Bias and Fairness in Conversational User Interfaces

Guy Laban
Department of Computer Science and Technology
University of Cambridge
Cambridge, United Kingdom
guy.laban@cl.cam.ac.uk

Julian Hough
School of Mathematics and Computer Science
Swansea University
Swansea, United Kingdom
julian.hough@swansea.ac.uk

Minha Lee
Industrial Design
Eindhoven University of Technology
Eindhoven, Netherlands
m.lee@tue.nl

Alva Markelius
Department of Computer Science and Technology
University of Cambridge
Cambridge, United Kingdom
ajkm4@cam.ac.uk

Mary Ellen Foster
School of Computing Science
University of Glasgow
Glasgow, United Kingdom
maryellen.foster@glasgow.ac.uk

Jane Stuart-Smith
School of Critical Studies
University of Glasgow
Glasgow, United Kingdom
jane.stuart-smith@glasgow.ac.uk

Muneeb Imtiaz Ahmad
School of Mathematics and Computer Science
Swansea University
Swansea, United Kingdom
m.i.ahmad@swansea.ac.uk

Abstract

Conversational User Interfaces (CUIs), including chatbots, virtual agents and social robots, are increasingly shaping how we communicate, seek support and access services. Yet, as these systems grow more sophisticated, concerns about bias and fairness in their design and deployment have become increasingly urgent. We propose a multidimensional approach to bias and fairness in CUIs that spans four interconnected themes: conceptual grounding, verbal communication, multimodal expression and interactional dynamics. Rather than framing bias merely as a technical flaw, we argue that it should be understood as a relational, interactional and design-based phenomenon. Accordingly, in this workshop, we aim to foster critical discussion around how CUIs encode social norms, perpetuate or mitigate exclusion, and shape perceptions of fairness through their language, embodiment and behaviour. By bringing together researchers, designers and policymakers, the workshop will explore pathways towards more equitable and transparent CUIs. The goal is to promote a relational understanding of fairness, one that centres user experience and social context, to guide future work in conversational AI.

CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**; **Natural language interfaces**; **Empirical studies in HCI**; • **Social and professional topics** → **User characteristics**.

CUI '25, Waterloo, ON, Canada

ACM ISBN 979-8-4007-1527-3/25/07
https://doi.org/10.1145/3719160.3728629

Keywords

Conversational User Interfaces, Bias, Fairness, Ethics, Human–Agent Interaction, Multimodal Interaction, Large Language Models, Conversational AI

ACM Reference Format:
Guy Laban, Julian Hough, Minha Lee, Alva Markelius, Mary Ellen Foster, Jane Stuart-Smith, and Muneeb Imtiaz Ahmad. 2025. Bias and Fairness in Conversational User Interfaces. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25), July 08–10, 2025, Waterloo, ON, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3719160.3728629

Background

Conversational user interfaces (CUIs) are rapidly transforming the way we interact with technology, reshaping our everyday experiences through embodied (e.g., social robots and virtual humans) and disembodied (e.g., chatbots) conversational agents. They have become part of our everyday routines, popping up in fields as diverse as healthcare, education, finance, and entertainment [22]. At the heart of their success are rapid advances in language processing, voice recognition, and machine learning, which allow these systems to understand and respond to us with increasing accuracy and sophistication [71].

However, as CUIs become more common and more powerful, it's crucial to examine how bias and fairness come into play. When these systems are trained on large datasets, they can accidentally pick up on stereotypes or other forms of bias hidden in that data [26, 52]. Sometimes, the algorithms themselves amplify these biases [8], and in other cases, the ways people interact with CUIs might trigger different responses for different groups [50]. The consequences can be especially significant when CUIs are used in critical areas like healthcare, where biased advice could affect someone's well-being

[5, 40, 42, 52], or education, where slanted interactions could impact students' access to learning resources [16, 29].

This workshop will offer a deep dive into understanding and addressing bias and fairness in CUIs. One of our primary goals is to create a clear narrative that connects broad ideas about ethics and fair design in CUI studies and human-agent interaction (HAI) research at large with both the ways we can measure language-based performance and the non-verbal signals CUIs pick up or project. Examining verbal cues—such as specific words, phrases, or emotional tones—can reveal patterns of bias in a system's responses. But we also want to look at non-verbal elements that might shape the verbal projection of the CUI, since these subtler behaviours can reinforce unfair treatment.

During the workshop, we'll unpack big questions about what "*bias*" and "*fairness*" really mean in the context of conversational interfaces, and how we can test for and measure them. We will also consider how these technologies shape user experiences and how we can utilise the data collected from these systems more fairly.

Another aspect we'll explore is how policies and regulations could guide the development of CUIs. By bringing researchers, industry professionals, and policymakers together, we hope to spark meaningful conversations about the best ways to keep bias in check, preserve fairness, and maintain accountability. The workshop will include talks, discussions, and group exercises that spotlight cutting-edge research, identify gaps where more work is needed, and propose practical steps toward making CUIs more equitable.

Ultimately, as CUIs continue to influence our personal lives and society at large, we want to ensure they do so in a way that respects and protects all users. This workshop is a chance to share knowledge, foster collaboration, and drive thoughtful innovation, with the aim of creating conversational systems that reflect our highest values of inclusion, fairness, and social responsibility.

To unpack these challenges systematically, both here and in the workshop, we structure our exploration around four interrelated themes. Each theme addresses a distinct yet interconnected dimension of how bias and fairness operate in CUIs: from conceptual foundations to linguistic and multimodal expressions, and finally to the dynamics of interaction. Together, these themes offer a comprehensive lens for understanding bias not merely as a static design flaw, but as an evolving, socially embedded process.

1 Theme 1: Conceptual Grounding

At the heart of any meaningful discussion about bias and fairness in CUIs lies the need for **conceptual clarity**. Bias and fairness are not monolithic properties that can be universally measured or removed; rather, they are deeply contextual, interpretive, and relational constructs that emerge through the complex interaction between technology design, language, and social norms. Under this theme, we seek to lay the groundwork for a theoretical approach to understanding how bias is embedded into, and enacted through, conversational systems—not merely at the level of outputs, but through their underlying design assumptions and communicative structure.

1.1 The Need for Conceptual Grounding

Recent attention to fairness in conversational AI often focuses on measurable harms [67]: uneven performance across demographic groups, harmful stereotypes reproduced in responses, or inequities in accessibility [2, 28, 65]. While such concerns are urgent, this *"detection paradigm"* risks treating bias as a deviation from a presumed neutrality, rather than asking what neutrality itself presupposes. CUIs are never neutral, they are shaped by a chain of design decisions that reflect implicit norms and value systems, whether acknowledged or not. Therefore, we argue that bias in CUIs is not simply a technical issue but also a design-related phenomenon. In this view, fairness is not just about correcting for representational gaps or algorithmic skew, but about interrogating how systems are designed to "*understand*" and "*respond*" in the first place.

1.2 Intentional and Unintentional Biases in Design

Biases in CUIs can result from both intentional design decisions (e.g., selecting a default voice or interaction script) [59] and unintentional omissions (e.g., failing to account for users whose language patterns deviate from the training data norm) [23]. But even these categories, intentional vs. unintentional, might oversimplify it. Often, the problem lies not in the presence of malice or neglect, but in the absence of theoretical sensitivity to cultural and social variation in the context of CUI. The question is not whether a design is biased or not, but what kinds of users and interactions it is designed to accommodate, and which it excludes or misinterprets.

We therefore propose a situated perspective on fairness in CUIs, one that recognizes fairness as context-dependent, user-relative, and mediated by design choices. In practice, this means attending to how fairness is perceived, negotiated, and contested within specific human–CUI interactions.

1.3 Toward a Shared Vocabulary

Moreover, we call for the establishment of a shared conceptual vocabulary that can ground future empirical work. Such a vocabulary would include distinctions between:

- Design-time bias vs. runtime bias.
- Interactional bias vs. distributional bias.
- Normative fairness goals (e.g., equity, inclusion, transparency) vs. operational fairness metrics.

By creating a common theoretical grounding, we aim to facilitate cross-disciplinary dialogue and position future studies of CUI bias.

This conceptual framing sets the stage for the following themes, which extend this theory into verbal, nonverbal, and interactional dimensions. Together, they propose an integrated framework for understanding how CUIs can perpetuate or challenge social asymmetries, via *what they say* (theme 2), *how they say it* (theme 3), and *how they shape the broader dynamics of interaction* (theme 4).

2 Theme 2: Verbal Communication

Language is the primary medium through which CUIs engage with users [53], and thus, it is one of the most consequential spaces where bias is encoded, perceived, and negotiated [55]. While Theme 1 positioned bias and fairness as socio-technical constructs rooted in conceptual and design decisions, Theme 2 extends this theoretical lens to the **verbal** layer of interaction, asking how linguistic choices

in CUIs, whether scripted, generated, or emergent, shape normative assumptions about users, roles, and relations.

2.1 Language is Never Neutral

CUIs are often evaluated on metrics such as clarity, efficiency, or fluency [14]. However, such metrics abstract language from its pragmatic and social dimensions [68]. Every utterance made by a CUI carries with it an implied stance toward the user: about what is appropriate, who holds authority, and what kind of interaction is expected. Bias at this level is not only about offensive or discriminatory content, but about who gets to speak, how they are spoken to, and how their input is interpreted or constrained.

2.2 Bias in Dialogue Structure

Bias can also manifest in the structure and flow of dialogue. CUIs are designed with assumptions about turn-taking [61, 62], topic relevance [12, 41, 43], and conversational goals [44, 48, 49], assumptions that may not be shared across all users or contexts. When a system prematurely redirects a user, fails to follow up on emotionally charged content, or overly constrains user input through narrow question and answer formats, it may unintentionally reproduce patterns of marginalisation or epistemic injustice, such as denying certain users the opportunity to express themselves meaningfully or to be understood on their own terms.

2.3 Computational Perspectives on Verbal Bias and Fairness

Verbal bias in conversational systems is often conceptualized in terms of distributional disparities, stereotyped language associations, and differential system behaviour across user groups. Large language models (LLMs) that power CUIs are typically trained on massive corpora that reflect the biases, both overt and subtle, present in those data sources [55]. As a result, CUIs may reproduce or amplify stereotypes depending on perceived user characteristics (e.g., [1]), or systematically misunderstand certain dialects and language varieties [57].

Efforts to quantify verbal bias often involve benchmarking models using carefully curated test sets. These may include stereotype sentence completions, sentiment evaluations across demographic prompts, or turn-level analyses of linguistic style and tone in generated responses [9]. While these computational techniques can reveal important patterns, it is also essential to move beyond the model's output and consider the interactional dynamics of the conversation. A CUI may avoid overtly biased phrasing yet still subtly frame one group's concerns as more valid than another's through word choice, topic control, or failure to follow up. Thus, fairness in verbal CUIs cannot be reduced to lexical parity alone; it requires attention to the pragmatics of conversation, including *what is said*, *how it is said*, *what remains unsaid*, and how these patterns intersect with identity and power.

2.4 Toward Fair Linguistic Design

To conceptualize **verbal** fairness in CUIs, we argue that we should move beyond word-level content moderation or keyword filtering. Fairness in this domain should entail:

- Supporting linguistic pluralism, by enabling diverse dialects, registers, and speech styles.
- Acknowledging interactional asymmetries, such as power dynamics embedded in directive vs. invitational utterances.
- Designing for conversational responsiveness, including sensitivity to when a user is signalling uncertainty, distress, or social discomfort.
- Incorporating technical methods for fairness auditing in generation pipelines.

Importantly, fairness in verbal interaction is not only about the user-facing output, but also about the system's interpretive logic. It is therefore equally important to consider how the system parses input, frames intent, and selects among possible responses. This highlights the need for further discussion around open, explainable, and human-moderated technical developments (e.g., [18–21]; rather than relying solely on closed systems), as well as the advancement of theoretical models that treat CUIs as co-participants in dialogue, rather than merely reactive tools.

3 Theme 3: Multimodal Expression

While CUIs are often understood through their verbal outputs, an increasing number of systems, especially those with physical (e.g., social robots) or virtual embodiments (e.g., avatars), communicate in **multimodal** ways [42, 64]. Facial expressions, gaze direction, gestures, posture, spatial positioning, and paralinguistic features like prosody or intonation are integral to how CUIs interact with users [60, 64, 66]. These nonverbal signals are not just supplementary; they actively shape the meaning of the verbal message [27], the relational framing of the interaction, and accordingly they could also shape the user's interpretation of system intent and fairness.

In this theme we aim to explore bias and fairness in CUIs as multimodal phenomena, acknowledging that nonverbal cues often reinforce (or sometimes contradict) the social roles, affective meanings, and interactional expectations constructed by language [60, 66].

3.1 Embodiment as Normative Representation

The physical or visual embodiment of a CUI, whether as a humanoid robot, animated avatar, or abstract interface, always carries implicit messages about identity, authority, and social role [34, 35, 43, 56]. The design of facial features, skin tone, gender presentation, voice, and body movement encode normative assumptions about what a service provider, teacher, assistant, or companion "should" look and act like [6, 15, 70]. These choices can reproduce stereotypes or exclusionary ideals, especially when they default to Western, white, able-bodied, or cisgender representations. Rather than treating embodiment merely as an engineering or aesthetic decision, we argue that embodiment must be seen also as a cultural and political act. Bias can emerge not only from who is represented, but from who is systematically not represented, or from how bodies are simplified, stereotyped, or decontextualized.

3.2 Multimodal Cues and Perceived Fairness

Bias can also arise through incongruities between verbal and nonverbal signals [54]. These mismatches can erode trust, especially in sensitive settings such as counselling [58], education [29], or

healthcare [47], where subtle affective signals carry significant interpersonal meaning. More subtly, multimodal cues can differentially affect user perception based on cultural expectations [51]. A direct gaze may be seen as confident in some contexts and aggressive in others; expressive hand gestures may signal warmth in one culture and impropriety in another. A system that treats nonverbal communication as universal risks creating interactions that feel alienating or unjust to certain users.

3.3 Computational Perspectives on Nonverbal Bias and Fairness

In multimodal and embodied CUIs, bias often arises not only through language, but through visual and auditory channels: facial expressions, gaze patterns, gestures, tone of voice, and physical embodiment. From an AI perspective, these components are modelled through affective computing pipelines, including emotion recognition systems, gesture generation models, synthetic voices, that introduce new points at which bias can manifest [10]. Fairness and bias discrepancies originating from multimodal features can lead to the misattribution of emotional states, which in turn affects how a CUI perceives users and responds to them [1, 7].

Computational research in this space has developed metrics for performance disparity across demographic groups (see [10]). Researchers should not assume that there is a single correct interpretation of an expression, or that fairness means parity of system behaviour regardless of individual difference [32]. Such assumptions risk erasing the very diversity that fairness seeks to protect.

A more robust approach acknowledges that multimodal fairness is not about removing difference, but about supporting contextual intelligibility [33]. In other words, we should design systems that can adapt to, rather than normalize away, user variation (e.g., [13]). This includes training on expressive datasets that span identities and interactional styles, incorporating user control and moderation over nonverbal behaviours, and treating emotional and embodied responses as socially situated, not universal. From this perspective, fairness in multimodal CUIs becomes a matter of interpretive flexibility and responsive adaptation, rather than statistical balance alone.

3.4 Designing for Multimodal Fairness

The challenge, then, is not to erase embodiment or minimize nonverbal expressivity, but to design CUIs whose **multimodal behaviour** is adaptive, context-sensitive, and inclusive. This entails:

- Critically reflecting on representational diversity in embodiment, including whose identities are modelled or prioritized.
- Building systems that can engage in multimodal alignment—adjusting gaze, gesture, and tone to match the interactional and emotional context.
- Considering interactional asymmetries introduced by embodiment, especially when physical form implies authority, surveillance, or affective labour.
- Implementing fairness-aware multimodal models that evaluate and mitigate expressive and perceptual performance disparities.

Fairness in multimodal CUIs requires acknowledging that social meaning is co-constructed across channels [17]. Designers and researchers must attend not only to what the system says, but how it looks, sounds, moves, and "feels" to interact with. These embodied aspects of CUIs are not cosmetic, they are central to how fairness is interpreted and experienced.

4 Theme 4: Interactional Dynamics

Bias and fairness in CUIs do not reside solely in system outputs or datasets, they emerge through ongoing interactional processes. Even well-intentioned and technically sound systems can produce biased experiences when design assumptions misalign with users' interactional expectations, affective needs, or social positioning [29, 42]. In this final theme, we aim to focus on how fairness and bias are experienced, negotiated, navigated, and at times repaired within the dynamic flow of conversational exchanges.

4.1 Interaction as the Site of Social Meaning

CUIs are not static tools but co-participants in interaction [44, 63]. Their behaviour unfolds turn by turn, and users develop expectations about their responsiveness, consistency, tone, and adaptability. As in human-human interaction, trust, rapport, and perceived fairness are not established in a single utterance—they are built over time [42]. When a CUI interrupts, ignores, misreads, or talks over a user, the harm may not reside in any single response but in the pattern that emerges across multiple interactions [24, 30, 31]. Thus, bias is not only about content, it is also about timing, acknowledgment, sequencing, and responsiveness. A system that routinely offers more clarification to one type of user and not another, or that treats similar statements from different users unequally, may reproduce bias **interactionally**, even if its responses are lexically similar.

4.2 Perceived Fairness and Moral Attribution

Fairness in CUIs is not determined solely by what the system does, but also by how that behaviour is *perceived* by users. A CUI that produces equivalent responses across demographic groups may still be experienced as unfair if users feel dismissed, judged, or differently treated [37, 39]. Conversely, a system with minor output disparities may be interpreted as fair if users believe it is responsive, caring, or transparent in its actions [45]. These perceptions are shaped not only by the immediate interaction, but by users' prior experiences with technology, cultural expectations, and social identity [15]. Accordingly, users bring their own expectations, identities, and past experiences to the interaction, which in turn shape how fairness is inferred, even in the absence of any objectively biased behaviour.

This means that fairness is often perceived through more than just operational dynamics like timing, fluency, or tone. A CUI that offers neutral or supportive content may still be judged as biased if its message challenges the user's world-view, engages with morally sensitive topics, or evokes discomfort [39]. Conversely, a system may be seen as fair and trustworthy simply because it conforms to the user's social norms or validates their perspective, even if its responses are generically generated or behaviourally imprecise [38, 69]. This underscores the need to examine fairness and bias as projected qualities, not just measured ones.

In this view, CUIs become sites of moral projection. Users may anthropomorphize them, attribute intent to their responses, or evaluate their fairness based on how closely the system aligns with their expectations, regardless of the system's internal mechanics [36]. These projections are intensified in contexts involving sensitive or value-laden content, such as mental health support, political advice, or interpersonal feedback [3]. However, users do not always distinguish between the system and its creators. Biases perceived in CUIs are often seen as reflecting broader institutional, corporate, or societal values [25].

This highlights the importance of studying user perception alongside behavioural and algorithmic audits. Investigating how users interpret system behaviour, where they locate responsibility for fairness, and how these attributions evolve over time is essential for developing CUIs that are perceived by users as morally intelligible and trustworthy. Hence, perceived fairness is neither reducible to design nor to evaluation metrics, it is an emergent property of human interpretation, shaped by *what the system says*, *how it says it*, and *who is listening*.

4.3 Repair, Correction, and the Possibility of Fairness

Interaction is not only where bias manifests, it is also where fairness can be mitigated and repaired. CUIs that are able to recognize misunderstandings, solicit clarification, or reflect back uncertainty can create openings for users to reassert control, redirect the conversation, or resist problematic framings [4]. This underscores the importance of designing for interactional sensitivity: Can the CUI detect when a user seems dissatisfied, confused, or emotionally affected? Can it invite elaboration without imposing? Can it adapt over time to better fit the user's preferences, communication style, or context? These questions align with broader theoretical commitments to interactional justice, which emphasizes not only what is said, but how participants are treated in the process of speaking and being heard [52].

4.4 From Static Ethics to Relational Ethics

Frameworks for fairness in AI often focus on pre-deployment assessments [10, 28]. These approaches have been instrumental in raising awareness and accountability. However, in conversational settings, fairness often evolves through interaction and should be seen as relational, contingent, and emergent, something that unfolds in time, through exchange. Therefore, fairness must also account for small-scale, multimodal, and temporally situated dynamics that shape how engagement with a CUI is perceived, interpreted, and evaluated throughout an interaction [11]. This calls for extending the scope of bias and fairness research in CUIs from static ethics, which focus on outcomes and compliance, to relational ethics, which emphasise practices, encounters, and lived experiences.

5 Conclusion: Toward a Relational Theory of Bias and Fairness in CUIs

Together, these four themes present a theoretical framework for understanding bias and fairness in CUI as multidimensional, socially embedded, and interactionally emergent. Rather than isolating technical failures or linguistic anomalies, we call for a broader rethinking of CUIs as systems that shape and are shaped by social life. Bias should not be understood simply as an error to be fixed, but as a reflection of deeper assumptions about communication, identity, and social legitimacy. Similarly, fairness is not just a static goal to be checked off, but an ongoing design commitment, one that calls for attentiveness to diverse user perspectives, responsiveness to interactional dynamics, and reflexivity throughout the system's development and use.

By addressing these four themes, we aim to support a growing body of work that approaches CUIs not only as engineering challenges, but also as spaces of social interaction, ethical responsibility, and human-centred design.

6 Workshop Plan

The workshop aims to provide a venue for multidisciplinary CUI researchers interested in bias and fairness. It is organised by several members of the CUI research community who have previously organised workshops on similar topics for the CUI community (e.g., [46, 50]). The workshop will feature two keynote speakers, addressing both technical and interactional perspectives on bias and fairness in CUIs, while also offering insights into the ethical foundations of the topic.

The workshop will be publicised on social media, relevant mailing lists, and a dedicated website, which will provide all necessary information and the schedule. A call for papers will be issued, supported by programme committee members and the workshop organisers, who will peer review submissions. Submitters may also propose topics for inquiry and discussion in the groups brainstorming discussions, designed to facilitate dialogue before the workshop; these topics will be displayed on the workshop website ahead of the event.

Accepted papers will be presented orally during the workshop and disseminated via the website. Additionally, discussions from the groups brainstorming activity will be documented and presented in the 'Reflection' session before the workshop concludes, culminating in a collaborative theoretical paper that encapsulates the workshop's insights and will be submitted to a leading journal in our field.

6.1 Schedule

The half-day workshop will begin with an introduction, followed by the first keynote. This will be followed by a paper presentation session. After a coffee break, a second paper session will take place, leading into the second keynote. This keynote will transition into a brainstorming session, during which participants will be divided into groups for interactive discussions. The workshop will conclude with a collective reflection session moderated by the organisers.

6.2 Keynote Speakers

Micol Spitale is an Assistant Professor at the Department of Electronics, Information and Bioengineering at the Politecnico di Milano (Polimi), as well as a Visiting Affiliated Researcher at the University of Cambridge. Her research has been focused on the field of Social Robotics, exploring ways to develop robots that are socio-emotionally adaptive and provide 'coaching' to promote wellbeing,

Table 1: Tentative schedule

5 min	Introduction and Welcome
35 min	*Keynote 1*
45 min	*Paper Presentations 1*
15 min	Break and Networking
45 min	*Paper Presentations 2*
35 min	*Keynote 2*
40 min	Interactive group discussion and brainstorming activity
20 min	Reflections/Take-aways
5 min	Final Remarks and Conclusion

while being unbiased and fair. Previously, she was a Postdoctoral Researcher at the Affective Intelligence & Robotics Laboratory (AFAR Lab) of the University of Cambridge.

Dorian Peters is a Research Fellow at Imperial College London, an Associate Fellow at the Leverhulme Centre for the Future of Intelligence (LCFI) at the University of Cambridge and a Senior Research Associate at the Intellectual Forum, Jesus College Cambridge. She is a human-computer interaction researcher who specialises in design for digital health and wellbeing, human autonomy, learning, and translation of research to practice. She has worked in participatory digital health across age groups with communities in the UK, Australia and South America. Her current work explores how conversational AI might be leveraged to tackle health disparities in the majority world in ways that acknowledge model bias and power dynamics. Her books include Positive Computing: Technology for Wellbeing and Human Potential (MIT Press), and Interface Design for Learning (Pearson).

6.3 Organisers

Guy Laban is a Research Associate in the Department of Computer Science & Technology at the University of Cambridge, and an incoming Senior Lecturer (Assistant Professor) at the department of Industrial Engineering and Management of Ben-Gurion University of the Negev. His research explores how people share emotions with artificial agents and how these interactions shape social dynamics and well-being. He investigates the semantic and social implications of conversational AI, including how large language models interpret and generate emotional expressions and the biases that influence these interactions. He was awarded the International Society for Research on Emotion (ISRE) 2024 Dissertation Award (2nd Runner-Up), served as publicity chair for ACII 2024, associate editor for ICSR 2024, and has co-organized workshops on conversational AI, including Robo-Identity 2 (HRI 2022) and Ethics of CUIs (CHI 2022).

Julian Hough is an Associate Professor in Computer Science at Swansea University. His research focuses on applying Artificial Intelligence and Natural Language Processing techniques to interaction data and within interactive systems to improve the quality of HAI. He has won an outstanding paper award at COLING 2020 and was appointed a Turing Fellow at the Alan Turing Institute (2019-2021). Since 2019, he is co-chair of the SemDial conference series and has served in roles in international conference venues such as General Chair (SemDial), Area Chair (InterSpeech, EACL) and Poster chair (HAI), and leads the UK EPSRC-funded FLUIDITY project.

Minha Lee is an Assistant Professor at the Eindhoven University of Technology in the Department of Industrial Design, with a background in philosophy, digital arts, and HCI. Her research concerns morally relevant interactions with various agents like robots or chatbots. Her work explores how we can explore our moral self-identity through conversations with digital entities, e.g., via acting compassionately towards a chatbot. She co-leads the steering committee of the ACM CUI conference series after serving as one of the general chairs of the CUI 2023 conference.

Alva Markelius is a PhD candidate and Cambridge Trust Scholar at the Department of Computer Science & Technology at the University of Cambridge. She is also recipient of Top 100 Brilliant Women in AI Ethics™ award 2024 and committee chair of the Centre for Human Inspired AI Early Career Community. Her main research interests are the ethics of social AI and robotics, specifically in the intersection of embodied robotics and large language models. She also holds an MSt degree in AI Ethics & Society at the Leverhulme Centre for the Future of Intelligence at the University of Cambridge.

Mary Ellen Foster is a Senior Lecturer in the School of Computing Science at the University of Glasgow. Her primary research interests are human-robot interaction, social robotics, and embodied conversational agents. Her recent work focusses on consulting the widest possible range of stakeholders when designing and deploying social robots, to ensure that the system takes into account the needs of all potential users and operators. She has previously co-organised a workshop series on Public-Space Social Robotics.

Jane Stuart-Smith is Professor of Phonetics and Sociolinguistics, Co-Director of the Glasgow University Laboratory of Phonetics at the University of Glasgow. Her main research interests are on the interfaces between language and society, especially relating to speech of stigmatised varieties, e.g. Working-class vernaculars and ethnolects (phonetics, sociophonetics, sociolinguistics). She has a long-term interest in speech and language accommodation beyond typical face-to-face interactions, especially without typical human interaction, e.g. experiencing broadcast media, Human-Robot-Interaction.

Muneeb Imtiaz Ahmad is a Senior Lecturer in Computer Science at Swansea University's (SU) Computational Foundry, and his research is situated within the Human-Centred Computing domain and is at the Intersection of Human-Robot Interaction (HRI), Intelligent Robotics and Human-Centred Artificial Intelligence (AI). He has recently received funding from Morgan advances studies institute to establish best practices for designing healthcare social robots with culturally sensitive AI, focusing on reducing communication bias and promoting efficient communication and engagement. He acted as a general co-chair at HAI 2024 and is acting as a publicity chair at the ICMI 2025 conference.

References

[1] Nida Itrat Abbasi, Fethiye Irmak Dogan, Guy Laban, Joanna Anderson, Tamsin Ford, Peter B. Jones, and Hatice Gunes. 2025. Robot-Led Vision Language Model Wellbeing Assessment of Children. (4 2025). https://arxiv.org/abs/2504.02765v1

[2] Andrea Aler Tubella, Flavia Barsotti, Rüya Gökhan Koçer, and Julian Alfredo Mendez. 2022. Ethical implications of fairness interventions: what might be hidden behind engineering choices? *Ethics and Information Technology* 24, 1 (3 2022). doi:10.1007/S10676-022-09636-Z

[3] Theo Araujo, Natali Helberger, Sanne Kruikemeier, and Claes H. de Vreese. 2020. In AI we trust? Perceptions about automated decision-making by artificial

intelligence. *AI and Society* 35, 3 (9 2020), 611–623. doi:10.1007/S00146-019-00931-W/TABLES/4

[4] Minja Axelsson, Micol Spitale, and Hatice Gunes. 2024. "Oh, Sorry, I Think I Interrupted You": Designing Repair Strategies for Robotic Longitudinal Well-being Coaching. *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (3 2024), 13–22. doi:10.1145/3610977.3634948

[5] Minja Axelsson, Micol Spitale, and Hatice Gunes. 2024. Robots as Mental Well-being Coaches: Design and Ethical Recommendations. *ACM Transactions on Human-Robot Interaction* 13, 2 (6 2024). doi:10.1145/3643457/ASSET/EBE92B19-CC4D-4F8C-88F8-8BABC7886953/ASSETS/GRAPHIC/THRI-2022-0075-PAGE50.JPG

[6] Philip Baillargeon, Jina Yoon, and Amy Zhang. 2024. Who Puts the "Social" in "Social Computing"?: Using A Neurodiversity Framing to Review Social Computing Research. (10 2024). https://arxiv.org/abs/2410.15527v1

[7] Brandon M. Booth, Louis Hickman, Shree Krishna Subburaj, Louis Tay, Sang Eun Woo, and Sidney K. D'Mello. 2021. Bias and Fairness in Multimodal Machine Learning: A Case Study of Automated Video Interviews. *ICMI 2021 - Proceedings of the 2021 International Conference on Multimodal Interaction* (10 2021), 268–277. doi:10.1145/3462244.3479897

[8] Begum Celiktutan, Romain Cadario, and Carey K. Morewedge. 2024. People see more of their biases in algorithms. *Proceedings of the National Academy of Sciences of the United States of America* 121, 16 (4 2024), e2317602121. doi:10.1073/PNAS.2317602121/SUPPL{_}FILE/PNAS.2317602121.SAPP.PDF

[9] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. 2024. A Survey on Evaluation of Large Language Models. *ACM Transactions on Intelligent Systems and Technology* 15, 3 (3 2024), 39. doi:10.1145/3641289

[10] Jiaee Cheong, Sinan Kalkan, and Hatice Gunes. 2021. The Hitchhiker's Guide to Bias and Fairness in Facial Affective Signal Processing: Overview and techniques. *IEEE Signal Processing Magazine* 38, 6 (11 2021), 39–49. doi:10.1109/MSP.2021.3106619

[11] Jiaee Cheong, Micol Spitale, and Hatice Gunes. 2023. "It's not Fair!" - Fairness for a Small Dataset of Multi-modal Dyadic Mental Well-being Coaching. *2023 11th International Conference on Affective Computing and Intelligent Interaction, ACII 2023* (2023). doi:10.1109/ACII59096.2023.10388118

[12] Sophie Chiang, Guy Laban, Emily S. Cross, and Hatice Gunes. 2025. Comparing Self-Disclosure Themes and Semantics to a Human, a Robot, and a Disembodied Agent. (4 2025). https://arxiv.org/abs/2504.06374v1

[13] Nikhil Churamani, Ozgur Kara, and Hatice Gunes. 2023. Domain-Incremental Continual Learning for Mitigating Bias in Facial Expression and Action Unit Recognition. *IEEE Transactions on Affective Computing* 14, 4 (10 2023), 3191–3206. doi:10.1109/TAFFC.2022.3181033

[14] Riccardo Coppola and Luca Ardito. 2021. Quality Assessment Methods for Textual Conversational Interfaces: A Multivocal Literature Review. *Information 2021, Vol. 12, Page 437* 12, 11 (10 2021), 437. doi:10.3390/INFO12110437

[15] Emily S. Cross and Richard Ramsey. 2021. Mind Meets Machine: Towards a Cognitive Science of Human–Machine Interactions. *Trends in Cognitive Sciences* 25, 3 (3 2021), 200–212. doi:10.1016/J.TICS.2020.11.009

[16] Deema Dakakni and Nehme Safa. 2023. Artificial intelligence in the L2 classroom: Implications and challenges on ethics and equity in higher education: A 21st century Pandora's box. *Computers and Education: Artificial Intelligence* 5 (1 2023), 100179. doi:10.1016/J.CAEAI.2023.100179

[17] Annamaria Silvana de Rosa. 2024. Social representations and multi-channel communication on crucial societal issues through the modelling approach. *Discover Global Society* 2, 1 (12 2024), 1–22. doi:10.1007/S44282-024-00126-9/FIGURES/1

[18] Fethiye Irmak Doğan, Gaspar I Melsión, and Iolanda Leite. 2023. Leveraging explainability for understanding object descriptions in ambiguous 3D environments. *Frontiers in Robotics and AI* 9 (2023), 937772.

[19] Fethiye Irmak Dogan, Umut Ozyurt, Gizem Cinar, and Hatice Gunes. 2025. GRACE: Generating Socially Appropriate Robot Actions Leveraging LLMs and Human Explanations. In *IEEE International Conference on Robotics and Automation (ICRA)*. https://arxiv.org/abs/2409.16879v2

[20] Fethiye Irmak Dogan, Maithili Patel, Weiyu Liu, Iolanda Leite, and Sonia Chernova. 2024. A Model-Agnostic Approach for Semantically Driven Disambiguation in Human-Robot Interaction. *arXiv preprint arXiv:2409.17004* (2024).

[21] Fethiye Irmak Doğan, Ilaria Torre, and Iolanda Leite. 2022. Asking Follow-Up Clarifications to Resolve Ambiguities in Human-Robot Conversation. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 461–469. doi:10.1109/HRI53351.2022.9889368

[22] Asbjørn Følstad, Theo Araujo, Effie Lai-Chong Law, Petter Bae Brandtzaeg, Symeon Papadopoulos, Lea Reis, Marcos Baez, Guy Laban, Patrick McAllister, Carolin Ischen, Rebecca Wald, Fabio Catania, Raphael Meyer von Wolff, Sebastian Hobert, and Ewa Luger. 2021. Future directions for chatbot research: an interdisciplinary research agenda. *Computing 2021* 103 (2021), 2915–2942. doi:10.1007/S00607-021-01016-7

[23] Isabel O. Gallegos, Ryan A. Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, and Nesreen K. Ahmed. 2024. Bias and Fairness in Large Language Models: A Survey. *Computational Linguistics* 50, 3 (9 2024), 1097–1179. doi:10.1162/COLI{_}A{_}00524

[24] Romi Gideoni, Shanee Honig, and Tal Oron-Gilad. 2024. Is It Personal? The Impact of Personally Relevant Robotic Failures (PeRFs) on Humans' Trust, Likeability, and Willingness to Use the Robot. *International Journal of Social Robotics* 16, 6 (6 2024), 1049–1067. doi:10.1007/S12369-022-00912-Y/FIGURES/5

[25] Hangzhi Guo, Pranav Narayanan Venkit, Eunchae Jang, Mukund Srinath, Wenbo Zhang, Bonam Mingole, Vipul Gupta, Kush R. Varshney, S. Shyam Sundar, and Amulya Yadav. 2024. Hey GPT, Can You be More Racist? Analysis from Crowdsourced Attempts to Elicit Biased Content from Generative AI. (10 2024). https://arxiv.org/abs/2410.15467v1

[26] Yufei Guo, Muzhe Guo, Juntao Su, Zhou Yang, Mengqiu Zhu, Hongfei Li, Mengyang Qiu, and Shuo Shuo Liu. 2024. Bias in Large Language Models: Origin, Evaluation, and Mitigation. (11 2024). http://arxiv.org/abs/2411.10915

[27] Judith A. Hall, Terrence G. Horgan, and Nora A. Murphy. 2019. Nonverbal Communication. *Annual Review of Psychology* 70, Volume 70, 2019 (1 2019), 271–294. doi:10.1146/ANNUREV-PSYCH-010418-103145/CITE/REFWORKS

[28] Matthew G. Hanna, Liron Pantanowitz, Brian Jackson, Octavia Palmer, Shyam Visweswaran, Joshua Pantanowitz, Mustafa Deebajah, and Hooman H. Rashidi. 2025. Ethical and Bias Considerations in Artificial Intelligence/Machine Learning. *Modern Pathology* 38, 3 (3 2025), 100686. doi:10.1016/J.MODPAT.2024.100686

[29] Anna Henschel, Guy Laban, and Emily S Cross. 2021. What Makes a Robot Social? A Review of Social Robots from Science Fiction to a Home or Hospital Near You. *Current Robotics Reports* 2 (2021), 9–19. doi:10.1007/s43154-020-00035-0

[30] Shanee Honig and Tal Oron-Gilad. 2018. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in Psychology* 9, JUN (6 2018), 351644. doi:10.3389/FPSYG.2018.00861/BIBTEX

[31] Shanee Honig and Tal Oron-Gilad. 2021. Expect the Unexpected: Leveraging the Human-Robot Ecosystem to Handle Unexpected Robot Failures. *Frontiers in Robotics and AI* 8 (7 2021), 656385. doi:10.3389/FROBT.2021.656385/XML/NLM

[32] Rachael E. Jack, Oliver G.B. Garrod, Hui Yu, Roberto Caldara, and Philippe G. Schyns. 2012. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America* 109, 19 (5 2012), 7241–7244. doi:10.1073/PNAS.1200155109/SUPPL{_}FILE/SM02.AVI

[33] Tonni Das Jui and Pablo Rivas. 2024. Fairness issues, current approaches, and challenges in machine learning models. *International Journal of Machine Learning and Cybernetics 2024 15:8* 15, 8 (1 2024), 3095–3125. doi:10.1007/S13042-023-02083-2

[34] Dimosthenis Kontogiorgos, Andre Pereira, Olle Andersson, Marco Koivisto, Elena Gonzalez Rabal, Ville Vartiainen, and Joakim Gustafson. 2019. The effects of anthropomorphism and non-verbal social behaviour in virtual assistants. *IVA 2019 - Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents* (7 2019), 133–140. doi:10.1145/3308532.3329466

[35] Dimosthenis Kontogiorgos, Andre Pereira, and Joakim Gustafson. 2021. Grounding behaviours with conversational interfaces: effects of embodiment and failures. *Journal on Multimodal User Interfaces* 15, 2 (6 2021), 239–254. doi:10.1007/S12193-021-00366-Y/FIGURES/18

[36] Guy Laban. 2021. Perceptions of Anthropomorphism in a Chatbot Dialogue: The Role of Animacy and Intelligence. In *Proceedings of the 9th International Conference on Human-Agent Interaction*. ACM, New York, NY, USA, 305–310. doi:10.1145/3472307.3484686

[37] Guy Laban and Theo Araujo. 2020. The Effect of Personalization Techniques in Users' Perceptions of Conversational Recommender System. In *IVA '20: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA '20)*. ACM, Virtual Event, Glasgow, Scotland UK, 3. doi:10.1145/3383652.3423890

[38] Guy Laban and Theo Araujo. 2020. Working Together with Conversational Agents: The Relationship of Perceived Cooperation with Service Performance Evaluations. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Asbjørn Følstad, Theo Araujo, Symeon Papadopoulos, Effie L Law, Ole-Christoffer Granmo, Ewa Luger, and Petter B Brandtzaeg (Eds.). Chatbot Research and Design, Vol. 11970. Springer International Publishing, Cham, 215–228. doi:10.1007/978-3-030-39540-7{_}15

[39] Guy Laban and Theo Araujo. 2022. Don't Take it Personally: Resistance to Individually Targeted Recommendations from Conversational Recommender Agents. In *HAI 2022 - Proceedings of the 10th Conference on Human-Agent Interaction*. Association for Computing Machinery, New York, NY, USA, 57–66. doi:10.1145/3527188.3561929

[40] Guy Laban, Ziv Ben-Zion, and Emily S. Cross. 2022. Social Robots for Supporting Post-traumatic Stress Disorder Diagnosis and Treatment. *Frontiers in psychiatry* 12 (2022). doi:10.3389/FPSYT.2021.752874

[41] Guy Laban, Sophie Chiang, and Hatice Gunes. 2025. What People Share With a Robot When Feeling Lonely and Stressed and How It Helps Over Time. (4 2025). https://arxiv.org/abs/2504.02991v1

[42] Guy Laban and Emily S. Cross. 2024. Sharing our Emotions with Robots: Why do we do it and how does it make us feel? *IEEE Transactions on Affective Computing* (2024), 1–18. doi:10.1109/TAFFC.2024.3470984

[43] Guy Laban, Jean-Noël George, Val Morrison, and Emily S. Cross. 2021. Tell me more! Assessing interactions with social robots from speech. *Paladyn, Journal of*

Behavioral Robotics 12, 1 (2021), 136–159. doi:10.1515/pjbr-2021-0011

[44] Guy Laban, Arvid Kappas, Val Morrison, and Emily S Cross. 2024. Building Long-Term Human–Robot Relationships: Examining Disclosure, Perception and Well-Being Across Time. *International Journal of Social Robotics* 16, 5 (2024), 1–27. doi:10.1007/s12369-023-01076-z

[45] Guy Laban, Tomer Laban, and Hatice Gunes. 2024. LEXI: Large Language Models Experimentation Interface. In *Proceedings of the 12th International Conference on Human-Agent Interaction*. ACM, New York, NY, USA, 250–259. doi:10.1145/3687272.3688296

[46] Guy Laban, Sebastien Le Maguer, Minha Lee, Dimosthenis Kontogiorgos, Samantha Reig, Ilaria Torre, Ravi Tejwani, Matthew J Dennis, and Andre Pereira. 2022. Robo-Identity: Exploring Artificial Identity and Emotion via Speech Interactions. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, 1265–1268. doi:10.1109/HRI53351.2022.9889649

[47] Guy Laban, Val Morrison, and Emily Cross. 2024. Social Robots for Health Psychology: A New Frontier for Improving Human Health and Well-Being. *European Health Psychologist* 23, 1 (2 2024), 1095–1102. https://www.ehps.net/ehp/index.php/contents/article/view/3442

[48] Guy Laban, Val Morrison, Arvid Kappas, and Emily S. Cross. 2025. Coping with Emotional Distress via Self-Disclosure to Robots: An Intervention with Caregivers. *International Journal of Social Robotics* (2025). doi:10.1007/s12369-024-01207-0

[49] Guy Laban, Julie Wang, and Hatice Gunes. 2025. A Robot-Led Intervention for Emotion Regulation: From Expression to Reappraisal. (3 2025). https://arxiv.org/abs/2503.18243v1

[50] Minha Lee, Jaisie Sin, Guy Laban, Matthias Kraus, Leigh Clark, Martin Porcheron, Benjamin R Cowan, Asbjørn Følstad, Cosmin Munteanu, and Heloisa Candello. 2022. Ethics of Conversational User Interfaces. 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, 1–7. doi:10.1145/3491101.3503699

[51] Velvetina Lim, Maki Rooksby, and Emily S. Cross. 2021. Social Robots on a Global Stage: Establishing a Role for Culture During Human–Robot Interaction. *International Journal of Social Robotics* 13, 6 (9 2021), 1307–1333. doi:10.1007/S12369-020-00710-4/FIGURES/1

[52] Alva Markelius. 2025. *An Empirical Design Justice Approach to Identifying Ethical Considerations in the Intersection of Large Language Models and Social Robotics*. Oxford University Press. doi:10.1093/9780198945215.003.0013

[53] Michael F. McTear. 2002. Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys (CSUR)* 34, 1 (3 2002), 90–169. doi:10.1145/505282.505285

[54] Shu Morioka, Michihiro Osumi, Mayu Shiotani, Satoshi Nobusako, Hiroshi Maeoka, Yohei Okada, Makoto Hiyamizu, and Atsushi Matsuo. 2016. Incongruence between Verbal and Non-Verbal Information Enhances the Late Positive Potential. *PLoS ONE* 11, 10 (10 2016), e0164633. doi:10.1371/JOURNAL.PONE.0164633

[55] Roberto Navigli, Simone Conia, and Björn Ross. 2023. Biases in Large Language Models: Origins, Inventory, and Discussion. *J. Data and Information Quality* 15, 2, Article 10 (June 2023), 21 pages. doi:10.1145/3597307

[56] Mark Paterson. 2024. Why robot embodiment matters: questions of disability, race and intersectionality in the design of social robots. *Medical Humanities* 50, 4 (1 2024), e013028. doi:10.1136/MEDHUM-2024-013028

[57] Remi Poivet, Mélanie Lopez Malet, Catherine Pelachaud, and Malika Auvray. 2023. The influence of conversational agents' role and communication style on user experience. *Frontiers in Psychology* 14 (2023), 1266186. doi:10.3389/FPSYG.2023.1266186

[58] Torr Polakow, Guy Laban, Andrei Teodorescu, Jerome R. Busemeyer, and Goren Gordon. 2022. Social robot advisors: effects of robot judgmental fallacies and context. *Intelligent Service Robotics* (8 2022), 1–17. doi:10.1007/s11370-022-00438-2

[59] Malak Sadek, Rafael A. Calvo, and Céline Mougenot. 2024. The Value-Sensitive Conversational Agent Co-Design Framework. *International Journal of Human–Computer Interaction* (11 2024). doi:10.1080/10447318.2024.2426737

[60] Katie Seaborn, Norihisa P. Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. 2021. Voice in Human–Agent Interaction: A Survey. *ACM Comput. Surv.* 54, 4, Article 81 (May 2021), 43 pages. doi:10.1145/3386867

[61] Gabriel Skantze. 2021. Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Computer Speech & Language* 67 (5 2021), 101178. doi:10.1016/J.CSL.2020.101178

[62] Gabriel Skantze and Bahar Irfan. 2025. Applying General Turn-taking Models to Conversational Human-Robot Interaction. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction* (Melbourne, Australia) *(HRI '25)*. IEEE Press, 859–868.

[63] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2022. A longitudinal study of human–chatbot relationships. *International Journal of Human-Computer Studies* 168 (2022), 102903. doi:10.1016/j.ijhcs.2022.102903

[64] Micol Spitale, Minja Axelsson, Sooyeon Jeong, Paige Tuttosı, Caitlin A. Stamatis, Guy Laban, Angelica Lim, and Hatice Gune. 2024. Past, Present, and Future: A Survey of The Evolution of Affective Robotics For Well-being. (7 2024). doi:10.48550/arXiv.2407.02957

[65] Justyna Stypinska. 2022. AI ageism: a critical roadmap for studying age discrimination and exclusion in digitalized societies. *Ai & Society* 38, 2 (4 2022), 665. doi:10.1007/S00146-022-01553-5

[66] Jacqueline Urakami and Katie Seaborn. 2023. Nonverbal Cues in Human–Robot Interaction: A Communication Studies Perspective. *ACM Transactions on Human-Robot Interaction* 12, 2 (3 2023), 1–21. doi:10.1145/3570169

[67] Angelina Wang, Aaron Hertzmann, and Olga Russakovsky. 2024. Benchmark suites instead of leaderboards for evaluating AI fairness. *Patterns* 5, 11 (11 2024), 101080. doi:10.1016/J.PATTER.2024.101080

[68] Nigel Ward and Divette Marco. 2024. A Collection of Pragmatic-Similarity Judgments over Spoken Dialog Utterances. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue (Eds.). ELRA and ICCL, Torino, Italia, 154–163. https://aclanthology.org/2024.lrec-main.14/

[69] Joel Wester, Sander de Jong, Henning Pohl, and Niels van Berkel. 2024. Exploring people's perceptions of LLM-generated advice. *Computers in Human Behavior: Artificial Humans* 2, 2 (8 2024), 100072. doi:10.1016/J.CHBAH.2024.100072

[70] Katie Winkle, Donald McMillan, Maria Arnelid, Madeline Balaam, Katherine Harrison, Ericka Johnson, and Iolanda Leite. 2023. Feminist human-robot interaction: disentangling power, principles and practice for beter, more ethical HRI. *ACM/IEEE International Conference on Human-Robot Interaction* (3 2023), 72–82. doi:10.1145/3568162.3576973

[71] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wensen Cheng, Qi Zhang, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huang, and Tao Gui. 2023. The Rise and Potential of Large Language Model Based Agents: A Survey. (9 2023). https://arxiv.org/abs/2309.07864v3