Speech AI for All: The What, How, and Who of Measurement

Kimi Wenzel
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
kwenzel@andrew.cmu.edu

Alisha Pradhan
Department of Informatics, Ying Wu College of Computing
New Jersey Institute of Technology
Newark, New Jersey, USA
alisha.pradhan@njit.edu

Maria Teleki
Department of Computer Science & Engineering
Texas A&M University
College Station, Texas, USA
mariateleki@tamu.edu

Tobias M Weinberg
Computer Science - Matter of Tech Lab
Cornell Tech
New york, New York, USA
Tmw88@cornell.edu

Robin Netzorg
EECS
University of California, Berkeley
Berkeley, California, USA
robert_netzorg@berkeley.edu

Alyssa Hillary Zisk
AssistiveWare
Amsterdam, Netherlands
alyssa@assistiveware.com

Anna Seo Gyeong Choi
Cornell University
Ithaca, New York, USA
sc2359@cornell.edu

Jingjin Li
AImpower.org
Mountain View, California, USA
jingjin@aimpower.org

Raja Kushalnagar
Gallaudet University
Washington, District of Columbia USA
raja.kushalnagar@gallaudet.edu

Colin Lea
Apple
Cupertino, California, USA
colin_lea@apple.com

Abraham Glasser
Science, Technology, Accessibility, Mathematics, and Public Health (STAMP)
Gallaudet University
Washington, District of Columbia USA
abraham.glasser@gallaudet.edu

Christian Vogler
Technology Access Program
Gallaudet University
Washington, District of Columbia USA
christian.vogler@gallaudet.edu

Ly Xīnzhèn M. Zhngsūn Brown
Disability Studies
Georgetown University
Georgetown, District of Columbia USA
lyxinzhen.zhangsunbrown@autistichoya.com

Nan Bernstein Ratner
Neuroscience and Cognitive Science
University of Maryland, College Park
College Park, Maryland, USA
nratner@umd.edu

Allison Koenecke
Cornell University
Ithaca, New York, USA
koenecke@stanford.edu

Karen Nakamura
Anthropology
University of California, Berkeley
Berkeley, California, USA
knak@berkeley.edu

Shaomei Wu
AImpower.org
Menlo Park, California, USA
shaomei@aimpower.org

Abstract

Optimized for "typical" and fluent speech, today's speech AI systems perform poorly for people with speech diversities, sometimes to an unusable or even harmful degree. These harms play out in daily life through household voice assistants and workplace meeting services, in higher stakes scenarios like medical transcription, and in emerging applications of AI in augmentative and alternative communication. Standard metrics aiming to quantify these inequities, however, fail to comprehensively understand the impact

CHI EA '26, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/26/04
https://doi.org/10.1145/3772363.3778768

of speech AI on diverse user groups, and furthermore do not easily generalize to newer speech language and speech generation models. To address these social inequities and measurement limitations, this workshop brings academics, practitioners, and non-profit workers together in proactive dialogue to improve measurement of speech AI performance and user impact. Through a poster session and breakout group discussions, our workshop will extend current understanding on how to best leverage existing metrics, like Word Error Rate, within the HCI design ecosystem, and also explore new innovations in speech AI measurement. Key outcomes of this workshop include: a research agenda for CHI community to guide and contribute to speech AI development, groundwork for new papers on speech AI measurement, and a diversity-centered benchmark suite for external evaluators.

CCS Concepts

• **Human-centered computing → Accessibility**; **Human computer interaction (HCI)**; • **Computing methodologies → Artificial intelligence**; • **Social and professional topics → Computing / technology policy**.

Keywords

AI FATE, automatic speech recognition, augmentative and alternative communication, speech technology, disability, accessibility, speech diversity

ACM Reference Format:

Kimi Wenzel, Alisha Pradhan, Maria Teleki, Tobias M Weinberg, Robin Netzorg, Alyssa Hillary Zisk, Anna Seo Gyeong Choi, Jingjin Li, Raja Kushalnagar, Colin Lea, Abraham Glasser, Christian Vogler, Ly Xīnzhèn M. Zhngsūn Brown, Nan Bernstein Ratner, Allison Koenecke, Karen Nakamura, and Shaomei Wu. 2026. Speech AI for All: The What, How, and Who of Measurement. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3772363.3778768

1 Motivation — Why this workshop?

Speech AI — artificial intelligence technologies that interpret and generate human-like speech — have accelerated rapidly, both in their technical abilities and in user adoption. Powered by advancements in automatic speech recognition (ASR), speech language models (SLM), and speech synthesis, speech AI products are popular both as consumer technologies (ex. home smart speakers and car interfaces) as well as in professional domains (ex. customer service agents, AI interviewers, and meeting transcription software). Despite their proliferation, speech AI are limited due to their training on typical[1] speech, often underserving those with diverse patterns such as stuttering [14, 15], d/Deaf and Hard-of-Hearing (DHH) speech [6], speech patterns of individuals experiencing age-related changes [22, 26], aphasia [11], second language speech [10, 31], gendered speech [25], as well as regional vernaculars and ethnic dialects [12]. Materially, these deficiencies have far-reaching consequences, such as worse medical transcriptions for people with aphasia [2], adverse hiring outcomes for DHH individuals [20], and psychological harms [3, 30, 31]. For the augmentative and alternative communication (AAC) community, users often communicate through synthesized voices or with atypical prosody that current systems fail to recognize, marginalizing them in contexts where speech AI is becoming infrastructure for daily interaction [28]. Recent work has shown how voice representation and accent bias in synthetic voices shape identity and inclusion in digital spaces [5, 19, 21], while AAC-specific studies highlight how timing, rhythm, and cultural nuance are central to expressive communication [27, 29]. Given the proliferation of AAC tools and speech AI systems more broadly, advancing accessible speech AI is not only a technical challenge but also a socio-cultural imperative that directly impacts users' agency, identity, and participation in society. While CHI and other venues have published studies addressing the underperformance of speech AI systems in practice, little has been done to overview or propose new methods for measurement of harms. For example, speech AI systems rely primarily on sparsely populated leaderboards that overindex on singular metrics [24].

The first iteration of this workshop [32] brought together researchers, practitioners, policy makers, and community advocates to build a cross-sector coalition for fair and accessible speech AI. Together, we laid the groundwork and identified the most pressing gaps in the designing and developing speech AI that is accessible for people with speech diversities. In the second annual workshop, we will address the primary research gap that our expert participants identified: *measurement*.

Current ways of assessing how well speech recognition works typically considers word error rate (WER), the number of transcription errors divided by the total words. Yet, encompassing different types of errors into a single WER fails to capture *who* experiences these errors and *how* the errors impact access. For example, a commercial speech recognition system with "good" average WER (7%) might still have higher errors for older adults or people with diverse speech abilities or patterns [6, 26]. In this context, there is an opportunity to develop new metrics that capture these nuanced, inclusive, and user-centered dimensions of performance. For instance, it might be fruitful to consider approaches such as abandonment rate, type of breakdown, or psychological impact. While empirical studies with diverse user groups in HCI are surfacing some promising alternatives, such as interaction-error rates, retries-to-success, and subgroup-specific measures, we still lack systematic, widely adopted metrics that bridge insights from these user studies into the core evaluation of speech recognition technologies. Developing and debating such user-centered metrics for inclusive speech AI is a central goal of this workshop.

More recently, speech language models (SLMs) [1, 4] – conditioned on both speech and text – have been proposed as an alternative approach to modular ASR-based systems, offering the promise of fully-trainable end-to-end systems that can integrate recognition, understanding, and generation in a single framework. These models are particularly promising for adapting to diverse user speech, potentially enabling more equitable interactions [16]. However, equity in SLMs remains extremely underexplored: to date, only a handful of studies have explicitly addressed fairness concerns in these models [1, 17, 33]. Both ASR and SLM paradigms are important, but they require different strategies for auditing, intervention, inclusive design, and measurement.

[1]In this workshop, we define typical speech to include fluent speech, speech with a dominant accent (often embodied by those with higher socioeconomic classes), and dominant vernaculars.

1.1 Workshop Themes and Objectives

During the workshop, we will discuss the following topics, with a goal of developing research agendas and collaborations to pursue these lines of inquiry long-term.

- **Connecting existing metrics to downstream user impact.** Previous work has demonstrated the disparate cognitive and emotional burdens of speech AI [7, 9, 13, 18, 30] on people with speech diversities. We aim to bridge the gap between engineers and designers by elucidating connection between engineer-relevant metrics and user impact. Looking towards established metrics, such as word error rate (WER), we will discuss potential research projects and user studies to answer the questions: What level of WER is (un)acceptable, for whom, in what context? How does WER specifically link to the social and emotional impact of the end user?
- **Developing new holistic metrics for different stages of the speech AI lifecycle.** As the development cycle of speech AI continues to change, ways of measuring and evaluating accuracy and downstream user impact must also evolve. While WER can be useful at the model level (millions of words) for ASR outputs in modular systems, it is often limited at the individual level (few words, high context) and in end-to-end speech systems. What new methods of evaluation and auditing may be appropriate for modular vs. end-to-end speech AI systems?

2 Length of the workshop

This workshop will span 2 sessions (180 minutes total). See Table 1 for the complete schedule.

3 Organizers

Kimi Wenzel is a PhD candidate at Carnegie Mellon University's Human-Computer Interaction Institute. Her research centers on understanding the downstream representational harms of speech AI, with a focus on psychological impacts. Her work examining speech diversities in AI has won awards from CHI and the Center for Machine Learning and Health. She has previously co-led workshops at CHI and FAccT.

Alisha Pradhan is an Assistant Professor in the Department of Informatics in the Ying Wu College of Computing at the New Jersey Institute of Technology. Her research has examined design and use of conversational voice technologies by older adults, in particular identifying the accessibility benefits and barriers posed by these technologies, and engaging older adults in design of equitable voice technologies. She has co-organized workshops and panels at conferences including, ACM CHI and UbiComp.

Maria Teleki is a PhD student in the Department of Computer Science & Engineering at Texas A&M University. Her research sits at the intersection of spoken language processing and human-AI interaction, focusing on disfluencies, spontaneous speech, and conversational systems. Recently, her work at ICWSM investigates how large language models reflect and perpetuate biases in gendered speech, with implications for more equitable conversational AI design.

Tobias Weinberg is a PhD candidate at Cornell Tech whose research focuses on augmentative and alternative communication (AAC) and the socio-technical dimensions of AI-mediated speech. As both an AAC user and researcher, his research investigates how personalization, timing, and cultural nuance shape communicative agency, identity, and privacy. Drawing on his Argentinian and bilingual background, his research also examines how cultural and linguistic identity are negotiated in AI-mediated communication. This dual perspective—lived experience and technical research—positions him to contribute unique insights on how speech AI can better recognize, represent, and support communicators with speech diversities.

Robin Netzorg (she/her) is a PhD student at the University of California at Berkeley, advised by Professors Gopala Anumanchipalli and Bin Yu. Her research focuses on interpretable models of speaker identity that quantifies expert perception of voice quality, with an aim to apply these methods to gender-affirming voice training and speech therapy as assistive tools. She believes strongly in the potential for machine learning to support, not replace, human expertise, and has tried to accomplish this via community-driven design and collaboration.

Alyssa Hillary Zisk (they/them) is the AAC (augmentative and alternative communication) research team lead at AssistiveWare. Their work addresses a variety of topics related to AAC use and access, including some discussion of AI applications in AAC. Dr. Zisk is also a mostly-speaking part-time AAC user. When they are speaking, they speak near-mainstream American English, second language Mandarin Chinese, and scattered words from Hebrew and Yiddish. When using AAC, they struggle to access linguistically appropriate speech output beyond English and gender-aligned speech output in any language.

Anna Seo Gyeong Choi is a 5th-year PhD Candidate in Information Science at Cornell University, where she works at the intersection of AI Ethics, Algorithmic Fairness, and Speech Technologies. Her research focuses on evaluating and improving the fairness of Automatic Speech Recognition systems, with particular emphasis on high-stakes speech applications and linguistically minoritized populations. Anna's work has been published in leading conferences including ACM FAccT, Interspeech, ICASSP, and AIES. She is currently affiliated with University of Pennsylvania, Linguistic Data Consortium as a Visiting Scholar.

Jingjin Li is a research fellow at AImpower.org where she leads research on co-designing inclusive Speech AI and videoconferencing tools with the stuttering community. She has co-organized panels and served on the organizing committee for CSCW and DIS.

Raja Kushalnagar is a Deaf Professor and Director of the Artificial Intelligence, Accessibility and Sign languages Center with deaf-accented speech. His research explores the accuracy and usability of conversational voice technologies by deaf and hard of hearing people, and fairness, equity and inclusion in AI models and platforms. He has served on the program committee at ASSETS, ICCHP and CSUN.

Colin Lea is a research scientist and manager at Apple. His group focuses on making interactive technologies — especially speech — more inclusive for people with disabilities. His work is at the intersection of machine learning, HCI, and accessibility and emphasizes data collection/curation and ML modeling.

Abraham Glasser is a faculty member in the Accessible Human-Centered Computing and Policy (AHCP) program at Gallaudet University, where he is also co-director of the Rehabilitation Engineering Research Center on Technology for the Deaf and Hard of Hearing (DHH RERC). He and his students conduct Human Computer Interaction (HCI) research involving AI, speech and sign language technologies, immersive technologies, and accessible computing for Deaf and Hard of Hearing (DHH) users. He is also a member of the Coalition for Sign Language Equity in Technology (CoSET) and has contributed to published resources supporting standards work, e.g. for AI-based interpreting. Overall, Dr. Glasser has published numerous works and delivered award-winning presentations at prestigious international venues, including ACM, IEEE, and other events.

Christian Vogler is a deaf person who speaks English with an accent that derives both from his German roots and his deafness. He is the director of the Technology Access Program at Gallaudet University and has led numerous grants and projects on accessible technologies for the DHH. Some of his most recent work focuses on both text-to-speech and speech-to-text technologies for DHH people. He has served on the organizing committee of conferences and workshops, including ASSETS, Gesture Workshops, AI-related workshops, and others.

Ly Xīnzhèn M. Zhǎngsūn Brown (also published widely as Lydia X. Z. Brown) is an internationally recognized advocate, community organizer, community builder, and scholar-activist whose work addresses interpersonal, corporate, and state violence targeting disabled people at intersections of race, class, gender, sexuality, faith, language, and nation. Ly Xīnzhèn is Assistant Teaching Professor of Disability Studies at Georgetown University, founding Executive Director of The Autistic People of Color Fund, and is creating Disability Justice Wisdom Tarot. They serve as past president and treasurer of the Disability Rights Bar Association, Disability Justice Committee representative to the National Lawyers Guild board, and a gubernatorial appointee to the Maryland Commission on LGBTQIA+ Affairs. Outside of their scholarship and organizing work, Ly Xīnzhèn is director of public policy at National Disability Institute.

Nan Bernstein Ratner is a professor in the Department of Hearing and Speech Sciences at the University of Maryland, College Park. Her primary areas of research are fluency development and disorder, psycholinguistics, and child language development. Nan is a co-founder and co-manager of FluencyBank [23], a corpus of annotated disfluent speech that has been highly influential in both fluency research and speech AI development. She is a longstanding organizer of sessions for the Annual Meeting of the American Association for the Advancement of Science and the recipient of multiple NIH Conference grants (R15).

Allison Koenecke is an Assistant Professor of Information Science at Cornell University. Her research on algorithmic fairness includes auditing disparities in ASR system performance, especially among underrepresented speech populations including African American English, d/Dhh, and aphasic speakers.

Karen Nakamura is a cultural and visual anthropologist and the Haas Distinguished Chair in Disability Studies at the University of California Berkeley, where she also runs the UC Berkeley Disability Lab. Her research is on disability, sexuality, and other minority social movements in contemporary Japan and the United States. Her books, films, and articles have resulted in numerous prizes including the John Whitney Hall Book Prize, the SVA Short Film Award, and David Plath Media Award. Nakamura is currently collaborating on research involving the impact of artificial intelligence / machine learning (AI/ML) on disability communities.

Shaomei Wu is a person who stutters and the founder and CEO of AImpower.org – a tech non-profit that researches and co-develops disfluency friendly AI technologies *for*, *with*, and *by* the stuttering community. Her research explores fairness and ethical issues in mainstream technologies including social media, AI, and videoconferencing. She has co-organized workshops at CHI and CSCW, and has served as a workshop co-chair for ICWSM.

4 Publishing plans

Participants will be encouraged to publish their workshop papers on arXiv. Accepted papers will be distributed and archived at workshop proceedings with an index submission, as well as on the workshop website.

5 Accessibility

Given the workshop's focus on the accessibility and equity in speech AI, we will welcome and expect attendees with speech diversities, such as DHH individuals [8], AAC users and people who stutter. Sign language interpreters and live transcriptions will be requested and arranged to facilitate the full participation of all attendees in our workshop. Leading up to the conference, we will solicit accessibility requests from participants to assess their specific needs and preferences. At the beginning of both sessions, we will review community guidelines on respectful communication.

6 Offline Materials

The workshop keynote will be recorded and posted online. (Due to the discussion based nature of the other workshop activities, they will not be recorded.) Accepted workshop papers will be posted on our website for the public to engage with.

7 Activities

The first session of the workshop will foster productive engagement. A keynote address will introduce the central themes and lay the foundation for our discussions. Following the keynote, a poster session featuring participants' accepted papers will provide an opportunity for interaction and exchange, setting the stage for collaboration.

The second session will build upon the connections made with structured group discussions. These discussions will be centered upon our two primary research themes. To aid the productivity of these discussions, participants will be sorted into groups based on a pre-survey to ensure those with common interests and diverse backgrounds are able to collaborate. Walk-in participants will be sorted based on an expedited version of this survey. In their groups, participants will identify a scribe and be given a set of specific questions to address. After these discussions, there will be a shareout from each group and closing remarks.

Table 1: Workshop Timetable (180 minutes)

Minutes	Activity
30 min.	**Invited Keynote:** Summarize current work in this domain, introduce research questions, establish community guidelines for respectful and productive discussion.
60 min.	**Poster Session:** Participants will present their accepted workshop papers in a poster format.
—	Break
30 min.	**Guided Discussion #1:** In breakout groups, participants will discuss the first research objective: *Connecting existing metrics to downstream user impact.*
30 min.	**Guided Discussion #2:** In breakout groups, participants will discuss the second research objective: *Developing new holistic metrics for different stages of the speech AI lifecycle..*
20 min.	**Shareout**
10 min.	**Closing Remarks**

8 Post-Workshop Plans

We will close the workshop with a discussion that reflects our collective learnings and build a plan to create momentum in fair and inclusive speech AI for all.

The participants will be encouraged to publish their optional papers on arXiv and the workshop's website for archival and visibility purposes. The participants will have the opportunity to co-author research papers built on the insights from the guided discussions and form cross-sector, interdisciplinary working groups around the most important research questions proposed at the workshop.

The research papers will be published on the workshop's website and submitted to CHI or ACM Interactions. The working groups will be self-organized through mailing list and regular virtual meet-ups to drive forward the research agenda developed at the workshop. Collectively, the work groups will also carry out follow-up activities such as identifying funding and collaboration opportunities (e.g. joint grant proposal, shared projects), planning for additional workshop on fair and accessible speech AI at CHI or other venues (e.g. NeurIPS, CSCW, FAccT), organizing special issues in relevant journals, and establishing the infrastructure for industry-academic-community partnership.

9 Call for Participation

This double session CHI 2026 workshop invites researchers, practitioners, policy makers, and community members interested in fair and inclusive speech AI technologies to explore new opportunities for speech AI measurement for people with speech diversities and disabilities.

Participants will span experts from industry and academic, centering the experience of users with speech diversities to better understand the gap between existing measurement of speech AI systems, and what we collectively ideate as a more representative set of standards for inclusive speech AI. These standards, which more realistically map onto diverse user experiences, will lead to the co-creation of more comprehensive benchmarking and measurement strategies for speech AI developers to study and implement.

Participants are encouraged, but not required, to submit position papers, tech demo papers, technical research papers, policy papers, as well as experience reports and briefs from other fields, under the general theme of understanding and improving speech AI technologies for and with people with speech diversities. Accepted papers will be presented during the workshop as a poster presentation, and published on the workshop's website.

Interested participants can apply by completing a digital form to share their background and motivation for participation, with an option to attach a paper of up to ten pages (plus references) in the ACM single-column format. The workshop organizing team will select participants based on the alignment between the participants' backgrounds and the workshop themes, while striving to assemble a diverse group across a range of disciplines, methodologies, and seniorities.

10 Note about Past Workshops

This workshop is a sequel to last year's workshop, Speech AI for All: Promoting Accessibility, Fairness, Inclusivity, and Equity [32].

11 Expected Size of Attendance

Based on paper submissions and attendance of last year's workshop, we expect an interest of 30-50 participants, which we can cap based on room capacity.

Acknowledgments

Some contents of this paper were developed under a grant from the U.S. National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number 90REGE0027). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents of this paper do not necessarily represent the policy of NIDILRR, ACL, HHS, and you should not assume endorsement by the Federal Government. Additionally, this material is based upon work supported by the U.S. National Science Foundation under Grants 2348221, 2440601, 2427710, and 2447704.

References

[1] Siddhant Arora, Kai-Wei Chang, Chung-Ming Chien, Yifan Peng, Haibin Wu, Yossi Adi, Emmanuel Dupoux, Hung-Yi Lee, Karen Livescu, and Shinji Watanabe. 2025. On the landscape of spoken language models: A comprehensive survey. *arXiv preprint arXiv:2504.08528* (2025).

[2] Garance Burke and Hilke Schellmann. 2024. Researchers say an AI-powered transcription tool used in hospitals invents things no one ever said. *AP News* (2024).

[3] Geoffrey A. Coalson, Alexus Crawford, Shanley B. Treleaven, Courtney T. Byrd, Lauren Davis, Lillian Dang, Jillian Edgerly, and Alison Turk. 2022. Microaggression and the adult stuttering experience. *Journal of Communication Disorders* 95 (2022), 106180. doi:10.1016/j.jcomdis.2021.106180

[4] Wenqian Cui, Dianzhi Yu, Xiaoqi Jiao, Ziqiao Meng, Guangyan Zhang, Qichao Wang, Yiwen Guo, and Irwin King. 2024. Recent advances in speech language models: A survey. *arXiv preprint arXiv:2410.03751* (2024).

[5] Andreea Danielescu, Sharone A Horowit-Hendler, Alexandria Pabst, Kenneth Michael Stewart, Eric M Gallo, and Matthew Peter Aylett. 2023. Creating inclusive voices for the 21st century: A non-binary text-to-speech for conversational assistants. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[6] Abraham Glasser. 2019. Automatic Speech Recognition Services: Deaf and Hard-of-Hearing Usability. In *Extended Abstracts of the 2019 CHI Conference on Human*

Factors in Computing Systems (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3290607.3308461

[7] Abraham Glasser, Kesavan Kushalnagar, and Raja Kushalnagar. 2017. Deaf, hard of hearing, and hearing perspectives on using automatic speech recognition in conversation. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 427–432.

[8] Abraham Glasser, Kesavan Kushalnagar, and Raja Kushalnagar. 2017. Deaf, Hard of Hearing, and Hearing Perspectives on Using Automatic Speech Recognition in Conversation. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 427–432. doi:10.1145/3132525.3134781

[9] Abraham T Glasser, Kesavan R Kushalnagar, and Raja S Kushalnagar. 2017. Feasibility of using automatic speech recognition with voices of deaf and hard-of-hearing individuals. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 373–374.

[10] Seung-Eun Kim, Bronya R. Chernyak, Olga Seleznova, Joseph Keshet, Matthew Goldrick, and Ann R. Bradlow. 2024. Automatic recognition of second language speech-in-noise. *JASA Express Letters* 4, 2 (02 2024), 025204. doi:10.1121/10.0024877 arXiv:https://pubs.aip.org/asa/jel/article-pdf/doi/10.1121/10.0024877/19669029/025204_1_10.0024877.pdf

[11] Allison Koenecke, Anna Seo Gyeong Choi, Katelyn X Mei, Hilke Schellmann, and Mona Sloane. 2024. Careless Whisper: Speech-to-Text Hallucination Harms. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1672–1681.

[12] 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.

[13] Raja S Kushalnagar and Christian Vogler. 2020. Teleconference accessibility and guidelines for deaf and hard of hearing users. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 1–6.

[14] Colin Lea, Zifang Huang, Jaya Narain, Lauren Tooley, Dianna Yee, Dung Tien Tran, Panayiotis Georgiou, Jeffrey P Bigham, and Leah Findlater. 2023. From User Perceptions to Technical Improvement: Enabling People Who Stutter to Better Use Speech Recognition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 361, 16 pages. doi:10.1145/3544548.3581224

[15] Qisheng Li and Shaomei Wu. 2024. Towards Fair and Inclusive Speech Recognition for Stuttering: Community-led Chinese Stuttered Speech Dataset Creation and Benchmarking. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 331, 9 pages. doi:10.1145/3613905.3650950

[16] Yi-Cheng Lin, Wei-Chih Chen, and Hung-yi Lee. 2024. Spoken stereoset: on evaluating social bias toward speaker in speech large language models. In *2024 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 871–878.

[17] Yi-Cheng Lin, Tzu-Quan Lin, Chih-Kai Yang, Ke-Han Lu, Wei-Chih Chen, Chun-Yi Kuan, and Hung-yi Lee. 2024. Listen and speak fairly: a study on semantic gender bias in speech integrated large language models. In *2024 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 439–446.

[18] 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* 4 (2021), 725911.

[19] Shira Michel, Sufi Kaur, Sarah Elizabeth Gillespie, Jeffrey Gleason, Christo Wilson, and Avijit Ghosh. 2025. "It's not a representation of me": Examining Accent Bias and Digital Exclusion in Synthetic AI Voice Services. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. 228–245.

[20] American Civil Liberties Union of Colorado. 2025. Complaint of Discrimination against Intuit, Inc. and HireVue, Inc. Complaint filed with the Colorado Civil Rights Division and U.S. Equal Employment Opportunity Commission. https://assets.aclu.org/live/uploads/2025/03/Redacted-HireVue_Intuit-Complaint-of-Discrimination_Redacted.pdf Filed March 19, 2025.

[21] Alisha Pradhan, Sheena Erete, Shaan Chopra, Pooja Upadhyay, Oluwaseun Sule, and Amanda Lazar. 2025. 'No, not that voice again!': Engaging Older Adults in Design of Anthropomorphic Voice Assistants. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW141 (May 2025), 30 pages. doi:10.1145/3711039

[22] Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–27.

[23] Nan Bernstein Ratner and Brian MacWhinney. 2018. Fluency Bank: A new resource for fluency research and practice. *Journal of fluency disorders* 56 (2018), 69–80.

[24] Vaibhav Srivastav, Somshubra Majumdar, Nithin Koluguri, Adel Moumen, Sanchit Gandhi, et al. 2023. Open Automatic Speech Recognition Leaderboard. https://huggingface.co/spaces/hf-audio/open_asr_leaderboard.

[25] Maria Teleki, Xiangjue Dong, Haoran Liu, and James Caverlee. 2025. Masculine Defaults via Gendered Discourse in Podcasts and Large Language Models. In *ICWSM 2025*.

[26] Ravichander Vipperla, Steve Renals, and Joe Frankel. 2010. Ageing Voices: The Effect of Changes in Voice Parameters on ASR Performance. *EURASIP Journal on Audio, Speech, and Music Processing* (2010). https://asmp-eurasipjournals.springeropen.com/articles/10.1155/2010/525783

[27] Tobias Weinberg, Claire O'Connor, Ricardo E Gonzalez Penuela, Stephanie Valencia, and Thijs Roumen. 2025. One Does Not Simply'Mm-hmm': Exploring Backchanneling in the AAC Micro-Culture. *arXiv preprint arXiv:2506.17890* (2025).

[28] Tobias Weinberg, Ricardo E Gonzalez Penuela, Stephanie Valencia, and Thijs Roumen. 2025. I, Robot? Socio-Technical Implications of Ultra-Personalized AI-Powered AAC; an Autoethnographic Account. *arXiv preprint arXiv:2509.13671* (2025).

[29] Tobias M Weinberg, Kowe Kadoma, Ricardo E Gonzalez Penuela, Stephanie Valencia, and Thijs Roumen. 2025. Why So Serious? Exploring Timely Humorous Comments in AAC Through AI-Powered Interfaces. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.

[30] Kimi Wenzel, Nitya Devireddy, Cam Davison, and Geoff Kaufman. 2023. Can voice assistants be microaggressors? Cross-race psychological responses to failures of automatic speech recognition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[31] Kimi Wenzel and Geoff Kaufman. 2024. Designing for Harm Reduction: Communication Repair for Multicultural Users' Voice Interactions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[32] Shaomei Wu, Kimi Wenzel, Jingjin Li, Qisheng Li, Alisha Pradhan, Raja Kushalnagar, Colin Lea, Allison Koenecke, Christian Vogler, Mark Hasegawa-Johnson, et al. 2025. Speech AI for All: Promoting Accessibility, Fairness, Inclusivity, and Equity. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–6.

[33] Chih-Kai Yang, Neo S Ho, and Hung-yi Lee. 2025. Towards holistic evaluation of large audio-language models: A comprehensive survey. *arXiv preprint arXiv:2505.15957* (2025).