Explorable Explainable AI: Improving AI Understanding for Community Health Workers in India

Ian René Solano-Kamaiko
Cornell Tech
United States
irs24@cornell.edu

Dibyendu Mishra
Cornell University
United States
dm926@cornell.edu

Nicola Dell
Cornell Tech
United States
nixdell@cornell.edu

Aditya Vashistha
Cornell University
United States
adityav@cornell.edu

Abstract

AI technologies are increasingly deployed to support community health workers (CHWs) in high-stakes healthcare settings, from malnutrition diagnosis to diabetic retinopathy. Yet, little is known about how such technologies are understood by CHWs with low digital literacy and what can be done to make AI more understandable for them. This paper examines the potential of explorable explanations in improving AI understanding for CHWs in rural India. Explorable explanations integrate visual heuristics and written explanations to promote active learning. We conducted semi-structured interviews with CHWs who interacted with a design probe in which AI predictions of child malnutrition were accompanied by explorable explanations. Our findings show that explorable explanations shift CHWs' AI-related folk theories, help develop a more nuanced understanding of AI, augment CHWs' learning and occupational capabilities, and enhance their ability to contest AI decisions. We also uncover the effects of CHWs' sociopolitical environments on AI understanding and argue for a more holistic conception of AI explainability that goes beyond cognition and literacy.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**; **HCI design and evaluation methods**.

Keywords

Explainable AI, Human-centered Explainable AI, Explorable Explanations, user study, community health workers, Global South

ACM Reference Format:
Ian René Solano-Kamaiko, Dibyendu Mishra, Nicola Dell, and Aditya Vashistha. 2024. Explorable Explainable AI: Improving AI Understanding for Community Health Workers in India. In *Proceedings of 2024 ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '24).* ACM, New York, NY, USA, 21 pages. https://doi.org/XXXXXXX.XXXXXXX

CHI '24, May 11-16, 2024, Hawaii

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Malnutrition accounts for 45% of the deaths among children under the age of five worldwide [95]. Given its prevalence and dangerous consequences, a key target of the United Nations Sustainable Development Goals is to "end all forms of malnutrition by 2030" [123]. Of the three indicators used to measure malnutrition— i.e., stunting (low height for age), wasting (low weight for height), and underweight (low weight for age)—India has one of the highest proportions of stunted and wasted children [52, 124]. To fight malnutrition, India employs a large network of community health workers who are trained to provide maternal and neonatal care in hard-to-reach regions [2, 115]. A crucial component of this network is Anganwadi Workers (AWWs) who are women recruited from local communities to provide a number of essential services, such as supplementary nutrition, preschool education, growth monitoring, immunizations, and antenatal and postpartum care [2]. While AWWs have been shown to improve maternal and neonatal outcomes [122], they experience several challenges including limited training and upskilling opportunities [132], unpredictable and long work hours [83], and overburdened schedules resulting from inadequate infrastructure, workforce shortages, and budgetary cuts [28, 55, 77].

To optimize their operational efficiency, AWWs are increasingly experiencing occupational digitalization from a variety of for-profit and not-for-profit organizations aimed at developing AI tools to automate parts of their work. These tools range from dynamically designing AWWs' work schedules [78], to helping them diagnose diseases such as malnutrition and diabetic retinopathy [15, 19, 56], to analyzing rapid diagnostic tests [98, 105]. With the burgeoning integration of AI into AWWs' workflows, it is critical to ensure that AWWs understand AI and use these tools safely and responsibly.

The field of explainable AI (XAI), which provides a set of methods aimed at making AI decisions more understandable and trustworthy to human stakeholders, can provide important design affordances for developing responsible AI-driven systems that assist AWWs [11, 46, 69]. However, Okolo et al. [92] found that very little XAI research has been done with users in the Global South who operate within unique technological, economic, and sociocultural contexts. Moreover, there are distinctive questions regarding XAI for AWWs given the fact that many of these stakeholders are digital novices (i.e., they have only started to interface with digital technologies in a significant way within the past decade) [53] and possess low levels of digital literacy [8]. To date, there is a scarcity of research investigating AI-related needs and perceptions of AWWs as well as the benefits and challenges associated with the integration of AI into their workflows [54, 92, 93]. Little is known about how AI technologies are understood by AWWs and what can be done to make AI more understandable for them [92, 93].

In this paper, we examine the potential of *explorable explanations* at improving AI understanding for AWWs in rural India. Explorable explanations incorporate human-centered, interactive visual heuristics and written explanations to promote a more active form of learning and calibrated trust [41, 61, 67, 127]. Researchers have argued that XAI methods should be interactive so that explanations can help facilitate follow-up questions to close the understanding gap [7, 69, 76, 82]. Along with interactivity, explorable explanations provide a simple and approachable method for promoting end-users' active learning [41] and more deliberative thinking [21, 58]. For these reasons, we studied the role explorable explanations can play in improving AI understanding for AWWs. To investigate our approach, we conducted semi-structured interviews with 30 AWWs who interacted with a design probe in which AI predictions of child malnutrition were accompanied by explorable explanations. We chose child malnutrition as our study context given the prevalence of the condition in India and the steady increase of AI-driven anthropometry tools in AWWs' workflows [1, 15]. Given AWWs low levels of AI literacy and the fact that XAI is a fairly nascent field, we decided to used a design probe to help facilitate our research. Design probes have been shown to be particularly effective in stimulating elaborate user feedback from respondents who lack technological knowledge and are especially useful when the design space is amorphous [51, 85, 131].

We built our design probe as a *Figma* [40] prototype, which enabled AWWs to interact with an AI-driven child malnutrition prediction tool accompanied by explorable explanations. The probe did not run a real machine learning (ML) model to make predictions but instead was *instrumented* to mimic how an AI-driven application might behave in production. In addition to the *Figma* prototype, we also used a physical doll as a proxy for a child with malnutrition. We used the doll and prototype in tandem to emulate the process of an AWW using an AI-driven application to predict the nutritional status of a child. In our study, participants began by assessing the nutritional status of the child (doll) and then interacted with the design probe. They first reviewed the AI prediction, and then engaged with the explorable explanations, which consisted of four main components: *Feature Information Modals*, *Edit Measurements Section*, *Feature Importance Section*, and *Comparison Screens*. These explorable XAI methods expand on the existing literature on human-centered XAI, explorable explanations, and inclusive design for digitally novice users. The design probe allowed us to capture AWWs' AI-related folk theories and understanding, in addition to their critical reflections on the benefits and limitations of these explorable explanations.

Through thematic analysis of observations and interviews with AWWs who engaged with the design probe, we found that **explorable explanations shaped the AI-related folk theories of AWWs and helped them develop a more nuanced understanding of AI**. As a result of these shifts in AI-related folk theories and understanding, we observed that explorable explanations increased AWWs' skepticism of AI and enhanced their ability to contest AI predictions when they perceived the system made mistakes. AWWs also found explorable explanations to be a useful pedagogical tool that could augment their domain knowledge and help to collaboratively improve their accuracy of diagnosing malnutrition. We also found that AWWs' social, political, and economic environment heavily shaped their notions related to AI adoption and understanding. Based on these findings, we discuss the need for a more holistic conception of AI explainability that goes beyond cognition and literacy to consider aspects such as user training and community engagement as core tenants of explainability efforts. Taken together, this work makes three contributions to HCI and XAI scholarship:

- We conduct one of the first studies to examine the role explorable explanations can play in shifting AI-related folk theories of AI end-users.
- Through qualitative interviews, observations, and engagement with a design probe, we provide empirical evidence showing that explorable explanations can improve AI understanding for AWWs who engage with an AI-driven malnutrition prediction application.
- We discuss the need to situate AI interventions and efforts to improve AI understanding within the sociocultural, sociopolitical, and socioeconomic realities of the context within which these systems are integrated.

2 Related Work

In this section, we contextualize our work by discussing the pivotal role played by AWWs as the primary point of contact for millions of marginalized people in hard-to-reach areas in rural India. We explore the rationale behind our study by positioning our research within the realm of AI interventions aimed at supporting community health workers (CHWs) and conclude by situating our work in current scholarship on making AI explainable to novice users.

2.1 AI and Community Health Workers

Anganwadi Workers (AWWs) are an important part of the frontline workforce of the Integrated Child and Development Services (ICDS), which is a flagship government program that supports the health and developmental needs of women and children in rural India. AWWs' responsibilities include monitoring children's nutritional status, providing antenatal care, administering supplementary nutrition, conducting immunizations, and engaging in community health counseling [122]. Although AWWs' have a formal role within the government's workforce and receive a stipend as opposed to the incentive based pay that other CHWs receive, their compensation remains far below similar occupations in the country [97]. Despite numerous unionization efforts, AWWs still struggle for basic bargaining power and continue to receive below-market compensation as well as untimely wage payments [55, 97]. These labor conditions are critical to holistically understanding the environment in which these types of AI-driven systems are being deployed and the subsequent sociotechnical challenges.

In order to make determinations regarding the nutritional status of children, AWWs use a variety of measuring devices such as infantometers, tape measures, stadiometers, and growth charts as well as assess behavioral indicators. They do this work primarily by visiting families at their places of residence or meeting with them at Anganwadi Centers (AWC), which are generally located in rural town centers. Although these assessments were previously recorded in physical registers, the current process has been completely digitized. A number of mobile applications [91] have been introduced in an effort to increase operational and administrative efficiency. Our study participants, who are AWWs located in the northern state of Uttar Pradesh, India use at least four different mobile applications on a day-to-day basis to complete their work. These devices include a nutrition data entry application called *Poshan Tracker*, an infrastructural needs and requests tracking application called *Ek Sang*, an application to interface with supervisors called *Sahyog*, and an educational application called *Bal Pitara* [101].

Given these digitization efforts, there exists a large body of work within HCI that has examined the use of technologies to support CHWs in India. For example, researchers have designed audio-visual tools and mobile applications to improve information dissemination [64, 103], explored how technologies can improve CHWs' workflows and productivity [33, 78], examined data collection and digitization tools [35, 96], and assessed decision support systems [34, 35]. Recent HCI scholarship has also studied AI tools and their use by CHWs in resource-constrained settings. For instance, Ismail and Kumar [54] outline the increasing use of AI technologies by CHWs in India, which is driven by equipment and shortages of medical personnel. Several projects also aim to improve child and maternal health using anthropometric tools for monitoring nutritional indicators [15, 19, 25]. Others have looked at AI-powered risk assessments related to obstetrics [117] and early detection of prevalent diseases, such as breast cancer [104], diabetic retinopathy [56] and tuberculosis [80]. The COVID-19 pandemic also spurred the development of AI-based tools to help CHWs to perform contact tracing and diagnosis [98, 105].

However, a growing number of studies have underscored the challenges posed by digital and AI systems on CHW workflows. Researchers have highlighted that these applications do not capture the invisible work done by CHWs and disregard their domain expertise by reducing them to mere data collectors [111, 126]. Often, the care work takes a backseat and data work becomes an end in itself, given the sheer amount of energy related to data collection and the coercive environment in which CHWs operate [96, 109]. While a study regarding the effectiveness of such digitization tools in large-scale public health programs has yet to be seen, recent research has begun to identify some of the potential harms and challenges concerning these systems including patient misdiagnoses, hyper-surveillance of CHWs, infrastructural mismatches, increases in invisible labor for CHWs, and the obfuscation of CHW concerns [53, 109, 126]. Subsequently, researchers have called for a strong human-centered understanding of CHWs' needs and desires, advocating for co-designing tools with CHWs in order to incorporate their domain expertise into these systems [9, 54, 65, 93, 99]. Even prominent global health foundations, policy think tanks, and government institutions, who otherwise promote the widespread use of AI-driven technologies [5, 42, 94], have called for a more responsible use of AI citing issues related to direct harm, transparency, inclusion, and evaluation [6, 94]. Through our research, we add to this body of work by expanding the limited literature on the understanding of AI-driven systems by novice users such as CHWs and brings the perspectives of these frontline workers in the Global South to a field dominated by the viewpoints of those in the Global North [71, 92].

The proliferation of AI-driven applications in community health work calls for a serious inquiry into AI literacy for low-digitally literate users like AWWs, who are both the primary users and one of the most directly affected stakeholders of these systems. Long and Magerko [74] define AI literacy as *"a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace"*. They see digital literacy, i.e., the competency needed to use computational devices, as a prerequisite but do not find computational literacy (i.e., the ability to program and understand the internal mechanics of a computer) necessary for AI literacy. The authors underscore several design considerations, including developing XAI tools and the implementation of "low barrier to entry" approaches as effective ways to increase AI literacy [74]. In this paper, we extend these recommendations by exploring interactive, human-centered approaches for developing XAI methods that focus on increasing AI literacy for AWWs who are digital novice users with low AI know-how.

2.2 Human-Centered Explainable AI

XAI refers to the set of methods and techniques to make AI decisions more understandable to human users [11, 46, 69]. While the term explainability is used by the research community with varying scope, for the purpose of this research, we are focusing on XAI that allows stakeholders to understand and calibrate appropriate trust in the results of machine learning (ML) algorithms. Typically defined as post-hoc explainability, three of the most popular methods include *LIME* [107], *SHAP* [75], and *SAGE* [31]. These approaches are generally divided into three categories: local, cohort, and global explainability. Local explainability is used to explain individual predictions, cohort is used to understand predictions over a subset of data, and global explains the model's behavior across the whole dataset [11, 69]. Although our work generates inspiration from all three approaches, we focus predominantly on local explainability as these techniques are typically the most relevant for end users [14].

Given the increasing prevalence of AI-driven systems that are responsible for high-stakes decisions in critical domains such as healthcare [57, 79, 134], AI researchers have sought to better explain these systems to human stakeholders [45, 73], highlighting issues around cognitive overload, non-comprehensibility of explanations, and excessive reliance on AI. Despite these efforts, the discourse around XAI has predominantly focused on algorithm-centered approaches, which may not fully address the diversity of AI literacy levels, the concerns from end users, and may exacerbate issues of algorithmic opacity [39]. As a result, XAI researchers have begun to emphasize the importance of developing human-centered approaches that center the technical development of XAI methods on people's explainability needs [38, 39, 70, 130]. In 2020, Ehsan and Riedl [38] introduced "Human-centered Explainable AI" (HCXAI), a framework that prioritizes human considerations and advocates developing a holistic understanding of diverse stakeholders, considering the *"interplay of values, interpersonal dynamics, and the socially situated nature of AI systems"* [38]. Liao and Varshney [70] further highlight centering XAI design on individual stakeholders moving away from one-size-fits-all solutions, challenging technocentric assumptions, and drawing insights from theories on human cognition and behavior [70, 130].

Despite AI rapidly becoming an integral part of people's daily lives all around the globe, current scholarship on making AI explainable focuses mainly on communities and contexts in the Global North. A systematic literature review conducted by Okolo et al. [92] identified that only 0.08% of approximately 18,000 articles published on XAI focused on contexts in the Global South, despite the region housing two-thirds of the global population [100]. Little is known about their needs, workflows, and contexts within which XAI methods must be designed, deployed, and tested. This is particularly concerning given the growing interest from governments, companies, and academics in using AI/ML in high-stakes and sensitive domains. Our study contributes to the emerging literature by centering the perspectives of AWWs and investigating design heuristics that increase their understanding of AI. In the next section, we delve into capturing AWWs' mental models of AI through folk theories, which is our method of choice for assessing AI understanding in the study.

2.2.1 AI Folk Theories and Mental Models Researchers have found value in eliciting users' mental models to examine their perceptions of AI explainability methods [44, 63, 108]. Mental models are a concept drawn from cognitive psychology, which are *"people's continuously evolving cognitive representations of a system that incorporate their beliefs regarding the way the system works"* [60]. These are different from conceptual models which are a more accurate and complete representation of the target system and are developed by experts or designers [89]. Given the variability in users' understanding of an AI system, researchers use methods of varying complexity to elicit these mental models. These range from intricate written explanations tailored for users with high system knowledge [63] to more accessible think-aloud activities used while playing games that involve the system [44].

For digitally novice users, like AWWs, who would have trouble articulating conceptual models of the system, providing avenues for folk theorization which are *"intuitive, informal theories that individuals develop to explain the outcomes, effects, or consequences of technological systems"* [36] is a more suitable approach. Since mental models rely heavily on participants' prior experiences and environmental factors, it is unclear whether existing methods, which have either focused on users with some level of familiarity with AI systems or have predominantly been conducted in the Global North, are appropriate for users from other demographics.

2.2.2 Explorable Explanations The term "explorable explanation" was first introduced in a 1994 paper by Brusilovsky [20] but it did not become commonly used until 2011 when Victor [127] published an eponymous essay on the subject. Victor [127] situates explorable explanations in the context of "active reading" of static documents, where he encourages design affordances that can empower readers to interact with and modify content, enabling them to question, explore and verify the presented information [127]. Importantly, explorable explanations contain interactive elements combined with static graphics, relying heavily on the importance of "play" as a learning didactic [41, 87]. They encourage users to discover things about concepts for themselves and test their expectations of its behavior against its actual behavior, promoting a more active form of learning [41]. In fact, several researchers have previously argued that XAI methods should be interactive so that explanations can help facilitate follow-up questions from users to close the understanding gap [7, 69, 76, 82].

In addition to interactivity, another advantage of explorable explanations is the simplicity and synchronous feedback of the approach. While other post-hoc explainability methods rely heavily on end-user mathematical and computational literacy [31, 75, 107], explorable explanations, as aforementioned, are grounded in "play" as an intuition building mechanism. As a result, explorable explanations have the potential to be more appropriate for users who possess limited digital and AI literacy. Additionally, the active learning component of explorable explanations requires users to thoughtfully interact with the prediction explanations, which subsequently helps them engage in System 2 (slow and deliberative) thinking [58]. Researchers such as Buçinca et al. [21], have advocated for AI explanations that require people to exert effort (deliberative thinking) to build AI understanding.

HCI researchers have been increasingly incorporating explorable explanations through Augmented Reality, audio guides, and interactive visualizations in the areas of data journalism [29, 30, 66, 67], education [27, 88] and programming [129]. These studies show that explorable explanations helped convert static and prescriptive media into knowledge creation tools, reduced cognitive overload and task-related stress among participants, and improved conceptual clarity of the target systems. Recently, researchers have started to implement explorable explanations as a way of teaching ML concepts to students and developers through visual analytics [49, 59] and output exploration [68, 110]. For example, *TensorFlow Playground* [118], a visual web application, is a popular tool that allows users to directly manipulate neural networks and build their intuitions about how they work.

To date, the work on explorable explanations has focused mostly on users in the West with computational literacy, such as students and software developers. Little is known about: (1) whether explorable explanations can improve AI understanding and (2) how well these explanations work for users with limited digital literacy and AI knowledge in the Global South. Our work extends the current scholarship on explorable explanations and human-centered XAI by examining the following research question: Do explorable explanations improve AI understanding for AWWs in rural India?

3 Methodology

We first present the design probe that we used to elicit AI folk theories of AWWs and then describe the design of the explorable explanations integrated within the probe to make the inner workings of AI more understandable to AWWs. Next, we present our study protocol, the demographic data of the 30 AWWs who participated in our research activities, the analysis we conducted on the collected data, and our author positionality statement.

3.1 Probe Design

Many HCI researchers have used design provocations and cultural probes to stimulate user feedback and discussions [17, 43, 51, 128]. These methods have proven to be valuable when working with participants to critically and concretely evaluate the design utility of technologies in real-world settings, such as community healthcare work [85]. More recently, researchers studying AI have begun to use design probes to gain insights into how to develop XAI systems for key stakeholders [48, 106, 121, 135]. Given the limited familiarity of AWWs with AI generally and XAI tools specifically, we designed a probe in *Figma* to enable AWWs to engage with an AI-driven tool to detect child malnutrition, where the AI system's predictions were accompanied by explorable explanations (discussed in detail below).

We selected child malnutrition as our focus for three main reasons. Firstly, numerous organizations are already actively developing and deploying AI-driven systems to identify malnutrition in the Global South. One such initiative, the "Child Growth Monitor Project" by Welthungerhilfe and Microsoft [1], served as inspiration for our design probe. This not only ensured that our study was grounded in a real-world AI application developed for AWWs in the Global South, it also allowed us to begin to understand some of AWWs' perspectives regarding the utility and appropriateness of these tools. Secondly, the concise visual indicators used to detect malnutrition make it well-suited for explorable explanations. Unlike many other common diseases, malnutrition attributes, such as height or weight, are visually apparent, aligning effectively with the visualizations and XAI techniques emphasized in our design exploration. Lastly, childhood malnutrition is prevalent in India and the government has prioritized it as a national health concern. Addressing malnutrition is a primary responsibility of AWWs, making it a crucial context for investigating AI-driven applications for community healthcare in the Global South. While we based our probe on the existing AI-driven malnutrition applications, more research is necessary to understand whether these tools are actually useful for AWWs especially given that AWWs are critically underrepresented in the design, development, and implementation of these AI-driven systems.

Our design probe represented two narrative directions, one where the probe classified the child as malnourished and the other where the child was classified as normal. We also used child body measurements and growth charts from online resources published by the World Health Organization (WHO) [4] and used the WHO's definition of malnutrition, which defines the condition as *"deficiencies or excesses in nutrient intake, imbalance of essential nutrients, or impaired nutrient utilization"* [3]. The probe without explorable explanations consisted of three screens: a launch screen, photo upload screen, and AI prediction screen (see Figure 4b). The probe did not run an actual machine learning (ML) model to make predictions. Instead, it was instrumented to mimic how an AI-driven application might behave in production. The prototype was accompanied by a set of pre-selected child photos (photos of three different dolls) and outputted predefined predictions accompanied by explorable explanations. Next, present the design of explorable explanations, which participants could interact with to test their expectations of the AI behavior and to understand the inner workings of the AI prediction.

3.2 Design of Explorable Explanations

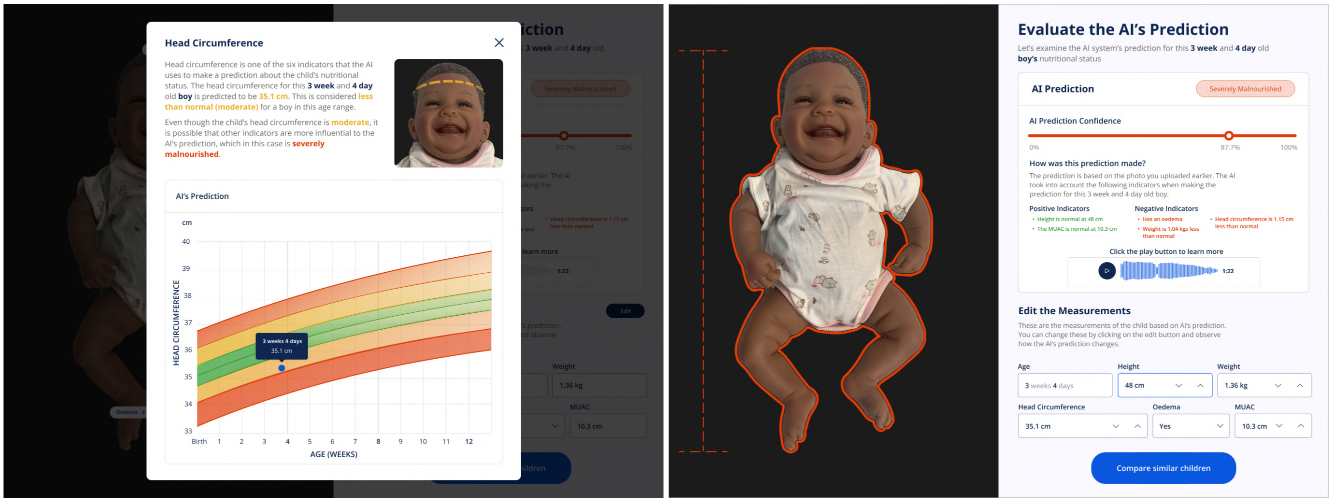
The explorable explanations consisted of four key design elements: *Feature Information Modals*, *Edit Measurements Section*, *Feature Importance Section*, and *Comparison Screens*. Each of these interfaces incorporated important design heuristics and affordances, which are grounded in the HCI and XAI literature, as is discussed below.

3.2.1 Feature Information Modals The feature informational modals explained the model features for the child and contextualized the corresponding feature measurement. As can be seen in Figure 2a, we overlaid clickable buttons on the child's photo, which corresponded to the different AI model features. Once a button was clicked, the modal overlay was triggered (see Figure 1a). The model contained informational content that was relevant to the specific features including: a textual overview, illustration of how the measurement was taken, and feature measurement chart indicating the various outcome thresholds. This design leveraged findings from Simkute et al. [116], which encourages the "ability to expand information and see 'noise' in the data" to better support explainability mechanisms for medical experts.

Researchers have also emphasized the value of data visualizations for AI explainability and transparency methods [12]. Initially, we designed this user interface (UI) element to incorporate a scatterplot diagram of the global feature measurements and corresponding child predictions, which was inspired by *SAGE* [31], a post-hoc explainability method for explaining the model's behavior across the entire dataset. However, in our pilot study, we observed that the plot confused participants who were not familiar with reading mathematical diagrams. In an effort to "meet people where they are" and adopt a culturally responsive design that uses familiar graphics [119], we iterated on the design and provided AWWs with a feature measurement chart that no longer captured global feature data but provided local explanations, which are more useful for end users [14].

3.2.2 Edit Measurements Section Prior work on XAI argues that explanations should help facilitate follow-up *what-if* style questions by users to close the understanding gap [7, 69, 76, 82]. Building on prior work on explorable explanations [41, 67, 127], we incorporated the ability of AWWs to modify the AI model measurements and see how the changes influenced the AI prediction (see Figure 1b). For example, if a participant modified the child's height, this change was reflected in the prototype by altering the outline of the child's photo and the value in the *Feature Importance Section* (see Section 3.2.3). Based on the modification to the measurement, the change would also affect the confidence and prediction of the model.

In the prototype, model confidence was designed as a UI slider, which communicated the probability that the AI prediction was correct. The work of Miller [82] found that probabilities or statistical relationships seem to not matter unless accompanied by causal explanations. As a result, we used color and UI interactivity in the design probe to connect the relationships between the model



(a) Explorable Explanations Feature Information Modal. **(b) Explorable Explanations Edit Measurements Section.**

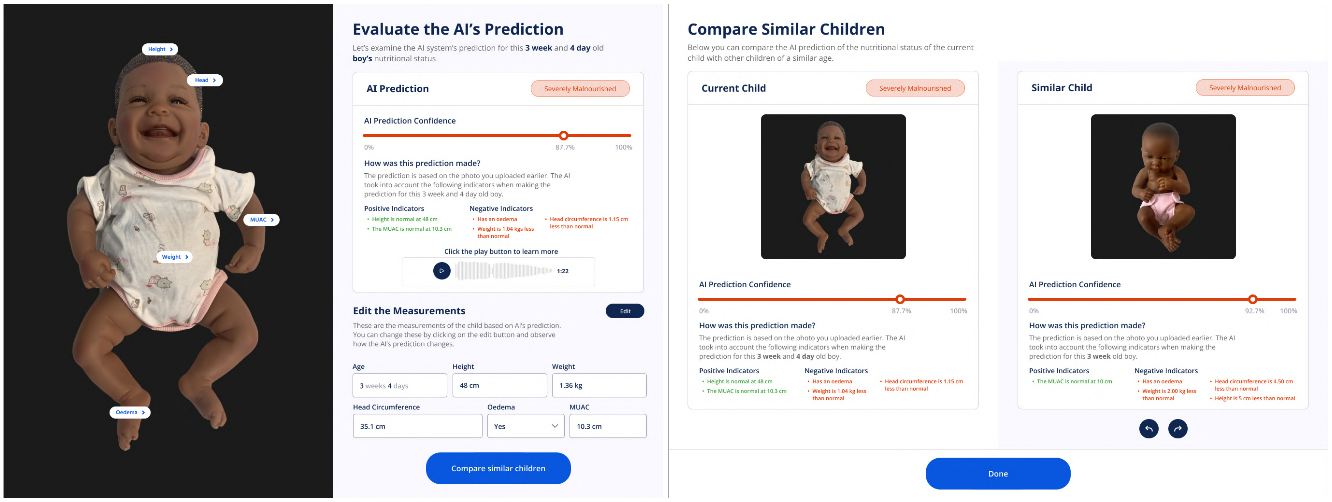
Figure 1: (a) The *Feature Information Modal* is showing the child's head circumference feature. At the top of the modal, in large dark blue text is the modal header ("head circumference") followed by a textual overview of the head circumference feature. The overview text color is primarily gray, with yellow and red colors highlighting different keywords and phrases. The first paragraph states that the child is three-weeks and four-days-old (displayed in bold-ed gray text) and has a head circumference measuring 35.1 centimeters (represented in yellow text). The information also states that the head circumference falls within the moderate range (shown in yellow text). In the second paragraph, the text cautions the user that despite the head circumference falling within the moderate range, the overall prediction for the child is severely malnourished (indicated by the red text). To the right of the header and overview text is a diagram illustrating how the head circumference measurement is taken. The diagram shows a cropped portrait image of the child with a dashed yellow line around the child's head, indicating where the head circumference measurement is taken. Below both the textual overview and illustration is a head circumference measurement chart, which shows the various outcome thresholds as displayed in red, yellow, and green color bands. The marker on the graph communicates the same measurement as is in the text (35.1 centimeters) and indicates where the marker falls in the malnutrition severity level bands (bottom yellow band). (b) The *Edit Measurements Section* is split into two panels. On the left hand side is a photo of the child with a red outline and measurement stick that corresponds to the modified height of the child. On the right side is the AI prediction panel and interactive heuristics. The right panel contains the AI prediction, confidence slider, textual overview, list of the model features, audio player, and modifiable measurement inputs. The design displays the overall prediction in a red colored pill-shaped icon, which indicates that the child is severely malnourished. The text overview lists that the height and upper-arm circumference are in the normal range (indicated by the green text) and oedema, weight, and head circumference in the severe range (indicated by the red text). At the bottom of the right panel are the editable input fields for each measurement: age, height, weight, oedema, upper-arm circumference, and head circumference.

features (measurements) and the confidence slider. We chose percentages to communicate confidence probabilities, as this approach is supported by the literature on inclusive design for low-literate users, which emphasizes the use of numerical knowledge [119]. In general, these playful interactive design affordances encouraged AWWs to discover AI properties and behaviors themselves and test their expectations of the systems' behavior against its actual behavior, promoting a more active form of learning.

3.2.3 Feature Importance Section *LIME* [107], *SHAP* [75], and *SAGE* [31] are three of the most popular post-hoc explainability methods used to explain the output of ML models. *LIME* and *SHAP* work similarly in that they both explain individual predictions (local interpretability), while *SAGE* explains the model's behavior across the entire dataset (global interpretability). All three of these tools use visualizations to try and explain "black-box" models to users. While these methods have been widely adopted by the XAI community [10, 23, 50], given our understanding of AWWs and their limited levels of AI literacy, we felt that the current methods would not resonate with AWWs.

Prior work in XAI and HCI4D advocates for presenting comprehensive evidence but in a concise manner [13, 69, 119, 133]. We therefore simplified the local feature explanation by listing the model features under two explicit categories: "positive" and "negative" (see Figure 2a). The categories represented whether the feature was associated with the positive (normal) or negative (severely malnourished) prediction outcome. We also built an audio feature into the probe that verbally explained the model prediction and feature importance to AWWs with limited reading skills [119].

Initially, the features were accompanied by icons, a green check mark and a red cross symbol for positive and negative features, respectively. We also listed the feature name along with the current



(a) Explorable Explanations Feature Importance Section.

(b) Explorable Explanations Comparison Screens.

Figure 2: (a) The *Feature Importance Section* is part of the AI prediction panel and displays local feature explanations by listing the model features under two explicit categories: "positive" and "negative". Under each category are the corresponding feature names as described in 1b along with their distances from the normal measurement range. For example, the text representing weight says, "The weight is 1.04 kg less than normal." (b) The *Comparison Screens* contain two different children, along with their prediction outcomes, ages, feature contributions, and model confidence measures. The left panel contains the "Current Child", which is three-weeks and four-days-old and is classified as severely malnourished (indicated by the red colored pill-shaped icon). The feature importance sections list height and upper-arm circumference in the normal range (indicated by the green text) and oedema, weight, and head circumference in the severe range (indicated by the red text). The right panel contains the "Similar Child", which is also categorized as "severely malnourished" and is exactly three-weeks-old. However, the "Similar Child" only has upper-arm circumference listed in the normal range (indicated by the green text) and has oedema, weight, head circumference, and height all listed in the severe range (indicated by the red text).

feature measurement or observation. Later, we iterated on the designs since during our pilot study we noticed that the iconography was confusing to AWWs. We thus removed the icons in favor of colored text and bullet points, with green representing positive and red representing negative features. Lastly, we changed the lists to contain the feature names along with their distances from the normal measurement range (e.g., Height is 2 cm less than normal).

3.2.4 Comparison Screens Miller [82] shows that explanations are contrastive. People explain the cause of an event more effectively relative to some other event. Similarly, Cramer et al. [32] show that algorithmic explanations that compare the likeness of similar predictions are the most compelling in justifying predictive systems. Building on these findings, we designed an interface that allowed AWWs to compare the current child to other similarly predicted children (see Figure 2b) and contrast their prediction outcomes, ages, feature contributions, and model confidence. The goal of this feature was to help AWWs build intuition and a relative understanding of nutritional status classifications.

3.3 Study Protocol

We conducted a qualitative study with 30 AWWs in rural regions in Uttar Pradesh, India. To recruit AWWs, we partnered with a grassroots organization focused on community healthcare initiatives. An organization staff member reached out to AWWs who participate in program activities and described the study to them. We then scheduled interviews with AWWs interested in participating in our research. All interviews and observations took place in-person at various locations in the field to make it easier for AWWs to participate. We continued recruiting participants until we reached theoretical saturation in our findings [113].

Our study began with a brief introduction of our research, followed by an informed consent process in which we requested participants' verbal consent. We then proceeded with a semi-structured interview activity, which was interlaced with the design probe that participants interacted with. All study activities were conducted in Hindi and recorded for research purposes. An author fluent in Hindi led the interviews, while another who was not fluent in Hindi took detailed observational notes and photos when participants interacted with the probe. The study protocol consisted of five key steps (Figure 3). First, AWWs assessed the nutritional status of a child (doll). They then engaged with the design probe that predicted the nutrition status of the child (doll) and compared it with their own assessment. We then captured AWWs' AI-related folk theories and sentiments regarding system trust and agreement. Next, AWWs interacted with explorable explanations. We then recaptured AWWs' folk theories and sentiments regarding system trust and agreement. Below we discuss these steps in more detail.

Step 1: Nutritional Assessment (Doll) → **Step 2:** Check AI Prediction → **Step 3:** Capture AI Folk Theories → **Step 4:** Explorable Explanations → **Step 5:** Capture AI Folk Theories

Figure 3: Research study protocol.

3.3.1 Step 1: Malnutrition Assessment via Doll

The interview began by asking demographic and experiential questions with the goal of understanding AWWs' age, educational levels, technological aptitude, and their occupational complexities and experience in diagnosing malnourished children. We then used a physical doll as a proxy for a child with malnutrition. Since using an actual malnourished child would have been irresponsible and dangerous, we used the doll and prototype in tandem to emulate the process of using an AI-driven application to predict the nutritional status of children. To make this process as realistic as possible, we specifically chose a doll (based on its size and physical features) that best represented a child whose nutritional status was ambiguous; we expected AWWs to classify the doll as either normal or malnourished based on their personal assessments. The doll we used was an anatomically correct, lifelike silicone vinyl doll, measuring 19 inches and representing a three-week and four-day-old baby girl (see Figure 4a). In the first step, the AWWs manually inspected the doll and classified its nutritional status based on the protocols they follow in the field. Throughout the study, we used two dolls as visual probes. For the first 10 interviews we used a doll, which elicited mostly malnourished assessment responses from AWWs due to its stature, posture, and facial expression. As a result, for the next 20 interviews, we used a different doll, which received more varied nutritional assessment responses from AWWs.

3.3.2 Step 2: Use Design Probe to Predict Malnutrition

We next asked participants about their familiarity with AI. Based on the AWWs' familiarity (or lack thereof), we provided a brief explanation of AI and some examples of familiar technologies that use AI. This information was in preparation for our next task, in which AWWs used the design probe to predict child malnutrition. The objective was to use the probe as a starting point to help establish a common understanding of what an AI-driven application to predict child malnutrition might look like. We chose to display the prototype on an iPad in contrast to a smartphone because of its larger form factor and subsequent ease of use. Based on AWWs' assessments of malnutrition, we showed them an AI prediction that either agreed or disagreed with their initial assessment with the goal of balancing both groups. Since we could not predetermine AWWs' initial assessments, we kept count of both the assessments made by AWWs and predictions shown. We alternated the predictions as necessary in each sequential interview to balance the two groups. At the end of our study, 16 AWWs saw an AI prediction that was different from their assessment and 14 AWWs saw an AI prediction that was the same as their assessment. We asked the AWWs to interact with the probe and think-aloud [24]. We also provided clarifications and assistance when appropriate. After interacting with the prototype, we asked the AWWs to explain the utility and functionality of the probe in their own words. We also captured AWWs' AI trust and agreement with the AI prediction by asking three structured questions to examine participant changes before and after interacting with the explorable explanations:

(1) Do you agree or disagree with the prediction shown here? Why?
(2) Do you trust this AI's prediction? Why?
(3) What prediction do you want to log-in the system? Your initial prediction or the AI's prediction?

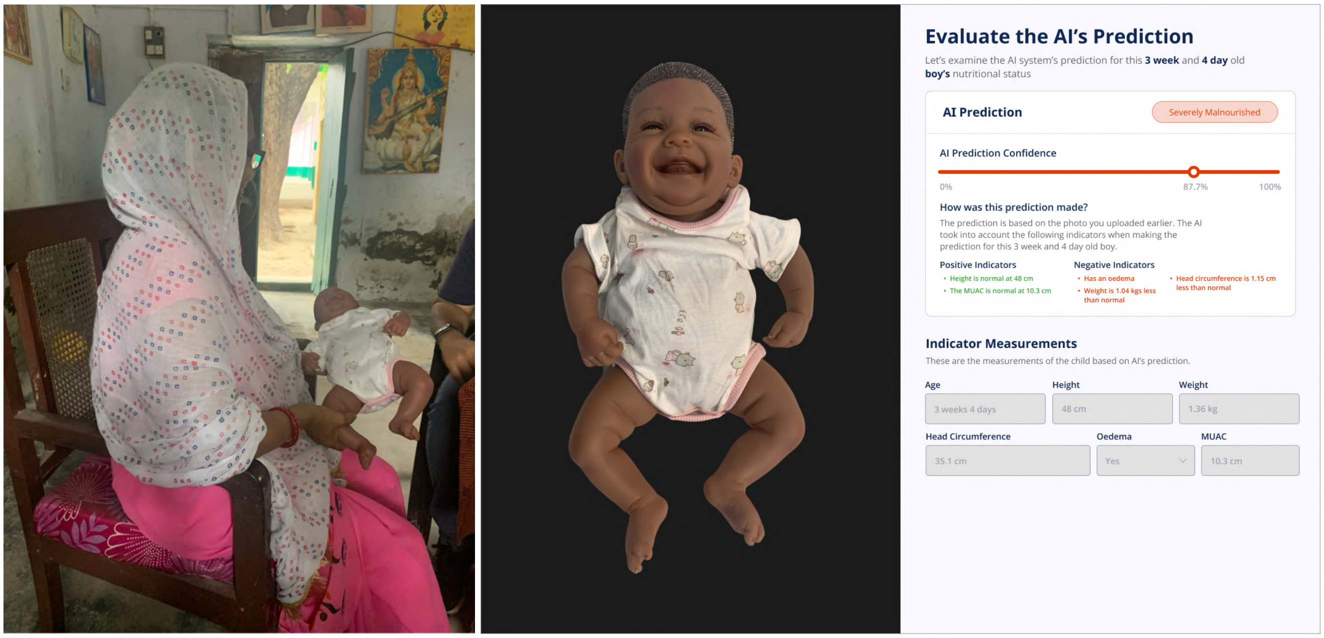
3.3.3 Step 3: Capturing AI Folk Theories

We then conducted activities to capture AWWs' folk theories around how the AI model in the probe predicts malnutrition [36]. We asked participants to imagine they had been selected to train their colleagues using the probe and asked them to verbally communicate how they would explain the system to the colleague. They were asked to include details about how they thought the system makes its predictions based on a photo and what constitutes the system's main components. We also asked them to describe how they would inform their colleagues about integrating the application into their existing workflows and how they were expected to use it. Lastly, we asked them to convey to their colleagues the strengths and potential harms of the system and any changes they would like to make.

The folk theories elicitation activity was a significant departure from what we had preliminarily envisioned. At the beginning, our goal was to capture folk theories using two subjective tasks: a drawing and writing activity as is used in several prior works [36, 60, 120]. However, participants constantly requested to skip these activities in favor of verbally communicating their thoughts due to a few different factors. Some AWWs had limited literacy levels and therefore, found writing particularly difficult. Drawing, on the other hand, was fairly unusual for many participants and subsequently was met with hesitation. For both activities, the concept of "blank page syndrome"—the mental block and anxiety of developing writing or drawing from an empty "base" state—invoked a level of intimidation [102]. AWWs' ingrained deference to authority also played a part in their reluctance to participate. Despite our continued efforts to reassure them that these activities were not a test and that we were not governmental officials monitoring them in any way, many AWWs felt that these activities were aimed at evaluating them and thus were extremely reluctant to engage in tasks that generated artifacts. Before modifying our assessment methods, only one participant perfunctorily completed both the drawing and writing activities. As a result, we decided to capture AWWs' AI folk theories qualitatively through verbal explanations.

3.3.4 Step 4: Engagement with Explorable Explanations

After capturing AWWs' initial folk theories, we presented the explorable explanations prototype, which included a variety of interactive simulations and design heuristics by drawing on established best practices [41, 61, 67, 74, 127]. We nudged the participants to



(a) (b)

Figure 4: (a) A participant evaluating the doll probe in order to make their initial nutritional assessment (Step 1). (b) The AI prediction screen in the prototype. On the left is a photo of the child (doll) and on the right is a textual overview, the AI prediction, model confidence, and feature measurements (Step 2) as was described in Figure 1.

engage with the explorable explanations and asked them to think-aloud and explain the interactions. We also asked AWWs a series of probing questions at key inflection points to elicit responses regarding the visual design elements, their confusion, their expectations of the prototype, their AI understanding, and their trust in the AI prediction. When necessary, we helped participants navigate the probe.

3.3.5 Step 5: Re-capturing AI Folk Theories Lastly, we asked participants several questions about their initial responses to the verbal elicitation prompts, in order to recapture their AI folk theories. For example, we asked participants whether there were any clarifications they would like to make to their initial statements after interacting with the explorable explanations. We also re-asked the same three questions regarding AI trust and agreement (see Section 3.3.2).

3.4 Participant Demographics

Table 1 shows the demographic information for 30 AWWs who participated in our study. All were women who provide maternal and neonatal care in their communities. They ranged in age from 32 to 60 years old, and on average had 20 years of experience working as an AWW. Eight participants had completed high school, 12 had a bachelor's degree, and 10 had a master's degree. All participants owned and used a smartphone, at least for work purposes, but nearly half (47%) expressed the need for some assistance when using them. Furthermore, despite the fact that AI applications are increasingly integrated into AWW workflows, most participants (93%) had no prior knowledge or experience with AI.

3.5 Data Collection and Analysis

We collected approximately 24 hours of audio data, 1,110 photographs of participant interactions, and 75 pages of detailed notes. The audio recordings were translated into English and transcribed. To analyze the data, we used inductive thematic analysis [18]. Two of us engaged in open coding of the transcripts. We first coded one interview together, generating a baseline set of codes. Next, we chose a different interview to code separately and met to organize, reconcile, and merge coding conflicts. We repeated this process for six more documents, at which point the codebook had stabilized. Subsequently, we separately coded the remaining 22 interviews, meeting regularly to discuss code additions, disagreements, ambiguities, and to iteratively refine the codes. Prolonged engagement with the data helped us establish credibility and reduce coding biases. This process resulted in a total of 219 codes. We then performed affinity diagramming to synthesize the codes into three high-level themes that shape the findings.

3.6 Positionality

Our mixed-gender team has four members, three from countries in the Global South. Three authors have a long history of working with CHWs and one author has worked with our partner organization for over a decade. Haraway [47] in her seminal work states that knowledge is shaped from a positional perspective and must be understood as being situated within a particular context. Haraway

Table 1: Demographic details of study participants (N=30).

Demographic attribute	Participant demographics
Age (years)	min: 32, max: 60, avg: 46.5, sd: 7.3
Education	high school: 8, bachelors: 12, masters: 10
AWW experience (years)	min: 12, max: 32, avg: 20.6, sd: 6
Number of children in their care	min: 18, max: 168, avg: 82.4, sd: 35.1
Children diagnosed malnourished (last month)	min: 0, max: 5, avg: 1.8, sd: 1.5
Device ownership	feature phone: 1, smartphone: 30, tablet: 1, computer: 0
Smartphone ownership (years)	min: 0.5, max: 15, avg: 4.8, sd: 3.1
Need tech assistance	no: 16, yes: 14
Prior AI knowledge	no: 28, yes: 2
Apps frequently used	work apps: 30, YouTube: 26, WhatsApp: 24, Amazon: 5, Facebook: 5

[47] calls for recognizing the influence of power dynamics as essential to produce more inclusive and socially relevant knowledge. Although we rely heavily on the cultural and language proficiency of our team and the partner organization, our gender, education, socioeconomic status, and urbanity placed us in an uneven power dynamic with our participants who were all low-income women working within a patriarchal system in rural India. Two male researchers (one Hindi-speaking Indian and one Latinx American) conducted the fieldwork. Given the gender differences between the researchers and participants, we tried to alleviate AWWs' hesitations by conducting our research in a room with the doors open and in the presence of a female staff member who had a longstanding relationship with the participants. However, some AWWs still showed mild signs of discomfort and took additional time to acclimate and speak openly. Due to the authors' different lived realities and power differentials, our work is only able to partially capture the perspectives of AWWs [47].

4 Findings

In this section, we illustrate how our analysis revealed that **explorable explanations shifted AWWs' AI-related folk theories and helped them develop a more nuanced understanding of AI** . In Section 4.1, we discuss how AWWs' assumptions about AI mimicking their processes changed and how explorable explanations improved their understanding of the inner workings of AI. In Section 4.2, we show how explorable explanations enabled AWWs to contest AI predictions and increased their skepticism in AI predictions, but struggled to improve AWWs' understanding of ML model confidence. In Section 4.3, we discuss AWWs' desire to use the explorable explanations for improved learning outcomes and improving their understanding of borderline cases of malnutrition.

4.1 Impact of Explorable Explanations on Shaping AI-Related Folk Theories

4.1.1 AI Works Differently Than AWWs Initially, after observing the baseline prototype, AWWs had a hard time expressing their thoughts about how the AI-driven system worked. Participants generally stated that the AI was estimating the child's measurements by looking at the severity levels of malnutrition but they were unsure how exactly the system was operating. AWWs who did try to articulate their thoughts regarding the baseline prototype tended to describe folk theories that were projections of their own protocols and procedures, often anthropomorphizing the AI in the process. For example, P1 thought that the AI must have accompanied another AWW going house-to-house taking analog measurements and making assessments. Others, such as P26, made analogies:

> *"Like our madam (referring to their supervisor who is an auxiliary nurse midwife) can just look at a baby and tell, it's also predicting by looking at the baby. We, however, only say anything after we have measured the baby. I think it's using the same formula as us."*

After interacting with the explorable explanations, AWWs' AI-related folk theories and understanding shifted. Although most participants still had trouble articulating the mechanisms behind the AI prediction, many recognized that the AI's processes were different from their own. For example, P29 said *"I don't completely understand how it is doing these classifications. I think it does things a bit differently from what we generally do"*. However, for some AWWs certain explorable explanations helped them develop nuanced understandings about the AI. A concise, pared-down interface with interactive explanations incorporated into both the *Edit Measurements Section* (Section 3.2.2) and *Feature Importance Section* (Section 3.2.3) allowed AWWs to "play" with the prototype and test their expectations of the AI's behavior against its actual behavior. For example, P18 was able to understand how the positive and negative indicators displayed in the *Feature Importance Section* contribute to the overall AI prediction. After interacting with the *Edit Measurements Section* and toggling back-and-forth the child's measurements to witness the changing predictions, P18 said:

> *"The AI is not just looking at weight or height in isolation and making a prediction. It is seeing if a couple of indicators are in the red (contributing towards a malnourished prediction) and then classifying them [the*

> *child]. It's not what we do but I think that's how this app works."*

Many AWWs expressed similar sentiments to P18. In addition, the *Feature Information Modals* (Figure 1a) communicated in simple, colloquial terms to participants how and why feature contributions differed (or did not) from the overall prediction. These expandable elements supported a more granular understanding of the system functionality and helped AWWs internalize the differences between their and the system's processes. As, P29 stated:

> *"It helped me go back and forth between the weight and mid-upper arm circumference (MUAC). Since I'd noticed that the weight was a bit lower than expected for a 3 week old baby, I could reason why the MUAC was smaller despite height-weight being in the normal ranges. The charts were also helpful since I use them regularly."*

4.1.2 Improved Understanding of the "AI Prediction"

The explorable explanations changed AWWs understanding of what constitutes an AI "prediction" and shifted their perceptions that were built on pre-existing knowledge of other digital technologies. We found that before interacting with the explorable explanations, many AWWs struggled to understand that the AI system is making a prediction, which could be wrong, instead of giving a definite diagnosis. After interacting with the explorable explanations, many participants were able to recognize the conceptual difference between an algorithmic prediction and analog measurements. This shift resulted from editing the measurements, during which the participants observed changes in the malnutrition prediction. The *Edit Measurements Section* interface enabled a certain amount of data fluidity, which challenged AWWs to think critically about what these interface changes reflected and how the system could be working. These data mutations meant that participants tacitly understood that the system prediction had to be an estimation rather than a ground truth, since the outcomes dynamically updated based on the underlying data changes. Some participants such as P26 were able to intellectualize the concept of a prediction more illustratively and spoke to their perceived limitations of the tool's abilities saying:

> *"...this app is also measuring and giving us the prediction. How can it say anything about the child without measuring it? It must have gone to various villages and measured various babies and come up with its own charts. No one can come up with predictions without doing any practicals. No one is so intelligent to just look at something and give correct predictions."*

Participants also witnessed unexpected predictive classifications when manipulating the measurement values. Because the explorable explanations did not follow rigid classification protocols that participants were familiar with, AWWs began to intuit that the system must be predicting classifications based on past data. For example, participants initially expected that the AI would make classifications based on the strict guidelines they use for weight and height, but when editing the measurements (*Edit Measurements Section*) they noticed this was not always the case—sometimes the AI system behaved differently than their expectations. The *Comparison Screens* also displayed different measurements and classifications per child, which added to AWWs' understanding that the system was not following one specific protocol and thus it must be making informed guesses using past malnutrition data. As, P18 said:

> *"Like I'd said earlier, the AI is not just focusing on one indicator and giving a result. It's looking at multiple things. The first similar child just had abnormal height according to the AI but the second child has oedema and is abnormal by weight. So it's classified as medium. If there were more indicators in red (a feature that contributes to the malnutrition prediction), it would have been classified into severely malnourished."*

4.2 Impact of Explorable Explanations on AI Contestability, Trust, and Confidence

4.2.1 Enhanced Contestability

Explorable explanations not only enabled AWWs to build intuitions about the AI prediction but also enhanced their ability to contest AI predictions when they perceived that it made mistakes. When AI prediction differed from AWWs' assessment of malnutrition (via inspecting the doll), some AWWs changed their assessment to match the AI prediction, either providing limited justification as to why or expressing a bias towards the AI system. AWWs who showed these biases would often liken the AI probe to a medical device such as an X-ray machine and assumed that the AI system has to be *"right."* Additionally, AWWs expressed optimism about technological advancements as a justification for their high levels of trust in them. In such cases, the AWWs provided fairly coarse explanations of how the AI prediction came to be and why it was more accurate.

After interacting with the explorable explanations, we witnessed a shift in AWWs' explanations and attitudes. In particular, they were able to communicate more granular and precise information about the underlying functionality of the predictive system and contest AI predictions when they became aware of the specific attributes the AI model took into account to arrive at a prediction. Certain design heuristics, such as the *Edit Measurements Section*, elicited more detailed and accurate responses regarding the prototype's functionality. For example, P26 initially agreed with the AI prediction. However, after interacting with the explorable explanations and witnessing the prediction classifications change when modifying the height and weight measurements in the *Edit Measurements Section*, she said:

> *"This should have at least been in medium malnutrition because there are babies who can be malnourished despite having normal weight and height. If this child still has oedema in its feet, then there is some form of malnutrition. This child is probably anemic."*

Other participants made more pronounced corrections after interacting with the explorable explanations. For example, P24 initially blamed herself for the incongruity between her assessment and the AI's prediction, stating: *"I think there is a difference in my and AI's assessment but I'm getting old now. How can a computer be wrong? It's more likely that I am wrong."* However, after interacting with the explorable explanations, her assumptions began to change. For example, she found it odd that the AI model classified the three week old baby as "normal" at 48 cm, as although newborns are typically born at 48 cm or longer, they tend to grow by a few more centimeters in the following weeks. To inspect this further, she increased the child's (doll) height from 48 centimeters to 50 centimeters and noticed that the prediction was still "normal", as she had expected. However, her skepticism grew when the prediction persisted as "normal" even after increasing the height further to 52 centimeters, a point where she expected a classification of "medium malnourishment". She became skeptical of AI's veracity and at one point, mentioned that she felt the prediction was not accurate given the updated measurement. When interacting with the *Feature Information Modals*, she felt that the weight predicted by AI should have been higher given the child's age. Later in the interview, she dismissed the AI's prediction stating:

> *"This AI thing just predicts and does not actually measure the baby. Although the height and weight [informational] modals indicate a normal baby, the weight should have been more than 2.4 kilograms, unlike what the AI shows, since the child grows rapidly in the first few weeks."*

4.2.2 Increased AI Skepticism We found that when AWWs' assessments differed from the AI prediction, many continued to believe in their own judgments. This was primarily a result of AWWs' skepticism in the probe's ability to make accurate predictions based solely on a photograph. AWWs expressed that the prototype was not holistically evaluating the child and that it was lacking an empirical analysis. These perceptions were reinforced when engagement with explorable explanations improved AWWs' understanding of the inner workings of the AI model. P18 elaborated:

> *"I'd go with what I see in reality. Pictures lose details and can't observe everything we can with our own eyes. Imagine I have some form of disability in the legs, my saree will completely hide that in the photograph. So we need to observe how a child is walking or whether it's playing. If the tool is just looking at the photo then it can't capture a child's behavior."*

After engaging with explorable explanations, some AWWs, such as P25, reasoned that the application was focusing only on certain features such as weight and height. She emphasized that the AI needed to take into account other factors, such as behavioral indicators, diet, and the child's healthcare history to determine the nutritional status. While they saw the utility of the application, given the high-stakes nature of the task, AWWs such as P25 did not want to *"blindly believe"* the AI system and instead preferred to confirm the child's nutritional status via analog methods.

We also found that participants' previous experiences with other digital technologies both in and outside of work seemed to have ingrained the idea that technologies and by proxy AI were not omniscient and thus could make mistakes. For example, AWWs referenced the *Poshan Calculator* mobile application [22] they currently use, which allows AWWs to input the weight and height of a child and receive the child's corresponding nutritional status. P20 made connections between the prototype and *Poshan Calculator* and said:

> *"But is it necessary [true] that the app will always be right? For example, in my experience, the Poshan Calculator app sometimes gives results that do not match the growth charts or our own assessment on looking at the baby."*

These findings align with the work of Schmidt et al. [114], which shows that providing detailed explanations can affect users' trust in AI, either because users find the explanations confusing or because the explanations have developed capabilities to critically contest the system. While our findings indicate that explorable explanations did help AWWs develop more precise contestability, some AWWs might be expressing skepticism because of their confusion about inner workings of the AI system as opposed to building a deeper AI understanding.

4.2.3 Impact on the Understanding of AI Model Confidence While some AWWs loosely understood the concept of ML model confidence, most were confused by what the designs were trying to communicate. The explorable explanations had little impact on improving AWWs' understanding of confidence. Rather, we observed that the confidence slider seemed to complicate their understanding of AI. In general, participants understood that changing the child's measurements could have an effect on model confidence but were unable to grasp what exactly the confidence slider element represented.

AWWs frequently misinterpreted confidence as representing the level of severity of malnutrition. For example, participants would alter data in the *Edit Measurements Section* and expect the confidence slider to change in accordance with their perceptions of the severity levels of child malnutrition. When the prediction did not reflect their expectation, AWWs expressed confusion. On occasion, the opposite occurred, and the confidence slider would move in a direction that AWWs expected. This unfortunately reinforced AWWs' misunderstanding until they played more with the data and noticed that model confidence was not behaving as they expected. Occasionally, a few AWWs seemed to understand the concept of the model confidence slider. P20 mentioned:

> *"Since it's taking a guess, maybe the percentage tells us how correct their guess is? But I don't understand how one can tell the correctness of one's guess without knowing the answer."*

The difficulties in understanding model confidence stemmed from AWWs' experiences with other digital technologies. For example, in the *Poshan Calculator* application , nutritional status information is communicated using a numerical measure to indicate severity. As a result, when AWWs saw numerical values that describe confidence, they assumed that the values represent the severity of malnutrition. We found it difficult to shift AWWs' inferences away from these anchoring biases [90]. These findings align with the work of Dogruel [37], who showed that people tend to form analogies based on similar domains to explain unfamiliar phenomena.

4.3 Explorable Explanations Are Learning and Augmentation Tools

4.3.1 Augmenting AWWs' Domain Learning AWWs viewed the explorable explanations as an educational tool that could help them learn more about child malnutrition. The interactive and didactic components of the probe nurtured AWWs' curiosity. For example, many AWWs toggled back-and-forth, interacting with the visual affordances in an intuition grounding process, which helped build understanding about the AI model features. AWWs felt that learning about the contributing features and their measurement ranges was knowledge that they could apply when in the field. This added information could help them better understand the indicators to look for in a child and the compounding factors that could lead to child malnutrition. For example, P28 elaborated:

> *"If the app comes with these buttons then they would be very useful not just to understand the predictions but also these charts will help us to know the right ranges for all parts of the body. These charts will help us better understand the predictions as we get to know the right ranges."*

The value of explorable explanations was especially noticeable when participants interacted with the *Feature Information Modals* for measurements such as head circumference, which was a metric they currently do not evaluate under their existing protocol. AWWs expressed how they thought these explorable explanations could provide them with additional contextual information that would help them perform more holistic and comprehensive assessments in the field. Many AWWs felt that AI could play a much larger role in cases that are complicated and are close to different classification boundaries of malnutrition. P28 elaborated:

> *"A lot of AWWs do not know what to do at border measurements. I think this section can tell us where a baby should be classified if the measurements are on the border of two categories."*

4.3.2 Augmenting AWWs' Capabilities AWWs felt that an AI tool accompanied with explorable explanations could help them more effectively diagnose child malnutrition. In particular, AWWs viewed the explorable explanations as a "sounding board", which could help them debug complicated child assessments and provide additional data points. For instance, P15 thought that the prototype could be best utilized as an additional verification step:

> *"It is possible that I could be wrong in taking measurements. For example, it can happen when a baby is not laid down correctly on the infantometer or a part of its body is away from the machine. Then the AI can keep me in check."*

In addition to serving as a sounding board, P20 mentioned that because explorable explanations displayed all relevant information on a single screen, it would save her time and help her avoid going through multiple physical registers to collate information. P20 also felt that the explanations could act as an effective presentation tool in meetings to help describe to others how the AWWs classified malnutrition. Other AWWs viewed the explorable explanations as a valuable communication tool for interfacing with parents. They felt that the explanations would allow AWWs to more concretely and precisely demonstrate areas of improvement to parents regarding their children's nutrition and help them make specific, action-oriented recommendations. Lastly, given the space and time constraints within which AWWs work, some AWWs felt that AI could prove especially useful in automating laborious and rote tasks and that explanations could help AWWs to keep AI in check in case these systems make mistakes.

Despite the potential of augmenting AWWs' capabilities, many AWWs continued to look at AI with skepticism due to engagement with explorable explanations, which demystified AI for them. They also highlighted how the sociopolitical complexities of their workplaces eventually determines whether and how such tools and explanations are integrated into their workflows. Given their precarious and extremely hierarchical work environments, some participants felt that they would have to suppress their critical opinions on AI in favor of "following orders" if the prototype was integrated into their workflows.

5 Discussion

Given the proliferation of AI applications and organizations' desire to leverage these predictive tools to tackle the world's most pressing societal issues (such as child malnutrition), it is critical to understand how to explain AI technologies to key stakeholders. We investigated whether explorable explanations—human-centered, interactive visual heuristics and written explanations—helped improve XAI methods for non-AI experts with low digital literacy such as AWWs. In this section, we reflect on our findings and discuss implications and considerations for future work.

5.1 Re-imagining XAI as Play

Current AI technologies are largely designed by a few people in the West who decide what values these tools will inherit and where they will be used. AWWs are often viewed as low-skilled laborers by AI interventionists who are generally unaware of the enormous amount of invisible labor done by frontline health workers [83, 126]. AI designers and developers frequently do not take into account stakeholders' values and viewpoints such as AWWs, who not only lack the power and legitimacy to influence the AI system's design but are also forced to integrate these tools into their workflows [81]. Such interventions reduce AWWs from being domain experts to data collectors, diminish their agency, and make their labor more scrutinized and measured by narrow success metrics.

Chirumamilla and Pal [26] warn against "developmental optic"—a narrative in ICTD specifically and Global Development more broadly, which envisions the primary audience such as AWWs, as perpetually "backward" and in need of "improvement". The developmental optic not only undermines the voice of communities in the creation and appropriation of initiatives, it also forces success to be measured through a static set of narrow goals that focus on improving "productivity and accuracy" rather than "agency". Chirumamilla and Pal [26] advocate for a re-imagination of research to center "non-productive" activities, such as having fun, to counter developmental narratives.

We see parallels between our research and the contentions outlined above. Currently, AWWs, who are generally unfamiliar with

AI-driven technologies, often experience these technological interventions imposed upon them without consultation and without avenues for stakeholder engagement [26, 81]. Moreover, there is little work being done to make AI understandable to novice users like AWWs, who are nevertheless expected to operate AI technologies safely within the world's poorest and most marginalized communities [92, 93]. In our work, we saw how explorable explanations enabled AWWs to organically explore the inner workings of AI and thereby improve their AI understanding. In particular, we observed how the notion of "play", a critical component of the explorable explanations, helped AWWs better articulate their thoughts regarding the AI model's functionality. For example, we saw how toggling measurements in the *Edit Measurements Section* allowed AWWs to reflect on the predictive changes in real-time and helped them build intuitions about AI based on these playful experiences.

We also found that participants expressed epistemic curiosity about the AI system [72, 86]. Epistemic curiosity is the *"desire for knowledge that motivates individuals to learn new ideas, eliminate information gaps, and solve intellectual problems"* [72]. As noted in Section 4.3, AWWs viewed the explorable explanations as a learning heuristic and tool to augment their existing workflows. The prototype's interactivity and the didactics seemed to nurture participants' intellectual curiosity. They encouraged a more active form of learning, which not only empowered them but also promoted a sense of agency and control amongst AWWs. We observed that because the data in the explorable explanations could be mutated, participants subsequently viewed the tool in a less prescriptive manner. Through engagement with essential design elements of explorable explanations, AWWs were also able to substantiate the relationships between the child's photo, measurements, and the prototype's malnutrition classification.

While more work is necessary to state whether explorable explanations can help participants appropriately *calibrate trust* in AI-driven systems [136], similar observations regarding deliberative thinking have been made by Buçinca et al. [21]. In particular, Buçinca et al. [21] found that *cognitive forcing functions*, which are design elements and interventions that elicit thinking at the decision making time have been shown to reduce overreliance on AI. Similar to cognitive forcing functions, explorable explanations also add friction to the AI prediction process and require AWWs to actively engage with the prediction explanation — engaging the System 2 (slow and deliberative) thinking process [58]. As a result, explorable explanations extend the findings of Buçinca et al. [21], which state that *"explainable AI researchers should ensure that people will exert effort to attend to those explanations"*.

The design elements that best resonated with AWWs were the *Feature Information Modals* (Figure 1a), *Feature Importance Section* (Figure 2a), *Edit Measurements Section* (Figure 1b), and *Comparison Screens* (Figure 2b). In general, we found these designs were successful because they surfaced information that was user-centered, appropriate, contextual, and concise. For example, while the design probe introduced new concepts to AWWs, many of the interactions, iconography, and colors referenced in explorable explanations were derived from best practices in designing UIs for digitally novice users [119] and contained existing design language from mobile applications and analog artifacts that were familiar to participants. For example, we designed measurement charts to accompany the *Feature Information Modals* (Figure 1a), which referenced designs from the current child weight and height charts used by AWWs. As a result, adopting a culturally responsive design enabled AWWs to absorb and reflect on new information as well as increased their familiarly and comfort when interacting with explorable explanations.

Given these findings, we advocate for the aforementioned design considerations to be incorporated into new XAI research focusing on users, such as AWWs, who possess little to no AI knowledge and low levels of digital skills. We underscore the importance of developing XAI methods that focus on building approachable and culturally responsive designs, recognize and celebrate participants' domain expertise, and in general advocate for methods that encourage and facilitate stakeholder curiosity.

5.2 Explainability is Only A Part of the Puzzle

Our findings also show that AI explainability is only a part of the puzzle and several other factors determine the integration of AI technologies into frontline healthcare workflows.

5.2.1 Power Differentials As discussed in 4.3.2, several AWWs felt forced to use AI technologies by their supervisors. AWWs lacked agency and negotiating power to dictate in what areas and how these technologies can support their needs. AWWs work under precarious conditions in environments with limited resources and opportunities [62, 84, 112]. The little compensation that AWWs receive is tied to their work performance, which remains under constant surveillance by the Indian government [109]. Integrating AI technologies into an environment already under high surveillance and that promotes outcome-based remuneration will only further reinforce or exacerbate these existing power imbalances [62, 109]. This is especially true when the AI technologies do not account for the needs of AWWs who are expected to become AI workers, and when AWWs do not have the necessary knowledge to understand the inner workings of AI. We clearly witnessed these power dynamics at play in our study. For example, many AWWs initially expressed a reluctance to outwardly object to the AI's decision, despite internally disagreeing with the prediction, until they interacted with explorable explanations which gave them the language to express hesitations and contest AI decisions.

Researchers should consider the domain contexts in which XAI systems are designed to operate and the stakeholders they are expected to support. Although the explorable explanations improved AWWs' AI understanding and enhanced AI contestability, we strongly believe that, given the current sociopolitical environment, the prototype would be severely limited in its efficacy as a result of the power disparities [109]. Any deployments of AI technologies need to take multi-pronged and holistic approaches to developing responsible technologies. XAI is only one piece of the puzzle and not a panacea.

5.2.2 Infrastructural Challenges AWWs are one of the most accessible arms of the government in rural India. Subsequently, they often receive tasks beyond their expected duties [109]. One of these tasks has been the increasing data work from new digital technologies that are integrated into their workflows to help them diagnose diseases and manage patient care. Many of these applications exist in silos, which exacerbate things further and result in a fractured and disparate system of work. For example, AWWs in our study not only had to use new digital applications such as the *Poshan Calculator* to identify child malnutrition but often had to keep manual records in parallel, doubling their workloads.

AWWs frequently contend with infrastructural challenges, such as poor Internet connectivity and power cuts, which reduced their willingness to use digital technologies for work. AWWs also highlighted the lack of compensation and their invisible labor that goes unrecognized by administrators. Some AWWs ended up paying for occupation-related expenses out-of-pocket, including cellular phones that are required for their jobs. Many AWWs underscored the fact that, in addition to having to pay for these work-related devices, they also received little to no technological assistance or support. Instead, they relied on family members to help with tech-related hurdles. These findings suggest that integrating a new application into their workflows, even if it makes AI understandable, may only add to their troubles. As other scholars have observed, these types of data work become ends unto themselves and as a result, often flatten the human work of care into a few data points [96, 109, 126].

5.2.3 Lack of Training and On-the-Ground Support One consistent theme that we observed was the requests from AWWs for training on the prototype. While explorable explanations shifted their AI-related folk theories and improved their AI understanding, many AWWs continued to express a strong need for training to improve their AI literacy. AWWs also requested that a support infrastructure be provided so that they could have resources for technological assistance and/or maintenance. Without developing these local capabilities, AWWs expressed that an AI intervention is doomed to fail, regardless of how explainable it is.

In addition to training and scaffolding structures, AWWs also emphasized the need for AI technologies to be accepted by local communities. They highlighted the importance of engaging with and educating guardians and the community at large about the prototype. Given the years that AWWs have spent building community trust, they felt that AI technologies in frontline healthcare must be robust and understandable to communities in order to prevent the erosion of trust due to AI malfunctions. They called for awareness programs, posters, and other analog artifacts to start a conversation and increase understanding of AI tools amongst the communities they serve. Taken together, these findings highlight: (1) the importance of a more holistic conception of AI explainability that goes beyond cognition and literacy to consider aspects such as user training and community engagement as core tenants of explainability efforts and, (2) the need to consider AI technologies as "sociotechnical" tools that are cognizant of the sociocultural, sociopolitical, and socioeconomic realities of the context within which they are integrated.

5.3 Limitations and Future Work

Our work has some limitations. First, while Figma is an excellent design tool for interactive prototyping, the probe that we built supported only limited interactions. Future work should explore the building of a feature-rich, fully functional prototype that would allow AWWs to use the application in a more natural way that better mimics their daily processes. Furthermore, having a fully interactive prototype may encourage greater exploration and "play", which we found to be critical in improving AI understanding for AWWs.

Next, given the qualitative nature of this research our findings lack generalizability and more work is needed to evaluate quantitative results in support of our findings. In addition to building a fully functional prototype with explorable explanations, future research efforts should focus on conducting controlled experiments to examine the effectiveness of explorable explanations as well as the specific features of explorable explanations that increase user agency, help calibrate trust, and improve AI understanding. One such example could involve conducting a controlled experiment where one group of AWWs would receive the AI-driven malnutrition application *without* explorable explanations (control) and another group would receive the application *with* explorable explanations (experiment). We would subsequently evaluate both groups regarding empirical task performance as well as on other dimensions such as calibrate trust, system understanding, and system confusion, to name a few.

During our study, we consistently observed that AWWs wanted more time to interact with the design probe. A remote, longitudinal study administered in collaboration with a local partner might result in findings that are less influenced by observer effects and participant response bias [125]. Furthermore, a long-term deployment of explorable explanations and controlled evaluation of an AI-driven application that is deployed ethically, responsibly, and safely might also address the limitations aforementioned in Section 3.6 regarding the researcher-participant gender and power differences. Research participants could use the prototype in their daily tasks and the research team could capture analytics and other inputs from the device. This longitudinal study would also necessitate building a mobile prototype, which would allow us to test our designs on a smartphone (a device AWWs are more familiar with) as opposed to an iPad. These studies, in concert with our qualitative findings, would generate rich data on the potential utility of explorable explanations in improving AI understanding for digitally novice AI workers such as AWWs in the Global South.

Finally, future work should also explore the use of participatory AI methods [16]. These methods would enable AWWs to actively engage in the design process and articulate their needs and preferences regarding XAI, in order to safely operate these tools, which are being increasingly integrated into their workflows. Not only would this approach help design culturally appropriate interventions that enhance the overall utility of AI for AWWs, it would also enable XAI designers and developers to better understand how sociopolitical, socioeconomic, and sociocultural forces impact AI adoption and understanding.

6 Conclusion

This work examines the effectiveness of explorable explanations at enhancing the AI understanding of AWWs in rural India. We conducted interviews with AWWs who engaged with a design probe to predict child malnutrition. Through the probe, which incorporated AI predictions accompanied by explorable explanations, we found strong evidence that explorable explanations shape AI-related folk theories of AWWs and improve their understanding of AI. While explorable explanations increased their skepticism in AI and made AWWs more agentic in contesting the AI prediction, they faced challenges grasping the design elements and meaning behind AI confidence. AWWs engaged in "play" when interacting with explorable explanations and perceived the explanations as having high utility in improving learning outcomes and augmenting their capabilities. Finally, we discussed the need for a more holistic conception of AI explainability that goes beyond cognition and literacy, and takes into account power differentials, infrastructural challenges, and the scaffolding of structures to aid in integration and adoption of AI technologies into frontline healthcare workflows.

Acknowledgments

We thank our participants who graciously shared their insights, experiences, and made this work possible. Also a special thank you to our external partner organization (NYST) for their critical assistance and to the Cornell Center for Social Science for funding this work.

References

[1] 2023. Child Growth Monitor: Using AI to solve world hunger and malnutrition. https://news.microsoft.com/en-in/features/child-growth-monitor-malnutrition-india-microsoft-ai/
[2] 2023. Integrated Child Development Services - ICDS Scheme | Ministry of Women & Child Development|IN|tydnz — wcd.nic.in. https://wcd.nic.in/integrated-child-development-services-icds-scheme. [Accessed 06-12-2023].
[3] 2023. Malnutrition. https://www.who.int/health-topics/malnutrition
[4] 2023. The WHO Child Growth Standards. https://www.who.int/tools/child-growth-standards/standards
[5] NITI Aayog. 2018. National Strategy for Artificial Intelligence.
[6] NITI Aayog. 2021. Principles of Responsible AI.
[7] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. 2018. Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–18.
[8] A Ali, AA Raza, and IA Qazi. 2023. Validated digital literacy measures for populations with low levels of internet experiences. Development Engineering, 8, 100107.
[9] 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.
[10] Plamen P Angelov, Eduardo A Soares, Richard Jiang, Nicholas I Arnold, and Peter M Atkinson. 2021. Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 11, 5 (2021), e1424.
[11] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* 58 (2020), 82–115.
[12] Andrew Bell, Oded Nov, and Julia Stoyanovich. 2023. Think about the stakeholders first! Toward an algorithmic transparency playbook for regulatory compliance. *Data & Policy* 5 (2023), e12.
[13] Andrew Bell, Ian Solano-Kamaiko, Oded Nov, and Julia Stoyanovich. 2022. It's just not that simple: an empirical study of the accuracy-explainability trade-off in machine learning for public policy. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 248–266.
[14] 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.
[15] Ayushi Bhotica. 2021. Building Our Anthropometry Solution, One Field Visit At a Time. https://www.wadhwaniai.org/2021/02/building-our-anthropometry-solution-one-field-visit-at-a-time/
[16] Abeba Birhane, William Samuel Isaac, Vinodkumar Prabhakaran, Mark Díaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Frameworks and Challenges to Participatory AI. https://arxiv.org/pdf/2209.07572.pdf
[17] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1077–1086.
[18] G Braun. 1977. Using thematic analysis in psychology. *Using thematic analysis in psychology* 3, 10 (Oct. 1977), 77–83.
[19] Bill Briggs. 2019. How technology can speed help to malnourished children in the effort to end global hunger. https://news.microsoft.com/transform/app-end-global-hunger/
[20] Peter Brusilovsky. 1994. Explanatory visualization in an educational programming environment: connecting examples with general knowledge. In *Human-Computer Interaction: 4th International Conference, EWHCI'94 St. Petersburg, Russia August 2–5, 1994 Selected Papers 4*. Springer, 202–212.
[21] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.
[22] Poshan Calculator. 2023. https://www.poshantracker.in/PTCalculator [Accessed 09-14-2023].
[23] Ahmad Chaddad, Jihao Peng, Jian Xu, and Ahmed Bouridane. 2023. Survey of explainable AI techniques in healthcare. *Sensors* 23, 2 (2023), 634.
[24] Elizabeth Charters. 2003. The Use of Think-aloud Methods in Qualitative Research An Introduction to Think-aloud Methods. *Brock Education Journal* 12, 2 (July 2003).
[25] Irene Y Chen, Shalmali Joshi, and Marzyeh Ghassemi. 2020. Treating health disparities with artificial intelligence. *Nature medicine* 26, 1 (2020), 16–17.
[26] Padma Chirumamilla and Joyojeet Pal. 2013. Play and Power: A Ludic Design Proposal for ICTD. In *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers - Volume 1* (Cape Town, South Africa) *(ICTD '13)*. Association for Computing Machinery, New York, NY, USA, 25–33. https://doi.org/10.1145/2516604.2516628
[27] Neil Chulpongsatorn, Mille Skovhus Lunding, Nishan Soni, and Ryo Suzuki. 2023. Augmented Math: Authoring AR-Based Explorable Explanations by Augmenting Static Math Textbooks. *arXiv preprint arXiv:2307.16112* (2023).
[28] Comptroller and Auditor General of India (CAG). 2012–2013. Performance Audit of Integrated Child Development Services(ICDS) Scheme. https://cag.gov.in/uploads/download_audit_report/2012/Union_Performance_Civil_ICDS_Scheme%20_Ministry_Women_Child%20Development_22_2012.pdf
[29] Matthew Conlen and Jeffrey Heer. 2022. Fidyll: A Compiler for Cross-Format Data Stories & Explorable Explanations. *arXiv preprint arXiv:2205.09858* (2022).
[30] Matthew Conlen, Megan Vo, Alan Tan, and Jeffrey Heer. 2021. Idyll studio: A structured editor for authoring interactive & data-driven articles. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 1–12.
[31] Ian Covert, Scott M Lundberg, and Su-In Lee. 2020. Understanding global feature contributions with additive importance measures. *Advances in Neural Information Processing Systems* 33 (2020), 17212–17223.
[32] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-adapted interaction* 18 (2008), 455–496.
[33] Nicola Dell, Jessica Crawford, Nathan Breit, Timóteo Chaluco, Aida Coelho, Joseph McCord, and Gaetano Borriello. 2013. Integrating ODK Scan into the community health worker supply chain in Mozambique. In *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers-Volume 1*. 228–237.
[34] Brian DeRenzi, Nicola Dell, Jeremy Wacksman, Scott Lee, and Neal Lesh. 2017. Supporting community health workers in India through voice-and web-based feedback. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 2770–2781.
[35] Brian DeRenzi, Neal Lesh, Tapan Parikh, Clayton Sims, Werner Maokla, Mwajuma Chemba, Yuna Hamisi, David S hellenberg, Marc Mitchell, and Gaetano Borriello. 2008. E-IMCI: Improving pediatric health care in low-income countries. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 753–762.
[36] Michael Ann DeVito. 2021. Adaptive Folk Theorization as a Path to Algorithmic Literacy on Changing Platforms. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 1–38.
[37] Leyla Dogruel. 2021. Folk theories of algorithmic operations during Internet use: A mixed methods study. *The Information Society* 37, 5 (2021), 287–298.
[38] 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* (Copenhagen, Denmark). Springer-Verlag, Berlin, Heidelberg, 449–466. https://doi.org/10.1007/978-3-030-60117-1_33
[39] Upol Ehsan, Philipp Wintersberger, Q Vera Liao, Martina Mara, Marc Streit, Sandra Wachter, Andreas Riener, and Mark O Riedl. 2021. Operationalizing human-centered perspectives in explainable AI. In *Extended Abstracts of the*

2021 CHI Conference on Human Factors in Computing Systems. 1–6.

[40] Figma. 2023. https://www.figma.com/. [Accessed 09-14-2023].

[41] Jesper Hyldahl Fogh. 2018. Explorable Explanations: What are they? What do they explain? How do we work with them? Let's find out.

[42] U.S.Agency for International Development. 2019. Artificial Intelligence in Global Health: Defining a Collective Path Forward. https://www.usaid.gov/cii/ai-inglobal-

[43] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: cultural probes. *interactions* 6, 1 (1999), 21–29.

[44] Katy Ilonka Gero, Zahra Ashktorab, Casey Dugan, Qian Pan, James Johnson, Werner Geyer, Maria Ruiz, Sarah Miller, David R Millen, Murray Campbell, Sadhana Kumaravel, and Wei Zhang. 2020. Mental Models of AI Agents in a Cooperative Game Setting. 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, 1–12.

[45] Marzyeh Ghassemi, Luke Oakden-Rayner, and Andrew L Beam. 2021. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health* 3, 11 (2021), e745–e750.

[46] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI—Explainable artificial intelligence. *Science robotics* 4, 37 (2019), eaay7120.

[47] Donna Haraway. 2020. Situated knowledges: The science question in feminism and the privilege of partial perspective. In *Feminist theory reader*. Routledge, 303–310.

[48] Fred Hohman, Andrew Head, Rich Caruana, Robert DeLine, and Steven M Drucker. 2019. Gamut: A design probe to understand how data scientists understand machine learning models. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.

[49] Fred Hohman, Minsuk Kahng, Robert Pienta, and Duen Horng Chau. 2018. Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE transactions on visualization and computer graphics* 25, 8 (2018), 2674–2693.

[50] Andreas Holzinger, Anna Saranti, Christoph Molnar, Przemyslaw Biecek, and Wojciech Samek. 2020. Explainable AI methods-a brief overview. In *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*. Springer, 13–38.

[51] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, et al. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 17–24.

[52] IIPS and MacroInternational. 2019–21. National Family Health Survey (NFHS—5). http://rchiips.org/nfhs/NFHS-5Reports/NFHS-5_INDIA_REPORT.pdf

[53] Azra Ismail and Neha Kumar. 2019. Empowerment on the margins: The online experiences of community health workers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.

[54] Azra Ismail and Neha Kumar. 2021. AI in Global Health: The View from the Front Lines. 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 598, 21 pages. https://doi.org/10.1145/3411764.3445130

[55] Azra Ismail, Deepika Yadav, Meghna Gupta, Kirti Dabas, Pushpendra Singh, and Neha Kumar. 2022. Imagining Caring Futures for Frontline Health Work. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 480 (nov 2022), 30 pages. https://doi.org/10.1145/3555581

[56] Astha Jain, Radhika Krishnan, Ashwini Rogye, and Sundaram Natarajan. 2021. Use of offline artificial intelligence in a smartphone-based fundus camera for community screening of diabetic retinopathy. *Indian Journal of Ophthalmology* 69, 11 (2021), 3150.

[57] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology* 2, 4 (2017).

[58] Daniel Kahneman. 2011. *Thinking, fast and slow*. macmillan.

[59] Minsuk Kahng and Duen Horng Polo Chau. 2020. How does visualization help people learn deep learning? Evaluating GAN Lab with observational study and log analysis. In *2020 IEEE Visualization Conference (VIS)*. IEEE, 266–270.

[60] Christie Kodama, Beth St. Jean, Mega Subramaniam, and Natalie Greene Taylor. 2017. There's a creepy guy on the other end at Google!: engaging middle school students in a drawing activity to elicit their mental models of Google. *Information Retrieval Journal* 20, 5 (Oct. 2017), 403–432.

[61] Jude Dzevela Kong, Ugochukwu Ejike Akpudo, Jake Okechukwu Effoduh, and Nicola Luigi Bragazzi. 2023. Leveraging Responsible, Explainable, and Local Artificial Intelligence Solutions for Clinical Public Health in the Global South. *Healthc. Pap.* 11, 4 (Feb. 2023), 457.

[62] Sneha Krishnan. 2022. Exploring female frontline health workers' role and capacities in COVID-19 response in India. *International Journal of Disaster Risk Reduction* 75 (2022), 102962.

[63] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. In *2013 IEEE Symposium on Visual Languages and Human Centric Computing*. 3–10.

[64] Neha Kumar, Trevor Perrier, Michelle Desmond, Kiersten Israel-Ballard, Vikrant Kumar, Sudip Mahapatra, Anil Mishra, Shreya Agarwal, Rikin Gandhi, Pallavi Lal, et al. 2015. Projecting health: community-led video education for maternal health. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*. 1–10.

[65] Tzu-Sheng Kuo, Hong Shen, Jisoo Geum, Nev Jones, Jason I. Hong, Haiyi Zhu, and Kenneth Holstein. 2023. Understanding Frontline Workers' and Unhoused Individuals' Perspectives on AI Used in Homeless Services. 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 860, 17 pages. https://doi.org/10.1145/3544548.3580882

[66] Shahid Latif, Hagen Tarner, and Fabian Beck. 2021. Talking realities: Audio guides in virtual reality visualizations. *IEEE Computer Graphics and Applications* 42, 1 (2021), 73–83.

[67] Sam Lau and Philip J Guo. 2020. Data Theater: A live programming environment for prototyping data-driven explorable explanations. In *Workshop on Live Programming (LIVE)*.

[68] Justin Li. 2018. Explorable Web Apps to Teach AI to Non-Majors. (2018).

[69] 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*. ACM, Honolulu HI USA, 1–15.

[70] Q Vera Liao and Kush R Varshney. 2022. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences.

[71] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI?. 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 143, 14 pages. https://doi.org/10.1145/3411764.3445488

[72] Jordan A Litman. 2008. Interest and deprivation factors of epistemic curiosity. *Personality and individual differences* 44, 7 (2008), 1585–1595.

[73] Hui Wen Loh, Chui Ping Ooi, Silvia Seoni, Prabal Datta Barua, Filippo Molinari, and U Rajendra Acharya. 2022. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine* (2022), 107161.

[74] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–16.

[75] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[76] Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. 2019. A grounded interaction protocol for explainable artificial intelligence. *arXiv preprint arXiv:1903.02409* (2019).

[77] BIPASHA MAITY. 2016. Interstate Differences in the Performance of "Anganwadi" Centres under ICDS Scheme. *Economic and Political Weekly* 51, 51 (2016), 59–66. http://www.jstor.org/stable/44166205

[78] Aditya Mate, Andrew Perrault, and Milind Tambe. 2021. Risk-Aware Interventions in Public Health: Planning with Restless Multi-Armed Bandits.. In *AAMAS*. 880–888.

[79] Niharikareddy Meenigea and Venkata Ravi Kiran Kolla. 2023. Exploring the Current Landscape of Artificial Intelligence in Healthcare. *International Journal of Sustainable Development in Computing Science* 1, 1 (2023).

[80] Syeda Shaizadi Meraj, Razali Yaakob, Azreen Azman, Siti Nuralain Mohd Rum, and AS Ahmad Nazri. 2019. Artificial intelligence in diagnosing tuberculosis: a review. *International Journal on Advanced Science, Engineering and Information Technology* 9, 1 (2019), 81–91.

[81] Gloria J Miller. 2022. Stakeholder roles in artificial intelligence projects. *Project Leadership and Society* 3 (2022), 100068.

[82] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.

[83] Joy Ming, Srujana Kamath, Elizabeth Kuo, Madeline Sterling, Nicola Dell, and Aditya Vashistha. 2022. Invisible Work in Two Frontline Health Contexts. In *Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS '22)*. Association for Computing Machinery, New York, NY, USA, 139–151. https://doi.org/10.1145/3530190.3534814

[84] Pavitra Mohan and Raman Kumar. 2019. Strengthening primary care in rural India: Lessons from Indian and global evidence and experience. *Journal of Family Medicine and Primary Care* 8, 7 (2019), 2169.

[85] Maletsabisa Molapo, Melissa Densmore, and Limpho Morie. 2016. Designing with community health workers: enabling productive participation through exploration. In *Proceedings of the First African Conference on Human Computer Interaction*. 58–68.

[86] Patrick Mussel. 2010. Epistemic curiosity and related constructs: Lacking evidence of discriminant validity. *Personality and Individual Differences* 49, 5 (2010), 506–510.

[87] Lisa Charlotte Muth. 2018. "The reward of interacting is understanding" — blog.datawrapper.de. https://blog.datawrapper.de/interview-maarten/. [Accessed 27-08-2023].
[88] Amanda N. Curtis. 2022. Interactive Research Artifacts: Interactive Technologies as Tools for Knowledge Creation. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–5.
[89] Donald A Norman. 2014. Some observations on mental models. In *Mental models*. Psychology Press, 15–22.
[90] Mahsan Nourani, Chiradeep Roy, Jeremy E Block, Donald R Honeycutt, Tahrima Rahman, Eric Ragan, and Vibhav Gogate. 2021. Anchoring bias affects mental model formation and user reliance in explainable AI systems. In *26th International Conference on Intelligent User Interfaces*. 340–350.
[91] Ministry of Women and Child Development. 2018. Information Computers Technology Enabled Real Time Monitoring of Schemes (ICT-RTM). , 9–12 pages. https://wcd.nic.in/sites/default/files/Administrative_Guidelines_NNM-26022018.pdf
[92] 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)* (Seattle, WA, USA) *(COMPASS '22)*. Association for Computing Machinery, New York, NY, USA, 439–452.
[93] Chinasa T. Okolo, Srujana Kamath, Nicola Dell, and Aditya Vashistha. 2021. "It Cannot Do All of My Work": Community Health Worker Perceptions of AI-Enabled Mobile Health Applications in Rural India. 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 701, 20 pages. https://doi.org/10.1145/3411764.3445420
[94] World Health Organization. 2021. Ethics & Governance of Artificial Intelligence for Health: WHO guidance. https://www.who.int/publications/i/item/9789240029200. [Accessed 17-08-2023].
[95] World Health Organization et al. 2021. Fact sheets—malnutrition. *World Health Organization [Internet]* (2021).
[96] Joyojeet Pal, Anjuli Dasika, Ahmad Hasan, Jackie Wolf, Nick Reid, Vaishnav Kameswaran, Purva Yardi, Allyson Mackay, Abram Wagner, Bhramar Mukherjee, Sucheta Joshi, Sujay Santra, and Priyamvada Pandey. 2017. Changing Data Practices for Community Health Workers: Introducing Digital Data Collection in West Bengal, India. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development* (Lahore, Pakistan) *(ICTD '17)*. Association for Computing Machinery, New York, NY, USA, Article 17, 12 pages. https://doi.org/10.1145/3136560.3136582
[97] Rajni PALRIWALA and N. NEETHA. 2010. Care arrangements and bargains: Anganwadi and paid domestic workers in India. *International Labour Review* 149, 4 (2010), 511–527. https://doi.org/10.1111/j.1564-913X.2010.00101.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1564-913X.2010.00101.x
[98] Chunjong Park, Alex Mariakakis, Jane Yang, Diego Lassala, Yasamba Djiguiba, Youssouf Keita, Hawa Diarra, Beatrice Wasunna, Fatou Fall, Marème Soda Gaye, Bara Ndiaye, Ari Johnson, Isaac Holeman, and Shwetak Patel. 2020. Supporting Smartphone-Based Image Capture of Rapid Diagnostic Tests in Low-Resource Settings *(ICTD '20)*. Association for Computing Machinery, New York, NY, USA, Article 2, 11 pages. https://doi.org/10.1145/3392561.3394630
[99] Sun Young Park, Pei-Yi Kuo, Andrea Barbarin, Elizabeth Kaziunas, Astrid Chow, Karandeep Singh, Lauren Wilcox, and Walter S Lasecki. 2019. Identifying challenges and opportunities in human-AI collaboration in healthcare. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 506–510.
[100] Stewart Patrick and Alexandra Huggins. 2023. The Term "Global South" Is Surging. It Should Be Retired. (2023).
[101] Govt. Of Uttar Pradesh. 2023. Bal Vikas Seva Evam Pushtahar Vibhag. https://www.icdsup.net/e-book [Accessed 09-14-2023].
[102] Geraldine A Price. 2013. Creative solutions to making the technology work: three case studies of dyslexic writers in higher education. In *Approaches to Developing Accessible Learning Experiences*. Routledge, 21–38.
[103] Divya Ramachandran, John Canny, Prabhu Dutta Das, and Edward Cutrell. 2010. Mobile-izing health workers in rural India. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1889–1898.
[104] Jenisa Ramprasad, Lakshmi Krishnan, Charitha Gangadharan, Gargi Deshpande, Himanshu Madhu, Siva Teja Kakileti, and Geetha Manjunath. 2022. Performance of artificial intelligence-based breast cancer screening in a community setting: a real-world evaluation study. *The Lancet Oncology* 23 (2022), S20.
[105] Jawad Rasheed, Akhtar Jamil, Alaa Ali Hameed, Usman Aftab, Javaria Aftab, Syed Attique Shah, and Dirk Draheim. 2020. A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. *Chaos, Solitons & Fractals* 141 (2020), 110337.
[106] Juan Rebanal, Jordan Combitsis, Yuqi Tang, and Xiang'Anthony' Chen. 2021. Xalgo: a design probe of explaining algorithms' internal states via question-answering. In *26th International Conference on Intelligent User Interfaces*. 329–339.
[107] 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.
[108] Heleen Rutjes, Martijn Willemsen, and Wijnand IJsselsteijn. 2019. Considerations on explainable AI and users' mental models. In *CHI 2019 Workshop: Where is the Human? Bridging the Gap Between AI and HCI*. Association for Computing Machinery, Inc.
[109] Azhagu Meena S P, Palashi Vaghela, and Joyojeet Pal. 2022. Counting to be Counted: Anganwadi Workers and Digital Infrastructures of Ambivalent Care. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (Nov. 2022), 1–36.
[110] Alpay Sabuncuoglu. 2020. Designing one year curriculum to teach artificial intelligence for middle school. In *Proceedings of the 2020 ACM conference on innovation and technology in computer science education*. 96–102.
[111] Nithya Sambasivan and Rajesh Veeraraghavan. 2022. The deskilling of domain expertise in AI development. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–14.
[112] Lipekho Saprii, Esther Richards, Puni Kokho, and Sally Theobald. 2015. Community health workers in rural India: analysing the opportunities and challenges Accredited Social Health Activists (ASHAs) face in realising their multiple roles. *Human resources for health* 13, 1 (2015), 1–13.
[113] Benjamin Saunders, Julius Sim, Tom Kingstone, Shula Baker, Jackie Waterfield, Bernadette Bartlam, Heather Burroughs, and Clare Jinks. 2018. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity* 52 (2018), 1893–1907.
[114] Philipp Schmidt, Felix Biessmann, and Timm Teubner. 2020. Transparency and trust in artificial intelligence systems. *Journal of Decision Systems* 29, 4 (2020), 260–278.
[115] Kerry Scott, SW Beckham, Margaret Gross, George Pariyo, Krishna D Rao, Giorgio Cometto, and Henry B Perry. 2018. What do we know about community-based health worker programs? A systematic review of existing reviews on community health workers. *Human resources for health* 16, 1 (2018), 1–17.
[116] Auste Simkute, Ewa Luger, Bronwyn Jones, Michael Evans, and Rhianne Jones. 2021. Explainability for experts: A design framework for making algorithms supporting expert decisions more explainable. *Journal of Responsible Technology* 7-8 (Oct. 2021), 100017.
[117] I Sini, D Djanas, M Oktania, A Polim, and S Ulman. 2022. Mobile Obstetrics Monitoring (MOM) as a model for community-based antenatal care delivery in a low-resource setting. *Philips* (2022). https://www.philips.com/c-dam/b2bhc/master/Products/Category/enterprise-telehealth/mobile-obstetrics-monitoring/452299112911_MOM_WhitePaper_HR.pdf
[118] Daniel Smilkov, Shan Carter, D Sculley, Fernanda B Viégas, and Martin Wattenberg. 2017. Direct-manipulation visualization of deep networks. *arXiv preprint arXiv:1708.03788* (2017).
[119] Ayushi Srivastava, Shivani Kapania, Anupriya Tuli, and Pushpendra Singh. 2021. Actionable UI Design Guidelines for Smartphone Applications Inclusive of Low-Literate Users. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–30.
[120] Miriam Sturdee and Joseph Lindley. 2019. Sketching & Drawing as Future Inquiry in HCI. In *Proceedings of the Halfway to the Future Symposium 2019* (Nottingham, United Kingdom) *(HTTF 2019, Article 18)*. Association for Computing Machinery, New York, NY, USA, 1–10.
[121] Jiao Sun, Q Vera Liao, Michael Muller, Mayank Agarwal, Stephanie Houde, Kartik Talamadupula, and Justin D Weisz. 2022. Investigating explainability of generative AI for code through scenario-based design. In *27th International Conference on Intelligent User Interfaces*. 212–228.
[122] BN Tandon. 1989. Nutritional interventions through primary health care: impact of the ICDS projects in India. *Bulletin of the World Health Organization* 67, 1 (1989), 77.
[123] UN General Assembly. 2015. Transforming our world: the 2030 Agenda for Sustainable Development. https://www.refworld.org/docid/57b6e3e44.html A/RES/70/1.
[124] International Bank for Reconstruction United Nations Children's Fund (UNICEF), World Health Organization (WHO) and Development/The World Bank. 2023. Levels and trends in child malnutrition: UNICEF / WHO / World Bank Group Joint Child Malnutrition Estimates: Key findings of the 2023 edition. https://data.unicef.org/resources/jme-report-2023/. [Accessed 17-08-2023].
[125] Aditya Vashistha, Fabian Okeke, Richard Anderson, and Nicola Dell. 2018. 'You Can Always Do Better!" The Impact of Social Proof on Participant Response Bias. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–13.
[126] Nervo Verdezoto, Naveen Bagalkot, Syeda Zainab Akbar, Swati Sharma, Nicola Mackintosh, Deirdre Harrington, and Paula Griffiths. 2021. The invisible work of maintenance in community health: challenges and opportunities for digital health to support frontline health workers in Karnataka, South India. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–31.

[127] Bret Victor. 2023. Bret Victor, beast of burden — worrydream.com. http://worrydream.com/#!/ExplorableExplanations. [Accessed 30-07-2023].

[128] Jayne Wallace, John McCarthy, Peter C. Wright, and Patrick Olivier. 2013. Making Design Probes Work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3441–3450. https://doi.org/10.1145/2470654.2466473

[129] April Yi Wang, Andrew Head, Ashley Ge Zhang, Steve Oney, and Christopher Brooks. 2023. Colaroid: A Literate Programming Approach for Authoring Explorable Multi-Stage Tutorials. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–22.

[130] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y Lim. 2019. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.

[131] Susan Wyche. 2019. Using cultural probes in new contexts: Exploring the Benefits of Probes in HCI4D/ICTD. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 423–427.

[132] Deepika Yadav, Anushka Bhandari, and Pushpendra Singh. 2019. LEAP: Scaffolding Collaborative Learning of Community Health Workers in India. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–27.

[133] Qian Yang, Yuexing Hao, Kexin Quan, Stephen Yang, Yiran Zhao, Volodymyr Kuleshov, and Fei Wang. 2023. Harnessing Biomedical Literature to Calibrate Clinicians' Trust in AI Decision Support Systems. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23, Article 14)*. Association for Computing Machinery, New York, NY, USA, 1–14.

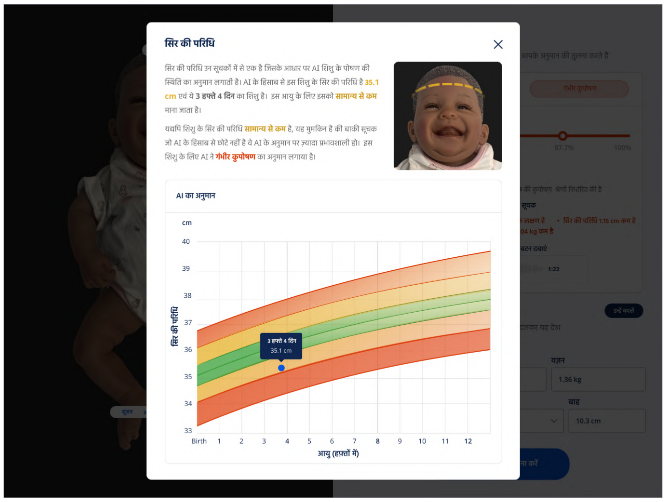
[134] Kun-Hsing Yu, Andrew L Beam, and Isaac S Kohane. 2018. Artificial intelligence in healthcare. *Nature biomedical engineering* 2, 10 (2018), 719–731.

[135] Angie Zhang, Alexander Boltz, Jonathan Lynn, Chun-Wei Wang, and Min Kyung Lee. 2023. Stakeholder-Centered AI Design: Co-Designing Worker Tools with Gig Workers through Data Probes. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

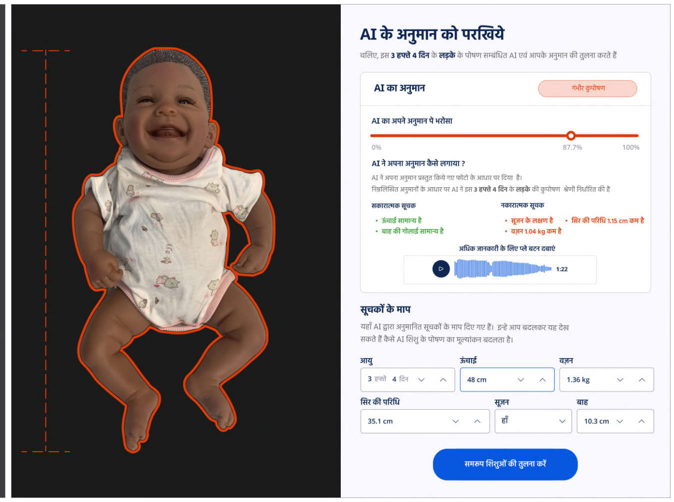
[136] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 295–305.

A DESIGN PROBE

This appendix contains the Hindi versions of the design probe we displayed in the Section 3 in Figure 1, Figure 2, and Figure 4.



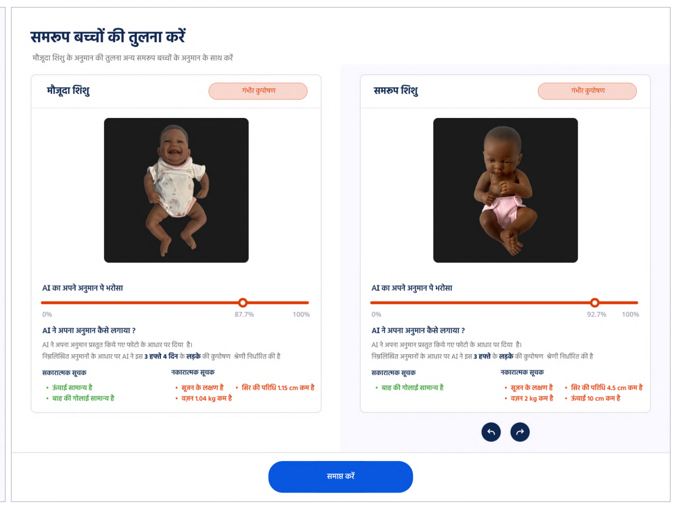
(a) Explorable Explanations Feature Information Modal.



(b) Explorable Explanations Edit Measurements Section.

Figure 5: The Hindi versions of (a) the *Feature Information Modal* and (b) the *Edit Measurements Section.*

(a) Explorable Explanations Feature Importance Section.



(b) Explorable Explanations Comparison Screens.

Figure 6: The Hindi versions of (a) the *Feature Importance Section* and (b) the *Comparison Screens.*

Figure 7: The Hindi version of the AI prediction screen in the prototype.