robinson. 2020. Roles for computing in social change. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 252–260.

[3] Accenture. 2023. AI ethics & governance. Retrieved November 16, 2023 from https://www.accenture.com/us-en/services/applied-intelligence/ai-ethics-governance

[4] Google AI. 2023. Responsible AI Practices. Retrieved November 16, 2023 from https://ai.google/responsibility/responsible-ai-practices/

[5] Emily Aiken, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E Blumenstock. 2022. Machine learning and phone data can improve targeting of humanitarian aid. *Nature* 603, 7903 (2022), 864–870.

[6] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[7] Josh Andres, Christine T Wolf, Sergio Cabrero Barros, Erick Oduor, Rahul Nair, Alexander Kjærum, Anders Bech Tharsgaard, and Bo Schwartz Madsen. 2020. Scenario-based XAI for humanitarian aid forecasting. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[8] Anthropic. 2023. Collective Constitutional AI: Aligning a Language Model with Public Input. Retrieved November 27, 2023 from https://www.anthropic.com/index/collective-constitutional-ai-aligning-a-language-model-with-public-input

[9] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[10] Jeffrey Bardzell and Shaowen Bardzell. 2015. What is Humanistic HCI? In *Humanistic HCI*. Springer, 13–32.

[11] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M Vardoulakis. 2020. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.

[12] R. K. E. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilović, S. Nagar, K. Natesan Ramamurthy, J. Richards, D. Saha, P. Sattigeri, M. Singh, K. R. Varshney, and Y. Zhang. 2019. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development* 63, 4/5 (2019), 4:1–4:15. https:

//doi.org/10.1147/JRD.2019.2942287

[13] Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José MF Moura, and Peter Eckersley. 2020. Explainable machine learning in deployment. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 648–657.

[14] Abeba Birhane. 2020. Algorithmic colonization of Africa. *SCRIPTed* 17 (2020), 389.

[15] Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022. The values encoded in machine learning research. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 173–184.

[16] Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A Killian. 2021. Envisioning communities: a participatory approach towards AI for social good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 425–436.

[17] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[18] Jenna Burrell and Kentaro Toyama. 2009. What constitutes good ICTD research? *Information Technologies & International Development* 5, 3 (2009), pp–82.

[19] NeurIPS Communication Chairs. 2021. Announcing the NeurIPS 2021 Datasets and Benchmarks Track. Retrieved June 12, 2023 from https://blog.neurips.cc/2021/04/07/announcing-the-neurips-2021-datasets-and-benchmarks-track/

[20] NeurIPS Program Chairs. 2023. 2023 Ethics Guidelines for Reviewers. Retrieved August 12, 2023 from https://nips.cc/Conferences/2023/EthicsGuidelinesForReviewers

[21] Yu-Neng Chuang, Guanchu Wang, Fan Yang, Zirui Liu, Xuanting Cai, Mengnan Du, and Xia Hu. 2023. Efficient xai techniques: A taxonomic survey. *arXiv preprint arXiv:2302.03225* (2023).

[22] Kate Conger, Richard Fausset, and Serge F Kovaleski. 2019. San Francisco bans facial recognition technology. Retrieved June 12, 2023 from https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html

[23] Andy Dearden. 2012. See no evil? Ethics in an interventionist ICTD. In *Proceedings of the fifth international conference on information and communication technologies and development*. 46–55.

[24] Nicola Dell and Neha Kumar. 2016. The ins and outs of HCI for development. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2220–2232.

[25] Wesley Hanwen Deng, Manish Nagireddy, Michelle Seng Ah Lee, Jatinder Singh, Zhiwei Steven Wu, Kenneth Holstein, and Haiyi Zhu. 2022. Exploring How Machine Learning Practitioners (Try To) Use Fairness Toolkits *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 473–484. https://doi.org/10.1145/3531146.3533113

[26] Susannah Kate Devitt, Jason Scholz, Timo Schless, and Larry Lewis. 2021. Developing a Trusted Human-AI Network for Humanitarian Benefit. *arXiv preprint arXiv:2112.11191* (2021).

[27] Shipi Dhanorkar, Christine T. Wolf, Kun Qian, Anbang Xu, Lucian Popa, and Yunyao Li. 2021. Who Needs to Know What, When?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1591–1602. https://doi.org/10.1145/3461778.3462131

[28] Fernão Reges dos Santos and Pollyana Notargiacomo. 2018. Intelligent educational assistant based on multiagent system and context-aware computing. *International Journal of Advanced Computer Science and Applications* 9, 4 (2018).

[29] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI Systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 82, 19 pages. https://doi.org/10.1145/3411764.3445188

[30] Upol Ehsan and Mark O Riedl. 2020. Human-centered explainable ai: Towards a reflective sociotechnical approach. In *HCI International 2020-Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*. Springer, 449–466.

[31] Upol Ehsan, Koustuv Saha, Munmun De Choudhury, and Mark O Riedl. 2023. Charting the Sociotechnical Gap in Explainable AI: A Framework to Address the Gap in XAI. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–32.

[32] Upol Ehsan, Ranjit Singh, Jacob Metcalf, and Mark Riedl. 2022. The algorithmic imprint. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1305–1317.

[33] Safa Elkefi and Safa Bhar Layeb. 2022. Artificial Intelligence and Operations Research in a Middle Ground to Support Decision-Making in Healthcare Systems in Africa. *Africa Case Studies in Operations Research: A Closer Look into Applications and Algorithms* (2022), 51–69.

[34] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (2021), 86–92.

[35] Ryan G Gomes, Bellington Vwalika, Chace Lee, Angelica Willis, Marcin Sieniek, Joan T Price, Christina Chen, Margaret P Kasaro, James A Taylor, Elizabeth M Stringer, et al. 2022. AI system for fetal ultrasound in low-resource settings. *arXiv preprint arXiv:2203.10139* (2022).

[36] Patrick Hall. 2016. Predictive modeling: Striking a balance between accuracy and interpretability. Retrieved June 16, 2023 from https://www.oreilly.com/content/predictive-modeling-striking-a-balance-between-accuracy-and-interpretability

[37] Karen Hao. 2020. The two-year fight to stop Amazon from selling face recognition to the police. Retrieved June 12, 2023 from https://www.technologyreview.com/2020/06/12/1003482/amazon-stopped-selling-police-face-recognition-fight/

[38] Christina Harrington, Sheena Erete, and Anne Marie Piper. 2019. Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–25.

[39] Roger Harris and Arul Chib. 2012. Perspectives on ICTD Research and Practice. *The Singapore Internet Research Centre (SiRC) initiates and conducts research* (2012), 3.

[40] Rebecca Heilweil. 2020. Big tech companies back away from selling facial recognition to police. That's progress. Retrieved June 12, 2023 from https://www.vox.com/recode/2020/6/10/21287194/amazon-microsoft-ibm-facial-recognition-moratorium-police

[41] Kashmir Hill. 2022. Microsoft Plans to Eliminate Face Analysis Tools in Push for 'Responsible A.I.'. Retrieved June 12, 2023 from https://www.nytimes.com/2022/06/21/technology/microsoft-facial-recognition

[42] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3290605.3300830

[43] Xia Hu, Lingyang Chu, Jian Pei, Weiqing Liu, and Jiang Bian. 2021. Model complexity of deep learning: A survey. *Knowledge and Information Systems* 63 (2021), 2585–2619.

[44] Gwo-Jen Hwang, Han-Yu Sung, Shao-Chen Chang, and Xing-Ci Huang. 2020. A fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors. *Computers and Education: Artificial Intelligence* 1 (2020), 100003.

[45] Helen Jiang and Erwen Senge. 2021. On two XAI cultures: A case study of non-technical explanations in deployed AI system. *arXiv preprint arXiv:2112.01016* (2021).

[46] Shivaram Kalyanakrishnan, Rahul Alex Panicker, Sarayu Natarajan, and Shreya Rao. 2018. Opportunities and challenges for artificial intelligence in India. In *Proceedings of the 2018 AAAI/ACM conference on AI, Ethics, and Society*. 164–170.

[47] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376219

[48] Benjamin Kellenberger, John E Vargas-Muñoz, Devis Tuia, Rodrigo C Daudt, Konrad Schindler, Thao TT Whelan, Brenda Ayo, Ferda Ofli, and Muhammad Imran. 2021. Mapping Vulnerable Populations with AI. *arXiv preprint arXiv:2107.14123* (2021).

[49] Os Keyes, Josephine Hoy, and Margaret Drouhard. 2019. Human-computer insurrection: Notes on an anarchist HCI. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[50] Sunnie SY Kim, Nicole Meister, Vikram V Ramaswamy, Ruth Fong, and Olga Russakovsky. 2022. HIVE: Evaluating the human interpretability of visual explanations. In *European Conference on Computer Vision*. Springer, 280–298.

[51] Satyapriya Krishna, Tessa Han, Alex Gu, Javin Pombra, Shahin Jabbari, Steven Wu, and Himabindu Lakkaraju. 2022. The Disagreement Problem in Explainable Machine Learning: A Practitioner's Perspective. *arXiv preprint arXiv:2202.01602* (2022).

[52] Frens Kroeger, Bianca Slocombe, Isa Inuwa-Dutse, Baker Kagimu, Beate Grawemeyer, and Umang Bhatt. 2022. Social Explainability of AI: The Impact of Non-Technical Explanations on Trust. In *IJCAI 2022 Workshop on Explainable Artificial Intelligence (XAI)*.

[53] Himabindu Lakkaraju, Dylan Slack, Yuxin Chen, Chenhao Tan, and Sameer Singh. 2022. Rethinking Explainability as a Dialogue: A Practitioner's Perspective. *arXiv preprint arXiv:2202.01875* (2022).

[54] Chiara Ledesma, Oshean Lee Garonita, Lorenzo Jaime Flores, Isabelle Tingzon, and Danielle Dalisay. 2020. Interpretable poverty mapping using social media data, satellite images, and geospatial information. *arXiv preprint arXiv:2011.13563* (2020).

[55] Paul S Levy and Stanley Lemeshow. 2013. *Sampling of populations: methods and applications*. John Wiley & Sons.

[56] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020*

CHI Conference on Human Factors in Computing Systems. 1–15.

[57] Yuan Liu, Zhongwei Cheng, Jie Liu, Bourhan Yassin, Zhe Nan, and Jiebo Luo. 2019. AI for earth: rainforest conservation by acoustic surveillance. *arXiv preprint arXiv:1908.07517* (2019).

[58] Loyani Loyani and Dina Machuve. 2021. A Deep Learning-based Mobile Application for Segmenting Tuta Absoluta's Damage on Tomato Plants. (2021).

[59] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence* 2, 1 (2020), 56–67.

[60] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 4768–4777.

[61] Michael Madaio, Lisa Egede, Hariharan Subramonyam, Jennifer Wortman Vaughan, and Hanna Wallach. 2022. Assessing the Fairness of AI Systems: AI Practitioners' Processes, Challenges, and Needs for Support. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–26.

[62] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229.

[63] Jessica Morley, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and engineering ethics* 26, 4 (2020), 2141–2168.

[64] Latifa M Mrisho, Neema A Mbilinyi, Mathias Ndalahwa, Amanda M Ramcharan, Annalyse K Kehs, Peter C McCloskey, Harun Murithi, David P Hughes, and James P Legg. 2020. Accuracy of a smartphone-based object detection model, PlantVillage Nuru, in identifying the foliar symptoms of the viral diseases of cassava–CMD and CBSD. *Frontiers in plant science* 11 (2020), 590889.

[65] Saima Munawar, Saba Khalil Toor, Muhammad Aslam, and Muhammad Hamid. 2018. Move to smart learning environment: exploratory research of challenges in computer laboratory and design intelligent virtual laboratory for eLearning Technology. *EURASIA Journal of Mathematics, Science and Technology Education* 14, 5 (2018), 1645–1662.

[66] United Nations. 2016. The 17 Goals. Retrieved June 12, 2023 from https://sdgs.un.org/goals

[67] NIST. 2023. AI Risk Management Framework. Retrieved November 16, 2023 from https://www.nist.gov/itl/ai-risk-management-framework

[68] José Anchieta CC Nunes, Igor CS Cruz, André Nunes, and Hudson T Pinheiro. 2020. Speeding up coral reef conservation with AI-aided automated image analysis. *Nature Machine Intelligence* 2, 6 (2020), 292–292.

[69] ACLU of Illinois. 2022. In Big Win, Settlement Ensures Clearview AI Complies With Groundbreaking Illinois Biometric Privacy Law. Retrieved June 12, 2023 from https://www.aclu.org/press-releases/big-win-settlement-ensures-clearview-ai-complies-with-groundbreaking-illinois

[70] Chinasa T. Okolo. 2023. *AI Explainability in the Global South: Towards an Inclusive Praxis for Emerging Technology Users*. Ph. D. Dissertation. Cornell University.

[71] Chinasa T. Okolo. 2023. AI in the Global South: Opportunities and challenges towards more inclusive governance. (2023).

[72] Chinasa T. Okolo. 2023. Towards a Praxis for Intercultural Ethics in Explainable AI. *arXiv preprint arXiv:2304.11861* (2023).

[73] Chinasa T Okolo, Dhruv Agarwal, Nicola Dell, and Aditya Vashistha. 2024. "If it is easy to understand then it will have value": Examining Perceptions of Explainable AI with Community Health Workers in Rural India. In *Proceedings of the ACM on Human-Computer Interaction (CSCW '24)*. Association for Computing Machinery.

[74] Chinasa T. Okolo, Kehinde Aruleba, and George Obaido. 2023. Responsible AI in Africa—Challenges and Opportunities. *Responsible AI in Africa: Challenges and Opportunities* (2023), 35–64.

[75] Chinasa T. Okolo, Nicola Dell, and Aditya Vashistha. 2022. Making AI Explainable in the Global South: A Systematic Review. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '22)*. Association for Computing Machinery, 439–452.

[76] Google PAIR. 2021. Google People+AI Guidebook. Retrieved August 12, 2023 from https://pair.withgoogle.com/guidebook/

[77] Meredith S Palmer, Sarah E Huebner, Marco Willi, Lucy Fortson, and Craig Packer. 2021. Citizen science, computing, and conservation: How can "Crowd AI" change the way we tackle large-scale ecological challenges? *Human Computation* 8, 2 (2021), 54–75.

[78] Dorian Peters, Karina Vold, Diana Robinson, and Rafael A Calvo. 2020. Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society* 1, 1 (2020), 34–47.

[79] Dragutin Petkovic. 2023. It is not "Accuracy vs. Explainability"—we need both for trustworthy AI systems. *IEEE Transactions on Technology and Society* 4, 1 (2023), 46–53.

[80] Anthony Poon. 2020. Practitioners and ICTD: Communities of Practice Theory in Technology Interventionism. In *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*. 160–173.

[81] Rida Qadri, Renee Shelby, Cynthia L Bennett, and Emily Denton. 2023. AI's Regimes of Representation: A Community-centered Study of Text-to-Image Models in South Asia. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 506–517.

[82] Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2021. Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 7 (apr 2021), 23 pages. https://doi.org/10.1145/3449081

[83] Divya Ramesh, Vaishnav Kameswaran, Ding Wang, and Nithya Sambasivan. 2022. How Platform-User Power Relations Shape Algorithmic Accountability: A Case Study of Instant Loan Platforms and Financially Stressed Users in India. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1917–1928. https://doi.org/10.1145/3531146.3533237

[84] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[85] Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence* 1, 5 (2019), 206–215.

[86] Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi, and Vinodkumar Prabhakaran. 2021. Re-imagining algorithmic fairness in India and beyond. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 315–328.

[87] Dipto Sarkar and Colin A Chapman. 2021. The smart forest Conundrum: Contextualizing pitfalls of sensors and AI in conservation science for tropical forests. *Tropical Conservation Science* 14 (2021), 19400829211014740.

[88] Laura Sartori and Andreas Theodorou. 2022. A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. *Ethics and Information Technology* 24, 1 (2022), 4.

[89] Michael Gomez Selvaraj, Alejandro Vergara, Henry Ruiz, Nancy Safari, Sivalingam Elayabalan, Walter Ocimati, and Guy Blomme. 2019. AI-powered banana diseases and pest detection. *Plant Methods* 15 (2019), 1–11.

[90] Donghee Shin. 2021. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies* 146 (2021), 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

[91] Julia Stoyanovich and Bill Howe. 2019. Nutritional labels for data and models. *A Quarterly bulletin of the Computer Society of the IEEE Technical Committee on Data Engineering* 42, 3 (2019).

[92] Nenad Tomašev, Julien Cornebise, Frank Hutter, Shakir Mohamed, Angela Picciariello, Bec Connelly, Danielle CM Belgrave, Daphne Ezer, Fanny Cachat van der Haert, Frank Mugisha, et al. 2020. AI for social good: unlocking the opportunity for positive impact. *Nature Communications* 11, 1 (2020), 2468.

[93] Kentaro Toyama and M Bernardine Dias. 2008. Information and communication technologies for development. *Computer* 41, 6 (2008), 22–25.

[94] John Traxler. 2012. Ethics and ICTD research. *The Singapore Internet Research Centre (SiRC) initiates and conducts research* (2012), 68.

[95] Rama Adithya Varanasi and Nitesh Goyal. 2023. "It is currently hodgepodge": Examining AI/ML Practitioners' Challenges during Co-production of Responsible AI Values. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[96] Durai Raj Vincent, Natarajan Deepa, Dhivya Elavarasan, Kathiravan Srinivasan, Sajjad Hussain Chauhdary, and Celestine Iwendi. 2019. Sensors driven AI-based agriculture recommendation model for assessing land suitability. *Sensors* 19, 17 (2019), 3667.

[97] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. 2021. "Brilliant AI doctor" in rural clinics: Challenges in AI-powered clinical decision support system deployment. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–18.

[98] Qiaosi Wang, Michael Madaio, Shaun Kane, Shivani Kapania, Michael Terry, and Lauren Wilcox. 2023. Designing Responsible AI: Adaptations of UX Practice to Meet Responsible AI Challenges. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[99] Human Rights Watch. 2023. Poverty Targeting Algorithms Harm Rights. Retrieved June 14, 2023 from https://www.hrw.org/news/2023/06/13/world-bank/jordan-poverty-targeting-algorithms-harm-rights

[100] Oliver R Wearn, Robin Freeman, and David MP Jacoby. 2019. Responsible AI for conservation. *Nature Machine Intelligence* 1, 2 (2019), 72–73.

[101] Hilde Weerts, Miroslav Dudík, Richard Edgar, Adrin Jalali, Roman Lutz, and Michael Madaio. 2023. Fairlearn: Assessing and Improving Fairness of AI Systems. arXiv:2303.16626 [cs.LG]

[102] Lily Xu, Shahrzad Gholami, Sara McCarthy, Bistra Dilkina, Andrew Plumptre, Milind Tambe, Rohit Singh, Mustapha Nsubuga, Joshua Mabonga, Margaret Driciru, et al. 2020. Stay ahead of Poachers: Illegal wildlife poaching prediction and patrol planning under uncertainty with field test evaluations (Short Version). In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 1898–1901.

[103] Nur Yildirim, Mahima Pushkarna, Nitesh Goyal, Martin Wattenberg, and Fernanda Viégas. 2023. Investigating How Practitioners Use Human-AI Guidelines: A Case Study on the People + AI Guidebook. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 356, 13 pages. https://doi.org/10.1145/3544548.3580900

[104] Nur Yildirim, Mahima Pushkarna, Nitesh Goyal, Martin Wattenberg, and Fernanda Viégas. 2023. Investigating How Practitioners Use Human-AI Guidelines: A Case Study on the People+ AI Guidebook. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.

[105] Rie Roselyne Yotsu, Zhengming Ding, Jihun Hamm, and Ronald Blanton. 2023. Deep learning for AI-based diagnosis of skin-related neglected tropical diseases: a pilot study. *medRxiv* (2023), 2023–03.

A Countries Our Participants Worked In

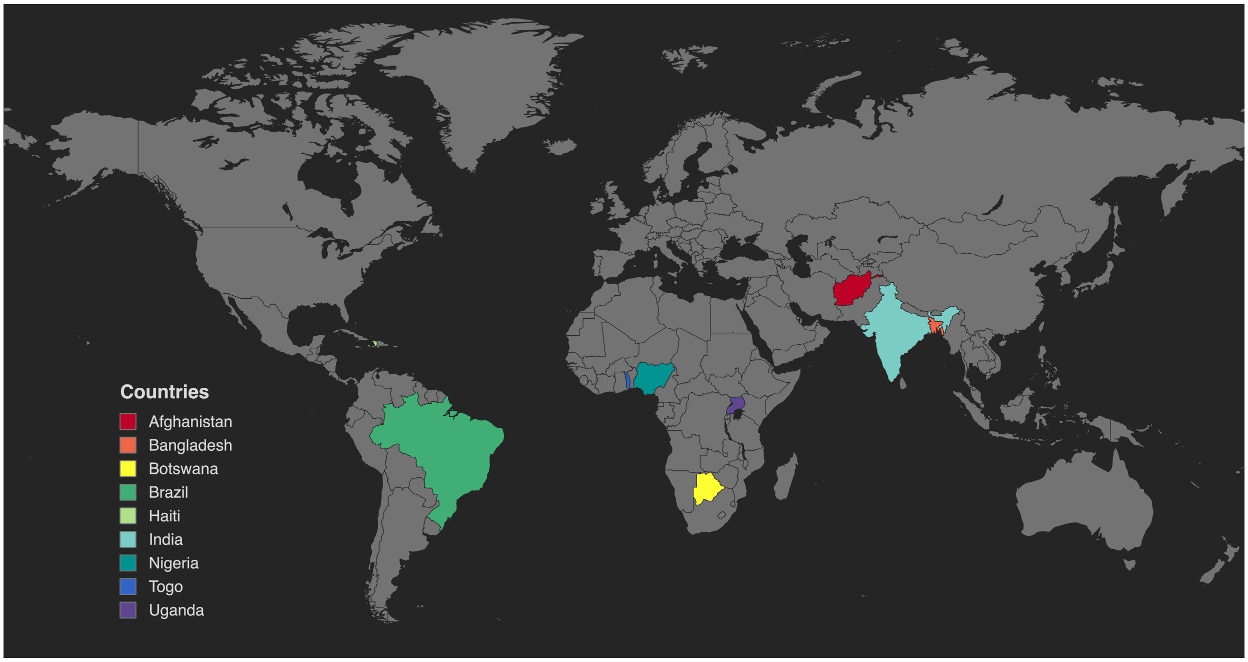

Figure 1: Countries where our participants conduct their AI4SG work. The countries include Afghanistan, Bangladesh, Botswana, Brazil, Haiti, India, Nigeria, Togo, and Uganda.

B Interview Protocol

- **Problem Identification, End User Engagement, and System Deployment**
  - What kinds of social good/impact problems do you address in your work?
    - Please provide examples of projects you've worked on.
  - How do you identify problems to address and determine what AI techniques are needed to solve these problems?
  - What kinds of end users does your work impact?
  - How do you currently engage with end users or collaborators when designing and building AI tools?
    - Do you use any frameworks/methodologies to guide your engagement with end users?
    - How do you understand and address the needs of end users?
    - How do these tools incorporate the expertise of the end users?
    - What do you think users/stakeholders need to know in order to effectively operate the AI tools you develop?
    - What challenges arise when engaging end-users in AI system design and development life cycles?
  - What challenges have you experienced in the deployment of AI or other systems in the field?
    - Who do users reach out to for help if they are confused?
    - Are the systems you deploy intended to be handed off to end users/system partners?
- **Understanding XAI Usage**
  - Do you have a good sense of how your algorithms come to their recommendations/ results?
  - Does your work involve model explainability? (if **yes**, continue questions, if **no** ask "What inhibits you from pursuing model explainability in your work?" and go to XAI Perceptions)
  - Can you discuss some approaches you have taken in your work to incorporate model explainability?
  - What is the approach and who is centered in these approaches (practitioners, end users, both)?
  - How effective were these approaches and did any issues occur when incorporating model explainability?
  - Are there other approaches you use to ensure that AI tools are understandable for end users?
- **XAI Perceptions**
  - Do you think model explainability is useful for end users?
    - Do your users ask about how your algorithms make such decisions?
    - Is it important to your users that they can understand the system?
  - Do you think model explainability is useful for practitioners and researchers (like yourself)? Why?
    - Useful could mean practical, effective in helping with decision-making, easy to implement, etc.

C End User Populations and Machine Learning Domains

ID	Targeted populations	ML methods
P01	Park rangers, conservation managers, community healthcare workers	SV, RL, DL
P02	Citizens	SV, US, DL, CV
P03	Policymakers, humanitarian aid recipients	SV, DL, CV
P04	Farmers, social media users	SV, SSV, US, RL, DL, CV, NLP, ASR
P05	Farmers, extension agents	SV, US, RL, DL, NLP
P06	General public	SV, DL, CV, NLP
P07	Farmers, translators, policymakers	SV, CV, NLP
P08	Educators, students, medical practitioners (doctors and nurses)	SV, SSV, US, DL
P09	Community healthcare workers, educators, linguists, farmers	SV, DL, CV, NLP
P10	Communities affected by air pollution (i.e., construction workers)	SV, DL, NLP , ASR
P11	Early-career IoT engineers and students	SV
P12	Farmers, social workers, policymakers	SV, SSV, US, CV, NLP

Table 2: The target end user populations of the AI4SG tools our participants built and the machine learning (ML) methods used within their work. The methods are abbreviated as follows: Supervised Learning (SV), Semi-supervised Learning (SSV), Unsupervised Learning (US), Reinforcement Learning (RL), Deep Learning (DL), Computer Vision (CV), Natural Language Processing (NLP), Automated Speech Recognition (ASR).