The ORBIT India Dataset: Understanding the Challenges of Collecting a Disability-First AI Dataset in Low-Resource Environments

Gesu India
gesu.india@swansea.ac.uk
Swansea University
Swansea, UK

Martin Grayson
mgrayson@microsoft.com
Microsoft Research
Cambridge, UK

Cecily Morrison
cecilym@microsoft.com
Microsoft Research EMEA
Boston, Massachusetts, USA

Daniela Massiceti
dmassiceti@microsoft.com
Microsoft Research
Sydney, Australia

Simon Robinson
s.n.w.robinson@swansea.ac.uk
Swansea University
Swansea, UK

Jennifer Pearson
j.pearson@swansea.ac.uk
Swansea University
Swansea, UK

Matt Jones
always@acm.org
Swansea University
Swansea, UK

Abstract

Computer vision systems are increasingly used by blind individuals to navigate their lives, helping, for example, locate objects such as doors or chairs. Yet these recognition systems do not work for many personal objects a blind user might want to find, such as keys or a special notebook. In response, efforts created personalized recognition systems, where individuals train their phones to identify and locate things, like a coffee mug or white cane, using example images/videos. However, these tools are trained on data from high-resource contexts, not necessarily reflecting India's material culture. This paper discusses the contribution of the ORBIT-India dataset, which extends these tools to the Indian context, home of the world's largest blind population. The ORBIT-India dataset comprises 105,243 images from 587 videos, representing 76 unique objects. We use this experience to examine dataset collection practices translated from high- to low-resource settings, providing recommendations to support cross-geography dataset collection.

CCS Concepts

• **Human-centered computing** → **Accessibility**.

Keywords

AI, accessibility, datasets, teachable object recognition, vision impairment, Global South

ACM Reference Format:
Gesu India, Martin Grayson, Cecily Morrison, Daniela Massiceti, Simon Robinson, Jennifer Pearson, and Matt Jones. 2026. The ORBIT India Dataset: Understanding the Challenges of Collecting a Disability-First AI Dataset in Low-Resource Environments. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3772318.3791099

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791099

1 Introduction

Artificial Intelligence (AI) has growing potential to improve the lives of people with disabilities by removing barriers to everyday tasks, from navigating a shop to finding one's keys [19, 31, 40]. However, the majority of AI tools designed to support disability are trained on datasets collected in the Global North [11, 34, 42, 44, 63], reflecting the associated material cultures and environments. As training data is critical to performance [42, 44, 49], this can result in effectiveness disparities for those using these tools in other contexts. The research literature has clearly shown that AI systems often have poor performance for users in culturally diverse contexts in the Global South [10, 24, 37]. This raises the question as to how we might extend data collection efforts to include Global South contexts.

By extending *Find My Things*, a teachable object recognition system for people who are blind, to the Indian context, this paper explores how a dataset collection protocol designed for the Global North must be adapted to work well in the Global South. Find My Things [74] allows people who are blind to identify and locate essential personal items, such as coffee mugs, white canes, or school bags, by teaching their AI app with example videos. Unlike generic object recognisers, teachable object recognisers are trained and tested under consistent, user-specific conditions, reducing variability between training and testing data and enabling greater adaptability to the user's context without cross-user interference [39].

Despite Find My Things' potential for localisation, previous research has shown that object-recognition systems fail in low-resource settings across the Global South due to differences in home environments, object availability, and camera usage behaviours [37]. These failures point to the need for training data that better reflects the lived realities of diverse user populations. In particular, systems trained on data collected by sighted individuals frequently underperform for blind or have low vision users, whose image data often includes challenges like low lighting, motion blur, and partially framed objects [16, 20, 32, 33, 50]. To develop object recognisers that are truly usable and inclusive, there is a need to train on a wider distribution of data to ensure people in a range of global contexts are served.

In this work, we introduce the ORBIT-India Dataset [36], the first teachable object recognition dataset produced by blind or have low vision data-collectors in India. This dataset consists of 105,243 images of 76 daily-use objects, contributed by 12 data collectors between July and December 2023. It has been annotated for use in a machine learning pipeline. The annotations capture whether an object is out of frame, and any personal identifiable information (PII) present in the images have been blurred and annotated accordingly. With ORBIT-India dataset creation, this paper aims to answer the following research questions: *A) How does the disability-first dataset collection protocol, originally designed for the Global North, need to be adapted for data collection in resource-variant contexts in the Global South, like India? B) What are the characteristics of collected data (visual, contextual, and environmental) and experiences of data collectors?* and lastly, *C) What lessons are learned from this intervention for future accessible AI dataset collection translation from the Global North to the Global South context?* In answering these questions, we make the following contributions:

- The ORBIT (Object Recognition for Blind Image Training)-India Dataset – a new dataset and associated documentation and descriptions to help train teachable object recognition systems for South Asian contexts;
- An analysis of the changes needed to shift a protocol from a Global North context to a Global South one, including: adaptation of data collection tools and the experiences of data collectors;
- A set of broader recommendations to support cross-geography dataset collection.

2 Prior Work

Here we discuss prior efforts related to data collection: (a) for teachable object recognisers involving blind or have low vision individuals; and, (b) for AI and other data-intensive technologies within the Global South.

2.1 Dataset Collection for Personalised Object Recognition with People Who Are Blind or Have Low Vision

Teachable object recognition enables people who are blind or have low vision to 'teach' AI systems, using a few examples, to recognise and find their personal objects [8, 39, 67, 74]. With teachable object recognisers, users can 'add' new objects, including those that are culturally specific and not usually recognised by generic object recognisers. Additionally, users can distinguish between items that are hard for them to differentiate by touch but are visually distinct. Such systems overcome the limitations of generic object recognition which only works for a small, fixed set of items [39]. Teachable object recognition offers opportunities for significant personalisation, regardless of context.

However, the training data used to create teachable object recognition systems does influence how well they can adapt to new content. For example, prior work has shown that systems trained on data largely gathered from sighted people do not work as well for users who are blind or have low vision [24, 37, 47, 67]. Images taken by people who are blind or have low vision tend to be blurrier, with objects that may be partially or completely out of frame and in insufficient lighting conditions, a phenomenon first characterized with the VizWiz visual question and answer dataset [16]. Secondly, certain objects (e.g., white cane, braille stickers, talking watch) used by people who are blind or have low vision are either not found or little-represented in datasets created by sighted people [33, 50]. Prior work on image dataset collection with individuals who are blind or have low vision has also reported several other challenges — including instances of personal identifiable information (PII) being captured in images, often unknowingly. For example, studies have estimated that up to 10% of images collected by blind users may contain such PII, highlighting additional privacy and curation concerns during dataset creation [16, 32]. Similarly, the adaptation of a teachable system to a highly different material culture is also likely to require new training data, motivating the dataset reported in this paper.

The ORBIT dataset [50, 67] was the first realistic dataset for teachable object recognisers that was collected with data contributors who are blind or have low vision, mainly in the UK and Canada. The dataset contains 3,822 videos of 486 objects recorded using iOS devices (e.g., iPhone or iPad) by 72 people who are blind or have low vision. It shows objects in a wide range of realistic conditions, including when objects are poorly framed, occluded by hands and other objects, blurred, and in a wide variation of backgrounds, lighting, and object orientations. In few-shot machine learning, models are trained to recognise completely novel objects from only a few examples, and teachable object recognisers neatly capture few-shot, high variation scenarios. An empirical study conducted on the ORBIT dataset benchmark [50] showed that training on existing few-shot learning datasets is not sufficient for good performance on the ORBIT benchmark, thus pushing innovation forward in few-shot as well as real-world recognition tasks. Though this paper does not evaluate the usefulness of the ORBIT-India dataset for teachable object recognition models, its collection was inspired by the promising research directions opened up by the ORBIT dataset. We describe the dataset's characteristics, also sharing insights into the data collection process, which may serve as a valuable reference for future work in this area.

Theodorou et al. [67] also talk about the disability-first approach taken for ORBIT dataset collection, prioritizing disability experiences above all. It employed a tool for data collection that enabled recording videos (instead of pictures), making the process more accessible for data contributors. In this paper, we describe the adaptation of this data collection infrastructure to the non-western and resource-diverse Indian context.

2.2 AI Dataset Collection in the Global South

The Global South, home to significant and rapidly growing economies such as Nigeria, India, and Vietnam [66] with increasing access to digital technologies, presents a crucial yet often overlooked context in AI development. Predominantly designed for the Global North, these technologies contribute to AI and performance divides affecting underrepresented communities, largely due to their limited involvement in AI system development [14, 29, 51, 57]. Data is the new 'oil' and the structural power asymmetries embedded within the design and development of AI systems and the extractive costs associated with it, is extended to the creation and use of datasets used to train such AI systems [45]. The growing social movement of "design justice" studies how design reproduces, is reproduced by, and/or challenges the matrix of domination (white supremacy, hetero-patriarchy, capitalism, and settler colonialism). To counter this, Sasha Costanza-Chock suggests aiming to ensure a more equitable distribution of design's benefits and burdens [21, 22]. Currently, the Global South bears this burden disproportionately, despite being the biggest global contributor to data-related work [30, 38].

To bridge this growing AI divide [60, 75], there is a growing emphasis on including under-representative communities within the AI development pipeline. Data feminism for AI systems, a feminist AI framework proposed by D'Ignazio & Klein [45], advocates for an intentional integration of pluralism by designing community-centred methodologies for participation and the visibility of marginalized labor. Our dataset protocol sets the beginning of exploring this pluralism within the dataset ecosystem for object-recognition technologies, uncovering the challenges and potential opportunities for using protocols developed for people in different contexts to be used in the global south. This step is necessary for rapid replication of the data collection infrastructure in varied contexts.

In recent years, several successful attempts have been made to create better representative datasets for various AI systems. The Dollar Street dataset, for instance, featuring 38,479 images of approximately 250 household items, collected from homes across 63 countries in Africa, America, Asia, and Europe [28], shows the variation in object appearance, branding, and placement within homes, with changing socio-economic and cultural context. DOSA, a community-generated dataset of 615 social artifacts from 19 Indian subcultures [63] aimed at improving the cultural awareness of Large Language Models, revealed the models gravitate towards web-dominant narratives when presented with "less common" artifacts rooted in marginalized and underrepresented cultures/regions. This also underscores the critical role of securing and preserving local metadata, such as object-labels as provided by data collectors, to ensure the objects are represented to reflect the authentic meaning and use within the community. Other examples of diversifying AI datasets for under-represented cultures include Masakhane [55], World Wide Dishes [48], and several more [9, 52, 53].

The Global South also encompasses the majority of the disability community, offering opportunities for social inclusion via AI. India alone accounts for the largest share of the world's blind and low-vision population [54], a demographic marked by socio-cultural diversity and varied access to resources. Various previous examples of tapping into the community's potential include [73] participants from the visually disabled community in India—many of whom were new to digital devices and literate mainly in local languages—who achieved a digitisation accuracy rate of 96.7% when transcribing handwritten Marathi and Hindi words using an Android app, outperforming crowd work platforms. Furthermore, studies also explore various methodologies for data work and community engagement [72]. However, this growing trend of data-related engagements with and collaborations in the Global South also raises ethical concerns regarding data collection and practices, such as ensuring data privacy and security within technological infrastructures, adopting culturally sensitive and participatory approaches to avoid reinforcing existing power imbalances and biases, and carefully documenting the provenance, creation processes, and intended uses of machine learning datasets to proactively mitigate discriminatory outcomes [25, 29, 62].

Our work adapts object-recognition probes, initially designed for the US and Canada [67], to the technological and infrastructural realities of India. In this paper, we report on the adaptations, while also critically examining nuances of our findings and proposing directions for AI dataset collection within the diverse multicultural landscape of the Global South. In the following section, we outline the specific modifications made to the data collection protocols and tools for ORBIT-India dataset collection.

3 ORBIT-India Dataset Collection

3.1 Data Collection Protocol

Data collectors were asked to use a modified version of the Find My Things app [74] to collect example videos of ten personal items that they might want to use a phone to find at a later point. These videos were screened for personally identifiable information (PII) and then added to the dataset, which participants were told would be open-sourced to enable technology companies and researchers to better adapt their work to disability contexts in India. Upon completion, data collectors could share their feedback on the data collection tools and protocol, in an audio-only one-to-one semi-structured interview. As per the participant's preference, the interviews were conducted over telephone/Zoom/Google Meet, and in Hindi or English. Each interview lasted for roughly 45-60 minutes and was audio-recorded upon participant's consent. As compensation, data collectors were offered two modes to choose from: either six months of exclusive access to the finding feature in the Find My Things app or a one-time monetary payment of Indian National Rupee (INR) 500 (~GBP 5) made via Unified Payments Interface (UPI) [1]. Those who participated in the feedback interview were offered additional compensation in their respective chosen modes (i.e., either three months of access to the finding feature or a UPI payment of INR 300 (~GBP 3)). The ethics approval for this work was obtained from Swansea University ethics review board. The dataset was collected between July and December 2023 in India and published at [36].

3.1.1 Object Selection. To ensure real-world relevance, participants were asked to select ten personal items they frequently use and often misplace—objects a phone-based assistive system would realistically help locate. While the modified Find My Things app provided suggestions, participants were encouraged to include culturally specific or locally relevant items to reflect Indian material culture and expand the diversity of object types. To support better

[1] Unified Payments Interface (UPI) is a secure instant funds transfer mechanism between two bank accounts in India, using UPI id, ensuring privacy of the parties involved [70]

model generalisation, we encouraged participants to include objects that varied in size, shape, and texture—from small items like keys and pen drives to medium-sized ones like handbags and water bottles.

3.1.2 Data Collection. The Find My Things app was used to record videos, using the specifically designed technique for users to capture the right video examples for training an object recogniser [50]. Following the Phase II ORBIT protocol, each object was recorded in eight short videos (15–20 seconds each):

- Six training videos, shot in isolation on uncluttered surfaces (e.g., table, chair, floor), providing clean object views.
- Two testing videos, recorded in natural contexts, with the object placed among other items in typical storage or usage locations.

Although the underlying model required only four training and one testing video per object, participants were encouraged to record the additional videos to address common issues, such as occlusion or poor lighting. This redundancy minimized the need for resubmissions and improved dataset robustness.

3.1.3 Data Quality Check. All videos underwent a two-stage verification process:

- Quality checks ensured object visibility, framing consistency, adequate lighting, and minimal occlusion.
- Privacy checks flagged and blurred any personally identifiable information (PII), including faces or documents, using a polygonal masking approach [32]. Each frame was annotated for PII presence and object visibility to support downstream model filtering and training. We discuss the data quality check procedures in detail later in the paper.

If issues arose, participants were asked to re-record specific videos. Support included daily email reminders, tutorial videos (in Hindi and English), and optional 1:1 or group assistance via phone or Zoom. More details on data quality check and dataset preparation are discussed in Section 3.4.

3.2 Adaptation of Find My Things for ORBIT-India dataset collection

To support dataset collection in India, we adapted the Find My Things app [74], previously used in UK and Canadian deployments [67, 74], making key modifications based on local context and accessibility needs.

3.2.1 Android Porting. Given that over 90% of smartphone users in India use Android devices [5, 35], we ported the app from iOS to Android (min OS 8+, API level 26+). While this increased reach, it also introduced trade-off, such as lower average camera quality and limited accessibility APIs. For instance, unlike iOS, Android does not support automatic flash light activation under low-light conditions, potentially impacting video quality.

3.2.2 Feature and Interface Updates. In the Find My Things app, we added screens for: a) Recording testing videos, b) Uploading data to secure cloud servers, c) User authentication. All app content was localised to English (India) and Hindi, a language spoken by ~45% of Indians [68], enhancing usability for native speakers. Instructions were contextualized with culturally relevant examples and included step-by-step guidance, object suggestions, and filming tips. Figure 1. shows screenshots from the English version of the Find My Things login screen and filming instructions to record a training video. Lastly, we stored and accessed the collected data through a secure cloud storage service, using cloud-based functions to manage user authentication (via user IDs) and other back-end processes.

3.2.3 Data Collection Workflow. After logging in with a unique user ID, the user enters an object name (which served as its label) and follows the app's audio-guided instructions. Each object is recorded in eight short videos: six training videos in isolation (e.g., object on floor, table), and two testing videos in realistic settings (e.g., mug on a crowded kitchen counter). During training video capture, the user begins by placing the object on a clear surface and positioning the phone close to the object. Upon detecting the object, the app provides an audio cue, after which the user moves the phone away while stepping back, capturing different angles. Throughout, audio-haptic feedback guides the user on object visibility, framing, and video duration. Each video lasts 10–15 seconds. Testing videos followed the same technique, with the object embedded in cluttered, everyday environments. Once all videos were recorded, users uploaded their data to our secure, password-protected cloud storage; local files were deleted to conserve device storage.

The first author maintained regular email check-ins with data collectors throughout the data collection period, providing consistent support with queries regarding filming instructions and troubleshooting app or device issues. We conducted online audio-only meetings on Zoom, lasting lasting 1.5–2 hours each, during which interested data collectors joined from the comfort of their homes and collaboratively collected data for one or two objects. These weekend "Datathons" were designed to encourage participation, foster a sense of community in an entirely online data collection, and provide opportunities for peer interaction, troubleshooting, and real-time support from the research team. Upon submission, the data was reviewed within 24–48 hours, and feedback on data quality and, if required, resubmission requests were made promptly. Prior work notes that data collectors who are blind or have low vision appreciate feedback on collected visual data [43, 67]. After each review, participants received a brief visual summary of their submission, including comments on data quality (for example, lighting conditions, object positioning, scenario setup) and whether the video met the dataset requirements. When it did not, participants were asked to re-record and were explained the reasons, for example, hand or body occlusion or the object being out of frame for more than half of the video length. Additionally, if prominent and multiple PIIs were present throughout the video, such as multiple human faces and bodies, that would require extensive masking, participants were asked to re-record. Such considerations were made to balance the privacy requirements with the effort taken by the data collectors. The detailed review feedback was intended to help participants assess their data quality and understand which behaviours to repeat or adjust. Each submission was used to reinforce effective practices, highlight areas for improvement, and provide tips for better-quality recordings. For instance, data collectors were

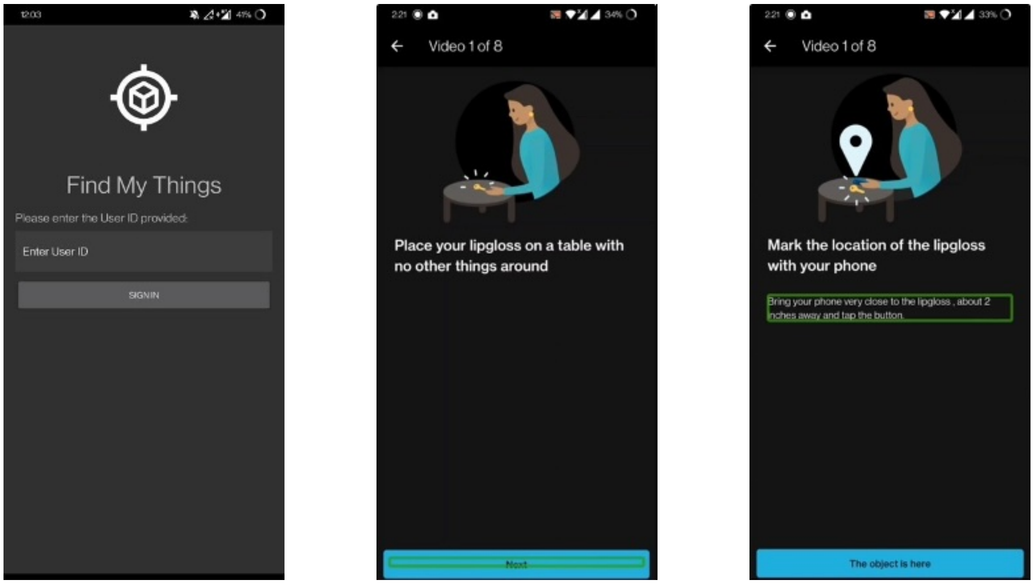

Figure 1: Screenshots from the Find My Things app: (a) Login screen for user to enter their User ID, (b) Screen with instructions to place the target object in a clear location before recording a training video, (c) Screen with instructions to bring the camera very close to the target object to start recording a training video.

reminded to avoid occlusions (e.g., hands or body), maintain adequate lighting, and prevent accidental capture of private or sensitive content.

3.3 Data Collectors

A total of 12 data collectors were recruited from different parts of India. Nine identified as men, and three as women, with average age of 27 ± 5.63 years. The majority of our participants (n = 11) came from tier-I Indian cities, as per the Government of India's three-tier city classification system [69] where tier-I cities have higher populations, more developed socio-economic schemes, and better technological infrastructure compared to tier-II and tier-III cities. Six data collectors held postgraduate degrees, four undergraduate degrees, and two had a high school diploma. Ten data collectors were full-time professionals and two were students. All data collectors, except P1, used the Find My Things app in English. The majority came from low-to-middle income families: five had annual family incomes below ₹5,00,000 (~$6,000), four were in the ₹5–10,00,000 range, and three were in the ₹10–25,00,000 range. The median household income in India is around ₹5,00,000 [4]. Detailed demographic data is shown in Table 1.

Data collectors were recruited via pan-India email lists and WhatsApp groups focused on discussing disability rights, job opportunities, and assistive tech for blind or low-vision communities. Eligibility criteria included: being 18 or older, legally blind or having low vision (with no other disabilities affecting participation), owning an Android smartphone (OS 8+), having ~700MB free storage, regular use of TalkBack for a minimum of over a year, and proficiency in English or Hindi. Data collectors also required ~600MB of internet data to download the app and upload their recordings, though internet access was not needed during the actual data recording process. Recruited data collectors gave informed consent and completed a short online survey capturing demographics, vision history, and usage of object recognition apps such as Lookout or Seeing AI.

3.4 Dataset Preparation for Public Release

The ORBIT-India dataset structure, including the organization of training and testing data and the associated annotations, was aligned with the original ORBIT dataset to ensure compatibility with existing pipelines and tools [50]. This alignment allows researchers and practitioners to extend AI models trained or evaluated on ORBIT to ORBIT-India without major pre-processing changes. It also supports reproducibility and makes the dataset readily usable for training and benchmarking computer vision models. The first author conducted all data quality checks and dataset preparation. Below, we briefly outline the process, organized into three main stages.

3.4.1 First stage: Data Categorization. The videos for each object were divided into two folders: 'clean' (for training data) and 'clutter' (for testing data). To preserve the authenticity of the data as envisioned by our data collectors, videos were placed into the categories they themselves intended. This means the dataset may contain images in the clean folder where one or more additional objects appear alongside the object of interest, and images in the clutter folder where no auxiliary objects are present. In doing so, the dataset reflects the mental models of the data collectors, supporting better model adaptability to realistic scenarios.

3.4.2 Second Stage: Labeling Instances of Personally Identifiable Information and Cases Where the Object Is Out of Frame. Upon data categorization, the next step involved identifying video data containing potential elements of Personally Identifiable Information

Table 1: Demographics and the dataset contributions by each data collector, with participants' original object labels. Abbreviations used a) for level of vision impairments (TB: Totally Blind; PS: Partially Sighted), and b) history of vision impairment (BB: Born Blind; LB: Late Blind).

P.Id.	Age	Sex	Level of VI (TB/PS)	History of VI (BB/LB)	Smartphone use (in yrs.)	Number of Objects	Number of images	Object Labels
P1	24	M	TB	BB	6	10	14122	earphone, sunglasses, water bottle, Steel glass, pin stapler, computer keyboard, mobile phone, cap, coffee mug, eye glasses
P2	26	M	TB	LB	11	2	3108	earphone, airpods pro
P3	28	M	PS	LB	10	11	10991	Box, Bangle, Black knife, Marker Pen, Tooth brush, Comb, White cane, Talking watch, Swich board, Spectacles, knife
P4	30	M	TB	LB	9	13	17850	headphone, basket, trofi, Amazon fire stick remote, wrist watch, belan, toy car, tooth brush, comb, cain, little toy car, trofi 1, hair clucher
P5	18	M	TB	BB	6	9	13936	Oneplus type a to type c cable red colour, Alexa third generation, perfume bottle, charger brick, wireless earphones, keys, spectacles, laptop, square shaped pillow
P6	29	M	TB	LB	11	1	1554	spectacles
P7	28	W	TB	BB	9	9	14676	flask, ear phones, speaker, charger, pen drive, cane, hand bag, wallet, scissors
P8	37	M	TB	BB	20	1	1547	bottle
P9	26	W	TB	BB	8	11	13666	headphone, earphone, glass container, emergency home phone mobile phone, currency note, laptop charger, mobile phone, mobile phone charger, towel, hair oil, laptop
P10	19	M	TB	BB	8	7	10891	bath towel, laptop charger, phone charger, house keys, car keys, wallet, bag
P11	36	W	TB	BB	4	1	1531	case
P12	30	M	TB	BB	7	1	1371	Spectacles case

(PII) and instances where the object of interest was completely out of frame, i.e., where not a single pixel of the object appeared. This step was necessary both because PIIs may compromise the privacy and identity of data collectors, and because out-of-frame instances can lead models to produce false positives by attempting to detect an object that is not present. Given the sensitivity of PIIs and the need for accuracy, all instances were identified via careful manual review rather than automated methods.

Two rounds of visual review were conducted on all videos, with each frame carefully screened. A total of 88 videos were flagged for potential PIIs. These were stored separately for further discussion among the authors to determine (a) whether the item in consideration truly constituted a PII and (b) the appropriate masking strategy. The review process prioritized retaining as much data as possible while balancing the masking effort against the video's value for model training. Some videos were ultimately deemed PII-free when the suspected content remained unidentifiable even at maximum zoom. Only one video was discarded, due to severe low lighting and multiple PIIs throughout, which would have required extensive editing. During the analysis, we categorized PII found in the dataset in accordance with the VizWiz-Priv framework [32]. The detected instances of PII in the dataset can be categorized into a) text-based PII, such as certificates and ID cards; b) object-based PII, such as family photographs or labelled belongings; or c) ambiguous PII, whose sensitivity depended on context. In total, 68 videos were found to contain PIIs and the identified PII items were masked using Adobe Premiere Pro.

In prior work, like [32], researchers have either blurred entire regions containing PIIs or selectively blurred only the sensitive elements. Although selective blurring is more time-consuming, it offers a better balance between privacy protection and data preservation by retaining contextual details and the overall semantics of the scene. For example, rather than blurring an entire family photograph, we manually outlined only the sensitive portions and applied masking. When PIIs appeared in dense clusters, five or more items in close proximity, we blurred the entire region. Although the impact of blurring on model performance is not fully known, such techniques are widely accepted as best practice in privacy-sensitive dataset curation and our approach is consistent with practices used in the ORBIT and VizWiz datasets [32, 50]. After masking PIIs, a marker labelled "PII" was added in Adobe Premiere Pro to each frame containing sensitive content. These markers serve as time-stamped annotations, so each "PII" marker recorded the exact frame location along with its label, and this information was exported as a CSV file for each video.

Similarly, we identified frames in each video where the target object was completely out of the frame. Nineteen videos contained at least one such instance, and for each, a marker labelled "OOOF" (Object-Out-of-Frame) was added and saved in a CSV file.

Alongside the visual analysis, the first author maintained a manual record of unique elements observed in the videos, including distinctive object characteristics (e.g., shape, size, branding) and their relative positioning within the environment. These records were used to conduct a comparative analysis between the ORBIT-India and ORBIT datasets to examine cultural and contextual nuances reflected in object appearance, auxiliary items, and household backgrounds. Due to the larger size of the ORBIT dataset, a qualitative sampling approach was adopted to obtain a representative overview of typical visual features. Specifically, 100 images were randomly selected from the ORBIT dataset, comprising 30 from the clean folders (training data) and 70 from the clutter folders (testing data). These images were jointly examined and discussed among the author team to identify recurring patterns such as household layouts, object types, and contextual characteristics. The interpretation of these differences were informed by our research team's diverse ethnic backgrounds and lived experiences. The resulting observations are detailed in the following section.

3.4.3 Third stage: Compiling Data and Data Annotations. Following the ORBIT dataset annotation format, each video in our dataset was annotated with a JSON file containing the keys: object_not_present_issue and pii_present_issue. For videos with PII and/or OOOF markers, the corresponding key values were set to true; otherwise, they were false. All videos were then split into individual JPEG frames, and a final visual review was conducted on every image to confirm correct masking and labeling of PII and OOOF instances. Annotation files were stored in a separate folder within the dataset.

Total number of images, across the clean and clutter folders, for each object and each data collector, is summarized in Table 1. The final dataset size is 5 GB and was released on July 1, 2024 [36]. In next section, an overview of the collected dataset is shared, including the general characteristics of the data and a simultaneous comparison with its parent dataset, ORBIT [50, 67], based on our object descriptions, images in the dataset, and findings described in [50, 67].

4 The ORBIT-India Dataset: An Overview and Comparison with the ORBIT Dataset

The ORBIT-India dataset contains a total of 105,243 images of 76 commonly used objects, collected by 12 data collectors who are blind or have low vision in India. The full annotated dataset can be found at [36]. Out of the total images, 75,994 belong to clean scenarios and 29,249 belong to clutter scenarios. On average, each data collector submitted 8,770.25 images. These images were extracted from a total of 587 videos, with an average of ~180 frames extracted per video.

4.1 Objects in the dataset

The ORBIT-India dataset comprises 76 total objects, representing 36 unique items frequently used by participants in their everyday lives. A complete list of all objects is given in Table 1. For each object, we collected an average of 5.58 clean videos and 2.14 clutter videos. Of these, 68 objects meet the criteria for training few-shot learning models, having at least four clean and one clutter video. The remaining eight objects—five with no clean videos and three with no clutter videos—do not meet the minimum threshold for few-shot training but still contribute valuable contextual information, such as background settings, object placement, and spatial relationships.

On average, each data collector contributed 6.33 objects, slightly fewer than in the original ORBIT dataset (eight objects per data collector). Although data collectors were instructed to collect ten objects each, the final number varied because participants dropped out due to time constraints and unforeseen personal circumstances. In such scenarios, participants were compensated for each object data submission. Figure 2 shows example images from the dataset.

4.2 Cultural and Contextual Uniqueness

While many objects overlap with those found in the Global North (e.g., ORBIT), several objects in our dataset reveal the distinct cultural practices, consumer preferences, and socio-economic conditions specific to India. For instance, objects like hair oil, a staple linked to South Asia's Ayurvedic heritage, and bangles, represent a culturally significant accessory. Other examples—such as a 'OnePlus type-A to type-C red cable'—reflect local branding and consumer preferences. Even globally common objects take on distinct forms: for instance, steel glasses (as shown in the last image in Fig. 2), widely used for drinking water in India, especially in low-to-medium income households, differ from ceramic or glass alternatives typically found in Western households. This cultural embeddedness of objects resonates with the Dollar Street dataset [28], that shows how the same object looks different depending on where one lives.

The dataset also highlights differences in household organization of the objects. For instance, in contrast to Western household norms, toothbrushes are sometimes stored near the kitchen sink in India. This practice is shaped by the collective Asian culture of living in joint families, where space is shared more intensively than in nuclear households, and also reflects resource constraints, such as the possibility of a single busy bathroom, making it practical to keep certain essentials elsewhere for easy access. Examples of such data, embodying cultural patterns of cohabitation and adaptation, improve cultural awareness of vision systems by better understanding how objects relate to their environments across cultures [24, 53]. The inclusion of Indian currency notes in the dataset also adds to the region-specific visual data.

The cultural uniqueness also extends to the local, or more personable, naming of objects. For example, a rolling pin labelled as 'belan' and cushion as 'square-shaped pillow'. Unless it was misleading, we retained the object labels as provided by the data collectors, to preserve the authenticity of data. For example, 'Trofi' for trophy, 'Cain' for cane, and 'Hair clucher' for hair clutcher. One participant labelled their eyeglasses as 'lamps' to bypass TalkBack-related typing

Figure 2: Sample images from the ORBIT-India dataset. Top row, from left to right: Training image for i) coffee mug, ii) bangle, iii) trophy, iv) handbag. Bottom row, from left to right: testing image for v) perfume bottle, vi) hair oil, vii) water bottle (with a manually blurred photo frame behind), and viii) steel glass.

issues and later requested a change, which we applied. Capturing such variations in data allows the systems to learn from natural inconsistencies, misspellings, and culturally-influenced naming practices [37, 50, 63], and improves model robustness.

Technological preferences also diverge. While the ORBIT dataset includes smart home devices like Alexa speakers, Apple AirPods, and Amazon Fire TV remotes, ORBIT-India reflects the prevalence of Android and OnePlus devices, which are more accessible and affordable for much of the Indian population. Unlike ORBIT which includes a few large-scale objects (e.g., front doors, cars), the ORBIT-India dataset is limited to small to medium-sized items. Two participants who attempted to capture large-sized objects (e.g., a bicycle), found it difficult to frame them effectively.

We made more attempts to preserve the authenticity of the data while balancing intent with model requirements throughout the analysis process. For instance, despite the frequent low-lighting issues, all images were left un-edited to represent the authentic middle-class Indian household scenarios. This also represents a real-world use-case scenario, when images are taken from lower-end Android phones, and needed for ensuring better generalisation of the model beyond the (comparatively) better iOS-quality images. Similarly, clutter and clean classifications of the data were also retained to reflect each data collector's mental model. In cases where cluttered videos were submitted separately (requested as part of the video review process), we reassigned the original labels to maintain consistency.

Taken together, the dataset provides valuable training data for building AI systems that reflect the diversity of Indian material culture. It captures: a) scene complexity consisting of diverse backgrounds and semantic cues, b) object relationships (co-occurrence, positioning, occlusions, and angles), c) environmental conditions like brightness and lighting, and d) cultural context that includes region-specific object use and affordances.

4.3 Presence of the object-of-interest in the frame

Out of 105,243 images, the object of interest was completely out of frame for 1,170 images (approximately 1.11%). Of these, 773 images belong to clean scenarios and 397 images belong to clutter scenarios.

During the analysis, we found major reasons for the object to be found out of frame were either hand occlusion or sudden camera movements, often unintentional, as confirmed by the data collectors later, but reflect real-world scenarios where users struggle with positioning objects for recording. In few-shot learning, annotated out of frame object images help the model focus on the true features, instead of, for example, the background, thus act as explicit negatives for the model to learn to distinguish between object present and absent scenarios. This, in turn, improves the generalisability of the model to novel settings [13, 33].

4.4 Occurrence of Personally Identifiable Information

Of the 105,243 frames in the ORBIT-India dataset, 3,995 (3.80%) were found to contain one or more instances of personally identifiable information (PII). This number is notably less compared to other datasets (e.g., VizWiz [16] had ~13% images with PII, leading to VizWiz-Priv [32]). We observed more instances of PII in clean video contexts (2,093 frames, or 2.75% of 75,994 clean frames) than in cluttered ones (1,902 frames, or 6.50% of 29,249 cluttered frames).

5 Experience of Data Collectors with Using the Find My Things App

Ten data collectors shared their experiences with using Find My Things app for data collection in the feedback interview. The remaining two (P6 and P8) did not participate due to time limitations. Three interviews (with P1, P4, and P10) were conducted in Hindi, later translated to English. The interview translation and transcription was manually conducted by the first author. For data analysis, thematic analysis [18, 65]. protocol was followed: the first author familiarised herself with the transcripts, systematically generating initial codes from relevant quotes resulting into a preliminary code-book. All authors discussed the relevance and relationship between the codes and the emerging thematic patterns. The final stage involved defining and producing the themes, resulting in a structured final code-book with clearly defined themes and their respective codes.

The data collectors were asked about the process and experiences with data collection, such as the thought process behind object selection, filming videos, following instructions, and creating scenarios for shooting videos. Data collectors had access to the application for an average of three weeks. Data collection was conducted across multiple sessions, during which participants typically collected data for one to two objects per session. Recording a single object—including the time needed to set up scenes—required an average of five minutes. The entire data collection process averaged 3.5 hours per participant (with a maximum of five hours). Two of the three data collectors (P3, P4, and P8) with access to the Find My Things app's finding feature were interviewed. Although none had used the feature, they appreciated the application's concept. P4 said, *"Find My Thing is a good thought...because the greater hardship a visually impaired faces is to find his own things. And one cannot always depend on someone else who will find my things for me, while wasting his time...Of course, there are ample applications like Google Lens, Lookout, they do similar kinds of things, but they cannot identify particular items, where we cannot have the option to record our things, upload it, and they will help us finding it back. So, in that sense, it's a unique application."* Data collectors appreciated the app in making the filming process accessible, and we detail out their feedback below.

5.1 Instructions for Filming

Participants found the filming instructions particularly helpful. For example, P2, who uses an iPhone for its accessibility features, was surprised by the accessibility embedded in an Android app.

> *"Those audio cues and the initial instructions were helpful. We don't get a lot of details whether you know the object that you are recording whether you know if it is exactly in the frame whether you know it is in center or is it in the right hand side of the frame or on the upper side of the frame so that information like in no application that is available where exactly the object is."*

Similarly, P4 found the step-by-step instructions guided him to hold the phone correctly, frame the object for the camera lens, and enabled him to record videos independently for the first time. He noted, *"Whenever I was not able to point accurately, within seconds it was guiding. The time we lose our focus, we won't be able to record properly, so that helped me a lot."*

Three participants had prior experience recording videos, which helped them navigate the process more easily. Except for two, most had also used an object recognition app before, which they felt made framing easier. As P3 said, *"The Lookout application is very similar to this application... because it tells us to move your phone right, move your phone bottom more, take your phone up, go away and come close so every instruction can be given to us through the camera. So recording video was actually not a difficult task for me."*

5.2 Object Selection

We encouraged participants to share their thought process when selecting objects. Participants put considerable thought into object selection. Their prior experiences with using object-recognition apps also influenced their choices, considering both size and personal relevance. While they were instructed to choose personally relevant items, many also considered objects important to the wider community. Some even discussed with friends and family members, often sighted, before deciding. P7 explained their process: *"I thought there are different types of dresses, right? What is today with me may not be there in the market itself tomorrow… also like the exact same dress, the pattern. So it's really going to be useless if I just record a dress. So, we were just on brainstorming calls with my friend, I just wanted something that everybody would have, and they would lose and they can't get it immediately."*

Some tried recording large objects (such as a bicycle) but struggled due to lack of instructions or framing issues. Others intentionally avoided large objects, feeling it was riskier. For example, P3 preferred smaller items, while P1 tried larger ones like an almirah (cupboard), sofa, and table but found the app less effective. Some participants had difficulties with very small objects, such as a pen drive (P7) or a bindi (P9). P4 deliberately recorded small objects, like toy cars, noting that even sighted people struggle to find them. Some participants also recorded cherished objects. P4, for instance, recorded toy cars belonging to his son, as well as his own trophies (with PII blurred). A few avoided sensitive objects to protect privacy. As P7 said, *"I actually thought of car keys and house keys. I don't know why, but I was worried about the key number. What if it's visible, right?"* Several found it easier to record away from family to avoid distractions or accidental appearances in videos. As P4 explained, *"Most videos I was able to record when my kid was not around. Whenever I make videos, or focus, he will suddenly take the phone or the object and run away."*

Some participants recorded culturally specific objects, such as a belan (rolling pin). As P4 noted, *"I wanted to try something that is fully ours, and used only in India. Like, the app would call it [belan] a stick but its shape is a bit different. It needs to recognise it."* P4 also emphasized precision in labelling: ' *'For a few objects I tried writing specifically, like I wrote wrist watch for my watch, because a watch can be a wall watch or table watch. Same for Amazon remote, because remotes are of different types, right? The suggestion for the application is not technically clear but like how Be My Eyes tells you, if everything gets developed like this then it will be good."*

5.3 Creating Scenes

Depending on the context, participants either created cluttered scenes or placed objects in natural settings. This helped avoid capturing PII in the frame. Some experimented with surfaces to improve accuracy. For instance, P4 avoided recording on tiled floors, which triggered app errors, and used a mat instead. P3 used a whiteboard to reduce patterned backgrounds. P9 recorded videos in the dark, hoping the app would learn to recognize objects in low light. Once a lighting prompt was added, participants like P2 found it helpful when recording in corners or sofa crevices.

Few participants mentioned taking PII into account while selecting scenes. P7, for instance, did not record videos of keys (house keys and car keys) due to the possibility of finding personal details on the keys. Similarly, P4 mentioned, *"I had to pay attention that no such thing gets captured that shouldn't be, because ultimately, it was our room. Like privacy-related, some pictures are not clicked. I did try for that."*

Ensuring well-lit conditions also played a role in selecting locations for shooting videos. Data collectors shot in familiar environments like their house or office spaces, and were aware of the locations with better lighting. P7 said, *"I know what my was room's layout. So I made sure that the lights were on wherever it's dark. And the other side I knew where the window is, that's how I ensure that the lighting was proper."* The in-app alert for low-lighting conditions also helped data collectors. For instance, P1, who often forgets to turn on lights, followed the app's prompt for creating well-lit conditions for filming.

5.4 Overcoming Recurring App or Device-Related Issues

Six out of 12 data collectors reported issues with battery drainage and phone heating. Since the app used the camera continuously, and participants often kept the internet on for notifications, they began switching it off upon our instructions. Battery issues were also common during 'Datathons,' where participants used the same device for multiple apps at once. P1 said, *"the phone would switch off when battery drainage was too much or heating. Like even during the Datathon, the hotspot was on, then the camera, and I was using Zoom, so everything was too much."*

Such issues added hurdles to the data collection process for some participants. As P4 noted, *"I wanted to do more (data collection). Sometimes after recording 4–5 videos, if we reach 3 or 4 or 5, the app used to get stuck. So, again we had to start and again we had to start."* On average, participants took 5–10 minutes to record one object, including scene creation. This was shorter (2–3 minutes) when there were no issues such as battery drainage or app errors.

5.5 Suggestions for Improvement Within the Find My Things app

Participants shared their suggestions for improving the usability of the FMT app. P5 found overlapping feedback (haptic, audio, TalkBack) overwhelming and suggested allowing users to customise preferences. P7 wanted an in-app review system for video quality and a *"lighting clarity percentage bar."* P1 preferred having all instructions at the beginning, plus specific guidance for large objects. Lastly, P3 requested more detailed feedback on object positioning: *"Totally blind people might require a few more descriptions... If the object is only half in the frame... it would be good."*

6 Discussion

The paper presents the ORBIT-India dataset [36], the first teachable object recognition dataset collected by individuals who are blind or have low vision in India. The dataset consists of 105,243 images of 76 daily-use objects, contributed by 12 data collectors between July and December 2023. We have described the dataset characteristics, its similarities and differences with the dataset it was inspired by (ORBIT [67]), and data collectors' experiences with the tools and protocol. We now discuss the challenges encountered during dataset collection and their implications for future AI data collection in low-resource communities. We also discuss the limitations of our dataset collection and outline directions for future work.

6.1 Barriers in Inclusive Dataset Collection in Low-resource Communities

Data collection in low-resource settings is shaped by unpredictable interactions between device variability, connectivity issues, digital literacy, and sociocultural or environmental constraints. However such challenges are not only technical, but also structural, aligning with the critiques from data colonialism that highlight how the global data practices often extract value from marginalised communities without sharing equitable benefit or agency [23].

Access to technology shapes its use and adoption. The hardware and resource demands of our tools effectively set eligibility boundaries for who can contribute their data or not. While it may seem counter-intuitive to collect data from someone whose device may not yet run the app (e.g., a teachable object recogniser), it is important to include data from people of diverse backgrounds (digital expertise and experience, gender, age, geographic and socio-cultural background). Data feminism reminds us that design choices about who gets to contribute to shaping technological systems are never neutral: they reflect power, privileges, and historical inequities in whose knowledge matters in the built systems[45]. The VizWiz dataset collection shows that even when users' devices produce low-quality or unconventional images, their participation is necessary to build systems that reflect real user needs [16, 32, 33].

Emergent users, a term first introduced by Devanuj and Joshi [12], are individuals—often in the Global South—who are just beginning to access modern mobile technologies. Including them in dataset design and representation is non-negotiable, even before tools fully support their devices. Their participation addresses representational gaps, ensuring systems evolve alongside their users. Supporting such contributors is key to building equitable datasets and, by extension, equitable AI systems for the global disability community.Creation of disability centred datasets like the ORBIT and VizWiz datasets show direct and early end user involvement crucial in capturing authentic, diverse, and realistic real-world data.We provide the following recommendations based on our data collection experiences for possible future data collection processes that may not be limited to our work, like with disability community and image datasets.

6.1.1 Overcoming network and Connectivity Issues. While internet access in India has improved significantly in the past decade [3], practices like data sharing via Bluetooth remain common, especially in rural areas where this method is often more reliable (and entirely free). To have equitable representation from rural areas, it might be beneficial to design for such practices—e.g., offering Bluetooth-based submission options—which can make data collection more inclusive, especially in other regions, where internet costs remain high (e.g., $19 per month for 3GB in South Africa) [61]. Data collection entities could also consider providing community-inclusive Wi-Fi hotspots, where people may visit to install the data collection tools and upload their data when recorded, free of charge. When such infrastructural and resource-related factors are overlooked, data collection efforts risk privileging urban, wealthier, and better-connected populations, leading to AI systems shaped by self-selection bias, where even within under-represented groups, only the most privileged voices are captured. Prior work such as VizWiz underscores that connectivity-dependent workflows can systematically exclude users in low-bandwidth contexts, further emphasizing the importance of offline or low-cost submission pathways.

6.1.2 Overcoming Device Limitations. In cases where shared hardware use is permitted, it offers promise as an inclusive solution. The StoryBank project [26, 27], for instance, deployed a walk-up-and-use kiosk in rural India for digital storytelling. Similarly, the Com-Me project [15] installed public tablets in rural South Africa to support media sharing. Such models can enable inclusive participation without requiring individual device ownership.

In our own dataset collection, a continuous feedback loop with participants helped navigate challenges around resource availability—for instance, by providing Wi-Fi hotspots, allowing flexible data submission timelines, and enabling low-light app notifications—however, remote data collection also meant the burden of troubleshooting largely fell on participants. Our data collectors, on average, had 9 ± 3.86 years of experience using smartphones and were all familiar with object-recognition apps, as well as with using cameras for object-recognition tasks. Not all potential participants will have this level of digital expertise or fluency. Designing inclusive tools requires anticipating these skill gaps and meeting participants where they are—through thoughtful design, flexible workflows, and responsive support. This is especially critical across socio-economic contexts, where even small differences in resource access can impact participation. The image datasets collected with individuals who are blind or have low vision in past prove that collected images may be unpredictably framed, have lighting or quality issues, reinforcing the importance of designing workflows that accommodate non-expert and low-resource contributors.

6.1.3 Language and Socio-Cultural Contexts . India alone has over 121 languages [1] and hundreds more dialects. People in Africa and Asia, for instance, often speak a tribal or native mother tongue, as well as a more widely-spoken national language, with some understanding of neighbouring community languages, and often some English [58, 59]. Our protocol supported Hindi and English, which shaped the inclusivity of both the data and its labels.Data colonialism and data feminism argues for elevating local, situated knowledge, especially from communities whose languages, cultures, and ontologies are often marginalised in global tech systems [23, 45]. Recruiting data collectors who speak regional and diverse languages contributes not only to greater inclusivity but also enhances the variety in the dataset. These participants are more likely to interact with region-specific objects—such as local-language newspapers, packaging, or signage—that reflect their linguistic and cultural environments. Including such data ensures that AI systems are trained to recognise a broader range of real-world objects, making them more relevant and usable across different language and cultural contexts.

6.1.4 Understanding of AI, Android Expertise, and Digital Fluency. Another challenge in recruitment arises when working with communities unfamiliar with complex AI concepts—such as teachable object recognition datasets—where instructions require more than just plain language translation, especially for low-literacy or less digitally fluent emergent users. Communicating goals, instructions, and technical requirements to non-expert users calls for thoughtful, accessible strategies. Co-design and participatory design approaches with potential data collectors can help the researchers design data collection protocols aligning to their understanding, knowledge levels, and digital exposure. Access restrictions, skill requirements, and support mechanisms shape who is able to contribute to a dataset—and, in turn, shape the quality and equity of representation within it.

6.2 Culture-Specific Perceptions of Privacy

Our dataset contained 3,999 images with personally identifiable information (PII)—roughly 3% of the total—significantly lower than other datasets, which have reported PII in about 10% of images. This relatively low proportion is reflective of the caution exercised by our data collectors. However, their interpretations of what constituted 'private' varied across individuals and were often shaped by socio-cultural norms. For instance, Participant P2 recorded videos in his family's living room, where items such as family photographs, graduation certificates, and other sensitive documents were openly displayed. The living room, in this context, became a space of blurred boundaries—public enough for guests, yet deeply personal. As P2 noted, *"that's where my guests come and sit,"* offering a glimpse into a communal understanding of privacy commonly observed in many Indian households.

In contrast, another participant chose to record exclusively in neutral parts of their home, deliberately avoiding identifiable elements while placing trust in the research team to remove any accidental PII. These contrasting strategies underscore how deeply privacy is embedded in cultural and spatial logics, rather than being a universal or uniform concern. Prior studies show that lack of accessibility features related to photos or videos often lead to people who are blind or have low vision unknowingly sharing images containing PIIs [6, 7, 17]. This can also be seen in the case of AccessShare study, where participants upon reviewing their data having PII agreed to share it[43].The reasons could be comfort with sharing private information as they do not consider it private, not understanding the risks of sharing private information with the strangers, or submitting data, especially when all data may have private information, for the greater good to help the community.

Understanding participants' perceptions of privacy—and their engagement with digital risks—is crucial, not just from an ethical standpoint, but also because it directly impacts the data collection pipeline. These perceptions influence recording behaviour and contribute to the downstream time, labour, and attention required to detect, mask, and annotate each PII instance.

Kamikubo et al. developed AccessShare, a system that screens images upon collection and provides automatic short descriptions, including information on image quality (blurriness, hand occlusion, object presence) and potential PIIs [41, 43]. The system enabled data collectors who are blind or have low vision to review their own contributions for teachable-object recognisers and decide whether they wanted to share them. Despite some inaccuracies, data collectors valued autonomy and agency over their data. Such interventions also help reduce the manual burden on researchers responsible for reviewing and screening data, saving considerable time and resources.

For instance, in case of ORBIT data collection, the team outsourced the work of cleaning and annotating the data. In our case, the first author undertook this work contributing a total 300 hours on carefully inspecting, labelling, and annotating the data for the dataset. The AccessShare study also showed that only a small number of data collectors sought assistance from sighted loved ones when reviewing their collected data [43]. This aligns with the works of Ahmed et al., that such reluctance may emerge from limited availability of sighted help, a desire for independence, or the wish to maintain privacy [6, 7].

Developing automated systems capable of detecting content in images collected by people who are blind or have low vision presents a chicken-and-egg challenge: the very data needed to train such systems is limited because collecting it requires those same systems. Nonetheless, AccessShare demonstrates that even imperfect tools can serve as meaningful first steps. In our study, participants appreciated receiving feedback about their submissions, which helped reinforce effective filming techniques and behaviours. These systems, or similar interventions, could be crucial for including individuals with lower digital expertise, thus supporting the creation of more representative datasets. More diverse data collected for teachable-object recognition can also support the training of other AI systems to better adapt to diverse cultural contexts.

One of the ways to improve the systems like AccessShare is to collect more ethical and disability-first 'privacy' datasets, like Blind-Priv [64] and Viz-Wiz-Priv [32], that can be used to train systems to detect objects with potential PIIs. Such datasets should be inclusive and should include region-specific cues—such as language scripts, household layouts, artifact styles, and visual markers of identity—to teach models to recognise both universal and culturally specific forms of sensitive information.

However, such systems may still require human oversight due to cultural nuances in how privacy is understood. For example, in our dataset, one data collector recorded videos of two trophies received from the workplace. Although the trophies displayed the data collector's full name and workplace address, the collector viewed them as symbols of pride and reputation and submitted the videos knowingly. In cases where cultural and personal perceptions of privacy conflict with personal security, human intervention becomes essential. This could be addressed by developing privacy datasets from varied cultural contexts. It is essential for accommodating differences in what is considered "safe" or appropriate to share. For instance, one data collector in our study was careful to ensure that her clothing items that were laying around were not captured in the video. Access to automated tools that can flag such elements would ease the workload on data collectors, reduce cognitive burden, and increase the likelihood of participation.

6.3 Locating Privacy Within Legal, Cultural, and Technological Contexts

Perceptions of privacy are not shaped in isolation—they are closely linked to individual experience, cultural values, local norms, and the broader regulatory environment. India's evolving legal landscape around data protection, particularly with the recent enactment of the Digital Personal Data Protection Act, 2023 (DPDPA) [46], brings this tension into sharper focus. While the DPDPA draws from global frameworks such as the European Union's General Data Protection Regulation (GDPR) [2], it lacks the same clarity, enforcement, and accountability, particularly around enforcement of rights like the "Right to be Forgotten." Such gaps affect how individuals conceptualize digital risks and influence the support structures they may—or may not—expect when sharing personal information.

These systemic factors must be accounted for in the design of privacy practices and consent processes. Researchers must be cautious not to impose Western-centric privacy models onto communities with different lived realities. Instead, the goal should be to co-create ethical protocols that align with local ways of living and sharing. This includes acknowledging that privacy, in many cultures, is not a strictly individual concern but one negotiated within families, communities, and social networks. Addressing this knowledge gap requires culturally grounded training and support mechanisms. Visual storytelling, localized case studies, and co-designed examples of PII can make abstract risks more tangible and contextually meaningful. However, this must be balanced against the cognitive burden placed on participants—especially when such guidance could inadvertently overwhelm or disengage them.

Finally, the ethical design of data collection processes must be grounded in the values of the communities involved. This requires critically rethinking our assumptions around consent, annotation pipelines, and what privacy protection looks like across different contexts. Robust national policies and institutional support—through legal enforcement and public awareness—are vital in building a privacy ecosystem that extends beyond individual effort or technological fixes. In conclusion, while our dataset highlights the potential for minimizing PII through culturally informed and diligent practices (as seen in prior work like ORBIT [67]), it also surfaces the hidden labour and complexity involved. Future datasets, particularly those involving blind and low-vision contributors from diverse global settings, must prioritize inclusive, ethical, and empowering data practices—ensuring technological innovation does not come at the cost of personal dignity and cultural integrity.

6.4 Limitations and Future Work

There are two main limitations of the ORBIT-India dataset. Firstly, the dataset is smaller when compared to other visual datasets like VizWiz and ORBIT. However small datasets can still be very useful. First, they can be incorporated into large pre-training datasets to provide at least some representation of content that would otherwise be completely absent. Even if imperfect, this gives models a better 'base' understanding of those concepts. Second, small datasets can seed the collection of larger datasets — whether through targeted web-scraping or, in the future, data synthesis as generative models improve. Finally, it is very common practice to adapt a model 'post-training' via methods like parameter-efficient fine-tuning (e.g., Low-Rank Adaptation of Large Language Models (LoRA)), in-context learning, and/or RAG. These approaches have been shown to be very successful with even small datasets — especially if they are high-quality/curated datasets [56, 76]. Thus, our dataset can be used with such post-training techniques to adapt models to this specific content, and would be a far more efficient and cost-effective way to inject new information than re-training from scratch.

A second limitation is that our dataset is contributed by 12 individuals who are blind or have low vision and cannot be broadly generalised to the realities of the country's entire population. For example, our data collectors represent a young age group (with the youngest being 18 and oldest 37), and 75% identified as male. As a result, the dataset skews toward the interests, needs, and environments of younger men, and may not fairly represent older adults or women. While the dataset does include household items regardless of gender, other items like saris, bindis, or women's accessories appear less frequently in the dataset. All data collectors were urban residents and objects and realities of rural communities are not captured in our dataset. For example, in many Indian villages, houses are traditionally constructed using locally available materials such as mud, thatch, or brick. Some households still traditionally use items like oil-fuelled lanterns (called lalten), as they are more environmentally friendly. Sleeping arrangements may include cots, wooden charpai, or mats on the floor. Utensils made of clay or earthenware are also common in some regions, particularly in hotter climates, as they help keep water and food naturally cool. India is a large and diverse country where culture, language, food, and daily practices change every hundred kilometer and urbanisation in the country is influenced by global, mostly western, style of housing, consumption, and lifestyle compared to rural regions. The essence of this cultural and regional diversity can be better represented when data collectors are recruited from rural communities. Intersectionality also plays a huge role: for instance, women who are blind are less likely to have the resources to participate in such a dataset collection. This dataset is a meaningful first step, demonstrating feasibility and providing a foundation for scaling to larger cohorts.

Future work in this direction should aim at creating a larger and more representative dataset, expanding to participants coming from diverse socio-economic backgrounds across gender, age, and geographic regions—including tier-III and rural areas of the country and beyond. To capture this diversity, the data collection tools and protocols have to be adapted to the infrastructural and technological realities of each region. For instance, our recruitment relied on online channels and urban-dominant languages, which limited participation from rural communities, who often speak regional dialects and may be more effectively reached via trusted local sources like village administrations or NGOs. Similarly, the FMT app relied on certain functionalities like ARCode to make data collection accessible, however this limits its usage to mid-to-high range Android devices. Individuals that are first-time smartphone owners and come from lower socio-economic background and/or from rural areas tend to have lower-end Android phones with limited functionalities. Additional support, such as internet connectivity, in-person troubleshooting assistance and workshops, may further contribute towards improving inclusivity of such protocols. It will also be beneficial to collect more PII-focused datasets [32, 71] for automated PII detection in collected image data, to reduce the cognitive burden on participants.

Lastly, an exciting step forward would be to explore how well ORBIT-trained models perform when fine-tuned or tested on our dataset [50]. This would help quantify generalization gaps and spur the development of domain-adaptive few-shot methods. For instance, can a model trained on ORBIT be adapted with just a few examples from our dataset to perform well in Indian settings? Additionally, a comparative evaluation between ORBIT and our ORBIT-India dataset could reveal systemic biases or assumptions in current few-shot models. The ORBIT-India dataset could be used as a cross-cultural generalization benchmark to evaluate how well models trained in one domain perform in a very different one.

7 Conclusion

This paper introduced the ORBIT-India dataset—the first teachable object recognition dataset collected by individuals who are blind or have low vision in India. The data collection tools and protocols are iteratively adapted to the local technological and socio-cultural context, resulting in a dataset that reflects realistic household environments and everyday objects in Indian settings. We demonstrated how the data collection process was extended to the Indian context, with tools iteratively designed and refined based on on-the-ground experiences. We uncovered insights from both the data collected and the data collectors' experiences using these tools, highlighting how the dataset reflects not only their lived realities but also the broader challenges and learnings encountered throughout the process. We also discussed the limitations of our work and proposed directions for future AI dataset collection within the diverse multicultural landscape of the Global South.

We are deeply grateful to the data collectors and visual disability community members in India who generously contributed their time, effort, and care for this dataset creation. The dataset stands as a testament to their labour, insights, and collaboration, and would not have been possible without them.

Acknowledgments

This work was supported by Engineering and Physical Sciences Research Council grant EP/Y010477/1 and by an EPSRC–Microsoft Research ICASE Award (EP/W522053/1). We are deeply grateful to the community of blind and low-vision individuals in India, especially our data collectors, who generously shared their time, effort, and care in contributing to this dataset. The dataset stands as a testament to their labour, insights, and collaboration, and would not have been possible without them.

References

[1] 2011. Census of India, Paper 1 of 2018, Language – India, States and Union Territories. http://www.censusindia.gov.in/2011Census/C-16_25062018_NEW.pdf

[2] 2016. General Data Protection Regulation (GDPR) – Legal Text. https://gdpr-info.eu/

[3] 2016. Reliance Jio launches cheap 4G services in India | India | The Guardian. https://www.theguardian.com/world/2016/sep/05/reliance-jio-launches-cheap-4g-services-in-india

[4] 2024. People Research on India's Consumer Economy & India's Citizen Environment - ICE360. https://www.price360.in/publication-details.php?url=incomes-up-since-covid-but-poorest-20-still-below-fy16-level-price-study

[5] 2024. Start.io | Low-End Android Users in India Audience. https://www.start.io/audience/low-end-android-users-in-india.

[6] T. Ahmed, R. Hoyle, K. Connelly, D. Crandall, and A. Kapadia. 2015. Privacy Concerns and Behaviors of People with Visual Impairments. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3523–3532.

[7] T. Ahmed, P. Shaffer, K. Connelly, D. Crandall, and A. Kapadia. 2016. Addressing Physical Safety, Security, and Privacy for People with Visual Impairments. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. 341–354.

[8] D. Ahmetovic, D. Sato, U. Oh, T. Ishihara, K. Kitani, and C. Asakawa. 2020. Recog: Supporting Blind People in Recognizing Personal Objects. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[9] Adekunle Anthony Akinrinmade, Emmanuel Adetiba, Joke A Badejo, and Oluwadamilola Oshin. 2023. NaijaFaceVoice: A Large-Scale Deep Learning Model and Database of Nigerian Faces and Voices. *IEEE Access* 11 (2023), 58228–58243.

[10] R. Alharbi, P. Lor, J. Herskovitz, S. Schoenebeck, and R.N. Brewer. 2024. Misfitting with AI: How Blind People Verify and Contest AI Errors. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–17.

[11] A. Arora, M. Barrett, E. Lee, E. Oborn, and K. Prince. 2023. Risk and the Future of AI: Algorithmic Bias, Data Colonialism, and Marginalization. *Information and Organization* 33, 3 (2023), 100478.

[12] Devanuj Kanta Balkrishan and Anirudha Joshi. 2013. Technology adoption by 'emergent' users: the user-usage model. In *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction* (Bangalore, India) *(APCHI '13)*. Association for Computing Machinery, New York, NY, USA, 28–38. doi:10.1145/2525194.2525209

[13] Andrei Barbu, David Mayo, Julian Alverio, William Luo, Christopher Wang, Dan Gutfreund, Josh Tenenbaum, and Boris Katz. 2019. Objectnet: A large-scale bias-controlled dataset for pushing the limits of object recognition models. *Advances in neural information processing systems* 32 (2019).

[14] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.

[15] Nicola J Bidwell, Simon Robinson, Elina Vartiainen, Matt Jones, Masbulele Jay Siya, Thomas Reitmaier, Gary Marsden, and Mounia Lalmas. 2014. Designing social media for community information sharing in rural South Africa. In *Proceedings of the Southern African Institute for Computer Scientist and Information Technologists Annual Conference 2014 on SAICSIT 2014 Empowered by Technology*. 104–114.

[16] J.P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R.C. Miller, R. Miller, A. Tatarowicz, B. White, S. White, and T. Yeh. 2010. VizWiz: Nearly Real-Time Answers to Visual Questions. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*. 333–342.

[17] S.M. Branham, A. Abdolrahmani, W. Easley, M. Scheuerman, E. Ronquillo, and A. Hurst. 2017. "Is Someone There? Do They Have a Gun?" How Visual Information about Others Can Improve Personal Safety Management for Blind Individuals. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. 260–269.

[18] V. Braun and V. Clarke. 2012. *Thematic Analysis.* American Psychological Association.

[19] D.D. Brilli, E. Georgaras, S. Tsilivaki, N. Melanitis, and K. Nikita. 2024. Airis: An AI-Powered Wearable Assistive Device for the Visually Impaired. In *2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*. 1236–1241.

[20] Tai-Yin Chiu, Yinan Zhao, and Danna Gurari. 2020. Assessing image quality issues for real-world problems. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3646–3656.

[21] Sasha Costanza-Chock. 2018. Design justice: Towards an intersectional feminist framework for design theory and practice. (2018).

[22] Sasha Costanza-Chock. 2020. *Design justice: Community-led practices to build the worlds we need.* The MIT Press.

[23] Nick Couldry and Ulises A Mejias. 2019. Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & new media* 20, 4 (2019), 336–349.

[24] T. De Vries, I. Misra, C. Wang, and L. Van der Maaten. 2019. Does Object Recognition Work for Everyone?. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 52–59.

[25] Remi Denton, Alex Hanna, Razvan Amironesei, Andrew Smart, Hilary Nicole, and Morgan Klaus Scheuerman. 2020. Bringing the people back in: Contesting benchmark machine learning datasets. *arXiv preprint arXiv:2007.07399* (2020).

[26] David M Frohlich, Ramnath Bhat, Matt Jones, Mounia Lalmas, Maxine Frank, Dorothy Rachovides, Roger Tucker, and Kiriaki Riga. 2009. Democracy, design, and development in community content creation: Lessons from the StoryBank project. *Information Technologies & International Development* 5, 4 (2009), pp–19.

[27] David M Frohlich, Dorothy Rachovides, Kiriaki Riga, Ramnath Bhat, Maxine Frank, Eran Edirisinghe, Dhammike Wickramanayaka, Matt Jones, and Will Harwood. 2009. StoryBank: mobile digital storytelling in a development context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1761–1770.

[28] W. Gaviria Rojas, S. Diamos, K. Kini, D. Kanter, V. Janapa Reddi, and C. Coleman. 2022. The Dollar Street Dataset: Images Representing the Geographic and Socioeconomic Diversity of the World. In *Advances in Neural Information Processing Systems*, Vol. 35. 12979–12990.

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

[30] Mary L Gray and Siddharth Suri. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass.* Harper Business.

[31] M. Grayson, A. Thieme, R. Marques, D. Massiceti, E. Cutrell, and C. Morrison. 2020. A Dynamic AI System for Extending the Capabilities of Blind People. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–4.

[32] D. Gurari, Q. Li, C. Lin, Y. Zhao, A. Guo, A. Stangl, and J.P. Bigham. 2019. VizWiz-Priv: A Dataset for Recognizing the Presence and Purpose of Private Visual Information in Images Taken by Blind People. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 939–948.

[33] D. Gurari, Q. Li, A.J. Stangl, A. Guo, C. Lin, K. Grauman, J. Luo, and J.P. Bigham. 2018. VizWiz Grand Challenge: Answering Visual Questions from Blind People. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3608–3617.

[34] A. Hagerty and I. Rubinov. 2019. Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence. *arXiv preprint arXiv:1907.07892* (2019).

[35] Josh Howarth. 2021. How Many People Own Smartphones? (2024-2029). https://explodingtopics.com/blog/smartphone-stats

[36] Gesu India, Martin Grayson, Daniela Massiceti, Cecily Morrison, Simon Robinson, Jennifer Pearson, and Matt Jones. 2024. *The ORBIT (Object Recognition for Blind Image Training)-India Dataset.* doi:10.5281/zenodo.12608444

[37] Gesu India, S. Robinson, J. Pearson, C. Morrison, and M. Jones. 2025. Exploring the Experiences of Individuals Who are Blind or Low-Vision Using Object-Recognition Technologies in India. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, Yokohama, Japan, 11. doi:10.1145/3706598.3713107

[38] Lilly Irani. 2016. The hidden faces of automation. *XRDS: Crossroads, The ACM Magazine for Students* 23, 2 (2016), 34–37.

[39] H. Kacorri, K.M. Kitani, J.P. Bigham, and C. Asakawa. 2017. People with Visual Impairment Training Personal Object Recognizers: Feasibility and Challenges. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5839–5849.

[40] R. Kamikubo, H. Kacorri, and C. Asakawa. 2024. "We Are at the Mercy of Others' Opinion": Supporting Blind People in Recreational Window Shopping with AI-Infused Technology. In *Proceedings of the 21st International Web for All Conference*. 45–58.

[41] Rie Kamikubo, Kyungjun Lee, and Hernisa Kacorri. 2023. Contributing to accessibility datasets: Reflections on sharing study data by blind people. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[42] R. Kamikubo, L. Wang, C. Marte, A. Mahmood, and H. Kacorri. 2022. Data Representativeness in Accessibility Datasets: A Meta-Analysis. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–15.

[43] Rie Kamikubo, Farnaz Zamiri Zeraati, Kyungjun Lee, and Hernisa Kacorri. 2024. AccessShare: Co-designing Data Access and Sharing with Blind People. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–16.

[44] A. Karamolegkou, P. Rust, Y. Cao, R. Cui, A. Søgaard, and D. Hershcovich. 2024. Vision-Language Models under Cultural and Inclusive Considerations. *arXiv preprint arXiv:2407.06177* (2024).

[45] Lauren Klein and Catherine D'Ignazio. 2024. Data Feminism for AI. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 100–112. doi:10.1145/3630106.3658543

[46] MINISTRY OF LAW and JUSTICE OF INDIA. 2023. Digital Personal Data Protection Act, 2023 (DPDPA). https://www.meity.gov.in/static/uploads/2024/06/2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf.

[47] K. Lee, J. Hong, S. Pimento, E. Jarjue, and H. Kacorri. 2019. Revisiting Blind Photography in the Context of Teachable Object Recognizers. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 83–95.

[48] Jabez Magomere, Shu Ishida, Tejumade Afonja, Aya Salama, Daniel Kochin, Yuehgoh Foutse, Imane Hamzaoui, Raesetje Sefala, Aisha Alaagib, Samantha Dalal, et al. 2025. The World Wide recipe: A community-centred framework for fine-grained data collection and regional bias operationalisation. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. 246–282.

[49] D. Massiceti, C. Longden, A. Slowik, S. Wills, M. Grayson, and C. Morrison. 2024. Explaining CLIP's Performance Disparities on Data from Blind/Low Vision Users. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12172–12182.

[50] D. Massiceti, L. Zintgraf, J. Bronskill, L. Theodorou, M.T. Harris, E. Cutrell, C. Morrison, K. Hofmann, and S. Stumpf. 2021. ORBIT: A Real-World Few-Shot Dataset for Teachable Object Recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10818–10828.

[51] Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology* 33 (2020), 659–684.

[52] Youssef Mohamed, Mohamed Abdelfattah, Shyma Alhuwaider, Feifan Li, Xiangliang Zhang, Kenneth Ward Church, and Mohamed Elhoseiny. 2022. Artelingo: A million emotion annotations of wikiart with emphasis on diversity over language and culture. *arXiv preprint arXiv:2211.10780* (2022).

[53] Shravan Nayak, Kanishk Jain, Rabiul Awal, Siva Reddy, Sjoerd Van Steenkiste, Lisa Anne Hendricks, Aishwarya Agrawal, et al. 2024. Benchmarking vision language models for cultural understanding. *arXiv preprint arXiv:2407.10920* (2024).

[54] World Health Organization. 2020. Blindness and Vision Impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[55] Iroro Orife, Julia Kreutzer, Blessing Sibanda, Daniel Whitenack, Kathleen Siminyu, Laura Martinus, Jamiil Toure Ali, Jade Abbott, Vukosi Marivate, Salomon Kabongo, et al. 2020. Masakhane–machine translation for Africa. *arXiv preprint arXiv:2003.11529* (2020).

[56] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems* 35 (2022), 27730–27744.

[57] J. Pearson, S. Robinson, T. Reitmaier, M. Jones, S. Ahire, A. Joshi, D. Sahoo, N. Maravi, and B. Bhikne. 2019. StreetWise: Smart Speakers vs Human Help in Public Slum Settings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[58] Madelaine Plauche, Udhyakumar Nallasamy, Joyojeet Pal, Chuck Wooters, and Divya Ramachandran. 2006. Speech recognition for illiterate access to information and technology. In *2006 International conference on information and communication technologies and development*. IEEE, 83–92.

[59] Madeline Plauché and Udhyakumar Nallasamy. 2007. Speech interfaces for equitable access to information technology. *Information Technologies & International Development* 4, 1 (2007), pp–69.

[60] Marie-Therese Png. 2022. At the tensions of south and north: Critical roles of global south stakeholders in AI governance. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1434–1445.

[61] Thomas Reitmaier, Electra Wallington, Dani Kalarikalayil Raju, Ondrej Klejch, Jennifer Pearson, Matt Jones, Peter Bell, and Simon Robinson. 2022. Opportunities and challenges of automatic speech recognition systems for low-resource language speakers. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–17.

[62] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–37.

[63] A. Seth, S. Ahuja, K. Bali, and S. Sitaram. 2024. DOSA: A Dataset of Social Artifacts from Different Indian Geographical Subcultures. *arXiv preprint arXiv:2403.14651* (2024).

[64] T. Sharma, A. Stangl, L. Zhang, Y.Y. Tseng, I. Xu, L. Findlater, D. Gurari, and Y. Wang. 2023. Disability-First Design and Creation of a Dataset Showing Private Visual Information Collected with People Who Are Blind. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[65] Vicki Squires. 2023. Thematic analysis. In *Varieties of qualitative research methods: Selected contextual perspectives*. Springer, 463–468.

[66] Chatham House – International Affairs Think Tank. 2023. Measuring the Power of the Global South. https://www.chathamhouse.org/publications/the-world-today/2024-02/measuring-power-global-south.

[67] L. Theodorou, D. Massiceti, L. Zintgraf, S. Stumpf, C. Morrison, E. Cutrell, M.T. Harris, and K. Hofmann. 2021. Disability-First Dataset Creation: Lessons from Constructing a Dataset for Teachable Object Recognition with Blind and Low Vision Data Collectors. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.

[68] Contributors to Wikimedia projects. 2004. List of languages by number of native speakers in India - Wikipedia. https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers_in_India [Online; accessed 2025-09-03].

[69] Contributors to Wikimedia projects. 2006. Classification of Indian cities - Wikipedia. https://en.wikipedia.org/wiki/Classification_of_Indian_cities

[70] Contributors to Wikimedia projects. 2016. Unified Payments Interface - Wikipedia. https://en.wikipedia.org/wiki/Unified_Payments_Interface [Online; accessed 2025-11-10].

[71] Yu-Yun Tseng, Tanusree Sharma, Lotus Zhang, Abigale Stangl, Leah Findlater, Yang Wang, and Danna Gurari. 2025. Biv-priv-seg: Locating private content in images taken by people with visual impairments. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 430–440.

[72] A. Vashistha, E. Brady, W. Thies, and E. Cutrell. 2014. Educational Content Creation and Sharing by Low-Income Visually Impaired People in India. In *Proceedings of the Fifth ACM Symposium on Computing for Development*. 63–72.

[73] A. Vashistha, P. Sethi, and R. Anderson. 2018. BSpeak: An Accessible Voice-Based Crowdsourcing Marketplace for Low-Income Blind People. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13. doi:10.1145/3173574.3173631

[74] L.Y. Wen, C. Morrison, M. Grayson, R.F. Marques, D. Massiceti, C. Longden, and E. Cutrell. 2024. Find My Things: Personalized Accessibility through Teachable AI for People Who Are Blind or Low Vision. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–6. doi:10.1145/3613905.3648641

[75] P.K. Yu. 2020. The Algorithmic Divide and Equality in the Age of Artificial Intelligence. *Florida Law Review* 72 (2020), 331.

[76] Chunting Zhou, Pengfei Liu, Puxin Xu, Srini Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, Susan Zhang, Gargi Ghosh, Mike Lewis, Luke Zettlemoyer, and Omer Levy. 2023. LIMA: Less Is More for Alignment. arXiv:2305.11206 [cs.CL] https://arxiv.org/abs/2305.11206