Disability-First Design and Creation of A Dataset Showing Private Visual Information Collected With People Who Are Blind

Tanusree Sharma*
tsharma6@illinois.edu
Informatics, University of Illinois Urbana Champaign
Champaign, IL, USA

Abigale Stangl
Human Centered Design and Engineering, University of Washington
Seattle, USA
astangl@uw.edu

Lotus Zhang
Human Centered Design and Engineering, University of Washington
Seattle, USA
hanziz@uw.edu

Yu-Yun Tseng
Computer Science, University of Colorado
Boulder, Colorado, USA
yu-yun.tseng@colorado.edu

Inan Xu
Computer Science, University of California at Santa Cruz
santa cruz, California, USA
inxu@ucsc.edu

Leah Findlater
Human Centered Design and Engineering, University of Washington
Seattle, USA
leahkf@uw.edu

Danna Gurari
Computer Science, University of Colorado
Boulder, Colorado, USA
danna.gurari@colorado.edu

Yang Wang
Information Sciences, University of Illinois at Urbana-Champaign
Champaign, USA
yvw@illinois.edu

Figure 1: An example of a prop object of private visual content, a pill bottle, at different stages of the dataset creation process: making the prop object, packaging it to send to participants, an image captured by a blind participant, and shown in a larger set of images in the final dataset, demonstrating the diversity of the captured content.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580922

ABSTRACT

We present the design and creation of a disability-first dataset, **"BIV-Priv,"** which contains **728** images and **728** videos of **14** private categories captured by 26 blind participants to support downstream development of artificial intelligence (AI) models. While best practices in dataset creation typically attempt to eliminate private content, some applications *require* such content for model development. We describe our approach in creating this dataset with

private content in an ethical way, including using *props* rather than participants' own private objects and balancing multi-disciplinary perspectives (e.g., accessibility, privacy, computer vision) to meet the tangible metrics (e.g., diversity, category, amount of content) to support AI innovations. We observed challenges that our participants encountered during the data collection, including accessibility issues (e.g., understanding foreground vs. background object placement) and issues due to the sensitive nature of the content (e.g., discomfort in capturing some props such as condoms around family members).

CCS CONCEPTS

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

KEYWORDS

dataset, accessibility, privacy, personal visual data, private visual content, visual assistance, visual interpretation, image description, computer vision, visual impairments, blind

ACM Reference Format:
Tanusree Sharma, Abigale Stangl, Lotus Zhang, Yu-Yun Tseng, Inan Xu, Leah Findlater, Danna Gurari, and Yang 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 (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3580922

1 INTRODUCTION

The development of artificial intelligence (AI) models is predicated on having access to *data*. That is because data is needed to *evaluate* whether models perform well and, for the case of machine learning (ML) based AI methods, to *train* the models. A trend for creating visual datasets, i.e., consisting of images or videos, is to remove or alter private visual information [22, 64] to protect people's privacy. Consequently, most vision datasets are limited to supporting the development of AI models that analyze common items, such as cups, animals, and furniture. Yet, there is a need for datasets with private visual content to support the creation of privacy-preserving AI models. In this paper, we aim to support the development of AI models (through ethical dataset creation), which can enable blind people to independently learn about private visual content captured in their images and videos (e.g., faces, medical prescriptions, addresses). This has the potential to benefit the estimated 39 million people worldwide who are blind [59], including nearly 8 million in the US [4].

Developing datasets containing private content to support our use case is challenging [18, 84]. Not only must data be collected that shows private visual content, but the data needs to be captured from blind individuals themselves. That is because blind and sighted users' images often exhibit different characteristics (e.g., blur, lighting, framing) [28] that, in turn, impacts the performance of downstream AI models [58].

In this paper, we present a privacy-preserving method for creating a *disability-first dataset* of private visual information and contribute lessons for future collection of sensitive datasets. Theodorou et al. [75] introduced the notion of disability-first datasets as *"An approach that is used to serve a disability community first but then could be generalized to serve all people through the innovation that it enables. It stands in opposition to mainstream ML datasets and approaches which are later augmented or co-opted to address issues of importance to disabled communities."* Collecting *private visual content* from blind users poses ethical and logistical challenges, such as how to engage blind users in contributing private visual content with consent and without violating their privacy, how to capture a diversity of content so the data can support a range of applications, and how to do so accessibly—given the known challenges with non-visual photography. We present a novel approach to address these challenges and, in doing so, examine two main research questions:

- RQ1: How to create a dataset of private visual content captured by blind users in a privacy-preserving way?
- RQ2: How do blind users think about the liabilities and opportunities with contributing images and videos they capture to a public dataset that can support the development of privacy-preserving technologies?

To address the ethical and privacy concerns, we designed a photography task so that blind participants could capture *prop* private objects in images and videos[1] that are demonstrative of real private visual content rather than participants' *own* private content. To facilitate the photography task, we created a set of 14 prop objects with fictitious private information to represent private visual content sometimes found in blind people's images and videos, mailed these props to participants, provided instructions and training for them to capture each prop with a custom app, and interviewed participants afterward. Our approach resulted in the **creation of a dataset of 728 images and 728 videos taken by 26 blind participants** who each captured 14 private objects (i.e., 28 images and 28 videos per category for 14 categories), which has the potential to support designing AI models that analyze private visual content.[2]

The dataset creation task also revealed insight into blind people's considerations about the collection and release of images and videos that contain prop private objects. For instance, participants' perceptions of what constitutes private visual content varied and they had different reactions to capturing certain prop types such as *tattoo* and *condom*. After completing the photography tasks, all participants were comfortable about the BIV-Priv dataset being released publicly and being used to develop AI tools because the data contains, as one participant said, *"Prop objects, NOT my own objects."* Nevertheless, participants encountered some challenges with the photography task. For instance, it was not clear to somehow to orient the props, avoid accidentally including their own private content in the *background* of their images and videos, or avoid others' negative impressions when they captured certain prop types (e.g., condom, pregnancy test). We provide recommendations for how to address these challenges in future dataset-creation efforts.

[1] As blind people commonly capture both images and videos when obtaining visual assistance; thus we determined that including both types of visual media to be important for the development of a range of future technologies [36].
[2] We will publicly share the dataset at https://home.cs.colorado.edu/~DrG/VisualPrivacy/BIV-Priv/

Our work makes two main contributions. First, designing and implementing our novel approach to the creation of a disability-first dataset containing with private content (e.g., creating/using props) produced lessons that should be useful for future dataset creation efforts with blind people and of sensitive information. For instance, we discuss disciplinary tensions, challenges, and trade-offs that arose when creating the dataset. Second, the dataset containing private visual content captured by blind users, which has diverse types of private visual content and compositions (foreground/background), can be used in the development of new AI models to better support blind users.

2 BACKGROUND

Below, we summarize existing literature related to the ethics of dataset creation, with a focus on challenges and methodology in collecting *private* data in a natural setting and datasets from people with disability.

2.1 Ethics of Dataset Construction

Datasets are critical to AI model development, serving as a basis to evaluate and train new models [26, 74]. A critique of prior dataset creation efforts is limited documentation about them [65], which in turn may have contributed to negative consequences from their use [17, 20, 31, 46, 55, 74]. Negative examples include intensifying discrimination, violating the value of inclusiveness and representation, and breaching legal rules (e.g., privacy, intellectual property licenses, consumer rights), including when data is obtained without proper consent (e.g., scraped from the Internet [e.g., [81]). More generally, past work categorized dataset-induced harms into two types—allocative harms (i.e., opportunities or resources are withheld from certain groups) and representational harms (i.e., certain groups are stigmatized or stereotyped) [74].

In response, recent work increasingly has affirmed the importance of conscientious data stewardship practices [17, 26, 46, 74] and proposed recommendations for ethical dataset creation in a number of aspects, including: (1) analyzing taxonomies that a dataset is based on to examine potential dataset bias [17, 47]; (2) sampling data from the target population to include traditionally underrepresented groups [49]; (3) removing data collected without proper consensus or offering people whose data was already included without consent the opportunity to voluntarily contribute the data, potentially with financial compensation [17]; and (4) transparently documenting the dataset creation process to facilitate community engagement and critical reflection [17, 26, 46, 68, 74]. Recommendations center on providing documentation that covers: objectives, composition, genealogy, collection process, distribution, maintenance, deprecation, confidentiality, licensing, auditing, target population, and representativeness of the dataset. Further, dataset creators should pay attention to: (1) documenting how identity-related data features, such as gender and race, are defined and annotated, (2) specifying limitations of use associated with them [66], and (3) constantly reflecting on their processes. Such information can, in turn, be used by consumers to inform how they decide to use the dataset [26].

Despite the many recommendations proposed in the AI community, ethical dataset creation is an ongoing challenge, especially for data that is innately sensitive or from underrepresented groups. We, therefore, take this effort one step further by exploring best practices for creating datasets consisting of private visual content from blind people.

2.2 Datasets with Private Visual Content

Privacy is a major challenge with the datasets that support AI model development [12, 60, 86]. Historically, developers of large-scale image datasets took precautions to avoid collecting private content (e.g., VOC [22], ImageNet/ILSVRC [64], COCO [45], LSUN [89], Open Images [41]). More recent efforts include automated systems that adopt elimination-, perturbation- and confusion- approaches to remove or alter identifier information from datasets before using them [86]. However, some AI applications *require* the collection and use of personally private data. For example, creating a privacy-preserving AI tool to automatically detect and obfuscate private information in images or videos [12, 44, 60, 86] requires access to data with private content for evaluation.

Developers have created numerous datasets to support the development of privacy-preserving models. One key consideration is how to collect visual data. Some curated data that directly shows private visual content (e.g., [25, 56, 57, 90, 92]); however, such data typically was scraped from online content without consent from the content contributors. Alternatively, a privacy dataset originating from blind photographers was created by only sharing the context in which private content is found to support model development (i.e., VizWiz-Priv) [29]. In other words, all private content was removed from public viewing. However, with this approach, it is impossible to evaluate how well AI models perform in analyzing private content in visual data taken by blind photographers as well as for privacy categories not yet included in other privacy datasets. Another consideration for dataset creation is how to define what *is* private visual content and should thus be included in the dataset [44]. While Li et al. [44] offer a taxonomy that includes 28 categories of sensitive content, many papers (including Li et al.) have elaborated on the challenges of specifying a universally accepted definition of what is considered private, highlighting the importance of personal and environmental contexts [39, 44, 53, 54, 70].

An alternative is to generate synthetic privacy datasets based on statistical properties extracted from real sample data [40, 61]. As it cannot be mapped back to a real person, synthetic data has become increasingly popular for producing publicly shareable AI datasets [40, 61]. However, generating synthetic data still requires collecting and learning from real private data, which not only imposes privacy leakage risk in the process but is challenging for situations where the original data is difficult to obtain in the first place. Further, synthesized data is not always representative and may produce models with poor performance [40, 71]. These performance and privacy risks can be even more pronounced for underrepresented groups due to the lack of data and existing societal bias [86].

We aim to advance methods for ethically creating datasets to enable privacy-preserving tools that work well for underrepresented groups. Our novel data collection method does not rely on computer-generated data but instead is the first to employ physical 'prop' objects. Additionally, we focus on the unique needs of one underrepresented group–blind people–as data contributors. We drew on the aforementioned best practices for constructing privacy datasets [39, 44, 70].

2.3 Disability-First Datasets and Blind Visual Data Contributors

ML algorithms can disproportionately harm underrepresented groups *"along the intersecting axes of race, ethnicity, gender, ability, and position in global hierarchies"* [46]. Among these many groups, people with disabilities are often one of the earliest adopters of ML technologies [15], but at the same time, one of the most at risk of potential downstream harms [51, 58]. Intentionally developed *inclusive* ML datasets offer a way to mitigate this harms [84]. To encourage more data contributed directly from people with disabilities, Theodorou et al. [75] proposed the concept of "Disability-First Dataset", which guides our dataset creation effort (as described in the introduction).

The *Incluset* repository [35] documents a range of existing accessibility-related datasets, yet disability-first datasets are scarce. A core aspect to creating disability-first datasets is ensuring that the experience and perspectives of people with disabilities are represented within, which is challenging at scale [35, 58]. The *"long-tail nature of various disabilities in a population"* and *"the heightened need to protect the sensitive demographic attributes related to disabilities"* further contributes to the dearth of disability-first datasets [58]. To overcome these challenges, some efforts have simulated data from people with disability, which can yield inaccurate results and reinforce social stigmas and stereotypes [51, 76, 84]. Instead, Park et al. [58] took a more human-centered approach by interviewing people with disabilities to understand their motivations, concerns, and challenges with contributing data, resulting in three factors to consider when collecting data from people with disabilities: 1) the motivation of contributors, 2) what and how to communicate about the dataset collection goal, and 3) how to create an accessible technical architecture and user experience.

Constructing disability-first visual datasets with people who are blind—the focus of this paper—presents even more complexity. First, blind people's images and videos commonly differ from those taken by sighted people—such as being blurred, having inconsistent lighting or focus, and having objects partially or fully out-of-frame [29, 30] due to challenges associated with non-visual photography [10, 33]. Consequently, ML models trained on data from sighted users often produce inaccurate image captions or visual question answers on blind people's images [30]. Researchers have thus introduced two datasets of images and/or videos captured *by* blind people: the VizWiz dataset, which includes images and visual questions [30], and the Orbit dataset, which includes 3,822 videos of 486 objects [48]. While useful for many ML applications, these datasets by design do not include private visual content.

A second challenge is that blind people who want to contribute images and videos to a dataset may not be able to independently assess that there is no *unintended* private content in the image/video [29, 48]. Reliance on sighted people for inspection presents privacy liabilities and concerns for blind people [72]. In this paper, we present our approach to creating a dataset of private visual content collected by blind people, including how we addressed the aforementioned and other related ethical concerns. We draw on guidelines for non-visual data collection [29, 48] and blind people's visual privacy concerns during visual assistance [73] to define which types of private visual content to include as props in our disability-first, non-visual, and privacy-preserving dataset collection effort.

3 BIV-PRIV PHOTOGRAPHY TASK DESIGN, MATERIALS, AND TECHNICAL INFRASTRUCTURE

We present our approach to creating a dataset called ***B****lind People's* ***I****mages and* ***V****ideos of* ***Priv****ate Content* (BIV-Priv), which is the outcome of a team of multi-disciplinary researchers' (computer vision, accessible computing, and usable privacy and security) year-long, collaborative effort [93]. Our goal from the outset was to engage blind participants as contributors to a privacy dataset in a meaningful, accessible, and privacy-preserving fashion. Below, we describe the primary aspects involved in our approach.

3.1 Guiding Questions for the Data Collection Design

3.1.1 What categories of private visual content to include? We composed an initial list of 39 categories that prior work reports are the types of private content found in blind people's images [29, 44, 48, 73]. We then eliminated 25 of those object types through a series of considerations: (1) avoid possible risks to researchers who create the props like breaking the law (e.g., making fake official government documents), (2) avoid possible risks to participants (e.g., distributing controlled substances such as inhalers, insulin pens, guns, alcohol), (3) avoid duplication of data that could be found in existing datasets of images or videos captured by blind people [29, 48], (4) avoid duplication of data categories that state of art computer vision algorithms are already able to recognize (e.g., nudity and human faces) [29], (5) avoid socially sensitive or stigmatized objects (e.g., sex toys) [73], and finally (6) avoid collecting categories that proved to be overly intensive with respect to labor, cost, time, or material availability (e.g., insurance cards, clothing with logo, local street signs).[3] Our final set included 14 categories: local newspapers, bank statements, bills or receipts, business cards, condom boxes, credit or debit cards, doctor's prescriptions, letters with addresses, medical record documents, pregnancy tests, empty pill bottles, tattoo sleeves, transcripts, mortgage or investment reports.

3.1.2 What number and diversity of images and videos should be collected? We consulted prior work in AI model development to inform how many images and videos to collect. While mainstream AI video analysis development often relies on hundreds of short video clips (i.e., typically less than 10 seconds in duration) [62, 87], mainstream AI image analysis development often relies on thousands to millions of examples [45, 63]. Due to practicality constraints, for the latter image-based scenario, we chose to instead support the mainstream few-shot learning framework for which

[3] More details about how we excluded private visual content types initially identified can be found in the Supplementary Materials.

hundreds of example images are sufficient to support model evaluation [13, 23, 43, 50, 52, 67, 77]).[4] With that said, our work can still be used to support AI model development in other AI frameworks. In total, we aimed to collect hundreds of images and hundreds of videos.

When creating a dataset to analyze any type of content, the other critical aspect is to capture target content under a *diversity of appearances* (e.g., lighting, angle, setting, etc.) [45, 63]. This supports developing more robust AI models that can handle the diversity of appearances content can manifest in real-world applications. We aimed to achieve greater diversity by instructing our study participants to capture images and videos with the private content props first positioned in the *foreground* and in the *background* of the images/videos; blind people's images commonly present the private information in both the foreground and background of images and videos [30]. To further provide diversity, we considered the amount of participants to recruit based on the labor we anticipated the task requiring; more participants would distribute the effort needed while helping to further ensure a diversity of compositions and appearances of the prop categories in the background and foreground of their images and videos.

In total, we set out to recruit 30 participants and so collect 840 images and 840 videos with 28 images/videos per each of 14 private categories by 30 participants. This would enable our work to meet the criteria commonly used in the computer vision community with regards to number of images/videos per category, diversity, and number of categories.

3.2 Designing, Creating, and Distributing the Props of Private Content

We aimed to collect images and videos showing private visual content from blind participants in their place of residence (or where they might naturally capture images and videos of private visual content), but without real personal information involved. Therefore, we decided to instead send each participant a package containing all prepared 'private content' for them to photograph. We call each item in each package a *prop* to stress that the items did not belong to the photographers. In total, we focused on 14 types of props comprised of private text on paper (e.g., bank statement, medical record) and physical objects (e.g., condom, pregnancy test, tattoo sleeve, credit card), some of which included a combination of a physical object with private information on a piece of sticker (e.g., pill bottle with a text label).

Of note, the use of props for generating privacy datasets is one of the key novelties of our work. In subsequent sections, we discuss how we explore the significance of our prop object design work for private dataset creation, through applied interviews, surveys, and photography tasks.

3.2.1 Why Props? Various privacy risks could arise for our blind participants if they took pictures/videos that contained their own private/sensitive content and the resulting dataset was released publicly. All types of privacy threats presented in Solove's privacy taxonomy[69] (those regarding information collection, processing and dissemination as well as invasion) could be applicable in that case. In contrast, by using props rather than participants' own (private) objects, most of the aforementioned privacy threats can be minimized because the private/sensitive data was fictitious. For instance, our pregnancy test props used real objects (pregnancy test kits) but with fictitious content (faked test results) and thus can protect our participants from privacy risks related to information processing (e.g., insecurity, the pregnancy test results were leaked/hacked during the processing) and information dissemination (e.g., disclosure, disclosing the pregnancy test results in the downstream use of the dataset) because the test results are faked. However, there might be some residual privacy risks after using props (e.g., interrogation, where family members might ask/interrogate them about why they had pregnancy test kits). We will discuss these potential residual privacy threats in Section 5.5.

3.2.2 Prop Design Materials. A key challenge with creating props was determining how realistic they could be in representing the real private category. For some categories, we purchased real objects (e.g., condoms, newspapers). When not possible though, we identified resources for creating prop objects with a near identical visual appearance to real objects. For instance, when identifying resources to create *credit or debit card* props, we considered four approaches: (1) ask banks to send out simulated cards with fake numbers and names; (2) collect expired cards; (3) hand-make credit cards; and (4) order credit cards from sources, such as, Etsy [2], with simulated card information. We determined that it is infeasible to find a bank that would produce a small batch of fake cards, that expired cards would have real private visual content in them, and that handmade cards would lack the accuracy needed. In turn, we chose to move forward with ordering cards from Etsy that had high fidelity (option 4). Similarly, when making a prop of private visual content type *Tattoo*, we had several ideas: (1) purchase tattoo stickers and place them on top of a figure, doll, or mannequin; (2) capture and send a video of someone with a tattoo (with permission) and ask participants to take a picture of the visual media, and (3) purchase tattoo sleeves. We decided to purchase tattoo sleeves (option 3), having determined that we did not want the doll with the tattoo to be the object of focus and that capturing video content of the private visual content would be difficult and would provide low-resolution data.

Consequently, different props entailed different types of effort to support. For example, for *local newspapers*, we purchased newspapers from different cities to increase the diversity in the locations, city names, and states. For *pregnancy tests*, we purchased two types of pregnancy test kits where half of the tests were *fake positive* and the other half *fake negative*. Other props entailed filling in existing blank templates. For instance, for the private visual content type *pill bottle*, we used Photoshop [6] and PDF Filler [5] to design and fill a prop label with fictitious information, printed it on a sticky label, and stuck it onto one of many different colored bottles that we ordered from Amazon.

3.2.3 Persona-Based Text in Place of Personally Identifiable Information. For props showing private text, we added fabricated personally identifiable information. Ten of the 14 props included such text, which is exemplified on the pill bottle with the prescription label in Figure 1. To generate the text, we made-up a persona per study

[4] Few shot learning algorithms can be trained using images in the VizWiz datasets (e.g., [77]), which only show non-private content.

participant (i.e., 30 personas), each with a set of unique information such as a fictional name, address, and phone number. We used online software to generate some of this fictitious data, such as randomlist.com [7] to generate addresses and fakenumber.org [3] for phone numbers. We personally created other fictitious specialized details such as a medication type for *prescription bottle* props and a bank account number for *bank statement* props. For example, for *medical information*, multiple members of the research team collaboratively populated the personas by each visiting various health-related websites to find legitimate medical details to use for medical records and pill bottle labels; i.e., we created 30 different labels for pill bottles with information such as the patient's name, address, doctor's name (prescribed by), the pill quantity, pill name, instructions for taking the pill. To create greater diversity in the props, we chose six *medical record* templates to insert the information into, each formatted with different styles (both traditional black and white ones and the more contemporary colorful ones), health service provider logos, and names.[5] More generally, we attempted to create *diversity* in all persona information, such as age, gender, occupation, educational status, marital status, and financial information. For instance, for credit card information details of "holder since" and "valid till", we made the years span all decades and distributed months across the 30 personas to avoid biasing the dataset to specific numbers or years. Still, some of the information had constraints of being US-based, such as address, phone number, and educational institutes.

3.2.4 Packaging and Distributing Props to Blind Participants. Each prop was placed in a separate plastic bag with two types of labels attached to enable participants to identify each type of prop: braille and QR codes that could be scanned with our data collection app (Section 3.3.3). Both braille and QR codes are commonly used to make artifacts accessible in various settings, including museum displays to support multiple types of non-visual literacies [27, 32]. We contracted a professional braille designer [1] to produce braille labels to specify the name of each prop we included in the set. In total, we designed, made, and packaged 420 props for the main study and 42 for the pilot study. We then mailed 30 packages of *prop objects* to our 30 participants.

3.3 BIV-Priv Photography Task Instructional Design and Training Protocol

Prior to the main study efforts, we included a live training session to train participants on how to contribute. We now describe the design of the instructional materials (Section 3.3.1) and the protocol for training participants (Section 3.3.2).

3.3.1 Instructional Design. We consulted prior work on best practices for supporting disability-first dataset creation [75] and non-visual photography [34] to gain insights into how to (1) describe the purpose of the task (e.g., engaging people with disabilities in data creation phase as contributors), (2) explain why non-visual photography skills are needed (e.g., representative dataset to serve the community and current absence of private data from blind photographers), (3) balance the amount of effort required for participants to accomplish the task while ensuring diversity within the images/videos each participant would capture (e.g., captured from different angles), and (4) ensure the technical infrastructure we created supported the overall aim (e.g., accessibility guidelines).

Initially, we followed the instructions proposed by the ORBIT dataset creators for guiding blind participants in independently completing photography tasks [75]. Our only modifications were to incorporate unique characteristics of our study, such as instructing participants to wear the tattoo sleeves on their hands before data capture. Following our pilot study, we then refined the instructions based on questions from the participants. For example, to address the most common question of how to take images and videos with objects in the *foreground* vs. *background*, we added definitions of each to our instructions but did not instruct participants *how* to do the foreground or background object capture in order to preserve individuals natural photography skills. More generally, we aimed to design the instructions to simultaneously support task success without influencing our participants' natural photography practices.

The instructions[6] have three main sections. The first section introduced the overall *purpose of the study* (i.e., the use case of an AI-based end-user tool which can potentially identify and remove private information in images or videos captured by blind people and thus protect their privacy when sharing images/videos). In this section, we also emphasized the importance of variation in the data. Second, the instructions included a brief description of the implications of private images/videos captured by blind users in building the tool. Third, we created *step-by-step images/videos filming instructions* for participants to learn how to: (1) install the data collection app, (2) unbox the prop objects, (3) identify the prop objects with the braille label or QR code and select it from the list in the app, (4) set up the environment for foreground and background images/videos, and (5) review and edit if there is any unintentional private content in the images/videos. Each participant was asked to take photos and videos of the 14 private visual content types within 24 hours of the training session (Section 3.3.2).

3.3.2 Required Training Session. A training session via Zoom (30-60 minutes) was required for all participants based on pilot participant feedback on the usefulness of going over the test photography task with the researchers. This training covered: (1) reading aloud the photography instructions from the website and Q/A, (2) performing the test photography task using a coffee mug as a test object to become familiar with the app, and (3) identifying one of the prop objects with either the braille label or the QR code within the app.

3.3.3 Mobile App for Creating the BIV-Priv Dataset. We developed an accessible data collection iOS app for iPhone and iPad [7] so participants could use their own devices, in their everyday environments, to capture and upload images/videos of prop objects. We intentionally designed the app for use with iOS's built-in screen reader, called VoiceOver [8].

[5] The supplementary material includes an example of medical record prop object for a given persona.

[6] https://biv-priv.github.io/

[7] https://github.com/BIV-Priv/DataCollection-App

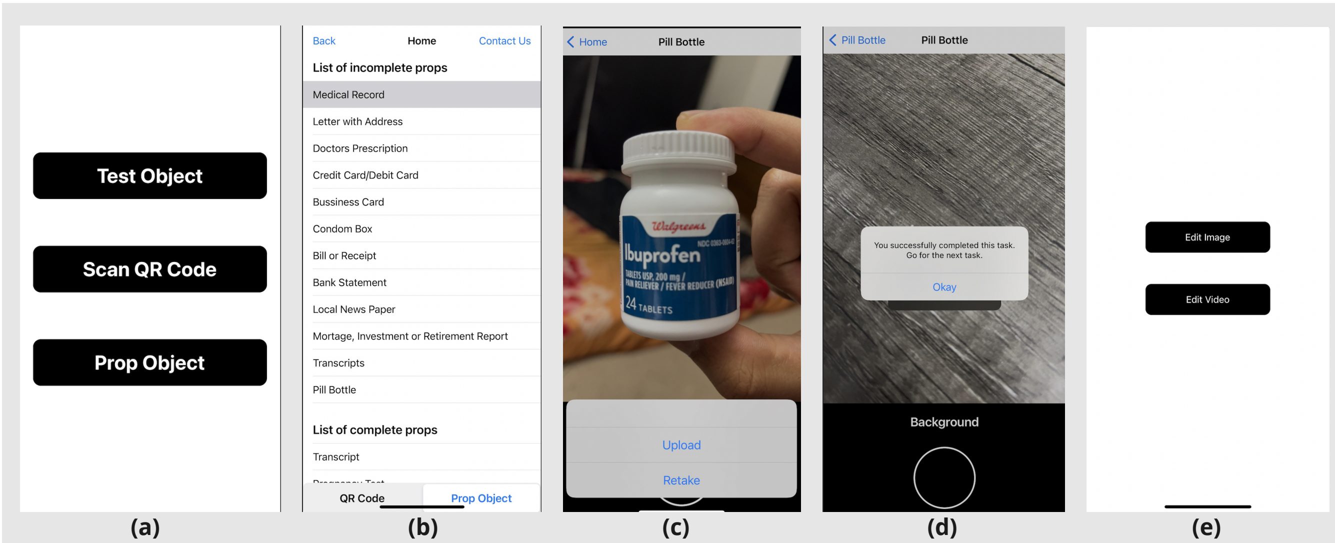

Figure 2: The data collection app BIV-Priv: (a) prop identification screen, (b) prop selection screen, (c) foreground/background images and videos recording screen, (d) images and videos upload screen with task confirmation, and e) edit images/videos screen if recapture needed. All screens are marked up with the adapted information hierarchy and touch-targets of the accessibility interface.

The BIV-Priv App guides users through the photography task, as shown in Figure 2. The user journey included on-boarding, identifying private visual content, selecting private visual content, capturing props, editing images and videos. Screen 'a' (Figure 2) allowed the participant to 1) choose an option to ***"identify prop object"*** and 2) use either the QR code or the braille label. After identifying the object type, participants were taken to ***"List of incomplete props"*** in screen 'b'. This screen keeps track of which props the participant completed and moved the props from the (1) *Incomplete Props* group, to the (2) *Completed Prop* objects group. Images and videos capture were handled in screens 'c' and 'd' in which participants could ***"capture and record the images and videos"*** for foreground and background compositions. The camera controls element gives the voice-over instructions to provide tips throughout the photography task. Once images and videos have been captured, the next screen allowed participants to *"retake"* or *"upload"*, in case the participants were not happy with the capture. It also provided cues with *"audio announcement"* in 25 seconds to indicate when the video recording should be stopped. To prevent inadvertently long videos, recordings were automatically ended after one minute. Each screen includes a button for enabling participants to ***"chat with researchers"*** in real-time for additional help. In some cases, when images and videos did not pass the validation criteria set by the researchers and so could not be added to the dataset, participants were asked to recapture. Participants then had an option to ***"edit images/videos"*** and re-upload from *"Prop object complete list"* by clicking on the intended objects and edit/upload it.

Of note, we designed this app based on feedback from two pilot studies and summarize some of our key resulting design choices here. We added the consent button on the same registration page to simplify the design in the second iteration. While our initial *"login"* design was based on real name and user password, after the pilots, we opted to begin by providing users with a *"vizid"*, rather than requiring real name/email (on the first prototype) as well as *"no password login"* to avoid participants' use of their password from real accounts to preserve their privacy to the greatest extent possible. For app *"Homescreen"*, pilot participants expected to have an in-app summary of photography instructions rather than going back to external resources, which we implemented in our second iteration. The initial prototype only had the *"PVC Identification"* (i.e., Scan QR Code, Select Prop Object) options as tabs in the bottom of the Homescreen. In contrast, after the pilots, the identification content was put on its own separate page to build this workflow more clearly. We also kept the *"Tabs"* at the bottom of the screen with both choices so users can change their preferences easily. Based on the experience of pilot participants, we implemented a real-time text option. In the second iteration, we incorporated *"Contact Us"* in each screen as a button that allows participants to chat with researchers in real-time (9 am-5 pm Central time) for additional instructions/ help.

3.4 Data Collection User Study

We designed and conducted a user study with blind participants in which they took pictures and videos with the 14 prop types. It consists of the following four key activities for the participants to perform, which are described in this section: (1) participant recruitment including a screening survey, (2) a pre-task survey, (3) a Photography Task training session, and (4) a post-task interview. The study design was informed by prior literature on dataset creation with/for blind users [29, 44, 75], infrastructure design for data collection [42, 58] and privacy in context [14, 70]. It was also approved by each participating institution's Institutional Review

Board (IRB). The images and videos collected from this study became our BIV-Priv dataset.

Participant Recruitment. At the outset of our recruitment efforts, we circulated an IRB-approved announcement on a listserv managed by organizations serving people who are blind. The announcement linked to an anonymous screening survey. It asked basic demographic questions about prospective participants' ages and locations of residence. We recruited adult blind participants (18-yr or older) living in the USA. Next, the survey asked about their current photography practices and use of visual assistance technologies to ensure (blind) participants had at least basic familiarity with cameras and capturing images and videos. The survey also asked about their willingness to take images and videos to share with the research team for the purposes of building a public disability-first dataset. We also checked their response to the consent of collecting their images/videos containing private visual content and for audio recording of Zoom sessions during the user study. The screening survey resulted in a pool of 83 blind individuals who were willing to take pictures and videos using our *"BIV-Priv App"* and share them with us.

From the pool of respondents we randomly selected 30 participants, using age as a criterion where we aimed to select 15 participants from 18 – 44 age range and 15 from 45 and older since prior research indicated that levels of photography skills may vary in different age groups [85]. We had 4 participants not complete the photography tasks and drop out of the study; 3 dropped out due to either family emergencies or issues with the iOS device while the fourth dropped out after completing the pre-survey and training session because her husband did not feel comfortable with the nature of some prop objects (e.g., condoms, pregnancy tests). In the end, we had 26 participants. Detailed demographics about our participants are shown in Table 1, with them bringing a diversity of ages, genders, timing for onset with vision loss (i.e., 70% participants were born blind and 30% had acquired blindness), and educational backgrounds (i.e., Master's degrees (46%), Bachelor's degrees (34%), some college credit (11%), Associate degree, and trade/technical). Some of this data can be used as a proxy for inferring socio-economic status; e.g., educational achievement has been shown to be correlated to socioeconomic status (SES) [83].

Pre-survey. We designed a pre-survey that included 28 questions to gain further understanding about participants' current photography practices. A subset of questions covers image/video capture frequency by blind users, challenges and experiences of sharing images/videos, applications/platforms used for sharing images/videos, and image/video sharing frequency by blind users. Some of the sample questions include- *"What applications or platforms do you use to post or share images and/or videos that you took?", "Have you ever shared private information in images/videos intentionally? If so, why and how did you know there was private information in the images/videos?"* The other subset of questions covers their (1) perception of privacy in content of image and video sharing, (2) preferences of taking/sharing images and videos with the research team. An example questions is, *"How comfortable would you be taking and sharing images and videos of prop objects with our research team [...]"*. As with all instruments, we obtained feedback from pilot testing that informed the survey design. The finalized Pre-Survey can be found in the Supplementary Materials and was administered through Qualtrics via email.

Photography Task Training . Once we selected participants and they completed the pre-survey, we asked them to read the Photography Task Instructions[8] (described in Section 3.3) and download the BIV-Priv app on their own. Once complete, we scheduled a training session (described in Section 3.3.2).

Post-Task Interview. After each participant completed the BIV-Priv Photography Task, we conducted a Post-Task Interview, in which we asked the participant about their experiences with the photography tasks as well as challenges and concerns in taking and uploading the images/videos. We also asked for their suggestions for researchers interested in collecting data for the development of privacy-preserving AI technologies and how such studies can be conducted in alignment with their values.

3.5 Data Analysis

Our data analysis involved thee types of data: Demographics and Quantitative Data, Qualitative Data, and Image and Video Data. We calculated descriptive statistics of the demographics data (e.g., age, gender, education, use of visual assistance technologies) and the pre-survey responses (e.g., self-reported attitudes, level of comfort for data collection, sharing and processing of prop images and videos).

For qualitative data from the interviews, we performed a thematic analysis [24]. Two authors performed open coding and deductive analysis to explore current photography practices of blind users and challenges during photography tasks. We regularly met and discussed the themes that emerged in the data, grouping data and concepts that belonged together into sub-clusters. Further abstraction of the data was performed by grouping sub-clusters into main clusters. We then organized them into high-level themes, and iterated this process to interpret the results of blind users' photography practices, challenges, and preferences in taking and sharing prop images and videos. Example themes include "future usage of their data," "local or global storage," and "individual norms of privacy."

We reviewed the collected data for any potential issues and emailed participants if we needed them to retake any images or videos. We applied three quality control criteria: (1) images and videos should not contain participants' own private content, (2) prop objects should be in the images and videos, and (3) videos should be 25 seconds long.

4 BIV-PRIV COLLECTION STUDY RESULTS

4.1 Pre-Survey: Participants' Experiences Before Data Collection

We now discuss participants' perspectives on contributing to our collection of the prop private visual content. Most study participants indicated a comfort with **sharing images and videos of prop objects with our research team** prior to starting the photography task; i.e, 21 out of 26 participants indicated they were comfortable

[8]We created a website of photography instructions for better accessibility rather than a Google document.

ID	*Gender*	*Age*	*Education*	*Assistive Technology Use*	*Image/Video Share on Social Media*	*Blindness*
P1	Female	25-34	Master's	NVDA and JAWS, VoiceOver, Braille display	Sometimes	BB
P2	Male	35-44	Master's	Tap-tap see, Seeing AI, ABBYY (OCR software), JAWs and NVDA. Seeing AI, ABBYY Fine Reader, Tap-Tap See	I go through periods of greater and lesser activity, but yes, I do share photos on social media. Sometimes, yes	BB
P3	Female	55-64	Master's	Aira	Sometimes	BB
P4	Male	45-54	Master's	Screen readers on various platforms	Sometimes, generally concerned that the contents of an image are displaying something different or more than I intended.	BB
P5	Female	45-54	Bachelor's	Screen reader, Braille display, screen recognition, Optical Character Recognition, and software that offers image description	Yes	BB
P9	Female	55-64	Bachelor's	Aira, Be My Eyes	Yes	BB
P10	Female	25-34	Trade, technical	Yes, screen reader, NVDA	No	BB
P11	Male	25-34	Some college credit	Voiceover, Jaws, NVDA	I share photos a couple of times a month	AB
P12	Female	65-74	Master's	Aira, Be My Eyes, Seeing AI, Voice Dream Scanner	No	BB
P13	Female	45-54	Bachelor's	voiceover, jaws NVDA Talkback	Yes	BB
P14	Male	35-44	Master's	Aira, Be My Eyes	Yes	BB
P15	No gender	35-44	Bachelor's	Voiceover	Yes	AB
P16	Female	55-64	Bachelor's	Jaws, NVDA, voiceover	Yes	BB
P17	Female	35-44	Master's	Jaws, VoiceOver, Seeing AI app, Aira, Be My Eyes app, Alexa	Yes	AB
P18	Male	25-34	Master's	JAWS, Brailliant BI 40, VoiceOver	No	AB
P19	Female	18-25	Some college credit	Jaws, voiceover, windows, NVDA and Narrator and a braille display	Yes	AB
P21	Male	35-44	Associates	Jaws, voiceover	Yes	BB
P22	Male	18-25	Bachelor's	VoiceOver, JAWS , screen reader	Yes	AB
P23	Male	55-64	Some college credit	Aira, Be My Eyes	Yes	BB
P24	Female	35-44	Master's	JAWS, NVDA, Voiceover, ID Mate, Braille display, Color identifier	Sometimes	BB
P26	Female	55-64	Master's	Be My Eyes, Aira, Seeing AI	Sometimes. I am reluctant if I can't get a good idea of what the photo or video is.	AB
P27	Female	65-74	Master's	Jaws	Yes	BB
P28	Female	25-34	Bachelor's	Screen reader, Braille display	Yes	BB
P29	Female	45-54	Bachelor's	JAWS, NVDA, iPhone SE 2020 with adaptive features and apps	Only if they are from somebody I know or I'm sure what they contain.	BB
P30	Female	25-34	Bachelor's	NVDA, VoiceOver	Yes	BB
P31	Female	25-34	Master's	Seeing AI, Aira	Sometimes	AB

Table 1: Participant demographics and background. (AB = acquired blindness; BB = born blind)

with sharing such data with our research team for the development of non-visual privacy-preserving photography and for publicly releasing the dataset. The four remaining participants asked for more details about data processing in removing *"accidental private content"* before public release. They were *somewhat comfortable* or *neutral* about sharing, largely stemming from a fear of accidentally capturing their own personal content due to lack of photography experience and the controversial nature or social stigma around some of the props (e.g. pregnancy tests, condoms). We addressed such concerns during the training Zoom session before participants started the photography tasks.

4.2 Participants' Experiences and Researchers' Reflections

In this section, we present participants' experiences and challenges in the photography task, their perceptions of ethical amount of labor/compensation for the study, and our reflections on the design of the *Prop Photography Task.*

4.2.1 Prop Photography Task Performance. On average, our participants took 2.2 hours to complete all photography tasks. Nine participants took 45 minutes to 1.5 hours, 14 participants took 2-3 hours, and three participants took 3.5-6 hours. Of the 26 participants, 23 were asked to retake photos. The majority of participants who were asked to retake photos had captured the blank side of a paper prop (50% of participants, 56 images, 57 videos), the object was not located in the frame (38% of participants, 33 images, 14 videos), the tattoo sleeve was not worn (53% of participants, 14 images, 14 videos), or unintentional private content such as the participant's face had been captured (8% of participants, 3 videos). Other less frequent reasons included: low light, objects not removed from plastic bag, and video recording less than 25s. Most participants attempted 2-4 times to complete the photography tasks while three participants completed the tasks in their first attempt. In general, the more prior experience they had with photography, the fewer times it took them to finish the study task.

4.2.2 Challenges Encountered During Photography Tasks and Solutions. Participants encountered several challenges during the photography task which were observed through the images and videos captured, reported through direct messages from participants through the app, and expressed during the post-interviews.

First, we observed challenges with collecting high-quality data. For example, some of the images and videos were blurry or had poor lighting, compositions, or framing, a common challenge of non-visual photography [79]. Participants also shared that knowing how to physically orient themselves to the props when capturing images and videos was challenging, another known challenge for non-visual photography [80, 91]. For instance, P03 explained, *"I'm not a person who takes videos and especially with handheld. It was really challenging about this, I kept thinking, 'Am I moving it too much, or am I not moving it enough?"* This suggests that the BIV-Priv App could include instructional messages to assist users by identifying the orientation in which they are holding the camera in relation to the object of interest. We also found two videos and three images (from two participants) which accidentally contained their own private content, and so deleted such data from our storage to avoid any privacy risks.

The second challenge was how to capture the props in the foreground and background of the images and videos. P2 noted - *"I made sure that there was nothing private in the background. So I felt perfectly comfortable regarding sharing those images and videos with the research team. However, that was challenging and took a lot of time."* The post-interviews revealed two insights: (1) the photography task instructions could have benefited from including interactive references for participants to understand the concepts of foreground vs. background; and (2) asking participants to capture the props in their images and videos in naturalistic settings (what the Orbit project did [75]), albeit without capturing their own private visual content, required an extensive amount of information management on their parts. Most participants said they set up their environment to avoid capturing their own private content in the field of view.

Third, participants struggled which side of paper props showed text (as opposed to a blank page). In fact, nine participants reported asking for help from sighted friends or family members to help with this, 17 participants used visual assistive technologies, and three participants used a combination of both. A related challenge arose after they captured the images and videos and they wanted to confirm if they had captured the text side of the paper props. All participants reported using visual assistance technologies, mostly Seeing AI, but also Viewfinder, Envision AI, OCR scanner, Aira, iPhone camera, and Supersense.

Finally, some participants experienced a sense of social stigma when completing the task, which delayed their completion of the task (i.e. outside of the 24-hour period requested in the instructions). In such cases, participants were not comfortable around other people (e.g,. niece and nephew) to perform the photography tasks due to the nature of objects, such as pregnancy test and condoms. P19 noted *"I was not comfortable to take photos of some of the props when I had some guest in my place. I had to wait. You know, even, it's a little uncomfortable to capture when my niece was around. It took a while for me to complete all the uploads."* One of the four participants who dropped out after completing pre-survey and training zoom session indicated a similar challenge; her husband did not feel comfortable with her task completion due to the nature of the prop objects. We overlooked impression management when we were designing the study because we felt they were all props. Future work should consider that privacy in the home is not just about the participants, but also the impressions of others in the home. We also recommend making the research intent more visible on the package–both visually and non-visually.

4.3 Liabilities and Opportunities of Collecting Data Showing Private Content

We now present participants' perceived opportunities and liabilities about capturing and contributing data to share in a public dataset.

4.3.1 Reasonable Amount of Compensation. We asked the participants if they felt the amount of compensation their study involvment was reasonable. Twenty-four participants indicated that the compensation was fair according to the overall study time; two of them went on to say they would even participate in the study without any compensation. Two participants noted the difficulties they faced during recapturing and constantly making sure that objects are in the frame which took more of their time than they initially expected and thus, suggested raising the compensation. In words of P17: *"It does take time. 3 phases of study, prop objects photography, exit interview. You could increase a little a bit more, for retaking some props. Was nervous if I am meeting expectations of photo criteria. Scheduling for interviews, touch base in email. Those takes effort."* Future work should consider the additional effort of possible data recapture when considering an appropriate amount of compensation.

4.3.2 Perceived Representativeness of Props as Private Content.

Nearly half (i.e., 45%) of participants indicated the props objects accurately represent their consideration of private objects (i.e., at a level of 10 where 1 is less accurate and 10 is most accurate). They explained those are *"Objects that I frequently get help by visual assistance technologies for identification. (P22)"* In words of P1: *"Those are something that everybody universally deals with, bank statements and prescriptions, it's a good representation, financial stuff to personal things, condom and the pregnancy test, that I wouldn't necessarily want people to see."* Some (i.e., 37%) of participants provided a score of 9 out of 10 for the accuracy of the props objects, with a rationale that context matters. For example, P5 discussed that she felt the "pregnancy test" prop object is irrelevant: *I don't have to worry about pregnancy test, because I had ovarian cancer. But if I were in a situation like that, I'd want the results to be private, [...] it really depend on specific situation."* Other participants provided scores from 5 to 7, such as P13 who noted: *"There were some, just objects, and it's not that as big of a deal. Tattoo stuff? Its more common now a days. Condom or pill bottle are not as big of a deal."*

Our findings also revealed types of prop objects that were not presented in existing taxonomies [44, 73] and could be included in future research. Most prominently, 10 participants spoke about Social Security Number, Photo ID, Driver License. Some suggested *passport, credit card offer/activation mail, paystub, legal document (marriage/divorce decree), taxes, adult magazine, clothing (under garment), sanitary napkin/tampoon* as potential private visual content types.

4.3.3 Most & Least Privacy Concerning Props.

Participants ranked the most concerning prop objects based on two main factors *"information to be targeted by bad actors,"* and *"social stigma"*. In word of P15: *"credit card, where people could easily get your banking information, and steal money from you, [...] And I've had that happen to me before."* P24 noted social stigma around personal/sexual or health related objects: *"if I hadn't told anybody I was pregnant and then there's a pregnancy test in the background, [..] it's sensitive to me. I would say the same with the condom, I would feel really uncomfortable, people knowing, I'm using condoms, or what brand of condoms."* Our participants ranked financial props (e.g., bank statement (17), credit card (15), mortgage (10)) as most concerning followed by personal/sexual health props (e.g., pregnancy test (13), condom (9)), medical props (medical record(6), pill bottle (2), doctor's prescription(2)), and then props identifications (e.g., letter with address (1), tattoo(1)) which conforms to previous literature [73]. Participants frequently indicated business card, newspaper, bill/receipt, letter with address as least concerning private content based on the type of information included in those object as well as possible risk if those information get out to public.

4.3.4 Comfort and Concerns with Capturing Prop Private Content.

As for the opportunities of contributing to the BIV-Priv dataset, all participants consistently indicated to be comfortable in sharing prop images and videos with the research team since they captured prop objects instead of their own objects. P10 noted - *I knew these were props but I knew I figured I knew that your goal as a team was going to be to working on something that would help protect the privacy of the community."* Another factors mentioned by the majority of the participants was collective good by contributing to the the development of AI tools. The photography task experience made participants internalize the purpose of the data collection and comfortable with sharing and the public release. To emphasize the importance of the design of prop objects, many of them indicated that they would not share images and videos if those objects were their own. In the words of P1 -*"I'm glad these are prop items, and not my own. Because a lot of those things, if they were my own, I wouldn't necessarily be wild about uploading them anywhere, probably won't share with research team. So, it was it made me think about privacy."*

As for the concerns, few participants mentioned their concerns if they would accidentally capture own private content during the photography tasks. To add more contexts, participants discussed about the background images / videos set up where the camera tended to capture a wide view having different objects in the frame that they might not be aware of. P21 noted- *"I constantly was thinking-I hope I'm not taking pictures of something I'm not supposed to, inappropriate, tiny bit nervous. It was just me sitting at my desk or at the table and I could capture me somehow. I've been born blind, I don't really know all the things and view that camera can see and pick up."*

4.4 The BIV-Priv Dataset

We designed the creation of the BIV-Priv dataset so that blind participants capture a diversity of private content types that vary in appearances (e.g., different templates, colors) to support AI model development.

Our final BIV-Priv dataset consists of **728** images and **728** videos showing 14 types of prop objects taken by 26 blind participants, with each private category having 26 foreground and 26 background images/videos with considerable diverse appearance and surroundings. Example images from the BIV-Priv dataset are shown in Figure 3. As discussed in our design of this dataset (Section 3.2), it was created to match the standard expectations for datasets in the AI community, specifically that there are a sufficient number of examples and diversity of data.[9] Moreover, the private visual content included in the BIV-Priv dataset represents a variety of information blind people consider to be private, identified in the previous literature [29, 73]. We will publicly-share all data deemed suitable for public sharing.[10] Such data may facilitate AI model development beyond privacy-preserving AI as it captures phenomena often observed in real-world scenarios from blind people (e.g., low quality content [19]) that are rare in existing AI datasets.

5 DISCUSSION

Research on the subject of fair and inclusive development of AI has highlighted the benefits and responsibility of engaging blind users in creating datasets of their images and videos for AI innovation [29, 39, 75, 77]. In this paper, we extended prior work focused on developing datasets from blind users [30, 48] by identifying best

[9] The total number of categories is comparable to the existing literature as a few-shot query set, where there are five categories in the query set of PASCAL-5^i [13, 50], 20 categories in the query set of COCO-20^i [52], 260 categories in the query set of FSS-1000 [43], and 200 categories in the query set of FSOD [23]. In the 14 categories, our dataset contains 52 images per category on average, which is valid compared to that of the existing datasets (e.g., 987 for PASCAL-5^i [13, 50], 2559 for COCO-20^i [52], 10 for FSS-1000 [43], and 70 for FSOD [23]).

[10] https://home.cs.colorado.edu/~DrG/VisualPrivacy/BIV-Priv/

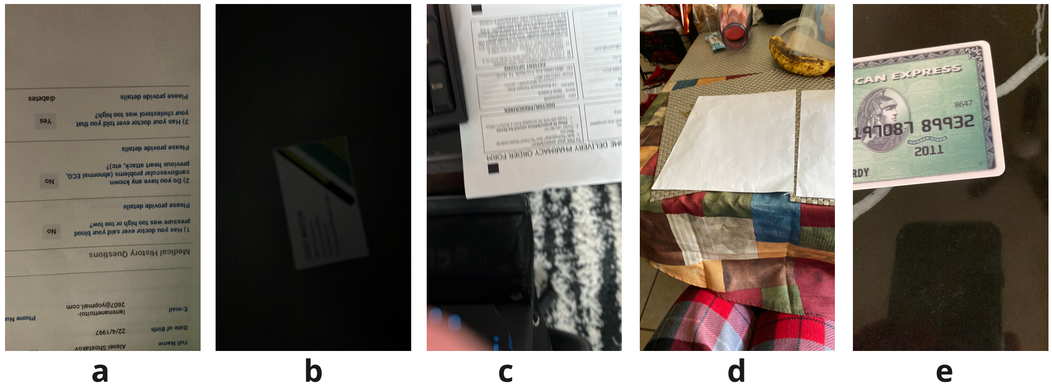

Figure 3: Example images exhibiting different characteristics such as (a) prop object is up-side down, (b) poor lighting, (c) blur, (d) blank side of paper props, and e) small objects partially in the frame.

practices for engaging blind people in capturing images and videos that contain private visual content. We presented our design decisions and reflections to highlight lessons learned through designing 14 types of *prop* private visual content, disseminating the props to 26 blind participants, who ultimately captured 728 images and 728 videos which we will publicly share as BIV-PRIV. Our approach focused on how to conduct the visual data collection in an ethical and accessible manner towards the utility of supporting AI models (e.g., few-shot learning) and the development of privacy-preserving tools for people who are blind. Here, we further discuss how key insights from our study can inform future data collection efforts of private and sensitive content from underrepresented groups in a privacy-respecting way.

5.1 Engage Blind Users in Dataset Creation

From the HCI literature related to dataset creation [37, 58, 75, 88], we found that collecting a dataset for AI innovation would work well when participants understand the study purpose and the tangible outcomes from it. Following these guidelines, our efforts to explain the goals of dataset collection to participants at the outset of the study helped orient participants to the task of capturing images/videos of private visual content and helped engage them in completing the task. For instance, we addressed participants' questions, concerns, and confusions about the study in training sessions via Zoom (e.g., where would the data would be released, how would we handle the pre-processing procedure, and would the props contain actual private information). Nevertheless, the protocol we developed and followed, also highlighted challenges in ensuring privacy. For instance, despite the training and instructions we provided, several participants accidentally captured their own private content in the images and videos they submitted. Some participants also requested additional support to understand non-visual photography and the task we asked them to accomplish (e.g., image composition, including capturing prop private visual content in the foreground and background of their images and videos). Based on these observations, we suggest that future dataset creation efforts should engage people who are blind to design a protocol with more interactive learning resources. Moreover, we observed a range of unexpected photography challenges through the images and videos submitted by participants (e.g., difficulty with the orientation of paper documents) despite careful planning with accessibility researchers. We, therefore, advocate for engaging blind people with non-visual photography experience with the research team in co-developing the study protocol as an additional guideline that extends disability-first dataset creation methods [29, 58, 75].

5.2 Contextual Privacy & Impression Management

Our findings highlight that individuals' privacy perceptions and contextual factors dictate how they feel about the privacy of different prop types (e.g., condoms, pregnancy test kits). This conforms to the key argument from Helen Nissenbaum's theory of *"Contextual Integrity"* [53], which views privacy as the appropriate flow of information rather than a static act of sharing. In addition, our findings on privacy perception towards different prop objects suggest an important distinction between privacy and sensitivity. For example, while performing the photography tasks with the prop private objects, our participants reported being mindful of the people around them, which often made them delay the tasks strategically to a later time when they could ensure that no one was around. Drawing from prior literature, our participants had underlying causes for privacy desires that result in the form of impression management [78], including getting more control and knowledge of the surroundings to avoid social stigma and interrogation. A major factor that influences their impression management was trust in the social circle [38] in combination with a number of contextual factors, such as age, sex, ties with individuals [82]. For instance, some participants seemed to be uncomfortable around their nephews/nieces who were younger, as well as people with close ties (e.g., husbands). These privacy concerns and impression management, unfortunately, also challenged our participants in performing the photography tasks as naturalistic as possible. We overlooked this impression management aspect of privacy when designing the props, which is an important lesson we have learned. Future research and creation of datasets of private contents should explicitly remind participants this potential challenge and make the research intent more visible in the study instructions and the artifacts (e.g., packaging).

5.3 Multi-disciplinary Efforts

Through our study, we affirmed the importance of multi-disciplinary efforts for sensitive dataset creation, especially when the target users are underrepresented. Our team's multi-disciplinary perspectives (accessibility, privacy, computer vision) were particularly helpful in shaping the dataset creation goals (e.g., deciding what data is needed), evaluating privacy risks and accessibility of the dataset creation procedure, as well as supporting each other in considering and protecting our own privacy during the research process. We were able to converge on an approach that was conservative in trying to minimize the risks to participants and researchers but still allowed us to create a useful dataset in terms of the diversity of content (i.e., number of categories) and size (i.e., number of images per category) expected in mainstream AI test datasets. Future AI work will need to investigate the utility of this dataset over alternative private visual content datasets, such as those created through image/video synthesis (discussed in Section 2.2).

5.4 Balance Between Accessibility and Utility Considerations

Our data collection infrastructure (e.g., mobile app, database) was designed to support the requirements of our intended AI application development. Therefore, we did not enable mobile phone camera features that may distort or blur the images and videos captured by participants. For example, we did not use auto-focus to mimic the diversity of appearances often reportedly captured in images and videos taken by blind photographers that many are blurry, poorly lit, and show partially obscured objects [9, 11, 30, 34]. We chose this to rectify findings [75] from authors of the recently created dataset, ORBIT, that they didn't see blurry or poorly lit videos, which they indicated could be an artefact of the camera functionality available in iOS devices. Despite detangling our data collection approach from iOS's proprietary auto-focus, our work has the potential to generalize beyond iPhones as well as with future technology as autofocus may change. That is because AI models can mimic different focusing functionality by converting RBG images into depth-aware style to make particular objects look sharp and everything else out-of-focus [21]. For similar reasons, we also used the default camera app without providing additional cues such as audio to note where the object is in relation to the camera. We chose this to mimic the diversity of challenges often reported by our target population around capturing the content of interest in their images and videos [16, 80, 91]. An interesting area for future research would be to explore adoption of iOS's recent innovations in auditory and haptic feedback. With that said, we ensured a basic level of accessibility of the task by providing audio instructions for navigating only the app and offered the opportunity for real-time chat in case participants needed specific help.

5.5 Limitations and Risks

Our dataset creation has a number of limitations. First, while the 14 types of private visual content in our study were informed by the prior literature (e.g., [44, 73, 75]), we did not include all the private content types due to various ethical and logistic reasons (e.g., guns). We also cannot claim that the 14 categories of private content can be generalized to everyone because people often have different views on which content is private. Second, our study only focused on blind users from a limited diversity of backgrounds which, in turn, could mean our findings generalize poorly beyond our sample population. To capture more diversity in the image and video data and so limit marginalizing some members of our target population from AI development, future work can include a greater diversity of users including with a greater range of visual impairments, ages, and socio-economic backgrounds. Third, we made choices about the camera settings and photography situation that can limit their realism for some scenarios. For instance, it is possible that turning off auto-focus would limit this dataset's applicability for current iPhone users. Additionally, specific instructions for taking photos and videos may have introduced bias and skewed the data away from what would be observed naturally from such photographers. Fourth, we overlooked the social implications (impression management) of some prop objects (e.g., pregnancy test kits). As discussed in section 3.2.1, we didn't anticipate residual risk relating to the discomfort and interrogation from family members asking about certain prop objects (i.e., condom, pregnancy test) where participants had to explain that those props are not their own possessions. However, in doing so, they had to reveal they were participating in a study that potentially poses a risk to their confidentiality. Finally, though our intent in creating a public dataset is to support the development of AI tools that empower blind users to independently protect their private/sensitive content, it is possible such downstream tools would be used improperly in society. Malicious individuals could instead use such tools to accelerate finding private visual content in blind people's images and videos in order to threaten or harm such individuals, through means such as blackmail or public sharing.

6 CONCLUSION

Many people with disabilities, including those who are blind, use AI for accessibility purposes. However, bias of AI models can have huge impacts on blind users' life, and in turn users' trust towards AI systems. In context of an AI system identifying private content in images and videos for blind users, the level of confidence of AI system depends on the representativeness of the data used to develop the AI system. We designed a method for engaging blind users to collect visual data in a privacy preserving way that can be used in a dataset to develop AI models. Our work resulted in the first disability-first dataset with private visual content, **BIV-Priv**. We conclude with reflections on challenges and opportunities for future research on the creation of inclusive datasets to drive AI innovation.

ACKNOWLEDGMENTS

We thank our participants for their contributions and sharing their insights. We also thank Yaman Yu and Smirity Kaushik for their help with the user study. This research was in part supported by the National Science Foundation (NSF) grants #2126314, #2028387, #2125925, and #2148080 and the CRA CIFellows program.

REFERENCES

[1] [n. d.]. braille designer,". In *http://luciahasty.com/index.html, 2022.*
[2] [n. d.]. etsy. In *https://www.etsy.com/, 2022, (Accessed on 01/08/2022.*
[3] [n. d.]. fakenumber.org,". In *https://fakenumber.org/, 2022.*
[4] [n. d.]. N. F. of the Blind, "Blindness statistics,". In *https://www.nfb.org/resources/blindness-statistics.*
[5] [n. d.]. PDFFiller,". In *https://www.pdffiller.com/, 2022.*
[6] [n. d.]. Photoshop,". In *https://www.adobe.com/products/photoshop.html, 2022.*
[7] [n. d.]. randomlist.com,". In *https://www.randomlists.com/random-addresses, 2022.*
[8] [n. d.]. VoiceOver. https://developer.apple.com/documentation/accessibility/supporting_voiceover_in_your_app
[9] Dustin Adams, Sri Kurniawan, Cynthia Herrera, Veronica Kang, and Natalie Friedman. 2016. Blind photographers and VizSnap: A long-term study. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. 201–208.
[10] Dustin Adams, Lourdes Morales, and Sri Kurniawan. 2013. A Qualitative Study to Support a Blind Photography Mobile Application. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments* (Rhodes, Greece) *(PETRA '13)*. Association for Computing Machinery, New York, NY, USA, Article 25, 8 pages. https://doi.org/10.1145/2504335.2504360
[11] Dustin Adams, Lourdes Morales, and Sri Kurniawan. 2013. A qualitative study to support a blind photography mobile application. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*. 1–8.
[12] Mohammad Al-Rubaie and J Morris Chang. 2019. Privacy-preserving machine learning: Threats and solutions. *IEEE Security & Privacy* 17, 2 (2019), 49–58.
[13] Zhen Liu Irfan Essa Amirreza Shaban, Shray Bansal and Byron Boots. 2017. One-Shot Learning for Semantic Segmentation. In *Proceedings of the British Machine Vision Conference (BMVC)*. Article 167, 13 pages.
[14] Karla Badillo-Urquiola, Yaxing Yao, Oshrat Ayalon, Bart Knijnenurg, Xinru Page, Eran Toch, Yang Wang, and Pamela J Wisniewski. 2018. Privacy in Context: Critically Engaging with theory to guide privacy research and design. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 425–431.
[15] Jeffrey P Bigham and Patrick Carrington. 2018. Learning from the front: People with disabilities as early adopters of AI. *Proceedings of the 2018 HCIC Human-Computer Interaction Consortium* (2018).
[16] Jeffrey P Bigham, Chandrika Jayant, Andrew Miller, Brandyn White, and Tom Yeh. 2010. VizWiz:: LocateIt-enabling blind people to locate objects in their environment. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 65–72.
[17] Abeba Birhane and Vinay Uday Prabhu. 2021. Large image datasets: A pyrrhic win for computer vision?. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 1536–1546.
[18] Abagayle Lee Blank. 2019. Computer vision machine learning and future-oriented ethics. (2019).
[19] 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.
[20] Kate Crawford and Trevor Paglen. 2019. Excavating AI : The politics of training sets for machine learning. https://www.excavating.ai
[21] Saikat Dutta. 2021. Depth-aware blending of smoothed images for Bokeh effect generation. *Journal of Visual Communication and Image Representation* 77 (2021), 103089. https://doi.org/10.1016/j.jvcir.2021.103089
[22] Mark Everingham, Luc Van Gool, C. K. I. Williams, J. Winn, and Andrew Zisserman. 2010. The PASCAL Visual Object Classes (VOC) challenge.
[23] Qi Fan, Wei Zhuo, Chi-Keung Tang, and Yu-Wing Tai. 2020. Few-Shot Object Detection with Attention-RPN and Multi-Relation Detector. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
[24] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods* 5, 1 (2006), 80–92.
[25] Andrea Frome, German Cheung, Ahmad Abdulkader, Marco Zennaro, Bo Wu, Alessandro Bissacco, Hartwig Adam, Hartmut Neven, and Luc Vincent. 2009. Large-scale privacy protection in google street view. In *2009 IEEE 12th international conference on computer vision*. IEEE, 2373–2380.
[26] 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 (nov 2021), 86–92. https://doi.org/10.1145/3458723
[27] Leandro Soares Guedes, Luiz André Marques, and Gabriellen Vitório. 2020. Enhancing interaction and accessibility in museums and exhibitions with Augmented Reality and Screen Readers. In *International Conference on Computers Helping People with Special Needs*. Springer, 157–163.
[28] Anhong Guo, Ece Kamar, Jennifer Wortman Vaughan, Hanna Wallach, and Meredith Ringel Morris. 2020. Toward fairness in AI for people with disabilities SBG@ a research roadmap. *ACM SIGACCESS Accessibility and Computing* 125 (2020), 1–1.
[29] Danna Gurari, Qing Li, Chi Lin, Yinan Zhao, Anhong Guo, Abigale Stangl, and Jeffrey 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.
[30] Danna Gurari, Qing Li, Abigale J Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey 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.
[31] Adam Harvey and Jules LaPlace. 2021. Exposing A. https://www.excavating.ai
[32] Annette Haworth and Peter Williams. 2012. Using QR codes to aid accessibility in a museum. *Journal of Assistive Technologies* (2012).
[33] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting Blind Photography. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (Dundee, Scotland, UK) *(ASSETS '11)*. Association for Computing Machinery, New York, NY, USA, 203–210. https://doi.org/10.1145/2049536.2049573
[34] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P Bigham. 2011. Supporting blind photography. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. 203–210.
[35] Hernisa Kacorri, Utkarsh Dwivedi, Sravya Amancherla, Mayanka Jha, and Riya Chanduka. 2020. IncluSet: A data surfacing repository for accessibility datasets. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4.
[36] Amlan Kar, Nishant Rai, Karan Sikka, and Gaurav Sharma. 2017. Adascan: Adaptive scan pooling in deep convolutional neural networks for human action recognition in videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3376–3385.
[37] Nicolas Kaufmann, Thimo Schulze, and Daniel Veit. 2011. More than fun and money. worker motivation in crowdsourcing–a study on mechanical turk. (2011).
[38] Alfred Kobsa, Sameer Patil, and Bertolt Meyer. 2012. Privacy in instant messaging: An impression management model. *Behaviour & Information Technology* 31, 4 (2012), 355–370.
[39] Pavel Korshunov and Touradj Ebrahimi. 2013. PEViD: privacy evaluation video dataset. In *Applications of Digital Image Processing XXXVI*, Vol. 8856. SPIE, 578–586.
[40] Aditya Kuppa, Lamine Aouad, and Nhien-An Le-Khac. 2021. Towards improving privacy of synthetic datasets. In *Annual Privacy Forum*. Springer, 106–119.
[41] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, et al. 2020. The open images dataset v4. *International Journal of Computer Vision* 128, 7 (2020), 1956–1981.
[42] Kyungjun Lee and Hernisa Kacorri. 2019. Hands holding clues for object recognition in teachable machines. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
[43] Xiang Li, Tianhan Wei, Yau Pun Chen, Yu-Wing Tai, and Chi-Keung Tang. 2020. FSS-1000: A 1000-Class Dataset for Few-Shot Segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020).
[44] Yifang Li, Nishant Vishwamitra, Hongxin Hu, and Kelly Caine. 2020. Towards a taxonomy of content sensitivity and sharing preferences for photos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
[45] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.
[46] Alexandra Sasha Luccioni, Frances Corry, Hamsini Sridharan, Mike Ananny, Jason Schultz, and Kate Crawford. 2022. A Framework for Deprecating Datasets: Standardizing Documentation, Identification, and Communication. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 199–212. https://doi.org/10.1145/3531146.3533086
[47] Giancarlo Manzi and Martin Forster. 2019. Biases in bias elicitation. *Communications in Statistics-Theory and Methods* 48, 18 (2019), 4656–4674.
[48] Daniela Massiceti, Luisa Zintgraf, John Bronskill, Lida Theodorou, Matthew Tobias Harris, Edward Cutrell, Cecily Morrison, Katja Hofmann, and Simone 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.
[49] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)* 54, 6 (2021), 1–35.
[50] Claudio Michaelis, Ivan Ustyuzhaninov, Matthias Bethge, and Alexander S. Ecker. 2018. One-Shot Instance Segmentation. *ArXiv* (2018).
[51] Meredith Ringel Morris. 2020. AI and Accessibility. *Commun. ACM* 63, 6 (may 2020), 35–37. https://doi.org/10.1145/3356727
[52] Khoi Duc Minh Nguyen and Sinisa Todorovic. 2019. Feature Weighting and Boosting for Few-Shot Segmentation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 622–631.

[53] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.

[54] Helen Nissenbaum. 2020. Protecting privacy in an information age: The problem of privacy in public. In *The Ethics of Information Technologies*. Routledge, 141–178.

[55] Curtis G Northcutt, Anish Athalye, and Jonas Mueller. 2021. Pervasive label errors in test sets destabilize machine learning benchmarks. *arXiv preprint arXiv:2103.14749* (2021).

[56] Tribhuvanesh Orekondy, Mario Fritz, and Bernt Schiele. 2018. Connecting pixels to privacy and utility: Automatic redaction of private information in images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8466–8475.

[57] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. 2017. Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In *Proceedings of the IEEE international conference on computer vision*. 3686–3695.

[58] Joon Sung Park, Danielle Bragg, Ece Kamar, and Meredith Ringel Morris. 2021. Designing an online infrastructure for collecting AI data from people with disabilities. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 52–63.

[59] Donatella Pascolini and Silvio Paolo Mariotti. 2012. Global estimates of visual impairment: 2010. *British Journal of Ophthalmology* 96, 5 (2012), 614–618.

[60] Arpita Patra and Ajith Suresh. 2020. BLAZE: Blazing Fast Privacy-Preserving Machine Learning. https://doi.org/10.14722/ndss.2020.24202

[61] Haoyue Ping, Julia Stoyanovich, and Bill Howe. 2017. DataSynthesizer: Privacy-Preserving Synthetic Datasets. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management* (Chicago, IL, USA) *(SSDBM '17)*. Association for Computing Machinery, New York, NY, USA, Article 42, 5 pages. https://doi.org/10.1145/3085504.3091117

[62] J Qi, Y Gao, Y Hu, X Wang, X Liu, X Bai, S Belongie, A Yuille, P Torr, and S Bai. 2021. Occluded video instance segmentation: dataset and challenge. NeurIPS.

[63] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision* 115, 3 (2015), 211–252.

[64] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *arXiv:1409.0575 [cs]* (Jan. 2015). http://arxiv.org/abs/1409.0575 arXiv: 1409.0575.

[65] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone Wants to Do the Model Work, Not the Data Work": Data Cascades in High-Stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 39, 15 pages. https://doi.org/10.1145/3411764.3445518

[66] Morgan Klaus Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed R Brubaker. 2020. How we've taught algorithms to see identity: Constructing race and gender in image databases for facial analysis. *Proceedings of the ACM on Human-computer Interaction* 4, CSCW1 (2020), 1–35.

[67] Amirreza Shaban, Shray Bansal, Zhen Liu, Irfan Essa, and Byron Boots. 2017. One-shot learning for semantic segmentation. *arXiv preprint arXiv:1709.03410* (2017).

[68] Rachel N Simons, Danna Gurari, and Kenneth R Fleischmann. 2020. " I Hope This Is Helpful" Understanding Crowdworkers' Challenges and Motivations for an Image Description Task. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–26.

[69] Daniel J Solove. 2005. A taxonomy of privacy. *U. Pa. l. Rev.* 154 (2005), 477.

[70] Eleftherios Spyromitros-Xioufis, Symeon Papadopoulos, Adrian Popescu, and Yiannis Kompatsiaris. 2016. Personalized privacy-aware image classification. In *Proceedings of the 2016 ACM on international conference on multimedia retrieval*. 71–78.

[71] Theresa Stadler, Bristena Oprisanu, and Carmela Troncoso. 2022. Synthetic data–anonymisation groundhog day. In *31st USENIX Security Symposium (USENIX Security 22)*. 1451–1468.

[72] Abigale Stangl, Kristina Shiroma, Nathan Davis, Bo Xie, Kenneth R Fleischmann, Leah Findlater, and Danna Gurari. 2022. Privacy Concerns for Visual Assistance Technologies. *ACM Transactions on Accessible Computing (TACCESS)* 15, 2 (2022), 1–43.

[73] Abigale Stangl, Kristina Shiroma, Bo Xie, Kenneth R Fleischmann, and Danna Gurari. 2020. Visual content considered private by people who are blind. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.

[74] Harini Suresh and John Guttag. 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In *Equity and Access in Algorithms, Mechanisms, and Optimization* (–, NY, USA) *(EAAMO '21)*. Association for Computing Machinery, New York, NY, USA, Article 17, 9 pages. https://doi.org/10.1145/3465416.3483305

[75] Lida Theodorou, Daniela Massiceti, Luisa Zintgraf, Simone Stumpf, Cecily Morrison, Edward Cutrell, Matthew Tobias Harris, and Katja Hofmann. 2021. Disability-first Dataset Creation: Lessons from Constructing a Dataset for Teachable Object Recognition with Blind and Low Vision Data Collectors. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.

[76] Shari Trewin, Sara Basson, Michael Muller, Stacy Branham, Jutta Treviranus, Daniel Gruen, Daniel Hebert, Natalia Lyckowski, and Erich Manser. 2019. Considerations for AI fairness for people with disabilities. *AI Matters* 5, 3 (2019), 40–63.

[77] Yu-Yun Tseng, Alexander Bell, and Danna Gurari. [n. d.]. VizWiz-FewShot: Locating Objects in Images Taken by People With Visual Impairments.

[78] Sonja Utz and Nicole Krämer. 2009. The privacy paradox on social network sites revisited: The role of individual characteristics and group norms. *Cyberpsychology: Journal of psychosocial research on cyberspace* 3, 2 (2009), 2.

[79] Marynel Vázquez and Aaron Steinfeld. 2012. Helping visually impaired users properly aim a camera. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. 95–102.

[80] Marynel Vázquez and Aaron Steinfeld. 2014. An assisted photography framework to help visually impaired users properly aim a camera. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 5 (2014), 1–29.

[81] James Vincent. [n. d.]. Transgender YouTubers had their videos grabbed to train facial recognition software. *The Verge* ([n. d.]). https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset

[82] Jessica Vitak. 2015. Balancing privacy concerns and impression management strategies on Facebook. In *Symposium on usable privacy and security (SOUPS)*. 22–24.

[83] Karl R White. 1982. The relation between socioeconomic status and academic achievement. *Psychological bulletin* 91, 3 (1982), 461.

[84] Meredith Whittaker, Meryl Alper, Cynthia L Bennett, Sara Hendren, Liz Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Rankin, Emily Rogers, Marcel Salas, et al. 2019. Disability, bias, and AI. *AI Now Institute* (2019).

[85] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1180–1192.

[86] Runhua Xu, Nathalie Baracaldo, and James Joshi. 2021. Privacy-Preserving Machine Learning: Methods, Challenges and Directions. (08 2021).

[87] Linjie Yang, Yuchen Fan, and Ning Xu. 2019. Video Instance Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

[88] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.

[89] Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao. 2015. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365* (2015).

[90] Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, and Elena Demidova. 2012. Privacy-aware image classification and search. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. 35–44.

[91] Yu Zhong, Pierre J Garrigues, and Jeffrey P Bigham. 2013. Real time object scanning using a mobile phone and cloud-based visual search engine. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–8.

[92] Jizhe Zhou, Chi-Man Pun, and Yu Tong. 2020. Privacy-sensitive objects pixelation for live video streaming. In *Proceedings of the 28th ACM International Conference on Multimedia*. 3025–3033.

[93] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 493–502.