# Envisioning Equitable Speech Technologies for Black Older Adults

ROBIN N. BREWER, University of Michigan, USA
CHRISTINA N. HARRINGTON, Carnegie Mellon University, USA
COURTNEY HELDRETH, Google, USA

There is increasing concern that how researchers currently define and measure fairness is inadequate. Recent calls push to move beyond traditional concepts of fairness and consider related constructs through qualitative and community-based approaches, particularly for underrepresented communities most at-risk for AI harm. One in context, previous research has identified that voice technologies are unfair due to racial and age disparities. This paper uses voice technologies as a case study to unpack how Black older adults value and envision fair and equitable AI systems. We conducted design workshops and interviews with 16 Black older adults, exploring how participants envisioned voice technologies that better understand cultural context and mitigate cultural dissonance. Our findings identify tensions between what it means to have fair, inclusive, and representative voice technologies. This research raises questions about how and whether researchers can model cultural representation with large language models.

**ACM Reference Format:**
Robin N. Brewer, Christina N. Harrington, and Courtney Heldreth. 2023. Envisioning Equitable Speech Technologies for Black Older Adults. In *2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23), June 12–15, 2023, Chicago, IL, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3593013.3594005

## 1 INTRODUCTION

Recently, researchers have struggled to come up with accurate ways to define and measure AI fairness. Attempts to define fairness have been static, focusing on training data to prevent "unfair" decisions without considering the social context [16]. Current approaches can hyper-focus on quantitative methods that ignore structural inequity (e.g., group vs. individual fairness)(e.g., [7]). Often, notions of fairness are only associated with technical aspects such as machine learning models, inputs, and outputs [41]. As such, there have been calls to move beyond traditional conceptualizations of fairness and commonly related constructs (e.g., accuracy) to create more equitable systems that also consider the social contexts that inform decision-making systems (e.g., [6, 19, 41]). Research suggests that conceptualizing fairness requires that we consider factors such as demographic information in data and that *achieving* fairness [41] requires that we consider the social and technical system components, particularly amongst marginalized communities [8]. In this paper, we use voice technologies as a case study to investigate how underrepresented communities within ML research conceptualize fairness and related constructs of inclusivity and representation.


*FAccT '23, June 12–15, 2023, Chicago, IL, USA*

ACM ISBN 979-8-4007-0192-4/23/06.
https://doi.org/10.1145/3593013.3594005

Despite the benefits of voice technologies (e.g., voice assistants), previous research has shown how disparities exist, causing less-than-ideal experiences for different groups. Specifically, marginalized groups such as older adults or Black Americans have both reported having experiences with automated speech recognition software that they feel might be biased. These lived experiences are also grounded in data as recent work shows that the average word error rate for Black adults is significantly higher than for white adults [29]. In addition to race, prior work suggests that age also plays a role in user experiences with speech technologies [11, 37]. Research also demonstrates how older adults prefer different voices than younger adults and that age-related disability can impact speech [10, 15, 49]. When investigating the intersection of age and race, researchers have found that Black older adults perceive voice technologies to be inequitable and designed for others [23]. While few studies have focused on this specific intersection, there is a unique opportunity to frame fairness among a historically disadvantaged group [22].

The primary goal of this project is to explore new ways to engage impacted communities in conversations about concepts related to equity. We also seek to define values and envision futures for fair and equitable AI technologies with and for Black older adults (ages 55+). Our primary research questions are:

- RQ1: How do Black older adults expect voice technologies to respond in equitable and culturally responsive ways?
- RQ2: How do Black older adults operationalize 'fairness', 'representation', and 'inclusivity' with voice and AI technologies?

To address these research questions, we conducted community-based design workshops and interviews with 16 Black older adults to explore their experiences with voice technologies and their expectations of fairness, inclusivity, and representation. We worked with a community organization focused on engaging Black older adults in research to develop a research agenda that would ethically obtain perspectives, experiences, and insights on fairness in AI and voice technologies. We build upon prior literature that suggests inequities in voice technology experiences among Black older adults by identifying ideal system responses and how social constructs of identity impact system interactions. We contribute a sociotechnical perspective of fairness based on an in-depth look at the social context of how Black older adults experience voice technologies. Through this work, we argue for increased authentic cultural representations as a core component of fairness and raise questions about data disclosure boundaries for underrepresented communities.

## 2 RELATED WORK

### 2.1 Older Adults' Experiences with Voice Technologies

Voice technologies provide feasible means of task support among older adults [9, 31, 38] that can also serve the purpose of a companion

- addressing several challenges of older adults attempting to age-in-place. Research in this area has explored the potential benefits of voice technology use for intelligent information search [11, 12, 23] and as a way to aid with daily activities [9, 31, 38]. Also, the voice modality potentially has a lower learning curve for novice technology users [14, 37, 40]. Voice technologies stand to increase technology interaction among a group of users that have long felt that technology is out of reach due to the ability to engage with voice technology devices in a conversational dynamic.

As voice technologies and stand-alone conversational devices become more pervasive in older adult households [1], researchers have identified potential and observed barriers and roadblocks that disrupt what is perceived to be an easy-to-use tool. In focus groups with 38 older adults, Trajkova and Martin-Hammond [45] found that many older adults could see the potential value in engaging with Amazon Echo. However, participants abandoned these devices because they felt that there was little value beyond entertainment and that they should not rely on these devices for tasks they could complete themselves. Other research in this area has similarly determined that while older adults perceive initial value in these devices, this may be different in practice due to perceived privacy concerns [45], natural language processing limitations [11], and obstacles in attempted use [23, 31]. In summary, older adults may feel that the nature of speech recognition and voice interactions can be more frustrating than helpful.

While many studies have explored older adults' perceptions of voice technologies [23, 32, 37, 45], more recent studies have used scenario-based feasibility testing to understand actual experiences, or surveyed users that have previous experience with these devices [11, 23, 37, 42]. Previous work exploring voice technology use for online health information seeking has determined that communication breakdowns and frustrations among older users stem from expectations of conversational-like interactions [11, 23]. Sin et al. [42] highlight the primitive nature of voice technologies that limits their capabilities compared to the promise of their functionality. More specifically, among Black older users, speech recognition within voice technologies may not understand those who speak outside of what is considered standard English [23]. The challenges with speech recognition and information depth suggest the need for further research into what causes these devices to fall short and the implications of these experiences on perceptions of inclusivity, fairness, and representation among voice and other AI technologies.

## 2.2 Racial Disparities in Speech Recognition

Recent research on speech recognition and voice technology accessibility has surfaced tensions between voice technology benefits and racial disparities in automated speech recognition (ASR) [29]. Current ASR approaches may perpetuate stereotypes among racially minoritized users [13, 35] or those with non-American English dialects [26]. The research community continues to investigate inadequate dialect representation despite emerging initiatives to help machines learn how people speak [4, 29, 33]. Some of the most widely available and used voice technologies have had disproportionate error rates between Black and white users [24, 29, 33]. In a study comparing commercial voice technology devices, researchers found substantial errors among Black users' transcriptions compared to white users [29, 34]. This study also found that Black men had significantly more challenges when compared to other subgroups and that those who spoke African-American Vernacular English (AAVE) experienced higher error rates. Mengesha et al. found that not only did Black Americans experience higher rates of errors and resulting frustrations when engaging with voice technologies, but these experiences directly contributed to feelings of being "othered" or excluded [34]. Here, researchers surveyed voice technology users and observed reports of participants feeling ASR systems fail when they speak with AAVE, which may be normative to their speech patterns. Others explored ASR accuracy when capturing speakers across different dialects, genders, and races and found that while there were no significant differences across genders, there were significant error rates for non-white users [44]. Similar research on user experiences by Harrington et al. showed that older Black Americans face more challenges with ASR. Here, participants indicated that these poor experiences with voice technologies have led to feelings of exclusion and assumptions "that these technologies are not intended for them" [23]. Although limited qualitative studies exist in this area, all seem to highlight that experiencing error rates or challenges with voice technologies cause feelings of exclusion or inequitable experiences among Black users.

While most research on racial disparities in ASR and voice technologies has been inclusive of Black users, we see few empirical or exploratory research studies that examine how age and race intersect. Harrington et al. argue for considering Black older adults' experiences with voice technologies, particularly among a group that may also experience normative age-related changes in speech. [23]. This work also points to challenges with inclusivity and representation in voice technology experiences that may persist across other intersecting identity groups (e.g., disability). Along with other emerging research, we argue that the research community needs to understand Black older adults' expectations of fairness, inclusivity, and representation and where current experiences fall short.

## 2.3 Fairness and Voice Technologies

Most ASR and voice technologies have been deemed to be "unfair" simply because they do not operate in equal ways for all users [39]. These systems have higher error rates for certain groups. Unfairness is usually attributed to a lack of diverse representation in speech training data, which could be an ethical and a legal concern [17]. Researchers agree that demographic group affiliation should not change the accuracy of ASR. Prior work uses a counterfactual fairness approach where researchers trained ASR systems to achieve "equivalent output label distributions" for speakers whose voices had different protected attributes [39].

Systems are often trained on datasets that are not inclusive of variations of English. Most datasets include "American English" or "Standard American English", but few have considered African American English (AAE) or African American Vernacular English (AAVE) [33]. In one study, researchers tested whether a novel attacker could exploit racial and gender biases in ASR, considering Standard American English, Korean, and Nigerian with varying genders [47]. Findings confirm that racial biases exist, particularly for female speakers with accented speech. Other fairness literature confirms gender speech bias in [17]. Much of this work explores fairness by measuring accuracy to calculate disparity, or manipulating datasets to weight underrepresented data or collect more data [36]. Most of these findings highlight gender and racial bias but neglect age. In this paper, we take a sociotechnical approach to voice technology bias, particularly at the intersection of race and age.

## 3 METHODS

Researchers have proposed justice (e.g., [28]) and reparations-based approaches (e.g., [43]) to fairness that recognize the role of power imbalances [5] and structural inequities [21]. This work calls for centering marginalized groups through community-based research [21, 30, 41]. Because research shows increased voice technology adoption by older adults [1] and highlights racial disparities that exist with speech technologies [25, 29, 33, 44], we focus our study on understanding how Black older adults envision fair and equitable futures with AI-powered voice technologies by conducting design workshops and interviews with 16 Black older adults.

### 3.1 Data Collection

After an internal ethics review from our organization, two co-authors conducted two two-part design workshops (four workshops total) in-person and remote post-interviews from July to August 2022 with 16 Black older adults. We asked each participant to participate in one Workshop 1 and one Workshop 2. Each workshop lasted 1.5 - 2 hours. Participants were compensated $125 at the end of each workshop. Each workshop was audio recorded, with two of the co-authors taking observational notes and memos throughout the workshops. Post-interviews were at most 30 minutes and we compensated participants $50 at the end of each interview. Each post-interview was audio or video recorded.

**Workshop 1:** In the first design workshop (Figure 1), we sought to understand everyday experiences with voice technologies, probing for experiences where conflict occurred. Instances of conflict could include system misinterpretation, user frustration, or the user perceiving the system as exhibiting bias. At the beginning of each workshop, we asked participants to introduce themselves and briefly share how they wanted to engage with voice technologies on a phone or computer. After these introductions, we provided a grounding definition of the phrase "conversational assistant" or "voice assistant" as "devices that can talk back to you" and shared examples of common voice assistants such as Alexa, Siri, and Google Assistant. Next, we asked participants to share their initial reactions to voice assistants, their perceived purpose, and any concerns they might have. To understand individual experiences with voice assistants that people may not have felt comfortable sharing aloud, we also asked participants to write a brief diary entry on a sheet of paper, describing one experience they had with a voice assistant. We intentionally presented participants with a neutral prompt to encourage them to describe positive or negative experiences. For participants who had not used a voice assistant, we asked them to use the sheet of paper to describe something they know or have heard about voice assistants or technologies. We invited participants to share what they had written.

Fig. 1. Participants in the first design workshop writing on sticky notes


Fig. 2. Participant completing the storyboarding activity

Next, we conducted a role-playing storyboarding activity with participants to probe for and understand conflict (Figure 2). We defined storyboarding, engaged participants in an example storyboarding activity, and grouped participants in pairs. Within each pair, we asked one person to play the role of the voice assistant and the other to play the user. We asked each pair to draw a storyboard about an interaction they might have or have had with a voice assistant where the assistant did not understand them. We instructed each pair to draw their interactions in storyboard frames and write any additional notes to accompany their storyboard. After the activity, we invited participants to share their storyboards with the group.

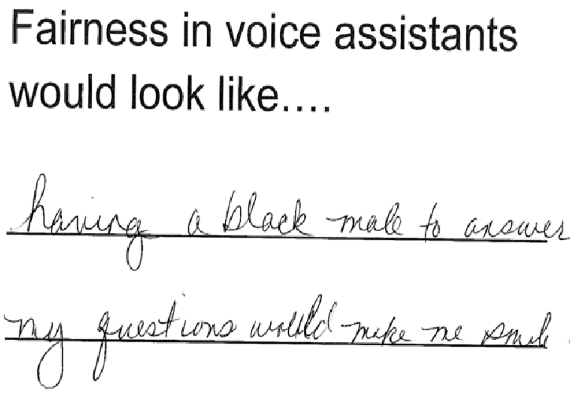


Fig. 3. Fill-in-the blank response about fairness in voice assistants

As our goal was to understand how conflict and identity with voice technologies intersected, we ended each Workshop 1 with an activity to probe for how one's identity might affect voice technology use. We gave each participant a small group of sticky notes for this activity. We asked them to write down if there are any parts of their identity (1) that they think the voice assistant should know about them to respond better, and (2) that they think makes it hard for the voice assistant to respond to them. We invited participants to share what they had written and for others to reflect on their responses. As participants wrote their notes and shared them verbally, one of the workshop facilitators grouped them into common themes on a large sheet of paper towards the front of the room. We used these discussions as a launching point for Workshop 2.

**Workshop 2:** The second workshop investigated how Black older adults defined and conceptualized concepts of fairness, representation, and inclusivity in voice technologies and AI-powered technologies, more broadly.

After a brief icebreaker, participant introductions, and Workshop 1 summary, we asked participants to define terms we would use throughout the workshop. We gave participants sticky notes and asked them to write down any word(s) or phrases they associated with "artificial intelligence", "voice assistant", "fairness", "representation", and "inclusivity". Once participants finished writing, we engaged them in a discussion to identify patterns and group sticky notes.

Next, we gave each participant six sheets of paper, each containing a fill-in-the-blank prompt to explore how participants envision FATE-related concepts with AI and voice technologies (Figure 3). We invited participants to share what they had written and comment on others' responses. The prompts were:

- Fairness in artificial intelligence would look like:
- Fairness in voice assistants would look like:
- Representation in artificial intelligence would look like:
- Representation in voice assistants would look like:
- Inclusivity in artificial intelligence would look like:
- Inclusivity in voice assistants would look like:

Lastly, we conducted an activity to connect identity to fairness, representation, and inclusivity. We asked participants to discuss how their identity as a Black person, an older adult, or the intersection of being a Black older adult impacted their experiences using voice assistants. We also asked how technologists could design voice assistants to better support people in these identity groups (ethnicity, late-life) or others (e.g., gender, education, income).

**Post-Interviews:** Shortly after each workshop, we invited each participant to complete a brief post-interview to learn more about their individual voice technology experiences, contextualize reactions made during the workshops, or ask for further clarification on workshop statements. We also used these post-interviews to invite further reflection on workshop discussions about fairness, representation, and inclusivity.

## 3.2 Participants

To recruit Black older adult participants, we partnered with the Healthier Black Elders Center (HBEC) [20], a community organization whose goal is to increase Black older adult representation in research (ages 55+). Through our partnership with HBEC, we engaged with a community advisory board of four older adults who advised the researchers on recruitment strategies, study location accessibility, and the research protocol. As such, we started recruitment with a verbal announcement made by HBEC staff at an in-person member meeting. After this announcement, those who expressed interest (n = 72) were asked to complete a brief screening questionnaire online or by phone with a research team member. This screener questionnaire asked participants about demographic information (age, gender, disability, race, ethnicity) and voice technology experience. Of those who completed the screening questionnaire (n = 24), we contacted 16 to participate in the study. We prioritized our selection to include a range of diversity, including gender, age, disability, and variation in voice technology experiences. Also, we used the screening survey to ensure that all participants were over the age of 55 and identified as Black. Sixteen Black older adults participated in our research study (ages 57-86, avg. age = 69, 3 = male, 12 = female; additional demographics in the Appendix).

## 3.3 Analysis

Data from this study consisted of two audio recordings from the workshops, design artifacts from the workshop activities, including storyboards, post-its and flipcharts, written stories and fill-in-the-blanks, and post-interview recordings. A third-party organization transcribed all workshops and interviews. One co-author reviewed the workshop transcripts to analyze the design workshop data and created a codebook representing themes from the workshop discussions. After discussions with the research team, we iteratively refined the codes in the workshop codebook. There were 17 codes in this codebook, reflecting discussion patterns such as *Fairness in Voice Tech*, *Representation in Voice Tech*, *Inclusivity in Voice Tech*, F*uture Design for Equity*, *Cultural Dissonance*, and *Race Knowledge* [1]. Next, one co-author used this codebook to qualitatively code both Workshop 1 transcripts. Similarly, another co-author used the codebook to qualitatively code both Workshop 2 transcripts.

[1]We share our full codebook in the Appendix

Another co-author developed a separate codebook representative of the workshop artifacts, including their voice technology experiences expressed in diary form (in Workshop 1) and fill-in-the-blank responses (in Workshop 2). Voice technology experiences were coded based on sentiment (positive, negative, neutral) and topic (directions, music, health, "how-to", technical support, general search, and games). A research member categorized participants' fill-in-the-blank responses with nine codes, including 'Humanizing Text-to-Speech' (e.g., "allow more voice types and be friendly") and 'Equitable Access' (e.g., "available for all"). We provide our full codebook in the appendix. Because discussion and reflection took place during the interviews, we captured photos of the large flip charts where the workshop facilitators grouped the sticky notes into themes, yet did not do a separate analysis of this data, relying on the interview transcripts to capture this data.

We report patterns observed from this qualitative coding process in our findings. As such, we focus less on quantifying the number of people who made certain arguments and instead acknowledge that the argument exists. We take a qualitative approach to data collection and analysis as recent arguments made in the FAccT community discuss how race and fairness are"contested" constructs that may not be quantifiable [21, 27, 50].

As a practice of research reflexivity, we include our positionality with respect to the participants included in these workshops. Our research team is composed of three college-educated Black-identifying women from across the United States. None of our research team identifies as an older adult but we have each had previous experience doing research with this population. Two research team members had prior experience collaborating with the community organization. It was from existing collaborations with this community that we decided to partner with them for this project as the makeup of HBEC fits the target population of Black older adults. We believe that these existing collaborations also helped to build trust with the participants, the community advisory board of older adults, and community partner staff.

## 4 FINDINGS

### 4.1 Current Voice and AI Technology Expectations (RQ1)

During workshops, participants described how they expected to interact with voice technologies based on their current design. In much of this discussion, they did not expect voice technologies to understand cultural knowledge or their voices. At a higher level, participants expected that using AI technologies could lead to overuse or exclude communities, engaging in discussions about community boundaries and equitable participation in voice and AI technology design.

*4.1.1 Misunderstanding Cultural Information.* During the design workshops, several participants (n = 5) described how they did not expect the current state of voice technologies to understand cultural knowledge. One participant explicitly stated, "*there's certain cultural knowledge or ethnic knowledge or something that we wouldn't necessarily expect a voice assistant to know*" (P1). This knowledge included information about holidays important to the African American community, common sayings, and regional knowledge. In describing a scenario where a user might ask for cultural information, one participant details how she envisions the system responding:

> "*In our scenario, we assumed—we decided that Alexa might not know all the cultural stuff...She asked, "What is Juneteenth?" Alexa said, "I did not understand. Could you please rephrase?" She hollered, "What is Juneteenth?" Alexa said, "Juneteenth is the name of a Richard Wright novel...Is Juneteenth a national holiday? Then Alexa could figure it out.*" (P1)

P1 acknowledged how voice technologies can answer factual questions common to mainstream audiences, such as information about books. However, she did not expect voice technologies to know "*all the cultural stuff*" related to a holiday celebrating Black heritage such as Juneteenth, which might include its historical significance or celebratory traditions. She attributed this to community size, saying that voice technologies are "not programmed to know, simply because they are niche information for smaller communities" (P1). Implicitly, this statement reveals an assumption that voice and AI technologies only know information for historically represented communities.

Similarly, other participants (n = 5) described how voice technologies might not have appropriate regional knowledge to answer questions about locations or local holiday celebrations. For example, one participant described the storyboard scenario they created:

> "*Our question was, she asked, "How large is Bell Isle?" I'm like, "How do you spell that?" Then she said, "I-S-L-E." Then I said, "How long is Bell Island?" Then she was like, "No, Isle. I-S-L-E." It was kind of like when you're hearing "isle" and then "island," playing off those kind of words* [sic] (P9)"
>
> "*Because it's so used to the common I-S-L-A-N-D.* (P12)"

Here, P9 and P12 show how they expected voice technologies to understand certain location-related information based on word popularity (isle vs. island). In these quotes, participants describe how words and phrases that are important to their community may be ignored by other communities. As such, they expected voice technologies to misunderstand community-specific cultural and regional knowledge.

*4.1.2 Excluding Certain Voices.* Often, participants described how the technologies did not understand their voice (n = 9). These types of interactions led to participants expecting that the voice assistant would not understand them with their initial voice query. For example, P11 described an experience with needing to rephrase a query several times for deleting a song from a playlist. She was unsure why her voice assistant did not understand her initial query. In another example, P12 felt confident that ethnicity affected how the system understood her voice:

> "*When I speak to my voice assistant, I'm like maybe I'm using too much Ebonics. Let me speak using my king's English. Let me use my white girl voice. 'I said I would like to—' sometimes you have to be specific and articulate with it. Your eloquence of speech matters it seems because oftentimes, it'll say, 'I don't understand what you said.' I'm like, 'Maybe you don't like the way I'm talking.'*"

Similar to prior work [23], P12 described how she "code switches" or alternates between English that sounds more "white" than Black. This quote suggests that there is an expectation that white-sounding voices are understood better than Black-sounding voices. Voice technologies often have training processes where the user can fine-tune speech recognition according to specific voice features. Current features that are intended to improve natural language processing did not work well for participants. Two participants were aware of these training features. For example:

> "*It doesn't recognize every voice. The cell phone even tells you to train your voice to the phone. I did get a notification on this device that told me, 'Train your voice to the phone.' I know I did that when I got the phone, but it asked me to do it again, so I did it again.* (P13)"

These quotes show that participants expect voice technologies to favor certain speech patterns over others. While it may be unclear what type of speech leads to better responses, others suspect ethnicity can play a role. Moreover, current strategies for mitigating speech recognition errors have yet to work well for participants. These findings raise questions about how to mitigate speech-related bias and other forms of bias with AI technologies.

*4.1.3 Leaving Communities Behind.* As voice technologies are one type of AI technology, we also engaged participants in higher-level discussions about AI expectations. Their discussions unpacked concerns about historically underrepresented communities being excluded from AI systems and design processes, which could negatively impact data representation.

Several participants echoed concerns about becoming overly reliant on AI technologies (n = 7). Concerns about reliance and dependence were part of larger conversations about humans becoming "*obsolete*" (P2) and AI technologies feeling "*impersonal*" (P16). For example, P4 asked, "*What is the need for humans? Because artificial intelligence becomes [sic] so human, do they need us?*" Although researchers continue to develop models that can better mimic human behavior and speech (e.g., with natural language processing), participants were concerned that machines would replace humans. Other participants such as P13 envisioned a more robotic world with the future of AI, saying, "Now what they got to do is put a chip in us. And everywhere we go, you can scan the chip, scan yourself." To participants, AI is, and will continue to be, inauthentic.

The workshop coordinators observed how such conversations reflected participants' confusion. about AI technology boundaries or how to describe AI. This confusion connected with P16's concerns when she said, "*My concern is that everything is going to artificial identification or artificial intelligence. And I am just so afraid that a lot of blacks are going to be left behind.*" During the design workshops, 3 participants were concerned that researchers would exclude Black people from the future of AI and voice technology design. For example, P5 wanted "*more people from diverse backgrounds having input into the programming process*" as this could mitigate exclusion and lead to more equitable voice and AI futures. These participants did not want developers and researchers to exclude underrepresented communities.

## 4.2 Envisioning Future Voice Technologies (RQ1)

To address these concerns, participants described how voice and AI technologies should work with users to mitigate cultural dissonance, proactively engage with users, and be more transparent about conversational boundaries.

*4.2.1 Mitigating Cultural Dissonance.* Participants (n = 7) wanted voice technologies to understand cultural and regional differences and work to mitigate cultural dissonance. For example, P5 wanted voice technologies to know that "*if someone's talking about sugar, they're not talking about dominos, they're talking about diabetes.*" To do so, voice technologies would be programmed with "*a dictionary of idioms that only basically black people usually use*" (P5). This would mean that voice technologies would need to know the user's ethnicity to provide culturally relevant information.

In addition to identity-related information, participants (n = 4) wanted voice technologies to use location data to provide regional responses. When describing his experience with voice assistants, P2 recounted a time when he tried to ask his GPS to travel to a city in his current state, but it continued to route him to a different city in a nearby state. To address such instances, P5 asked, "[...] *does it give you any way of saying a region you can refer to?*" Similar to existing voice training processes, enabling one's location to improve responses may already exist, but this process was not apparent to participants.

These examples raise questions about when to provide voice technologies with this information and whether they should automatically learn or infer identity-related information or have users manually provide such information. We return to these open questions in the discussion section.

*4.2.2 Being Proactive and Transparent.* Participants (n = 6) also described how voice technologies could be more proactive when they could not understand a user's voice. Aligning with expectations in prior work [11], participants wanted voice technologies to ask them to "*repeat the question*", ask "*could you be clearer?*" (P8), or "*ask them to spell it*" (P9). These participants wanted voice technologies to prompt the user with strategies that could improve system performance and provide them with strategies they could use in subsequent searches for better outcomes.

Participants (n =5) expressed how voice technologies, similar to other AI technologies, could better leverage their learning capabilities. P2 said, "*If you keep asking it the same thing and it gives you the wrong answer, there ought to be a way for it to correct itself.*" Currently, when participants re-ask questions after receiving an incomplete or incorrect response, voice technologies do not change their answer, leading to a continued frustrating experience. Similarly, P10 said, "*I think the machines don't try to help us. If you're talking to a person, they would immediately try to help you. They just wouldn't stupidly say, 'I don't understand you.' Ask me again.*" These quotes point towards envisioning voice and AI technologies as tools that can proactively guide users.

Moreover, three participants envisioned voice technologies that could be more transparent. In one workshop, participants discussed what they should expect from voice technologies. P1 said, "*we can't know which things it doesn't know the answer to.*" When using voice technologies, some participants expected to be able to ask any question and receive a response. Others considered voice technology limitations, e.g., "*I think it's kind of disingenuous to sell me something that's supposed to be able to answer all my questions*" (P1). In both cases, participants wanted voice and AI technologies to share knowledge boundaries. These knowledge boundaries can be about general information or identity-related information specific to communities.

## 4.3 Fairness, Representation, and Inclusivity Perceptions (RQ2)

We asked Black older adults to operationalize equity-related concepts of fairness, inclusion, and representation. We observed identity-related nuances in how they conceptualize each term. Discussions about these definitions revealed tensions about how voice technologies could best represent Black speech.

*4.3.1 Defining Fairness, Representation, and Inclusion.* The FAccT community has done well at critiquing fairness, with open questions about how to define and whether to measure fairness. We argue that there is value in understanding other equity-related constructs (e.g., representation and inclusion). Our findings highlight similarities and differences between these constructs, particularly around how participants spoke about different facets of identity.

In one activity, participants completed fill-in-the-blank prompts about 'fairness', 'representation', and 'inclusivity' with AI and voice technologies. Participants' responses about fairness (25 prompt responses) highlighted the importance of humanizing text-to-speech systems (7/25) and understanding all voices or topics (7/25). In their completed prompts, many had references to being "*equal*", "*all*", or "*anyone*". For example, participants wrote that fairness means "availability to all and an understanding of all voices" (P16) and "*being able to understand all voices regardless to gender, tones, accents.*" These statements rarely included information about race or ethnicity (3/25), or age (0/25).

We also asked participants to reflect on inclusion in voice and AI technologies (26 prompt responses). Similarly, these responses focused on understanding all voices or topics (n = 13) or humanizing text-to-speech (7/26). For example, inclusivity means "*the community addressing both seniors, youth, and middle age*" (P14), and "*all types of voices would be used - men, women, even children if appropriate, different accents*" (P11). Many of these responses focused on identity (e.g., gender, age, race, ethnicity), understanding differences, and personalization.

We also asked participants to define representation in voice and AI technologies (32 prompt responses). Participants defined representation as understanding all voices and topics (14/32), cultural representation (10/32), and personalization (10/32). For example, participants defined representation as "*lack of cultural/racial bias in responses. using varied examples when appropriate*" (P1) "*artificial intelligence would know more about my ethnic group*" (P6), and "*represent[ing] the community it serves. [Anonymized City] community majority African American*" (P14). There were also several responses about voice diversity, including "*having diversity and choices with how you are able to interact with different devices. Like how you can change the hand colors in Zoom!*" and "*you being able to choose who answers your questions such as Midwestern, Southern, older mature voice or young with high-pitched voice; being able to customize it to your own personal settings*" (P15). In these responses about representation, there was more focus on ethnicity, regions, and languages than when defining fairness. There were also fewer responses about gender and age compared to inclusion.

We use these statements to define how participants perceive these three equity-related concepts for voice and AI technologies. Our findings show how participants understood each term slightly differently. These patterns justify why aligning with different equity concepts is important and help contextualize tensions with voice technologies equitably representing Black older adults.

*4.3.2 Tensions with Voice Representation.* Participants discussed how to represent voice diversity without reflecting stereotypes and provide identity information to improve speech interactions while preserving privacy.

Some participants (n = 5) expressed how representative and inclusive voice technologies would sound like a Black person speaking. Participants like P4 said, "*As an African American, I would like to see blacks being used for voiceover for voice assistants*" because "*I think you feel it's more friendly when you're talking to a Black person*" (P15). P10 said, "*I feel ethnicity is important, and I think all the voices are white. They all sound like white people. I've had enough of white people in my lifetime.*" This quote shows how P10 had not encountered diverse AI voices and wanted speech that better reflected her identity. Doing so in a way that does not reflect stereotypes is tricky as P3 said, "*They don't have to be 'Yo, baby,' or whatever, but just you sense that this is not someone either above you or beneath you, just somebody probably like your friends and family.*" To P3 and others, Black sounding voices would be comforting.

Other participants (n = 6) wanted to personalize the voice, e.g., with "*a Mississippi accent*" or "*a man that sounds young and fine, like Tupac*" (P12). These quotes show how participants wanted to have a single Black voice option or choose from a range of Black voice options. Many voice technologies offer options for switching the default voice, and we reviewed these voices with participants in one workshop. However, no one thought the options sounded Black or older. Although one solution could be to increase the types of voices offered, there were some concerns with doing so.

One participant strongly disagreed that voice technology options should "sound Black." P1 said:

> "*I think it's a little bit dangerous, though, to assume that there's some certain way that all Black people want to be spoken to. I think that there's a wide range within each—any ethnic group about the way people speak and what have you. I don't know how seriously I would take some source of information if it was trying to pretend to be Black just because I'm Black…I don't like ages, and I don't like sexes, and I don't like racism. Don't discriminate me because I'm a Black woman, because I'm older.*"

Similarly, she reflected on the diversity of Black speech, saying, "*Even if it was a different language, there's no language every single Black person speaks. That's my point, that to assume some kind of monolithic Black culture is not really appropriate*" (P1). These quotes challenge the assumption that increasing voice options could address representation, fairness, or inclusion. Instead, these conversations highlight how there is no single, representative Black voice and how trying to create one could be discriminatory. Further, these quotes suggest that some participants do not trust that developers would appropriately represent the nuance of Black speech.

*4.3.3 Tensions with Disclosing Identity for Inclusion.* Similar to voice representation tensions, participants disagreed on whether voice and AI technologies should collect or store identity-related information to provide more fair, inclusive, or representative responses. Some participants (n = 9) said it would be helpful for voice assistants to know some aspects of their identity, particularly for "cultural stuff" (P3). P13 said:

> "*I really don't want to know about things that aren't specific to me. I want to know about Black haircare, how to cook collard greens if that's what I want to know. I don't want to know how to play Euchre, and I don't want to know how to put a perm in someone else's hair that has straight hair, or how to get a tan.*"

In this example, P13 notes certain contexts where it would be important for voice and AI technologies to know her ethnicity (hair, food, entertainment, skin care). Health was another context where participants indicated that understanding one's identity would be important. P16 wanted voice and AI technologies to "*have representation to the black communities. If you're asking a question related to medical, it should be related to blacks.*"

However, not all participants agreed and wanted more neutral responses (n = 3). P8 did not "*necessarily feel that voice assistants should be concerned with my ethnicity. I think my articulation should be on its own merit. I ask a clear, concise question, and it gives me a clear, concise response.*" In part, neutrality is aligned with maintaining anonymity. P12 said, "*I don't want anything geared towards knowing who I am on this other side. You don't need to know if I'm old or if I'm Black or if I'm white or if I'm young. Just give me my answer, whoever gives it to me. I don't want you to distinguish how to give me the answer.*" There were also questions as to how to disclose identity information ("*how does it know if you're Black or white?*", P12).

These quotes highlight tensions with disclosing identity, where some participants acknowledged that doing so could lead to more inclusive results, whereas others wanted system neutrality. These quotes provide evidence that designing more inclusive and representative systems is complex. We return to these tensions in the discussion section.

## 5 DISCUSSION

Returning to our initial research questions, Black older adults expect voice and AI technologies to mitigate cultural dissonance and be transparent about knowledge limitations to facilitate equitable speech interactions (RQ1). We contribute to the ongoing discourse around "fairness" in the FATE community with an analysis of how Black older adults operationalize fairness, as well as representation, and inclusivity (RQ2). These terms seem to be conceptualized differently, raising tensions with equitable voice representation and identity-related disclosure. In the remainder of this section, we unpack the complexity and nuance needed for cultural representation in AI technologies, proposing an authenticity lens to fairness. We also use participants' concerns about disclosure to raise open questions for the FAccT community on how to involve communities in research and incorporate cultural knowledge into large models.

### 5.1 Cultural Representation in Voice Technology

Our findings reveal tensions between what Black older adults wanted from voice and AI technologies and what currently exists. Although developers and ML researchers have made efforts to allow people to choose voices that are more representative, participants did not see voices that resonate with their culture (e.g., tone, dialect, knowledge)represented in these options. Further, participants had several discussions about how to represent Black-sounding voices without reinforcing existing stereotypes about Black people and speech. For participants, *fairness*, *representation*, and *inclusivity* in voice technologies meant authenticity.

Authenticity is a nuanced construct, particularly for intersectional fairness. To address identity-based differences, researchers have argued for group fairness. While group fairness does define what is 'fair' differently depending on subgroups, it still does not adequately recognize structural and systemic differences that exist for these groups [21]. Similarly, Kong (2022) acknowledges that intersectional fairness is challenging because current measures focus on attributes rather than inequitable systems, and acknowledging each combination of one's identities would result in too many subgroups [30]. Rather than focus on quantifiably fair systems [18, 39], we argue that ML and fairness researchers should understand how to build systems that provide authentic cultural representations. Applications can go beyond speech technologies, such as authenticity in image or text generation. Doing so raises open questions for FAccT researchers. Who decides what is authentic? Participants raised concerns that large tech company employees usually did not represent their identities (as Black or older). As such, how would employees at these companies be able to build systems with truly authentic voices? Even beyond speech, how would these companies be able to train models to produce authentic responses to search questions about culturally important information (e.g., Juneteenth)? Cherumanal points to the challenge of voice tech not being able to answer complex questions [36]. Culturally-tailored information adds another layer of complexity to these questions. User-centered and community-based approaches are one way to mitigate these challenges. However, we argue for a more expansive view of 'community.' Could a community-based approach also mean that smaller community-owned tech organizations that are more representative of users' identities be the ones responsible for developing authentic language models? Or, could policies be created to require large tech companies to consult with these smaller community-based organizations for continued model and algorithmic auditing? We raise these questions to the FAccT community as it includes members from small nonprofit research organizations to large, global organizations.

*5.1.1 Incorporating Communities into Large Models.* Based on our findings, we argue for better incorporating cultural, identity-based knowledge into large models. Prior work highlights the need for large models and datasets to be more demographically balanced [17], yet there is no understanding of what this means for cultural references of data and information. Our findings raise questions about whether and how participants wanted voice and AI technologies to collect or store information about their identities. Using identity data could improve search results and responses from voice technologies. Participants discussed how identity-based personalization could be beneficial (1) when asking health questions for conditions that are more prevalent in Black communities or (2) when seeking detailed information about cultural events unique to Black people.

Prior work in the FAccT and accessibility communities describes how researchers need to question whether more data about people will improve fairness. Such data can misrepresent communities, collection can be a form of surveillance, and using such data has privacy risks [3, 46, 48]. Researchers propose a less data-centric view and instead raise questions about when such data is appropriate to collect and why [2]. Aligning with our findings, we raise open questions about when and how to incorporate cultural and identity data in large language models. Although machine learning approaches are scalable, how might researchers incorporate more manual approaches where users in less represented communities can teach models niche cultural or regional differences? Similarly, are there (or should there be) boundaries of how much identity information users should disclose (manual) or machines should learn (automated)? Moreover, could manual or automated approaches to learning more about underrepresented communities create fairness disparities when compared to majority groups?

### 5.2 Limitations

Throughout this paper, we often refer to participants' intersecting identities as Black and older. However, we recognize that participants likely had other identities that affected their perceptions of equitable voice technologies, such as gender and disability (visible and less visible). Similarly, most of our participants were women. Future work with older Black men could explore themes we unpack in the paper.

## 6 CONCLUSION

This study presents how Black older adults define and envision equitable voice and AI technologies based on constructs of fairness, inclusivity, and representation. We find that they want systems to better mitigate cultural dissonance and responsibly use identity-related cultural data in large language models. Our findings point towards authenticity as being an important fairness-related construct to consider, suggest ways to involve communities in model design and auditing, and raise open questions as to whether researchers need data boundaries with underrepresented communities.

## ACKNOWLEDGMENTS

We thank the Healthier Black Elders Center (HBEC) for partnering with us for this research. HBEC is supported by a grant from the National Institutes of Health, 5P30 AG015281, and the Michigan Center for Urban African American Aging Research.

## REFERENCES

[1] 2020. . https://www.nationalpublicmedia.com/uploads/2020/04/The-Smart-Audio-Report_Spring-2020.pdf

[2] McKane Andrus, Elena Spitzer, Jeffrey Brown, and Alice Xiang. 2021. What We Can't Measure, We Can't Understand: Challenges to Demographic Data Procurement in the Pursuit of Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 249–260. https://doi.org/10.1145/3442188.3445888

[3] McKane Andrus and Sarah Villeneuve. 2022. Demographic-Reliant Algorithmic Fairness: Characterizing the Risks of Demographic Data Collection in the Pursuit of Fairness. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1709–1721. https://doi.org/10.1145/3531146.3533226

[4] Rosana Ardila, Megan Branson, Kelly Davis, Michael Henretty, Michael Kohler, Josh Meyer, Reuben Morais, Lindsay Saunders, Francis M Tyers, and Gregor Weber. 2019. Common voice: A massively-multilingual speech corpus. *arXiv preprint arXiv:1912.06670* (2019).

[5] Chelsea Barabas, Colin Doyle, JB Rubinovitz, and Karthik Dinakar. 2020. Studying up: Reorienting the Study of Algorithmic Fairness around Issues of Power. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 167–176. https://doi.org/10.1145/3351095.3372859

[6] Reuben Binns. 2018. Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency*. PMLR, 149–159.

[7] Reuben Binns. 2020. On the Apparent Conflict between Individual and Group Fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 514–524. https://doi.org/10.1145/3351095.3372864

[8] Abeba Birhane, Elayne Ruane, Thomas Laurent, Matthew S. Brown, Johnathan Flowers, Anthony Ventresque, and Christopher L. Dancy. 2022. The Forgotten Margins of AI Ethics. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 948–958. https://doi.org/10.1145/3531146.3533157

[9] Kenneth A Blocker, Travis Kadylak, Lyndsie M Koon, Christopher E Kovac, and Wendy A Rogers. 2020. Digital home assistants and aging: initial perspectives from novice older adult users. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 64. SAGE Publications Sage CA: Los Angeles, CA, 1367–1371.

[10] Robin Brewer, Raymundo Cornejo Garcia, Tedmond Schwaba, Darren Gergle, and Anne Marie Piper. 2016. Exploring Traditional Phones as an E-Mail Interface for Older Adults. *ACM Trans. Access. Comput.* 8, 2, Article 6 (jan 2016), 20 pages. https://doi.org/10.1145/2839303

[11] Robin Brewer, Casey Pierce, Pooja Upadhyay, and Leeseul Park. 2022. An Empirical Study of Older Adult's Voice Assistant Use for Health Information Seeking. *ACM Trans. Interact. Intell. Syst.* 12, 2, Article 13 (jul 2022), 32 pages. https://doi.org/10.1145/3484507

[12] Robin N. Brewer. 2022. "If Alexa Knew the State I Was in, It Would Cry": Older Adults' Perspectives of Voice Assistants for Health. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 442, 8 pages. https://doi.org/10.1145/3491101.3519642

[13] Joy Buolamwini. [n.d.]. Voicing erasure. https://www.ajl.org/voicing-erasure

[14] Simon Coghlan, Jenny Waycott, Lei Nui, Kelly Caine, and Brodrick Stigall. 2022. Swipe a Screen or Say the Word: Older Adults' Preferences for Information-Seeking with Touchscreen and Voice-User Interfaces. In *Proceedings of the 33rd Australian Conference on Human-Computer Interaction (OzCHI '21)*. Association for Computing Machinery, New York, NY, USA, 130–143. https://doi.org/10.1145/3520495.3520512

[15] Dorothy D Dunlop, Jing Song, Larry M Manheim, Martha L Daviglus, and Rowland W Chang. 2007. Racial/ethnic differences in the development of disability among older adults. *American journal of public health* 97, 12 (2007), 2209–2215.

[16] Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and Removing Disparate Impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*. Association for Computing Machinery, New York, NY, USA, 259–268. https://doi.org/10.1145/2783258.2783311

[17] Gianni Fenu, Giacomo Medda, Mirko Marras, and Giacomo Meloni. 2020. Improving Fairness in Speaker Recognition. In *Proceedings of the 2020 European Symposium on Software Engineering (ESSE 2020)*. Association for Computing Machinery, New York, NY, USA, 129–136. https://doi.org/10.1145/3393822.3432325

[18] Gianni Fenu, Giacomo Medda, Mirko Marras, and Giacomo Meloni. 2020. Improving Fairness in Speaker Recognition. In *Proceedings of the 2020 European Symposium on Software Engineering (ESSE 2020)*. Association for Computing Machinery, New York, NY, USA, 129–136. https://doi.org/10.1145/3393822.3432325

[19] Jessie Finocchiaro, Roland Maio, Faidra Monachou, Gourab K Patro, Manish Raghavan, Ana-Andreea Stoica, and Stratis Tsirtsis. 2021. Bridging Machine Learning and Mechanism Design towards Algorithmic Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 489–503. https:

//doi.org/10.1145/3442188.3445912

[20] LaToya N Hall, Lisa J Ficker, Letha A Chadiha, Carmen R Green, James S Jackson, and Peter A Lichtenberg. 2016. Promoting retention: African American older adults in a research volunteer registry. *Gerontology and geriatric medicine* 2 (2016), 2333721416677469.

[21] Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud. 2020. Towards a Critical Race Methodology in Algorithmic Fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 501–512. https://doi.org/10.1145/3351095.3372826

[22] Christina N. Harrington. 2020. The Forgotten Margins: What is Community-Based Participatory Health Design Telling Us? *Interactions* 27, 3 (apr 2020), 24–29. https://doi.org/10.1145/3386381

[23] Christina N. Harrington, Radhika Garg, Amanda Woodward, and Dimitri Williams. 2022. "It's Kind of Like Code-Switching": Black Older Adults' Experiences with a Voice Assistant for Health Information Seeking. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 604, 15 pages. https://doi.org/10.1145/3491102.3501995

[24] Wiebke Toussaint Hutiri and Aaron Yi Ding. 2022. Bias in Automated Speaker Recognition. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 230–247. https://doi.org/10.1145/3531146.3533089

[25] Wiebke Toussaint Hutiri and Aaron Yi Ding. 2022. Bias in Automated Speaker Recognition. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 230–247. https://doi.org/10.1145/3531146.3533089

[26] Chinaemere Ike, Seth Polsley, and Tracy Hammond. 2022. Inequity in Popular Speech Recognition Systems for Accented English Speech *(IUI '22 Companion)*. Association for Computing Machinery, New York, NY, USA, 66–68. https://doi.org/10.1145/3490100.3516457

[27] Abigail Z. Jacobs and Hanna Wallach. 2021. Measurement and Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 375–385. https://doi.org/10.1145/3442188.3445901

[28] Maximilian Kasy and Rediet Abebe. 2021. Fairness, Equality, and Power in Algorithmic Decision-Making. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 576–586. https://doi.org/10.1145/3442188.3445919

[29] Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, and Sharad Goel. 2020. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences* 117, 14 (2020), 7684–7689. https://doi.org/10.1073/pnas.1915768117 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1915768117

[30] Youjin Kong. 2022. Are "Intersectionally Fair" AI Algorithms Really Fair to Women of Color? A Philosophical Analysis. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 485–494. https://doi.org/10.1145/3531146.3533114

[31] Lyndsie M Koon, Sean A McGlynn, Kenneth A Blocker, and Wendy A Rogers. 2020. Perceptions of digital assistants from early adopters aged 55+. *Ergonomics in Design* 28, 1 (2020), 16–23.

[32] Jarosław Kowalski, Anna Jaskulska, Kinga Skorupska, Katarzyna Abramczuk, Cezary Biele, Wiesław Kopeć, and Krzysztof Marasek. 2019. Older Adults and Voice Interaction: A Pilot Study with Google Home. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312973

[33] Joshua L. Martin and Kevin Tang. 2020. Understanding Racial Disparities in Automatic Speech Recognition: The Case of Habitual "be". In *Proc. Interspeech 2020*. 626–630. https://doi.org/10.21437/Interspeech.2020-2893

[34] Zion Mengesha, Courtney Heldreth, Michal Lahav, Juliana Sublewski, and Elyse Tuennerman. 2021. "I don't Think These Devices are Very Culturally Sensitive."—Impact of Automated Speech Recognition Errors on African Americans. *Frontiers in Artificial Intelligence* (2021), 169.

[35] Cade Metz. 2020. There is a racial divide in speech-recognition systems, researchers say. https://www.nytimes.com/2020/03/23/technology/speech-recognition-bias-apple-amazon-google.html

[36] Sachin Pathiyan Cherumanal. 2022. Fairness-Aware Question Answering for Intelligent Assistants. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 3492. https://doi.org/10.1145/3477495.3531682

[37] Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of Intelligent Voice Assistants by Older Adults with Low Technology Use. *ACM Trans. Comput.-Hum. Interact.* 27, 4, Article 31 (sep 2020), 27 pages. https://doi.org/10.1145/3373759

[38] Wendy Rogers, George Mois, Kenneth Blocker, and Raksha Mudar. 2022. EXPLORING THE POTENTIAL OF VOICE INTERFACES FOR EMERGING TECHNOLOGIES: OPPORTUNITIES AND CHALLENGES. *Innovation in Aging* 6, Supplement_1 (2022), 216–216.

[39] Leda Sarı, Mark Hasegawa-Johnson, and Chang D. Yoo. 2021. Counterfactually Fair Automatic Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021), 3515–3525. https://doi.org/10.1109/TASLP.2021.3126949

[40] Sergio Sayago, Barbara Barbosa Neves, and Benjamin R Cowan. 2019. Voice Assistants and Older People: Some Open Issues. In *Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 7, 3 pages. https://doi.org/10.1145/3342775.3342803

[41] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 59–68. https://doi.org/10.1145/3287560.3287598

[42] Jaisie Sin, Dongqing Chen, Jalena G. Threatt, Anna Gorham, and Cosmin Munteanu. 2022. Does Alexa Live Up to the Hype? Contrasting Expectations from Mass Media Narratives and Older Adults' Hands-on Experiences of Voice Interfaces. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 18, 9 pages. https://doi.org/10.1145/3543829.3543841

[43] Wonyoung So, Pranay Lohia, Rakesh Pimplikar, A.E. Hosoi, and Catherine D'Ignazio. 2022. Beyond Fairness: Reparative Algorithms to Address Historical Injustices of Housing Discrimination in the US. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 988–1004. https://doi.org/10.1145/3531146.3533160

[44] Rachael Tatman and Conner Kasten. 2017. Effects of Talker Dialect, Gender & Race on Accuracy of Bing Speech and YouTube Automatic Captions.. In *Interspeech*. 934–938. https://doi.org/10.21437/Interspeech.2017-1746

[45] Milka Trajkova and Aqueasha Martin-Hammond. 2020. "Alexa is a Toy": Exploring Older Adults' Reasons for Using, Limiting, and Abandoning Echo. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376760

[46] Shari Trewin, Sara Basson, Michael Muller, Stacy Branham, Jutta Treviranus, Daniel Gruen, Daniel Hebert, Natalia Lyckowski, and Erich Manser. 2019. Considerations for AI Fairness for People with Disabilities. *AI Matters* 5, 3 (dec 2019), 40–63. https://doi.org/10.1145/3362077.3362086

[47] Payton Walker, Nathan McClaran, Zihao Zheng, Nitesh Saxena, and Guofei Gu. 2022. BiasHacker: Voice Command Disruption by Exploiting Speaker Biases in Automatic Speech Recognition *(WiSec '22)*. Association for Computing Machinery, New York, NY, USA, 119–124. https://doi.org/10.1145/3507657.3528558

[48] Meredith Whittaker, Meryl Alper, Cynthia L Bennett, Sara Hendren, Liz Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Rankin, Emily Rogers, Marcel Salas, et al. 2019. Disability, bias, and AI. *AI Now Institute* (2019), 8.

[49] Georgia Zellou, Michelle Cohn, and Bruno Ferenc Segedin. 2021. Age-and gender-related differences in speech alignment toward humans and voice-AI. *Frontiers in Communication* 5 (2021), 600361.

[50] Marilyn Zhang. 2022. Affirmative Algorithms: Relational Equality as Algorithmic Fairness. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 495–507. https://doi.org/10.1145/3531146.3533115