

Designing an Online Infrastructure for Collecting AI Data From People With Disabilities

Joon Sung Park
Microsoft Research – Redmond,
Stanford University
Stanford, CA, USA
joonspk@stanford.edu

Ece Kamar
Microsoft Research – Redmond
Redmond, WA, USA
eckamar@microsoft.com

Danielle Bragg
Microsoft Research – New England
Cambridge, MA, USA
danielle.bragg@microsoft.com

Meredith Ringel Morris
Microsoft Research – Redmond
Redmond, WA, USA
merrie@microsoft.com

ABSTRACT

AI technology offers opportunities to expand virtual and physical access for people with disabilities. However, an important part of bringing these opportunities to fruition is ensuring that upcoming AI technology works well for people with a wide range of abilities. In this paper, we identify the lack of data from disabled populations as one of the challenges to training and benchmarking fair and inclusive AI systems. As a potential solution, we envision an online infrastructure that can enable large-scale, remote data contributions from disability communities. We investigate the motivations, concerns, and challenges that people with disabilities might experience when asked to collect and upload various forms of AI-relevant data through a semi-structured interview and an online survey that simulated a data contribution process by collecting example data files through an online portal. Based on our findings, we outline design guidelines for developers creating online infrastructures for gathering data from people with disabilities.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility design and evaluation methods.**

KEYWORDS

AI FATE, datasets, inclusion, representation, accessibility, disability

ACM Reference Format:

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 *Conference on Fairness, Accountability, and Transparency (FAccT '21)*, March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3442188.3445870>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

FAccT '21, March 3–10, 2021, Virtual Event, Canada

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8309-7/21/03...\$15.00

<https://doi.org/10.1145/3442188.3445870>

1 INTRODUCTION

The increasing power and pervasiveness of AI technologies pose opportunities for enabling people with disabilities (PWD) to engage in previously inaccessible activities. Existing AI products like Microsoft Seeing AI or Google Lookout audibly describe images to enable blind people to interpret visual information [12, 20], and DragonDictate, which transcribes speech into text, enables those who experience physical disabilities or dyslexia to communicate more efficiently in written language [11]. However, though some AI technologies may be effective for PWD, others may actually aggravate disability-based discrimination [38]. Computer vision tools may not realize their potential benefits to PWD if they are less accurate at recognizing images taken by blind users than by sighted users [25], and automated transcription tools risk leaving behind users with varied speech patterns, such as "deaf accent" [21, 50, 52]; further, there is a risk that future AI systems like motion detectors in a self-driving car may pose physical danger if they fail to recognize pedestrians using mobility aids such as wheelchairs [57].

One important source of AI systems' inclusivity issues is the lack of representation for PWD in the datasets used to train and benchmark them [10, 23, 38] (though it is also important to note other sources of inclusivity issues, such as lack of representation of PWD on the teams creating AI technologies). But collecting a representative dataset for PWD is hard. PWD are minorities in the general population, making it infeasible or expensive to collect large enough data from PWD in an in-person setting [23]. This is further aggravated as traveling to an in-person data collection site may be very challenging for some PWD. Scraping online data of PWD is also problematic not only because of the long-tail nature of various disabilities but also because of the ethical considerations and the heightened need to protect the sensitive demographic attributes related to disabilities [23, 46]. Although some tried to overcome these challenges by having participants without disabilities simulate disability (e.g., covering one's eyes while performing a task), such "simulated disability data" are inaccurate and can reinforce negative stereotypes around disabilities [44, 57], making ML techniques that rely on "synthetic data" [13, 40] fraught for this demographic.

In order to mitigate this issue, we envision an online infrastructure for enabling controlled data collection in which PWD can willingly contribute their data to create more representative datasets

for training or testing AI systems. If effective, such an infrastructure could be an important step toward improving the representation of PWD in AI datasets by making the data collection process scalable and cost-effective, with direct implications for fairness in AI systems for PWD who are often discriminated against due to their disability status [16]. However, the open questions around how to create a data contribution process that is sufficiently motivating, privacy-sensitive, and accessible to a diverse array of abilities need to be answered for its success [23, 24].

To this end, we conducted a two-part study with 30 participants who experience a wide range of disabilities, to gain generalizable insights on PWD’s motivations, concerns, and challenges with regards to contributing data typically needed for a variety of AI systems through an online infrastructure. In the first part, we interviewed our participants (through a video call due to COVID-19) to qualitatively probe their willingness and concerns for contributing their data to hypothetical AI datasets. In the second part, we asked our participants to complete an online survey, during which they were asked to collect and upload to an online form six common types of AI training data (e.g., a self-portrait photo like those used for face authentication on a phone, or a video of them speaking that can be auto-transcribed by a speech-to-text system) to simulate a data contribution process. They were asked to describe challenges they faced or help they needed, and complete an online survey that revisited several themes from the interview in light of having experienced a sample data contribution process.

We find that many of our participants were open to contributing their data, even on a voluntary basis, if it meant that AI could become more effective for PWD, suggesting that seeking data contributions from PWD could be a promising direction for creating more inclusive AI datasets. However, our findings also suggest that the process of motivating and receiving PWD’s data contribution has to meet the unique needs and circumstances that they may experience. For example, when determining compensation forms and rates for PWD’s data contribution, one may want to consider the sensitivity of the data as well as the effort and resources it takes for PWD to collect and upload those data. It is also important to be cognizant of the fact that people experience disabilities on various spectrums; some experience non-apparent disabilities such as ADHD or dyslexia and might be more cautious about contributing data that might disclose their disabilities, while some experience more prominent disabilities within a disability category that calls for a carefully designed data collection process that is accessible for them. Based on our findings, we end with guidelines for designing data collection processes for PWD.

2 RELATED WORK

In this section, we cover prior work that investigated the performance disparities and biases of AI systems based on demographic traits, and how the lack of representation in training and testing data of certain populations can fuel such outcomes. We also cover opportunities and challenges to addressing this issue.

2.1 Performance Disparities of AI Systems

2.1.1 Disparities impacting marginalized groups. An increasing number of evaluations have examined performance disparities of

AI systems for people with demographic attributes, particularly attributes that are marginalized in the technology industry specifically and/or in Western societies more generally. Such studies have investigated how one’s gender [10, 34, 42, 47, 50], race [10, 34], and socioeconomic status [1] could negatively affect the performance of an AI system such as facial recognition or natural language processing systems. In the landmark Gender Shades study, for example, authors show that many commercial AI systems that are used to label binary gender classes based on one’s face in an image did not work well for women with darker skin color [10].

One source of disparities is the lack of representation of marginalized populations in the datasets used to train AI systems. Whittaker et al. note that “AI systems model the world based on what’s in the data they’re given. If something is missing from the data, say images of people with dark skin, these people will be missing from the AI model, and thus won’t be recognized or included” [57]. When the Gender Shades study identified the problem of higher error rates for women with darker skin color, the developers updated their system to mitigate this disparity by creating a more balanced training dataset; this reduced the error rate by nearly ten-fold when measured using a testing dataset similar to the one used in the Gender Shades study [45]. In response, there is a growing effort to create more balanced datasets that represent different demographic attributes such as the creation of the FairFace dataset that has equal representation of various races and genders [30].

2.1.2 Disparities impacting PWD. Despite the recent advances investigating discriminatory AI models based on race, gender, and socioeconomic status, the dimension of disability received less attention [52]. Nonetheless, a series of anecdotal evidence highlights the potentially devastating impact non-inclusive AI can have on PWD. One example considers the AI models created to make autonomous vehicles recognize pedestrians. In 2018, Treviranus tested one of these models with a video recording of a friend propelling backward in a wheelchair, and found that all AI models studied chose to run over her friend [51]. Treviranus concludes that models trained on datasets that mainly represent those not in a wheelchair fail to parse even a slightly unusual activity performed in a wheelchair such as propelling backward. Similarly, people found a food delivery robot to not yield the curb cut to a pedestrian in a wheelchair [48] while others cited autonomous vehicles crashing into those with bicycles [22, 35] as cautionary tales to further support that AI models will likely fail to recognize those with mobility aids [41, 57].

Beyond the case of autonomous vehicles that are not yet widely commercialized, these anecdotes also describe how PWD are affected by today’s non-inclusive, mainstream AI products. For example, speech recognition systems that take speech input and output text could be useful tools for people who are deaf or hard of hearing as they can transcribe a live conversation or videos [19, 56]. In addition, these systems can also be used to make traditional input devices more accessible to those with limited mobility [3]. However, speech recognition systems often fails for those who have atypical speech [2]. Such bias exists even in the gender dimension where speech recognition systems are reported to perform better for men than women [47, 50], and worse for accents including those originating from disabilities like deaf accent [21, 50]. Similarly, computer vision systems built to help blind users could fail to accurately

caption images taken by these users as the images could differ in terms of their framing, angle, or lighting when compared to images taken by sighted people [25] used to train many of these systems [14]. Finally, smart sensor systems can fail to work properly for people with different body shape because of growth differences, amputation, or because they are seated in a wheelchair [39].

2.2 Efforts to Create Representative Datasets

2.2.1 Calls-to-action for inclusive AI. In response to the growing AI inclusivity concerns for PWD, there have recently been multiple calls-to-action advocating for a rigorous investigation into creating more disability-inclusive AI technology [39]. These calls-to-action encouraged inspecting whether AI applications are trained on datasets that represented PWD, whether they had reasonable accuracy when used by PWD, and what sorts of societal biases they can incur against PWD and other vulnerable groups [6, 17, 24, 28, 52, 57]. In particular, they highlighted the creation of inclusive public datasets for PWD used to train and test modern AI applications as one of the key areas for future work [24].

Unfortunately, however, creating representative datasets that include PWD remains difficult as common data-collection methods introduce disproportional barriers and risks for PWD. In-person data collection can be prohibitively expensive, due to the low representation of many disability groups in the general population due to the "long tail of disability" [23]. Some disabilities may prevent potential participants from traveling to onsite data collection events due to limited mobility (a challenge exacerbated by the COVID-19 pandemic, since many with disabilities are in high-risk groups and must maintain strict social distancing). Scraping online data sources to collect data of PWD is also difficult due to the low representation of disability groups, the heightened needs for privacy when the data concerns PWD, the ethical issues of scraping data without consent, and the fact that accurate descriptions of disability status are rarely associated with scraped content [23]. Finally, simulating disabilities can bear inaccurate data that may also reinforce societal prejudices and stereotypes of experiencing disabilities [44, 57]. Such challenges pose roadblocks for well-intentioned AI practitioners who are motivated to understand and mitigate biases in their systems but lack sufficient expertise to work with PWD.

2.2.2 Online data collection from PWD. Given these challenges, creating online infrastructures that centralize the data collection from PWD can be an alluring solution for sourcing data directly from PWD by allowing a semi-controlled data collection process at scale [9]. Indeed, online infrastructures have been powerful and scalable means of sourcing data from people for many years [37, 59]. Games with a purpose, such as the "ESP Game" for labeling images, were an early success that leveraged the power of people on the Internet to collect data with a direct implication to improving accessibility by captioning images on the web [54, 55]. The VizWiz app allows blind users to upload photos and receive descriptions from paid crowdworkers [7]; many VizWiz users have opted to share images to create public datasets for AI training [25, 26]. Similarly, citizen scientists connected on the web have proven themselves to be effective, for example at tracking and sharing the radiation data after the Fukushima nuclear disaster in Japan [27, 29]. Finally, more recent efforts such as the "data dignity" project have started

to envision how users can be compensated for data they contribute for a variety of commercial and/or academic purposes [33].

However, the following open questions need to be answered for designing successful processes for collecting data from PWD [24]. What would motivate PWD to contribute their data to an AI dataset? What are their concerns for contributing their data to an AI dataset as it relates to ethics and privacy? And what challenges would they face when collecting and uploading their data online? We approach these questions from the perspectives of PWD to inform the design of online infrastructures that can be used to collect data from PWD to create more representative AI datasets.

3 METHOD

To better understand PWD's concerns, challenges, and motivations for contributing data for AI development, we conducted a two-part study that included a semi-structured interview followed by an online survey in which the participants were asked to collect and upload sample data to an online portal. The questions used in the interview and the survey are included in the Appendix of the paper.

3.1 Interviews

Our interviews took place through a video call (due to COVID-19 social distancing). With Shepherd Center's (a rehabilitation hospital in Atlanta, GA) help, we recruited participants with diverse abilities and collected their self-descriptions of disability status and information about accessibility needs for participation. The interviews were approximately 45 minutes and structured as follows:

Stage 1: Defining AI. We first asked our participants about their understanding of AI, probing the types of AI applications they are aware of, to get a sense of their familiarity with AI. We then shared our definition of AI (inspired by [43, 49]) as "a computing system or software that can learn from existing data to perceive and function appropriately in various environments," and described the importance of inclusive datasets for creating AI applications that work for everyone. We then provided them with examples of AI applications like voice assistants, image recognition systems, and self-driving cars. We followed up our description with a few questions on inclusion like "Have you heard any reports or stories about biases or discrimination that are associated with any AI applications, or how some AI applications do not work for certain users?" to gauge their knowledge about AI inclusivity issues.

Stage 2: Motivations for contributing to an AI dataset. We then explored what might motivate our participants to contribute to an AI dataset. We presented various forms of compensation mechanisms such as monetary compensation (i.e. the data collecting organization pays the participants money for the data contributed), matching donation (i.e. the data collecting organization donates to a non-profit of the participants' choosing), and voluntary contribution. For each, we probed conditions for the participants to feel motivated to contribute their data. For example, we asked questions like "Would you expect to be paid for contributing your data to an AI dataset? If so, how much do you think you should be paid for contributing [various types of data]," or what other information about the purpose of data collection they would want to know.

Stage 3: Concerns about contributing to an AI dataset. Finally, we asked our participants about their concerns with contributing to

Table 1: Self-Descriptions of Types of Disability That Our Participants Experienced

Physical Disability		Hearing Loss or Impairment		Blindness or Visual Impairment		Cognitive Disability	
P1	Spinal Cord Injury	P10	Mild Hearing loss	P16	Blindness	P25	ADHD and Dyslexia
P2	Cerebral Palsy	P11	Profound hearing loss	P17	Blindness	P26	ADHD and Dyslexia
P3	Spinal Cord Injury	P12	Profound hearing loss	P18	Blindness	P27	Dyslexia
P4	Rheumatoid Arthritis	P13	Mild Hearing loss	P19	Congenital Glaucoma	P28	Autism
P5	Shaking Hands	P14	Profound hearing loss	P20	Blindness	P29	Dyslexia
P6	Cerebral Palsy	P15	Profound hearing loss	P21	Blindness	P30	ADHD and Dyslexia
P7	Torsion Dystonia			P22	Mild Blindness		
P8	Spinal Cord Injury			P23	Blindness		
P9	Paralyzed			P24	Visually Impaired		

an AI dataset. For this stage, we narrowed down the context of data contribution by applying a scenario-based method [32, 58]. We presented to our participants a scenario in which a nonprofit organization working for disability justice wants to collect data from PWD to create more inclusive AI applications. We explained to our participants that the types of data collected might include those relevant to disabilities they experience (eg. photos taken of or by blind people, or voice samples of those with a deaf accent). We then asked if there are any physical, psychological, or privacy concerns they might have if they choose to contribute to such an effort. Such scenario-based methods are widely used in social psychology and ethics research, and considered to be an effective way of investigating participants' opinions and beliefs [32, 58].

3.2 Online Survey

After the interview, we provided our participants with a link to an online survey and asked them to complete it within two days; the survey took approximately 45 minutes to complete. The survey allowed us to observe the challenges that our participants might experience when collecting and uploading common AI training data, and to acquire quantitative responses grounded in the experience of having performed a number of specific data-uploading tasks as a followup to the conversation during the interview.

Stage 1: Simulating data contribution. We chose six data types to ask our participants to collect and upload: 1) a self-portrait photo that is relevant for training facial authentication systems on a smartphone, 2) a photo of any one household object that they took with a caption that is relevant for training photo description systems like Seeing AI, 3) a video of them reading out loud a short paragraph either in English or in American Sign Language (ASL) that is relevant for building speech recognition or ASL interpretation systems, 4) a video of them moving six feet to the left and right that is relevant for training motion recognition systems, 5) typing what they hear in a video that reads out loud a short English paragraph with ASL interpretation that might be relevant for AI-based spelling, grammar, or other writing-support systems and 6) clicking on small dots appearing on the browser using a mouse or a trackpad, which can be relevant for inferring motor abilities or age to automatically adjust interface properties (i.e., similar to data collected on the LabintheWild platform by Gajos et al [18]).

Although these are not a comprehensive set of data types used to train AI, they are cases in point for discussing how to build accessible data collection infrastructures for PWD. In addition, it is also important to note that we informed our participants that they could skip a task if they were unable to complete it, and that the data they provided would not be used to train real AI systems.

Stage 2: Closing questions. After completing the data-uploading tasks, we asked our participants about their motivations and concerns for contributing data. This included questions like "How acceptable are the following uses of the data that you contributed" accompanied with examples of data usage like "Used to develop new AI-powered accessibility tools" or "Used to make existing, general-audience AI applications." Unlike during the interview, we took a more quantitative approach by asking our participants to answer many of the questions on a Likert scale. We ended the study with a demographics survey. Appendix 2 includes the full set of questions.

3.3 Designing an Accessible Study

It was important to ensure that all portions of our study were accessible. All our forms and surveys were voice recorded and interpreted into ASL (American Sign Language) and the video of the interpretation was posted alongside the written instructions for the participants who found this form of instructions more accessible. During the interviews with deaf participants, an ASL interpreter was present to interpret the conversation in real-time. We also ensured our survey was screen-reader accessible. In addition, we communicated frequently with Shepherd Center to find out any other accessibility needs our participants had. Finally, the study protocol and materials were IRB-approved.

3.4 Participants

Instead of focusing on people with a particular type of disability, we sought to understand the motivation, concerns, and challenges that PWD might face across various types of disabilities. This was done to uncover the more generalizable framework for enabling data collection from PWD given that technologies present different barriers to people with different disabilities. To this end, we collaborated with Shepherd Center, a non-profit hospital in the U.S. to recruit a participant pool that represented as diverse a set of disabilities as possible so long as a potential participant's disability did not bar them from providing meaningful informed consent. Ultimately, we

recruited 30 participants who experienced four broad categories of disabilities: physical disability such as spinal cord injuries and cerebral palsy (n=9), blindness or visual impairment (n=9), hearing loss to a varying degree (n=6), and cognitive disabilities such as ADHD, dyslexia, and autism (n=6). Table 1 summarizes participants' self-described disability status.

We conducted the study from June to July of 2020. All interviews were conducted by the first author of the study through video calls. Per the participants' approval, the interviews were video recorded, and later anonymized and transcribed by the first author for analysis. Once the participants had finished all portions of the study, the participants were paid a \$75.00 gift card through the recruiting organization for an hour and half of their time (45 minutes for the interview and another 45 minutes for the uploading tasks and survey). Shepherd Center determined the rate of compensation based on their knowledge of the participants. Our participants were on average 39.83 years old (STD=11.97) with an age range from 22 to 70. They identified racially as 60% Caucasian, 16.67% Black, 13.33% Hispanic, 3.33% Asian, and 6.67% other. Regarding gender, 12 identified as women and 18 as men. For brevity, we refer to the interview participants as P1-P30.

3.5 Analysis

We analyzed transcripts of the interviews and the results of the survey that included both quantitative and shorter qualitative responses, as well as the task completion rate and the rate at which our participants needed help. For qualitative data, we conducted an interpretative qualitative analysis [36] that started with open coding in two phases [31]. During the first phases, we coded the transcripts on a line-by-line basis to ensure that our coding of the data closely reflects the data. For example, codes that were generated in this phase included "Spectrum of disability support" and "Behavior altering medication might be needed for cognitive disabilities." In the next phase, we synthesized the codes from the previous phase to extract the higher-level themes that could be observed in our data. These themes included "Interest in data contribution," "Privacy concerns," and "Perceived challenges with data contribution." This coding process was iterative in that we continued to review our themes and data throughout the process to accurately synthesize our findings. Finally, for the quantitative responses, we conducted a one-way ANOVA to find the main effect in the data and ran post-hoc analysis where it was appropriate.

4 RESULTS

We summarize the findings about PWD's motivations, concerns and challenges to contributing to an AI dataset through an online infrastructure. In subsection 4.1, we start by describing our participants' knowledge and usage of AI prior to the study in order to provide a context for our subsequent summary of results. We then focus on our participants' motivations, concerns, and challenges for contributing to an AI dataset in subsections 4.2, 4.3, and 4.4.

4.1 Prior Knowledge and Use of AI

4.1.1 Understanding of AI. Almost all our participants expressed some degree of familiarity with AI applications when asked how they understood AI prior to the interview (n=29). They cited their

awareness of existing AI applications including general audience applications like voice assistants (e.g. Amazon Alexa and Google Home) or chatbots they encounter while browsing the web (n=15), and accessibility-related AI applications like Seeing AI and Dragon Dictate (n=17). Some also cited AI's appearance in popular culture such as movies and fiction (n=6). Several participants (n=8) even had at least a basic understanding of how AI applications are created, using terms such as training and model to describe AI. Only one participant had never heard the term "AI" prior to the interview.

After describing AI and example applications to our participants, we asked whether they used any AI applications in the past. Nearly all our participants had exposure to at least one type of commercial AI application (n=28). These included general audience-facing applications like voice assistants (n=16), online search or feed algorithms (n=5), and narrower AI applications like facial recognition for logging into a smartphone (n=3). In addition, our participants reported having used AI for accessibility (n=17). For example, participants who experience blindness or visual impairment mentioned that they used object recognition systems like Microsoft's Seeing AI or Google's Lookout to interpret visual information (n=7; 77.8%).¹ Similarly, those who experience hearing loss mentioned using text-to-speech systems like auto-captions or Dragon Dictate for transcribing conversations or captioning audio (n=4; 66.7%). The same system was also used by those who experienced dyslexia to generate written content (n=3; 50%). Finally, some found general-audience AI helpful for their accessibility needs, like those who experienced physical disability who used voice assistants to control smart home devices without moving (n=2; 22.2%).

4.1.2 AI failures. When asked their knowledge of reports of biases or discrimination by commercial AI applications, only a few participants (n=3) noted their familiarity with the topic, citing public news stories such as the ones about how a machine vision system fails to recognize those with a darker skin tone. However, many of our participants acknowledged that they have personally experienced AI failures as it relates to accessibility. Their experiences included many that resonate with the growing inclusivity concerns with AI that are being reported in anecdotal evidence. For example, blind participants reported that object recognition systems often failed to recognize the photos that they took (n=5; 55.6%), while our participants who experience dyslexia found some speech-to-text systems time out too fast for them to formulate their thoughts (n=2; 33.3%). Meanwhile, a few participants also expressed that some applications are simply not accessible for them, such as people who are deaf and communicate via English text or ASL rather than speech being unable to use AI-tools like smart-speaker-based voice assistants (n=4; 66.6%).

Interestingly, we also observe that new inclusivity issues arise as AI applications make more activities accessible for PWD. Two blind participants, for example, noted that they hand write more since object recognition applications made reading and writing handwritten words more accessible. But recognizing handwriting using AI is still difficult, with both participants expressing sentiments that it would "definitely work better for somebody who is sighted" (P14).

¹We report percentages for findings relevant to a particular disability category (vision/hearing/motor/cognitive), giving the percentage relative to participants identifying with that category rather than relative to the overall participant pool.

But in cases like these where AI failed to be inclusive, some viewed the failure to have been caused by their misuse rather than by the AI system’s shortcomings (n=5). This made them think that there is little to be done to improve AI applications; “I’m not going to fault [the AI application] because I mean, everyone’s handwriting is so different. So I don’t think there’s too much to be improved (P13)”.

4.1.3 Hopes for Future AI applications. A majority of the participants saw instances in which AI applications in the future might help them in their daily lives as it relates to their disabilities (n=17). Many of the types of AI applications that the participants brought up were those that exist today but are not fully effective. These included more powerful voice assistants and smart home devices that could automate more functions of living space for those with physical disabilities (n=3; 33.3%), advanced event calendars for those with ADHD (n=2; 33.3%), and more accurate speech-to-text applications for those who prefer to type using their voice (n=3; 50%). Other applications that the participants brought up have not yet been commercialized but have been discussed in the media as near-future AI applications (n=10), such as self-driving cars and robotics tools that can help those with physical disabilities to move around (n=3; 33.3%), ASL translators for facilitating conversation between ASL communicators and non-signers (n=4; 66.7%), and visual navigation tools that can help blind pedestrians navigate streets (n=3; 33.3%).

4.2 Motivations for Contributing Data

4.2.1 Willingness to contribute. Having explained to our participants how we define AI, its training process, and the importance of representative datasets, we probed their perception of contributing data to an AI dataset. Many were open to contributing so long as their data is used to help them or their disability communities (n=24), some even without compensation (n=14). But their willingness to contribute still depended on a number of elements. For instance, some considered the type of data that is being collected to be important when deciding whether to contribute (n=6). Unsurprisingly, they were hesitant if the collected data was more personal, such as their photos or legal documents. This concern was particularly stressed by those who experienced a “hidden disability” [53] such as attention-deficit hyperactivity disorder (ADHD), dyslexia, or post-traumatic stress disorder (PTSD), as these participants saw their disclosure of disability as “optional.” We go deeper into these concerns for contributing data in 4.3.

We also found that the purpose of the data collection influenced our participants’ willingness to contribute. Not only did they want their contribution to help disability communities, but some also pointed out that they did not want their data to be used in certain ways (n=3), including for law enforcement and improving public relations to make a for-profit company appear inclusive only on the surface (i.e., “ethics washing” or “diversity theater” [4]). A participant noted: “Maybe they get one or two people with disability and like, OK, well, we met the quota and so we can move on with it but not really testing to see how does it really... work for someone with a disability” (P30). In the post-interview survey, we probed deeper into this topic by presenting to our participants five ways their data might be used to create inclusive AI (e.g. to develop AI-powered accessibility tools, to raise accessibility awareness in AI) and asked whether these were acceptable on a seven-point Likert scale. The

Table 2: Level of comfort for contributing data reported on 7pt. Likert scales (1 = strongly disagree; 7 = strongly agree).

By the Usage of Contributed Data	<i>Average</i>	<i>STD</i>
Develop AI accessibility tools	6.82	0.39
Make general AI applications more inclusive	6.17	1.63
Used to test AI for fairness	6.41	1.02
Used to teach how to develop AI	6.10	1.47
Used to raise accessibility awareness in AI	6.54	1.07
By the Type of Data		
Self-portrait photo	5.30	1.56
Photo of an object	6.52	1.28
Video of speaking	5.60	1.78
Video of moving	5.13	1.70
Type what you hear	6.16	1.60
Browser dexterity	6.42	1.63
By the Type of Metadata		
Name	5.14	1.96
Contact Info	5.00	1.87
Age	6.31	1.11
Disability	6.38	1.27
Gender	6.29	1.41
Ethnicity	6.39	1.17
Education	6.39	1.07
Income	5.19	2.00
By the Data-Collecting Organization		
Large tech company	5.59	1.57
Small start-up	4.86	1.75
Public university	5.62	1.37
Private university	5.62	1.50
Disability focused university	6.24	1.15
Disability advocacy organization	6.28	1.13
Government	4.56	1.87

five goals we presented induced enthusiastic responses from our participants with the average responses for these use cases scoring higher than six out of seven as summarized in Table 2.

Finally, the importance of the impact of their contribution was a theme amongst the participants (n=5), with some explicitly noting that they would be more willing to contribute their data if they are a member of a “harder to access” group whose contribution would bear a significant weight or if the AI application that is being built is in a post-prototype stage and its success is more likely with the proper data contribution.

4.2.2 Monetary compensation. We subsequently explored how monetary compensation might or might not motivate our participants to contribute their data to an AI dataset. Although a few mentioned that monetary compensation was not necessary if their contribution helps others and themselves (n=4), a majority of the participants saw monetary compensation as an added incentive for contributing their data (n=19). For many, the reason was straightforward: “Yeah, I mean, why not? If you get some money in the process,

that's a good incentive, right?" (P11). But for some, the need for monetary compensation was closely tied to their disability status, which made monetary compensation more appealing. A participant noted, "You know, being disabled, you are always on the budget... the monetary, you know, would make me wanna [contribute] more" (P18). But this caused some to see monetary compensation as not a constructive source of motivation for contributing their data as some people might feel compelled to contribute their data because of their financial difficulties, as expressed by this participant: "... what's even more important with personal data is people who are in need of money might feel compelled to go further than they would otherwise with sharing their data or sharing their information" (P5). Another participant mentioned perhaps a graver concern by noting that "... that might be a little bit more... pressure point for people... It sometimes feels like you are selling yourself out a little" (P9).

4.2.3 Determining compensation amount. With the monetary compensation still being viewed as an effective motivator by many of our participants, we proceeded to investigate how the participants reasoned about determining the appropriate level of compensation by asking them how much they should be paid for a given data type. Many of the factors that our participants considered when reasoning about this resonated with what they considered when determining whether they were interested in contributing in the previous subsection. They considered what type of data was being collected (n=8): is it identifiable or anonymous, is the data already out there on the internet (e.g. on social media), and how comfortable am I personally to share this type of data? They also considered the purpose of the data collection (n=4), explicitly or implicitly concluding that they would be willing to contribute for less compensation if the data collected will be used to better their lives or the lives of those in disability communities.

Additionally, the participants found the resources needed to collect and upload their data to be relevant, noting that PWD may need to invest more time and even recruit help to complete the contribution task (n=6). Similarly, some noted that it would be helpful for them to be aware of how people are paid in other seemingly related contexts such as photographers submitting their images to a media company; "if I were a freelance photographer, and I work, you know, taking pictures for deaf people, and I had a price already set for things that I did for as a photographer, then I could have an estimate of what I might ask in return for" (P10).

However, even though our participants had ideas for what would be relevant points of consideration when determining the compensation, they found the activity of coming up with the exact compensation level daunting. When asked to provide an exact dollar amount for four different types of data (a selfie, a video of them moving, a video of them speaking, and a photo of a household object), many responded that they were not sure how to provide an answer (n=15): "Oh man, I don't know. I've always been terrible with money" (P15). What helped many participants to answer this question was to actually try out the data contribution tasks. Whereas only 9.75 (STD=1.26) or 32.5% of the participants on average across the four types of data provided a dollar compensation amount prior to actually completing the contribution tasks, 21 (STD=2) or 70.0% of the participants on average did after completing the tasks.

We analyzed the compensation amount that the participants indicated as appropriate after completing the sample data capture and upload tasks. The overall trend suggests that a simple photo of an object that does not reveal the participants is considered the least valuable (median=\$8.75), followed by a self-portrait photo (median=\$20.00), videos of speech (median=\$40.00) and motion (median=\$50.00) which are considered the most valuable data type in this set. Interestingly, however, the participants' response to what is a reasonable compensation contained a few very high outliers where the suggested amount was higher than \$1,000 for a single contribution (e.g. an image or a video file). There were three out of 23 such responses for selfie portrait photos, one out of 23 about a video of them speaking, and one out of 23 about a video of them moving. Although these participants did not provide clear reasoning for this, we suspect that this might be an illustration of them not wanting to contribute that particular type of data. As a case in point, there were no such dramatic outlier values for contributing a photo of an object, which was considered to be less privacy sensitive.

4.2.4 Other forms of compensation. We also explored other types of compensation. We started by suggesting an option in which the data-collecting organization offers to donate money to a non-profit organization of our participants' choosing (e.g. an organization advocating for disability justice). This was acceptable for a large portion of our participants (n=14), with a few of them expressing their enthusiasm as they are involved with disability-related non-profit organizations and are familiar with the work done by these organizations (n=5). However, this form of compensation was met with a lukewarm response from others who noted that it would not necessarily generate additional incentive to contribute to a dataset (n=4). This was because it seemed difficult to confirm how the money was used, and it's always possible for the participants to directly donate the money themselves if they wanted to.

Beyond donation-based compensation, some also noted their interest in other non-monetary forms of compensation such as credit attribution for the data they contributed or offers for publicity by appearing in advertisements for accessible technology (n=3). Additionally, some participants suggested that they would like to be able to observe the progress of AI in return for their contribution (n=7) or even get early access to use the trial version of the technology that was built with their data (n=5). For them, it was more important that they were a part of the process for developing new beneficial technology: "... the excitement of participating in the technology, making it better... I would value probably a long term relationship on developing the technology more than the monetary compensation. That is... hey, you were one of the early people in study. We've got a piece of AI we would like you to try out. We'd like to find out if it's screen reader accessible [for blind users]" (P8).

4.3 Concerns about Contributing Data

4.3.1 Physical or psychological concerns. To understand our participants' concerns about contributing to an AI dataset, we described to them a scenario in which a nonprofit organization advocating for disability justice wants to gather data from PWD to create a more inclusive AI application. Many noted that if the data contribution task is not beyond what is often done online (e.g. uploading a photo or a video), they were not concerned that the process might

TASK	Completion Time (Average in Seconds)	Completion Rate (%)	Task Difficulty (7pt. Likert Average)
1. Upload a self-portrait	254; 121; 109; 330; 217	78; 100; 100; 100; 93	2.66; 4.60; 4.00; 2.86; 3.35
2. Upload a photo and a caption	191; 109; 98; 189; 155	78; 100; 100; 100; 93	3.14; 1.75; 1.00; 2.00; 2.28
3. Upload a video of speaking	348; 139; 434; 335; 326	78; 100; 100; 89; 90	2.50; 2.40; 1.50; 3.17; 2.58
4. Upload a video of moving	222; 212; 366; 391; 303	78; 67; 100; 78; 80	4.43; 4.00; 4.50; 4.00; 4.23
5. Type what you hear	384; 210; 155; 352; 294	89; 100; 100; 100; 97	2.71; 5.67; 2.33; 4.86; 4.09
6. Click the dots on the screen	64; 55; 34; 87; 62	44; 100; 100; 100; 83	5.20; 2.80; 1.00; 2.75; 3.31

Table 3: Statistics of the six data contribution tasks presented in this order: participants experiencing blindness or visual impairment; cognitive disability; hearing loss or impairment; physical disability; all participants. For the completion time, the averages are without the outliers (three STD from the mean). For task difficulty, seven represents the most difficult.

harm them physically or psychologically (n=13). But some noted important caveats. For example, a deaf participant stressed that the data collection process needs to be respectful towards the specific cultures that a disability community might have (P6). To the Deaf² community, a sign language is considered not only a means of communication but also a cherished cultural element that identifies the community [8]. Therefore, when collecting data from members of the Deaf community, it would be important to offer a sign language interpretation for the instructions even when the participants are fluent in a written language. In addition, some noted that the data collection process could induce performance anxiety, particularly for PWD (n=3): “So you are dealing with psychologically injuring people. I look normal, but I’ve had a lot of difficulty in school and failed out of college. So there’s a fear of failure that follows” (P20). Another participant who shared a similar sentiment stressed that for this reason, the data collection should be “done in such a way that it’s not punitive. If you make a mistake... you are not told that you are out or we are never doing business with you again” (P7).

4.3.2 Privacy concerns. Our participants showed a bimodal reaction in terms of their privacy concerns, with some expressing much stronger concerns than others. Many noted that they were not overly concerned about privacy when contributing their data (n=16) and justified this lack of concerns by mentioning that these types of data are already online, for example on social media (n=6). These participants also noted that they are open about the disabilities they experience, and any potential disclosure of disability from the data shared is less of a concern (n=6). However, for those less open about their disabilities, especially when they experienced non-apparent forms of disabilities like ADHD or dyslexia and saw their disclosure of disabilities as optional in their daily lives, their concerns were more prominent (n=3; 50.0%): “Disclosure is optional for some people... And it’s an option that we don’t take lightly...” (P9).

Our participants mentioned that if the data being collected is less identifiable and unlikely to reveal non-apparent forms of disability, they would be less concerned about their privacy (n=9). This was also represented in their answers to the survey questions summarized in Table 2. One-way ANOVA on our participants’ response to the questions that asked how comfortable they would

be with contributing different types of data shows that the data type significantly influenced the degree of comfort for contributing ($F(4.232, 102.4)=4.293, p=0.025$). The followup Tukey’s test for post-hoc analysis indicates that uploading a photo of an object was more comfortable than uploading a self-portrait ($p=0.012$), or videos of them talking ($p=0.018$) or moving ($p=0.0051$). We observe a similar trend when analyzing the participants’ level of comfort for sharing their demographic information as metadata for a dataset. Once again, one-way ANOVA shows that the type of metadata plays a significant role in determining the comfort level for sharing the data ($F(3.456, 92.81)=9.837, p<0.0001$), and the post-hoc analysis indicates that the participants were significantly less comfortable to share identifiable information like their name and contact information.

Another recurring theme related to privacy concerns was the participants’ trust in the data-collecting organization (n=5). Many thought if the organization worked to advocate for PWD, they felt safer to share data and more confident that any terms in the consent form would be kept; “The Arthritis National Research Foundation, which I am involved with, I would trust [with my data]... Yeah, it would really depend on how reputable they are” (P25). We investigated this topic more thoroughly in the survey portion; Table 2 summarizes the results. Once again, one-way ANOVA shows that the type of data-collecting organization played a significant role in determining the participants’ level of comfort for sharing their data ($F(3.331, 91.05)=12.38, p<0.0001$) and the Tukey’s test for post-hoc analysis support that organizations that are oriented towards supporting disability communities are seen as more reliable for handling the data than a small start-up company or the government.

4.3.3 Intersectional concerns. Our results indicate additional intersectional identity concerns that may affect PWD’s willingness to contribute. These include those related to age and ethnicity, and show how different norms and social risks for disclosing disability can factor into PWD’s decision to contribute. Regarding age, our participants noted that there are generational differences in terms of how open one is to sharing their disabilities, that older people are going to share less; “you know, 50 and older... because they grew up in a generation where having a disability was supposed to be something that is secretive and private... [when] they have an illness or they become disabled, then they are not usually willing to share it until it becomes a problem” (P30). Additionally, for ethnicity, when one is a part of a historically disadvantaged ethnic

²We use “Deaf” with a capital “D” when referring to Deaf culture and with a lowercase “d” when referring to hearing status.

group, that person may be more likely to not want to disclose their disability due to the more acute fear of discrimination; “just being a person of disability, part of the disability community, also being an African American in today’s US climate... I wouldn’t want to give anyone an opportunity to discriminate against me” (P9).

4.4 Challenges Contributing Data

4.4.1 Data capture. We asked our participants to capture and upload six different types of data as summarized in Table 3 and asked whether they experienced challenges or needed external help. We note that needing external assistance is not necessarily problematic, and indeed that *interdependence* [5] is important to and valued by many PWD; however, it is important for system designers to understand if and how assistance may be required for data upload tasks to create an accessible system. We find that a majority of our participants (over 80% of the participants across the six data types) were able to collect and upload their data as requested. However, perhaps as expected, they faced various accessibility issues in the process that were heavily depended upon each of the participants and the type and prominence of disabilities they experienced. For example, participants with physical disabilities reported challenges in steadily holding a phone when taking videos or photos when collecting data (although this could be desirable if the goal of data collection is to train models that can recognize wobbly pictures and videos to better support this population) (n=3; 33.3%). Furthermore, physical disability also made it more challenging for the participants to type, with some reporting that they needed multiple pauses when typing or experienced pain when typing (n=4; 44.4%).

Blind participants reported difficulty with taking photos of an object or themselves as they could not be sure if the photos were properly focused or framed (n=3; 33.3%). Also, the multi-step process of having to take the photo on a phone and then transferring the file to another machine with which they were taking the survey added an extra dimension of accessibility challenges for these participants (n=4; 44.4%); “Had to take the picture with my phone... then open the ‘Photos’ app, email the photo to my desktop, open in Outlook, save the attachment to my local folder, then upload. Would be easier if everything was in one data collection app” (P8). The task of clicking dots that appear on a screen for measuring dexterity, which relies heavily on visual signal, was simply very difficult for many of the blind participants, with only three of the nine blind participants finishing the task.

Those with cognitive disabilities such as ADHD and dyslexia, on the other hand, documented challenges with tasks involving reading and typing (n=5; 55.6%). Furthermore, beyond the challenges experienced in this study, a participant who experiences ADHD noted that some with ADHD, such as himself, would find it challenging to sit through a data collection procedure that takes longer than a certain amount of time (P22). This may require them to take behavior-altering medications, which might help them to get through the data collection process but compromise the validity of the collected data depending on the purpose of the data collection.

Finally, for some deaf participants, aside from written English not being the primary language, they were concerned that the spectrum of disability and cultural identity can pose a challenge for data contribution (though we offered ASL interpretation of all content,

they were cognizant that many data collection platforms might not). One remarked: “There are different levels of hearing and hearing loss. Genetic deafness, people who are hard of hearing, people who have lost their hearing, people who function orally, people who function with sign language and cultural Deafness” (P6). In the case of P20, the participant described himself as culturally Deaf, that he was born without hearing in a family heavily involved in the Deaf culture. This meant that he was not only fluent in sign language but also signing was a part of the cherished culture. But for P21, the hearing loss was gradual with “80% hearing loss in my left ear and 20% in my right ear... there’s a point some years from now where I will probably be totally deaf” (P28). This participant chose to communicate using spoken English and expressed that sign language is difficult. This suggests that an inclusive data collection process needs to be mindful of the cultural aspects around disabilities, and be aware that even when collecting data from two deaf participants (e.g. videos of them communicating via ASL), there could be a difference in the fluency of ASL between the participants.

4.4.2 Resources needed. Depending on the type of disability our participants experienced and the tasks they were given, the participants had to invest more resources (e.g. more time or the need to recruit extra help) to complete the tasks than others. When completing tasks that asked the participants to take photos or videos of themselves, for instance, participants who are blind or experience physical disabilities noted that they asked a member of their family to help them with the task (n=7; 38.9%). However, not all participants had someone else available to help them complete the task even when they thought they needed help either because they physically had no one else around, or their family members also experienced a similar form of disability (n=3; 16.7%). Finally, some noted that their disabilities made tasks more cumbersome and time-consuming (n=4): “It took more time than it would take someone who is sighted to align the camera with my face” (P20).

5 DISCUSSION

Seeking data contributions from PWD could help create more inclusive AI datasets if the contribution process is designed to respect, protect, and motivate those who participate. In this section, we synthesize our findings and outline concrete design guidelines for building a process to collect data from PWD.

5.1 Design Guidelines for Data Collection

5.1.1 How to motivate. Monetary compensation could be effective for motivating PWD to contribute data for AI. If the participants were to receive monetary compensation, it would be useful to provide them with a relevant benchmark to justify its amount. For example, some of our participants compared the data collection process to a media company receiving photos from a professional photographer. Although a media company may not be the most appropriate point of comparison (i.e., since a professional photographer is being paid based on artistic skill), a proper benchmark can help the participants reason about the compensation amount. In the absence of such a benchmark, they might set their own benchmark that may not be fitting for a real running system. We suspect that this was the case in the portion of our study that tried to understand our participants’ expected compensation for their data, which

likely illustrated the *relative value* of different tasks and data types accurately but not the *market price* of such data.

Estimates of the time or effort required for the data collection will also help participants form the correct mental model; it is important to ensure such estimates are based on trial data collections by PWD using the types of assistive technologies that target participants would use. Additionally, certain data collection and uploading tasks can take more resources for PWD in the forms of longer completion time and the need to recruit help. To use monetary compensation, it may be necessary to determine the true market price of such data by studying how the different price points affect the participation rate and find a price that is effective, but not unethically low.

However, we also stress that monetary compensation might not always work. Offering monetary compensation may pressure PWD who are unemployed or have a lower income to participate. This can skew datasets towards certain socioeconomic brackets while being exploitative. In addition, motivating the participants with high monetary compensation may lead to lower data quality by encouraging them to complete the data contribution task without fully engaging in them, as is sometimes the case with crowd workers [15]. Therefore, it may be advisable to also consider providing non-monetary forms of compensation such as matching donations to a nonprofit organization of the participant's choosing, a free trial for the technology being developed with the collected data, or creating a sense of community effort for building AI technology that meets the needs of people with disabilities.

5.1.2 What to communicate. Be upfront about the data collection goal and indicate any privacy concerns there might be, especially if the data collected may reveal non-apparent forms of disabilities like ADHD and dyslexia. Though many participants indicated that they were not too concerned with issues of privacy, note that there might be limitations for how well laypeople can anticipate the privacy impacts of the data they share. People's public information on social media may be used by AI and injure them in ways they cannot anticipate (e.g. different rates for services and job discrimination) or used in ways they do not condone down the line (e.g. for data surveillance and persecution). We need an ongoing conversation between data contributors and AI developers or privacy experts on what is the appropriate use of any collected data.

5.1.3 An accessible process. In addition to following the standard accessible design practices, the online infrastructure for enabling data contribution from PWD should work to make the entire process accessible. Contributing data is often a multi-step process with multiple devices. To contribute a photo, a participant might use a phone to take the photo and send it to the machine from which the collection is done to upload. Each step can be a barrier for PWD and should be simplified as much as possible, for example, by building a mobile app with which the participant can collect and upload data on a single device. Also take into consideration how the spectrum of difference within disability categories might affect participants' ability to contribute. We discussed this when presenting our findings from deaf participants who ranged from those identifying as culturally Deaf to those who were getting accustomed to the Deaf culture gradually. This resonates with Bragg et al.'s findings [8] and we conjecture that the spectrum of disabilities will play an important role across other forms of disabilities (e.g., photos collected

from someone who is legally blind but has some functional vision may differ substantially from those with no vision at all). This indicates that simplistic metadata about disability categories may not suffice for interpreting data and/or ensuring diverse coverage of target populations; however, collecting more granular descriptions of disability status further increases privacy risks. Finally, the process of data collection should not be punitive towards its participants by blocking them for their mistakes or setting up stringent time limits as there can be more conscious performance anxiety for PWD. Unfairly rejecting PWD's contribution due to minor mistakes or slow performance time as observed in existing practices by Zyskowski et al. [60] can place an unnecessary burden on contributors.

5.2 Limitations and Future Work

Our study is based on the premise that more inclusive AI datasets and the more accessible AI applications will benefit PWD. While this will be true in many contexts, there are important normative questions around the role of AI and cases where disability-aware AI applications will not be the right solution. Though we do not consider this directly here, we need an ongoing conversation about what AI will mean for various disability communities and what role inclusive AI datasets will play for them. Additionally, we note limitations to our study that point to concrete directions for future work. First, the self-reported measures that we used, though widely accepted, may not fully capture people's behaviors when interacting with a deployed system; our findings can inform the design of data collecting infrastructures, but they should be updated as such infrastructures are deployed (also note that the results may vary if the infrastructure is launched outside the US). In addition, we focused our study on data contributors' perspectives of data contribution. Although this is an important part of designing an inclusive data-collection infrastructure, our findings need to be supplemented by the perspectives of domain experts in privacy and law. Finally, the details of implementing such infrastructure that combines the ethical and technical issues such as where and for how long the data should be stored, and how to effectively remove already contributed data if the participant changes their decision about their contribution, need to be carefully considered.

6 CONCLUSION

An online infrastructure for enabling data contribution from PWD can open up a path toward creating more inclusive AI datasets at scale. However, for such an effort to succeed, the process of data contribution needs to be designed so that it is motivating, safe, and accessible for PWD. To better understand these design challenges, we analyzed our interviews with 30 participants who experience a diverse array of disabilities as well as the results of an online questionnaire that included six sample data contribution tasks. We identified what might motivate our participants, and what concerns and challenges they might have for contributing data to an AI dataset. Finally, we synthesized what we learned into guidelines for designing such an infrastructure, and posed questions that need to be considered when soliciting data from PWD. We hope our findings will help the creation of inclusive AI datasets that benefit PWD by directly and ethically sourcing data from this community.

REFERENCES

- [1] Facebook AI. 2019. *Does object recognition work for everyone? A new method to assess bias in CV systems*. <https://ai.facebook.com/blog/new-way-to-assess-ai-bias-in-object-recognition-systems/>
- [2] Google AI. 2020. *Google Euphonia*. <https://sites.google.com/view/project-euphonia/>
- [3] Apple. 2019. *Introducing Voice Control. Your all-access to all devices*. Retrieved Aug 1, 2020 from www.apple.com/macros/catalina-preview/#accessibility
- [4] Yochai Benkler. 2019. *Don't let industry write the rules for AI*. <https://go.gale.com/ps/anonymou?id=GALE%7CA584200722&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=00280836&p=HRCA&sw=w>
- [5] Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. 2018. Interdependence as a Frame for Assistive Technology Research and Design. In *ASSETS '18: Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*.
- [6] Cynthia L. Bennett and Os Keyes. 2020. What is the Point of Fairness? Disability, AI and the Complexity of Justice. In *SIGACCESS Access*, Vol. 125. <https://doi.org/10.1145/3386296.3386301>
- [7] Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, and Tom Yeh. 2010. VizWiz: Nearly Real-time Answers to Visual Questions. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology*.
- [8] Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, Christian Vogler, and Meredith Ringel Morris. 2019. Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*.
- [9] Danielle Bragg, Oscar Koller, Naomi Caselli, and William Thies. 2020. Exploring Collection of Sign Language Datasets: Privacy, Participation, and Model Performance. In *In The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*.
- [10] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *In Conference on Fairness, Accountability and Transparency*. 77–91.
- [11] Wikipedia Commons. 2020. *Dragon NaturallySpeaking*. https://en.wikipedia.org/wiki/Dragon_NaturallySpeaking
- [12] Wikipedia Commons. 2020. *Seeing AI*. https://en.wikipedia.org/wiki/Seeing_AI
- [13] Wikipedia Commons. 2020. *Synthetic Data*. https://en.wikipedia.org/wiki/Synthetic_data
- [14] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *In 2009 IEEE conference on computer vision and pattern recognition*. 248–255.
- [15] Julie S. Downs, Mandy B. Holbrook, Steve Sheng, and Lorrie Faith Cranor. 2010. Are Your Participants Gaming the System? Screening Mechanical Turk Workers. In *Proceedings of the 2010 CHI Conference on Human Factors in Computing Systems*.
- [16] Equality and Human Rights Commission. 2020. *Disability discrimination*. <https://www.equalityhumanrights.com/en/advice-and-guidance/disability-discrimination>
- [17] Leah Findlater, Steven Goodman, Yuhang Zhao, Shiri Azenkot, and Margot Hanley. 2020. Fairness Issues in AI Systems That Augment Sensory Abilities. In *SIGACCESS Access*, Vol. 125. <https://doi.org/10.1145/3386296.3386304>
- [18] Krzysztof Gajos. 2016. *Lab in the Wild and Other Tools for Large-scale Behavioral Research*. <http://www.eecs.harvard.edu/~kgajos/research/>
- [19] Google. 2019. *YouTube Help – Use automatic captioning*. Retrieved Aug 1, 2020 from <https://support.google.com/youtube/answer/6373554?hl=en>
- [20] Google. 2020. *Google Play*. https://play.google.com/store/apps/details?id=com.google.android.apps.accessibility.reveal&hl=en_US
- [21] Linda G Gottermeier and S Kushalnagar Raja. 2016. User Evaluation of Automatic Speech Recognition Systems for Deaf-Hearing Interactions at School and Work. In *Audiology Today*, Vol. 28. 20–34.
- [22] Troy Griggs and Daisuke Wakabayashi. 2018. How a Self-Driving Uber Killed a Pedestrian in Arizona. *The New York Times* (2018). <https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html>
- [23] Anhong Guo, Ece Kamar, Jennifer Wortman Vaughan, Hanna Wallach, Sasa Junuzovic, Besmira Nushi, Jacquelyn Krone, and Meredith Ringel Morris. 2020. Evaluating Face-Based AI Systems for Fairness: Challenges and Tradeoffs. (2020).
- [24] Anhong Guo, Ece Kamar, Jennifer Wortman Vaughan, Hanna Wallach, and Meredith Ringel Morris. 2020. Toward Fairness in AI for People with Disabilities: A Research Roadmap. In *SIGACCESS Access. Comput*, Vol. 125. <https://doi.org/10.1145/3386296.3386298>
- [25] 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 *Computer Vision and Pattern Recognition*.
- [26] Gurari Gurari, Qing Li, Chi Lin, Yanan 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 (CVPR)*.
- [27] Sara Hussein. 2018. 'Citizen scientists' track radiation seven years after Fukushima. (2018). <https://phys.org/news/2018-03-citizen-scientists-track-years-fukushima.html>
- [28] Ben Hutchinson, Vinodkumar Prabhakaran, Emily Denton, Kellie Webster, Yu Zhong, and Stephen Denuyl. 2020. Unintended Machine Learning Biases as Social Barriers for Persons with Disabilities. In *SIGACCESS Access*, Vol. 125.
- [29] Joke Kenens, Michiel Van Oudheusden, Go Yoshizawa, and Ine Van Hoyweghen. [n.d.]. Science by, with and for citizens: rethinking 'citizen science' after the 2011 Fukushima disaster. ([n. d.]).
- [30] Jungsoek Joo Kimmo Karkkainen. 2019. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age. *arXiv preprint* (2019).
- [31] Hanna Krasnova, Helena Wenninger, Thomas Widjaja, and Peter Buxmann. 2013. Envy on Facebook: a hidden threat to users' life satisfaction? *International Conference on Wirtschaftsinformatik* (2013).
- [32] Petrino L. O'Neill P, and Jorgensen M. 1993. An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology* (1993).
- [33] Jaron Lanier. 2019. *Jaron Lanier Fixes the Internet*. <https://www.nytimes.com/interactive/2019/09/23/opinion/data-privacy-jaron-lanier.html>
- [34] Steve Lohr. 2018. Facial Recognition Is Accurate, if You're a White Guy. *The New York Times* (2018). <https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html>
- [35] Aarian Marshall and Alex Davies. 2019. Uber's Self-Driving Car Didn't Know Pedestrians Could Jaywalk. *Wired* (2019). <https://www.wired.com/story/ubers-self-driving-car-didnt-know-pedestrians-could-jaywalk/>
- [36] Sharan B. Merriam and Associates. 2002. *Introduction to qualitative research. Qualitative research in practice: Examples for discussion and analysis*. Jossey-Bass.
- [37] S. Mirri, C. Prandi, P. Salomoni, F. Callegati, and A. Campi. 2014. On Combining Crowdsourcing, Sensing and Open Data for an Accessible Smart City. In *2014 Eighth International Conference on Next Generation Mobile Apps, Services and Technologies*.
- [38] Meredith Ringel Morris. [n.d.]. AI and Accessibility: A Discussion of Ethical Considerations. ([n. d.]).
- [39] Meredith Ringel Morris, Shaun Kane, and Anhong Guo. [n.d.]. Sense and Accessibility. ([n. d.]).
- [40] Craig S. Mullins. 2020. *Data Management Today by Craig Mullins*. https://web.archive.org/web/20090721111006/http://www.neon.com/blog/blogs/cmullins/archive/2009/02/05/What-is-Production-Data_3F00_.aspx
- [41] Karen Nakamura. 2019. My Algorithms Have Determined You're Not Human: AI-ML, Reverse Turing-Tests, and the Disability Experience. In *21st International ACM SIGACCESS Conference*. 1–2.
- [42] Antony Nicol, Christopher Casey, and Stuart J MacFarlane. 2002. Children are ready for speech technology-but is the technology ready for them. *Interaction Design and Children* (2002). https://www.researchgate.net/publication/228574580_Children_are_ready_for_speech_technology-but_is_the_technology_ready_for_them
- [43] Nils J. Nilsson. 2010. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge University Press, Cambridge, UK.
- [44] Robert Pear. 2019. On Disability and on Facebook? Uncle Sam Wants to Watch What You Post. *The New York Times* (2019). <https://www.nytimes.com/2019/03/10/us/politics/social-security-disability-trump-facebook.html>
- [45] Ruchir Puri. 2018. *Mitigating Bias in AI Models*. Retrieved Aug 1, 2020 from <https://www.ibm.com/blogs/research/2018/02/mitigating-bias-ai-models/>
- [46] Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. 2010. Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing. In *AAAI/ACM AI Ethics and Society conference*.
- [47] James A. Rodger and Parag C. Pendharkar. 2004. A field study of the impact of gender and user's technical experience on the performance of voice-activated medical tracking application. *International Journal of Human-Computer Studies* 60, 5-6 (2004), 529–544. <https://doi.org/10.1016/j.ijhcs.2003.09.005>
- [48] Bill Schackner. 2020. *Pitt benches food-delivery robots after complaint from a student*. <https://www.post-gazette.com/business/tech-news/2019/10/23/robot-starship-university-of-pittsburgh-disabilities-higher-education-campus-life/stories/201910230137>
- [49] Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, and Astro Teller. 2016. *Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*. Retrieved Aug 1, 2020 from <http://ai100.stanford.edu/2016-report>
- [50] Rachael Tatman. 2017. Gender and Dialect Bias in YouTube's Automatic Captions. In *In Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. 53–59.
- [51] Jutta Treviranus. 2018. *Sidewalk Toronto and Why Smarter Is Not Better*. Retrieved Aug 1, 2020 from <https://medium.com/datadriveninvestor/sidewalk-toronto-and-why-smarter-is-not-better-b233058d01c8>

- [52] Shari Trewin. 2018. AI Fairness for People with Disabilities: Point of View. *arXiv preprint arXiv* (2018). <https://arxiv.org/pdf/1811.10670.pdf>
- [53] Aimee Valeras. 2020. “We don’t have a box”: *Understanding Hidden Disability Identity Utilizing Narrative Research Methodology*. <https://dsq-sds.org/article/view/1267>
- [54] Luis von Ahn and Laura Dabbish. 2004. Labeling Images with a Computer Game. In *CHI*.
- [55] Luis von Ahn and Laura Dabbish. 2008. Designing Games With A Purpose. In *communications of the acm*, Vol. 51.
- [56] Chris Welch. 2019. Android Q’s Live Caption feature adds real-time subtitles to any audio or video playing on your phone. *The Verge* (2019). <https://www.theverge.com/2019/5/7/18528447/google-android-q-live-caption-video-transcription-io-2019>
- [57] Meredith Whittaker, Meryl Alper, Cynthia L. Bennett, Sara Hendren, Liz Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Rankin, Emily Rogers, Marcel Salas, and Sarah Myers West. 2019. Disability, Bias, and AI. *AI Now Institute* (2019). <https://ainowinstitute.org/disabilitybiasai-2019.pdf>
- [58] S.N. Woods, Michael Leonard Walters, Kheng Lee Koay, and Kerstin Dautenhahn. 2006. Comparing human robot interaction scenarios using live and video based methods: Towards a novel methodological approach. In *In: Proceedings of 9th IEEE international workshop on advanced motion control*.
- [59] John Zimmerman, Anthony Tomasic, Charles Garrod, Daisy Yoo, Chaya Hiruncharoenvate, Rafae Aziz, Nikhil Ravi Thiruvengadam, Yun Huang, and Aaron Steinfeld. 2011. Field trial of Tiramisu: crowd-sourcing bus arrival times to spur co-design. In *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
- [60] Kathryn Zyskowski, Meredith Ringel Morris, Jeffrey P Bigham, Mary L Gray, and Shaun K Kane. 2015. Accessible crowdwork? Understanding the value in and challenge of microtask employment for people with disabilities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work Social Computing*.