Speech AI for All: Designing Speech Technologies for Interactive Communication

KIMI V. WENZEL, Carnegie Mellon University, USA
MARIA TELEKI, Texas A&M University, USA
DYLAN GAINES, Kennesaw State University, USA
SERAY IBRAHIM, King's College London, UK
JEFF HIGGINBOTHAM, University at Buffalo, USA
ANNA SEO GYEONG CHOI, New Jersey Institute of Technology, USA
KHIET P. TRUONG, Radboud University, The Netherlands
ABRAHAM GLASSER, Gallaudet University, USA
CHRISTIAN VOGLER, Gallaudet University, USA
RAJA KUSHALNAGAR, Gallaudet University, USA
SHAOMEI WU, AImpower.org, USA
NICHOLAS CUMMINS, King's College London, UK
THIJS ROUMEN, Cornell Tech, USA
ALLISON KOENECKE, Cornell Tech, USA

Many speech technologies are designed to support a transmissive "sender-receiver" model of communication. However, this model fails to capture many of the collaborative nuances that form a social interaction between conversational participants. This workshop aims to improve our understanding of how speech AI systems can be designed to accommodate more collaborative models of social interaction. This workshop unites participants from academia, industry, and non-profit organizations to discuss how we can work together towards this common goal. We will encourage cross-discipline collaboration through mini keynotes and interactive fishbowl activities centered on both social theory and hardware/technology. Following these structured discussions, we will host a speech AI poster session. The primary outcome from this workshop will be a recurring annual digest on speech AI research that will support the long-term tracking of progress in the field on topics such as communication modeling.

CCS Concepts: • **Human-centered computing** → **User models**; *Natural language interfaces*; **Interaction design**; Accessibility technologies.

Authors' Contact Information: Kimi V. Wenzel, Carnegie Mellon University, Pittsburgh, PA, USA, kwenzel@andrew.cmu.edu; Maria Teleki, Texas A&M University, College Station, TX, USA, mariateleki@tamu.edu; Dylan Gaines, Kennesaw State University, Marietta, GA, USA, dgaine20@kennesaw.edu; Seray Ibrahim, King's College London, London, UK, seray.ibrahim@kcl.ac.uk; Jeff Higginbotham, University at Buffalo, Buffalo, NY, USA, cdsjeff@buffalo.edu; Anna Seo Gyeong Choi, New Jersey Institute of Technology, Newark, NJ, USA, seo.choi@njit.edu; Khiet P. Truong, Radboud University, Nijmegen, The Netherlands, khiet.truong@ru.nl; Abraham Glasser, Gallaudet University, Washington, D.C., USA, abraham.glasser@gallaudet.edu; Christian Vogler, christian.vogler@gallaudet.edu, Gallaudet University, Washington, D.C., USA; Raja Kushalnagar, Gallaudet University, Washington, D.C., USA, raja.kushalnagar@gallaudet.edu; Shaomei Wu, AImpower.org, Menlo Park, USA; Nicholas Cummins, King's College London, London, UK; Thijs Roumen, Cornell Tech, New York City, USA, thijs.roumen@cornell.edu; Allison Koenecke, Cornell Tech, New York City, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
Manuscript submitted to ACM

Additional Key Words and Phrases: speech ai, speech technologies, social interaction, communication model

ACM Reference Format:
Kimi V. Wenzel, Maria Teleki, Dylan Gaines, Seray Ibrahim, Jeff Higginbotham, Anna Seo Gyeong Choi, Khiet P. Truong, Abraham Glasser, Christian Vogler, Raja Kushalnagar, Shaomei Wu, Nicholas Cummins, Thijs Roumen, and Allison Koenecke. 2026. Speech AI for All: Designing Speech Technologies for Interactive Communication. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Motivation

Speech AI is increasingly used to facilitate communication, both between multiple people and between people and technology. Yet, the models of social interaction that make up its design do not always reflect the interactional nature of communication [3, 6, 9, 11]. That interaction includes the co-construction of meaning, the situatedness of talk, and the contingent, moment-by-moment actions through which parties build shared understanding [4, 7, 12]. However, many designs—including augmentative and alternative communication (AAC) devices—draw on a conventional sender-receiver model that treats communication as information transmitted by one person and received by another. This model of social interaction overlooks how people reach mutual understanding in real time through collaboration, speech, gesture, gaze, and the ongoing work of checking and repairing understanding [1, 8, 16]. That narrow view has consequences beyond individual technologies: it shapes what speech AI is imagined to be for, and whose communicative practices it is built to support. For example, it is beyond the scope of the sender-receiver model in AAC to deal with issues of communication interdependence and authorship which are at the heart of the social interaction model [5, 7].

This pattern extends across application areas. For example, in speech AI for AAC, speech-rate efficiency is frequently prioritized as the target for improving existing devices, sidelining opportunities to design for broader properties of social interaction [5, 14, 16], including the wider interactional ecology, the dynamic and shifting roles of multiple parties in a conversation (human and AI alike), and how those parties work together across speech and other communicative resources to build shared understanding [10]. A similar reduction is present in how speech AI models the speaker: user models often assume a single, stable identity, flattening variation in accent, dialect, disfluency, or communicative style, and treating deviation from a normative profile as error to be corrected rather than as legitimate communicative diversity [2, 15, 17, 20]. Finally, in the case of everyday conversational agents, the same sender-receiver logic underlies the prevailing single-turn, command-response design of voice assistants. These systems often struggle to accommodate repair, overlap, or the incremental sequencing that ordinary talk depends on, leaving little room for the assistant to participate as a genuine co-constructing party in interaction.

A richer interaction model also creates practical hardware tensions. Capturing multiple speakers, sensing embodied actions, and responding in time may call for additional microphones, cameras, displays, and processing capacity, with consequences for portability, cost, and power use. Where the AI components run matters too; for a person using AAC, reliance on remote processing services can expose sensitive conversational data and make communication dependent on connectivity and service availability. These choices shape what the system can perceive, what partners can see, and whether support is available when it is needed. It also invites to be creative about new opportunities in hardware design to support expressing diversity or expressiveness through the very design of the tools used to support speech.

Taken together, these cases point to a broader question the workshop takes up:

> *What would it mean to design speech AI from a richer model of social interaction from the outset, rather than retrofitting interactional nuance onto systems built only for transmission?*

The **first iteration** of this workshop [19] convened researchers, practitioners, policymakers, and community advocates into a cross-sector coalition for fair and accessible speech AI, identifying key gaps in designing accessible speech AI for people with speech diversities. The **second iteration** [18] addressed the top-priority gap our community identified: measurement. It produced a shared understanding of measurement taxonomies' value and limitations, and a cross-disciplinary position paper [13] critiquing three problems: speech AI's incomplete models of communication and identity, and its use of the wrong metrics. The **current workshop** iteration builds on these identified challenges, addressing the next priority gap: designing speech technologies for interactive communication. Our goal for this workshop is to improve our community's collective understanding and approaches towards designing speech AI that acknowledge more expansive, interactive models of communication.

2 Organizers

Kimi Wenzel is a PhD Candidate at the Human-Computer Interaction Institute at Carnegie Mellon University. Her research focuses on psychosocial harm mitigation in language technologies. She has previously organized workshops at both CHI and FAccT.

Maria Teleki is a PhD Candidate in Computer Science at Texas A&M University, advised by James Caverlee. Her research develops methods, benchmarks, and audits for *Socially Robust AI*, with a special focus on speech disfluencies: testing whether speech AI systems are robust to how people naturally speak. Her work has appeared at INTERSPEECH, ICASSP, ACL, and FAccT, and she was a Student Researcher at Google DeepMind in 2026.

Dylan Gaines is an Assistant Professor of Computer Science at Kennesaw State University. His research combines innovation in natural language processing and human-computer interaction to develop assistive communication technologies. His work has appeared at venues such as CHI, IUI, ASSETS, PETRA, and EMNLP. He has served on the program committee for ACM ASSETS and on the organizing committee for ACM MobileHCI.

Seray Ibrahim is an Assistant Professor of Applied Digital Health at King's College London, with clinical training in Speech and Language Therapy. Her research explores the design and use of digital technology for supporting social interaction, mental wellbeing, and communication disability. She has served on program committees for SIGCHI conferences (CHI, IDC) and organised SIGCHI workshops.

Jeff Higginbotham is Professor in the Department of Communicative Disorders and Sciences at the University at Buffalo and Director of the Communication and Assistive Device Laboratory (CADL). His research examines, through microanalysis of naturally occurring interaction, how people with disabilities who use augmentative and alternative communication (AAC) coordinate conversation with their partners. His work also focuses on the development and evaluation of AI-enhanced conversational AAC technologies designed to better support participation, interaction, and communication agency.

Anna Seo Gyeong Choi is an Assistant Professor of AI for Business at Martin Tuchman School of Management, New Jersey Institute of Technology. Her research focuses on algorithmic fairness and accountability in speech AI, including auditing speech recognition systems for clinical speech and legal/institutional contexts.

Khiet Truong is an associate professor at the Centre for Language Studies, Radboud University and at the Human Media Interaction group, University of Twente in the Netherlands. Her research interests lie in modeling paralinguistic aspects that are indicative of speaker characteristics and conversation dynamics among speakers for application domains such as human-robot interaction, multimedia retrieval, and digital health. She has served numerous chairing roles for conferences such as Interspeech and ACM ICMI, and editing and reviewing roles as well for IEEE Transactions on Affective Computing, IEEE ACII, and ACM/IEEE HRI.

Abraham Glasser is a Deaf person who speaks English with an accent that derives from his deafness. He 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 code-switches among American Sign Language, spoken English and spoken German. He speaks with an accent induced by his deafness. While it substantially changed over time as a result of receiving cochlear implants, it persists in creating problems with speech technologies, which manifest differently depending on which language he uses. He is the co-director of the AHCP program at Gallaudet University and also directs the Technology Access Program there, which is a research group focused on making technology-mediated communication accessible for DHH people. He is involved in projects and that include deaf friendly text-to-speech design, voice cloning of deaf-accented speakers, AI-assisted text communications, sign language communications in videoconferencing, and next-generation assistive listening technologies. He has co-organized conferences, including ASSETS 2021 and 2022.

Raja Kushalnagar is a Deaf person who communicates in American Sign Language and spoken English, and speaks with an accent that derives from his deafness. He is a Professor and Director of the Information Technology Program at Gallaudet University, where he also directs the AI, Accessibility, and Sign Language (AIASL) Center and co-directs the Accessible Human-Centered Computing and Policy (AHCP) M.S. and Ph.D. programs. He also serves as Special Assistant to the Provost for CBTL and AI Integration. His research focuses on accessible computing and sign language AI, including ASL recognition, captioning, and signing avatar technologies, as well as HCI more broadly. He is PI or co-PI on several NSF-funded projects, including the Universal AI NSF Research Traineeship (NRT), a Gallaudet–RIT institutional partnership pilot, and the REU AICT undergraduate research site. Dr. Kushalnagar has published at venues including ACM ASSETS, CHI, TACCESS, W4A, and SIGCSE.

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 people with verbal diversities. Her recent work focuses on fairness and inclusivity of speech and community technologies, data ethics, and grassroots AI movement. She has co-organized workshops at CHI and CSCW, and has served as a workshop co-chair for ICWSM.

Nicholas (Nick) Cummins is a Senior Lecturer in Speech Analysis and Responsible AI in Health at King's College London. He leads the Voice and Speech Processing for Health Group, whose research focuses on integrating speech technologies into clinical research and practice. His work combines speech processing, machine learning, and patient involvement to develop reliable speech-based tools that support health innovation. He holds numerous leadership and organisational roles, including serving on the ISCA Diversity Panel, co-leading the Speech and Language Processing for Health track at Interspeech 2026, and serving as an editor for Computer Speech & Language.

Thijs Roumen is Assistant Professor of Information Science at Cornell Tech. He studies Computational Fabrication and Accessibility within the Matter of Techlab. He is PI of the Cornell Initiative on Accessible Technology and AI (CATAI) and organizer of the first two annual Accessibility Summits at Cornell Tech in 2025 and 2026. A core focus of some of his work has been on speech accessibility, specifically on how to make AAC interfaces more expressive.

Allison Koenecke is an Assistant Professor of Information Science at Cornell Tech. Her research on algorithmic fairness includes auditing disparities in ASR system performance, especially among underrepresented speech populations including African American English, d/Dhh, and aphasia speakers. She has previously organized workshops at CHI.

3 Activities

The workshop will consist of two 90-minute sessions with a short break between. They may be shortened to 75 minutes per the conference's final schedule. The sessions are described in Table 1.

Session	Activity		Description
Session 1	Keynote	(50 min)	The workshop will begin with two mini keynotes from experts taking different approaches to communication modeling: One focused on social theory, and one focused on hardware and technology. The keynotes are intended to set the tone of the session, and inform participants from different disciplines about one another's work for more cross-disciplinary action.
	Fishbowl	(40 min)	Following the keynotes, participants will have the opportunity to engage in a fishbowl discussion about communication modeling. Discussion prompts will be sourced from three areas: (a) solicited from participants ahead of the workshop, (b) gathered through a Google form during the keynote talks, (c) identified by workshop chairs in advance. Chairs will be arranged in a semi-circle to foster the discussion, and individual conversation turns will be limited to maximum 1 minute to allow time for everyone to participate. Participants can opt-in to join as a discussant or choose to simply listen to the live discussion.
Session 2	Poster Session (90 min)		The second half of the workshop will be dedicated to a poster session. Posters will be submitted in advance as part of participants' workshop application. A carousel slidedeck will play at the front of the room during the poster session briefly advertising each poster with text, such that participants may appropriately identify posters of interest.

Table 1. Descriptions of the sessions and activities planned for the workshop.

4 Call for Participation

Join us for the third annual Speech AI for All workshop at CHI! This year's workshop will contain three primary activities: keynote sessions, a fishbowl activity, and a poster session. We welcome participants in-person from all sectors interested in improving communication modeling in speech AI, including (but not limited to) researchers, engineers, designers, social psychologists, speech therapists, and non-profit organizers. Interested participants should complete the interest form on our website, answering questions about their motivation for joining the workshop and optionally submitting a poster. Accepted posters will be published on the workshop webpage. For the workshop outcome, we will seed a lightweight annual digest that surfaces key statistics and achievements. Concretely, we envision compiling notable audits and fairness-oriented benchmarking results published on speech AI over the preceding year into a short, accessible summary (tentatively, a *Speech AI Index*). Issuing this as a recurring outcome, rather than a one-off position paper, will let the workshop community track shifts in the field's concerns and progress across successive iterations and will lower the barrier to participation for members who cannot commit to co-authoring a full paper. We expect approximately 30 workshop participants to join us. http://speechai4all.org/

References

[1] Michael Argyle, Mark Cook, and Duncan Cramer. 1994. Gaze and mutual gaze. *The British Journal of Psychiatry* 165, 6 (1994), 848–850.

[2] Anna Seo Gyeong Choi, Maria Teleki, Miguel del Rio, James Caverlee, Corey Miller, and Allison Koenecke. 2026. SpeechSpectrum: A Linguistic Fidelity Spectrum for Accountable Speech-to-Text. https://mariateleki.github.io/pdf/2026SpeechSpectrum.pdf

[3] Joel E. Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for voice interface design. In *Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3342775.3342788

[4] Charles Goodwin and John Heritage. 1990. Conversation Analysis. *Annual Review of Anthropology* 19, 1 (1990), 283–307. _eprint: https://doi.org/10.1146/annurev.an.19.100190.001435. doi:10.1146/annurev.an.19.100190.001435

[5] Seray B Ibrahim, Tom Griffiths, Michael Clarke, Simon Judge, Petr Slovak, Graham Pullin, and Jeff Higginbotham. 2026. Before the Technological Fix: Scoping AI and AAC for Social Futures. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3772363.3798436

[6] Korbinian Kuhn, https://orcid.org/0009-0005-1296-4987, View Profile, Verena Kersken, https://orcid.org/0009-0000-5007-6327, View Profile, Benedikt Reuter, https://orcid.org/0009-0004-4618-376X, View Profile, Niklas Egger, https://orcid.org/0009-0003-9866-2175, View Profile, Gottfried Zimmermann, https://orcid.org/0000-0002-3129-1897, and View Profile. 2024. Measuring the Accuracy of Automatic Speech Recognition Solutions. *ACM Transactions on Accessible Computing* 16, 4 (Jan. 2024), 1–23. doi:10.1145/3636513

[7] Stephen C. Levinson. 2025. *The Interaction Engine: Language in Social Life and Human Evolution* (1 ed.). Cambridge University Press. doi:10.1017/9781009570343

[8] Maja Sigurd Pilesjö. 2014. Creating meaning through the coordination of gaze direction and arm/hand movement. *Journal of Interactional Research in Communication Disorders* 5, 1 (2014), 63–96. Num Pages: 34 Place: Lafayette, United Kingdom. http://www.proquest.com/docview/1541479629/abstract/339842D8BCE2446EPQ/1

[9] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 640:1–640:12. doi:10.1145/3173574.3174214

[10] Francesco Possemato, Antara Satchidanand, and Jeff Higginbotham. 2026. Repair Under Constraint: Interaction Ecologies and Composition Delay in AAC-Mediated Conversation. https://osf.io/preprints/socarxiv/kwsze_v1/

[11] Sam O'Connor Russell, Iona Gessinger, Anna Krason, Gabriella Vigliocco, and Naomi Harte. 2024. What automatic speech recognition can and cannot do for conversational speech transcription. *Research Methods in Applied Linguistics* 3, 3 (Dec. 2024), 100163. doi:10.1016/j.rmal.2024.100163

[12] Lucille Alice Suchman. 1987. *Plans and situated actions: The problem of human-machine communication.* Cambridge university press.

[13] Maria Teleki, Kimi Wenzel, Anna Seo Gyeong Choi, Tobias Weinberg, Shree Harsha Bokkahalli Satish, Stephanny Sanchez, Belu Ticona, Ariadna Sanchez, Yash Sonkar, Aarti Mathur, Christoph Minixhofer, Abraham Glasser, Raja Kushalnagar, James Caverlee, Minha Lee, Shaomei Wu, Alyssa Hillary Zisk, Éva Székely, Dylan Gaines, Angelika Seeschaaf Veres, Seray Ibrahim, Nicholas Cummins, and Allison Koenecke. 2026. A Cross Community Agenda for Speech AI. arXiv:2609.13168 [cs.HC] https://arxiv.org/abs/2609.13168

[14] Stephanie Valencia, Jessica Huynh, Emma Y Jiang, Yufei Wu, Teresa Wan, Zixuan Zheng, Henny Admoni, Jeffrey P Bigham, and Amy Pavel. 2024. COMPA: Using Conversation Context to Achieve Common Ground in AAC. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. doi:10.1145/3613904.3642762

[15] Tobias Weinberg, Aaleyah Lewis, Ricardo E. Gonzalez Penuela, Weicong Hong, Jennifer Mankoff, and Thijs Roumen. 2026. Me, Myself, and My Voice: Exploring Cultural and Linguistic Identity in AAC AI-generated Voices. arXiv:2605.24337 [cs.HC]. doi:10.48550/arXiv.2605.24337

[16] Tobias M 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. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3663547.3746381

[17] 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*. ACM, Honolulu HI USA, 1–17. doi:10.1145/3613904.3642900

[18] 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. Zhāngsūn Brown, Nan Bernstein Ratner, Allison Koenecke, Karen Nakamura, and Shaomei Wu. 2026. Speech AI for All: The What, How, and Who of Measurement. In *Proceedings of the Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3772363.3778768

[19] Shaomei Wu, Kimi Wenzel, Jingjin Li, Qisheng Li, Alisha Pradhan, Raja Kushalnagar, Colin Lea, Allison Koenecke, Christian Vogler, Mark Hasegawa-Johnson, Norman Makoto Su, and Nan Bernstein Ratner. 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 (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3706599.3706746

[20] Maike Züfle, Maria Teleki, Fabian Retkowski, Vilém Zouhar, Oliver Grabner, Alexander Waibel, James Caverlee, and Jan Niehues. 2026. The Role of Disfluencies in Speech Translation. arXiv:2608.02138 [cs.CL] https://arxiv.org/abs/2608.02138