Conversational User Interfaces in Healthcare: A Scoping Review of ACM CUI Conference Contributions

Nidhi Kothari
Auckland Bioengineering Institute
The University of Auckland
Auckland, New Zealand
nidhikot1511@gmail.com

Andreia Valente
Auckland Bioengineering Institute
The University of Auckland
Auckland, New Zealand
aval430@aucklanduni.ac.nz

Mark Billinghurst
Auckland Bioengineering Institute
The University of Auckland
Auckland, New Zealand
mark.billinghurst@adelaide.edu.au

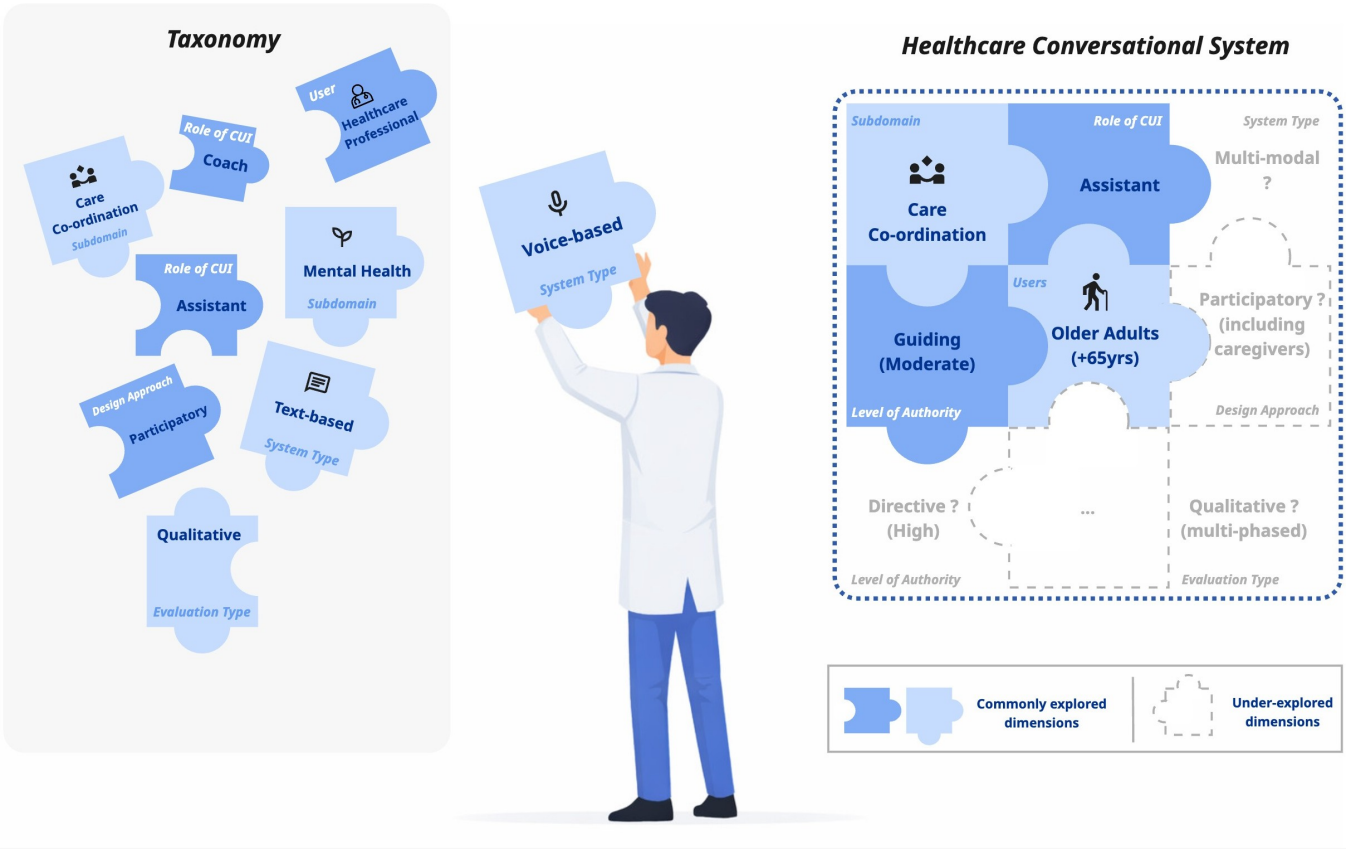


Figure 1: Conceptual overview of the review framework. The puzzle pieces illustrate how healthcare context, target users, modality, system role, authority, and evaluation combine into complete healthcare CUI configurations.

Abstract

Conversational user interfaces (CUIs) are increasingly explored in healthcare, yet their roles, agency, and evaluation practices remain fragmented. This scoping review analyses 22 healthcare-related full papers from the ACM Conference on Conversational User Interfaces (2019–2025) across four pillars: healthcare context, CUI design and interaction, agency, and evaluation. Findings show that healthcare CUIs cluster in prevention and behaviour change, rely on structured conversational strategies, and operate with constrained authority, with no system exercising directive control. Evaluation remains predominantly short-term and usability-focused, with limited attention to clinical or longitudinal outcomes. We outline gaps and directions for designing and assessing healthcare CUIs with greater attention to agency and context.

CUI '26, Bremen, Germany

ACM ISBN 979-8-4007-2741-2/26/07
https://doi.org/10.1145/3816046.3816289

CCS Concepts

• **Human-centered computing** → **Natural language interfaces**;
• **Applied computing** → *Health care information systems.*

Keywords

Conversational User Interfaces, Healthcare, Scoping Review

ACM Reference Format:

Nidhi Kothari, Andreia Valente, and Mark Billinghurst. 2026. Conversational User Interfaces in Healthcare: A Scoping Review of ACM CUI Conference Contributions. In *ACM Conversational User Interfaces 2026 (CUI '26), July 21–24, 2026, Bremen, Germany*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3816046.3816289

1 Introduction

Conversational User Interfaces (CUIs) are becoming a prominent mode of interaction in digital systems. With the growing use of large language model (LLM)-based technologies, CUIs increasingly support interactions that are more flexible, adaptive, and open-ended than earlier task-oriented systems [24, 30, 44]. In healthcare, this shift has fuelled interest in conversational systems for delivering health information, supporting behaviour change, assisting with everyday care, and reducing administrative burden [26, 41], positioning CUIs as part of the digital health ecosystem.

Healthcare is a high-stakes context in which conversational interactions can shape users' understanding, decisions, sense of agency, and trust in care. A cross-dimensional view is needed because decisions about where a healthcare CUI is used, how it communicates, what authority it assumes, and how it is evaluated are interdependent. For example, a CUI used for low-risk wellbeing reflection raises different design and evidence requirements from one that supports triage, clinical explanation, or long-term care. Although healthcare has become a visible theme within the ACM International Conference on Conversational User Interfaces (CUI), there is still no clear account of how this body of work frames healthcare problems, designs interactions, distributes agency, or evaluates success. Prior reviews have mostly focused on specific technologies or health areas, such as voice assistants, mental health, or chronic condition management [6, 36, 40], leaving a gap around venue-specific patterns, priorities, and blind spots within ACM CUI. A scoping review suits this aim because it supports the systematic mapping of heterogeneous and emerging work without requiring the homogeneity assumed by systematic reviews [3, 39].

To address this gap, we conducted a venue-bounded scoping review of 22 full papers published in the ACM CUI proceedings and analysed them through four pillars: healthcare context, CUI design and interaction, agency, and evaluation. We treat ACM CUI as a bounded corpus offering a focused view of how this venue has approached healthcare, rather than as a proxy for all healthcare conversational systems. The aim is not to represent the full CUI field, but to identify how one research community has framed healthcare-related CUI work over time. This boundary allows us to examine community-specific assumptions, priorities, and blind spots that may be diluted in broader cross-venue syntheses. Our analysis shows that healthcare research in ACM CUI concentrates on mental health and behaviour change, while many other contexts remain underexplored. These systems commonly act as guides for reflection, support, and self-management, but differ in how agency is distributed across users, practitioners, and conversational systems. Notably, no system in the corpus exercises directive authority, revealing a constrained design space where CUIs advise but never decide. Evaluation focuses mainly on usability, engagement, and short-term user experience, with limited attention to long-term, clinical, or ecologically valid outcomes. This work contributes (1) a structured overview of healthcare-related research in ACM CUI, (2) the identification of recurring design patterns and gaps, and (3) directions for future healthcare CUI research around agency, evaluation, and the range of contexts addressed.

2 Related Work

Prior research on healthcare CUIs has focused strongly on specific application areas, particularly mental health. Systems such as Woebot, Wysa, Youper, and Phoenix illustrate how conversational agents have delivered structured support through cognitive behavioral therapy, mood tracking, and personalised coping strategies [5, 14, 16, 21]. Other work has extended to particular user groups through multilingual mental health chatbots, or examined interaction-level choices such as voice-based delivery, response structure, and answer formatting [13, 27, 33]. Emerging research has also explored CUIs for healthcare professionals, especially in reducing administrative burden, while raising questions around accuracy, trust, and clinical integration [26].

Prior reviews have typically been organised around particular technologies or specific health domains such as mental health and chronic condition management [6, 36, 40]. More recent work has examined ACM CUI proceedings at a broader level, offering a general overview of conversational interface research within the venue [30]. While such reviews capture the evolution and scope of CUI research, they are not specific to healthcare in terms of system role, interaction design, agency, or evaluation, and rarely engage with how a single research community frames healthcare CUIs over time. This gap motivates a cross-cutting, venue-bounded perspective on healthcare-related CUI research within ACM CUI.

3 Scope

This paper presents a venue-bounded scoping review of full papers published at the ACM Conference on Conversational User Interfaces between 2019 and 2025, examining healthcare-related research within a shared community. To set the boundaries of the review, we clarify what counts as a conversational user interface and how healthcare is understood as an application domain.

Conversational User Interfaces. CUIs are front-end systems that enable interaction through speech, text, touch, and other forms of input and output [28]. Related terms such as chatbots, voice assistants, and conversational agents fall under the broader CUI umbrella, differing in modality and interaction type.

Healthcare as a Domain. We use healthcare as an umbrella domain for conversational systems concerned with health-related needs, outcomes, or care practices. Within this umbrella, we distinguish healthcare as diagnosis, treatment, and illness management from care as ongoing or long-term support, and from wellbeing as broader physical, mental, and social health, consistent with the World Health Organization's definition of health [42]. This distinction was not a strict exclusion boundary. Instead, it guided inclusion and charting decisions by identifying whether each paper was primarily clinical, care-oriented, or wellbeing-oriented. To avoid

excluding relevant work framed outside strictly clinical settings, we included papers situated in healthcare, care, or wellbeing contexts only when the conversational system addressed a health-related need, user group, activity, or outcome.

3.1 Research Questions

Grounded in these definitions, we ask:

RQ1. What healthcare subdomains and stages of care are represented in ACM CUI research?

RQ2. How are healthcare CUIs designed in terms of modality, conversation strategy, and target users?

RQ3. How are roles, agency, and authority distributed between users and conversational systems in healthcare CUIs?

RQ4. How are healthcare CUIs evaluated, and how is user acceptance reported as part of those evaluations?

4 Methodology

We follow the scoping review framework proposed by Arksey and O'Malley [3] and refined by Levac et al. [22], reporting our process in line with the PRISMA Extension for Scoping Reviews (PRISMA-ScR) [39]. The five-stage process comprises identifying research questions, identifying relevant studies, study selection, charting data, and collating and summarising results. No a priori protocol was registered, which we acknowledge as a limitation.

Identification. The search was conducted in March 2026 via the ACM Digital Library, complemented by manual enumeration of each year's proceedings. Of 377 records identified, only full papers were retained, resulting in 133 papers for screening.

Table 1: Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
Full papers published in ACM CUI proceedings (2019–2025)	Workshop, demo, poster, or non-archival papers
Papers situated in healthcare, care, or wellbeing contexts	Papers outside healthcare, care, or wellbeing scope

Screening and eligibility. The 133 full papers were screened in two stages by a single reviewer. First, titles and abstracts were reviewed. A keyword-assisted screening using terms including *health*, *healthcare*, *care*, and *wellbeing* identified 14 candidates, with one excluded due to an ambiguous match (e.g., "care" in "career"). This pilot showed that keyword screening alone could miss papers that did not use explicit health keywords but were framed through specific conditions, life stages, or activities. We therefore conducted a manual pass over all 133 full papers. During this pass, papers were included only when the CUI addressed a health-related need, user group, activity, or outcome. This added 9 papers with implicit healthcare relevance, including work on dementia support, postpartum care, exercise planning, and care coordination. The final corpus contains 22 papers. Figure 2 illustrates the selection process. No eligible papers appeared in 2019, 2020 and 2022, therefore, the included papers come from 2021, 2023, 2024, and 2025. The use of a single reviewer for screening and charting is acknowledged as a limitation in Section 7.

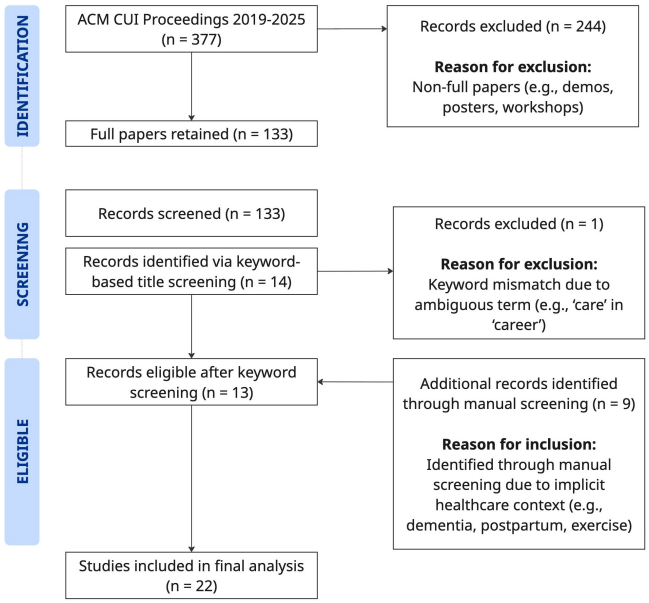


Figure 2: PRISMA-ScR flow diagram for ACM CUI full papers (2019–2025).

Data charting and analysis. A charting form was developed iteratively around four pillars: healthcare context, CUI design and interaction, agency, and evaluation. These pillars capture linked decisions in healthcare CUI design. Healthcare context defines the risk level and care setting, CUI design and interaction describe how the system communicates with users, agency captures how much influence the system has over user decisions, and evaluation shows what evidence is used to validate the system. Analysing these dimensions together allows us to assess whether context, interaction style, authority, and evidence are aligned rather than as separate features. Pillar 1 draws on the World Health Organization's classification of digital health interventions [43]. Pillars 2 and 4 were developed from recurring descriptors in the screened papers, informed by prior CUI taxonomies [11, 19, 29]. Pillar 3 (agency) was developed inductively to capture how systems position themselves in relation to users, an aspect not consistently addressed in frameworks. The form was piloted on 4 papers and refined before being applied to the full corpus. Charting was conducted by a single reviewer. We also conducted an inductive thematic analysis [7] of limitations and challenges following Braun and Clarke's six-phase approach, to surface recurring human-centered and system-centered concerns. A summary table of all studies and their charted attributes can be found in Table 2 and Table 3.

5 Results

We analysed the corpus using two complementary approaches: a taxonomy-driven analysis across the four pillars (healthcare context, CUI design and interaction, agency, and evaluation), and an inductive thematic analysis [7] of recurring human-centered and system-centered challenges.

Table 2: Descriptive overview of the 22 included studies. Country reflects the reported study location. *N* = participants; n pairs = paired studies; n_1, n_2 = multi-phase samples.

ID	First Author	Year	Ref.	Focus	N	Country
1	Kocielnik	2021	[20]	Conversational vs. form-based social-needs screening in clinical settings	41	USA
2	Maharjan	2021	[27]	Pen/paper WHO well-being questionnaire vs. discrete and open-ended CUIs	60	Denmark
3	Peltola	2023	[32]	ACT-based text virtual coach for preventive mental well-being in adolescents	216	Finland
4	Albert	2023	[2]	Voice CUIs in homecare and patient-caregiver interaction	1 pair	UK
5	Desai	2023	[12]	Storytelling VUI for informal health learning in older adults	10	USA
6	Oewel	2023	[31]	Long-term Alexa use by older adults in assisted living	-	USA
7	Lupetti	2023	[25]	Embodied agent for periconception care and trust	25	Netherlands
8	Ahire	2024	[1]	Proactive voice assistant for knowledge workers' well-being	15	Germany
9	Ruitenburg	2024	[34]	Speculative dialogue for couples managing reality disjunction in dementia	6 pairs	Netherlands
10	Cooney	2024	[8]	Practitioners' views on speech agents for mood logging in blended care	15	Ireland
11	Rey	2024	[33]	Voice assistant answer structures for personal health data queries	82	*via Prolific*
12	Kjaerulff	2024	[18]	Voice vs. multimodal CUIs for mindfulness-based smoking cessation	9	Denmark
13	Sun	2024	[38]	RCT on humour in chatbot interventions for physical activity	66	Europe
14	Balaji	2024	[4]	Couples-based chatbot prototype to promote safe sex	23, 4	Netherlands
15	De Souza	2025	[10]	Voice assistants for postpartum mothers' daily routines and care	55	Brazil
16	Samimi	2025	[35]	Visual-conversational decision support for diabetes risk prediction	30	*via Prolific*
17	Czech	2025	[9]	Three-month deployment of a multimodal voice assistant in households	14	UK
18	Liang	2025	[23]	Customisable vs. non-customisable conversational diet recommenders	214	*via Prolific*
19	Hu	2025	[15]	Co-designing voice health interactions with older adults	20	USA
20	Elwahsh	2025	[13]	Co-designing multilingual mental well-being chatbots with mothers	6	UK
21	Shin	2025	[37]	LLM-driven agent for guideline-based exercise plan generation	13, 18	-
22	Joshi	2025	[17]	User preferences for conversational symptom diagnosis	16	-

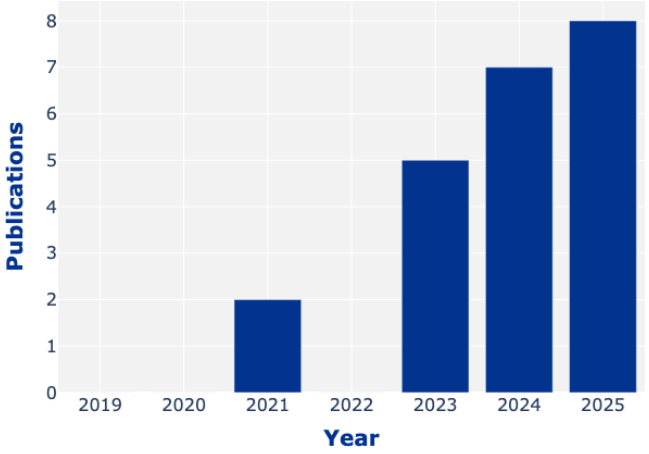


Figure 3: Healthcare-related ACM CUI full papers per year (2019–2025).

5.1 Overview of Healthcare Conversational User Interfaces

The corpus spans 2019 to 2025 (n=0 in 2019 and 2020, n=2 in 2021, n=0 in 2022, n=5 in 2023, n=7 in 2024, n=8 in 2025), with activity rising from 2023 onward as seen in Figure 3.

5.1.1 Pillar 1: Healthcare Context. The first pillar addresses **RQ1** across dimensions: subdomain and stage. Subdomains were adapted from the World Health Organization's digital health intervention categories, covering mental health, health literacy, self-management and behaviour change, screening and triage, clinical decision support, care co-ordination, rehabilitation, and public/community health [43]. Stages include prevention, assessment, treatment or therapy, rehabilitation, and long-term management. Across the corpus, as seen in Figure 4(a) CUIs are most often designed for behaviour intervention (n=7/22, 31.8%), with fewer systems addressing mental health, health literacy, clinical decision support, rehabilitation, or population health. By stage, systems primarily operate in prevention (n=12/22, 54.5%) or long-term management (n=6/22, 27.3%), with few in post-diagnosis treatment or acute clinical contexts. CUIs are therefore concentrated in early-stage, behaviour-oriented contexts rather than deeply embedded clinical applications.

5.1.2 Pillar 2: CUI Design and Interaction. The second pillar addresses **RQ2** by examining who healthcare CUIs are designed for and how, across five categories: target user, age group, modality or system type, conversation strategy, and design approach. CUIs primarily target patients and the general population, with fewer systems designed for caregivers or healthcare professionals, who appear as secondary or dyadic users. This aligns with the prominence of preventive and behaviour-oriented interventions, largely user-facing and self-managed. User groups are predominantly adults and young adults, with fewer focusing on older adults (n=4/22, 18.2%), most of whom (n=3/4, 75%) engage through voice, likely reflecting its hands-free nature. There is limited exploration of users with distinct sensory requirements, such as deaf or blind populations, suggesting most designs assume a broadly able user base.

Despite increasing modality diversity, conversational design remains predominantly structured (n=9/22, 40.9%), including scripted, questionnaire-driven, or task-specific systems. We distinguish structured from generative strategies based on observable interaction behaviour rather than underlying implementation, given inconsistent reporting of AI techniques. Among hybrid approaches (n=5/22, 22.7%), the source of training or response data is often not reported,

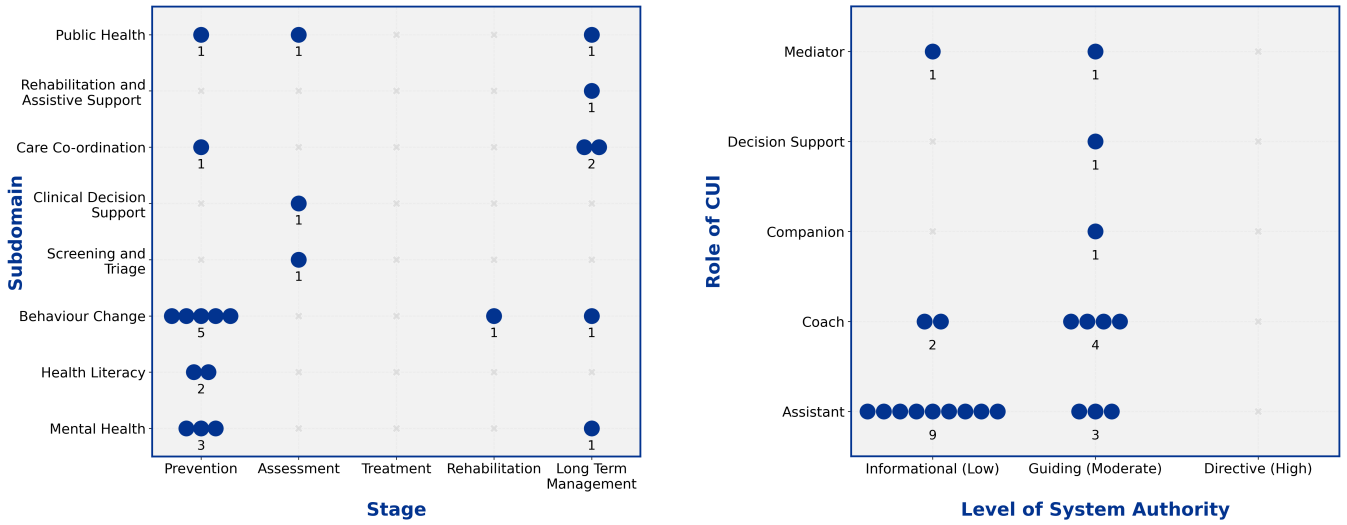
Table 3: Four-pillar charting of the 22 included studies.

ID	Subdomain	Stage	Modality	Strategy	Role	Authority	Evaluation	Acceptance
1	Public/community health	Assessment	Multimodal	Structured	Assistant	Informational	Experimental	Not evaluated
2	Mental health	Prevention	Voice	Hybrid	Assistant	Informational	Experimental	Mixed
3	Mental health	Prevention	Text	Structured	Coach	Guiding	Survey	Mixed
4	Care coordination	Long-term management	Voice	Not specified	Mediator	Informational	Qualitative	Mixed
5	Health literacy	Prevention	Voice	Hybrid	Assistant	Guiding	Qualitative	Positive
6	Public/community health	Long-term management	Voice	Not specified	Companion	Guiding	Qualitative	Mixed
7	Care coordination	Prevention	Embodied	Structured	Assistant	Informational	Qualitative	Mixed
8	Behaviour change & self-management	Prevention	Multimodal	Hybrid	Coach	Guiding	Qualitative	Positive
9	Rehabilitation & assistive	Long-term management	Not specified	Not specified	Mediator	Guiding	Qualitative	Mixed
10	Mental health	Long-term management	Voice	Not specified	Assistant	Informational	Qualitative	Mixed
11	Behaviour change & self-management	Prevention	Voice	Structured	Assistant	Informational	Experimental	Positive
12	Behaviour change & self-management	Rehabilitation	Multimodal	Structured	Coach	Informational	Qualitative	Mixed
13	Behaviour change & self-management	Prevention	Text	Structured	Coach	Informational	Experimental	Mixed
14	Public/community health	Prevention	Text	Not specified	Assistant	Informational	Qualitative	Not evaluated
15	Behaviour change & self-management	Long-term management	Voice	Not specified	Assistant	Informational	Survey	Mixed
16	Clinical decision support	Assessment	Multimodal	Hybrid	Decision support	Guiding	Experimental	Positive
17	Care coordination	Long-term management	Multimodal	Not specified	Assistant	Informational	Qualitative	Mixed
18	Behaviour change & self-management	Prevention	Text	Structured	Assistant	Guiding	Experimental	Positive
19	Health literacy	Prevention	Voice	Not specified	Assistant	Informational	Qualitative	Not evaluated
20	Mental health	Prevention	Text	Structured	Coach	Guiding	Qualitative	Not evaluated
21	Behaviour change & self-management	Prevention	Multimodal	Hybrid	Coach	Guiding	Experimental	Positive
22	Screening & triage	Assessment	Text	Structured	Assistant	Guiding	Qualitative	Not evaluated

raising concerns around explainability and trust. While many studies (n=9/22, 40.9%) did not report their design approach, those that did mostly employed user-centered (n=7/13, 53.8%) or participatory methods (n=6/13, 46.1%), indicating scope for more explicit and context-sensitive frameworks.

5.1.3 Pillar 3: Agency. The third pillar addresses **RQ3** by examining how agency is distributed between users and CUIs along two dimensions: the role of the CUI and its level of system authority. Roles capture how the system is positioned in interaction, such as an assistant, coach, or mediator. Authority reflects how much the CUI influences user decisions: *informational* systems present information without prescribing action, *guiding* systems offer recommendations the user can accept, modify, or ignore, and *directive* systems issue instructions the user is expected to follow or that the system enacts without confirmation. Figure 4(b) maps this relationship and shows a clear alignment between positioning and influence. Assistant-based systems are concentrated at the informational level, while coaching roles more often occupy guiding authority. No system extends into directive authority, indicating CUIs are designed to support and influence behaviour without assuming responsibility for decision-making. Agency is therefore constrained, with systems operating within advisory boundaries rather than as decision-makers.

5.1.4 Pillar 4: Evaluation. The fourth pillar addresses **RQ4** by capturing evaluation approaches and user acceptance as an outcome. Evaluation includes experimental designs, qualitative methods such as interviews, surveys, usability-focused UX assessments, and technical benchmarks. Evaluation is predominantly qualitative (n=13/22, 59.1%), relying heavily on interviews, compared to experimental approaches (n=7/22, 31.8%) and survey-based evaluations (n=2/22, 9.1%). A large subset embedded usability-related tasks within experimental and qualitative designs [27, 34]. Factors examined include usability, engagement, perceived usefulness, motivation, willingness to use, satisfaction, and interaction preferences, with some studies considering behavioural outcomes. Because acceptance was reported through instruments and qualitative indicators, we coded it descriptively as positive, mixed, or not evaluated. Acceptance was reported as positive when users described the system as useful, usable, or desirable for future use, and mixed when studies reported substantial concerns, preferences for alternatives, or divergent responses across participants. Few studies evaluate technical robustness. Only a small number report longer-term or multi-stage evaluations such as follow-ups or phased studies [2, 9, 31], with most remaining short-term or single-session. Direct comparison between acceptance outcomes is limited due to varied metrics. Issues underlying mixed acceptance are discussed in Section 5.2



(a) Subdomain × Stage of Care

(b) CUI Role × System Authority

Figure 4: Healthcare CUI distribution across contexts and role-authority patterns.

5.2 Challenges: A Thematic Analysis

Our thematic analysis (Figure 5) identified eight challenge areas across human-centered and system-centered concerns. On the human-centered side, efficiency and customization issues show that CUIs often increase cognitive burden, require redundant interactions, and offer limited personalization. Rather than adapting to users, systems rely on static behaviours, shifting interactional effort onto users. Response design and naturalness concerns reveal that replies are repetitive, poorly paced, and lacking coherence or empathy, reducing engagement. Trust, transparency, and privacy concerns further highlight users' hesitation to share sensitive health information when data practices or system explainability are unclear.

System-centered challenges point to technological limitations. Voice-based modality constraints restrict flexibility and are prone to recognition errors due to vocabulary, language, and accent variability. System failures are prevalent, with breakdowns in interaction flow and difficulty handling complex or multi-step requests. Limited context awareness restricts effectiveness, as CUIs struggle to track user history or adapt to context, yielding inappropriate or disconnected responses. Multi-user challenges expose the single-user assumption embedded in many systems, making it difficult to support dyadic or triadic use and sometimes disrupting human communication. These findings suggest that healthcare CUIs are constrained by technical limitations and a mismatch with expectations of conversation, care, and context.

6 Discussion

The most striking finding across the four pillars is that no system in this corpus exercises directive authority. Healthcare CUIs in ACM CUI advise, guide, and support, but never decide. This absence should not be read as evidence that decision-making or decision support is impossible in healthcare AI. Rather, it shows that within this ACM CUI corpus, conversational systems have so far been positioned at the boundary between informing and advising users, rather than taking responsibility for decisions. This constrained agency is not incidental. It mirrors a broader pattern in which systems are situated in precautionary contexts such as behaviour change and prevention, operate with low to moderate authority, and rely on structured conversational strategies.

Several barriers may explain why CUIs remain within this advisory boundary. Technically, directive systems require reliable interpretation of symptoms, context, and user history, which is difficult given breakdowns, modality constraints, and limited context awareness identified in our analysis. Ethically, moving toward directive authority raises questions of accountability, safety, consent, and clinician oversight. Interactionally, directive CUIs may undermine user agency if recommendations are presented confidently but without adequate transparency or opportunities for correction.

Keeping CUIs in supportive roles therefore preserves user control and lowers risk, but it also limits what these systems can contribute in scenarios where timely guidance or structured decision support may be valuable. Examples in the corpus, including symptom diagnosis preferences, diabetes risk explanations, and guideline-based exercise planning, point to cases where stronger decision-support roles could be explored [17, 35, 37]. Such systems would require different design and evaluation standards, including clearer escalation pathways, explicit limits of authority, provenance for recommendations, and clinical, longitudinal, or ecologically valid evidence.

This positioning limits integration into clinical workflows and reduces opportunities to explore roles such as decision support or mediation. Limited healthcare professional involvement in design and evaluation reinforces this separation, suggesting that many CUIs are developed at the periphery of formal healthcare infrastructures [4, 35]. Although CUIs are framed as conversational systems, their interaction design remains structured, indicating a preference for predictability over adaptive dialogue.

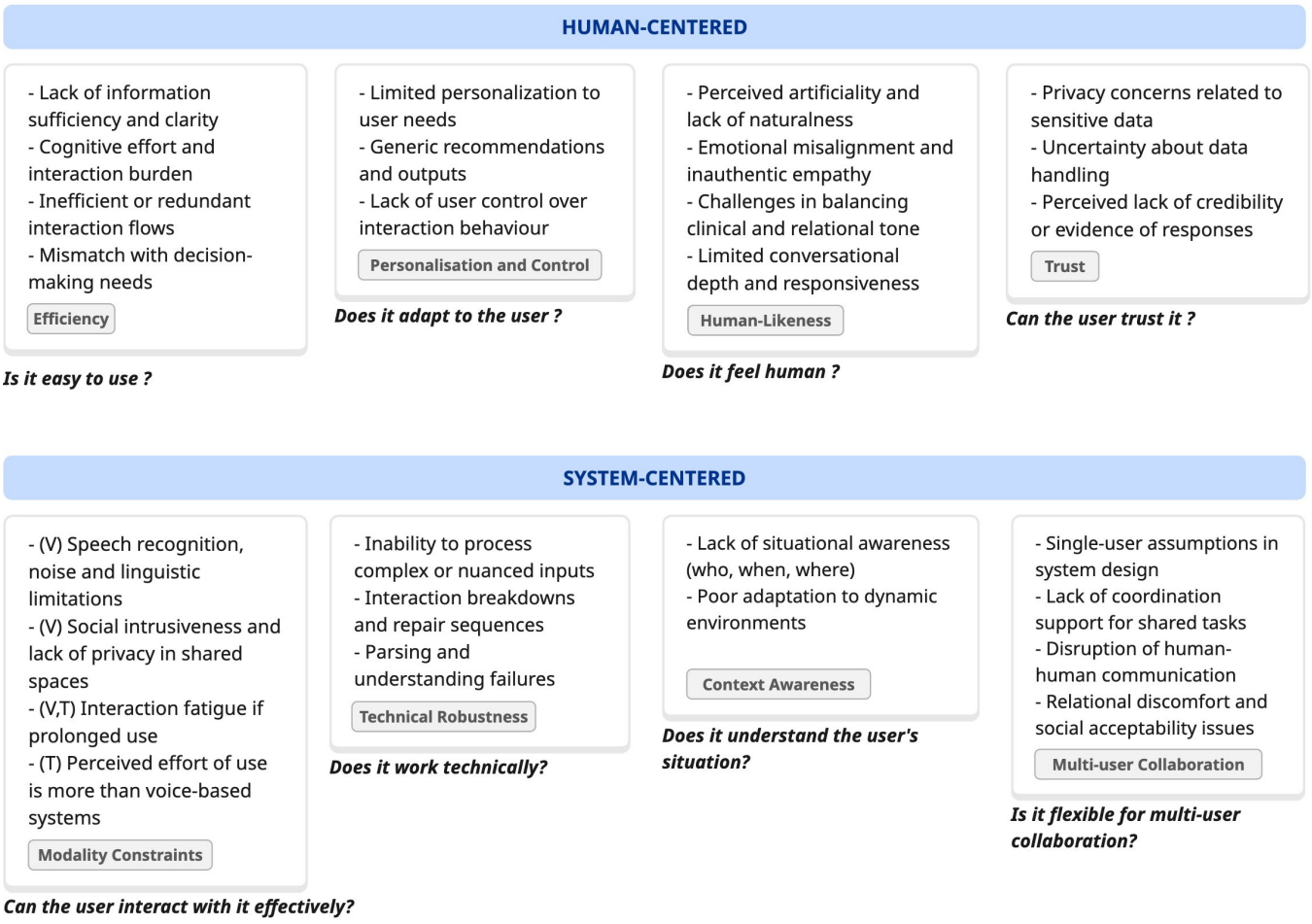


Figure 5: Eight challenge areas from the thematic analysis, grouped into human-centered and system-centered concerns. (V = voice, T = text)

Evaluation approaches are predominantly qualitative, with limited use of standardized or modality-specific frameworks, making findings difficult to compare or assess beyond immediate user experience. Despite deployment in healthcare contexts where reliability and impact matter, evaluations remain largely short-term and usability-focused, pointing to a mismatch between expected system roles and the evidence used to validate them.

The challenge areas identified in Section 5 help explain this mismatch between role, authority, and evaluation. Human-centered concerns such as privacy, transparency, response naturalness, and customization affect whether users can appropriately trust and rely on a healthcare CUI. System-centered concerns such as modality constraints, interaction breakdowns, limited context awareness, and difficulty supporting multi-user situations make higher-authority roles harder to justify. These challenges therefore connect directly to the agency findings. CUIs remain mostly advisory not only because healthcare is high-risk, but also because current systems are not yet consistently evaluated for the reliability, contextual sensitivity, and accountability that more directive roles would require.

Modality exploration is uneven. Voice-based CUIs are common (n=9/22, 40.9%), while embodied and multimodal systems remain underexplored. Cultural representation is skewed towards the Global North, with most studies conducted in the USA, UK, or Europe, and others relying on online recruitment platforms (e.g., Prolific) or not reporting location, limiting the applicability of current design approaches across diverse contexts.

Building on these patterns, we suggest three directions for healthcare CUI research within ACM CUI. First, future work should explicitly position systems on the authority spectrum and justify this relative to the healthcare context, rather than leaving agency implicit. Second, hybrid systems combining structured and generative components should report the source and provenance of their training and response data to support explainability and user trust. Third, evaluation should extend beyond short-term usability and include at least one longitudinal, clinical, or ecologically valid

measure when systems are deployed in care contexts. These shifts would help close the gap between the supportive roles healthcare CUIs are designed to play and the evidence used to validate them.

7 Limitations

This review has several limitations. First, the analysis is restricted to 22 papers from the ACM CUI proceedings, enabling a focused cross-dimensional examination but potentially excluding relevant work from adjacent venues such as CHI, CSCW, IMWUT, or medical informatics outlets. The findings should be interpreted as representative of the ACM CUI community rather than healthcare conversational systems as a whole. Second, screening, charting, and thematic analysis were conducted by a single reviewer, and no a priori protocol was registered, introducing a risk of interpretation bias where reporting was ambiguous. We mitigated this by iteratively developing the charting form and grounding pillar definitions in established frameworks, but inter-rater reliability was not assessed. Third, the review relies on authors' descriptions rather than independent assessment of system performance, limiting conclusions about real-world effectiveness and potentially undercounting under-reported details. Finally, the corpus reflects existing biases, including a concentration of studies in the Global North and a predominance of early-stage systems, which may limit generalizability to broader healthcare settings.

8 Conclusion

This work offers a cross-dimensional view of healthcare CUIs in ACM CUI by examining how they are positioned across context, conversational role, system authority, and evaluation. Role and authority emerge as critical considerations in healthcare settings, where decisions carry varying levels of risk and responsibility. Our findings show that CUIs in this corpus are predominantly designed as supportive systems with constrained authority, with no system exercising directive control. This raises important questions about how agency should be distributed between users and conversational systems in care contexts. Designing safer and more effective healthcare CUIs will require deliberate alignment between context, conversational design, and system authority, alongside evaluation approaches that reflect the complexity of real-world care, including how emerging capabilities can support more adaptive, accountable, and context-aware interactions across patients, clinicians, and caregivers.

References

[1] 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* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 5, 14 pages. https://doi.org/10.1145/3640794.3665561

[2] Saul Albert, Magnus Hamann, and Elizabeth Stokoe. 2023. Conversational User Interfaces in Smart Homecare Interactions: A Conversation Analytic Case Study. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/3571884.3597140

[3] Hilary Arksey and Lisa O'Malley. 2005. Scoping Studies: Towards a Methodological Framework. *International Journal of Social Research Methodology* 8, 1 (feb 2005), 19–32. https://doi.org/10.1080/1364557032000119616

[4] Divyaa Balaji, Gert-Jan De Bruijn, Tibor Bosse, Carolin Ischen, Margot Van Der Goot, and Reinout Wiers. 2024. Designing a Couples-Based Conversational Agent to Promote Safe Sex in New, Young Couples: A User-Centred Design Approach. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 4, 11 pages. https://doi.org/10.1145/3640794.3665556

[5] J. Benita, S. Jaswanth, N. Bhuvaneshwar, R. Yuvaraj, and Y. Lakshmi Narayana. 2025. Phoenix: A Conversational Agent for Emotional Well-Being and Psychological Support. In *2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI)*. IEEE, Erode, India, 1137–1142. https://doi.org/10.1109/ICMSCI62561.2025.10894579

[6] 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. *J Med Internet Res* 23, 3 (29 Mar 2021), e25933. https://doi.org/10.2196/25933

[7] 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

[8] Orla Cooney, Kevin Doherty, Marguerite Barry, David Coyle, Gavin Doherty, and Benjamin R. Cowan. 2024. Using Speech Agents for Mood Logging within Blended Mental Healthcare: Mental Healthcare Practitioners' Perspectives. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 10, 11 pages. https://doi.org/10.1145/3640794.3665540

[9] Elaine Czech, Ewan Soubutts, Ian Craddock, and Aisling Ann O'Kane. 2025. Understanding the Multimodal Voice Assistant as an Informal and Social Care Support Tool in the UK. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 61, 14 pages. https://doi.org/10.1145/3719160.3736605

[10] Jessica de Souza, Chen Chen, Tanya Punater, Ariana Talai, Bárbara Tideman Sartorio Camargo, Ana Carolina Lavio Rocha, Kelly Pereira Coca, and Edward Jay Wang. 2025. Understanding the Challenges and Design Opportunities of Using Voice Assistants to Support Postpartum Mothers in Brazil. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 51, 17 pages. https://doi.org/10.1145/3719160.3736620

[11] Kerstin Denecke and Richard May. 2023. Investigating conversational agents in healthcare: Application of a technical-oriented taxonomy. *Procedia Computer Science* 219 (2023), 1289–1296. https://doi.org/10.1016/j.procs.2023.01.413 CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2022.

[12] Smit Desai, Morgan Lundy, and Jessie Chin. 2023. "A Painless Way to Learn:" Designing an Interactive Storytelling Voice User Interface to Engage Older Adults in Informal Health Information Learning. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 5, 16 pages. https://doi.org/10.1145/3571884.3597141

[13] Sarah Elwahsh, Nora Stern, Aneesha Singh, and Amid Ayobi. 2025. Linguistic Diversity and Mental Well-Being: Co-Designing Custom AI Chatbots with Multilingual Mothers. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 65, 17 pages. https://doi.org/10.1145/3719160.3736615

[14] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health* 4, 2 (06 Jun 2017), e19. https://doi.org/10.2196/mental.7785

[15] Xinhui Hu, Smit Desai, Morgan Lundy, and Jessie Chin. 2025. Beyond Functionality: Co-Designing Voice User Interfaces for Older Adults' Well-being. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 64, 15 pages. https://doi.org/10.1145/3719160.3736613

[16] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. 2018. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR Mhealth Uhealth* 6, 11 (23 Nov 2018), e12106. https://doi.org/10.2196/12106

[17] Rutuja Joshi, Yu-Jou Lee, and Klaus Bengler. 2025. User Preferences in Conversational AI for Healthcare: Insights from an Interview Study. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 67, 13 pages. https://doi.org/10.1145/3719160.3736631

[18] Simon Bak Kjærulff, Stefan Bjerregaard Pedersen, Tummas Johan Sigvardsen, Niels van Berkel, and Eleftherios Papachristos. 2024. Exploring VUI-Supported Mindfulness Techniques for Smoking Cessation. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1, 11 pages. https://doi.org/10.1145/3640794.3665546

[19] Ahmet Baki Kocaballi, Liliana Laranjo, and Enrico Coiera. 2019. Understanding and Measuring User Experience in Conversational Interfaces. *Interacting with Computers* 31, 2 (05 2019), 192–207. https://doi.org/10.1093/iwc/iwz015 arXiv:https://academic.oup.com/iwc/article-pdf/31/2/192/29109111/iwz015.pdf

[20] Rafal Kocielnik, Raina Langevin, James S. George, Shota Akenaga, Amelia Wang, Darwin P. Jones, Alexander Argyle, Callan Fockele, Layla Anderson, Dennis T. Hsieh, Kabir Yadav, Herbert Duber, Gary Hsieh, and Andrea L. Hartzler. 2021. Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 4, 10 pages. https://doi.org/10.1145/3469595.3469599

[21] Kate Sangwon Lee, Janice Yeung, Arianni Kurniawati, and Dillon Tingyang Chou. 2025. Designing Human-Centric AI Mental Health Chatbots: A Case Study of Two Apps. In *Information Systems for Intelligent Systems*, Andres Iglesias, Jungpil Shin, Bharat Patel, and Amit Joshi (Eds.). Springer Nature Singapore, Singapore, 435–452.

[22] Danielle Levac, Heather Colquhoun, and Kelly K. O'Brien. 2010. Scoping Studies: Advancing the Methodology. *Implementation Science* 5, 1 (sep 2010), 69. https://doi.org/10.1186/1748-5908-5-69

[23] Minhui Liang, Jinping Wang, and Yuhan Luo. 2025. SmartEats: Investigating the Effects of Customizable Conversational Agent in Dietary Recommendations. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 63, 16 pages. https://doi.org/10.1145/3719160.3736635

[24] Jiaxi Liu. 2024. ChatGPT: perspectives from human–computer interaction and psychology. *Frontiers in Artificial Intelligence* Volume 7 - 2024 (2024), 1418869. https://doi.org/10.3389/frai.2024.1418869

[25] 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* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 25, 14 pages. https://doi.org/10.1145/3571884.3597128

[26] A Luke MacNeill, Lillian MacNeill, Alison Luke, and Shelley Doucet. 2024. Health Professionals' Views on the Use of Conversational Agents for Health Care: Qualitative Descriptive Study. *J Med Internet Res* 26 (25 Sep 2024), e49387. https://doi.org/10.2196/49387

[27] Raju Maharjan, Darius Adam Rohani, Per Bækgaard, Jakob Bardram, and Kevin Doherty. 2021. Can we talk? Design Implications for the Questionnaire-Driven Self-Report of Health and Wellbeing via Conversational Agent. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 5, 11 pages. https://doi.org/10.1145/3469595.3469600

[28] Michael F. McTear. 2017. The Rise of the Conversational Interface: A New Kid on the Block? In *Future and Emerging Trends in Language Technology: Machine Learning and Big Data*, Jose F. Quesada, Francisco J. Martin Mateos, and Teresa Lopez-Soto (Eds.). Springer, Cham, Switzerland, 38–49. https://doi.org/10.1007/978-3-319-69365-1_3

[29] Quim Motger, Xavier Franch, and Jordi Marco. 2022. Software-Based Dialogue Systems: Survey, Taxonomy, and Challenges. *ACM Comput. Surv.* 55, 5, Article 91 (dec 2022), 42 pages. https://doi.org/10.1145/3527450

[30] 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)*. Association for Computing Machinery, New York, NY, USA, Article 45, 18 pages. https://doi.org/10.1145/3719160.3736621

[31] Bruna Oewel, Tawfiq Ammari, and Robin N. Brewer. 2023. Voice Assistant Use in Long-Term Care. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 24, 10 pages. https://doi.org/10.1145/3571884.3597135

[32] 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* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 1, 16 pages. https://doi.org/10.1145/3571884.3597142

[33] Bradley Rey, Yumiko Sakamoto, Jaisie Sin, and Pourang Irani. 2024. Understanding User Preferences of Voice Assistant Answer Structures for Personal Health Data Queries. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 17, 15 pages. https://doi.org/10.1145/3640794.3665552

[34] Yvon Ruitenburg, Minha Lee, Wijnand IJsselsteijn, and Panos Markopoulos. 2024. Seeking Truth, Comfort, and Connection: How Conversational User Interfaces can help Couples with Dementia Manage Reality Disjunction. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 6, 15 pages. https://doi.org/10.1145/3640794.3665547

[35] Reza Samimi, Aditya Bhattacharya, Lucija Gosak, Gregor Stiglic, and Katrien Verbert. 2025. Visual-Conversational Interface for Evidence-Based Explanation of Diabetes Risk Prediction. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 52, 18 pages. https://doi.org/10.1145/3719160.3736616

[36] Theresa Schachner, Roman Keller, and Florian v Wangenheim. 2020. Artificial Intelligence-Based Conversational Agents for Chronic Conditions: Systematic Literature Review. *J Med Internet Res* 22, 9 (14 Sep 2020), e20701. https://doi.org/10.2196/20701

[37] Donghoon Shin, Gary Hsieh, and Young-Ho Kim. 2025. PlanFitting: Personalized Exercise Planning with Large Language Model-driven Conversational Agent. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 66, 19 pages. https://doi.org/10.1145/3719160.3736607

[38] 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* (Luxembourg, Luxembourg) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 3, 19 pages. https://doi.org/10.1145/3640794.3665555

[39] Andrea C. Tricco, Erin Lillie, Wasifa Zarin, Kelly K. O'Brien, Heather Colquhoun, Danielle Levac, David Moher, Micah D. J. Peters, Tanya Horsley, Laura Weeks, Susanne Hempel, Elie A. Akl, Christine Chang, Jessie McGowan, Lesley Stewart, Lisa Hartling, Adrian Aldcroft, Michael G. Wilson, Chantelle Garritty, Simon Lewin, Christina M. Godfrey, Marilyn T. Macdonald, Etienne V. Langlois, Karla Soares-Weiser, Jo Moriarty, Tammy Clifford, Özge Tunçalp, and Sharon E. Straus. 2018. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine* 169, 7 (oct 2018), 467–473. https://doi.org/10.7326/M18-0850

[40] Lorainne Tudor Car, Daryl A 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 (2020), e17158. https://doi.org/10.2196/17158

[41] Anke Versluis, Kyma Schnoor, Niels H Chavannes, and Esther PWA Talboom-Kamp. 2022. Direct Access for Patients to Diagnostic Testing and Results Using eHealth: Systematic Review on eHealth and Diagnostics. *J Med Internet Res* 24, 1 (12 Jan 2022), e29303. https://doi.org/10.2196/29303

[42] World Health Organization. 1948. Constitution of the World Health Organization. https://www.who.int/about/governance/constitution

[43] World Health Organization. 2023. *Classification of digital interventions, services and applications in health: a shared language to describe the uses of digital technology for health, 2nd ed.* World Health Organization, Geneva, Switzerland. https://www.who.int/publications/i/item/9789240081949

[44] Tom Young, Frank Xing, Vlad Pandelea, Jinjie Ni, and Erik Cambria. 2022. Fusing Task-Oriented and Open-Domain Dialogues in Conversational Agents. *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 10 (Jun. 2022), 11622–11629. https://doi.org/10.1609/aaai.v36i10.21416