The Art of Talking Machines: A Comprehensive Literature Review of Conversational User Interfaces

Mohammad (Matt) Namvarpour
matt.namvarpour@drexel.edu
Drexel University
Philadelphia, Pennsylvania, USA

Afsaneh Razi
afsaneh.razi@drexel.edu
Drexel University
Philadelphia, Pennsylvania, U.S.A

Abstract

Conversational User Interfaces (CUIs) enable human-like interactions via voice, text, and multimodal communication, driven by natural language processing and machine learning. Prior literature reviews have primarily focused on specific application domains or design aspects, lacking an integrated, multi-dimensional analysis. This study addresses this gap by providing a structured framework synthesizing CUI research into interface design, system development, and ethical considerations. Our analysis highlights advancements in CUI design, such as dialogue structure, multimodal interactions, and adaptability. It also reveals persistent challenges, including bias in persona design, trust calibration, and data privacy. System development benefits from improvements in NLP, conversation memory, and multilingual capabilities. Ethical considerations, including social bias, user autonomy, and transparency, remain central to discussions on responsible CUI design. By analyzing existing research, we identify key gaps and suggest future directions, including multilingual and culturally adaptive CUIs, privacy-preserving AI techniques, and enhanced reasoning mechanisms for context-aware interactions.

CCS Concepts

• **Human-centered computing → Interaction design**.

Keywords

Conversational User Interfaces, CUI, Literature Review, Survey

ACM Reference Format:
Mohammad (Matt) Namvarpour and Afsaneh Razi. 2025. The Art of Talking Machines: A Comprehensive Literature Review of Conversational User Interfaces . In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25), July 8–10, 2025, Waterloo, ON, Canada.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3719160.3736621

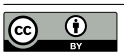
1 Introduction

The origins of CUIs can be traced back to the development of ELIZA in the mid-1960s, a groundbreaking chatbot designed to simulate therapeutic conversations through simple pattern matching [145]. Since then, CUIs have evolved dramatically, with significant milestones such as Siri's launch in 2011 and the widespread adoption of Alexa in smart homes, showcasing the growing integration of these technologies into daily life and being integral to industries like healthcare and education [104, 118, 138]. This evolution has been fueled by advancements in natural language processing (NLP), machine learning (ML), and neural networks, enabling more sophisticated and human-like interactions [5, 104, 124]. Modern CUIs now support multimodal interactions and dynamically adapt to user preferences to enhance accessibility and user experience [104, 138, 142].

Previous literature reviews on Conversational User Interfaces (CUIs) have primarily examined their applications within specific domains such as healthcare [26, 124, 138], education [113], and task automation [142]. Other studies have explored interface design and user experience [65, 104, 140], as well as ethical concerns such as privacy, security, and bias [28, 61, 67, 121, 127, 128]. While these reviews provide valuable insights, they often focus on isolated aspects of CUIs rather than offering a broad, integrated perspective. In contrast, our work builds on these existing reviews by conducting a more comprehensive analysis that provides a more structured and holistic understanding of CUIs, consolidating key trends, challenges, and research gaps in the field.

The ACM International Conference on Conversational User Interfaces (CUI) is an annual event dedicated to research and innovation in human-centered conversational technologies. Established in 2019, ACM CUI emerged from successful workshops at ACM CHI and ACM Mobile HCI, signaling the growing importance of CUIs in both academia and industry [1]. The conference attracts researchers, practitioners, and industry professionals from diverse fields such as artificial intelligence (AI), human-computer interaction (HCI), and linguistics, fostering collaboration across disciplines [2, 3]. Key topics include chatbots, voice assistants, multimodal interfaces, ethics, and user experience design [1, 3]. ACM CUI aims to establish a diverse and dynamic research community, driving innovation, and setting high standards in conversational technologies while addressing challenges like accessibility, inclusivity, and ethical design [2, 4]. With six iterations of ACM CUI completed, the field has matured, making this the right time to consolidate findings and identify trends and opportunities to guide future innovation.

Just like any other interaction with a system, interaction with CUIs consist of three fundamental elements: *the system* (a set of interconnected elements that influence one another), *the interface* (the boundary across which two systems exchange information), and *the user* (an entity that interacts with a system through an interface to derive value from it)[77]. Understanding CUIs requires a close examination of these three elements, as they collectively influence how users interact and feel while using them, drive the advancement of new technologies to improve efficiency and effectiveness, and ensure that systems respect human dignity and uphold fundamental values. Building on this perspective, we introduce a structured framework (figure 1) for analyzing CUIs, which extends these three fundamental elements into three key dimensions of



CUI '25, July 8–10, 2025, Waterloo, ON, Canada

ACM ISBN 979-8-4007-1527-3/2025/07
https://doi.org/10.1145/3719160.3736621

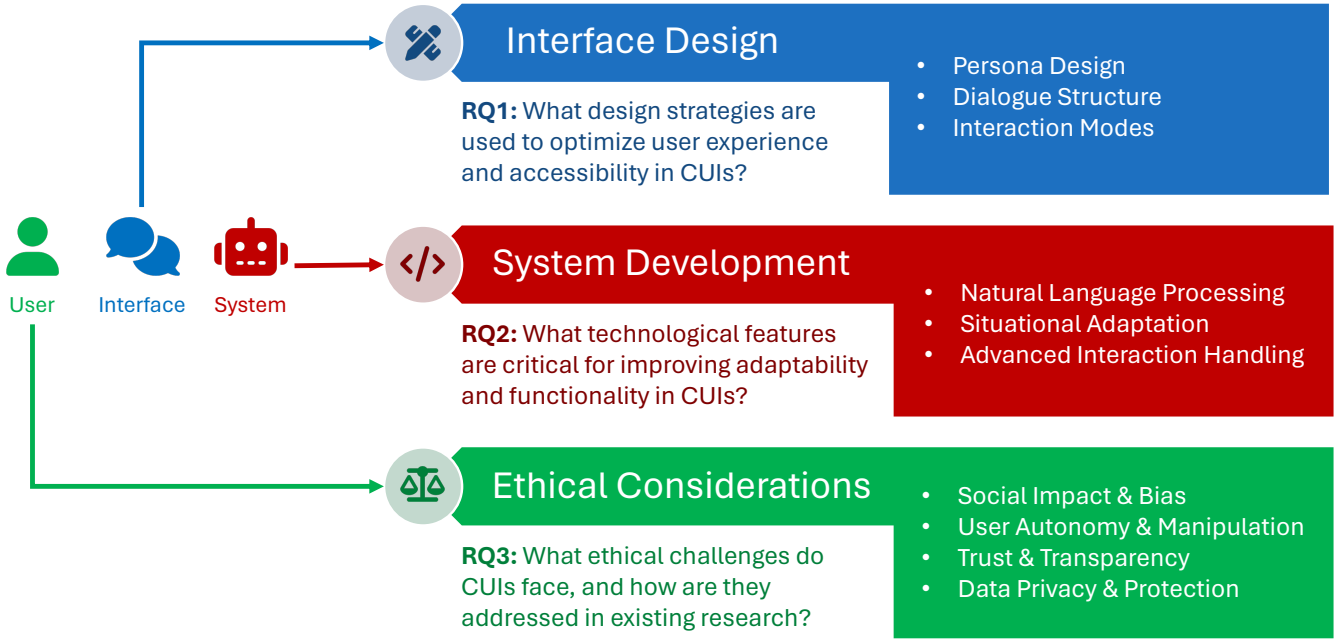


Figure 1: Our proposed framework for analyzing CUI research, which connects the three fundamental elements of interaction—User, Interface, and System—to three key research dimensions: Interface Design (RQ1), System Development (RQ2), and Ethical Considerations (RQ3). This framework provides a structured approach to examining CUIs by categorizing relevant research topics under each dimension.

research: *Interface Design*, *System Development*, and *Ethical Considerations*. Just as interaction with a system relies on the system, the interface, and the user, these dimensions reflect critical areas of study in CUIs.

Interfaces determine how users engage with a system and how intuitive or accessible that engagement is. A thorough review can aggregate insights to develop cohesive principles for designing CUIs that facilitate seamless and effective user interactions. This leads to the question:

> **RQ1 - Interface Design:** *What design strategies are used to optimize user experience and accessibility in CUIs?*

Technological advancements such as AI-driven large language models and multimodal systems have expanded the adaptability and usability of CUIs. Examining past studies can help identify the most impactful technological features and provide a roadmap for leveraging these advancements in system development. This raises the question:

> **RQ2 - System development:** *What technological features are critical for improving adaptability and functionality in CUIs?*

Ethical considerations such as privacy, transparency, and bias mitigation are at the forefront of responsible AI design involving human users. While ACM CUI discussions frequently address these concerns, there are still gaps in comprehensively tackling ethical challenges. Reviewing existing research will clarify how these issues are currently handled and highlight areas for improvement in ethical CUI development. This brings us to the question:

> **RQ3 - Ethical Considerations:** *What ethical challenges do CUIs face, and how are they addressed in existing research?*

To provide a comprehensive overview of CUI research as presented at the ACM CUI conference, we reviewed all 279 publications from its six editions (2019–2024), analyzing them across the three core dimensions of Interface Design, System Development, and Ethical Considerations. Our review intentionally focuses on this specific venue to examine how the ACM CUI community has evolved, what key themes it has addressed, and how it reflects the maturity of CUI research. While we acknowledge that related work exists in other venues, our aim is to assess the scope and development of this dedicated community. The main contributions of our work are as follows:

- **Framework for Structured Analysis of CUI Research:** We provide a structured framework that enables a comprehensive and balanced analysis of CUIs, offering a foundation for organizing and evaluating future research in the field.
- **Structured Review of CUI Research:** We analyze trends, advancements, and gaps in CUI research over the past five years, synthesizing insights to provide a cohesive understanding of the field.
- **Analysis of Interface Design:** Our study examines how dialogue structure, multimodal interactions, and persona design impact user engagement and accessibility, offering a foundation for improving CUI usability.
- **Technological Advancements in CUIs:** We highlight developments in system development, such as conversation memory, multimodal interaction handling, and NLP-based

emotion recognition, which enhance CUIs' adaptability and effectiveness.
- **Ethical Considerations:** Our work identifies ethical concerns, including bias, privacy, trust, and transparency, emphasizing the importance of responsible AI practices in CUI development.
- **Identification of Research Gaps:** We uncover gaps in dataset availability, interdisciplinary collaboration, and cultural representation, pointing to areas requiring further exploration.
- **Future Research Directions:** Based on our findings, we propose improvements in reasoning mechanisms, multilingual accessibility, and privacy-preserving AI techniques to enhance the effectiveness, usability, and inclusivity of CUIs.

By structuring findings across multiple dimensions within the context of ACM CUI, our work serves as a foundation for future research in this community, supporting the development of more efficient, inclusive, and ethically responsible conversational systems.

2 Related Work

CUIs have gained academic and commercial attention, leading to a growing body of review studies examining their various aspects. Existing reviews have primarily focused on distinct domains, such as healthcare applications [26, 124, 138], AI-driven systems designed to support learning by acting as tutors, mentors, or interactive educational tools (also referred to as pedagogical conversational agents, or PCAs) [113], and task automation [142]. Other studies examined interface design and user experience [65, 104, 140], as well as ethical considerations, privacy, and security [28, 61, 67, 121, 127, 128]. These reviews provide valuable insights but often adopt a narrow scope, lacking a comprehensive, multi-dimensional analysis of CUIs across different research areas and application domains.

Healthcare is a promising domain for CUI applications, with several reviews focusing on their use in chronic disease management, mental health support, and patient education [124, 138]. The reviews highlighted the potential of CUIs to enhance self-care and patient engagement [26, 124, 138]. Patient satisfaction rates are generally high, with many studies reporting positive user experiences and increased engagement with healthcare chatbots [138]. Future research could further explore how CUIs can be optimized for personalized care, improving long-term adherence to treatment plans and enhancing interactions between patients and healthcare providers [5].

Other review studies have examined pedagogical conversational agents (PCAs) and task automation [113, 113, 142]. PCAs have been shown to improve student engagement and motivation in learning environments, but challenges such as the uncanny valley effect—where overly human-like agents evoke discomfort—remain a concern [113]. In task automation, intelligent agents assist users in streamlining daily activities, yet usability challenges persist, necessitating the development of better heuristics and guidelines for their implementation [142]. These studies, while insightful, do not integrate their findings into a broader analysis of CUIs, limiting their applicability beyond specific contexts.

A major area of study in past reviews concerns interface design and user experience in CUIs[65, 140]. Research has explored how human-like characteristics—such as avatars, communication styles, and multimodal interactions—affect user perceptions and engagement [65]. Some studies suggested that anthropomorphic design elements enhance trust and rapport, while others indicated that excessive human-likeness can lead to discomfort [140]. Additionally, conversational style adaptability plays a critical role in user satisfaction, with research demonstrating that CUIs capable of adjusting elaborateness and indirectness improve interaction quality [104].

Ethical concerns in CUIs have also been reviewed, particularly regarding privacy, bias, inclusivity, and security [28, 61, 127, 128]. CUIs frequently reinforce gender stereotypes through default female voices, raising concerns about representation and social bias [67]. Privacy is a dominant issue, as users often express distrust due to data collection practices, unauthorized data access, and lack of transparency in voice assistants [61, 127]. Trust-building mechanisms, such as personalization and transparency, are also highlighted as key factors influencing user adoption [121].

Despite the growing body of research on CUIs, existing reviews remain narrow in scope, focusing on specific domains without providing a holistic and structured overview of CUI research. Many studies examined only one dimension of CUI design—such as trust, communication styles, or usability—but failed to integrate insights across multiple research areas. To address these gaps, this review provides a comprehensive, multi-dimensional analysis of all publications from the ACM Conference on Conversational User Interfaces from 2019 to 2024. Unlike previous studies, our work examines trends, methodologies, application areas, and ethical considerations, offering a structured synthesis of key topics in CUI research. By identifying understudied areas, research gaps, and emerging opportunities, this review contributes to a more inclusive and interdisciplinary understanding of CUIs, ensuring that future research and development efforts are better informed and ethically grounded.

3 Methodology

To understand the body of work presented at the ACM International Conference on CUI, we performed a literature review on all the publications at CUI to uncover trends, insights, and gaps in the field. We retrieved all publications from the CUI conference proceedings from 2019 to 2024 from ACM Digital Library. To ensure comprehensive coverage of the conference's published work, all papers were included in the review. This resulted in 279 papers.

To structure our analysis, we employed thematic analysis [35] to identify patterns across the collected papers. Our analysis was guided by a predefined conceptual structure based on the three fundamental elements of interaction with a system: the interface, the system, and the user [77]. Informed by this foundation, we examined CUI research through four dimensions: Research Overview, Interface Design, System Development, and Ethical Considerations. The first provided broader contextual insights, while the latter three formed the core analytical framework shaping our analysis.

The Research Overview dimension captured broad characteristics of the collected papers, including their application domains, target audiences, and methodological approaches. This provided a

Table 1: Codebook detailing the framework's dimensions, codes, and subcodes used to analyze research on Conversational User Interfaces (CUIs). The table presents the structured coding framework for Interface Design, Technological Development, and Ethical Considerations, while the Research Overview dimension, which captures fundamental characteristics of the papers, is described in the text.

Dimension	Code	Subcode	Definition
Interface Design	Persona Design (N= 21, 7.5%)	Gender representation (N= 13, 4.7%)	Design of CUIs' gender.
		Visual appearance (N= 10, 3.6%)	Design of CUIs' visual representation.
		Name (N= 1, 0.4%)	Naming conventions of CUIs.
		Backstory (N= 1, 0.4%)	Development of CUIs' narrative.
	Dialogue Structure (N= 100, 35.8%)	Clarity & Simplicity (N= 44, 15.8%)	Simplicity & ease of CUI interactions.
		Error Handling & Repair Strategies (N= 26, 9.3%)	Management of misunderstandings in CUIs.
		Dialogue State Tracking & Turn-Taking (N= 23, 8.2%)	Flow & context management in CUIs.
		Conversational Style (N= 21, 7.5%)	Tone, humor, & expressiveness in CUIs.
	Interaction Modes (N= 58, 20.8%)	Auditory (N= 61, 21.9%)	Use of voice-based interaction in CUIs.
		Visual (N= 30, 10.8%)	Use of graphical elements in CUIs.
		Textual (N= 19, 6.8%)	Use of text-based interaction in CUIs.
System Development	Natural Language Processing (N= 51, 18.3%)	Prosody, Emotion, & Sentiment Analysis (N= 21, 7.5%)	Recognition & response to emotions in CUIs.
		Data Strategies & Training Techniques (N= 17, 6.1%)	Data collection & model training for CUIs.
		Advanced Reasoning (N= 10, 3.6%)	Logical inference & adaptability in CUIs.
		Fact-Checked Responses & Factual Accuracy (N= 7, 2.5%)	Reliability & accuracy of CUI responses.
		Speech Recognition & Understanding (N= 4, 1.4%)	Accuracy & inclusivity of CUI voice input.
		Efficiency & On-device Optimization (N= 2, 0.7%)	Resource efficiency in CUIs.
	Situational Adaptation (N= 82, 29.4%)	Conversation Memory (N= 66, 23.7%)	Retention & recall of past interactions in CUIs.
		Environmental Context (N= 14, 5.0%)	Adaptation of CUIs to surroundings.
		Integration of Knowledge Bases (N= 11, 3.9%)	Use of external data for CUI responses.
	Advanced Interaction Handling (N= 56, 20.1%)	Multi-Modal Interactions (N= 33, 11.8%)	Combination of multiple input/output methods in CUIs.
		Multi-Language & Code-Switching (N= 16, 5.7%)	Support for multilingual interactions in CUIs.
		Multi-Person (Multi-Party) Interactions (N= 8, 2.9%)	Handling of conversations with multiple users.
Ethical Consideration	Social Impact & Bias (N= 73, 26.2%)		Fairness & inclusivity in CUIs.
	User Autonomy & Manipulation (N= 70, 25.1%)		Balance between guidance & control in CUIs.
	Trust & Transparency (N= 49, 17.6%)		Explainability & reliability of CUIs.
	Data Privacy & Protection (N= 49, 17.6%)		Security & consent in CUI data handling.

general understanding of how research in CUIs has been structured and contextualized before examining aspects more directly tied to our specific research questions. Within this dimension, contribution types were classified using the seven established categories in HCI [147], and analysis methods were categorized into three predefined groups: qualitative, quantitative, and mixed methods.

The remaining three dimensions served as the structured lens through which we developed our framework. Each paper was analyzed through these perspectives, with codes and subcodes emerging through thematic analysis. These codes were structured into a framework that categorizes studies based on their focus on user experience, technological implementation, or ethical implications. This approach ensures that the framework reflects not only the breadth of existing research but also the underlying concerns and advancements shaping the field. Table 1 presents the detailed classification of codes and subcodes developed within this framework.

Under the *interface design* dimension, we reviewed the literature on user-facing aspects of CUIs, focusing on persona design, dialogue structure, and interaction modes. *Persona design* includes how CUIs convey gender, appearance, and backstory. While conversation style and interaction patterns influence persona perception, we treat dialogue structure separately due to its distinct role. *Dialogue structure* refers to how conversational elements are organized and sequenced. *Interaction modes* ensure the framework applies across different CUI types, including text-based, voice, and multimodal systems. Identifying these modes helps maintain adaptability across technological platforms and use cases.

The *system development* dimension is focused on how CUIs function and adapt to different contexts. Three codes emerged in this category, including natural language processing, situational adaptation, and advanced interaction handling. *Natural language processing* examines how CUIs process and generate text or speech, including language models, dialogue management, and response generation techniques. *Situational adaptation* reflects how CUIs handle memory retention, recognize user preferences, and adjust responses based on prior interactions to create more personalized experiences. *Advanced interaction handling* encompasses systems that incorporate multimodal, multilingual, and multi-user interactions. This includes studies focusing on different input-output

methods, multilingual processing, and conversational dynamics involving multiple participants.

Ethical considerations captures aspects that focus on responsible CUI design and its societal impact, such as social impact and bias, user autonomy and manipulation, trust and transparency, and data privacy and protection. *Social impact and bias* include discussions on how CUIs may reinforce stereotypes or contribute to broader societal biases in AI-generated interactions. *User autonomy and manipulation* address concerns about persuasive design and user control over interactions. *Trust and transparency* focus on ensuring that CUI decision-making processes are explainable and understandable to users, thereby enhancing credibility and fostering responsible adoption. Lastly, *data privacy and protection* is a concern in CUI research, highlighting issues related to data collection policies, security risks, and privacy-preserving techniques.

4 Findings

4.1 Overall Research

4.1.1 Contribution Types. Figure 2 presents CUI papers' contribution types from 2019 to 2024. The largest focus has been on *artifact or system* contributions, particularly in 2020 (N=24, 42.1%) and 2024 (N=23, 34.8%), indicating strong interest in practical applications. *Empirical studies about people* have steadily grown, from (N=4, 15.4%) in 2019 to (N=12, 19.4%) in 2023, with a notable peak in 2022 (N=10, 33.3%). Similarly, *empirical studies about how people use a system* have been consistently explored, peaking in 2024 (N=15, 22.7%). However, *dataset* contributions remain limited, with no studies in 2021 and a small peak in 2023 (N=5, 8.1%). Contributions to *methods* and *theory* have been modest but notable, with methods reaching (N=7, 10.6%) in 2024 and theories peaking at (N=5, 16.7%) in 2022. Contributions categorized as *essay or argument* or *meta-analysis or literature surveys* have consistently been the least explored areas, with essay contributions at (N=3, 10.0%) in 2022 and meta-analyses peaking at (N=3, 4.8%) in 2023.

4.1.2 Applications. Research on CUIs has largely concentrated on two key applications: *healthcare and accessibility* (N= 47, 16.8%) and *education and learning* (N= 35, 12.5%). In healthcare and accessibility, studies highlighted the role of CUIs in supporting people with disabilities, promoting mental health, and encouraging healthier lifestyles [6, 9, 48]. Similarly, in education and learning, research showed that CUIs are fostering personalized and collaborative experiences, helping students engage creatively and enhance their critical thinking [52, 63].

While healthcare and education dominate the landscape, researchers also explored CUIs in less prevalent but equally intriguing domains. In *customer engagement and business operations* (N= 9, 3.2%), studies examined task-oriented chatbots designed to improve user interactions [155]. *Recommender and decision support systems* (N= 8, 2.9%) have also been studied, with tools like collaborative restaurant selection systems that streamline group decision-making [89]. In *entertainment and media engagement* (N= 7, 2.5%), CUIs have been investigated for applications like storytelling and interactive gaming, exemplified by SlugBot, which engages users with narrative-driven games [31]. Research in *safety, security, and law enforcement* (N= 5, 1.8%) has focused on tools such as systems for enhancing emergency response interactions [12], while *sustainability and environmental impact* (N= 2, 0.7%) studies included frameworks that encourage eco-friendly behaviors in smart homes [74].

4.1.3 Target Audiences. Among the various audience classifications (e.g., age, education, health, language), *Age-Based Groups* were the most frequently studied (N=49, 17.6%). Within this category, older adults made up the largest subgroup (N=18, 6.4%), with studies focusing on accessibility, digital inclusion, and empathetic support [17, 30]. Young adults (N=14, 5.0%) were also a prominent subgroup, often featured in educational, social, and professional contexts such as financial literacy and career development [131]. In contrast, adolescents (N=3, 1.1%) and children (N=6, 2.1%) appeared less frequently, with research in these groups typically addressing creativity, mental health, and participatory design [94].

In the *Educational Background* category (N=37, 13.3%), lifelong learners were the most represented group (N=17, 6.1%), with CUIs supporting skill development for novice and non-technical users [78]. Undergraduate students (N=14, 5.0%) were also frequently studied, especially in relation to collaborative learning and computational thinking [110]. Only one study focused on school-level students (N=1, 0.4%), highlighting a significant gap in CUI design for younger learners [24].

Research involving participants with *Health Conditions* (N=24, 8.6%) primarily targeted physical disabilities (N=10, 3.6%) and mental health support (N=6, 2.1%). For example, speech synthesis innovations have improved access for visually impaired users [90], while CUIs for mental health offered tools for mood support and personal reflection [20]. Areas such as maternal health (N=2, 0.7%) and speech diversity (N=1, 0.4%) remain largely underexplored.

In the *Language and Cultural Background* category (N=19, 6.8%), most studies examined multilingual users (N=15, 5.4%), focusing on issues like code-switching and linguistic support for non-native English speakers [120]. However, previous work has shown that the vast majority of CUI studies recruit participants from Europe or North America, with limited reporting on linguistic diversity or geographic origin [129]. Technologies developed for non-Western languages or contexts are rarely included in the literature. Although a few studies addressed cultural adaptation (N=4, 1.4%)—notably in regions like South Africa and Brazil—there is still a significant lack of attention to low-literacy users and underrepresented languages [101]. These findings point to a broader need for more inclusive and globally representative CUI research.

4.1.4 Research and Analysis Methods. Figure 3 presents the intersections and frequencies of research and analysis methods used in ACM CUI papers. It is based on papers that explicitly employed research or analysis methods (N=203, 72.8%), with purely theoretical works, such as arguments, excluded. *Experimental Research* is the most frequently used method (N=89, 31.9%), reflecting the field's focus on rigorously testing hypotheses in controlled environments. *Collaborative Engagement Methods* (N=52, 18.6%), *Surveys* (N=46, 16.5%), and *Interviews* (N=48, 17.2%) are also prominent, showcasing the ACM CUI community's commitment to understanding users' perspectives and designing for their needs. Significant intersections, such as Collaborative Engagement Methods paired with Interviews or Surveys, highlighted an emphasis on gathering diverse insights

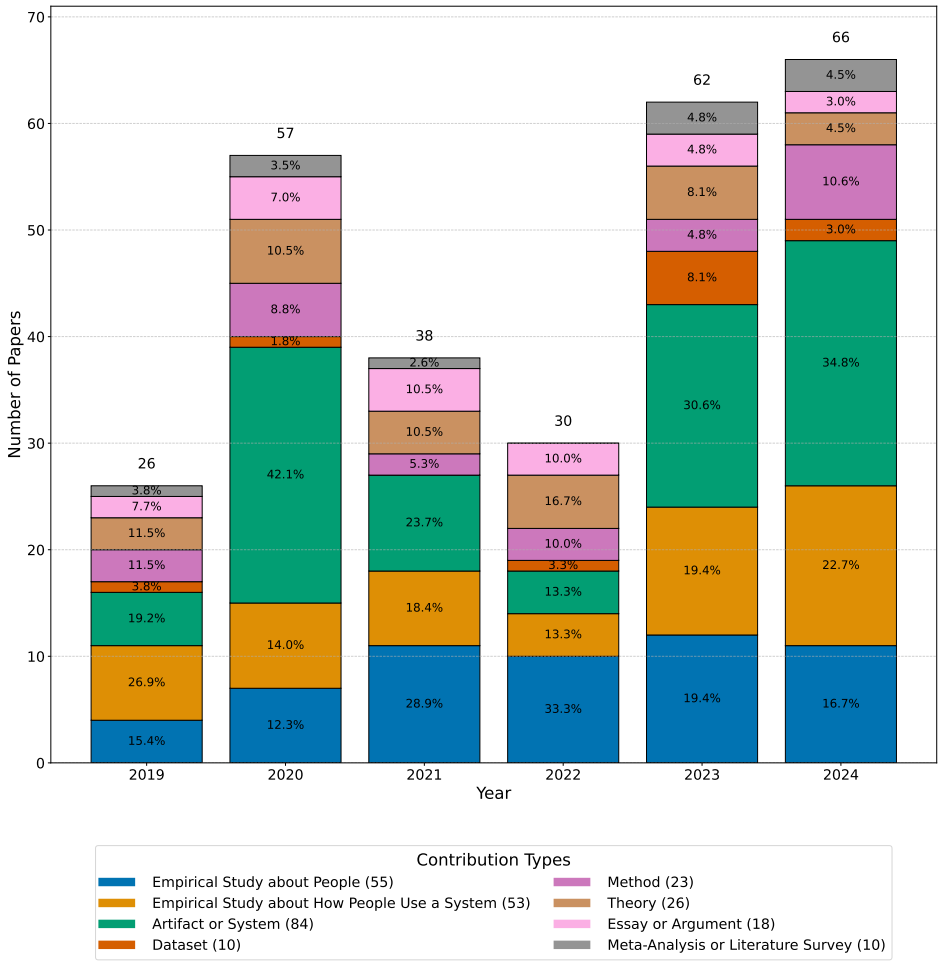


Figure 2: Contribution types in ACM CUI publications from 2019 to 2024.

to refine user-centered designs. Experimental Research is often combined with Usability Testing, illustrating a layered approach where qualitative insights guide iterative testing and validation.

Less frequently used methods, like *Observatory Studies* (N=29, 10.4%), suggest limited exploration of user behavior in naturalistic settings. Intersections involving Observatory Studies are especially rare, likely due to challenges in observing and analyzing real-world interactions. Similarly, combinations such as Usability Testing with Collaborative Engagement Methods are underexplored, even though they could provide valuable insights into collaborative problem-solving processes in realistic environments.

Mixed Methods dominated the analysis landscape (N=121, 43.4%), demonstrating a preference for integrating qualitative and quantitative insights to address complex research questions. For example, Mixed Methods are frequently used in Experimental Research to combine statistical rigor with in-depth understanding. *Qualitative Methods* (N=67, 24.0%) are central to exploring user experiences and interaction patterns, particularly in studies involving Interviews or Collaborative Engagement. Thematic Analysis and Conversation Analysis are the most used qualitative methods, enabling researchers to identify patterns and explore nuanced interaction dynamics. However, purely *Quantitative Methods* (N=15, 5.4%) remain underutilized, limiting opportunities to generalize findings across larger populations.

4.2 Interface Design

4.2.1 Persona Design. Persona design in CUIs involves creating distinct, relatable virtual agents by integrating gender representation, visual appearance, naming conventions, and backstories to enhance user engagement and trust. *Gender representation* (N= 13, 4.7%) influences how users perceive CUIs, particularly in terms of trustworthiness. Studies showed that users often rate female virtual agents as more trustworthy than male agents [139]. However, users bring their own biases into interactions, interpreting even gender-ambiguous voices through binary lenses. Research indicated that unless all design elements are carefully considered, societal expectations influence how CUIs are perceived [137]. To avoid reinforcing gender stereotypes while fostering inclusivity, designers must select voices and portrayals that ensure diverse representation. Studies

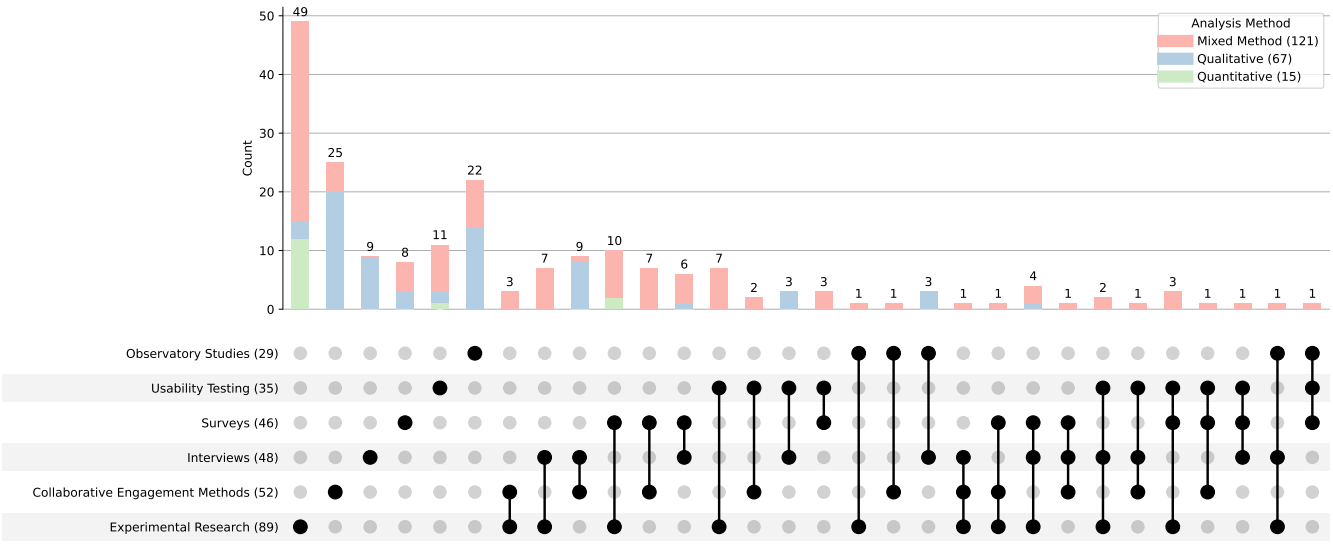


Figure 3: Upset plot depicting the intersections and frequencies of research and analysis methods used in ACM CUI papers. The vertical bars represent how frequently specific combinations of methods appear with color coding, distinguishing qualitative (pink), quantitative (blue), and mixed methods (green). The lower left column lists individual research methods and their total frequencies. The lower right plot shows method intersections with filled circles between the intersecting methods and the total number of papers and their analysis method in the upper bar chart.

suggested that virtual assistants should avoid reinforcing submissive roles for female voices while also ensuring representation for non-binary and marginalized identities [53]. Moving beyond binary norms requires designing CUIs that challenge traditional gender expectations, offering users a more nuanced and inclusive interaction [137]. Reflecting user diversity through a range of voices helps promote inclusivity, particularly benefiting underrepresented communities such as LGBTQ+ users [53]. One approach to fostering inclusivity is the development of gender-neutral voices, such as the Q voice, which researchers explored as a means of providing a more comfortable and inclusive experience for all users [53].

Visual appearance (N= 10, 3.6%) also plays an important role in shaping user trust and engagement. Research showed that while users rate the functionality of virtual agents similarly, their preferences vary significantly based on visual representation [29]. Integrating both visual and auditory cues enhances trust, as users respond more positively when facial expressions and tone of voice align [139]. In immersive environments, well-designed avatars and environmental cues increase engagement by fostering a stronger sense of presence [27]. Personalized avatars create a more socially interactive experience, benefiting users who may experience isolation by transforming CUIs into supportive companions [134]. Since users have diverse expectations, customizable design elements allow them to choose between human-like or non-human representations based on their preferences for trust or practicality [29]. Researchers found that striking the right balance in embodiment levels is essential, as too much embodiment can be distracting while too little can make the interface feel impersonal [134].

Name selection (N= 1, 0.4%) affects how users perceive and connect with CUIs. Research indicated that factors such as the agent's name, organizational background, and perceived attributes like race and gender contribute to user trust and relatability [91]. Furthermore, *backstory* (N= 1, 0.4%) enhances user familiarity and engagement by giving virtual agents personality traits that align with user expectations. Researchers found that carefully crafted backstories help users form stronger connections with voice assistants, making interactions more enjoyable and immersive [60].

4.2.2 Dialogue Structure. Dialogue structure refers to the organized sequence and arrangement of conversational elements within CUIs. A key factor in achieving this is *clarity and simplicity* (N= 44, 15.8%), which makes CUIs effective and easy to use. Studies showed that instead of developing complex, human-like conversations, designers should focus on simple and effective systems that maintain a smooth flow, as even basic responsive features significantly improve user engagement without requiring sophisticated technology [18]. Using accessible language without jargon ensures inclusivity and empowers users by improving their understanding and decision-making abilities. Research indicated that clear and simple explanations are essential, especially when providing fact-based information, as avoiding technical jargon enhances accessibility and media literacy [114].

Another important aspect of clarity is understanding how users interact with CUIs. Studies found that users often rely on direct, command-like language, highlighting the need for systems to recognize and process simplified inputs efficiently [21]. Presenting information concisely while avoiding overload is essential for maintaining engagement. Research showed that balancing the amount of information provided prevents users from feeling overwhelmed, ensuring that responses remain relevant and digestible [58, 80].

To further enhance usability, minimizing cognitive load is necessary, particularly in environments requiring immediate comprehension. Research indicated that guiding users through tasks by breaking them into smaller steps improves interaction efficiency, reducing mental effort [51]. This strategy is especially beneficial in high-stress situations, where managing workload effectively is critical. Studies showed that voice assistants that divided complex tasks into manageable steps improved user engagement in learning and aviation contexts [76, 154]. Additionally, older users benefit greatly from simplified communication processes in CUIs, as clear instructions and responses enhance their ability to interact comfortably with technology [123].

While clarity and simplicity make CUIs more intuitive, *error handling and repair strategies* (N= 26, 9.3%) ensure that interactions remain smooth even when misunderstandings occur. Researchers suggested that CUIs should provide transparent feedback to help users understand system limitations, as detailed error explanations to foster reliability [22]. In interactive environments like gaming, ensuring accurate command recognition is important to maintain engagement and reduce frustration. Studies showed that responsive error-handling mechanisms improve user experience by clarifying misunderstandings and reinforcing system reliability [54].

To manage errors effectively, using multiple repair strategies proves to be more effective than relying on a single approach. Research suggested that incorporating diverse techniques to handle misunderstandings leads to better outcomes [13]. Additionally, prompting users for more context when their inputs are incomplete helps make CUIs more intuitive, as this mirrors the way humans naturally ask for clarification [6].

For a truly engaging experience, *dialogue state tracking and turn-taking* (N= 23, 8.2%) must be designed to feel natural and intuitive. Studies showed that improving turn-taking mechanisms allows CUIs to better mimic human conversations, leading to smoother and more engaging interactions [17, 18]. Additionally, incorporating elements such as visual cues enhances user engagement by aligning with natural human communication patterns [85].

Ensuring a logical flow in conversations is another important consideration. Research indicated that structuring dialogue in a way that mirrors natural conversation progression improves coherence and usability [34] and focusing on progressivity—ensuring that conversations move forward rather than feeling stagnant—is makes voice interfaces feel more interactive [68].

Finally, *conversational style* (N= 21, 7.5%) influences how users perceive CUIs, with natural-sounding interactions making chatbots feel more engaging and approachable. Studies indicated that mimicking human communication patterns, such as turn-taking and emoticon use, makes chatbot interactions feel more familiar and enjoyable [57]. Additionally, adapting humor and communication styles to user preferences ensures that interfaces feel comfortable and inviting [136]. When catering to specific demographics, such as adolescents, using language that feels familiar and relatable further enhances engagement [112].

4.2.3 Interaction Modes. We identified the main modalities in the design of CUIs, which are shown in Figure 4. The majority of papers (N=221, 79.2%) focused on general, modality-agnostic designs, emphasizing broad usability. A substantial number of papers (N=61, 21.9%) addressed auditory CUIs, followed by visual CUIs (N=30, 10.8%), and textual modalities (N=19, 6.8%). We also found overlaps between modalities, with auditory and visual CUIs often being combined to enhance multimodal interactions. These overlaps reflect a growing interest in designing CUIs that integrate multiple modalities to improve user engagement and accessibility.

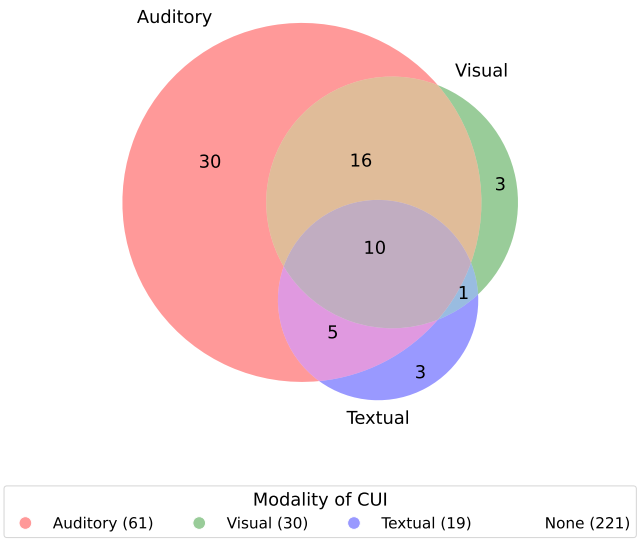


Figure 4: Distribution of papers by modality focus in ACM CUI publications. Note that 221 papers did not specify a modality for their discussed CUIs.

4.3 System Development

4.3.1 Natural Language Processing. NLP encompasses computational techniques for analyzing and generating human language, enabling machines to understand, interpret, and interact with human speech and text. One key aspect of NLP in CUIs is *Prosody, Emotion, and Sentiment Analysis* (N= 21, 7.5%), which influences how users engage with conversational systems by allowing them to recognize and respond to emotional cues. Research showed that CUIs with emotional intelligence can proactively manage tensions, using interventions like humor to enhance user satisfaction and diffuse uncomfortable situations [36]. Studies indicated that incorporating design tools that analyze pitch and tone helps tailor responses to users' emotions, making interactions more natural and meaningful [39]. Emotion recognition also plays a critical role in trust-building, particularly during errors. Research showed that users appreciate voice assistants that acknowledge frustration through apologies or adapted responses, which improves trust [14]. However, adjusting CUIs' tone to match negative emotional states must be handled carefully to maintain efficiency and effectiveness [151].

To make CUIs more adaptable, *Data Strategies and Training Techniques* (N= 17, 6.1%) play a crucial role in improving how they process user inputs. Research showed that training models with datasets that include interrupted queries improves CUIs' ability to handle incomplete queries effectively [6]. Studies also indicated that increasing diversity in expression within datasets makes CUIs

more responsive by mimicking natural conversation dynamics [46]. Addressing data imbalances is vital, as research showed that larger and more diverse datasets improve classifier performance, allowing CUIs to interpret a broader range of user inputs accurately [102]. When real user data is scarce, researchers found that synthetic data, such as that generated by GPT models, serves as a useful supplement but should not replace real interactions due to potential gaps in nuanced expression [103].

Advanced Reasoning (N= 10, 3.6%) extends CUIs' ability to handle more sophisticated interactions. Studies indicated that users expect chatbots to manage complex inquiries, which suggests the need for systems with advanced functionalities to enhance customer service [69]. Researchers found that integrating self-reflection mechanisms improves response accuracy by allowing CUIs to evaluate their own outputs [135]. Additionally, research showed that fallback mechanisms using LLMs help CUIs generate approximate answers for unrecognized queries, enhancing usability instead of merely returning error messages [75].

Ensuring that CUIs provide reliable information is another essential consideration, making *Fact-Checked Responses and Factual Accuracy* (N= 7, 2.5%) an area of focus. Studies indicated that prioritizing fact-checked news in Voice User Interfaces is essential for maintaining journalistic integrity, particularly in areas like politics and business [117]. Research suggested that robust fact-checking mechanisms are necessary in sensitive fields such as education and healthcare, where misinformation can have serious consequences [153]. Integrating external information sources improves the accuracy of CUIs, particularly in specialized domains where internal knowledge might be insufficient [135]. Recent methods such as Retrieval-Augmented Generation (RAG) help maintain factual consistency while improving conversational engagement [82].

Studies on *Speech Recognition and Understanding* (N= 4, 1.4%) tried to understand how CUIs interpret user inputs to create more natural interactions. Studies indicated that aligning speech recognition with real-world conversational patterns improves user engagement and authenticity [34]. Embracing conversational imperfections, such as disfluencies like "um" and "uh," enhances realism and relatability [22]. Research showed that CUIs should also recognize and process non-standard speech features like mumbling and slang to ensure accessibility for diverse users [24].

Finally, Efficiency and On-Device Optimization (N= 2, 0.7%) ensures that CUIs function effectively while minimizing computational costs. Research showed that simpler machine learning models, such as Random Forest, can achieve similar accuracy as deep learning models while using fewer resources, making them suitable for resource-limited settings [78]. Studies indicated that optimizing response times and costs in RAG systems is essential for practical deployment in real-world applications [135].

4.3.2 Situational Adaptation. Situational adaptation in CUIs refers to their ability to adjust dynamically to prior interactions, environmental factors, and external knowledge sources, ensuring more relevant and effective user experiences. A key aspect of this adaptability is *conversation memory* (N= 66, 23.7%), which allows CUIs to recall previous exchanges, improving personalization and efficiency. Research indicated that CUIs should dynamically modify responses based on historical interactions, improving communication over time [15]. Efficient memory management enhances continuity in conversations, even after breaks, making interactions feel smoother. Studies showed that implementing an 'on-hold state' allows chatbots to resume discussions naturally by reminding users of past exchanges [20]. Researchers found that chatbots designed to remember previous messages provide a more coherent and satisfying experience, especially when users refer back to prior parts of the conversation—such as variables, years, or visual elements—compared to chatbots that do not track conversation history [86].

Beyond maintaining conversation flow, memory in CUIs contributes to learning experiences by tracking user progress and feedback. Research showed that continuity in interactions creates a more personalized learning environment, strengthening user engagement [55]. CUIs that adapt to user input can reflect complex identities by generating responses aligned with social contexts [62]. Moreover, memory enhances accuracy in responses by providing necessary context. Studies indicated that dialogue history significantly improves retrieval accuracy, helping CUIs better understand intent [135]. Researchers found that integrating LLMs further refines these interactions by allowing CUIs to personalize responses dynamically, making conversations more immersive [32].

CUIs could also benefit from *environmental context* (N= 14, 5.0%), which allows them to align assistance with users' surroundings and activities. Research showed that incorporating environmental awareness, such as sensing emotions or surroundings, improves CUIs' ability to provide timely and relevant support, enhancing task performance [37]. Studies indicated that integrating location and activity data into CUIs makes interactions more effective, increasing user satisfaction [84]. For example, in-car assistants benefit from proactive suggestions, such as recommending nearby locations or mindfulness exercises, enhancing overall user experience during their journeys [99].

Environmental factors also influence how users communicate with CUIs. Research showed that the nature of a task and its setting shape language use, emphasizing the importance of designing adaptable CUIs that adjust to different contexts [49]. Proactive design enables CUIs to act on environmental cues rather than waiting for user input. Studies indicated that proactive conversational agents encourage beneficial behaviors, such as promoting sustainability, by prompting users at the right moments [74]. Additionally, research showed that in noisy or busy settings, multimodal CUIs improve accessibility by allowing users to interact through the most suitable method for their environment [100].

Some research has focused on *knowledge bases* (N= 11, 3.9%), which enables CUIs to provide accurate, context-aware responses by linking to evolving information sources. Researchers suggested that CUIs should rely on dynamic knowledge bases to deliver timely insights, particularly in fast-paced industries like manufacturing, where linking CUIs to real-time factory systems improves decision-making by allowing assistants to analyze and respond based on current operational conditions [87]. Research also showed that leveraging structured knowledge bases like Wikidata improves response accuracy, helping CUIs better understand queries in context [6]. Furthermore, advanced methods such as RAG enhance CUIs' conversational abilities by balancing factual consistency with social

interaction quality. Studies indicated that while RAG strengthens accuracy, it is equally important to maintain conversational fluency to ensure natural and engaging interactions [82].

4.3.3 Advanced Interaction Handling. Advanced Interaction Handling refers to the design and implementation of sophisticated communication methods in CUIs, leveraging multi-modal interactions, multi-language capabilities, and multi-party dynamics to create personalized, intuitive, and seamless user experiences. *Multi-modal interactions* (N= 33, 11.8%) enhance CUIs by integrating multiple sensory inputs and outputs, making interactions more flexible and engaging. Studies showed that allowing users to customize how they receive notifications—whether through vibrations, visuals, or voice prompts—improves comfort and engagement [7]. Researchers found that incorporating various feedback types, including speech, visuals, and behavioral cues, creates richer and more immersive user experiences [21]. CUIs that combine different modalities, such as text, video, and audio, enhance interactivity and context awareness, leading to more effective communication [37]. Beyond verbal communication, research indicated that nonverbal cues such as gestures and facial expressions align CUIs more closely with human interaction, making conversations feel more natural [70].

Multi-language and code-switching (N= 16, 5.7%) capabilities are essential for CUIs to accommodate bilingual and multilingual users. Studies indicated that bilingual speakers frequently switch between languages within conversations, requiring CUIs to recognize and respond to these shifts for more natural interactions [23]. Research showed that CUIs designed to handle code-switching allow smoother interactions, particularly for users who may not be fluent in their second language [47]. Recognizing these language shifts in real time reduces frustration and improves user confidence. Studies found that integrating language identification into Automatic Speech Recognition systems enhances transcription accuracy, leading to higher user satisfaction [81]. Further research indicated that fine-tuning models like Whisper to manage both language identification and transcription simultaneously improves performance, making CUIs more adaptable for multilingual users [81].

Multi-person (multi-party) interactions (N= 8, 2.9%) enable CUIs to support conversations involving multiple users, increasing engagement in social and educational settings. Studies indicated that CUIs capable of managing multi-user interactions create a more dynamic and interactive experience [16] and better resemble real human interactions [18]. Understanding conversational dynamics in multi-user scenarios is crucial for effective design. CUIs should be able to distinguish between speakers and tailor responses accordingly to maintain clarity and relevance [152]. To enhance multi-party interactions, researchers suggested that implementing a 'Dedicated Mode for Interaction' allows users to converse without needing to repeatedly activate the system [8].

4.4 Ethical Considerations

Social Impact and Bias (N= 73, 26.2%) influence how CUIs shape user perceptions and societal norms. Research showed that conversational agents often reflect the demographics they are modeled after, which can unintentionally reinforce social stereotypes [116]. To foster inclusivity, studies indicated that incorporating diverse voices in voice assistants can challenge traditional gender norms and support LGBTQ+ representation [53]. However, even gender-ambiguous voices are often interpreted through a binary lens unless all design elements are carefully considered [137].

Additionally, CUIs can promote critical thinking by encouraging users to reflect on different perspectives. Studies showed that features designed to challenge existing beliefs help mitigate biases and support informed decision-making [11]. Addressing hidden biases in dialogue systems requires diverse testing methods, as traditional evaluation approaches may overlook them [33]. CUIs fostering multi-perspective discussions create richer conversations and better decision-making outcomes [126].

User Autonomy and Manipulation (N= 70, 25.1%) is another ethical CUI development investigated by researchers. Research showed that designers must anticipate ethical issues from the outset, particularly privacy concerns and manipulation, to create responsible and user-friendly interactions [10]. To prevent deceptive practices, CUIs must avoid 'dark patterns' that manipulate users into unintended decisions. Research found that transparency about data collection and system capabilities is essential to maintaining ethical interactions [10]. Studies also showed that CUIs should empower users by allowing them to adjust features and customize their interactions based on their preferences, fostering a sense of control [9, 59].

Ensuring users do not become overly dependent on CUIs requires careful trust calibration. Research indicated that adjusting how much trust users place in CUIs prevents over-reliance while ensuring they remain in control of their choices [59]. Transparency about how AI systems process user behavior further strengthens trust and allows users to make informed decisions [66]. Ethical CUIs should also be capable of refusing unethical requests. Studies found that CUIs programmed to reject morally problematic requests help users recognize their biases and reflect on ethical considerations [146]. However, research showed that human oversight remains necessary, particularly in sensitive areas like investigations, to ensure ethical compliance and accountability [72].

Trust and Transparency (N= 49, 17.6%) are critical for ensuring users feel confident in CUIs. Studies indicated that explanatory dialogues, where users can ask "why" and receive informative responses, enhance transparency and make the technology more trustworthy [14]. Chatbots providing clear explanations for their responses help users understand system reasoning, reinforcing trust in AI decisions [40]. Accountability is equally important, particularly in error handling. CUIs should be designed to explain their actions, especially when mistakes occur, so users feel assured that the system is reliable and that corrections can be made [56]. Studies also indicated that transparency throughout AI design, including data collection and model development, is necessary for ensuring fairness and accountability [92].

Users appreciate clarity regarding CUI capabilities and limitations. Research showed that when systems explicitly state their constraints, users develop more realistic expectations and privacy concerns are addressed [96]. Studies indicated that chatbots should clearly identify themselves as non-human to prevent misleading interactions and promote ethical communication [97].

Data Privacy and Protection (N= 49, 17.6%) is another important factor which should be integrated from the outset of CUI design process [10]. Studies indicated that avoiding manipulative design

tactics and maintaining transparency about data collection prevent unethical exploitation of user information [10].

Continuous listening devices pose unique privacy challenges, requiring clear policies on when they can collect data. Research found that defining when CUIs are actively listening reassures users and builds trust [17]. Compliance with privacy regulations like GDPR is also essential for ethical CUI development. Research indicated that ensuring secure data handling practices aligns with legal standards and strengthens user confidence [44]. Studies found that developers should stay informed on evolving legal frameworks to ensure CUIs meet regulatory and ethical expectations [105].

User empowerment through control over data privacy settings enhances security and trust in CUIs. Research showed that enabling users to manage their conversation logs and disclosure preferences fosters transparency and confidence in interactions [117]. Studies indicated that ethical data collection practices, including obtaining informed consent, are particularly important in sensitive applications like healthcare, where user privacy is paramount [71].

5 Discussion

In this section, we highlight gaps in CUI research, opportunities to critically analyze current CUI interactions, and ways to advance the development of more effective and user-centered conversational systems.

5.1 Research Overview: Need for Establishing Ecologically Valid Datasets and Benchmarks with Broaden Participants

Despite advancements in CUI research, several gaps remain that must be addressed to enhance the field's impact and inclusivity. One major limitation is the lack of high-quality, standardized datasets for CUI development. Prior studies have emphasized the need for richer conversational datasets that reflect real-world interaction complexities [73] and broader domain coverage for training utterances [115]. To overcome this challenge, future research should prioritize large-scale data collection across diverse languages, demographics, and application domains such as healthcare and education. While gathering such datasets requires significant effort to ensure privacy and data anonymization, it is worthwhile for establishing benchmark datasets with standardized evaluation metrics that enable reproducibility and comparative studies. Collaborations between academia and industry can also support access to real-world conversational logs, provided ethical data handling and transparent consent policies are in place.

Another critical issue is the narrow focus of much CUI research on adult users from Western regions, often overlooking how children, adolescents, and culturally diverse populations interact with these systems. Recent work analyzing the ACM CUI venue has highlighted that while ethics and inclusion are often mentioned, actual participation from underrepresented groups remains rare [129]. Technologies developed outside the Western world are also underrepresented. These patterns raise important concerns about whose needs are prioritized in the design of CUIs. We emphasize the importance of inclusive research that engages broader and more diverse populations. Young users, for example, are increasingly interacting with CUIs [133], yet we know little about their experiences. While studies have found young people prefer Generative AI responses to some help-seeking messages, their preferences on more sensitive mental health topics still need more investigations [150]. Recent findings show that students perceive AI tools as collaborative learning partners, not just information sources, and use them to test ideas and reflect on feedback—highlighting the potential of CUIs to support exploratory learning and metacognition when appropriately designed [119]. Future studies should examine their use in educational and social settings, and investigate long-term effects, such as on creativity or social behavior. There is also a growing emphasis on pedagogy readiness and the need for critical AI literacy to ensure that such technologies support rather than undermine educational goals [119].

CUIs must be thoroughly studied before being deployed in high-stakes domains like healthcare or education. More research is needed on how CUIs can augment, rather than replace, healthcare professionals, particularly in supporting personal informatics through conversational data collection that captures meaningful self-reported insights. Research should move beyond controlled lab settings and focus on real-world interactions. Cross-cultural investigations are also needed to better understand how people from different regions expect and engage with CUIs. Prior work has already stressed the importance of designing for linguistic and cultural differences [125, 129], and this direction remains critical for creating globally relevant systems.

Finally, current CUI research tends to rely on short-term studies in artificial environments. To gain a realistic understanding of how CUIs are used in daily life, future work should adopt methods such as passive monitoring, diary studies, and long-term deployments in homes, hospitals, or public institutions [34, 93]. Integrating approaches from ethnography and communication research can also deepen our understanding of everyday CUI use and help build systems that are more responsive to real human contexts.

5.2 Interface Design: Consistency, Accessibility, and Character Depth

The design of CUIs lacks standardized principles, leading to inconsistencies in usability, accessibility, and evaluations. While previous research has identified the need for foundational guidelines [43] and attempted to adapt heuristics from graphical interfaces [106], there remains no comprehensive framework tailored specifically for CUIs. Future research should focus on developing standardized design principles through large-scale usability studies across domains such as healthcare, education, and customer service. Expert-driven workshops involving designers, linguists, and accessibility specialists can refine these principles to ensure inclusivity. Additionally, an open-source benchmarking system should be established to evaluate CUIs based on key usability metrics, such as response latency, conversational clarity, and error-handling efficiency. By creating and validating these standards, CUIs can be designed with greater consistency, improving overall trust and usability.

Another major gap in CUI research is the lack of tailored conversational styles for different demographics. Current CUIs often adopt a uniform interaction model, while different user groups may have distinct communication needs. Prior studies have highlighted how

adolescents use conversational agents for emotional support but have not provided specific implementation strategies for adapting communication styles [94]. On the other hand, research on chatbot formality has shown that conversational tone influences engagement, yet demographic variations have not been explored [50]. To address this, future studies should conduct controlled experiments analyzing how different age groups, cognitive abilities, and cultural backgrounds respond to variations in tone, language complexity, and response timing. These studies can help CUI researchers create a framework to adjust conversational style based on user profiles. Implementing such strategies would allow CUIs to provide more inclusive and personalized experiences, ensuring accessibility for diverse user groups.

The design of virtual agent personas also remains an underexplored area, particularly concerning name selection, backstory development, and persona adaptation. While research has examined chatbot personality traits such as agreeableness [141], CUIs often lack depth and consistency in persona construction. Insights from other disciplines offer valuable strategies for refining persona design. Theater and acting studies emphasize emotional depth, motivation, and adaptability in designing a persona [95, 132]. Game design techniques for iterative character development could be adopted to ensure that personas remain engaging and functionally coherent over time [41]. Literary studies and creative writing provide frameworks for crafting nuanced, socially aware personas that enhance the realism of interactions [98]. By drawing from these fields, CUI designers can create virtual agents with richer, more adaptive personalities that foster deeper user engagement.

5.3 System Development: Enhancing Expressivity, Reasoning, and Efficiency in CUIs

One major limitation of current CUIs is their reliance on text-based processing, which fails to capture prosody, intonation, and nonverbal cues essential for natural conversations. Research showed that CUIs struggle with non-standard speech forms such as mumbling, reducing their effectiveness in real-world interactions [24]. Future CUIs could incorporate textless NLP, a technique that enables direct processing of raw speech without relying on written transcriptions. This approach allows models to understand expressive features like rhythm, tone, and speaker identity, making interactions more natural [88]. Textless NLP is especially beneficial for low-resource and unwritten languages, as it eliminates the need for large, annotated text datasets [79]. To make this technology practical for CUIs, researchers should develop models trained on diverse spoken interactions. Additionally, hybrid models that combine speech-based and text-based processing should be explored to improve both expressivity and semantic accuracy [109].

Another major challenge is reasoning, as current CUIs struggle to handle complex queries requiring logical progression. Unlike humans, who break down problems into clear steps, most CUIs retrieve single-turn responses without deeper inference. To improve reasoning, future research should integrate Chain-of-Thought (CoT) prompting, which enables CUIs to explain their reasoning step by step [144]. This can be combined with Self-Consistency techniques, where multiple reasoning paths are generated, and the most consistent answer is selected [143]. For more advanced decision-making, Tree-of-Thought (ToT) reasoning should be explored, allowing CUIs to evaluate multiple possible solutions before finalizing a response [148].

Another critical limitation of CUIs that rely on LLMs is that they often forget past interactions, lose track of long conversations, and provide inconsistent responses due to the fixed-size context window. One way to address this is through adaptive note-taking mechanisms, which dynamically summarize and retain essential details—such as user preferences, unresolved questions, and key points—while discarding redundant information [42]. Another approach is memory layers, which function like an internal knowledge base by storing factual details in a structured key-value format, allowing CUIs to retrieve past information instantly without overloading the context window [25]. Additionally, hierarchical memory architectures introduce short-term, long-term, and persistent memory, enabling CUIs to differentiate between immediate conversation flow, retained user-specific details, and general knowledge that remains unchanged [42]. These memory enhancements can improve CUIs by maintaining context more effectively and ensuring continuity across interactions, making them feel more intelligent, personalized, and reliable. These capabilities are especially relevant for longitudinal, user-driven contexts like mental health support, where continuity and reflection over time may enhance users' ability to make sense of their own experiences.

Finally, CUIs must be optimized for low-resource environments, where computational power and internet connectivity are limited. Large-scale AI models require significant resources, making them impractical for many real-world applications. Prior studies highlighted the need for lightweight NLP models, but current solutions remain inadequate for offline and low-bandwidth scenarios [101]. To address this, researchers should focus on model compression techniques such as pruning and quantization, which reduce computational load while preserving accuracy. Knowledge distillation can be used to transfer knowledge from large models to smaller, more efficient versions suited for embedded systems. Hybrid processing approaches should also be developed, where on-device CUIs handle basic tasks while more complex queries are offloaded to cloud-based systems when connectivity is available. This is especially important for collecting personal data in sensitive contexts such as healthcare and mental health, where local processing can reduce privacy risks while still enabling helpful interactions.

5.4 Ethical Considerations: Mitigating Bias, Refining Trust, and Strengthening Privacy in CUIs

CUIs have made significant progress in reducing gender bias, yet broader ideological, cultural, and political biases remain underexplored. Research has shown that language models embedded in CUIs are shaped by the values and perspectives of their developers, influencing how they present political figures, economic policies, and governance models [38]. Previous studies have focused on gender representation and the exclusion of non-binary identities in CUIs [53] but have not examined how these systems frame sociopolitical issues. Future research should focus on increasing

transparency in model design by documenting training data origins, ideological influences, and decision-making processes. Researchers should also explore mechanisms that allow users to access multiple AI perspectives, fostering a pluralistic AI ecosystem that prevents ideological homogenization. Large-scale cross-cultural studies are needed to assess how different CUIs present politically sensitive information and how these representations impact user trust and decision-making.

The issue of trust calibration in CUIs is equally critical, as users must balance confidence in the system's capabilities with a healthy level of skepticism. Prior research has highlighted how trust influences user reliance on CUIs, proposing design strategies to improve trust calibration through transparency and feedback mechanisms [59]. However, these studies have not examined how trust evolves over extended use or how calibration strategies impact decision-making in high-stakes environments such as healthcare [64]. Longitudinal studies should be conducted to measure shifts in user trust over time, particularly when interacting with CUIs that provide recommendations in domains where errors carry significant consequences. Future research should also explore adaptive trust calibration models, where CUIs dynamically adjust their confidence indicators, explanations, and user feedback mechanisms based on the individual's experience level and contextual risks. By integrating insights from psychology, human-computer interaction, and AI ethics, researchers can develop standardized trust metrics that differentiate between appropriate trust and over-reliance.

Another critical area requiring attention is how CUIs handle unethical requests. Emerging research has shown that current systems can act unethically, engaging in inappropriate or harmful behaviors without user prompting. For example, analyses of user reviews of the Replika chatbot have documented cases of AI-induced sexual harassment, where chatbots made unsolicited sexual advances or failed to respect user boundaries [107]. Addressing unethical behavior should not be left to users, as this shifts the burden of managing AI failures onto those most affected by them. Recent work highlights how CUI research must consider the distribution of labor in human-AI interactions and take responsibility for preventing harmful outputs [108]. Many existing systems rely on evasive refusals, which can frustrate users and reduce trust in AI-driven interactions. Previous work has argued for greater moral transparency in CUIs but has not developed structured frameworks for how these systems should phrase refusals or ensure consistency in ethical decision-making [146]. Other research has proposed high-level guidelines for identifying ethical concerns in virtual assistants but has not addressed the need for adaptive frameworks that can evolve with changing ethical expectations [10]. Future work should integrate Constitutional AI and Constitutional Classifiers to enhance ethical refusals by ensuring CUIs provide principled, educational responses rather than simple denials [19, 130]. Constitutional AI enables CUIs to critique their own responses based on predefined ethical guidelines, refining refusals to ensure clarity, fairness, and constructive engagement. Constitutional Classifiers add an additional layer of filtering, allowing CUIs to block unethical prompts while ensuring that refusals include informative justifications. Researchers should empirically test different refusal strategies across cultural contexts, evaluating user reactions, trust levels, and comprehension of ethical reasoning. Additionally, domain-specific ethical refusal frameworks should be developed for CUIs used in legal, medical, and educational contexts to ensure nuanced, context-aware decision-making.

Privacy and data protection in CUIs remain ongoing challenges, particularly in balancing personalization with security and compliance with evolving regulations like GDPR. These concerns become even more critical when CUIs are used to collect sensitive information—such as therapeutic journaling or mental health reflections—which require rigorous safeguards for consent, storage, and access. Existing research has highlighted the need for greater transparency in how CUIs handle data but has not established concrete privacy-preserving methodologies that can be widely implemented across different systems [83]. Prior studies have also examined privacy risks associated with conversational agents but have not proposed scalable solutions that allow CUIs to learn from user interactions while minimizing data exposure [122]. Federated Learning offers a promising solution by allowing CUIs to train on decentralized user data while preserving privacy [149]. Unlike traditional centralized learning, Federated Learning ensures that raw user data remains on users' devices, significantly reducing exposure risks. However, Federated Learning alone is insufficient unless combined with additional security measures. Future research should optimize Federated Learning for CUIs using secure aggregation techniques such as differential privacy [111] to further protect user data during training [45]. Researchers should also develop standardized Federated Learning-based privacy protocols that ensure CUIs comply with global regulatory frameworks while maintaining usability. Hybrid privacy models should be explored, combining Federated Learning with on-device personalization techniques that allow CUIs to adapt to individual user preferences without centralizing sensitive data. Comparative studies should assess how different privacy-preserving methods impact CUI performance and user trust, to improve the security measures in a way to balance with the quality of user interactions.

6 Limitations and Future Work

This study provides valuable insights into CUI research trends while highlighting opportunities for future work. It intentionally focuses on ACM CUI conference papers from 2019 to 2024, and therefore does not include relevant research from major interdisciplinary venues like CHI, ACL, and NeurIPS. Future studies can broaden the scope to include these venues for a more comprehensive and cross-disciplinary analysis. Incorporating quantitative citation analysis could also offer deeper insights into the most influential works shaping the field. Expanding future reviews with broader sources, interdisciplinary perspectives, and structured methodologies will further strengthen our understanding of CUI advancements.

7 Conclusion

In this literature review, we systematically analyzed all ACM CUI conference publications (2019–2024) across interface design, system development, and ethical considerations. Research on interface design explores dialogue structure, persona design, and multimodal interactions to enhance user engagement. System development advancements focus on conversation memory, multimodal interactions, and natural language processing, improving adaptability and

responsiveness. Ethical concerns, including bias, user autonomy, trust, and privacy, remain key challenges.

By offering a structured framework, we synthesized key trends and uncovered gaps in CUI research. Our study specifically focuses on research published at the ACM CUI conference to examine how this dedicated venue has developed over time and addressed critical themes in conversational user interfaces. We identify trends and gaps in design principles, reasoning mechanisms, and ethical frameworks to guide future developments. Moving forward, it is important for the CUI research community to prioritize strengthening ethical safeguards, establishing standardized ecologically valid benchmark datasets and design principles, improving reasoning capabilities with AI, and enhancing multilingual accessibility.

References

[1] 2019. *CUI '19: Proceedings of the 1st International Conference on Conversational User Interfaces.* Association for Computing Machinery, Dublin, Ireland.

[2] 2020. *CUI '20: Proceedings of the 2nd Conference on Conversational User Interfaces.* Association for Computing Machinery, Bilbao, Spain.

[3] 2023. *CUI '23: Proceedings of the 5th International Conference on Conversational User Interfaces.* Association for Computing Machinery, Eindhoven, Netherlands.

[4] 2024. *CUI '24: Proceedings of the 6th ACM Conference on Conversational User Interfaces.* Association for Computing Machinery, Luxembourg, Luxembourg.

[5] Alaa Abd-Alrazaq, Zeineb Safi, Mohannad Alajlani, Jim Warren, Mowafa Househ, and Kerstin Denecke. 2020. Technical Metrics Used to Evaluate Health Care Chatbots: Scoping Review. *Journal of Medical Internet Research* 22, 6 (June 2020), e18301. https://doi.org/10.2196/18301

[6] Angus Addlesee and Marco Damonte. 2023. Understanding and Answering Incomplete Questions. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23).* Association for Computing Machinery, New York, NY, USA, Article 10. https://doi.org/10.1145/3571884.3597133

[7] Shashank Ahire, Saeid Othman, and Michael Rohs. 2024. Dual-Mode Interventions: Giving Agency to Knowledge Workers in Proactive Health Interventions. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24).* Association for Computing Machinery, New York, NY, USA, Article 37. https://doi.org/10.1145/3640794.3665578

[8] Shashank Ahire and Michael Rohs. 2020. Tired of Wake Words? Moving towards Seamless Conversations with Intelligent Personal Assistants. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20).* Association for Computing Machinery, New York, NY, USA, Article 20. https://doi.org/10.1145/3405755.3406141

[9] Shashank Ahire, Benjamin Simon, and Michael Rohs. 2024. WorkFit: Designing Proactive Voice Assistance for the Health and Well-Being of Knowledge Workers. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24).* Association for Computing Machinery, New York, NY, USA, Article 5. https://doi.org/10.1145/3640794.3665561

[10] Sanju Ahuja and Jyoti Kumar. 2022. Assistant or Master: Envisioning the User Autonomy Implications of Virtual Assistants. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22).* Association for Computing Machinery, New York, NY, USA, Article 4. https://doi.org/10.1145/3543829.3544514

[11] Annalena Bea Aicher, Daniel Kornmüller, Wolfgang Minker, and Stefan Ultes. 2023. Self-Imposed Filter Bubble Model for Argumentative Dialogues. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23).* Association for Computing Machinery, New York, NY, USA, Article 23. https://doi.org/10.1145/3571884.3597131

[12] Saul Albert, William Housley, and Elizabeth Stokoe. 2019. In Case of Emergency, Order Pizza: An Urgent Case of Action Formation and Recognition. In *Proceedings of the 1st International Conference on Conversational User Interfaces (Cui '19).* Association for Computing Machinery, New York, NY, USA, Article 15. https://doi.org/10.1145/3342775.3342800

[13] Essam Alghamdi, Martin Halvey, and Emma Nicol. 2024. System and User Strategies to Repair Conversational Breakdowns of Spoken Dialogue Systems: A Scoping Review. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24).* Association for Computing Machinery, New York, NY, USA, Article 28. https://doi.org/10.1145/3640794.3665558

[14] Fatemeh Alizadeh, Peter Tolmie, Minha Lee, Philipp Wintersberger, Dominik Pins, and Gunnar Stevens. 2024. Voice Assistants' Accountability through Explanatory Dialogues. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24).* Association for Computing Machinery, New York, NY, USA, Article 29. https://doi.org/10.1145/3640794.3665557

[15] Sungeun An, Robert Moore, Eric Young Liu, and Guang-Jie Ren. 2021. Recipient Design for Conversational Agents: Tailoring Agent's Utterance to User's Knowledge. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21).* Association for Computing Machinery, New York, NY, USA, Article 30. https://doi.org/10.1145/3469595.3469625

[16] Dimitra Anastasiou, Benjamin Gâteau, Luc Vandenabeele, Patrick Gratz, and Johannes Hermen. 2024. MagiChat: An AI-based Card-Guessing Chatbot. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24).* Association for Computing Machinery, New York, NY, USA, Article 47. https://doi.org/10.1145/3640794.3665574

[17] Matthew Peter Aylett, Andrea Carmantini, Christoper J Pidcock, Eric Nichols, and Randy Gomez. 2023. A Pilot Evaluation of a Conversational Listener for Conversational User Interfaces. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23).* Association for Computing Machinery, New York, NY, USA, Article 40. https://doi.org/10.1145/3571884.3605871

[18] Matthew Peter Aylett and Marta Romeo. 2023. You Don't Need to Speak, You Need to Listen: Robot Interaction and Human-like Turn-Taking. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23).* Association for Computing Machinery, New York, NY, USA, Article 11. https://doi.org/10.1145/3571884.3603750

[19] Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R. Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. 2022. Constitutional AI: Harmlessness from AI Feedback. https://doi.org/10.48550/arXiv.2212.08073 arXiv:2212.08073 [cs]

[20] Erkan Basar, Divyaa Balaji, Linwei He, Iris Hendrickx, Emiel Krahmer, Gert-Jan de Bruijn, and Tibor Bosse. 2023. HyLECA: A Framework for Developing Hybrid Long-Term Engaging Controlled Conversational Agents. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23).* Association for Computing Machinery, New York, NY, USA, Article 56. https://doi.org/10.1145/3571884.3604404

[21] Mitchell Baxter, Anna Bleakley, Justin Edwards, Leigh Clark, Benjamin R. Cowan, and Julie R. Williamson. 2021. "You, Move There!": Investigating the Impact of Feedback on Voice Control in Virtual Environments. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21).* Association for Computing Machinery, New York, NY, USA, Article 14. https://doi.org/10.1145/3469595.3469609

[22] Sophie Becker, Philip Doyle, and Justin Edwards. 2022. Embrace Your Incompetence! Designing Appropriate CUI Communication through an Ecological Approach. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22).* Association for Computing Machinery, New York, NY, USA, Article 13. https://doi.org/10.1145/3543829.3544531

[23] Casey C. Bennett. 2024. Findings from Studies on English-Based Conversational AI Agents (Including ChatGPT) Are Not Universal. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24).* Association for Computing Machinery, New York, NY, USA, Article 41. https://doi.org/10.1145/3640794.3665879

[24] Casey C. Bennett and Minha Lee. 2023. Would People Mumble Rap to Alexa?. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23).* Association for Computing Machinery, New York, NY, USA, Article 21. https://doi.org/10.1145/3571884.3603757

[25] Vincent-Pierre Berges, Barlas Oğuz, Daniel Haziza, Wen-tau Yih, Luke Zettlemoyer, and Gargi Ghosh. 2024. Memory Layers at Scale. https://doi.org/10.48550/ARXIV.2412.09764

[26] Caterina Bérubé, Theresa Schachner, Roman Keller, Elgar Fleisch, Florian V Wangenheim, Filipe Barata, and Tobias Kowatsch. 2021. Voice-Based Conversational Agents for the Prevention and Management of Chronic and Mental Health Conditions: Systematic Literature Review. *Journal of Medical Internet Research* 23, 3 (March 2021), e25933. https://doi.org/10.2196/25933

[27] Anna Bleakley, Vincent Wade, and Benjamin R. Cowan. 2020. Finally a Case for Collaborative VR? The Need to Design for Remote Multi-Party Conversations. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20).* Association for Computing Machinery, New York, NY, USA, Article 25. https://doi.org/10.1145/3405755.3406144

[28] Tom Bolton, Tooska Dargahi, Sana Belguith, Mabrook S. Al-Rakhami, and Ali Hassan Sodhro. 2021. On the Security and Privacy Challenges of Virtual Assistants. *Sensors* 21, 7 (March 2021), 2312. https://doi.org/10.3390/s21072312

[29] Michael Bonfert, Nima Zargham, Florian Saade, Robert Porzel, and Rainer Malaka. 2021. An Evaluation of Visual Embodiment For Voice Assistants On Smart Displays. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui*

'21). Association for Computing Machinery, New York, NY, USA, Article 16. https://doi.org/10.1145/3469595.3469611

[30] Rezvan Boostani, Cosmin Munteanu, and Anastasia Kuzminykh. 2024. Unravelling the Paradigm: Prioritizing Process over Results—Sports Training as a Catalyst for Age-Friendly CUI Design. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 58. https://doi.org/10.1145/3640794.3665882

[31] Kevin K. Bowden, Jiaqi Wu, Wen Cui, Juraj Juraska, Vrindavan Harrison, Brian Schwarzmann, Nicholas Santer, Steve Whittaker, and Marilyn Walker. 2019. Entertaining and Opinionated but Too Controlling: A Large-Scale User Study of an Open Domain Alexa Prize System. In *Proceedings of the 1st International Conference on Conversational User Interfaces (Cui '19)*. Association for Computing Machinery, New York, NY, USA, Article 24. https://doi.org/10.1145/3342775.3342792

[32] Efe Bozkir, Süleyman Özdel, Ka Hei Carrie Lau, Mengdi Wang, Hong Gao, and Enkelejda Kasneci. 2024. Embedding Large Language Models into Extended Reality: Opportunities and Challenges for Inclusion, Engagement, and Privacy. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 38. https://doi.org/10.1145/3640794.3665563

[33] Anouck Braggaar, Linwei He, and Jan De Wit. 2024. Our Dialogue System Sucks - but Luckily We Are at the Top of the Leaderboard!: A Discussion on Current Practices in NLP Evaluation. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 48. https://doi.org/10.1145/3640794.3665889

[34] Adam Brandt, Spencer Hazel, Rory Mckinnon, Kleopatra Sideridou, Joe Tindale, and Nikoletta Ventoura. 2023. From Writing Dialogue to Designing Conversation: Considering the Potential of Conversation Analysis for Voice User Interfaces. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 22. https://doi.org/10.1145/3571884.3603758

[35] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[36] Alexandra W.D. Bremers, Natalie Friedman, Sam Lee, Tong Wu, Eric Laurier, Malte F Jung, Jorge Ortiz, and Wendy Ju. 2024. (Social) Trouble on the Road: Understanding and Addressing Social Discomfort in Shared Car Trips. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 52. https://doi.org/10.1145/3640794.3665580

[37] Alexandra W.D. Bremers, Manaswi Saha, and Adolfo G. Ramirez-Aristizabal. 2024. Situated Conversational Agents for Task Guidance: A Preliminary User Study. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 51. https://doi.org/10.1145/3640794.3665575

[38] Maarten Buyl, Alexander Rogiers, Sander Noels, Guillaume Bied, Iris Dominguez-Catena, Edith Heiter, Iman Johary, Alexandru-Cristian Mara, Raphaël Romero, Jefrey Lijffijt, and Tijl De Bie. 2025. Large Language Models Reflect the Ideology of Their Creators. https://doi.org/10.48550/arXiv.2410.18417 arXiv:2410.18417 [cs]

[39] Julia Cambre and Chinmay Kulkarni. 2020. Methods and Tools for Prototyping Voice Interfaces. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 43. https://doi.org/10.1145/3405755.3406148

[40] Heloisa Candello, Gabriel Meneguelli Soella, Cassia Sampaio Sanctos, Marcelo Carpinette Grave, and Adinan Alves De Brito Filho. 2023. "This Means Nothing to Me": Building Credibility in Conversational Systems. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 29. https://doi.org/10.1145/3571884.3603759

[41] Cassiano Canheti, Flávio Andalo, and Milton Luiz Horn Vieira. 2019. Case Study: Game Character Creation Process. In *Advances in Human Factors in Wearable Technologies and Game Design*, Tareq Z. Ahram (Ed.). Vol. 795. Springer International Publishing, Cham, 343–354. https://doi.org/10.1007/978-3-319-94619-1_34

[42] Edoardo Cetin, Qi Sun, Tianyu Zhao, and Yujin Tang. 2024. An Evolved Universal Transformer Memory. https://doi.org/10.48550/ARXIV.2410.13166

[43] Fermin Chavez-Sanchez, Gloria Adriana Mendoza Franco, Gloria Angelica Martínez de la Peña, and Erick Iroel Heredia Carrillo. 2020. Beyond What Is Said: Looking for Foundational Principles in VUI Design. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 28. https://doi.org/10.1145/3405755.3406145

[44] Chen Chen, Khalil Mrini, Kemeberly Charles, Ella Lifset, Michael Hogarth, Alison Moore, Nadir Weibel, and Emilia Farcas. 2021. Toward a Unified Metadata Schema for Ecological Momentary Assessment with Voice-First Virtual Assistants. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 31. https://doi.org/10.1145/3469595.3469626

[45] Jingxue Chen, Hang Yan, Zhiyuan Liu, Min Zhang, Hu Xiong, and Shui Yu. 2024. When Federated Learning Meets Privacy-Preserving Computation. *Comput. Surveys* 56, 12 (Dec. 2024), 1–36. https://doi.org/10.1145/3679013

[46] Yoonseo Choi, Hyungyu Shin, Toni-Jan Keith Monserrat, Nyoungwoo Lee, Jeongeon Park, and Juho Kim. 2020. Leveraging the Crowd to Support the Conversation Design Process. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 50. https://doi.org/10.1145/3405755.3406155

[47] Helin Cihan, Yunhan Wu, Paola Peña, Justin Edwards, and Benjamin Cowan. 2022. Bilingual by Default: Voice Assistants and the Role of Code-Switching in Creating a Bilingual User Experience. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22)*. Association for Computing Machinery, New York, NY, USA, Article 12. https://doi.org/10.1145/3543829.3544511

[48] Leigh Clark, Benjamin R. Cowan, Abi Roper, Stephen Lindsay, and Owen Sheers. 2020. Speech Diversity and Speech Interfaces: Considering an Inclusive Future through Stammering. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 24. https://doi.org/10.1145/3405755.3406139

[49] Benjamin R. Cowan, Philip Doyle, Justin Edwards, Diego Garaialde, Ali Hayes-Brady, Holly P. Branigan, João Cabral, and Leigh Clark. 2019. What's in an Accent? The Impact of Accented Synthetic Speech on Lexical Choice in Human-Machine Dialogue. In *Proceedings of the 1st International Conference on Conversational User Interfaces (Cui '19)*. Association for Computing Machinery, New York, NY, USA, Article 23. https://doi.org/10.1145/3342775.3342786

[50] Samuel Rhys Cox and Wei Tsang Ooi. 2022. Does Chatbot Language Formality Affect Users' Self-Disclosure?. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22)*. Association for Computing Machinery, New York, NY, USA, Article 1. https://doi.org/10.1145/3543829.3543831

[51] Pietro Crovari, Fabio Catania, Pietro Pinoli, Philipp Roytburg, Asier Salzar, Franca Garzotto, and Stefano Ceri. 2020. OK, DNA! A Conversational Interface to Explore Genomic Data. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 38. https://doi.org/10.1145/3405755.3406163

[52] Claudia Maria Cutrupi, Salvatore Fadda, Giovanni Valcarenghi, Giulia Cosentino, Fabio Catania, Micol Spitale, and Franca Garzotto. 2020. Smemo: A Multi-Modal Interface Promoting Children's Creation of Personal Conversational Agents. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 36. https://doi.org/10.1145/3405755.3406162

[53] Andreea Danielescu. 2020. Eschewing Gender Stereotypes in Voice Assistants to Promote Inclusion. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 46. https://doi.org/10.1145/3405755.3406151

[54] Sam Davern, Orla Cooney, Iona Gessinger, and Benjamin R. Cowan. 2023. Investigating the Usability of a Speech Agent-Based Role-Playing Game. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 38. https://doi.org/10.1145/3571884.3604306

[55] Smit Desai and Jessie Chin. 2021. Hey Google, Can You Help Me Learn?. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 6. https://doi.org/10.1145/3469595.3469601

[56] Smit Desai, Tanusree Sharma, and Pratyasha Saha. 2023. Using ChatGPT in HCI Research—a Trioethnography. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 8. https://doi.org/10.1145/3571884.3603755

[57] Jayati Dev and L. Jean Camp. 2020. User Engagement with Chatbots: A Discursive Psychology Approach. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 52. https://doi.org/10.1145/3405755.3406165

[58] Mateusz Dubiel, Alessandra Cervone, and Giuseppe Riccardi. 2019. Inquisitive Mind: A Conversational News Companion. In *Proceedings of the 1st International Conference on Conversational User Interfaces (Cui '19)*. Association for Computing Machinery, New York, NY, USA, Article 6. https://doi.org/10.1145/3342775.3342802

[59] Mateusz Dubiel, Sylvain Daronnat, and Luis A. Leiva. 2022. Conversational Agents Trust Calibration: A User-Centred Perspective to Design. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22)*. Association for Computing Machinery, New York, NY, USA, Article 30. https://doi.org/10.1145/3543829.3544518

[60] Mateusz Dubiel, Smit Desai, Nima Zargham, and Anuschka Schmitt. 2024. Voicecraft: Designing Task-Specific Voice Assistant Personas. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 66. https://doi.org/10.1145/3640794.3670000

[61] Jide S. Edu, Jose M. Such, and Guillermo Suarez-Tangil. 2021. Smart Home Personal Assistants: A Security and Privacy Review. *Comput. Surveys* 53, 6 (Nov. 2021), 1–36. https://doi.org/10.1145/3412383

[62] Justin Edwards, Leigh Clark, and Allison Perrone. 2021. LGBTQ-AI? Exploring Expressions of Gender and Sexual Orientation in Chatbots. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 2. https://doi.org/10.1145/3469595.3469597

[63] Justin Edwards, Andy Nguyen, Marta Sobocinski, Joni Lämsä, Adelson de Araujo, Belle Dang, Ridwan Whitehead, Anni-Sofia Roberts, Matti Kaarlela, and Sanna Jarvela. 2024. MAI - a Proactive Speech Agent for Metacognitive Mediation in Collaborative Learning. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 46. https://doi.org/10.1145/3640794.3665585

[64] Justin Edwards and Elaheh Sanoubari. 2019. A Need for Trust in Conversational Interface Research. In *Proceedings of the 1st International Conference on Conversational User Interfaces (Cui '19)*. Association for Computing Machinery, New York, NY, USA, Article 21. https://doi.org/10.1145/3342775.3342809

[65] Edona Elshan, Naim Zierau, Christian Engel, Andreas Janson, and Jan Marco Leimeister. 2022. Understanding the Design Elements Affecting User Acceptance of Intelligent Agents: Past, Present and Future. *Information Systems Frontiers* 24, 3 (June 2022), 699–730. https://doi.org/10.1007/s10796-021-10230-9

[66] Philip Gregory Feldman. 2024. The Voice: Lessons on Trustworthy Conversational Agents from 'Dune'. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 54. https://doi.org/10.1145/3640794.3665890

[67] Janik Festerling and Iram Siraj. 2022. Anthropomorphizing Technology: A Conceptual Review of Anthropomorphism Research and How It Relates to Children's Engagements with Digital Voice Assistants. *Integrative Psychological and Behavioral Science* 56, 3 (Sept. 2022), 709–738. https://doi.org/10.1007/s12124-021-09668-y

[68] 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, Article 26. https://doi.org/10.1145/3342775.3342788

[69] Asbjørn Følstad and Marita Skjuve. 2019. Chatbots for Customer Service: User Experience and Motivation. In *Proceedings of the 1st International Conference on Conversational User Interfaces (Cui '19)*. Association for Computing Machinery, New York, NY, USA, Article 1. https://doi.org/10.1145/3342775.3342784

[70] Mary Ellen Foster. 2019. Face-to-Face Conversation: Why Embodiment Matters for Conversational User Interfaces. In *Proceedings of the 1st International Conference on Conversational User Interfaces (Cui '19)*. Association for Computing Machinery, New York, NY, USA, Article 13. https://doi.org/10.1145/3342775.3342810

[71] Maria J. Galvez Trigo, Martin Porcheron, Joy Egede, Joel E. Fischer, Adrian Hazzard, Chris Greenhalgh, Edgar Bodiaj, and Michel Valstar. 2021. ALTCAI: Enabling the Use of Embodied Conversational Agents to Deliver Informal Health Advice during Wizard of Oz Studies. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 26. https://doi.org/10.1145/3469595.3469621

[72] Mirai Gendi and Cosmin Munteanu. 2021. Towards a Chatbot for Evidence Gathering on the Dark Web. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 3. https://doi.org/10.1145/3469595.3469598

[73] Emer Gilmartin and Christian Saam. 2020. Pragmatics Research and Non-Task Dialog Technology. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 22. https://doi.org/10.1145/3405755.3406142

[74] Mathyas Giudici, Pietro Crovari, and Franca Garzotto. 2022. CANDY: A Framework to Design Conversational AgeNts for Domestic sustainabilitY. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22)*. Association for Computing Machinery, New York, NY, USA, Article 28. https://doi.org/10.1145/3543829.3544515

[75] Marcos Gomez-Vazquez, Jordi Cabot, and Robert Clarisó. 2024. Automatic Generation of Conversational Interfaces for Tabular Data Analysis. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 33. https://doi.org/10.1145/3640794.3665577

[76] Sarah Gosper, Johanne R. Trippas, Harriette Richards, Fraser Allison, Cynthia Sear, Sara Khorasani, and Fabio Mattioli. 2021. Understanding the Utility of Digital Flight Assistants: A Preliminary Analysis. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 32. https://doi.org/10.1145/3469595.3469627

[77] Erika Hall. 2018. *Conversational Design*. Number No 25 in A Book Apart. A Book Apart, New York.

[78] Xu Han, Michelle Zhou, Yichen Wang, Wenxi Chen, and Tom Yeh. 2023. Democratizing Chatbot Debugging: A Computational Framework for Evaluating and Explaining Inappropriate Chatbot Responses. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 39. https://doi.org/10.1145/3571884.3604308

[79] Michael Hassid, Tal Remez, Tu Anh Nguyen, Itai Gat, Alexis Conneau, Felix Kreuk, Jade Copet, Alexandre Defossez, Gabriel Synnaeve, Emmanuel Dupoux, Roy Schwartz, and Yossi Adi. 2024. Textually Pretrained Speech Language Models. https://doi.org/10.48550/arXiv.2305.13009 arXiv:2305.13009 [cs]

[80] Diana C. Hernandez-Bocanegra and Jürgen Ziegler. 2021. Conversational Review-Based Explanations for Recommender Systems: Exploring Users' Query Behavior. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 1. https://doi.org/10.1145/3469595.3469596

[81] Leopold Hillah, Mateusz Dubiel, and Luis A. Leiva. 2024. "¿Te Vienes? Sure!" Joint Fine-Tuning of Language Detection and Transcription Improves Automatic Recognition of Code-Switching Speech. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 62. https://doi.org/10.1145/3640794.3665579

[82] Sviatlana Höhn, Jauwairia Nasir, Ali Paikan, Pouyan Ziafati, and Elisabeth André. 2024. Using Large Language Models for Robot-Assisted Therapeutic Role-Play: Factuality Is Not Enough!. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 60. https://doi.org/10.1145/3640794.3665886

[83] Trung Dong Huynh, William Seymour, Luc Moreau, and Jose Such. 2023. Why Are Conversational Assistants Still Black Boxes? The Case for Transparency. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 52. https://doi.org/10.1145/3571884.3604319

[84] Razan Jaber and Donald McMillan. 2020. Conversational User Interfaces on Mobile Devices: Survey. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 10. https://doi.org/10.1145/3405755.3406130

[85] Razan Jaber and Donald McMillan. 2022. Cross-Modal Repair: Gaze and Speech Interaction for List Advancement. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22)*. Association for Computing Machinery, New York, NY, USA, Article 25. https://doi.org/10.1145/3543829.3543833

[86] Bernhard Jordan, Laura Koesten, and Torsten Möller. 2023. Chatting about Data - Interacting with Voice Interfaces to Engage with Election Panel Data. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 34. https://doi.org/10.1145/3571884.3597126

[87] Samuel Kernan Freire, Mina Foosherian, Chaofan Wang, and Evangelos Niforatos. 2023. Harnessing Large Language Models for Cognitive Assistants in Factories. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 44. https://doi.org/10.1145/3571884.3604313

[88] Eugene Kharitonov, Jade Copet, Kushal Lakhotia, Tu Anh Nguyen, Paden Tomasello, Ann Lee, Ali Elkahky, Wei-Ning Hsu, Abdelrahman Mohamed, Emmanuel Dupoux, and Yossi Adi. 2022. Textless-Lib: A Library for Textless Spoken Language Processing. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: System Demonstrations*. Association for Computational Linguistics, Hybrid: Seattle, Washington + Online, 1–9. https://doi.org/10.18653/v1/2022.naacl-demo.1

[89] Matthias Kraus, Stina Klein, Nicolas Wagner, Wolfgang Minker, and Elisabeth André. 2024. A Pilot Study on Multi-Party Conversation Strategies for Group Recommendations. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 31. https://doi.org/10.1145/3640794.3665569

[90] Sébastien Le Maguer and Benjamin R. Cowan. 2021. Synthesizing a Human-like Voice Is the Easy Way. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 19. https://doi.org/10.1145/3469595.3469614

[91] Minha Lee, Donald McMillan, Illaria Torre, Joel E. Fischer, and Yvon Ruitenburg. 2024. Between Trust and Identity: Form, Function, and Presentation. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 65. https://doi.org/10.1145/3640794.3669999

[92] Andreas Liesenfeld, Alianda Lopez, and Mark Dingemanse. 2023. Opening up ChatGPT: Tracking Openness, Transparency, and Accountability in Instruction-Tuned Text Generators. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 47. https://doi.org/10.1145/3571884.3604316

[93] Ye Liu, Stefan Ultes, Wolfgang Minker, and Wolfgang Maier. 2023. Unified Conversational Models with System-Initiated Transitions between Chit-Chat and Task-Oriented Dialogues. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 33. https://doi.org/10.1145/3571884.3597125

[94] Irene Lopatovska, Olivia Turpin, Jessika Davis, Ellen Connell, Chris Denney, Hilda Fournier, Archana Ravi, Ji Hee Yoon, and Eesha Parasnis. 2022. Capturing Teens' Voice in Designing Supportive Agents. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22)*. Association for Computing Machinery, New York, NY, USA, Article 23. https://doi.org/10.1145/3543829.3543838

[95] Mary Luckhurst and Sandra Mayer. 2020. THEATRE AND PERSONA: CELEBRITY AND TRANSGRESSION. *Persona Studies* 5, 2 (Feb. 2020), 1–8. https://doi.org/10.21153/psj2019vol5no2art911

[96] Maria Luce Lupetti, Emma Hagens, Willem Van Der Maden, Régine Steegers-Theunissen, and Melek Rousian. 2023. Trustworthy Embodied Conversational Agents for Healthcare: A Design Exploration of Embodied Conversational Agents for the Periconception Period at Erasmus MC. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 25. https://doi.org/10.1145/3571884.3597128

[97] Tamas Makany, Sungjong Roh, Kotaro Hara, Jie Min Hua, Felicia Goh Si Ying, and Wilson Teh Yang Jie. 2023. Beyond Anthropomorphism: Unraveling the True Priorities of Chatbot Usage in Smes. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 46. https://doi.org/10.1145/3571884.3604315

[98] Marta M. Maslej, Keith Oatley, and Raymond A. Mar. 2017. Creating Fictional Characters: The Role of Experience, Personality, and Social Processes. *Psychology of Aesthetics, Creativity, and the Arts* 11, 4 (Nov. 2017), 487–499. https://doi.org/10.1037/aca0000094

[99] Lesley-Ann Mathis, Can Günes, Kathleen Entz, David Lerch, Frederik Diederichs, and Harald Widlroither. 2024. Generating Proactive Suggestions Based on the Context: User Evaluation of Large Language Model Outputs for in-Vehicle Voice Assistants. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 43. https://doi.org/10.1145/3640794.3665568

[100] Linglong Meng and Stefan Schaffer. 2020. A Reporting Assistant for Railway Security Staff. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 39. https://doi.org/10.1145/3405755.3406164

[101] Leina Meoli, Francisco Nunes, Beatriz Félix, Joana Couto da Silva, Xolani Ntinga, and Melissa Densmore. 2024. A Not so Chatty "Chatbot": Co-designing to Support First-Time Parents in South Africa and Portugal. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 30. https://doi.org/10.1145/3640794.3665571

[102] Selina Meyer and David Elsweiler. 2023. Towards Cross-Content Conversational Agents for Behaviour Change: Investigating Domain Independence and the Role of Lexical Features in Written Language around Change. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 17. https://doi.org/10.1145/3571884.3597136

[103] Selina Meyer, David Elsweiler, Bernd Ludwig, Marcos Fernandez-Pichel, and David E. Losada. 2022. Do We Still Need Human Assessors? Prompt-based GPT-3 User Simulation in Conversational AI. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22)*. Association for Computing Machinery, New York, NY, USA, Article 8. https://doi.org/10.1145/3543829.3544529

[104] Juliana Miehle, Wolfgang Minker, and Stefan Ultes. 2022. When to Say What and How: Adapting the Elaborateness and Indirectness of Spoken Dialogue Systems. *Dialogue & Discourse* 13, 1 (April 2022), 1–40. https://doi.org/10.5210/dad.2022.101

[105] Thomas Mildner, Philip Doyle, Gian-Luca Savino, and Rainer Malaka. 2022. Rules of Engagement: Levelling up to Combat Unethical CUI Design. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22)*. Association for Computing Machinery, New York, NY, USA, Article 26. https://doi.org/10.1145/3543829.3544528

[106] Christine Murad, Cosmin Munteanu, Benjamin R. Cowan, and Leigh Clark. 2021. Finding a New Voice: Transitioning Designers from GUI to VUI Design. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 22. https://doi.org/10.1145/3469595.3469617

[107] Mohammad Namvarpour, Harrison Pauwels, and Afsaneh Razi. 2025. AI-induced Sexual Harassment: Investigating Contextual Characteristics and User Reactions of Sexual Harassment by a Companion Chatbot. https://doi.org/10.48550/ARXIV.2504.04299

[108] Mohammad Namvarpour and Afsaneh Razi. 2024. Uncovering Contradictions in Human-AI Interactions: Lessons Learned from User Reviews of Replika. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*. ACM, San Jose Costa Rica, 579–586. https://doi.org/10.1145/3678884.3681909

[109] Tu Anh Nguyen, Benjamin Muller, Bokai Yu, Marta R. Costa-jussa, Maha Elbayad, Sravya Popuri, Christophe Ropers, Paul-Ambroise Duquenne, Robin Algayres, Ruslan Mavlyutov, Itai Gat, Mary Williamson, Gabriel Synnaeve, Juan Pino, Benoit Sagot, and Emmanuel Dupoux. 2024. Spirit LM: Interleaved Spoken and Written Language Model. https://doi.org/10.48550/arXiv.2402.05755 arXiv:2402.05755 [cs]

[110] Abdessalam Ouaazki, Kristoffer Bergram, Juan Carlos Farah, Denis Gillet, and Adrian Holzer. 2024. Generative AI-enabled Conversational Interaction to Support Self-Directed Learning Experiences in Transversal Computational Thinking. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 13. https://doi.org/10.1145/3640794.3665542

[111] Ke Pan, Yew-Soon Ong, Maoguo Gong, Hui Li, A.K. Qin, and Yuan Gao. 2024. Differential Privacy in Deep Learning: A Literature Survey. *Neurocomputing* 589 (July 2024), 127663. https://doi.org/10.1016/j.neucom.2024.127663

[112] Johanna Peltola, Kirsikka Kaipainen, Katariina Keinonen, Noona Kiuru, and Markku Turunen. 2023. Developing a Conversational Interface for an ACT-based Online Program: Understanding Adolescents' Expectations of Conversational Style. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 1. https://doi.org/10.1145/3571884.3597142

[113] Diana Pérez-Marín. 2021. A Review of the Practical Applications of Pedagogic Conversational Agents to Be Used in School and University Classrooms. *Digital* 1, 1 (Jan. 2021), 18–33. https://doi.org/10.3390/digital1010002

[114] Lara Piccolo, Azizah C. Blackwood, Tracie Farrell, and Martino Mensio. 2021. Agents for Fighting Misinformation Spread on Twitter: Design Challenges. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 33. https://doi.org/10.1145/3469595.3469628

[115] David Piorkowski, Rachel Ostrand, Kristina Brimijoin, Jessica He, Erica Albert, and Stephanie Houde. 2024. Towards Interactive Guidance for Writing Training Utterances for Conversational Agents. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 14. https://doi.org/10.1145/3640794.3665553

[116] Alisha Pradhan and Amanda Lazar. 2021. Hey Google, Do You Have a Personality? Designing Personality and Personas for Conversational Agents. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 12. https://doi.org/10.1145/3469595.3469607

[117] Shruti Rao, Valeria Resendez, Abdallah El Ali, and Pablo Cesar. 2022. Ethical Self-Disclosing Voice User Interfaces for Delivery of News. In *Proceedings of the 4th Conference on Conversational User Interfaces (Cui '22)*. Association for Computing Machinery, New York, NY, USA, Article 9. https://doi.org/10.1145/3543829.3544532

[118] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The Human Side of Human-Chatbot Interaction: A Systematic Literature Review of Ten Years of Research on Text-Based Chatbots. *International Journal of Human-Computer Studies* 151 (July 2021), 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

[119] Afsaneh Razi, Layla Bouzoubaa, Aria Pessianzadeh, John S. Seberger, and Rezvaneh Rezapour. 2024. Not a Swiss Army Knife: Academics' Perceptions of Trade-Offs Around Generative Artificial Intelligence Use. https://doi.org/10.48550/ARXIV.2405.00995

[120] Thomas Reitmaier, Electra Wallington, Ondrej Klejch, Dani Kalarikalayil Raju, Nina Markl, Emily Esther Nielsen, Gavin Bailey, Jennifer Pearson, Matt Jones, Peter Bell, and Simon Robinson. 2024. UnMute Toolkit: Speech Interactions Designed with Minoritised Language Speakers. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 56. https://doi.org/10.1145/3640794.3665564

[121] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo. 2021. Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design. *International Journal of Human–Computer Interaction* 37, 1 (Jan. 2021), 81–96. https://doi.org/10.1080/10447318.2020.1807710

[122] Rahime Belen Sağlam and Jason R. C. Nurse. 2020. Is Your Chatbot GDPR Compliant? Open Issues in Agent Design. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 16. https://doi.org/10.1145/3405755.3406131

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

[124] Theresa Schachner, Roman Keller, and Florian V Wangenheim. 2020. Artificial Intelligence-Based Conversational Agents for Chronic Conditions: Systematic Literature Review. *Journal of Medical Internet Research* 22, 9 (Sept. 2020), e20701. https://doi.org/10.2196/20701

[125] Katie Seaborn, Iona Gessinger, Suzuka Yoshida, Benjamin R. Cowan, and Philip R. Doyle. 2024. Cross-Cultural Validation of Partner Models for Voice User Interfaces. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 19. https://doi.org/10.1145/3640794.3665537

[126] William Seymour and Emilee Rader. 2024. Speculating about Multi-User Conversational Interfaces and Llms: What If Chatting Wasn't so Lonely?. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 53. https://doi.org/10.1145/3640794.3665888

[127] William Seymour, Xiao Zhan, Mark Coté, and Jose Such. 2023. A Systematic Review of Ethical Concerns with Voice Assistants. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Montr\'{e}al QC Canada, 131–145. https://doi.org/10.1145/3600211.3604679

[128] William Seymour, Xiao Zhan, Mark Cote, and Jose Such. 2023. Who Are CUIs Really For? Representation and Accessibility in the Conversational User Interface Literature. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. ACM, Eindhoven Netherlands, 1–5. https://doi.org/10.1145/3571884.3603760

[129] William Seymour, Xiao Zhan, Mark Cote, and Jose Such. 2023. Who Are Cuis Really for? Representation and Accessibility in the Conversational User Interface Literature. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 26. https://doi.org/10.1145/3571884.3603760

[130] Mrinank Sharma, Meg Tong, Jesse Mu, Jerry Wei, Jorrit Kruthoff, Scott Goodfriend, Euan Ong, Alwin Peng, Raj Agarwal, Cem Anil, Amanda Askell, Nathan Bailey, Joe Benton, Emma Bluemke, Samuel R. Bowman, Eric Christiansen, Hoagy Cunningham, Andy Dau, Anjali Gopal, Rob Gilson, Logan Graham, Logan Howard, Nimit Kalra, Taesung Lee, Kevin Lin, Peter Lofgren, Francesco Mosconi, Clare O'Hara, Catherine Olsson, Linda Petrini, Samir Rajani, Nikhil Saxena, Alex Silverstein, Tanya Singh, Theodore Sumers, Leonard Tang, Kevin K. Troy, Constantin Weisser, Ruiqi Zhong, Giulio Zhou, Jan Leike, Jared Kaplan, and Ethan Perez. 2025. Constitutional Classifiers: Defending against Universal Jailbreaks across Thousands of Hours of Red Teaming. https://doi.org/10.48550/arXiv.2501.18837 arXiv:2501.18837 [cs]

[131] Suraj Sharma, Joseph Brennan, and Jason Nurse. 2021. StockBabble: A Conversational Financial Agent to Support Stock Market Investors. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 25. https://doi.org/10.1145/3469595.3469620

[132] Lori Shyba and James Tam. 2005. Developing Character Personas and Scenarios: Vital Steps in Theatrical Performance and HCI Goal-Directed Design. In *Proceedings of the 5th Conference on Creativity & Cognition - C&C '05*. ACM Press, London, United Kingdom, 187. https://doi.org/10.1145/1056224.1056251

[133] Olivia Sidoti, Eugenie Park, and Jeffrey Gottfried. 2025. About a Quarter of U.S. Teens Have Used ChatGPT for Schoolwork – Double the Share in 2023. *Pew Research Center* (Jan. 2025).

[134] James Simpson and Deborah Richards. 2021. Taming the Interaction Jungle: Embodiment Is All You Need. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 17. https://doi.org/10.1145/3469595.3469612

[135] Abigail Sticha, Norbert Braunschweiler, Rama Sanand Doddipatla, and Kate M Knill. 2024. Advancing Faithfulness of Large Language Models in Goal-Oriented Dialogue Question Answering. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 32. https://doi.org/10.1145/3640794.3665573

[136] Xin Sun, Isabelle Teljeur, Zhuying Li, and Jos A. Bosch. 2024. Can a Funny Chatbot Make a Difference? Infusing Humor into Conversational Agent for Behavioral Intervention. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 3. https://doi.org/10.1145/3640794.3665555

[137] Selina Jeanne Sutton. 2020. Gender Ambiguous, Not Genderless: Designing Gender in Voice User Interfaces (Vuis) with Sensitivity. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 11. https://doi.org/10.1145/3405755.3406123

[138] Lorainne Tudor Car, Dhakshenya Ardhithy Dhinagaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, and Rifat Atun. 2020. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. *Journal of Medical Internet Research* 22, 8 (Aug. 2020), e17158. https://doi.org/10.2196/17158

[139] Natalia Tyulina, Tatiana Aloi Emmanouil, and Sarah Ita Levitan. 2024. Understanding Linguistic and Visual Factors That Affect Human Trust Perception of Virtual Agents. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 57. https://doi.org/10.1145/3640794.3665581

[140] Michelle M.E. Van Pinxteren, Mark Pluymaekers, and Jos G.A.M. Lemmink. 2020. Human-like Communication in Conversational Agents: A Literature Review and Research Agenda. *Journal of Service Management* 31, 2 (June 2020), 203–225. https://doi.org/10.1108/JOSM-06-2019-0175

[141] Sarah Theres Völkel and Lale Kaya. 2021. Examining User Preference for Agreeableness in Chatbots. In *Proceedings of the 3rd Conference on Conversational User Interfaces (Cui '21)*. Association for Computing Machinery, New York, NY, USA, Article 38. https://doi.org/10.1145/3469595.3469633

[142] Abdul Wali, Saipunidzam Mahamad, and Suziah Sulaiman. 2023. Task Automation Intelligent Agents: A Review. *Future Internet* 15, 6 (May 2023), 196. https://doi.org/10.3390/fi15060196

[143] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-Consistency Improves Chain of Thought Reasoning in Language Models. https://doi.org/10.48550/arXiv.2203.11171 arXiv:2203.11171 [cs]

[144] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. https://doi.org/10.48550/arXiv.2201.11903 arXiv:2201.11903 [cs]

[145] Joseph Weizenbaum. 1966. ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine. *Commun. ACM* 9, 1 (Jan. 1966), 36–45. https://doi.org/10.1145/365153.365168

[146] Joel Wester, Minha Lee, and Niels van Berkel. 2023. Moral Transparency as a Mitigator of Moral Bias in Conversational User Interfaces. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 12. https://doi.org/10.1145/3571884.3603752

[147] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (April 2016), 38–44. https://doi.org/10.1145/2907069

[148] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of Thoughts: Deliberate Problem Solving with Large Language Models. https://doi.org/10.48550/arXiv.2305.10601 arXiv:2305.10601 [cs]

[149] Rui Ye, Wenhao Wang, Jingyi Chai, Dihan Li, Zexi Li, Yinda Xu, Yaxin Du, Yanfeng Wang, and Siheng Chen. 2024. OpenFedLLM: Training Large Language Models on Decentralized Private Data via Federated Learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, Barcelona Spain, 6137–6147. https://doi.org/10.1145/3637528.3671582

[150] Jordyn Young, Laala M Jawara, Diep N Nguyen, Brian Daly, Jina Huh-Yoo, and Afsaneh Razi. 2024. The Role of AI in Peer Support for Young People: A Study of Preferences for Human- and AI-Generated Responses. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. https://doi.org/10.1145/3613904.3642574

[151] Sebastian Zepf, Arijit Gupta, Jan-Peter Krämer, and Wolfgang Minker. 2020. EmpathicSDS: Investigating Lexical and Acoustic Mimicry to Improve Perceived Empathy in Speech Dialogue Systems. In *Proceedings of the 2nd Conference on Conversational User Interfaces (Cui '20)*. Association for Computing Machinery, New York, NY, USA, Article 2. https://doi.org/10.1145/3405755.3406125

[152] Xiao Zhan, William Seymour, and Jose Such. 2024. Beyond Individual Concerns: Multi-user Privacy in Large Language Models. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 34. https://doi.org/10.1145/3640794.3665883

[153] Xiao Zhan, Yifan Xu, and Stefan Sarkadi. 2023. Deceptive AI Ecosystems: The Case of ChatGPT. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 3. https://doi.org/10.1145/3571884.3603754

[154] Minrui Zhang, Eleftherios Papachristos, and Timothy Merritt. 2023. Facilitating Mindful Eating with a Voice Assistant. In *Proceedings of the 5th International Conference on Conversational User Interfaces (Cui '23)*. Association for Computing Machinery, New York, NY, USA, Article 42. https://doi.org/10.1145/3571884.3604311

[155] Zhenqi Zhao, Mariët Theune, Sumit Srivastava, and Daniel Braun. 2024. Exploring Lexical Alignment in a Price Bargain Chatbot. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces (Cui '24)*. Association for Computing Machinery, New York, NY, USA, Article 40. https://doi.org/10.1145/3640794.3665576