"You can't build what you don't understand": Practitioner Perspectives on Explainable AI in the Global South

Chinasa T. Okolo
Center for Technology Innovation
The Brookings Institution
Washington, D.C., United States
cokolo@brookings.edu

Hongjin Lin
Computer Science
Harvard University
Cambridge, Massachusetts, United States
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. Our work reflects on the shortcomings of XAI for real-world use and advocates for a reimagined agenda for human-centered 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. "You can't build what you don't understand": Practitioner Perspectives on Explainable AI in the Global South. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3613905.3651080

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, 32]. 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 [5, 7, 35], 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 [13, 31]. 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 [22]. Given that the concepts of AI "responsibility", "transparency," and "accountability" widely encompass various aspects and how "explainability" is a key facet of prominent responsible AI frameworks from the U.S. National Institute of Standards & Technology [21], 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.

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 [12, 17, 19, 20, 33, 34, 36]. 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 [23], a lack of access to relevant datasets [1], and a lack of local AI developers [26], 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. [27], 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 [25, 27], it is important to examine explainability practices to better understand ethical considerations around the fairness, trustworthiness, and potential misuse of AI systems deployed in this region.

To address the gap in research examining XAI in the Global South, 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 consider and implement model explainability in designing, developing, and deploying their AI systems in the Global South?

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3651080

RQ3: How do AI for Social Good practitioners working in Global South contexts perceive the usefulness of model explainability for themselves and their end users?

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 expertise 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 do not necessarily address the challenge of communicating with end users with different domains of expertise and varying levels of technical fluency.

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 detail how AI4SG practitioners consider model explainability when developing and deploying their technologies.
- We present considerations on reshaping existing notions of explainability to meet the needs of AI4SG practitioners and their end users.

2 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.

Participants Our study recruited AI4SG practitioners in academia, industry, nonprofit, 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 [15]. 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 [28], 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. To ensure that participants met the inclusion criteria for our study, we leveraged purposive sampling [18] 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). We initially recruited 14 participants, but two participants discontinued their participation in the study. They 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 about our participants in Table 1. In the appendix, we provide information about the end users our participants work with (Appendix C) and the countries they work in (Appendix A).

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). Appendix C details each participant's specific usage of ML techniques. 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.

Interview Procedure We conducted semi-structured interviews over Zoom, which were split into two parts: (1) understanding the social impact problems addressed by practitioners, the types

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

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. 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.

Analysis We collected 7.7 hours of audio recordings from our interviews. After transcribing the interviews, we used inductive thematic analysis [9] 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".

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.

Limitations This project is subject to several potential limitations. The primary constraint of this project lies in our small participant sample size. 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.

3 FINDINGS

Our interviews focused on practitioners' perspectives using XAI for and experiences developing AI4SG projects *in* the Global South; thus, all findings reflect experiences specific to the contexts/countries they worked in this region. This section begins by investigating AI4SG practitioners' motivations for engaging in AI4SG work and how AI4SG practitioners identify social impact problems to address (Section 3.1). We then examined the kinds of end users AI4SG practitioners engage within their work and what tools, methodologies, or strategies they use to engage these users (Section 3.2). Next, we explore how practitioners perceive XAI usefulness in their work and for end user engagement (Section 3.3). Lastly, we unpack how AI4SG practitioners incorporate XAI techniques in practice (Section 3.4).

3.1 What problems are AI4SG practitioners targeting in their work?

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.

3.2 How do AI4SG practitioners engage end users?

Who are the 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 pipeline development. 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. We also find that 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 (P04) and humanitarian aid recipients (P03). In Agriculture, our interviewees worked with banks as direct users (P04), extension agents as direct end users (P05), and farmers as both direct (P12) and indirect end 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).

End user engagement methods and challenges. Our practitioners engaged with users 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 [33], 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 cause practitioners to feel unsupported and take on additional tasks on top of their existing workloads [11, 33, 36].

AI4SG product deployment and challenges. 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"*. 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), and the fast turnaround of AI4SG projects and publication timelines (P01, P04, P12). 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 [27, 29, 30].

System hand-off. 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.

3.3 How do AI practitioners perceive XAI in their work and for their users?

Effectiveness of XAI approaches in AI4SG work. 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 that it is right in translating gender-neutral sentences into a gender-specific sentence"* and often *"bends towards a specific [gender] stereotype"*.

Perceptions of XAI usefulness for AI4SG practitioners. 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.

Perceptions of XAI usefulness for end users. 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 [25, 27]. 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."* 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.

3.4 How are AI4SG practitioners incorporating explainability into their work?

AI4SG practitioners' 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.

While all of our participants agreed that XAI would be helpful for their work (Section 3.3), 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 2) but did not provide specific examples of XAI tools or methods. If participants were still unable to name XAI methods, we considered dubious responses to indicate 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).

Who is centered in XAI approaches? We found that indirect end users were rarely centered in practitioners' usage of XAI approaches. P03 and P07 mentioned that their use of XAI centered on both end users and practitioners. In P03's AI4SG projects, the direct end users are policymakers who are in charge of operating humanitarian social assistance algorithms. In contrast, the citizens who are the targets of these social assistance programs are indirect end users because these algorithms impact their lives. P07 worked across multiple AI4SG domains and thus interacted with various end users. While P03 mentioned that XAI was mainly used by their research team and for communicating results to policymakers, they expressed interest in developing alternative, nontechnical explanation methods to engage indirect end users affected by the algorithms developed in their work.

Nontechnical notions of XAI. From the interviews with our participants, *we 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 that their research team provides *"a set of instructions during the deployment that is available in web or mobile format* and makes *"simple instructions to ensure the user understands what's happening."*

Given that our findings highlight multiple dimensions of explainability outside of technical methods, we find that further development of "sociotechnical XAI" [14], XAI that incorporates social and intercultural aspects [24] 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 3.2 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 implementing XAI in AI4SG work. Our participants who implemented XAI methods in their work faced various challenges, including 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, factors that impeded their use of XAI were the high efforts 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 [10]. 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.

4 DISCUSSION

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.

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 [6, 16]. Our study shows similar findings, detailing the challenges that affect how AI4SG practitioners choose to use XAI when developing 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 [8].

A Future Agenda for Explainability Research While one of the aims of our study was to understand how AI4SG practitioners incorporate XAI into their work, some of our participants noted that "explainability" is often 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 to understand the needs of both practitioners and end users when using explainability in practice. Such work could leverage the framework developed by Ehsan et al. [14] to actively engage stakeholders, particularly end users, to understand what they need explained from these systems and how such aspects could be interpreted. 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 sociotechnical aspects of explanations while being effective for a broad spectrum of users and use cases.

5 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.

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] 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.

[6] 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.

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

[8] 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.

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

[10] 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).

[11] 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

[12] 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

[13] 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

[14] 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.

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

[16] 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).

[17] 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).

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

[19] 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.

[20] 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.

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

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

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

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

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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

[30] 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.

[31] 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

[32] 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.

[33] 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.

[34] 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.

[35] 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

[36] 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.

A COUNTRIES OUR PARTICIPANTS WORKED IN

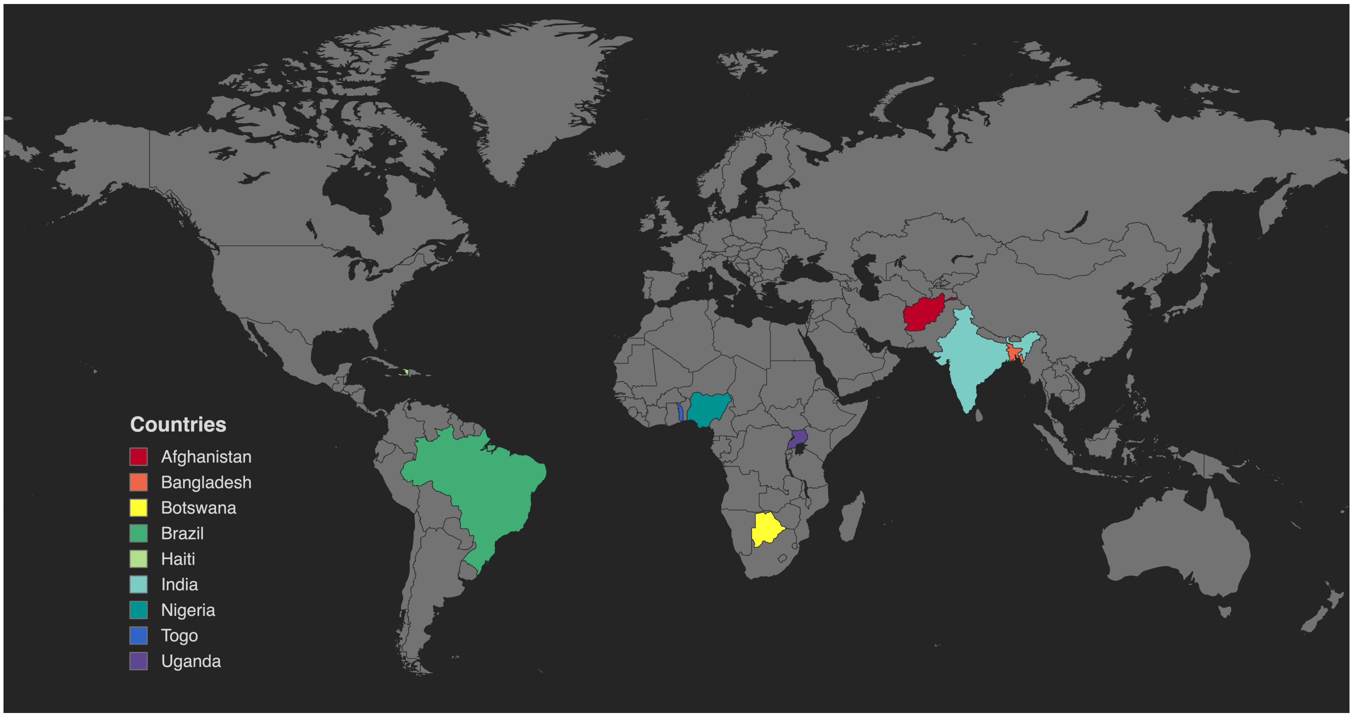

Figure 1: Countries where our participants conduct their AI4SG work. The countries include Afghanistan, Bangladesh, Botswana, Brazil, Haiti, India, Nigeria, Togo, and Uganda.

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).