Published in final edited form as:
ASSETS. 2022 October ; 2022: . doi:10.1145/3517428.3544826.

Data Representativeness in Accessibility Datasets: A Meta-Analysis

Rie Kamikubo,
College of Information Studies, University of Maryland, College Park, United States

Lining Wang,
Department of Computer Science, University of Maryland, College Park, United States

Crystal Marte,
College of Information Studies, University of Maryland, College Park, United States

Amnah Mahmood,
Department of Mathematics, University of Maryland, College Park, United States

Hernisa Kacorri
College of Information Studies, University of Maryland, College Park, United States

Abstract

As data-driven systems are increasingly deployed at scale, ethical concerns have arisen around unfair and discriminatory outcomes for historically marginalized groups that are underrepresented in training data. In response, work around AI fairness and inclusion has called for datasets that are representative of various demographic groups. In this paper, we contribute an analysis of the representativeness of age, gender, and race & ethnicity in accessibility datasets–datasets sourced from people with disabilities and older adults—that can potentially play an important role in mitigating bias for inclusive AI-infused applications. We examine the current state of representation within datasets sourced by people with disabilities by reviewing publicly-available information of 190 datasets, we call these accessibility datasets. We find that accessibility datasets represent diverse ages, but have gender and race representation gaps. Additionally, we investigate how the sensitive and complex nature of demographic variables makes classification difficult and inconsistent (*e.g.*, gender, race & ethnicity), with the source of labeling often unknown. By reflecting on the current challenges and opportunities for representation of disabled data contributors, we hope our effort expands the space of possibility for greater inclusion of marginalized communities in AI-infused systems.

Keywords

AI FATE; datasets; inclusion; diversity; representation; accessibility; aging

rkamikub@umd.edu .

Keywords

Human-centered computing → Human computer interaction (HCI); Accessibility; Social and professional topics → People with disabilities; Age; Gender; Race and ethnicity

1 INTRODUCTION

As AI-infused systems[1] become ubiquitous, ensuring that they work for a diversity of groups is vital [29, 56, 108]. Performance disparities in these systems could lead to unfair or discriminatory outcomes for historically and culturally marginalized groups, such as on the basis of gender, race, or disability [12, 18, 44, 149, 162, 172]. One fundamental source of disparities is the lack of representation in datasets used to train machine learning models and benchmark their performance [108, 162, 179]. A notable example comes from Treviranus [166], where during a simulation, she found that machine learning models for autonomous vehicles would run over someone who propels themselves backward in a wheelchair. Merely adding training examples of people using wheelchairs did not have the intended effect in this case; the algorithm failed with a higher confidence [166]. Treviranus suspected 'backward propelling' was still an outlier.

In this important discussion on AI fairness and inclusion, tensions around data representativeness involving disability [60, 79, 118] have also arisen. Data sourced from accessibility datasets can help AI-infused systems work better when deployed in real-world scenarios, both for assistive and general-purpose contexts [29, 75, 169]. However, privacy and ethical concerns are especially pronounced in this community, as disclosure of disability can pose risks associated with re-identification and further discrimination *e.g.*, for one's healthcare and employment [169, 179]. People who have distinct data patterns, like in the case of disability, are also more susceptible to data abuse and misuse [1, 60, 167]. In addition, even if AI-infused systems are trained with diverse data, this does not inherently challenge the power structures in which these systems are embedded, which may be the actual source of harm and marginalization for disabled people [7]. For example, a more equitable AI-infused system for diagnosing autism does not necessarily correspond to greater well-being of autistic people, because it may cement the power that medical institutions have to diagnose and gatekeep [7].

We contribute to these discussions via our exploration of representation in accessibility datasets, which reveal nuanced patterns of representation and marginalization along intersectional lines. In this work, we conducted a metadata analysis of existing accessibility datasets (1984–2021, N=190) spanning multiple communities of focus and data types to understand the representation and reporting of demographic attributes including age, gender, and race & ethnicity of data contributors. We used the publicly available documentation and resources of these datasets to explore the potential opportunities and limitations for increasing data representativeness.

[1]A term used by Amershi *et al.*, 2019 [4] to indicate "systems that have features harnessing AI capabilities that are directly exposed to the end user."

Our analysis shows mixed results for diverse representation of age, gender, and race & ethnicity. For age, we found that older adults are particularly well-represented, but this did not apply across all communities of focus (with Autism, Developmental, and Learning communities being notable exceptions). Gender representation skewed towards men/boys being more represented overall but varied widely by community of focus. We also found that well-documented structural marginalization in certain communities are reflected in accessibility datasets. For example, women/girls are underrepresented in Autism datasets, corresponding to existing diagnosis gaps [55, 130]. Marginalization is further embedded on a meta level, such as the case of binary categories for gender classification in the collection and reporting of gender data within datasets. Furthermore, we did not find consistent norms for reporting data, with the lack of standardized documentation, evolving practices, and variability of categories used across age, gender and race & ethnicity.

The contributions of this work are 1) a systematic examination of whether those sourcing data from the disability community are succeeding in representing diverse demographics, via an intersectional analysis along the axes of age, gender, and race & ethnicity as well as a meta-analysis of reporting methods; 2) codes of 190 existing accessibility datasets annotated with demographic metadata[2]; and 3) connections to larger conversations about the implications of representation, data stewardship, and epistemological challenges of data collection. We contend that data representativeness must be analyzed contextually using a critical lens, to accurately assess the potential and implications of greater inclusion of marginalized communities in AI-infused systems.

2 RELATED WORK

Sociocultural diversity has received attention in a wide range of disciplines, such as encouraging gender or ethnic diversity in teams or communities [21, 41, 74], with different concepts of diversity applied in research and applications [159]. More so, AI research has adopted diversity considerations deeply in the ongoing challenge of responsible and ethical AI [24, 42, 113]. Much conversation has been associated with the concepts around *balanced representation* of sub-groups (*e.g.*, equal participation of racial sub-groups within a focal group) [47]. A growing number of studies have explored bias and performance disparities of AI systems concerning representation [38, 108], especially influenced by demographic attributes like age [36, 97, 124], gender [18, 83, 142, 162], race [18, 96], socioeconomic status [34], and disability status [56, 179]. Often such evaluations found the source of concerns as the under-representation of certain demographic groups in the training data underlying predictive and inferential algorithm [108, 162, 179], calling for action to create more balanced datasets across different demographics. In response, we have seen efforts like constructing image datasets balanced in race, gender, and age (FairFace dataset [80]) or text corpora with gender-balanced labels (GAP [175]).

In support of the current discourse around diversity in AI data, researchers have argued that datasets sourced from people with disabilities and older adults can play an important role [75, 79, 118] such as improving speech recognition with stammering data [40] and

[2]Data codes available at https://www.openicpsr.org/openicpsr/project/174761/version/V1/view.

object recognition with photos taken by blind people [75]. Calls for action from this community often center around including disability in AI fairness discussions as it pertains to model performance, data excellence, and privacy [48, 77, 126, 168]. Increasing disability representation, however, is complex; there are myriads of challenges in collecting and sharing datasets from this group [1, 143]. Consent and disclosure can be problematic regarding sensitive disability status. Ethical concerns also arise given that datasets collected to mitigate AI bias for people with disabilities can be used against them by detecting their disabilities, leading to further discrimination risks [118]. There are also existing social biases and stereotypes reflected in data representing disability (*e.g.*, [63, 70]), which may produce AI-infused systems that reinforce greater harms and marginalization of people with disabilities [7]. Efforts aiming to increase inclusion thus need to be carefully considered [163].

To recognize the opportunities and limitations of accessibility datasets in the conversation of diversity in broader AI, we first need to understand the current status of representation in accessibility datasets. Prior work investigating issues associated with diversity in AI datasets has mostly focused on examining differences in model performance across pre-defined demographic attributes to draw implications for diversity [18, 34, 162]. This often leaves inquiries about the benefits and appropriate implementation of diversity in data unanswered [47], except for a few exceptions (as shown in Table 1) that explicitly analyzed datasets or issues related to datasets in terms of demographic representation like gender and other sociocultural attributes (*e.g.*, language) to explore the root causes of bias and misrepresentation. These studies concluded that such AI datasets (often image datasets) are skewed towards certain demographics, uncovering under-representation of older adults [109, 128], darker-skin, and females [109, 185], and lack of geographical diversity [148].

While representation has been discussed broadly across HCI and accessibility [1, 100] or within specific communities [114, 138], we have only seen a few studies analyzing representation and characteristics pertained to AI training datasets in related work [15, 82]. They are yet constrained to very specific tasks and applications. Additionally, discussions of biases against people with disabilities are found to be manifested in complex ways that require intersectional attention [63, 150]. This research complements prior work, by analysing existing accessibility datasets across the communities, to encourage holistic, societal implications for data representativeness including people with disabilities and older adults.

3 METHOD

Our aim is to conduct a broad investigation of what and how demographic attributes are represented in accessibility datasets—not only in terms of disability representation but also age, gender, and race. To this end, we leverage a recently compiled collection of accessibility datasets, sourced from people with disabilities and older adults. We analyze any available information on the data contributors' demographics in associated academic publications, sharing sites, and documentation. Here, we discuss the dataset collections, explain our coding and analysis approach, and reflect on our method and limitations. Reflecting on author positionality, we note that this research was conducted by Asian, Afro-Latina, and white scholars, four of whom identified as women, one identified as non-binary, and two identified as disabled. Research in accessibility ranged from first year grad students to a professor who has been publishing accessibility research for about thirteen years.

3.1 Accessibility Datasets in Our Collection

Recently, Kacorri *et al.* (2020) launched a data surfacing repository, called *IncluSet*, as a result of putting together a collection of datasets sourced from people with disabilities and older adults that were manually located over a multi-year period [76]. An underlying promise of these datasets is their potential for training, testing, or benchmarking machine learning models. The work was later extended to investigate the risks and benefits of collecting, reporting, and sharing accessibility datasets, analyzed in terms of 10 communities of focus, 7 data formats, and 3 data access methods [79]. We leveraged the accessibility datasets (1984–2021, N=190) included in the existing collection of *IncluSet* and their groupings (*i.e.*, communities of focus) as the basis for our investigation. Figure 1a illustrates the distribution of the datasets across the communities of focus. The datasets, including their annotations, are of different data types, as shown in Figure 1b. For example, there are voice recordings of people with speech impairments [25], video recordings of Deaf signers [69], text written by people with dyslexia [134], stroke gestures by people with motor impairments [171], photos of everyday objects taken by blind people [88], eye-tracking data from autistic children [43], and activity data from older adults [91].

Identifying publicly available documentation for these datasets often depended on how they were shared. Out of 190 datasets, about 84 can be downloaded directly and 41 can be accessed upon request—*e.g.*, through a webpage from the dataset creators or an online repository with a summary of the dataset. Summaries vary highly from a few lines to detailed descriptions of the contents of the dataset and how it was collected. Even though none of the datasets had explicitly adopted standardized documentation such as datasheets for datasets [54], some followed a systematic documentation dictated by the platforms where the datasets were stored such as Synapse.org. Associated academic publications were often referred to in the web documentation to link more detailed information about the data collected, though these sources did not always come with consistent information such as the number of data contributors, which could be easily updated on the web documentation. Dataset downloads sometimes came with relevant summary files, including a spreadsheet listing demographic information about people represented in the data. The remaining 65 datasets in the collection did not include any sharing intent with no sources available other than their academic publications. We still include these datasets in our analysis, in accordance with prior work analyzing accessibility datasets [77, 79].

3.2 Manual Coding and Analysis

We conducted an exploratory analysis where our formulation of what-to-code was based on (a) whether demographic information about the data contributors is available, (b) how is it collected and reported, and (c) how are accessibility datasets distributed among demographic groups within communities of focus.

Specifically, beyond the existing codes in Kamikubo *et al.* [79], we extracted information related to demographic attributes following prior surveys on datasets and studies in accessibility and AI that examined diversity and representation (summarized in Section 2.3). A total of three annotators (a PhD student in Information Studies, a Masters' student in HCI, and an undergraduate student in Math) were involved in the process, where at least two reviewed the documentation for each dataset and discussed to correct any disagreement and error. They had different levels of familiarity with accessibility and AI. We extracted the following diversity-related information from the documentation, when available:

Age.—We note how any age-related information is obtained (*e.g.*, self-reported, inferred, or unknown), reported (*e.g.*, individual level, year of birth, age bins, and/or aggregate statistics), and shared (*e.g.*, a separate file). We only calculate aggregated statistics from individual-level data when reporting findings and plotting distributions.

Gender.—We note the labels used (*e.g.*, sex, gender), if any; the categories used; the number of data contributors that belong to the categories used; and how metadata was obtained (*e.g.*, self-reported or inferred) and shared (*e.g.*, spreadsheet or publication). In response to concerns raised by trans and information science scholars that the sex/gender distinction can invalidate trans and intersex identities while veiling the socially constructed nature of sex categories, for this paper we use the term "gender" to refer to discussions of characteristics of data contributors (that may be labeled by researchers as either gender or sex) [46, 142, 146].

Race and ethnicity.—Race is a multidimensional and complex concept, not a singular, biological construct with distinct limits into which people can be classified. Alone, race and ethnicity, do not reveal much about an individual's experiences. As race and ethnicity can be viewed through multiple socially constructed lenses [17], we started with broad coding techniques to identify any information that pertains to these demographic attributes, including potential ethnic and cultural descriptors like geography and language. Manly [104] suggests that these attributes are proxies for or interrelated with unexamined variables, such as education and socioeconomic status. To better our understanding of race/ethnicity, it is central to deconstruct and examine the confounding influences of ethno-racial factors. We note any categories used to refer to data contributors' racial groups, such as those defined in the census [19] and group ethnic and cultural metadata like nationality, geography, and language under other sociocultural information. Based on the metadata identified, we update the annotation scheme by specifically going over how this information is obtained and shared. Metadata related to education included information in terms of how it is obtained, reported, and shared; language included information on dialect and skills earned which may interact with education; geography included information on data contributors' birthplaces and the recruitment location; and other information such as nationality or socioeconomic status when available.

3.3 Reflections on Limitations

Annotation consistency.—Annotation tasks are notably difficult, especially if they involve manual inspection of large data requiring particular skills and knowledge. Given that we inspected both dataset documentations and scholarly articles from various publication venues across many research disciplines and sub-disciplines (*e.g.*, Linguistics, Acoustics, Physiology, Computer Vision, HCI, Accessibility), it was unavoidable to go through a messy process to correct errors and disagreement in our codes. The annotators' varying levels of familiarity with accessibility and AI were also sources of difficulty. This is not a surprise. Even similar annotation tasks that were more limited in scope (*i.e.* within the field of accessibility), were characterized as "challenging and effortful" [100]. To address the challenges, as the coding process initially started with two annotators (PhD and undergraduate level), we invited a third member (Master's level) to have a detailed pass. The PhD student took a final pass to ensure that the annotations were agreed upon at least by two annotators.

We also experienced difficulty in programmatically extracting demographic-related metadata. This often created disparities among the annotators in identifying the relevant information from the documentation. We did not find a consistent, standardized method. For example, some methods we used included manually reviewing web documentation that provided summary statistics in writing [135] or table [2] formats; downloading files containing participants' demographic data (*e.g.*, age, gender) together with collected data points [164] or a separate csv file on participant demographics [6]; or extracting metadata from filenames [65]. Without standardized documentation and evolving practices, whether datasets contained demographic-related metadata was often unknown prior to downloads. In addition, without proper explanation of the labels used for demographic categories, such as in one dataset [6] that provided a supplementary spreadsheet with a label '1' under the Race column for each participant, we could not find the meaning of this information.

Lack of documentation.—As discussed in the Results, information on age, gender and race/ethnicity was in many cases sparse. When available, it was often unclear how the demographic-related metadata was obtained. Thus, we could not verify the source of classifications (such as for gender). Few datasets explicitly documented that the reported information was *e.g.*, "according to self-reports" [191]. Even fewer made inferences on these demographics *e.g.*, "using proprietary classifiers" [177] or "based on visual inspection" [151]; typically these inferences were employed on data collected over the web. Specifically, we observed that three datasets indicate estimations on data contributors' age; all three are solicited from user interactions with a web search engine with users' age reported being "over the age of 40 years inferred from their date of birth as reported at registration to Bing" [189] or "inferred using proprietary Bing classifiers"[177, 178].

White *et al.* [177, 178] employed a similar approach for gender. Whereas Shi *et al.* [151, 152] determine the gender of individuals by visually inspecting sign language videos from YouTube and the signers' social media; they used the code "Other" for videos including people whose gender was deemed unknown or where there were multiple signers. While we have included the codes for these datasets in our collection as a reference for future researchers, we don't include them in our analysis of 'reported' demographics; inferences can be inaccurate, perpetuate bias, and perpetuate exclusion (*e.g.* via binary classification of nonbinary individuals).

None of the datasets in the collection inferred or estimated demographics that pertain to race/ethnicity or other metadata related to nationality, geography, language, and education. Yet, this part of our analysis is the weakest one as it solely relies on a small number of datasets where the race/ethnicity information was specifically 'reported'; the majority (8) came from US institutions and one from UK even though the institutions of data stewards in the collection spanned across 42 countries from Asia, Africa, North America, South America, Europe, and Australia. Thus, our analysis of this demographic is inherently limited. Only limited reporting of race/ethnicity may be due to a number of factors, such as differences in census reporting among Western and non-Western countries, a prevailing consensus that racial designations do not identify genetically distinct populations, and the likelihood of misuse (e.g., privacy risks for disabled people) [84, 122, 147]. Cooper *et al.* suggest that "the correlation between the use of unsupported genetic inferences and the social standing of a group is glaring evidence of bias and demonstrates how race is used both to categorize and to rank order subpopulations." [31]. However, since federal and state legislation in the US have established evident discriminatory practices against African Americans, Hispanics, Asians, and other groups, racial categorization can be utilized to reflect intersectional gaps that are a product of racial stratification practices. Thus, considering the sociocultural and political contexts of different regions to further understand the decision to utilize racial categories is critical. We did not see within the scope of this paper a systematic way to report the somewhat sparse metadata across codes related to data contributors' nationality, geography, language, and education and tie them to sociocultural and political contexts of different regions. Nonetheless, we include these codes in our annotations for future reference.

Non exhaustive collection.—One of the main limitations of this work remains the fact that the list of datasets in the collection is not exhaustive. While somewhat systematic, the identification of these samples is itself noisy and prone to cascading biased decisions from the researchers collecting them and those that opt/know to include their datasets in the IncluSet repository. The lack of inclusion criteria related to *when* these datasets were introduced or *whether* they are currently in use and to *what* extent, could lead to systematic misalignment between current efforts and past trends. This is exacerbated by the fact that many datasets that are actually employed currently in commercial AI-infused products are not accessible for this type of analysis; representation of different demographic groups could be perhaps deduced via biased performance results (*e.g.*, [18]) but that is beyond the scope of this work. Thus, any insights from our analysis may not be generalizable beyond the research community.

4 RESULTS

Of 190 datasets whose publication and documentation we reviewed, the most commonly found types of demographic-related metadata are age (46.8%) and gender (54.2%), followed by few datasets reporting race (4.7%) and education (12.1%). We find that 71 datasets (37.4%) did not include any information related to the aforementioned types of metadata. These numbers differ from publications that also focus on health, wellness, accessibility, and aging, where few share data; when looking at 792 HCI studies, Abbott *et al.* (2019) found a distribution of 69.7%, 67.3% and 6.6% on age, gender, and ethnicity, respectively [1]. This difference could be due to tensions inherent in collecting “sensitive attribute data” [1, 11, 16] and concerns related to participant consent and re-identification risks [1]. A similar trend is seen among available metadata with respect to how others can access the datasets. Among those that are not publicly shared, 69.2% reported at least one of the demographics, compared to 57.1% for publicly shared and 53.7% for shared upon request.

In this section, we present our findings surrounding such “sensitive attribute data” in accessibility datasets across communities of focus (Figure 2). To better understand the current status in terms of reporting and including different demographic groups and variables, we focus on the following demographics: age, gender, and race and ethnicity. In our analysis, we compare with existing categories used to represent demographic variables in social data collection (*e.g.*, racial categories in census [174]), and investigate representativeness within accessibility datasets.

4.1 Age

A total of 6050 people within the communities of focus contributed data to the 89 datasets whose information on age was included. Their weighted average age was 43.6 (std=26.3). For the remaining of the report, statistics are reported at the dataset level (*i.e.* sampling distribution of the mean) even though the sample size across datasets varies highly from 1 to 990 people (mean=66.8, std=144.5). Data on age from control groups are not included in the analysis.

4.1.1 What Is Reported.—Datasets mostly reported such information in aggregate though some (36.0%) reported age at an individual level. Aggregate information includes minimum age (1.1%), range (15.7%), median (1.1%), average (20.2%), or a combination (25.8%). Typically, age was reported separately for target (*i.e.*, disability) and control groups (*e.g.*, [45]), contributors’ gender (*e.g.*, [170]), and dataset purpose (*e.g.*, training versus validation [86]). Few report on all groups together (*e.g.*, [22]). Data anonymization is a core component of data management to minimize risk of disclosure while preserving its utility for analysis [81]. However, we find that a majority of the datasets did not incorporate these strategies. For example, bucketing by age groups (*e.g.*, 18–30, 31–45, 46–60 years [107]) was only found in 7 datasets (7.9%).

Only 5 datasets reported median and 3 datasets reported both mean and median. More than half (58.4%) indicate standard deviation, including those reporting age at the individual level for which it can be calculated. All three, mean, standard deviation, and range, can be found for less than half (42.7%) of the datasets (*e.g.*, *“The mean age of the subjects was 54.9* ± *13.4 (SD) yr (range 36–70 yr)”* [64]). Meanwhile, some documentation noted only the minimum (*e.g.*, *“participants aged 50 or older”* [180]) or the age requirement for participation (*e.g.*, *“18 or older”* [13]).

4.1.2 Why Is It Reported.—Most often datasets did not specify why the ages were obtained and reported. It could be an effect of perceived norms and standards for questionnaires within the research community, which often include age questions [68, 161]. Age is an established variable that helps understand the general characteristics of participants. Its distribution may reflect the quality of data collection and analysis [5]; not accounting for age can threaten the generalizability of the work especially when there is a treatment effect heterogeneity in age or other factors that may covary with age (*e.g.*, [121]). Some datasets mention efforts to match age between target and control groups (*e.g.*, [26, 160]) or note age matching as not feasible (*e.g.*, [111]). Others mention age as a confounding variable *e.g.*, for early detection of Parkinson's disease based on touchscreen typing patterns [72]. Some datasets mentioned the goal of including data from diverse age groups to assess age-related decline of cognitive or mobility performance [91, 116]. For example, in a dataset acquiring age-related pen-based performance [116], participants were grouped based on cognition changes (*'young'* for 18–55, *'pre-old'* for 56–75, and *'old'* for 75+). Grouping varies across communities; in an attempt to build a diverse sign language corpus, researchers binned groups as 18–35 years, 36–50 years, 51–64 years, and 65+, rationalizing their decision based on language transmission variability within the Deaf community [141].

4.1.3 Representation Across Communities of Focus.—

Figure 3 illustrates with violin plots the sampling distribution of mean age in datasets across communities, where the white dot represents the median, the thick gray bar in the center indicates the interquartile range, and the thin gray line shows the rest of the distribution, except for points that are determined to be "outliers." Kernel density estimations on each side of the gray lines show the distribution shape. Wider sections indicate a higher probability that datasets will have a mean age of the given value; the skinnier sections indicate a lower probability. We note that datasets vary in their sample size, which is not accounted for by this visualization.

We find that mean age in datasets differs across communities, with some communities particularly inclining towards samples with a certain target age (*e.g.*, children, older adults). To better understand the age representation exhibited in accessibility datasets, the remainder of the section follows age groups discussed or referred to in prior literature in terms of technology (*e.g.*, 'older adults' as 65+, 'oldest-old adults' as 85+) [128], disability-related policies (*e.g.*, 'children' between 3 to 21 covered in IDEA [94]), and the communities of focus (*e.g.*, 'toddlers' of 18 to 36 months in developmental assessment [30]). Of course, variations exist across studies [154] as there is no rigid definition for these groupings.

Older adults.: Many accessibility datasets represent older adults. Among the datasets that contained some form of age-related information, 48.3% included at least one older adult (65+), and 6.7% at least one oldest-old adult (85+). The highest proportion of older adults was in the Cognitive and Health groups, reporting at least one older adult in 83.8% and 73.3% of their datasets, respectively. This may not be surprising, as these groups focus on cognitive and physical decline that can relate to age—*e.g.*, the risk of onset of dementia (*e.g.*, Alzheimer's disease) increases with older age [131]. Specifically, the Cognitive group had datasets with the highest mean of mean age (mean=61.7, std=12.4) which were often cross-listed with the Mobility and Speech groups including speech or motion data of patients with Parkinson's disease (*e.g.*, [71, 140]). The oldest participant, aged 89, was reported in the Cognitive and Health groups in the image dataset capturing daily activities of those with episodic memory impairment [89]. Communities that lack older adult representation are Autism, Developmental, and Learning, reflecting a broader gap in research pertaining to these groups [66, 73, 130, 139]. This can be due to many factors; for example, many autistic older adults experienced a severely delayed diagnosis [102]. Many adults with learning disabilities live in institutions such as nursing and residential homes, in which they arrive "before their 65th birthday" with "few opportunities to get out" [165].

Children and youth.: Children and youth are also represented in accessibility datasets; about a quarter (24.7%) of the datasets whose information on age was included contained data sourced by at least one person younger than 18 years old. It increases to 33.7% when including those 21 or younger, as the age criteria for study participation is often noted as 18 or older [13, 45]. Perhaps this reflects some of the ethical challenges in collecting data from children [32] as the process for obtaining consent, assent, or parental permission is more complex for those under the legal age [112]. While overall there are few datasets sourced from youth, they tend to concentrate in the Developmental (85.7% of datasets in this group include at least one person <18) and Learning (100.0%) groups. Datasets in the Learning group often focus on dyslexia (*e.g.*, [53, 115]), where diagnosis is critical at early ages. Data from toddlers (18 to 36 months old) are typically seen in the Development group for the purpose of developmental assessment (*e.g.*, [30]). They mostly involve speech data, sourced by stuttering children [58, 182] or late talkers [120]. The youngest reported age across all the accessibility datasets was 16 months, in a dataset sourced from autistic children [181], though not many (33.3%) datasets reporting age in the Autism group included those under the age of 18. The groups that lack data from children and youth are Vision, Hearing, and Mobility. We suspect that this is reflective of the most common purpose for collecting data such as image and video from this age group, which is to better assess and diagnose; disabilities related to one's vision, hearing, and mobility have long established methods and instruments that might not require such datasets.

Younger and middle-aged adults.: When looking at younger adults (over 18), we find that surprisingly, many (9) datasets with mean age in the Autism group tend to include people between the age of 18 and 44, with an overall mean of mean age 24.0 (std=13.8). This is in striking contrast with the broader research on autism, where the majority (94%) tends to focus on infants, toddlers, children, and adolescents [73] due to a focus on early diagnosis and intervention [117, 127]. Datasets including younger adults in this group were often collected in the context of assistive technologies (*e.g.*, evaluating text readability and comprehensibility via gaze fixations [45, 183, 184].)) Looking further at datasets skewed towards younger and middle-aged adults, the age range of Hearing and Vision groups was limited, even though visual and hearing impairments could be associated with older age [14, 95]. The datasets in the Hearing and Vision groups that reported age have an overall mean of mean age 28.3 (std=4.2) and 48.7 (std=3.6), respectively. This can be partially explained by how these datasets were collected. For example, the majority (66.7%) of datasets in the Vision group did not include any age information; they were collected from thousands of users via real-world applications (*e.g.*, [57, 78]), where user demographics may not be available or omitted due to privacy concerns. Similarly, in the Hearing group the majority of datasets do not include age information; they tend to collect sign language from online sources (*e.g.*, [92, 151]).

Diverse ages.: We observe that the Language group has the largest age variability. Among others, they include data sourced from children with epilepsy (*e.g.*, [160]), adolescents with language impairment (*e.g.*, [176]), and older adults with aphasia (*e.g.*, [3, 35]). Often datasets in this group come from clinical settings such as the FluencyBank found in TalkBank [101], a shared database established in 2002 for studying human communication. Perhaps this collaborative effort among a wide range of disciplines could explain the variability of datasets spanning across different communities over the years. Datasets in Speech also capture different age groups. Some can be found in TalkBank, including spoken phrases of older adults with Alzeimer's disease [105] as well as children [182] and adults [187] who stutter.

4.2 Gender

A total of 5598 people within the communities of focus contributed data to the 103 datasets whose information on gender was included. Again, we include information at a dataset level even though the sample size across datasets varies highly from 1 to 818 (mean=59.6, std=106.6). Data on gender for the control groups are not included in the analysis.

4.2.1 What Is Reported.—Gender metadata was commonly reported with the number of data contributors in the form of writing (*e.g.*, *"10 blind participants (5 female) ranging in age from 18 to 63 years old"* [9]) or table (*e.g.*, a M/F column [6]). Of datasets reporting such metadata, we observed that a binary classification was used (*female/male, women/men, girls/boys)*, with only one dataset in our collection reporting data on the *"other"* category [49]. However, it is difficult to draw conclusions from this alone, as few datasets reported their method of gendering contributors. Without this, we cannot distinguish between self-identification (*e.g.*, as part of a demographics questionnaire), or an external inference influenced by implicit assumptions (*e.g.*, by the study designers or validators). Furthermore, if participants were asked to self-identify, they may have been limited to choosing from binary options.

4.2.2 Why Is It Reported.—Similar to age being asked in standard demographic questions [68], datasets often included gender information as part of the data distribution, without specifically describing the goal of collecting such information.

Nonetheless, we can attempt to extrapolate the reasoning for some datasets, especially when they contain particular data formats. The highest presence of gender information was in datasets that collected *audio* (66%) compared to *video* (27%) or *image* (32%). Perhaps, this is reflective of an assumption of the influence of gender among those working with speech data. Datasets that capture *motion e.g.*, gait of Parkinson's disease patients [170], also attempt (about 50% of them) to account for physical measurement differences represented in data by using gender as a proxy.

In order to keep the study design as "*unbiased*" as possible, some datasets reported that gender (and/or age) was "*balanced*" in the test group (*e.g.*, *"roughly balanced for gender of the 249 participants, 52% (n= 129) were women"* [141]), but efforts to balance distribution between target and control groups were much more common (*e.g.*, [170], [125]).

4.2.3 Representation Across Communities of Focus.—

Gender demographics vary across the world, with most countries having a *female*[3] share of the population between 49% and 51% [137]. However, overall, accessibility datasets that include gender information tend to be imbalanced with men and boys (60.1%) who are more represented on average[4] than women and girls (39.9%). This is also evident in Figure 4a, which illustrates with violin plots the sampling distribution of gender representation in datasets across communities of focus, where the vertical dash lines indicate the quartiles and each side of the distribution shows kernel density estimations for 'women/girls' and 'men/boys'. This illustration also highlights how the gap is more prominent in some communities than others.

Specifically, we see a clear imbalance in the representation of data contributors in the Autism and Developmental groups; on average, 33.1% (std=8.1) and 27.9% (std=9.8) are women and girls, respectively. Such highly skewed representation has been actively discussed in the evaluation and diagnosis of autistic children, given that boys constituted 81% of the sample of children [55]. One widely cited *male-to-female* diagnosis ratio is approximately 4:1 [51]. However, when the ASD participants are controlled for cognitive impairments, this number changes [85, 98, 103, 106, 138]. About 50–55% of autistic children are estimated to be intellectually disabled (ID) [98]. Among ID autistic children, the *male-to-female* ratio is significantly smaller, at 2:1 [67]. In autistic children labeled as "high functioning", the existing literature points to a higher *male-to-female* ratio, about 6:1. Researchers have theorized an explanation for this relationship could be the tendency of (so-called) "high-functioning" autistic *females* to "mask" or "camouflage" core autistic traits [90, 133]. A growing body of evidence suggests that current diagnostic criteria for ASD may fail to account for these phenomena and the subtleties in behavior, leading to misdiagnosis and late-diagnosis for minority gender groups (*e.g.*, women, girls, non-binary) [87].

While many communities of focus portray gender disparity in their represented samples, it is not seen in the Vision group, with the average of 50.2% (std=3.2) consisting of women per dataset. According to 2018 U.S. disability statistics [186], 45.3% of visually disabled people were *male*, and 54.7 % were *female*. The slight skew towards women has been identified by researchers in this community as possibly attributable to differences in life expectancy by gender in addition to increased risk of visual impairments with age (*e.g.*, macular degeneration) [59], which women are noted to be at higher risk of than men [156].

4.3 Race & Ethnicity

Race is a complex and sensitive demographic variable [52, 145]. Only 9 (5%) accessibility datasets reported metadata on contributors associated with racial or ethnic groups, typically captured by demographic surveys (*e.g.*, [19]). Modern racial classification systems construct race using both observable physical features (*e.g.*, skin color) and nonobservable characteristics such as culture and language [27]. Thus, 'other' related demographic information we found could perhaps be utilized to draw some connections and inferences about race, including the place of birth [23], native language [72], or dialect [188]. However,

[3]When referring to data sourced from external collections, we follow the terminology used in their reports.

[4]With both gender-related and sex-related categories used in our collection of datasets, we report data for 'women/girls' or 'men/boys' combined with data for *e.g.*, 'female' or 'male'.

in past studies they have led to issues of forced classification and error [11, 123]. Therefore, in this section we don't make that connection. We report only on datasets with explicit racial and ethnic information.

4.3.1 What Is Reported.

4.3.1 What Is Reported.—The categories we found delineating racial composition were mostly *'White'* and *'Black'* [144], with variations of reporting them as *'White-Caucasian'* or *'Caucasian'* and *'African-American'* [160, 182, 191]. For other racial groups, data were ambiguously grouped together (*e.g.*, *"62% Caucasian, 30% African-American and 10% other"* [160]) or can be extrapolated by subtracting what was reported as the proportion of the *'white'* category only [190]. The use of these terms also highlight the limitations of the taxonomical racial categories; 'Caucasian', for example, is rather discussed as outdated and disproved [119].

Similar to age and gender, race was reported separately for target and control groups (*e.g.*, [190]). Notably, one speech dataset sourced from stuttering children aimed at a race-matched (as well as age- and gender-matched) cohort of children [132]—here, both stuttering and non-stuttering groups had 2 African American children and 1 child of mixed racial ancestry. This was also the only dataset in the collection reporting about mixed race, although we saw an attempt to collect data on race, including 'Mixed', from a demographic questionnaire in a study on Parkinson's disease [13].

4.3.2 Why Is It Reported.—Looking at datasets whose data on race was collected and/or reported, they are often related to medical research associated with studies on specific disorders. Specifically, they include speech samples collected from people with aphasia [144], Parkinson's disease[190], Alzheimer's disease [6], and epilepsy [160] to study early detection of impairments underlying cognitive disturbance. In medical research domains, there are controversies around collecting data on race, raising both benefits and risks given disparities in health outcomes established for racial minorities [50, 62]. Concerns also lie in the taxonomy of the categories used, which have brought efforts to standardize and improve methods of obtaining and reporting data on race [8, 50]. Recent guidelines [50] suggest including an explanation of who identified participant race & ethnicity and reasons for collecting the data. We did not find disclosure of the source of the classifications among the datasets included (*e.g.*, self-report, observation), nor a justification of why it was collected.

4.3.3 Representation Across Communities of Focus.—It was hard to distinguish the data between race and ethnicity or other sociocultural information, especially when the data spans multiple concepts and forms of classification (*e.g.*, *"129 of Caucasian, 14 of African American, 2 of Hispanic, and 2 of Asian origin"* [182]). For example, in US, guidelines that inform data collection for census note that the concept of race is separate from the concept of Hispanic origin [173].

For the few datasets that reported data contributors' race and ethnicity, the norms of how to report were highly inconsistent. Thus, with high variability and a small sample, we could not leverage standardized methods to analyze racial group composition among the communities of focus. The categories we saw (often in Cognitive and Language) were associated with *'white'* or *'non-white'*, portraying one group as primary over another. Mixed race was rarely indicated, which is problematic given changes in racial categories (*e.g.*, in the US census) reflecting racial mixture [20].

5 DISCUSSION

Our overarching goal lies in understanding the current state of representativeness of marginalized groups in AI datasets (along the axes of age, gender, and race & ethnicity) with a specific focus on disabled data contributors. This is relevant to the greater discourse around AI, ethics, and fairness, as marginalized communities tend to be under-represented in data [47], perpetuating cycles of exclusion as technology advances even for technologies that meant to promote inclusion such as assistive technology. We contribute to this important ongoing discussion through our analysis of 190 accessibility datasets. Specifically, we examine representation gaps and trends that can potentially lead down the road to further harm for the people who stand to be adversely affected by emerging, potentially ubiquitous technology. In this section, we recap and discuss the challenges and opportunities for representation while considering directions the accessibility field could take to carefully include marginalized communities in AI-infused systems.

5.1 Addressing Challenges and Seizing Opportunities for Representation

Our analysis revealed unique challenges in ensuring representation of intersecting demographics in accessibility datasets. Some representation gaps are attributable to societal and cultural norms and biases that operate intersectionally. For example, communities lacking older adult representation are Autism, Developmental, and Learning. This reflects not only a broader research gap on these groups [66, 73, 130, 139] but also discrimination at the intersection of disability and age; *e.g.*, many autistic older adults live without an accurate diagnosis [102]. Similarly, looking at the intersection of disability and gender, we observe a gap for Autism, Developmental, and Learning groups, where men and boys were often over-represented. These cases can have pernicious implications characterized not only by the communities of focus but also long established research frameworks that propagate existing societal marginalization, highlighting the importance of making gender-specific changes (*e.g.*, diagnostic criteria for autism [37, 87]).

In annotating accessibility datasets, we also surfaced how socially constructed identity categories such as race and gender are reproduced. Similar to Scheuerman's meta-analysis of gender in face datasets [142], by analyzing information such as reasons for reporting/data collection and labels used for metadata categories, we contribute a sociological meta-examination through which the research and data collection process itself can be analyzed for bias. For example, we found that the notion of a gender or sex binary was not explicitly challenged in our collection; only one dataset reported data on the *"other"* gender category. This may have downstream effects in shaping machine learning model design and subsequent problems/contexts—for example, in binary gender classification, which may harm nonbinary communities through technology-enabled misgendering [61].

We also found that there is very little reporting of how identity labels were associated with data contributors, whether through selfi-dentification or external assumption (*e.g.*, via preformed binary categories). We recommend greater transparency in disclosing these aspects of the data collection process, and for gender in particular, to include nonbinary, self-describe, and prefer not to disclose options, as recommended in the related literature [158].

At the same time, we acknowledge the implementation challenges that may need to be addressed to support transparency—*e.g.*, how to produce a set of questions which do not elicit information leading to unintentional misuse or unwanted societal biases for data contributors. We emphasize that careful reflection on this process is needed on the part of researchers who are collecting and reporting contributor data, including implications of use (*e.g.* surveillance) and any potential harms enacted by power structures through the systems we build. Aligning with recent research [110], we recommend an examination and contextualization of data representativeness grounded in political, economic, and socio-cultural lenses, integrating insights from scholars in fields such as critical disability studies [28], trans/gender studies [157], and histories of social movements [136] into an analysis of power relations. As an example, one could draw from recent work by disability studies scholars examining the context the data is collected in (*i.e.*, for AI systems vs for visibility and activism) and how representation impacts are also *context-dependent* [93].

5.2 Developing Participatory Approaches to Data Stewardship

This challenge of partitioning the pool of accessibility datasets into sub-communities was very real in our analysis, as the groupings that we opted for may not necessarily reflect the identities of individual data contributors. Recent work exploring challenges for collecting disability data suggests the voices of contributors to be reflected and provides best practices to ask about disability status [10]. Perhaps, to mitigate harms experienced by those from marginalized communities who are misclassified, we can extend this approach to other categories such as race and gender. Specifically, we urge researchers to come up with approaches for more meaningful engagement of data contributors in the data stewarding process. Echoing Shneiderman's motto [153], we recommend "*researchers in the loop, disabled contributors in the group*".

One way we could go about this is to employ participatory approaches to the data collection lifecycle in which users have the opportunity to enact their values in how their data is collected, maintained, shared, and interpreted in and out [33, 99]. Of course, this would require careful consideration of the many moving pieces in the Fairness, Accountability, Transparency, and Ethics (FATE) landscape both in terms of parties involved as well as exchange and access mechanisms; Bragg *et al.* [15] provide a wonderful starting point for this discussion in the context of the Deaf community. For example, to avoid inadvertently extractive approaches, and aligning with recent literature, we recommend meaningfully compensating participants for their work as data contributors [155]. In this vein, we also recommend developing long term relationships with data contributors and their communities (where possible) to facilitate sustainable and mutually beneficial collaboration, especially when designing and evaluating AI-infused systems that use contributor data [155, 163]. Disability community-led initiatives can help concentrate research efforts on those most likely to have a positive impact; the idea generation phase may be particularly fruitful when rooted in first person lived experience (*e.g.* as provided in [129]).

5.3 Addressing Epistemological Implications in Future Work

We encountered epistemological limitations at various stages in the annotation and analysis process. One such limitation is the extent to which strong claims can be made about overall representativeness, due to the lack of reporting and global statistics for disability, age, gender, and race. In addition, our findings are intrinsically linked to existing sociocultural contexts and hierarchies. Our analysis of accessibility datasets showcases these epistemological limitations. By acknowledging these limitations, we hope to spark conversations on the inclusion of marginalized communities in AI-infused systems and its myriad challenges. In future efforts, we recommend the following for broader research implications:

Exploration of disabled people's concerns around representation.—Increasing representativeness may not always be beneficial; it may perpetuate injustice as extensions of existing systems of oppression and power. As explored in the previous section, it is vital to include first person disabled perspectives on representativeness and inclusion, as well as data collection and sharing practices. Future work remains in exploring contributor concerns such as privacy [60, 77] and surveillance [7], especially for multiple marginalized contributors.

Analyzing other sociocultural factors.—A more in-depth analysis of the sociocultural contexts in which datasets were produced, not just what was reported, could lead to interesting insights. A quick inspection of our datasets revealed that when data involves children, specifically in studies of developmental disability, we sometimes find family information, such as socioeconomic status [132] or parental education [58, 160]. Future work could explore representation along axes of level of education, language, nationality, and socioeconomic status of the data contributors, as well as intersections between them. It would also be interesting to explore the influence of dataset origin (*i.e.* from the HCI vs medical research community) on demographic representation as they may opt for different models of disability.

Accounting for dataset impact.—Our analysis of the implications of representation is complicated by the fact that datasets vary in research impact. Potential indicators of impact include the number of citations, the models they are used to train or benchmark, the venues in which they are published, and whether they originate from academia or industry. Future work remains in investigating and defining impact indicators and metrics, and weaving those insights into discussions of representativeness.

Beyond accessibility datasets.—While any insights from our analysis may not be generalizable beyond the research community, our findings present an opportunity for broader AI communities to strive towards more representativeness—along disability and other dimensions—by including accessibility datasets in their training data. For example, AI datasets have been critiqued for being heavily skewed towards younger adults, and under-representing older adults [128]. In contrast, accessibility datasets yield a wide variability of age groups. In future research, we strive to connect our discussions of representation gaps with larger trends for broader AI datasets and investigate whether accessibility can be used as a lens to diversify representation for the broader AI community.

6 CONCLUSION

We conducted a detailed analysis of data representativeness among 190 accessibility datasets, with an emphasis on the intersections of disability with age, gender, and race & ethnicity. While we found diverse representation of age in accessibility datasets, we identified gaps in gender and race & ethnicity representation among these datasets. Our findings illustrate the implications of historical and social contexts. Although we acknowledge there are limitations when collecting these demographic variables, going forward, we propose a participatory approach when collaborating with disabled contributors and encourage transparency regarding data collection purpose and maintenance throughout the process. We hope our effort elucidates the current challenges in representation among the accessibility community while expanding the space of possibility for greater inclusion of marginalized communities in AI-infused systems more broadly. Finally, we hope that our efforts provoke conversations on data representativeness through a critical and epistemological lens.

ACKNOWLEDGMENTS

We thank Hal Daumé III for providing valuable feedback on our preliminary work. We also thank our anonymous reviewers for further strengthening this paper. This work is supported by National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR), ACL, HHS (#90REGE0008).

REFERENCES

1. Abbott Jacob, MacLeod Haley, Nurain Novia, Ekobe Gustave, and Patil Sameer. 2019. Local Standards for Anonymization Practices in Health, Wellness, Accessibility, and Aging Research at CHI. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery (ACM), 1–14. 10.1145/3290605.3300692
2. Aggarwal Gaurav and Singh Latika. 2018. Evaluation of Supervised Learning Algorithms Based on Speech Features as Predictors to the Diagnosis of Mild to Moderate Intellectual Disability. 3D Research 9, 4 (2018), 55. 10.1007/s13319-018-0207-6
3. Allen Meghan, Joanna McGrenere, and Barbara Purves. 2007. The Design and Field Evaluation of PhotoTalk: A Digital Image Communication Application for People with Aphasia. In Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '07). Association for Computing Machinery (ACM), 187–194. 10.1145/1296843.1296876
4. Amershi Saleema, Weld Dan, Vorvoreanu Mihaela, Fourney Adam, Nushi Besmira, Collisson Penny, Suh Jina, Iqbal Shamsi, Bennett Paul N., Inkpen Kori, Teevan Jaime, Kikin-Gil Ruth, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. Association for Computing Machinery, New York, NY, USA, 1–13. 10.1145/3290605.3300233
5. Andrews Frank M and Herzog A Regula. 1986. The quality of survey data as related to age of respondent. J. Amer. Statist. Assoc. 81, 394 (1986), 403–410.
6. Becker James T, Boiler François, Lopez Oscar L, Saxton Judith, and McGonigle Karen L. 1994. The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. Archives of neurology 51, 6 (1994), 585–594. 10.1001/archneur.1994.00540180063015 [PubMed: 8198470]
7. Bennett Cynthia L. and Keyes Os. 2020. What is the Point of Fairness? Disability, AI and the Complexity of Justice. 125, Article 5 (March 2020), 1 pages. 10.1145/3386296.3386301
8. Bhalla Rohit, Yongue Brandon G, and Currie Brian P. 2012. Standardizing race, ethnicity, and preferred language data collection in hospital information systems: results and implications for healthcare delivery and policy. Journal for Healthcare Quality 34, 2 (2012), 44–52.
9. Bigham Jeffrey P., Cavender Anna C., Brudvik Jeremy T., Wobbrock Jacob O., and Ladner Richard E. 2007. WebinSitu: A Comparative Analysis of Blind and Sighted Browsing

Behavior. In Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (Assets '07). Association for Computing Machinery (ACM), 51–58. 10.1145/1296843.1296854

10. Blaser Brianna and Ladner Richard E 2020. Why is Data on Disability so Hard to Collect and Understand?. In 2020 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT), Vol. 1. IEEE, 1–8. https://www.washington.edu/doit/sites/default/files/atoms/files/RESPECT_2020_DisabilityData.pdf
11. Bogen Miranda, Rieke Aaron, and Ahmed Shazeda. 2020. Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (Barcelona, Spain) (FAT* '20). Association for Computing Machinery, New York, NY, USA, 492–500. 10.1145/3351095.3372877
12. Bolukbasi Tolga, Chang Kai-Wei, Zou James Y, Saligrama Venkatesh, and Kalai Adam T. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Advances in neural information processing systems 29 (2016), 4349–4357.
13. Bot Brian M., Suver Christine, Neto Elias Chaibub, Kellen Michael, Klein Arno, Bare Christopher, Doerr Megan, Pratap Abhishek, Wilbanks John, Dorsey E. Ray, Friend Stephen H., and Trister Andrew D. 2016. The mPower study, Parkinson disease mobile data collected using ResearchKit. Scientific Data 3 (March 2016), 160011. 10.1038/sdata.2016.11 [PubMed: 26938265]
14. Bowl Michael R and Dawson Sally J. 2019. Age-related hearing loss. Cold Spring Harbor perspectives in medicine 9, 8 (2019), a033217. [PubMed: 30291149]
15. Bragg Danielle, Caselli Naomi, Hochgesang Julie A., Huenerfauth Matt, Katz-Hernandez Leah, Koller Oscar, Kushalnagar Raja, Vogler Christian, and Ladner Richard E.. 2021. The FATE Landscape of Sign Language AI Datasets: An Interdisciplinary Perspective. 14, 2, Article 7 (July 2021), 45 pages. 10.1145/3436996
16. Braun Lundy, Fausto-Sterling Anne, Fullwiley Duana, Hammonds Evelynn M, Nelson Alondra, Quivers William, Reverby Susan M, and Shields Alexandra E. 2007. Racial categories in medical practice: how useful are they? PLoS medicine 4, 9 (2007), e271. [PubMed: 17896853]
17. The Editors of Encyclopaedia Britannica. 2021. critical race theory. https://www.britannica.com/topic/critical-race-theory.
18. Buolamwini Joy and Gebru Timnit. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research, Vol. 81), Friedler Sorelle A. and Wilson Christo (Eds.). PMLR, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html
19. United States Census Bureau. 2021. QuickFacts United States. https://www.census.gov/quickfacts/fact/table/US/PST045221. Accessed: 2022-01-03.
20. Burhansstipanov Linda and Satter Delight E. 2000. Office of Management and Budget racial categories and implications for American Indians and Alaska Natives. American Journal of Public Health 90, 11 (2000), 1720. [PubMed: 11076238]
21. Campbell Lesley G, Mehtani Siya, Dozier Mary E, and Rinehart Janice. 2013. Gender-heterogeneous working groups produce higher quality science. PloS one 8, 10 (2013), e79147. [PubMed: 24205372]
22. Carette Romuald, Elbattah Mahmoud, Cilia Federica, Dequen Gilles, Guerin Jean-Luc, and Bosche Jérôme. 2019. Learning to Predict Autism Spectrum Disorder based on the Visual Patterns of Eye-tracking Scanpaths. In Proceedings of the 12th International Conference on Health Informatics. 103–112. 10.5220/0007402601030112
23. Caselli Naomi K, Sehyr Zed Sevcikova, Cohen-Goldberg Ariel M, and Emmorey Karen. 2017. ASL-LEX: A lexical database of American Sign Language. Behavior research methods 49, 2 (2017), 784–801. [PubMed: 27193158]
24. Celis L. Elisa, Deshpande Amit, Kathuria Tarun, and Vishnoi Nisheeth K.. 2016. How to be Fair and Diverse? arXiv:1610.07183 cs.LG
25. Cesari Ugo, De Pietro Giuseppe, Marciano Elio, Niri Ciro, Sannino Giovanna, and Verde Laura. 2018. A new database of healthy and pathological voices. Computers & Electrical Engineering 68 (May 2018), 310–321. 10.1016/j.compeleceng.2018.04.008

26. Chen Shi and Zhao Qi. 2019. Attention-based autism spectrum disorder screening with privileged modality. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 1181–1190.
27. Chou Vivian. 2017. How science and genetics are reshaping the race debate of the 21st century. Science in the News 17 (2017).
28. Clare Eli. 2015. Exile and pride. In Exile and Pride. Duke University Press.
29. Clark Leigh, Cowan Benjamin R., Roper Abi, Lindsay Stephen, and Sheers Owen. 2020. Speech Diversity and Speech Interfaces: Considering an Inclusive Future through Stammering. Association for Computing Machinery, New York, NY, USA. 10.1145/3405755.3406139
30. Clifford Jantina Rochelle. 2005. An evaluation of the technical adequacy of a parent-completed inventory of developmental skills. (2005).
31. Cooper Richard S. 2003. Race and genomics. The New England journal of medicine 348, 12 (2003), 1166. [PubMed: 12646675]
32. Coyne Imelda T. 1998. Researching children: some methodological and ethical considerations. Journal of Clinical Nursing 7, 5 (1998), 409–416. [PubMed: 9855992]
33. Davidson Jennifer L and Jensen Carlos. 2013. What health topics older adults want to track: a participatory design study. In Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility. 1–8.
34. de Vries Terrance, Misra Ishan, Wang Changhan, and van der Maaten Laurens. 2019. Does object recognition work for everyone?. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 52–59.
35. DePaul Roxanne. 2016. DementiaBank English PPA Corpus. 10.21415/T5ZH5T
36. Diaz Mark, Johnson Isaac, Lazar Amanda, Piper Anne Marie, and Gergle Darren. 2018. Addressing Age-Related Bias in Sentiment Analysis. (2018). 10.1145/3173574.3173986
37. Digitale Erin. 2022. Study finds differences between brains of girls, boys with autism. https://med.stanford.edu/news/all-news/2022/02/autism-brain-sex-differences.html/
38. Dixon Lucas, Li John, Sorensen Jeffrey, Thain Nithum, and Vasserman Lucy. 2018. Measuring and mitigating unintended bias in text classification. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. 67–73.
39. Dodge Jesse, Sap Maarten, Marasović Ana, Agnew William, Ilharco Gabriel, Groeneveld Dirk, Mitchell Margaret, and Gardner Matt. 2021. Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 1286–1305.
40. Doshi Rohan, Chen Youzheng, Jiang Liyang, Zhang Xia, Biadsy Fadi, Ramabhadran Bhuvana, Chu Fang, Rosenberg Andrew, and Moreno Pedro J.. 2021. Extending Parrotron: An End-to-End, Speech Conversion and Speech Recognition Model for Atypical Speech. In ICASSP 2021 – 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 6988–6992. 10.1109/ICASSP39728.2021.9414644
41. Dray Susan M, Peters Anicia N, Brock Anke M, Peer Andrea, Druin Allison, Gitau Shikoh, Kumar Janaki, and Murray Dianne. 2013. Leveraging the progress of women in the HCI field to address the diversity chasm. In CHI'13 Extended Abstracts on Human Factors in Computing Systems. 2399–2406.
42. Drosou Marina, Jagadish HV, Pitoura Evaggelia, and Stoyanovich Julia. 2017. Diversity in big data: A review. Big data 5, 2 (2017), 73–84. [PubMed: 28632443]
43. Duan Huiyu, Zhai Guangtao, Min Xiongkuo, Che Zhaohui, Fang Yi, Yang Xiaokang, Gutiérrez Jesús, and Le Callet Patrick. 2019. A Dataset of Eye Movements for the Children with Autism Spectrum Disorder. In Proceedings of the 10th ACM Multimedia Systems Conference (Amherst, Massachusetts) (MMSys '19). Association for Computing Machinery (ACM), New York, NY, USA, 255–260. 10.1145/3304109.3325818
44. Engler A. 2019. For some employment algorithms, disability discrimination by default. https://www.brookings.edu/blog/techtank/2019/10/31/for-some-employment-algorithms-disability-discrimination-by-default/

45. Eraslan Sukru, Yaneva Victoria, Yesilada Yeliz, and Harper Simon. 2019. Web users with autism: eye tracking evidence for differences. Behaviour & Information Technology 38, 7 (2019), 678–700. 10.1080/0144929X.2018.1551933

46. Fausto-Sterling Anne. 2000. Sexing the body: Gender politics and the construction of sexuality. Basic Books.

47. Fazelpour Sina and De-Arteaga Maria. 2021. Diversity in Sociotechnical Machine Learning Systems. CoRR abs/2107.09163 (2021). arXiv:2107.09163 https://arxiv.org/abs/2107.09163

48. Findlater Leah, Goodman Steven, Zhao Yuhang, Azenkot Shiri, and Hanley Margot. 2020. Fairness Issues in AI Systems That Augment Sensory Abilities. 125, Article 8 (mar 2020), 1 pages. 10.1145/3386296.3386304

49. Findlater Leah and Zhang Lotus. 2020. Input Accessibility: A Large Dataset and Summary Analysis of Age, Motor Ability and Input Performance. In The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (Virtual Event, Greece) (ASSETS '20). Association for Computing Machinery, New York, NY, USA, Article 17, 6 pages. 10.1145/3373625.3417031

50. Flanagin Annette, Frey Tracy, Christiansen Stacy L, AMA Manual of Style Committee, et al. 2021. Updated guidance on the reporting of race and ethnicity in medical and science journals. JAMA 326, 7 (2021), 621–627. [PubMed: 34402850]

51. Fombonne Eric. 2009. Epidemiology of pervasive developmental disorders. Pediatric research 65, 6 (2009), 591–598. [PubMed: 19218885]

52. Ford Marvella E and Kelly P Adam. 2005. Conceptualizing and categorizing race and ethnicity in health services research. Health services research 40, 5p2 (2005), 1658–1675. [PubMed: 16179001]

53. Gala Núria, Tack Anaïs, Javourey-Drevet Ludivine, François Thomas, and Ziegler Johannes C. 2020. Alector: A parallel corpus of simplified French texts with alignments of misreadings by poor and dyslexic readers. In Language Resources and Evaluation for Language Technologies (LREC).

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

55. Giarelli Ellen, Wiggins Lisa D, Rice Catherine E, Levy Susan E, Kirby Russell S, Pinto-Martin Jennifer, and Mandell David. 2010. Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. Disability and health journal 3, 2 (2010), 107–116. [PubMed: 21122776]

56. Guo Anhong, Kamar Ece, Vaughan Jennifer Wortman, Wallach Hanna, and Morris Meredith Ringel. 2020. Toward Fairness in AI for People with Disabilities: A Research Roadmap. SIGACCESS Accessible Computing 125, Article 2 (March 2020), 1 pages. 10.1145/3386296.3386298

57. Gurari Danna, Li Qing, Stangl Abigale J., Guo Anhong, Lin Chi, Grauman Kristen, Luo Jiebo, and Bigham Jeffrey P. 2018. VizWiz Grand Challenge: Answering Visual Questions from Blind People. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (Jun 2018). 10.1109/cvpr.2018.00380

58. Hakim Haya Berman and Ratner Nan Bernstein. 2004. Nonword repetition abilities of children who stutter: An exploratory study. Journal of fluency disorders 29, 3 (2004), 179–199. [PubMed: 15458830]

59. Hamedani Ali G, VanderBeek Brian L, and Willis Allison W. 2019. Blindness and visual impairment in the medicare population: disparities and association with hip fracture and neuropsychiatric outcomes. Ophthalmic epidemiology 26, 4 (2019), 279–285. [PubMed: 31062638]

60. Hamidi Foad, Poneres Kellie, Massey Aaron, and Hurst Amy. 2018. Who Should Have Access to My Pointing Data? Privacy Tradeoffs of Adaptive Assistive Technologies. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (Galway, Ireland) (ASSETS '18). Association for Computing Machinery, New York, NY, USA, 203–216. 10.1145/3234695.3239331

61. Hamidi Foad, Scheuerman Morgan Klaus, and Branham Stacy M.. 2018. Gender Recognition or Gender Reductionism? The Social Implications of Embedded Gender Recognition Systems. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–13. 10.1145/3173574.3173582
62. Romana Hasnain-Wynia and Baker David W. 2006. Obtaining data on patient race, ethnicity, and primary language in health care organizations: current challenges and proposed solutions. Health services research 41, 4p1 (2006), 1501–1518. [PubMed: 16899021]
63. Hassan Saad, Huenerfauth Matt, and Alm Cecilia Ovesdotter. 2021. Unpacking the Interdependent Systems of Discrimination: Ableist Bias in NLP Systems through an Intersectional Lens. CoRR abs/2110.00521 (2021). arXiv:2110.00521 https://arxiv.org/abs/2110.00521
64. Hausdorff Jeffrey M, Lertratanakul Apinya, Cudkowicz Merit E, Peterson Amie L, Kaliton David, and Goldberger Ary L. 2000. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. Journal of applied physiology (2000).
65. Hausdorff Jeffrey M, Mitchell Susan L, Firtion Renee, Peng Chung-Kang, Cudkowicz Merit E, Wei Jeanne Y, and Goldberger Ary L. 1997. Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease. Journal of applied physiology 82, 1 (1997), 262–269. 10.1152/jappl.1997.82.1.262 [PubMed: 9029225]
66. Heller Tamar, Stafford P, Davis LA, Sedlezky L, and Gaylord V. 2010. People with intellectual and developmental disabilities growing old: An overview. Impact: Feature Issue on Aging and People with Intellectual and Developmental Disabilities 23, 1 (2010), 2–3.
67. Holtmann Martin, Bölte Sven, and Poustka Fritz. 2007. Autism spectrum disorders: Sex differences in autistic behaviour domains and coexisting psychopathology. Developmental Medicine & Child Neurology 49, 5 (2007), 361–366. [PubMed: 17489810]
68. Howden Lindsay M, Meyer Julie A, et al. 2011. Age and sex composition: 2010.
69. Huenerfauth Matt and Kacorri Hernisa. 2014. Release of experimental stimuli and questions for evaluating facial expressions in animations of American Sign Language. In Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel, The 9th International Conference on Language Resources and Evaluation (LREC '14). 10.1007/978-3-642-39188-0_55
70. Hutchinson Ben, Prabhakaran Vinodkumar, Denton Emily, Webster Kellie, Zhong Yu, and Denuyl Stephen. 2020. Social Biases in NLP Models as Barriers for Persons with Disabilities. CoRR abs/2005.00813 (2020). arXiv:2005.00813 https://arxiv.org/abs/2005.00813
71. Iakovakis Dimitrios, Hadjidimitriou Stelios, Charisis Vasileios, Bostantjopoulou Sevasti, Katsarou Zoe, Klingelhoefer Lisa, Reichmann Heinz, Dias Sofia B, Diniz José A, Trivedi Dhaval, et al. 2018. Motor impairment estimates via touchscreen typing dynamics toward Parkinson's disease detection from data harvested in-the-wild. Frontiers in ICT 5 (2018), 28.
72. Iakovakis Dimitrios, Hadjidimitriou Stelios, Charisis Vasileios, Bostantzopoulou Sevasti, Katsarou Zoe, and Hadjileontiadis Leontios J. 2018. Touchscreen typing-pattern analysis for detecting fine motor skills decline in early-stage Parkinson's disease. Scientific reports 8, 1 (2018), 1–13. [PubMed: 29311619]
73. Jang Jina, Matson Johnny L, Adams Hilary L, Konst Matt J, Cervantes Paige E, and Goldin Rachel L. 2014. What are the ages of persons studied in autism research: A 20-year review. Research in Autism Spectrum Disorders 8, 12 (2014), 1756–1760.
74. Joshi Aparna and Roh Hyuntak. 2009. The role of context in work team diversity research: A meta-analytic review. Academy of management journal 52, 3 (2009), 599–627.
75. Kacorri Hernisa. 2017. Teachable Machines for Accessibility. SIGACCESS - Accessible Computing 119, 10–18. 10.1145/3167902.3167904
76. Kacorri Hernisa, Dwivedi Utkarsh, Amancherla Sravya, Jha Mayanka, and Chanduka Riya. 2020. IncluSet: A Data Surfacing Repository for Accessibility Datasets. Association for Computing Machinery (ACM). 10.1145/3373625.3418026
77. Kacorri Hernisa, Dwivedi Utkarsh, and Kamikubo Rie. 2020. Data Sharing in Wellness, Accessibility, and Aging. (2020).

78. Kacorri Hernisa, Mascetti Sergio, Gerino Andrea, Ahmetovic Dragan, Takagi Hironobu, and Asakawa Chieko. 2016. Supporting Orientation of People with Visual Impairment: Analysis of Large Scale Usage Data. In Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '16). Association for Computing Machinery (ACM), 151–159. 10.1145/2982142.2982178

79. Kamikubo Rie, Dwivedi Utkarsh, and Kacorri Hernisa. 2021. Sharing Practices for Datasets Related to Accessibility and Aging. In The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (Virtual Event, USA) (ASSETS '21). Association for Computing Machinery, New York, NY, USA, Article 28, 16 pages. 10.1145/3441852.3471208

80. Kärkkäinen Kimmo and Joo Jungseock. 2019. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age. CoRR abs/1908.04913 (2019). arXiv:1908.04913 http://arxiv.org/abs/1908.04913

81. Kaur Preet Chandan, Tushar Ghorpade, and Mane Vanita. 2016. Analysis of data security by using anonymization techniques. In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence). IEEE, 287–293.

82. Kaushal Amit, Altman Russ, and Langlotz Curt. 2020. Geographic distribution of US cohorts used to train deep learning algorithms. Jama 324, 12 (2020), 1212–1213. [PubMed: 32960230]

83. Kay Matthew, Matuszek Cynthia, and Munson Sean A.. 2015. Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. Association for Computing Machinery, New York, NY, USA. 10.1145/2702123.2702520

84. Kertzer David and Arel Dominique. n.d.. Census and identity. (n. d.).

85. Kirkovski Melissa, Enticott Peter G, and Fitzgerald Paul B. 2013. A review of the role of female gender in autism spectrum disorders. Journal of autism and developmental disorders 43, 11 (2013), 2584–2603. [PubMed: 23525974]

86. Klucken Jochen, Barth Jens, Kugler Patrick, Schlachetzki Johannes, Henze Thore, Marxreiter Franz, Kohl Zacharias, Steidl Ralph, Hornegger Joachim, Eskofier Bjoern, et al. 2013. Unbiased and mobile gait analysis detects motor impairment in Parkinson's disease. PloS one 8, 2 (2013), e56956. 10.1371/journal.pone.0056956 [PubMed: 23431395]

87. Lai Meng-Chuan, Lombardo Michael V, Auyeung Bonnie, Chakrabarti Bhismadev, and Baron-Cohen Simon. 2015. Sex/gender differences and autism: setting the scene for future research. Journal of the American Academy of Child & Adolescent Psychiatry 54, 1 (2015), 11–24. [PubMed: 25524786]

88. Lee Kyungjun and Kacorri Hernisa. 2019. Hands Holding Clues for Object Recognition in Teachable Machines. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery (ACM), 1–12. 10.1145/3290605.3300566

89. Lee Matthew L. and Dey Anind K. 2007. Providing Good Memory Cues for People with Episodic Memory Impairment. In Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (Assets '07). Association for Computing Machinery (ACM), 131–138. 10.1145/1296843.1296867

90. Lehnhardt Fritz-Georg, Falter Christine Michaela, Gawronski Astrid, Pfeiffer Kathleen, Tepest Ralf, Franklin Jeremy, and Vogeley Kai. 2016. Sex-related cognitive profile in autism spectrum disorders diagnosed late in life: implications for the female autistic phenotype. Journal of Autism and Developmental Disorders 46, 1 (2016), 139–154. [PubMed: 26319250]

91. Leightley Daniel, Yap Moi Hoon, Coulson Jessica, Barnouin Yoann, and McPhee Jamie S. 2015. Benchmarking human motion analysis using kinect one: An open source dataset. In Proceedings of the 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA '15). IEEE, 1–7. 10.1109/APSIPA.2015.7415438

92. LI DONGXU, Rodriguez Cristian, Yu Xin, and LI HONGDONG. 2020. Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).

93. Linton Megan Marie Quaglia. 2021. The Institutional Remains: Transinstitutionalization of Disability & Sexuality. (2021).

94. Lipkin Paul H, Okamoto Jeffrey, Norwood Kenneth W, Adams Richard C, Brei Timothy J, Burke Robert T, Davis Beth Ellen, Friedman Sandra L, Houtrow Amy J, Hyman Susan L, et al. 2015. The Individuals with Disabilities Education Act (IDEA) for children with special educational needs. Pediatrics 136, 6 (2015), e1650–e1662. [PubMed: 26620061]
95. Loh Keng Yin and Ogle J. 2004. Age related visual impairment in the elderly. The Medical journal of Malaysia 59, 4 (2004), 562–8. [PubMed: 15779599]
96. Lohr Steve. 2018. Facial recognition is accurate, if you're a white guy. New York Times 9, 8 (2018), 283.
97. Loi Daria and Lodato Thomas. 2020. On empathy and empiricism: addressing stereotypes about older adults in technology. Interactions 28, 1 (2020), 23–25.
98. Loomes Rachel, Hull Laura, and Mandy William Polmear Locke. 2017. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. Journal of the American Academy of Child & Adolescent Psychiatry 56, 6 (2017), 466–474. [PubMed: 28545751]
99. Lupton Deborah. 2017. Digital health now and in the future: Findings from a participatory design stakeholder workshop. Digital health 3 (2017), 2055207617740018. [PubMed: 29942616]
100. Mack Kelly, McDonnell Emma, Jain Dhruv, Wang Lucy Lu, Froehlich Jon E., and Findlater Leah. 2021. What Do We Mean by "Accessibility Research"? A Literature Survey of Accessibility Papers in CHI and ASSETS from 1994 to 2019. Article 371 (2021), 18 pages. 10.1145/3411764.3445412
101. MacWhinney Brian, Bird Steven, Cieri Christopher, and Martell Craig. 2004. TalkBank: Building an open unified multimodal database of communicative interaction. In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC '04). Evaluations and Language resources Distribution Agency, 525–528. http://www.lrec-conf.org/proceedings/lrec2004/pdf/392.pdf
102. David S Mandell, Lindsay J Lawer, Kira Branch, Edward S Brodkin, Kristin Healey, Robert Witalec, Donielle N Johnson, and Raquel E Gur. 2012. Prevalence and correlates of autism in a state psychiatric hospital. Autism 16, 6 (2012), 557–567. 10.1177/1362361311412058 arXiv:https://doi.org/10.1177/1362361311412058. [PubMed: 21846667]
103. Mandy William, Chilvers Rebecca, Chowdhury Uttom, Salter Gemma, Seigal Anna, and Skuse David. 2012. Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. Journal of autism and developmental disorders 42, 7 (2012), 1304–1313. [PubMed: 21947663]
104. Manly Jennifer J.. 2006. Deconstructing Race and Ethnicity: Implications for Measurement of Health Outcomes. Medical Care 44, 11 (2006), S10–S16. http://www.jstor.org/stable/41219499 [PubMed: 17060816]
105. Mantero José Luis Pérez. 2014. Interacción y predictividad: Los intercambios conversacionales con hablantes con demencia tipo alzhéimer. revista de investigación Lingüística 17 (2014), 97–118.
106. Matheis Maya, Matson Johnny L, Hong Esther, and Cervantes Paige E. 2019. Gender differences and similarities: Autism symptomatology and developmental functioning in young children. Journal of autism and developmental disorders 49, 3 (2019), 1219–1231. [PubMed: 30443700]
107. Matthes Silke, Hanke Thomas, Regen Anja, Storz Jakob, Worseck Satu, Efthimiou Eleni, Dimou Athanasia-Lida, Braffort Annelies, Glauert John, and Safar Eva. 2012. Dicta-Sign–building a multilingual sign language corpus. In Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon (LREC '12). https://www.sign-lang.uni-hamburg.de/lrec/lrec/pubs/12016.pdf
108. Mehrabi Ninareh, Morstatter Fred, Saxena Nripsuta, Lerman Kristina, and Galstyan Aram. 2021. A Survey on Bias and Fairness in Machine Learning. ACM Comput. Surv. 54, 6, Article 115 (July 2021), 35 pages. 10.1145/3457607
109. Merler Michele, Ratha Nalini, Feris Rogerio S., and Smith John R.. 2019. Diversity in Faces. arXiv:1901.10436 cs.CV
110. Miceli Milagros, Posada Julian, and Yang Tianling. 2021. Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power? arXiv:2109.08131 cs.HC

111. Novotný Roman Čmejla-Hana Růžičková Jiří Klempíř Michal, Rusz Jan and Růžicka Evžen. 2016. Hypernasality associated with basal ganglia dysfunction: evidence from Parkinson's disease and Huntington's disease. PeerJ 4 (2016), e2530. https://dx.doi.org/10.7717%2Fpeerj.2530 [PubMed: 27703866]

112. Minors Emancipated and Minors Self-Sufficient. 2017. Guidance and Procedures: Child Assent and Permission by Parents or Guardians. https://ora.research.ucla.edu/OHRPP/Documents/Policy/9/ChildAssent_ParentPerm.pdf. (2017).

113. Mitchell Margaret, Baker Dylan, Moorosi Nyalleng, Denton Emily, Hutchinson Ben, Hanna Alex, Gebru Timnit, and Morgenstern Jamie. 2020. Diversity and Inclusion Metrics in Subset Selection. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (New York, NY, USA) (AIES '20). Association for Computing Machinery, New York, NY, USA, 117–123. 10.1145/3375627.3375832

114. Mitchell Ross E, Young Travas A, Bachelda Bellamie, and Karchmer Michael A. 2006. How many people use ASL in the United States? Why estimates need updating. Sign Language Studies 6, 3 (2006), 306–335.

115. Modak Masooda, Ghotane Ketan, Siddhanth V, Kelkar Nachiket, and Iyer Prachi G Aravind. 2019. Detection of Dyslexia using Eye Tracking Measures. International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8 (2019), 1011–1014.

116. Moffatt Karyn Anne. 2010. Addressing age-related pen-based target acquisition difficulties. Ph.D. Dissertation. University of British Columbia. http://www.sigaccess.org/2010/01/addressing-age-related-pen-based-target-acquisition-difficulties/

117. Moore Vanessa and Goodson Sally. 2003. How well does early diagnosis of autism stand the test of time? Follow-up study of children assessed for autism at age 2 and development of an early diagnostic service. Autism 7, 1 (2003), 47–63. [PubMed: 12638764]

118. Morris Meredith Ringel. 2020. AI and accessibility. Commun. ACM 63, 6 (2020), 35–37.

119. Moses Yolanda. 2017. Why Do We Keep Using the Word "Caucasian"? https://www.sapiens.org/column/race/caucasian-terminology-origin/

120. Moyle Maura Jones, Weismer Susan Ellis, Evans Julia L, and Lindstrom Mary J. 2007. Longitudinal relationships between lexical and grammatical development in typical and late-talking children. (2007).

121. Munger Kevin, Gopal Ishita, Nagler Jonathan, and Tucker Joshua A.. 2021. Accessibility and generalizability: Are social media effects moderated by age or digital literacy? Research & Politics 8, 2 (2021), 20531680211016968. 10.1177/20531680211016968 arXiv:https://doi.org/10.1177/20531680211016968

122. Neal Karama C et al. 2008. Use and misuse of 'race'in biomedical research. Journal of Health Ethics 5, 1 (2008), 8.

123. Nerenz David R, McFadden Bernadette, Ulmer Cheryl, et al. 2009. Race, ethnicity, and language data: standardization for health care quality improvement. (2009).

124. Nicol Antony, Casey Chris, and MacFarlane Stuart. 2002. Children are ready for speech technology-but is the technology ready for them. Interaction Design and Children, Eindhoven, The Netherlands (2002).

125. Nikolopoulos Spiros, Georgiadis Kostas, Kalaganis Fotis, Liaros Georgios, Lazarou Ioulietta, Adam Katerina, Anastasios Papazoglou-Chalikias, Chatzilari Elisavet, Oikonomou P. Vangelis, Petrantonakis C. Panagiotis, Kompatsiaris I, Kumar Chandan, Menges Raphael, Staab Steffen, Müller Daniel, Sengupta Korok, Bostantjopoulou Sevasti, Katsarou Zoe, Zeilig Gabi, Plotnin Meir, Gottlieb Amihai, Fountoukidou Sofia, Ham Jaap, Athanasiou Dimitrios, Mariakaki Agnes, Comanducci Dario, Sabatini Eduardo, Nistico Walter, and Plank Markus. 2017. The MAMEM Project - A dataset for multimodal human-computer interaction using biosignals and eye tracking information. 10.5281/zenodo.834154

126. World Institute on Disability. 2021. AI and Accessibility. https://wid.org/2019/06/12/ai-and-accessibility/

127. Ozonoff Sally, Goodlin-Jones Beth L, and Solomon Marjorie. 2005. Evidence-based assessment of autism spectrum disorders in children and adolescents. Journal of Clinical Child and Adolescent Psychology 34, 3 (2005), 523–540. [PubMed: 16083393]

128. Park Joon Sung, Bernstein Michael S., Brewer Robin N., Kamar Ece, and Morris Meredith Ringel. 2021. Understanding the Representation and Representativeness of Age in AI Data Sets. CoRR abs/2103.09058 (2021). arXiv:2103.09058 https://arxiv.org/abs/2103.09058

129. Park Joon Sung, Bragg Danielle, Kamar Ece, and Morris Meredith Ringel. 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.

130. Piven Joseph, Rabins Peter, and on behalf of the Autism-in Older Adults Working Group. 2011. Autism Spectrum Disorders in Older Adults: Toward Defining a Research Agenda. Journal of the American Geriatrics Society 59, 11 (2011), 2151–2155. 10.1111/j.1532-5415.2011.03632.x arXiv:https://agsjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532–5415.2011.03632.x [PubMed: 22091837]

131. Prince Martin, Knapp Martin, Guerchet Maelenn, McCrone Paul, Prina Matthew, Comas-Herrera A, Wittenberg Raphael, Adelaja Bayo, Hu Bo, King Derek, et al. 2014. Dementia UK: -overview. (2014).

132. Ratner Nan Bernsteinand Silverman Stacy. 2000. Parental perceptions of children's communicative development at stuttering onset. Journal of Speech, Language, and Hearing Research 43, 5 (2000), 1252–1263.

133. Ratto Allison B, Kenworthy Lauren, Yerys Benjamin E, Bascom Julia, Wieckowski Andrea Trubanova, White Susan W, Wallace Gregory L, Pugliese Cara, Schultz Robert T, Ollendick Thomas H, et al. 2018. What about the girls? Sex-based differences in autistic traits and adaptive skills. Journal of autism and developmental disorders 48, 5 (2018), 1698–1711. [PubMed: 29204929]

134. Rello Luz, Baeza-Yates Ricardo, and Llisterri Joaquim. 2014. DysList: An Annotated Resource of Dyslexic Errors. In Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC '14). European Languages Resources Association (ELRA), 1289–1296. http://www.lrec-conf.org/proceedings/lrec2014/pdf/612_Paper.pdf

135. Rello Luz and Ballesteros Miguel. 2015. Detecting Readers with Dyslexia Using Machine Learning with Eye Tracking Measures. In Proceedings of the 12th Web for All Conference (W4A '15). Association for Computing Machinery (ACM), Article 16, 8 pages. 10.1145/2745555.2746644

136. Rembis Michael. 2021. Crip Camp: A Disability Revolution. Journal of American History 108, 3 (12 2021), 667–669. 10.1093/jahist/jaab339 arXiv:https://academic.oup.com/jah/article-pdf/108/3/667/41938029/jaab339.pdf

137. Ritchie Hannah and Roser Max. 2019. Gender Ratio. Our World in Data (2019). https://ourworldindata.org/gender-ratio.

138. Rivet Tessa Taylor and Matson Johnny L. 2011. Review of gender differences in core symptomatology in autism spectrum disorders. Research in Autism Spectrum Disorders 5, 3 (2011), 957–976.

139. Roestorf Amanda, Bowler Dermot M, Deserno Marie K, Howlin Patricia, Klinger Laura, McConachie Helen, Parr Jeremy R, Powell Patrick, Van Heijst Barbara FC, and Geurts Hilde M. 2019. "Older Adults with ASD: The Consequences of Aging." Insights from a series of special interest group meetings held at the International Society for Autism Research 2016–2017. Research in autism spectrum disorders 63 (2019), 3–12. [PubMed: 31275429]

140. Betul Erdogdu Sakar M Isenkul Erdem, Sakar C Okan, Sertbas Ahmet, Gurgen Fikret, Delil Sakir, Apaydin Hulya, and Kursun Olcay. 2013. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics 17, 4 (2013), 828–834. 10.1109/JBHI.2013.2245674 [PubMed: 25055311]

141. Schembri Adam, Fenlon Jordan, Rentelis Ramas, Reynolds Sally, and Cormier Kearsy. 2013. Building the British sign language corpus. Language Documentation & Conservation 7 (2013), 136–154.

142. Scheuerman Morgan Klaus, Wade Kandrea, Lustig Caitlin, and Brubaker Jed R. 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.

143. Sears Andrew and Hanson Vicki. 2011. Representing Users in Accessibility Research. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 2235–2238. 10.1145/1978942.1979268
144. Sebastian Rajani, Thompson Carol B, Wang Nae-Yuh, Wright Amy, Meyer Aaron, Friedman Rhonda B, Hillis Argye E, and Tippett Donna C. 2018. Patterns of decline in naming and semantic knowledge in primary progressive aphasia. Aphasiology 32, 9 (2018), 1010–1030. [PubMed: 30613121]
145. Sen Maya and Wasow Omar. 2016. Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics. Annual Review of Political Science 19, 1 (2016), 499–522. 10.1146/annurev-polisci-032015-010015
146. Serano Julia. 2013. Excluded: Making feminist and queer movements more inclusive. Seal Press.
147. Serre David and Pääbo Svante. 2004. Evidence for gradients of human genetic diversity within and among continents. Genome research 14, 9 (2004), 1679–1685. [PubMed: 15342553]
148. Shankar Shreya, Halpern Yoni, Breck Eric, Atwood James, Wilson Jimbo, and Sculley D. 2017. No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World. arXiv:1711.08536 stat.ML
149. Sharma Shanya, Dey Manan, and Sinha Koustuv. 2021. Evaluating Gender Bias in Natural Language Inference. arXiv preprint arXiv:2105.05541 (2021).
150. Shaw Linda R, Chan Fong, and McMahon Brian T. 2012. Intersectionality and disability harassment: The interactive effects of disability, race, age, and gender. Rehabilitation Counseling Bulletin 55, 2 (2012), 82–91.
151. Shi Bowen, Del Rio Aurora Martinez, Keane Jonathan, Michaux Jonathan, Brentari Diane, Shakhnarovich Greg, and Livescu Karen. 2018. American Sign Language Fingerspelling Recognition in the Wild. In 2018 IEEE Spoken Language Technology Workshop (SLT). 145–152. 10.1109/SLT.2018.8639639
152. Shi Bowen, Del Rio Aurora Martinez, Keane Jonathan, Brentari Diane, Shakhnarovich Greg, and Livescu Karen. 2019. Fingerspelling Recognition in the Wild With Iterative Visual Attention. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).
153. Shneiderman Ben. 2020. Human-centered artificial intelligence: three fresh ideas. AIS Transactions on Human-Computer Interaction 12, 3 (2020), 109–124.
154. Singh Ajay, Yeh Chia Jung, and Blanchard Sheresa Boone. 2017. Ages and stages questionnaire: a global screening scale. Boletín Médico Del Hospital Infantil de México (English Edition) 74, 1 (2017), 5–12.
155. Sloane Mona, Moss Emanuel, Awomolo Olaitan, and Forlano Laura. 2020. Participation is not a design fix for machine learning. arXiv preprint arXiv:2007.02423 (2020).
156. Smith W, Mitchell P, and Wang JJ. 1997. Gender, oestrogen, hormone replacement and age-related macular degeneration: Results from the Blue Mountains Eye Study. Australian and New Zealand journal of ophthalmology 25, 4 (1997), 13–15.
157. Spade Dean. 2009. Trans law and Politics on a Neoliberal landscape. Trans Law and Politics on a Neoliberal Landscape (June 26, 2009). Temple Political & Civil Rights Law Review 18 (2009), 09–05.
158. Spiel Katta, Haimson Oliver L., and Lottridge Danielle. 2019. How to Do Better with Gender on Surveys: A Guide for HCI Researchers. Interactions 26, 4 (jun 2019), 62–65. 10.1145/3338283
159. Steel Daniel, Fazelpour Sina, Gillette Kinley, Crewe Bianca, and Burgess Michael. 2018. Multiple diversity concepts and their ethical-epistemic implications. European journal for philosophy of science 8, 3 (2018), 761–780. [PubMed: 30956737]
160. Strekas Amy, Ratner Nan Bernstein, Berl Madison, and Gaillard William D. 2013. Narrative abilities of children with epilepsy. International journal of language & communication disorders 48, 2 (2013), 207–219. [PubMed: 23472960]
161. SurveyMonkey. n.d.. Gathering demographic information from surveys. https://www.surveymonkey.com/mp/gathering-demographic-information-from-surveys/. Accessed: 2022-01-03.

162. Tatman Rachael. 2017. Gender and Dialect Bias in YouTube's Automatic Captions. In Proceedings of the First ACL Workshop on Ethics in Natural Language Processing. Association for Computational Linguistics, Valencia, Spain, 53–59. 10.18653/v1/W17-1606
163. Theodorou Lida, Massiceti Daniela, Zintgraf Luisa, Stumpf Simone, Morrison Cecily, Cutrell Edward, Harris Matthew Tobias, and Hofmann Katja. 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 (Virtual Event, USA) (ASSETS '21). Association for Computing Machinery, New York, NY, USA, Article 27, 12 pages. 10.1145/3441852.3471225
164. Thiyagarajan Krishna. 2016. Parkinson's Disease Observations: Variables Regarding Parkinson's Disease. https://www.kaggle.com/krisht/parkinsonsdisease.
165. Thompson David. 2002. Misplaced and forgotten:. Housing, Care and Support 5, 1 (2022/01/13 2002), 19–22. 10.1108/14608790200200006
166. Treviranus Jutta. 2018. Sidewalk Toronto and Why Smarter is Not Better*. https://medium.datadriveninvestor.com/sidewalk-toronto-and-why-smarter-is-not-better-b233058d01c8
167. Treviranus Jutta. 2019. The Value of Being Different. In Proceedings of the 16th Web For All 2019 Personalization - Personalizing the Web (W4A '19). Association for Computing Machinery (ACM), Article 1, 7 pages. 10.1145/3315002.3332429
168. Trewin Shari. 2018. AI Fairness for People with Disabilities: Point of View. CoRR abs/1811.10670 (2018). arXiv:1811.10670 http://arxiv.org/abs/1811.10670
169. Trewin Shari, Basson Sara, Muller Michael, Branham Stacy, Treviranus Jutta, Gruen Daniel, Hebert Daniel, Lyckowski Natalia, and Manser Erich. 2019. Considerations for AI Fairness for People with Disabilities. AI Matters 5, 3 (Dec. 2019), 40–63. 10.1145/3362077.3362086
170. Vásquez-Correa Juan Camilo, Arias-Vergara Tomas, Rafael Orozco-Arroyave Juan, Eskofier Björn, Klucken Jochen, and Nöth Elmar. 2018. Multimodal assessment of Parkinson's disease: a deep learning approach. IEEE journal of biomedical and health informatics 23, 4 (2018), 1618–1630. 10.1109/jbhi.2018.2866873 [PubMed: 30137018]
171. Vatavu Radu-Daniel and Ungurean Ovidiu-Ciprian. 2019. Stroke-Gesture Input for People with Motor Impairments: Empirical Results & Research Roadmap. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery (ACM), 1–14. 10.1145/3290605.3300445
172. Vyas Darshali A, Eisenstein Leo G, and Jones David S. 2020. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. 874–882 pages.
173. Wallman Katherine K. 1998. Data on race and ethnicity: Revising the federal standard. The American Statistician 52, 1 (1998), 31–33.
174. Wallman Katherine K, Evinger Suzann, and Schechter Susan. 2000. Measuring our nation's diversity: developing a common language for data on race/ethnicity. American Journal of Public Health 90, 11 (2000), 1704. [PubMed: 11076235]
175. Webster Kellie, Recasens Marta, Axelrod Vera, and Baldridge Jason. 2018. Mind the gap: A balanced corpus of gendered ambiguous pronouns. Transactions of the Association for Computational Linguistics 6 (2018), 605–617.
176. Wetherell Danielle, Botting Nicola, and Conti-Ramsden Gina. 2007. Narrative skills in adolescents with a history of SLI in relation to non-verbal IQ scores. Child Language Teaching and Therapy 23, 1 (2007), 95–113. 10.1177/0265659007072322
177. White Ryen W, Doraiswamy P Murali, and Horvitz Eric. 2018. Detecting neurodegenerative disorders from web search signals. NPJ digital medicine 1, 1 (2018), 1–4. 10.1038/s41746-018-0016-6 [PubMed: 31304287]
178. White Ryen W and Horvitz Eric. 2019. Population-scale hand tremor analysis via anonymized mouse cursor signals. NPJ digital medicine 2, 1 (2019), 1–7. 10.1038/s41746-019-0171-4 [PubMed: 31304351]
179. Whittaker Meredith, Alper Meryl, Bennett Cynthia L, Hendren Sara, Kaziunas Liz, Mills Mara, Morris Meredith Ringel, Rankin Joy, Rogers Emily, Salas Marcel, et al. 2019. Disability, Bias, and AI. AI Now Institute, November (2019). https://wecount.inclusivedesign.ca/uploads/Disability-bias-AI.pdf

180. Wolters Maria K, Kilgour Jonathan, MacPherson Sarah E, Dzikovska Myroslava, and Moore Johanna D. 2015. The CADENCE corpus: a new resource for inclusive voice interface design. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 3963–3966.

181. Xu Dongxin, Richards Jeffrey A., Gilkerson Jill, Yapanel Umit, Gray Sharmistha, and Hansen John. 2009. Automatic Childhood Autism Detection by Vocalization Decomposition with Phone-like Units. In Proceedings of the 2nd Workshop on Child, Computer and Interaction (WOCCI '09). Association for Computing Machinery (ACM), Article 5, 7 pages. 10.1145/1640377.1640382

182. Yairi Ehud and Ambrose Nicoline Grinager. 1999. Early childhood stuttering I: Persistency and recovery rates. Journal of Speech, Language, and Hearing Research 42, 5 (1999), 1097–1112.

183. Yaneva Victoria, Temnikova Irina, and Mitkov Ruslan. 2015. Accessible Texts for Autism: An Eye-Tracking Study. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15). Association for Computing Machinery (ACM), 49–57. 10.1145/2700648.2809852

184. Yaneva Victoria, Temnikova Irina, and Mitkov Ruslan. 2016. A Corpus of Text Data and Gaze Fixations from Autistic and Non-Autistic Adults. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC '16). European Language Resources Association (ELRA). https://aclanthology.org/L16-1077

185. Yang Kaiyu, Qinami Klint, Fei-Fei Li, Deng Jia, and Russakovsky Olga. 2020. Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (Barcelona, Spain) (FAT* '20). Association for Computing Machinery, New York, NY, USA, 547–558. 10.1145/3351095.3375709

186. Yang K Lisa and Tan Hock E. 2019. Disability statistics: Online resource for US disability statistics. Accessed: 2022-01-12.

187. Yaruss J Scott and Quesal Robert W. 2006. Overall Assessment of the Speaker's Experience of Stuttering (OASES): Documenting multiple outcomes in stuttering treatment. Journal of fluency disorders 31, 2 (2006), 90–115. [PubMed: 16620945]

188. Yilmaz Emre, Ganzeboom MS, Beijer LJ, Cucchiarini Catia, and Strik Helmer. 2016. A Dutch dysarthric speech database for individualized speech therapy research. (2016).

189. Youngmann Brit, Allerhand Liron, Paltiel Ora, Yom-Tov Elad, and Arkadir David. 2019. A machine learning algorithm successfully screens for Parkinson's in web users. Annals of clinical and translational neurology 6, 12 (2019), 2503–2509. 10.1002/acn3.50945 [PubMed: 31714022]

190. Zhang Hanbin, Song Chen, Wang Aosen, Xu Chenhan, Li Dongmei, and Xu Wenyao. 2019. PDVocal: Towards Privacy-Preserving Parkinson's Disease Detection Using Non-Speech Body Sounds. In Proceedings of the 25th Annual International Conference on Mobile Computing and Networking (MobiCom '19). Association for Computing Machinery, Article 16, 16 pages. 10.1145/3300061.3300125

191. Zheng Hui, Mahapasuthanon Pattiya, Chen Yujing, Rangwala Huzefa, Evmenova Anya S, and Motti Vivian Genaro. 2021. WLA4ND: A Wearable Dataset of Learning Activities for Young Adults with Neurodiversity to Provide Support in Education. Association for Computing Machinery, New York, NY, USA. 10.1145/3441852.3471220

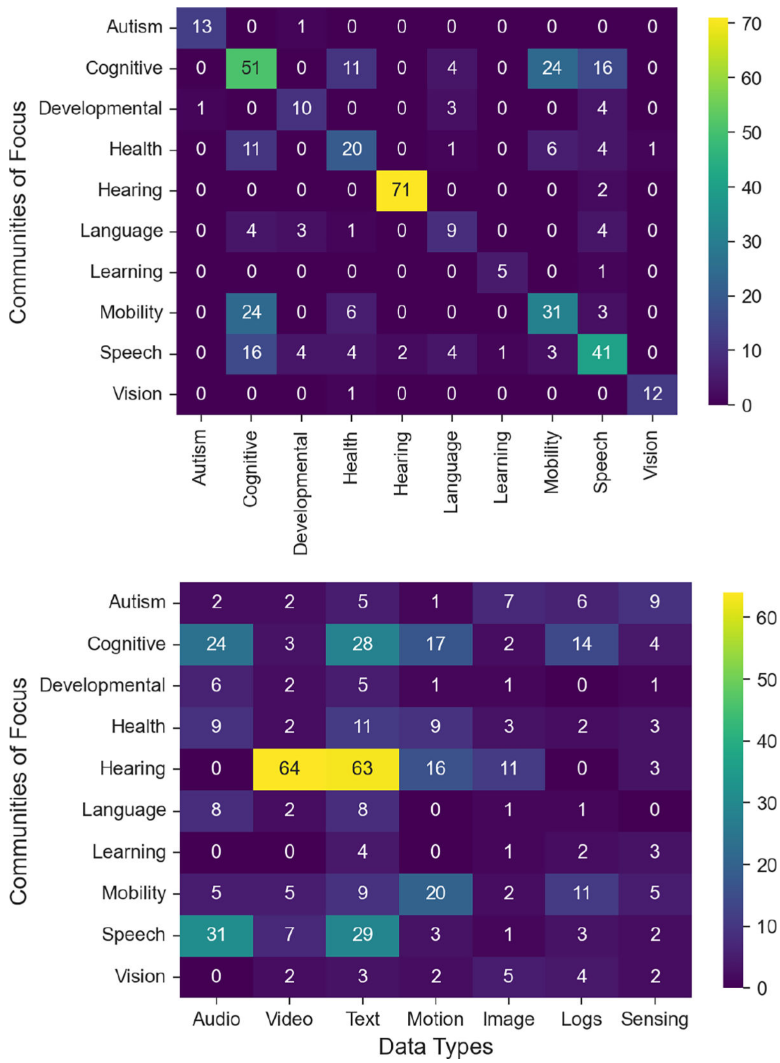

Figure 1:

Distribution of accessibility dataset count across all communities of focus (a) and data types (b).

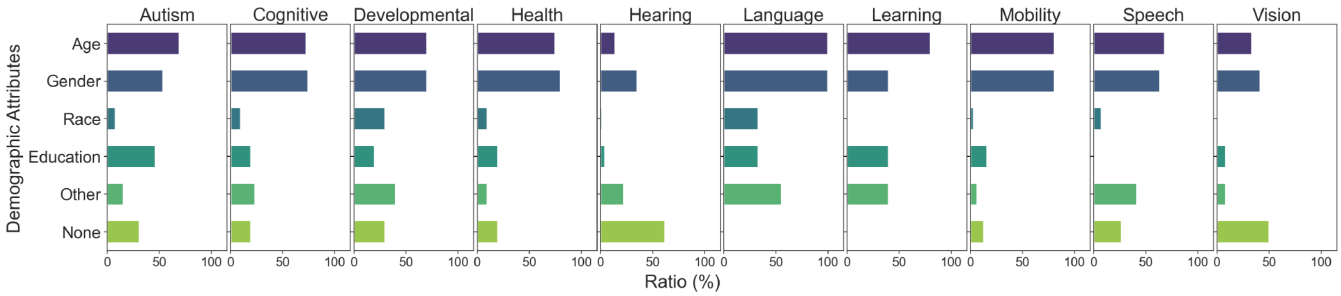

Figure 2:
Proportion of accessibility datasets across all communities including metadata related to the age, gender, race, education, or other sociocultural factors about their data contributors. Many datasets (*e.g.*, in the Hearing group) did not contain any metadata.

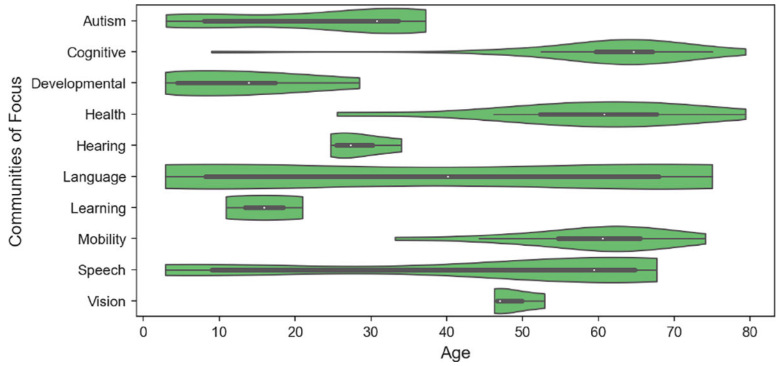

Figure 3:
Sampling distribution of ‘reported’ mean age, which differs across communities. Means are calculated on varying sample sizes.

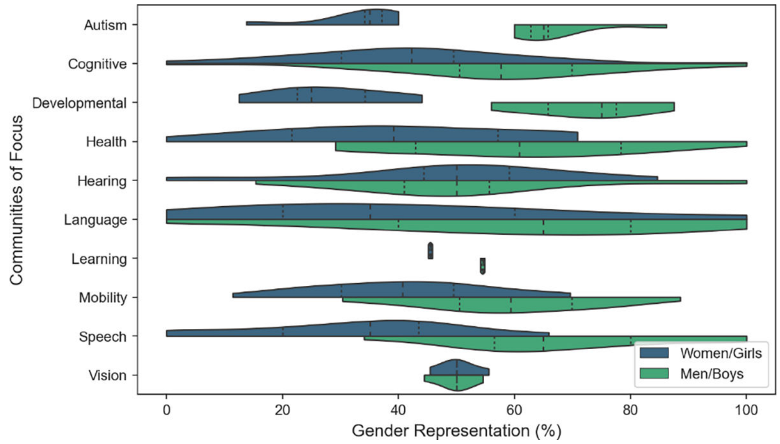

Figure 4:
Sampling distribution of gender representation across accessibility datasets. The representation gap is more prominent in some communities than others.

Table 1:

Prior work on analysis of broader AI and accessibility datasets with varying sample sizes.

	Data	# of Datasets	Age	Gender	Race	Skin Color	Geography	Sociocultural
Accessibility								
Bragg *et al.* [15]	Sign Language Datasets	n=NA						•
Kaushal *et al.* [82]	Clinical Image Datasets	n=74					•	
Broader AI								
Dodge *et al.* [39]	C4 Webtext Corpora	n=1					•	•
Merler *et al.* [109]	Face Image Datasets	n=7–8	•	•		•		
Park *et al.* [128]	Face Image Datasets	n=92	•					
Scheuerman *et al.* [142]	Face Image Datasets	n=92		•	•			
Shankar *et al.* [148]	Open Images, ImageNet	n=2					•	
Yang *et al.* [185]	ImageNet	n=1	•	•		•		