

Secondary Publication



Wagner, Nicolas; Kraus, Matthias; Minker, Wolfgang; u. a.

A Survey on Multi-User Conversational Interfaces

Date of secondary publication: 07.08.2026

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-116632x

Primary publication

Wagner, Nicolas; Kraus, Matthias; Minker, Wolfgang; u. a. (2025): A Survey on Multi-User Conversational Interfaces, in: Applied Sciences, Basel: MDPI, Vol. 15, No. 13, 7267, S. 1–29, doi: 10.3390/app15137267.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>

Review

A Survey on Multi-User Conversational Interfaces

Nicolas Wagner ^{1,2,*} , Matthias Kraus ³ , Wolfgang Minker ¹ , David Griol ²  and Zoraida Callejas ² 

- ¹ Dialogue Systems Group, Ulm University, Albert-Einstein-Allee 43, 89081 Ulm, Germany; wolfgang.minker@uni-ulm.de
- ² Departamento de Lenguajes y Sistemas Informáticos, E.T.S. de Ingenierías Informática y de Telecomunicación, Universidad de Granada, C/Periodista Daniel Saucedo Aranda S/N, 18071 Granada, Spain; dgriol@ugr.es (D.G.); zoraida@ugr.es (Z.C.)
- ³ Chair for Human-Centered Artificial Intelligence, Faculty of Applied Computer Science, University of Augsburg, Universitätsstraße 6, 86159 Augsburg, Germany; matthias.kraus@uni-a.de
- * Correspondence: wagner@correo.ugr.es
- [†] Current address: Natural Language Generation and Dialogue Systems Group, University of Bamberg, Kapuzinerstraße 16, 96047 Bamberg, Germany; nicolas.wagner@uni-bamberg.de.

Abstract

This paper investigates the evolving landscape of Multi-User Conversational Interfaces, addressing the limitations of traditional systems that primarily focus on single-user interactions. As real-world applications grow more complex, there is a pressing need for conversational systems capable of facilitating dialogues among multiple participants. Our systematic survey reviews recent advancements in the field, highlighting innovative architectures, application domains, user and dialogue modelling, and evaluation metrics tailored for multi-user contexts. We conduct a comprehensive analysis of relevant literature, employing both quantitative and qualitative methodologies to identify common patterns and challenges in multi-user interactions. The findings underscore the importance of developing robust interfaces that can effectively manage overlapping dialogues, ensure collaborative group work, and enhance overall conversational quality. This work contributes to the understanding of Multi-User Conversational Interfaces and may help as a basis for future research aiming to develop more natural, user-friendly, and effective conversational interfaces.



Academic Editor: Rui Araújo

Received: 6 May 2025

Revised: 11 June 2025

Accepted: 25 June 2025

Published: 27 June 2025

Citation: Wagner, N.; Kraus, M.; Minker, W.; Griol, D.; Callejas, Z. A Survey on Multi-User Conversational Interfaces. *Appl. Sci.* **2025**, *15*, 7267. <https://doi.org/10.3390/app15137267>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: multi-user interaction; multi-user evaluation; multi-user robotic agents; human–computer interaction; dialogue systems

1. Introduction

The vision of enabling humans to communicate with machines existed well before the past century. However, it was the development of advanced computers and processing methods that made the transition from concepts to reality possible. Lately, remarkable progress has been made in allowing one single user to access computer applications via Conversational User Interfaces (CUIs), yet most of contemporary work neglects an important aspect of natural human behaviour—interactions in groups. The goal of this paper is therefore to provide a comprehensive overview of research on CUIs that focus on interactions with multiple users simultaneously.

Research on CUIs aims to develop systems which simulate human-like interactions through natural language-processing techniques, enabling intuitive dialogues between humans and machines. In literature, there exist various terms to describe such a system: dialogue system, conversational agent, or chatbot, just to name a few. Although each

term is associated with a specific definition, they are often used interchangeably within the research community. In this work, a CUI is referred to as a user interface that allows its users to interact with a computer application using natural language dialogue, either via text, spoken language, or a combination of both. Figure 1 displays the traditional pipeline architecture of single-user CUIs, which consisted of 5 distinct modules responsible for completing domain-dependent tasks. First, the Automatic Speech Recognition (ASR) converts speech to text to enable spoken language interactions. In a next processing step, the Natural Language Understanding (NLU) module analyses the input text to extract meaning and semantic information, such as intent of the user and relevant entities. Subsequently, the dialogue management decides the next action or response in accordance with its dialogue policy by considering the current state of the conversation, the user intent combined with their preferences, and the context of the conversation. This component manages the flow of the conversation and also decides whether to access external services or knowledge sources relevant for the task, defining the behaviour of the system in a conversation. The system action (e.g., request for more relevant information, or details about ongoing tasks) is later processed by Natural Language Generation (NLG), which is in charge of transforming the semantic system actions back into natural language. In cases where a speech-based interface is used, the last step of the pipeline is the Text-To-Speech (TTS) component that converts the textual response into a speech signal. Contemporary systems also aim to employ end-to-end training to replace the modular architecture, or hybrid approaches to maintain control over the dialogue sequence and generations.

In contrast to traditional single-user interfaces, where the focus is on one-to-one interactions between a user and a system, Multi-User Conversational Interfaces (MUCIs) are required to interpret input and handle the conversation for several users at the same time. This competence introduces an additional level of complexity, which makes it necessary to equip MUCIs with skills for turn-taking between the system and different users, context awareness of all the implicated parties, and in-group dynamics, as can be seen in Figure 2.

To maintain a coherent dialogue flow, the system must be able to determine when it is addressed, whether it should respond, and to which user. Since these requirements pose major challenges for the development and implementation, it is essential to intensify research, enabling the components to recognise context within group interactions to appropriately manage the dialogue flow. However, apart from limited resources such as computing power or suitable training data, there is another complicating factor: the guidelines for the design of MUCIs are largely undefined or derived from concepts of human-machine interaction for a single user, which is impractical for many applications. Due to this, MUCIs are not implemented in a standardised way, which impedes further development and improvement of existing systems. As a consequence, the capabilities of contemporary systems are typically limited to specific use-cases, which may lead to an inadequate user experience. However, in this survey paper, we aim to identify common patterns and design guidelines to move towards more standardised approaches. Hence, the contributions of this survey paper are:

- Development of a survey paper following the PRISMA flow method [1]. Following this approach, we filter 181 articles, analysing in detail 59 final candidates. From these candidates, we provide an analysis of the venues and journals in which they were published. Additionally, we outline possible existing connections between authors and countries of the analysed collection. Although dispersed, we observe a growing community and indicate several networks of potential collaborators.
- Proposal of a definition of the MUCIs term based on related works and renowned dictionaries, given the lack of consensus in the literature.

- Identification of the most important aspects of contemporary MUCIs by highlighting patterns and common design approaches. To create a broad understanding, the relevant publications of the sample are categorised into four sections: Application Domains and Examples, User and Dialogue Modelling, Multi-Modality, and Evaluation Methods.
- Provision of guidelines for future implementations of MUCIs and their current challenges to address.

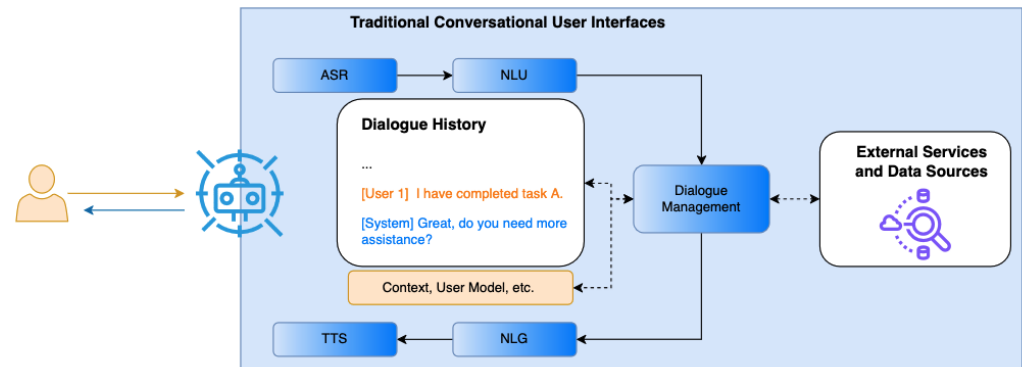


Figure 1. Traditional pipeline architecture of a conversational interface.

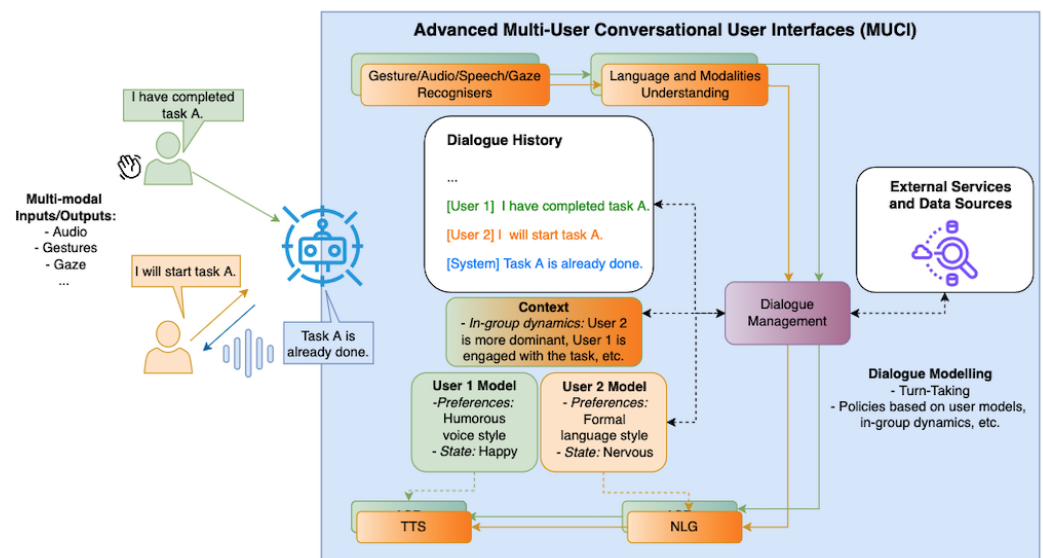


Figure 2. Modular architecture of Multi-User Conversational Interfaces and their components.

The paper is structured as follows: In Section 2, we define the scope and methodology of our survey, including the literature search strategy and its inclusion and exclusion criteria. Section 3 presents a comprehensive quantitative analysis of the relevant literature. In Section 4, we provide a detailed examination of the publications, organised by research fields and associated challenges. The key findings are subsequently discussed in Section 5, followed by a conclusion and an outlook on MUCIs in Section 6.

2. Scope and Methodology

As a starting point for our survey, we performed an initial search regarding MUCIs. The results revealed that the topic is highly relevant and that there exists considerable literature on the subject. However, it became clear that there is little correlation between the papers, and a cohesive research community could not be identified. The aim of this paper is therefore to give an overview of the research conducted on MUCIs by first providing a

common definition of the term, then examining the methods and approaches used, and finally a comparison of key findings and guidelines for future systems.

We conducted our literature review according to the PRISMA standards for systematic analysis [1]. For this, the publisher-independent global citation database ‘Web of Science’ (WoS) was used as a digital library. We chose WoS as our primary database since it is widely recognised for its strong indexing standards and comprehensive coverage of high-impact journals and conferences in computer science and engineering (including more than 34,865 journals, books, proceedings, patents, and datasets). Although no single database can claim exhaustive coverage, WoS offers several advantages: it ensures a baseline of quality by enabling manual adjustments of inclusion criteria, provides reliable citation data for impact analysis, and is frequently used in systematic reviews to identify research trends. Additional significant publications known to the authors were also included. After identifying keywords through a pilot search, we were able to determine a search string which is shown in Table 1. Our survey considers all results with a publication date between 1 October 2012 and 1 May 2025. This decision was made on the observation that significant research interest around MUCIs began to emerge in the early 2010s. By starting our review in 2012, we ensured coverage of the most relevant publications considering the rapid advances of conversational Artificial Intelligence (AI) in the last decade and reflecting the most current literature available at the time of this survey. Furthermore, we assumed that if techniques and methodologies published before 2012 are still in use in contemporary research, their approaches would be referenced or built upon in subsequent literature, and thus captured through our analysis of more recent works.

Table 1. Search query on Web of Science, retrieved on 1 May 2025 at 17:00 CET.

(TS = (Multi-user dialog* system*) OR TS = (Multi-party dialog* system*)
OR
TS = (Multi-user NEAR/15 (“conversation* agent” OR “user interface”
OR
chatbot* OR robot* OR multi-modal* OR assistan*))
OR
TS = (Multi-user NEAR/2 evaluation)) NOT TS=(network OR MIMO)

The initial sample contained 186 publications. After a more detailed analysis, the papers that were unrelated or had only marginal relation to the topic of MUCIs were removed. The remaining articles were systematically and carefully reviewed to minimise bias and provide reliable results for the survey. We had to remove three duplicates and another publication which was not in English. Two publications were not accessible due to being hosted by non-open-access publishers, which prevented us from including them in the detailed analysis. Regarding the inclusion criteria, we considered only those articles that offered a description of the design process, implementation, or empirical evaluation of a MUCI. Eligible publications were required to present either a (prototypical) dialogue interface or to position multi-user human–machine interactions as the central aspect of the investigation. As a result, 121 retrieved articles were excluded because they did not meet these requirements. We identified three primary criteria for exclusion: absence of a conversational interface, insufficient support of multi-user interaction, or an unrelated thematic focus on network technology. More specifically, 61 articles were excluded because they described interfaces that did not support natural language dialogues between users and the system, relying instead on alternative interaction modalities such as graphical controls or keyboard input. Although managing collaborative tasks and shared workspaces is central to group work and there exists overlap to the field of ‘Computer Supported Collaborative Work’, our survey specifically targets multi-user conversational interaction

and thus we excluded publications that did not incorporate interfaces with language capabilities as a core component. In addition, our focus was on multi-user interaction according to the definition provided in Section 4, which requires that two or more users are able to interact with the same system. Publications ($n = 34$) that addressed only single-user scenarios or did not investigate multi-user interactions were therefore excluded. Finally, to maintain thematic focus, we excluded 26 papers that primarily discussed technological advances in (mobile) networks, infrastructure, or protocols, as these did not provide substantive insights into conversational interface design or usage. The final working sample contained 59 articles, Figure 3 shows the process flow of the working sample.

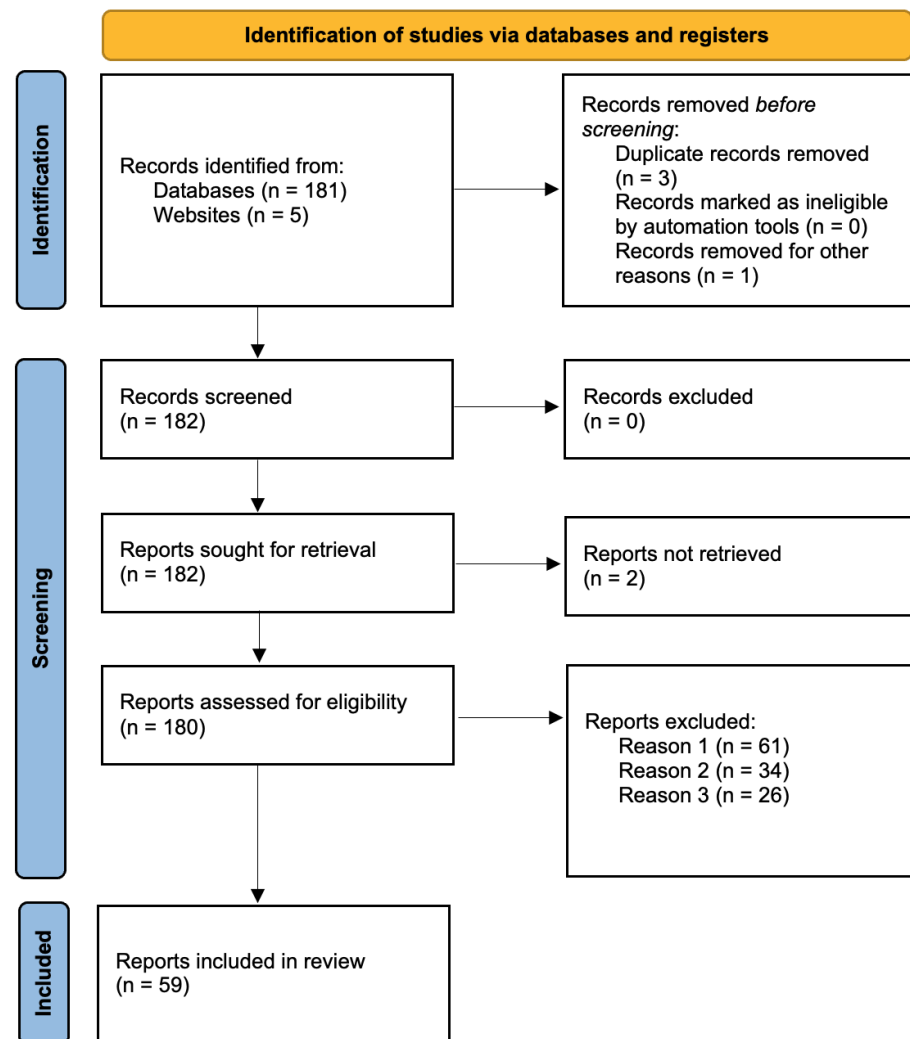


Figure 3. Visualisation of the PRISMA flow diagram.

3. Quantitative Paper Analysis

In this section, we provide details on the relevant venues identified from the filtered articles, as well as the key players and collaboration networks discovered in the context of MUCIs using cluster and graph-based approaches.

3.1. Relevance of Conferences and Journals

After determining all relevant publications, we moved to a second stage to analyse the samples quantitatively. As indicated, there is a wide variety of sources in which the papers in the topic are published. When taking a closer look at them divided into conference proceedings and journals, it can be observed that the 42 conference papers

have been published in a wide variety of sources (26 conferences), so there is no single big event that gathers the majority of publications in the area. All conferences that have published more than one paper from our sample are listed in Table 2. As can be observed, the ones with more papers are related to robotics, in particular the conferences which have published more documents from our sample are in the cross-roads of robotics and human-machine interaction (IEEE RO-MAN) or located in Conversational User Interfaces (CUI). These are followed by the conferences focused on speech and language processing and conversational interaction, those on human-computer multi-modal interaction and computer graphics, and finally conferences related to ambient intelligence and intelligent environments. Additionally, in Figure 4, it can be observed which publishers of the conferences were involved per year. Here, the main publishers are ACM, ACL, and IEEE, which appear mostly across all the years, highlighting that their databases also contain relevant articles related to MUCIs field.

Table 2. Conferences of the sample with more than one relevant publication.

Conference	Number of Papers
IEEE International Symposium on Robot and Human Interactive Communication (IEEE RO-MAN)	6
Conversational User Interfaces (CUI)	4
Annual Meeting of the Association for Computational Linguistics (ACL)	4
International Conference on Robotics and Automation (ICRA)	2
International Conference on Intelligent Environments (IE)	2
Text, Speech and Dialogue (TSD)	2
International Conference on Intelligent Robots and Systems (IROS)	2
ACM International Conference on Multi-Modal Interaction (ICMI)	2
Augmented Reality, Virtual Reality, and Computer Graphics (AVR)	2

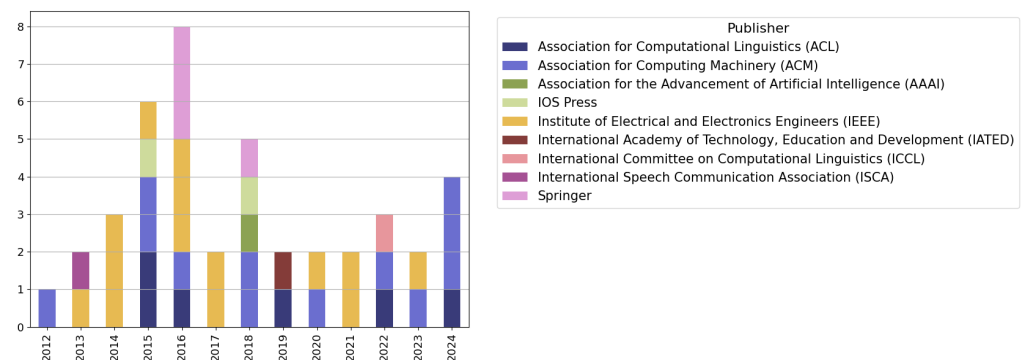


Figure 4. Conference articles ordered by publisher and year. The X-axis represents the publication year and the Y-axis the number of articles. Colours refer to the publisher displayed in the legend.

The scenario is quite similar when focusing on journals, the 16 journal papers and 1 book chapter, which have been published in 14 different journals, excluding the book chapter. Only two journals have published more than one document from our sample: ‘Frontiers in Robotics and AI’ with 2 papers, and ‘International Journal of Social Robotics’ with another 2. As commented before, most journals are related to the topic of robotics (e.g., Journal of Human-Robot Interaction, International Journal of Social Robotics, or International Journal of Humanoid Robotics), human-computer interaction, conversation, speech, and language (e.g., International Journal of Human-Computer Studies, or Computer Assisted Language Learning).

In context of a high-level analysis addressing the publishers of the 16 journals and the book chapter, Figure 5 reveals that for the last years (2021 to 2025), Elsevier is achieving a higher prominence versus the previous dominance of Springer in the period of 2016–2020.

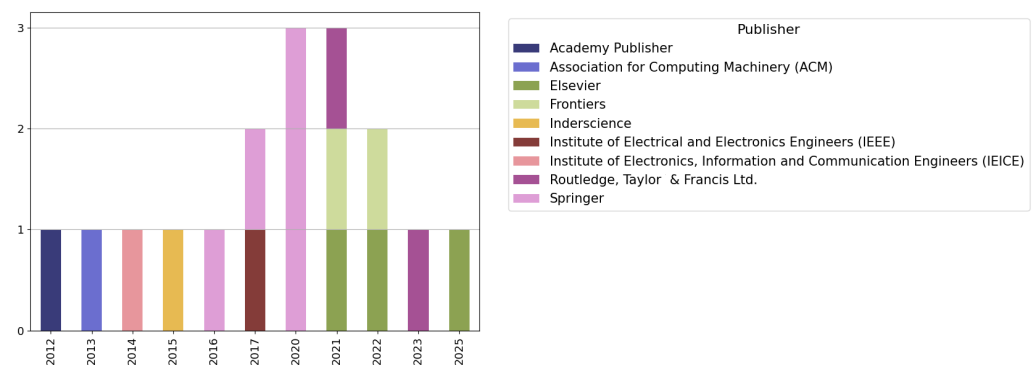


Figure 5. Journal articles ordered by publisher and year. The X-axis represents the publication year and the Y-axis the number of articles. Colours refer to the publisher displayed in the legend.

For exploring the most relevant articles in terms of bibliometrics (e.g., number of citations, downloads, or cited articles), we use the ResearchRabbit tool [2]. This tool is a publicly available AI-powered research platform that allows academic literature discovery and analysis through interactive graph visualisation and citation-based mapping tools.

Figure A1, listed in the Appendix A, shows a graph with the most relevant articles detected per year from the collection of 59 included publications. In this graph, each node represents one article with the surname of the last author, and the connections display the citation network between articles. The diameter of each node indicates the impact of the respective publication. As can be seen, some of the most impactful publications are the works supervised by G. Skantze in 2015 [3] and 2020 [4], as well as J. Beskow in 2015 [5]. From the connections, it can be derived that the more recent work of Skantze from 2021 [4] references to the previous publication in 2015 [3], which indicates an active continuation of the research line on MUCIs in this research group.

3.2. Networks of Research Collaborations

To analyse possible collaborations between researchers and countries, we again employed the ResearchRabbit tool. Based on the collection of 59 articles in this survey, possible existence of connections between 56 authors were identified, as shown in Figure 6.

On the top upper part, we can observe that O. Lemon (Heriot-Watt University, Edinburgh, UK) has one of the largest collaboration networks including S. Keizer (Cambridge Research Lab, Toshiba Europe Ltd., Cambridge, UK), M. E. Foster (University of Glasgow, Glasgow, UK), A. Isard (University of Edinburgh, Edinburgh, UK), A. Gaschler (Technische Universität München, München, Germany), K. Georgila (University of Southern California, Los Angeles, CA, USA), D. Traum (University of Southern California, Los Angeles, CA, USA), J-M. Odobez (Idiap Research Institute, Martigny, Switzerland), P. Motlicek (Idiap Research Institute, Martigny, Switzerland), O. Canévet (Idiap Research Institute, Martigny, Switzerland). In this cluster, there is a relationship between institutions of Germany-UK-Switzerland-USA.

On the right side of the graph, we can also observe the extended cluster of J. Fan (Vanderbilt University, Nashville, TN, USA), L. C. Mion (Ohio State University, Columbia, OH, USA), N. Sarkar (Vanderbilt University, Nashville, TN, USA), including L. Beuscher (Vanderbilt University, Nashville, TN, USA), and P. A. Newhouse (Vanderbilt University, Nashville, TN, USA). This cluster may indicate potential collaborations between three major researchers (N. Sarkar, P. A. Newhouse and L. C. Mion), all inside the USA.

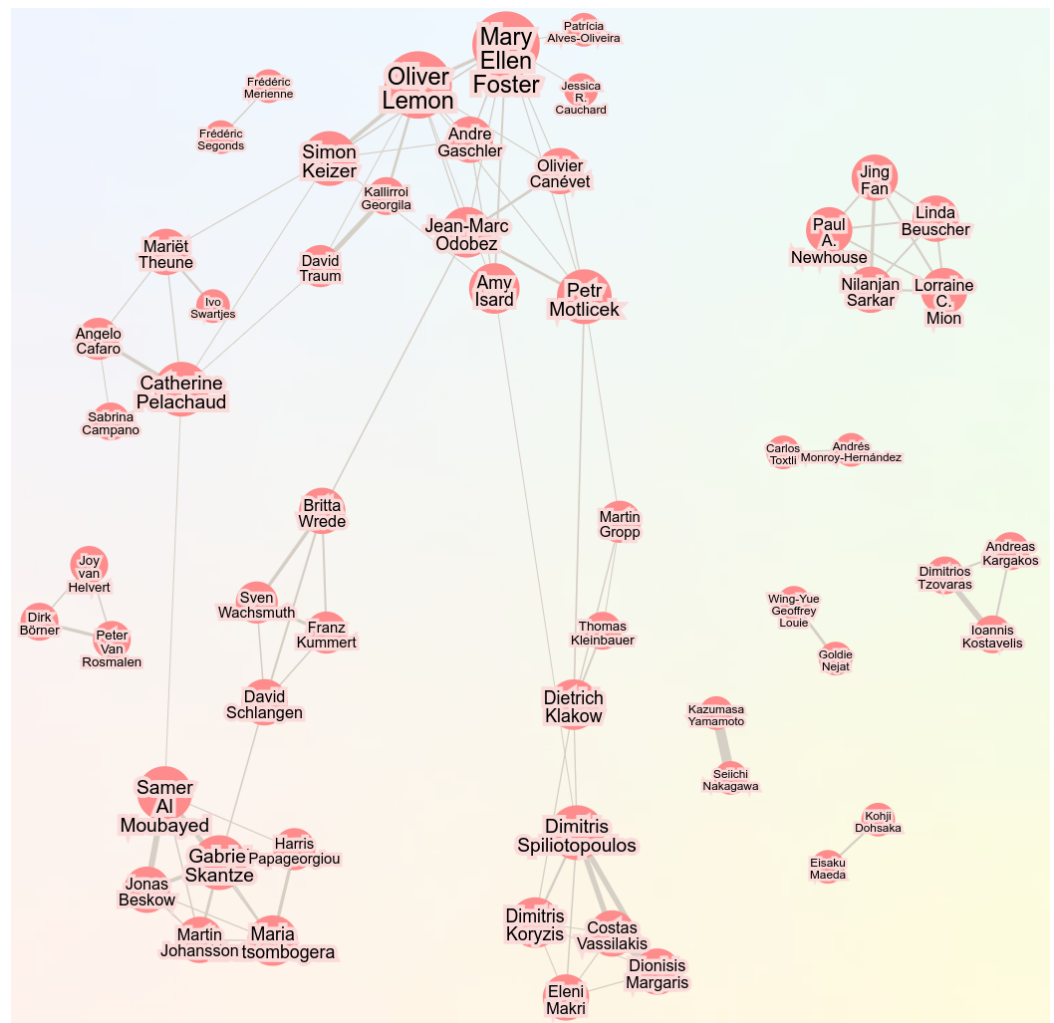


Figure 6. Distribution of the most relevant connections between 56 researchers of different institutions from the total of 435 authors and co-authors discovered by the ResearchRabbit tool.

Appearing at the mid-lower area of the figure, the relationship between S. Nakagawa (Toyohashi University of Technology, Japan), and K. Yamamoto (Chubu University, Kasugai, Japan), both based in Japan, illustrate an example of inter-institutional cooperation. In close proximity to this cluster, the collaboration between G. Nejat and W. Y. G. Louie represents another significant academic partnership.

In the lower central area, a distinct collaborative cluster can be identified, comprising C. Vassilakis, D. Spiliotopoulos, D. Koryzis, E. Makri (all affiliated with University of the Peloponnese, Greece), and D. Margaris (University of Athens, Greece). Within this cluster, D. Spiliotopoulos extends the connections beyond the national level, specifically to D. Klakow (Saarland University, Germany) and A. Isard (University of Edinburgh, UK), thereby linking regional and international research efforts.

The final cluster analysed highlights the collaborative network led by C. Pelachaud (Sorbonne University, France), which spans multiple countries. This cluster demonstrates active connections with researchers from Sweden (S. Al Moubayed, KTH Royal Institute of Technology), the United States (D. Traum, University of Southern California), the Netherlands (M. Theune, University of Twente) and United Kingdom (S. Keizer).

In conclusion, while collaborations between research groups do exist, the connections between them remain limited and underdeveloped. This suggests that, although some researchers operate across international collaborations, several clusters remain isolated from the main network, indicating the absence of a large and well-structured research

community on MUCIs, which is still growing. However, with this analysis we aim to expand this community and detect key researchers interested in the development of MUCIs.

4. Capturing Multi-User Conversational Interfaces

In this section, we provide an entry point for understanding approaches to state-of-the-art MUCI research. We outline and discuss the key aspects that characterise the prototypes and frameworks presented.

In the reviewed literature, the term Multi-User (MU) was used inconsistently, leading to ambiguity in its interpretation. For instance, in some studies, ‘Multi-User’ referred to an interface designed to sequentially engage with multiple users, rather than supporting simultaneous interactions [6,7]. This variation in terminology not only complicates comparative analysis, but also shows the need for clearer definitions and standardised usage within the field. Therefore, in this article, we refer to the definition according to renowned dictionaries such as Merriam-Webster (<https://www.merriam-webster.com/dictionary/multiuser>—last visit to website 1 May 2025): [the ability] “to be used by more than one person simultaneously”. This differs in particular from the field of Multi-Party CUIs where multiple system agents may participate in conversations [8]. As this can be seen as an extension, the focus of our analysis is primarily on scenarios in which multiple users interact with the same CUIs.

In this regard, research on MUCIs is centred on developing technologies that enable natural interaction between a conversational interface and multiple users simultaneously [9]. MUCIs have a wide range of applications, including virtual assistants, customer service chatbots, educational platforms, and collaborative or social robots. However, the design and implementation of MUCIs involve challenging tasks that cover a broad spectrum of technical aspects. The aim of this paper is thus to give an overview of the challenges and implications of patterns within the literature. For a better understanding of conversation analysis in MU scenarios, we recommend reading the survey of Gu et al. [8].

One of the primary challenges in MUCIs is the need to support multiple users with different conversational styles and preferences. In contrast to traditional conversational agents, it is not sufficient to just adapt to the preferences of a single user, but to join several of them. To achieve this, it seems to be necessary to develop algorithms that can process and analyse large amounts of conversational data in real-time. This includes developing techniques for detecting and classifying different conversational patterns, such as turn-taking, backchannelling, and topic changes, as well as off-topic talk awareness [10,11].

Another key challenge in capturing MUCIs is to design systems that are able to adapt to user needs and preferences. To achieve this, user-centred evaluation methods have to be developed that can assess the quality and effectiveness of MUCIs from the users’ perspective. This involves collecting user feedback and using it to iteratively enhance the system’s design, functionality, and performance [12,13].

Additionally, MUCIs often require the integration of hardware and software components. For instance, in robotics applications, researchers have to implement algorithms that can control both the robot’s hardware and software to enable natural and intuitive interactions with multiple users. This involves developing techniques for speech recognition, facial and gaze recognition, gesture recognition, and natural language processing.

Despite the challenges, research on MUCIs has made significant progress in recent years, with several novel systems and prototypes being developed and evaluated. In the following, we will present the innovations presented in the reviewed papers divided into four main sections: Application Domains and Examples, User and Dialogue Modelling, Multi-Modality, and Evaluation Methods.

4.1. Application Domains and Examples

As commented before, Multi-User Conversational Interfaces are often dependent on the application domain. For this reason, we created a first distribution of the reviewed articles in the context of their applications to understand the requirements and complexities associated with each scenario.

4.1.1. Health Care

Concerning an application in the health care domain, Snaith et al. [14] describe a multi-user dialogue game for achieving behaviour change in a patient by applying a goal-setting strategy. Here, two or more virtual health coaches have a conversation with a patient to reach agreement on a behaviour-change goal. For example, the agents and a patient may engage in a discussion to nudge the patient towards more healthy nourishment and apply different conversational strategies on how to achieve this. Although the author's work presents an interesting use case for applications in health care, they only provide a conceptual framework of the dialogue game without the implementation of an actual prototype.

4.1.2. Social Conversation

Todo et al. [15] present a spoken dialogue system prototype that can handle small talk between two agents having different characteristics and one human user. The subject of the small talk addresses the liking and disliking of different dishes, hence, making the agents' behaviour similar to human characteristics to study the interaction dynamics. During the dialogue, the agents can express opinions about specific dishes, e.g., "I really like Miso Ramen.", and engage in a conversation with users on their preferences. In doing so, the authors intended to create a more engaging conversation. Shibahara et al. [16] extended the system to be able to change the topic of small talk to create more variation in the dialogues.

Moreover, Dohsaka et al. [13] present a quiz-style multi-party dialogue system in which two artificial agents and two users engage in a conversation. One of the agents acted as the quiz-master, while the other agent acted as a peer which was engaging in finding the answer together with the human users. The goal of the quiz was to guess the name of a famous person. The quiz-master continuously provides hints until one of the human players or the peer agent finds the correct answers. The interface of the game was designed as a chat-like interface where users could type in their guesses.

Clavel et al. [17] developed a sophisticated virtual agent framework that was able to handle social interactions between one user and multiple embodied agents. Here, the user was represented as an avatar in a 3D serious game environment. The game's goal was to successfully join a group of four agents (two males, two females) where the agents expressed various degrees of social behaviour (in-group vs. out-of-group attitudes). Even though the agents could express various interaction strategies, the conversational aspect was quite limited in the context of the author's application.

In the realm of human-robot interaction, collaborative game-like scenarios have been explored, in which two users and a robot engage in discussions to find solutions [11]. The two visitors and a Furhat robot are given the task of sorting the cards according to specified criteria. For example, the task could be to sort a set of inventions in the order they were invented, or a set of animals based on how fast they can run. This is a collaborative game in which visitors have to discuss and find the solution together with a Furhat robot head.

Liu et al. [18] present a multi-party chatbot that can handle conversations based on logs from Ubuntu-related chat rooms, where multiple users discuss topics related to technical support with Ubuntu issues. Based on this Ubuntu-corpus, an annotated dataset intended for use in multi-user dialogue systems has been developed by Ouchi et al. [19], where

the authors included information about the addressee and response selection. For the realisation of the extraction model, the authors applied an end-to-end approach. Similarly, Ju et al. [20] describe a multi-party chatbot that can take the role of a character of a specific TV show and communicate with multiple other agents representing one of the TV show's characters depending on its relation to the other dialogue participants.

4.1.3. Service Robots

In Lee et al. [10], the authors present a dialogue-based robot service that can provide relevant services to multiple users simultaneously utilising a service-scheduling scheme. Although the available services are not explicitly mentioned in the paper, the author presents an example use case where the robot can provide weather information. Moreover, Keizer et al. [21,22] describe scenarios in which a robot bartender is designed to handle tasks such as taking customer orders and serving drinks. It also incorporates socially appropriate behaviours, including greeting customers and responding politely to expressions of thanks. While it does not physically serve drinks, it uses gestures to simulate this action, facilitating realistic multi-user human–robot interactions for research purposes. The primary goal of the system is to effectively engage, maintain, and conclude interactions with users, providing a balanced integration of functionality and social interaction. In the specified scenario, the robot bartender was able to interact with two users.

Louie et al. [12] present the social robot Tangy, which is capable of planning, scheduling, and facilitating group-based recreational activities. This involves two key objectives: first the planning and scheduling a series of recreational group activities throughout the day while taking into account the schedules of multiple users, and second the interaction with a group of users during these scheduled activities by offering both group-based and individualised assistance based on the current activity state and individual user needs. In this human–robot interaction scenario, Tangy is tasked with organising and running a series of bingo games. For this, Tangy schedules groups of individuals for different game sessions. Before each session, Tangy navigates to each user's location based on their schedule and invites them to attend the upcoming game, reminding them of the time and location. During game sessions, the robot takes on the role of a game moderator, calling out numbers and assisting players. Tangy effectively manages multiple users during the games and distinguishes players to provide individualised assistance when required. Although the robot was able to interact with a range of 4–20 users, conversations were not always conducted simultaneously.

In the work of Ben et al. [23], interactions between a robot and three users within a domestic scenario are observed. The robot functions as a shared smart home application, primarily serving a cleaning task. Here, the focus is set on the examination of various control schemes in the context of multi-user interaction. The authors set up a shared living environment and study small groups of three people, such as roommates or family members, who are accustomed to sharing household devices. To create a realistic setting, a desktop model of an apartment is designed, with common spaces including a living room, kitchen, and bathroom, as well as individual bedrooms. It is examined how users interact with the robot in different scenarios, including cases where all participants share a joint goal of cleaning communal spaces, and situations with conflicting cleaning goals that can occur if each person wants their room cleaned first.

4.1.4. Tutoring

Another discipline in MUCIs involves educational settings. Divekar et al. [24] observe a tutoring scenario in which two students interact with two embodied AI agents within a virtual world called 'Cognitive Immersive Language Learning Environment'. This interac-

tion serves the purpose of familiarising students with vocabulary and cultural knowledge in a playful manner. As a specific use case, the authors modelled a learning scenario where users are sent to virtualised street markets in China.

Koutsombogera et al. [25] also discuss the integration of embodied dialogue tutors in an educational context. Here, their aim is to facilitate MU discussions, encourage collaboration, and ensure equal participation among children. They also assess children's responses to text-based questions, providing insights into their comprehension and engagement. In the created educational setting, a tutor agent engages in a conversation with two children. The tutor asks questions related to texts the children have studied and assesses their reading comprehension through these questions. Additionally, the tutor promotes collaboration between the children and assists them in finding the correct answers within the text. It is important to note that the integration described in the paper is currently a conceptual framework designed primarily for data collection purposes, with no real system implementation in place at the time of publication.

Moreover, Laine et al. [26] describe a tutoring system to overcome the fear of public speaking using a virtual reality environment. Here, a participant is standing on a stage and tasked to speak in front of a mixed audience consisting of humans and computer-controlled avatars.

To help learners develop their argumentation skills effectively, a tutoring system called Metalogue has been developed to provide real-time feedback during debates [27,28]. This feedback enhances the learning experience by adapting posture, volume, and tone adjustments. The system generates various performance analyses while ensuring that real-time feedback is delivered in a balanced manner to prevent cognitive overload. A reflection dashboard allows for detailed review and analysis of learner performance, fostering awareness of the specific aspects and behaviours being trained.

4.1.5. Negotiation and Conflict Resolution

A variety of multi-party dialogue scenarios have been explored in the context of negotiation dialogues and conflict resolution. For negotiation dialogues, game-like structures such as 'Settlers of Catan' [29–31] have been used where multiple agents are expected to negotiate trades to be successful in the game. However, these scenarios are designed as purely simulated environments, with no direct involvement of human participants, to study uncooperative interaction strategies.

In a separate theoretical framework, a dialogue game has been conceptualised, involving multiple artificial agents engaged in resolving belief conflicts through argumentation [32]. Notably, this framework does not involve human participants, focusing exclusively on artificial agents. It primarily serves as a theoretical construct without a corresponding implementation.

Nurgaliyev et al. [33] discuss the development of a virtual assistant for conflict resolution within a smart home environment. The MU system adapts its operations to individual routines and preferences of its users, creating profiles to personalise interactions and recommendations. It continuously monitors user activities and scheduled events, offering timely prompts and reminders to maintain routines and safety. The system is also designed to detect and respond to potential risks in the home environment, issuing alerts to prevent harm. Overall, the proposed system integrates personalised assistance with proactive safety management to enhance the user experience.

The analysis of application domains also reveals that the design and implementation of MUCIs are influenced by various factors, with two key dimensions standing out:

- MU collaborative interactions vs. MU competitive interactions: For example, in social conversation contexts, tasks are often collaborative in nature [11,13], with users

working together toward a shared goal. In contrast, competitive scenarios consist of a different dynamic, where users may interact with the system in parallel, without being explicitly aware of each other's intentions and ambitions. This can lead to implicit or explicit competition, where users compete for the system's attention or resources, either unintentionally, as observed in [10], or deliberately, as demonstrated in [21,22]. Such interactions between users influence both the task and behaviour of the agent and thus require a different optimisation of the dialogue policy.

- Task-oriented designs vs. open-domain designs: Another meaningful way to categorise the reviewed works is by distinguishing between task-oriented and open-domain systems. Task-oriented systems are specifically designed and customised to achieve a specific objective, such as Tangy [12], which assists users in scheduling and playing bingo games. In contrast, open-domain systems aim to engage in open conversations without a narrowly defined goal, as shown in [15,16]. These differences in system goals also have significant implications for dialogue policy design and the handling of MU dialogues. While user interactions are often structured and goal-driven in task-oriented dialogues, reducing the need for extensive individual customisation, open-domain systems may require more nuanced adaptation to individual users and in-group dynamics to maintain engaging and contextually appropriate interactions. Moreover, the application of user models may facilitate the decision-making in domains such as tutoring [25,26] or conflict resolution [33].

As we can observe, the user and dialogue modelling plays a key role in the design of MUCIs. Consequently, the following section offers a more in-depth analysis of this topic.

4.2. User and Dialogue Modelling

While traditional systems employed models with pre-defined rules and modular approaches [15], current modern MUCIs use data-driven methods, substituting some of the components (e.g., in [34] or integrating end-to-end pipelines (e.g., in [18,19] where neural networks are used to model the addressee and response selection of the system). However, as commented before, the decision of using one approach or another also depends on the task and final target. For example, in critical domains such as mental health, complete data-driven approaches could result in catastrophic consequences; for this reason, the application domain plays a key role in the design decisions. In this section, we examine different approaches followed in the literature of our survey on user and dialogue modelling.

A user model is a representation of personal data collected from users, which usually consists of categories such as their preferences, relevant personal traits, or previous interactions. The aim is to equip a conversational interface with knowledge about users, thus enabling an adaptation of its conversation behaviour and response generation to specific characteristics. While traditional approaches contained only static information, contemporary models are updated dynamically during interactions. In parallel, dialogue modelling refers to the task of defining an interaction policy, thus determining which system action a conversational interface has to perform next. Since this decision-making frequently depends on previous user input, dialogue models typically include a dialogue history containing interaction-specific data.

One particular challenge in multi-user conversational systems is turn-taking, which means at which time and which message a system should send to its users [3]. Here, research faces the problem of detecting clear turn-taking signals, in contrast to a two-person (dyadic) conversation where indicators such as pauses and gaze directions are relatively easy to detect. In multi-user conversations, these signals are often subtle and ambiguous, making it difficult to accurately determine when one speaker has finished and another is likely to start. In a dyadic dialogue, it is usually clear when the role of the

speaker changes, whereas in multi-user dialogues, additional roles like side participant or overhearer further complicate the identification of the (next) speaker. Another issue is based on crosstalk: in spoken multi-user conversations, speakers often overlap, interrupt, and talk simultaneously, making it complex to distinguish between turn transitions and backchannelling (e.g., ‘ok’, or ‘well’) [11]. Moreover, the fact that people have different speaking styles, speech rates, and utterance durations makes it challenging to model turn-taking dynamics in a MU conversation. Some individuals tend to speak more often or for longer durations, while others may remain silent or speak less frequently. Turn-taking in MU conversations is also often influenced by contextual and social factors such as the topic of discussion, the participants’ roles and status, and their prior relationships. These factors can affect turn-taking behaviour, making it difficult to develop models that can accommodate these aspects. Multi-user conversations quickly become more complex as the number of users increases, since functions like face recognition and gaze tracking can put heavy computational requirements on the system. In general, there is so far no golden rule for modelling the turn-taking in Multi-User Conversational Interfaces, and as a result, different system prototypes were proposed in the reviewed literature. Table 3 presents an overview of the user and dialogue modelling approaches found in the reviewed publications.

Table 3. Characteristics of publications with regard to their user and dialogue modelling approaches.

Publication	User Model	Decision Making for Turn-Taking	Implemented System Prototype	Considering In-Group Dynamics
[21,35,36]	✓	✓	✓	✓
[3]	✓	✓	✓	
[25]	✓	✓		✓
[14]	✓	✓		
[11,12,33,37–39]	✓		✓	✓
[40,41]	✓			✓
[42]	✓			
[22,28]		✓	✓	✓
[15,16,34]		✓	✓	
[43,44]		✓		
[45]			✓	

Clavel et al. [17] present a survey of various methods for engaging conversational agents, including user-centred evaluation approaches. In the domain of multilingual MU dialogue systems, Gropp et al. [43] implement the Platon system, which reduces the complexity of modelling by creating a new dialogue engine instance for each user. These instances interact via shared variables or message exchanges, functioning similarly to text input handling. Platon employs a hierarchical task decomposition approach to manage multi-user scenarios by structuring sub-tasks as agents with defined input-handling rules. Originally developed for an interactive multi-user game, the system ensures modular and scalable dialogue management.

A key aspect of MUCI is turn-taking, which Johansson and Skantze [3] address by proposing a model for deciding when and what to say in a conversation. Moreover, Keizer et al. [22] investigate a bartender robot designed to handle multiple users by tracking social states, including customer attention-seeking behaviour and order fulfilment status. The system processes continuous input from low-level vision and speech components,

fusing these data sources to assign speech hypotheses and estimate customer engagement. The model captures state hypotheses with confidence scores and considers alternative interpretations of drink orders.

Participation of conversational interfaces in group interactions is investigated by Kim et al. [36], where the authors describe a chatbot that engages in discussions by moderating rather than providing expert knowledge. Meanwhile, Koutsombogera et al. [25] focus on the collection of a human–human multi-modal corpus featuring three participants engaged in a reading comprehension task within an educational setup. The study analyses multi-modal interaction strategies, communicative functions, and behavioural patterns, contributing insights into the design of conversational systems by identifying key conversational features such as dominance, attention, and engagement.

The work of Kumar et al. [41] presents a deep learning-based approach to person-specific response generation in multi-party conversations. Using the Multi-WoO dataset, the model constructs persona embeddings that capture individual conversational behaviours. The system encodes utterances using gated recurrent units and is tested with transformer models, seq2seq, and generative adversarial networks. A dyadic speaker–addressee model is employed to capture interaction dynamics, where utterances are classified as central or context, ensuring that responses align with a speaker’s conversational style.

In the paper of Li et al. [34], it is examined how a robot-assisted bingo game operates in a residential care setting, utilising user state modelling to enhance engagement. The Assistance Behavior Deliberation module determines the robot Tangy’s actions using a finite state machine, ensuring appropriate responses to player requests. When multiple assistance requests are made, they are queued in the order buttons are pressed, as users cannot interact via speech and only receive instructions from the robot. Similarly, Louie et al. [12] investigate user state identification in household conflicts, where a virtual assistant provides both group-based and individualised assistance based on scheduled activities. The assistant relies on a User Profile database to manage availability and location data, tailoring interactions accordingly. Extending this concept, Nurgaliyev et al. [33] introduce a conflict resolution virtual assistant, modelling user profiles across three age groups—young, adult, and elderly—while facilitating discussions among three individuals and the assistant.

Conversational AI is also addressed in Shibahara et al. [16], which presents a multi-party chit-chat system offering different levels of sympathy between two agents. Two dialogue system types are compared: an ‘Unsympathetic dialogue system’, where agent preferences differ, and a ‘Sympathetic dialogue system’, where both agents share the same preference. The system follows a non-goal-oriented structure, transitioning between information collection, feature extraction, and response generation. Turn-taking and user state handling in dialogue systems are further reviewed by Skantze [4], which discusses models for backchannelling, interrupt management, system-generated turn-taking signals, and multi-party interaction strategies.

In the domain of health applications, Snaith et al. [14] conceptualise a formal dialogue game in which multiple agent coaches adhere to predefined dialogue policies while negotiating health-related goals with a patient. They employ a rule-based dialogue modelling to select the next system action. In this context, Tahir et al. [35] investigate a humanoid robot acting as a social mediator through social state estimation. The study applies non-verbal speech analysis and pattern recognition to assess human behaviour, quantifying speech mannerisms and sociometric indicators such as interest, agreement, and dominance. Using a speech corpus of two-person face-to-face conversations, the study extracts two categories of low-level speech metrics: conversational cues (e.g., turn-taking frequency, pause in conversation, interruptions) and prosodic cues (e.g., speaking rate, response time). These measures provide insight into the dynamics of social interaction and mediation.

Todo et al. [15] present response generation in a one-user, two-agent setup, employing a decision tree to determine response type and timing based on prosodic features. If an utterance matches a predefined rule, the agent provides a comment before the system transitions to the next agent. Notably, the authors describe the setup as a three-person conversation, despite only involving one user and two systems. Furthermore, Wang et al. [44] explore spoken language understanding in dyadic conversations, focusing on determining and analysing the correct conversational context. The system adapts to different user needs and intentions across 38 domains. A machine-learning model trained on computer-addressed utterances categorises the human–human dialogue segments, filtering for topic distribution and contextual information.

Building on structured conversation modelling, Zhang et al. [42] introduce a tree-based framework for generating utterances in group conversations. The tree structure organises interaction threads, with branches representing different conversational pathways. The system encodes group conversations by processing tree branches sequentially, utilising a gated recurrent unit layer to compute utterance representations at each time step. Meanwhile, Zhu et al. [39] focus on multi-user empathetic dialogue generation, analysing textual conversations rather than speech. The study employs a complex emotion analysis model incorporating static–dynamic features, including speaker sensibility nodes and emotion-related utterance nodes, connected by various edge types. Each dialogue involves more than three speakers, with utterances labelled based on their empathy degree and emotional content. The model distinguishes between standard dialogue turns and emotionally driven interactions, enhancing its ability to classify user sensibility within conversations.

Besides, some publications focus on analysing conversation behaviour in MU scenarios, aiming to derive best practices or guidelines. Aylett and Romeo [46] list four key design elements for implementing MUCI: ensuring that required third party systems support MU interaction, identifying an open engineering path during development, focusing on concrete measurable use cases, and exploring less sophisticated systems that interact well.

4.3. Multi-Modality in Multi-User Conversational Interfaces

Several studies have explored the intricacies of multi-modal MUCIs, particularly in the context of intelligent and context-aware systems that act as embodied dialogue tutors. These systems aim to facilitate successful MU conversations by interpreting both verbal and non-verbal signals, effectively understanding the participants' states of mind and communication dynamics [27,47,48]. In this regard, in the articles, the multi-modal conversations were taking place mostly with virtual agents or with robotic agents.

4.3.1. Virtual Agents

For instance, in the experiments conducted by Koutsombogera et al. [25] the authors collect and annotate audio and visual data, using encoding schemes to capture conversational multi-modal behaviours. The annotation scheme includes various elements, such as head movements, facial expressions, and facial gestures, which play a crucial role in managing turn-taking and appropriate feedback in multi-user scenarios. The proposed MUCI is capable to engage in conversations with students using a combination of sensors and technologies, including Kinects for gesture and head orientation recognition, lapel microphones for speech capture, and a multi-channel loudspeaker array for spatialised audio presentation. In addition, Divekar et al. [24] propose a learning environment framework for tutoring systems. Their aim was to achieve a more natural interaction with AI agents without the need for wake-up words. The environment can be extended to multi-party scenarios, in which several agents may interject or even compete for the users' attention.

4.3.2. Robotic Agents

Similar to the application area of service robots, embodied interfaces play an important role in MUCIs. Addlesee et al. [49] present a large language model-based spoken dialogue system deployed on a social robot, facilitating conversations with multiple patients and their companions in a hospital setting. The system processes both speech and video inputs and generates multi-modal outputs, including speech and gestures (arm, head, and eye movements), to manage turn-taking, handle in-domain questions, and respond appropriately to out-of-domain requests. The authors describe the system's architecture and components, highlighting its ability to generate human-like clarification requests and adapt to unexpected patient utterances, with a demonstration video showcasing the system's real-time operability.

Kondo et al. [50] introduce a system designed for multi-party communication with adaptable gestures and facial expressions based on a speaker's location and situation. It utilises a special database for real-time generation of flexible gestures, combined with speech generation and gaze motion planning. The system incorporated sensory data, including the identification of speakers, using a camera, and featured motion parametrisation and gaze motion planning.

In the paper of Fan et al. [51], the ROCARE system is described, designed to facilitate multi-user engagement-based coaching. ROCARE aimed to adapt its behaviour to individual users' progress and assess user task performance, affective states, and gaze positions. It employed a modular architecture, including sensing, actuation, database, supervisory controller, and graphical user interface, with an emphasis on multi-modal human-robot interaction.

Moreover, Lee et al. [10] focused on a dialogue-based robot service that could provide relevant services to multiple users simultaneously. Their service-scheduling scheme analysed user intention based on factors like distance from the robot and vocal requests. This scheme allowed the robot to generate user-specific services and handle interactions with up to three users concurrently, recognizing and analysing their intentions.

To address socio-feedback in MUCIs, Tahir et al. [35] integrated and analysed non-verbal speech metrics to assess the social states of participants in two-person conversations with a humanoid Nao robot. The robot used this information to provide appropriate feedback in real-time, improving the quality of interactions. Speech mannerism and sociometrics, including interest, agreement, and dominance of the speakers, were quantified through non-verbal speech metrics.

Louie et al. [52] described the Tangy robot and its use in facilitating multi-step group recreational activities. Tangy utilised multi-modal interactions, including gesture mimicry, speech interaction, and a chest-mounted tablet for display. It retrieved world state parameters through various sensors, such as cameras and laser range finders, and was capable of real-time behaviour adaptation to accommodate user-specific activities.

In Tsamis et al. [53], the authors introduced an augmented reality (AR)-based system designed for human-robot collaborative environments. This system allowed workers to interact with a robot while receiving information related to the state of the robot and safety plans. For this, the system employed a heavy-duty mobile platform integrated with an arm manipulator, along with a Microsoft HoloLens for AR interaction. It tracked the user's head gaze, recognised hand gestures, and provided information about the robot's intended movements and navigation plans for enhanced safety and trust.

Keizer et al. [21] proposed a social multi-user interaction model using a robot bartender system. The system used vision- and speech-processing modules, along with a behaviour controller. It maintained a model of the social context to generate effective and socially appropriate responses based on observations about users in the environment. The

system aimed to engage, maintain, and successfully finish interactions with users and take their orders via spoken conversation. Moreover, Skantze et al. [11] explored a system for regulating turn-taking in multi-party situated interaction. They used various sensors, including Kinect cameras, close-talking microphones, and motion tracking markers, to monitor users' head rotations, hand movements, and speech. The system employed real-time multi-modal perception data to infer participants' attention and generate appropriate feedback behaviour, including gaze behaviour and speech activity.

Oertel et al. [54] developed an attentive listening system that generated multi-modal listening behaviour based on human–human analysis. The system used a GoPro camera for video capture and motion-tracking markers for head rotation inference. It tracked gaze behaviour and speech activity to categorise listeners, allowing the robot to adjust its listening behaviour accordingly, including gaze targets and feedback generation.

4.4. Evaluation Methods

In the final step of the paper analysis, we reviewed the methods used for the evaluation of MUCIs. Effective assessment must capture not only the linguistic capabilities and task performance of the system but also the individual, interpersonal, and social dimensions that emerge in group interactions. Similarly to the previous sections, no standardised process was identified. However, we observed two main approaches: user-centred and performance-based metrics. User-centred metrics focus on the quality of user experience and perception. These metrics are typically obtained by human evaluations, in which participants or annotators rate system outputs according to individual factors such as empathy, engagement, appropriateness, relevance, and naturalness. In multi-user contexts, additional criteria include the system's ability to address in-group dynamics, mediate between conflicts, and maintain inclusiveness. Only a few publications were also collecting this kind of feedback on the group level. Subjective evaluation often involves questionnaires or qualitative feedback. Contrarily, performance-based and automatic metrics rely on computational measures to evaluate system outputs against reference samples or predefined criteria. Common automatic metrics include BLEU, ROUGE, and METEOR, which quantify similarity between generated and reference responses, and accuracy-based metrics for tasks such as intent recognition or emotion classification. While these metrics are efficient and scalable, they often fail to extract conversational characteristics, social appropriateness, or the evolving context of multi-user dialogues. As a result, automatic metrics may not correlate well with actual user experience, especially in complex or empathetic MU scenarios. For future MUCIs, a comprehensive evaluation requires a combination of both approaches. The categorisation of evaluation metrics used in the retrieved publications is shown in Table 4. In summary, it strongly depended on the research group which evaluation methods were selected.

Table 4. Characteristics of publications with regard to their evaluation metrics.

Publication	User Study	Performance-Based Metrics	Group or User-Centred Metrics	Data Collection
[11,21,28]	✓	✓	user-centred	✓
[36]	✓	✓	group-centred	✓
[12,15,22,34,37,45,55,56]	✓	✓	user-centred	
[25,57]	✓		user-centred	✓
[35,58]	✓		user-centred	
[16,33,59]	✓		group-centred	
[3,39,60]		✓	group-centred	✓

Table 4. Cont.

Publication	User Study	Performance-Based Metrics	Group or User-Centred Metrics	Data Collection
[42]		✓	group-centred	
[40,41,44]		✓	user-centred	

To provide a clearer understanding of how evaluation has been approached in literature, we offer a more detailed description of the used evaluation methods in the following. Anbro et al. [37] validate the applicability of a virtual reality evaluation in clinical scenarios for MU applications. Similarly, Adikari et al. [40] examine the accuracy of system predictions regarding users' emotional reactions and the general emphatic state. In the domain of interactive storytelling for education, Alofs et al. [61] explore a MU tabletop interface designed to support social interaction. A qualitative study involving four pairs of children assessed the system through positive and negative feedback, as well as questions on social interaction. In human–robot interaction, Canevet et al. [62] evaluate conversation tracking and user identification in two settings: an 'easy' mode with two participants and a 'hard' mode with up to six participants. The evaluation, conducted on data recordings from 28 participants, measured face detection performance, tracking ability, and participant re-identification.

Divekar et al. [24] investigate a virtual environment for foreign language learning with multiple AI agents. This study evaluates the students' language performance through pre-, post-, and delayed post-tests, focusing on objective metrics such as vocabulary recognition, listening comprehension, transcription, and interactive conversation. Subjective feedback was gathered through a student experience questionnaire measuring self-reported learning gains and overall experience with 10 participants. Dohsaka et al. [13] present a quiz-style multi-party setting involving multiple users and agents, testing five systems under various experimental conditions. The evaluation considers communication activation through three perspectives: the number of utterances (objective), user satisfaction (subjective), and user opinions about the agent (subjective). Additionally, user memorisation of biographical facts conveyed in the dialogues is assessed as a secondary objective with 64 participants. In the context of socially assistive robotics, Fan et al. [38] present a robotic coach architecture that interacts with two elderly participants per session. Cognitive impairment is screened using the Mini-Mental State Examination, while a self-reported questionnaire captures perceptions of interacting with both the robot and the human partner, with eight participants divided into four groups.

Investigating the impact of robotic motion behaviour on collaborative tasks, Faria et al. [63] let groups of three participants engage in several interaction sessions, each corresponding to a different type of motion. Objective measures include reaction time and error count, while subjective measures assess collaboration, fluency, movement legibility, and predictability, as well as perceptions of animacy and intelligence. Additionally, personal preferences regarding study conditions are collected, with 33 participants in total. Moreover, Keizer et al. [21] evaluate social engagement and interaction in a bartender robot scenario. A pre-experiment questionnaire, based on the Godspeed questionnaire series, is followed by four rounds of interaction. Subjective metrics cover task success, ease of seeking attention, order comprehension, and interaction naturalness, while objective metrics include attention-seeking time, interaction time, serving time, speech input events, as well as speaker identification and speech-recognition failures. A total of 48 participants paired into 24 groups were recruited for the study. The focus is on user-based metrics. Kondo et al. [50] investigate body gesture planning for androids in a guide scenario, comparing a proposed

system to alternatives lacking motion interruptivity and motion parametrisation. The study involved four conditions with 1662 participants, measuring how many participants interacted with the system and residence time. A second study with 42 participants applied the Semantic Differential method using 28 pairs of antonymous adjectives.

Müller and Richert [64] evaluate the interaction of social robots in public spaces by conducting an in-the-wild study at a science museum, where a Furhat robot interacts with groups of visitors. They analyse 129 MU interaction sessions using qualitative video analysis to understand how the robot manages attention, turn-taking, and engagement across group members. The evaluation also examines in-group dynamics, which they consider crucial for the successful development of realistic robotic agents. The system introduced by Paetzel-Prüsmann et al. [59] allows individuals or groups to engage with a social robot. Following a scripted storyline, the users are able to interact with the robot several times and are asked to rate their satisfaction with the system.

The impact of conversational strategies in MU scenarios are addressed by Kraus et al. [55]. In their Wizard-of-Oz study, a conversational recommender system interacts with groups using two dialogue strategies. The evaluation measures task effectiveness and efficiency, user experience, and group dynamics, revealing that while the use of proactivity accelerates decision-making and reduces the need for chatbot moderation, it may also introduce a bias towards dominant individuals, potentially impacting overall group performance. The virtual agent developed by Makri et al. [48] is designed to teach metacognitive and individual- and community-level skills. For this, they conducted a pilot study in which five participants tested group interactions in English, evaluating visual signal processing and information load. A subsequent evaluation with 41 participants only considered single-user interactions.

Oertel et al. [54] present a study on an attentive listening system generating multi-modal listening behaviours. The first study measured user perception of the robot's social presence using the Social Presence Questionnaire and assessed the focus of attention through gaze time, involving 12 participants in single-user interactions with two robots. The second study had 21 third-party observers rate videos from the previous experiment. Furthermore, Ostrowski et al. [65] investigate smart speakers engaging with users at home, conducting three studies with voice-based interfaces used by families between 3–5 members in a long-term evaluation. The research on conversational AI by Shibahara et al. [16] assesses a chat-chat system that provides different levels of sympathetic responses. Their evaluation had participants watch a video, interact with the system for a few minutes, and subsequently fill in a questionnaire on user-centred metrics. This was repeated with a second system. They recruited 20 participants, each interacting with two agent behaviours. Straßmann et al. [56] analyse the effects of task-related robot discrimination. For this, they gave pairs of participants a book-sorting challenge. The users had to interact with a robot assistant to complete the task and fill in self-reported questionnaires on the attribution of blame and credit of the task outcome. Moreover, Tahir et al. [35] use a humanoid robot acting as a social mediator using an estimator for the social state of the group in their evaluation. The first study involved single-user dialogues in which 20 participants identified system feedback messages and completed the Godspeed questionnaire. The second study introduced two-person dialogues, with the robot facilitating scenario-based conversations in four different scenarios, pairing an expert user with a real participant. System feedback was rated via a questionnaire, followed by an additional Godspeed questionnaire. Lastly, Zhu et al. [39] investigate multi-user empathetic dialogue generation. The evaluation was carried out through both performance-based and user assessments. The automatic metrics evaluated how model-generated responses matched human reference

responses, which was subsequently extended by a human annotation of sample dialogues using subjective metrics.

5. Discussion

The reviewed literature indicates that the majority of MUCIs are highly tailored to specific application contexts. Across the five main application domains (health care, social conversation, service robots, tutoring, and negotiation and conflict resolution), we observed a diverse spectrum of prototype implementations, user roles, system capabilities, and conversational behaviour. However, several common patterns appear across the domains:

- **Role differentiation:** Most system developers assign distinct roles to users and system (e.g., patients and health coach [14], customers and assistant [3], learners and tutor [5]), which guide the flow of dialogues and facilitate to manage turn-taking. One reason for this may be that systems which support clear role distributions tend to manage MU input more effectively and ensure balanced participation.
- **Limited dialogue management:** Several publications, particularly in tutoring and conflict resolution, describe conceptual frameworks or systems that only partially implement dialogue logic [25,32]. As a result, dynamic conversational capabilities remain underdeveloped in contemporary approaches. Instead of employing data-driven approaches, systems often rely on scripted or semi-structured dialogues with minimal adaptability to unexpected user input or spontaneous conversational shifts. However, suitable training data for MUCIs remains scarce across many applications and use cases.
- **Rigid interaction behaviour:** A significant number of interfaces still rely on pre-defined and fixed conversation structures, which implies that the system is processing user input rather sequentially and separately from the group context. In contrast, some systems (e.g., [11,22]) demonstrate capabilities to handle truly simultaneous input from multiple users with dynamic turn-taking. Nevertheless, this seems necessary to enable natural interactions in group settings, where users may speak out of turn, overlap, or engage in side conversations. Without support for simultaneous input and context-aware coordination, systems risk selecting inappropriate or intrusive responses.
- **Prototypical systems:** Many systems are developed as research prototypes or simulations. Although the findings of their corresponding user studies offer valuable insights, fully functional and deployable real-world MUCIs have yet to be realised. Exceptions are seen in service robotics and personal assistants that operate in constrained environments, such as [64,65]. Bridging the gap between experimental prototypes and robust real-world MUCIs remains a critical challenge, requiring advances in dialogue management, multi-user conversation analysis, and multi-modal integration.
- **Multi-Modality:** The use of embodiment and multi-modality through robots or virtual avatars has shown great potential for maintaining engagement and managing attention in groups. Systems that integrate multi-modal input and output (e.g., gaze, speech, and gestures) tend to be more robust in turn-taking or user state tracking [24,49]. However, implementing and synchronising multi-modal channels remains technically demanding and resource-intensive, which may limit scalability. In addition, such systems may consist of hardware setups that are difficult to replicate and transfer to further use-cases.

In addition, more research gaps in MUCIs were identified which slow down their development and applicability. One of the most prominent gaps is the absence of reliable turn-taking mechanisms, which are an essential aspect of enabling natural interaction. Only a few systems incorporate robust conversational cues such as gaze, user state, or gestures to coordinate speaker transitions, leading systems to interrupt and confuse users. As a

consequence, user frustration and conversation breakdowns become more likely. Without mechanisms to interpret and respond to these non-verbal cues, systems struggle to manage the timing of their responses, resulting in misalignment between system actions and user intentions. Furthermore, although many systems operate in social or sensitive contexts, there is minimal modelling of in-group dynamics, such as dominance, engagement, or social roles, which have been shown to be fundamental for ensuring fair and inclusive interaction. Another persistent challenge is the limited number of real-world deployments and long-term evaluations. Most systems are evaluated in controlled lab settings or simulations, leaving open questions about how they perform over time in diverse, dynamic environments. For the research community, the use of fragmented evaluation practices with little standardisation across subjective user metrics, task performance measures, or interaction quality at the individual or group level makes it difficult to compare results and build cumulative knowledge. Moreover, ethical considerations and guidelines are also notably absent within literature, for instance in terms of privacy management, fairness in moderation, and bias in user recognition. Finally, research on MUCIs currently lacks modular and domain-independent architectural frameworks and best practices that could support scalable development across application domains.

To address these research gaps and overcome the limitations described, there is a growing need for structured, reusable toolkits that can support the systematic development of MUCIs. In addition to the reviewed literature, we identified IrisTK [66] as one of the few toolkits specifically developed for MU interaction. This toolkit stands out as an early and influential attempt to support real-time, multi-user dialogues by combining a modular architecture, statechart-based dialogue management, and multi-modal input and output channels in face-to-face conversations. While IrisTK provides a framework for understanding the architectural and interactional requirements of MUCIs, it reflects the capabilities and design assumptions of its time. Expanding on this foundation, we integrate findings from the reviewed literature to propose a set of best practices for the design, implementation, and evaluation of MUCIs:

- **Application-orientated design:** MUCIs should be adapted to the social and physical context in which they operate. This includes considerations such as the number of users, their spatial arrangement, and the interaction setting (e.g., public spaces, homes, classrooms). A good contextual fit improves relevance, usability, and user engagement.
- **Adaptive dialogue management:** Future systems should be able to handle interruptions, overlapping speech, and changing in-group dynamics. Effective dialogue policies require flexibility and, where possible, adaptive mechanisms that can respond to dynamic user behaviours, leaving behind fixed interaction patterns. While data-driven methods like reinforcement learning and transformer-based dialogue models have shown notable success in single-user CUIs, their potential in multi-user scenarios needs further exploration.
- **Natural turn-taking strategies:** As a characteristic of natural group interaction, users often communicate through a combination of speech, gaze, gestures, and facial expressions. MUCIs should integrate multi-modal channels in a coordinated way to manage turn-taking effectively, taking into account both individual and group-level cues.
- **Standardised evaluation metrics:** The evaluation of MUCIs remains fragmented, with a lack of consistent metrics and methodologies. To support meaningful comparison and progress in the field, it is necessary to apply shared evaluation standards that assess subjective user feedback, task performance, and group-level interaction quality. While short-term, lab-based studies provide valuable initial insights, future research should prioritise long-term, real-world deployments to fully capture user behaviour and social interaction over time.

- **Ethical considerations:** Although significant efforts on system development and dialogue management were presented in the reviewed articles, few explicitly address ethical aspects such as user privacy, fairness in interaction, and accessibility. These considerations are particularly critical in shared or public environments in which multiple users engage simultaneously and the system has to balance competing goals, complex social dynamics, and differing user expectations. Future work should include these principles in the design and evaluation of MUCIs.

Together, these best practices are intended to provide a solid basis for the design, implementation, and evaluation of MUCIs. Future systems should be not only technically robust, but also socially aware, contextually appropriate, and ethically grounded. As the field continues to progress, aligning future systems with these guidelines may contribute to overcome current limitations and to exploit the full potential of MUCIs across applications.

6. Conclusions and Outlook

Concerning research on MUCIs, there exist several factors that make it difficult for authors to communicate their findings effectively and establish collaborations. One of the main challenges is the lack of a unified community, as has been highlighted in this survey paper. As a result, researchers across MU topics often employ their own distinct terminology, utilise diverse methodologies, and approach research questions from different perspectives. However, this impairs mutual understanding of each other's work and discourages effective collaboration.

To address these challenges, it seems essential to enhance the perspicuity and visibility of publications. This involves using standardised terminology and avoiding ambiguous or personalised definitions, acronyms, and unique jargon that may confuse or mislead readers. By improving the perspicuity of their work, researchers can increase its impact and make it easier for others to build upon their findings. Additionally, research dissemination channels such as journals, conferences, and workshops can foster visibility and interdisciplinary collaboration by expanding platforms and discussions on Multi-User Conversational Interfaces. To build a stronger and more integrated community, it is crucial to connect people with diverse expertise, enabling them to face future challenges. As an initial step toward unifying the discourse, incorporating the definition proposed in this work could help align future publications under a shared understanding and consistent terminology.

In the context of the reviewed literature, most publications focus on scenarios in which two users interact with a single conversational interface. The reason for this may be that the complexity of the interface increases significantly as the number of supported users grows. In contrast, several papers had to be excluded from our survey because, despite being labelled as multi-user, the authors are actually referring to the sequential processing of single-user inputs and not to their simultaneous handling.

Despite significant research and development in the field, there is no implemented system that has gained widespread adoption, neither in an industrial or academic context. Instead, most researchers have focused on developing and evaluating prototypical systems or frameworks. These systems are designed to showcase new ideas and concepts and often feature novel technologies and experimental designs. However, the implementation of underlying dialogue models remains frequently incomplete, and many of these systems are not yet fully implemented for real-world deployment. Due to this, multi-user conversation is still more of a laboratory use case that is being explored in controlled environments. In particular, almost all of publications of our sample used different dialogue and user models. This poses a barrier to the generalisability and broader applicability of presented MUCIs.

Several publications investigated on equipping virtual agents or robots with conversational interfaces. For this, both hardware and software components of systems need capabilities to deal with multiple users. It requires expertise in robotics, artificial intelligence, and computer science, among other fields. Furthermore, robotics research often involves significant hardware development, which can be expensive and time-consuming. Robotics is also heavily depending on multi-modality, as effective communication with embodied user interfaces requires the integration of visual, audio, and tactile inputs to interpret user intents, respond appropriately, and adapt to dynamic environments in real-time. Another unresolved challenge lies in the evaluation of MUCIs, as current methods often lack consistency and fail to capture the complex interactions involved. Developing comprehensive and standardised evaluation frameworks is essential to ensure meaningful comparisons and guide future improvements.

In conclusion, the development of effective MUCIs remains an open challenge, particularly in addressing the requirements and expectations of users across varying contexts. Future research should prioritise the development of scalable, robust systems that can adapt to their users and the rapidly changing technological landscape. Considering the recent progress in conversational AI by integrating the generation capabilities of large language models, the performance of MUCIs has the potential to be significantly enhanced.

Author Contributions: Conceptualization, N.W.; methodology, N.W.; software, N.W.; validation, N.W.; formal analysis, N.W.; investigation, N.W.; resources, N.W.; data curation, N.W.; writing—original draft preparation, N.W., M.K.; writing—review and editing, N.W., M.K.; visualization, N.W.; supervision, W.M., D.G., Z.C.; project administration, D.G., Z.C.; funding acquisition, D.G., Z.C. All authors have read and agreed to the published version of the manuscript.

Funding: The research leading to these results has received funding from the Spanish R&D&i project TrustBoost (PID2023-150584OB-C21 and PID2023-150584OB-C22) financed by ICIU/AEI/10.13039/501100011033 and from FEDER, EU. It has also received funding the HORIZON EUROPE MSCA-Staff Exchanges Action entitled “CRYSTAL, Conversational Systems for Emotional Support and Customer Assistance” (Grant no. 101182965). Additionally, it has also received funding from the project FORSocialRobots, financed by the Bavarian Research Foundation (AZ1594-23).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The collection of the 59 final selected articles and the graph of author connections are available at the following link: <https://www.researchrabbitapp.com/collection/public/PLOVR1MGZG0> (last visit 24 June 2025). If the link expires, the list of selected publications will be provided upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
DS	Dialogue Systems
CUI	Conversational User Interface
MU	Multi-User
MUCI	Multi-User Conversational Interface

Appendix A. Most Relevant Articles

Figure A1 shows the most relevant papers of the collection of 59 included publications, clustered by the ResearchRabbit tool, ordered vertically by the publication (or online availability) year.

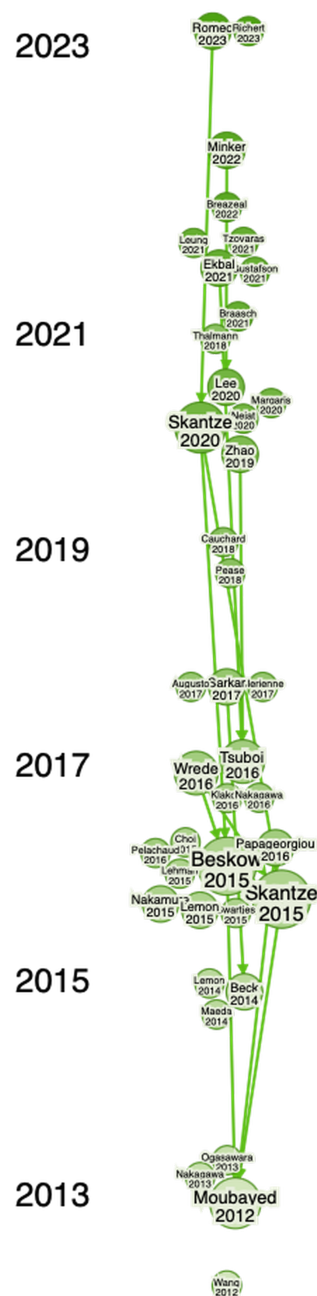


Figure A1. Distribution of the most influential publications per timeline from the selection based on bibliometrics of the articles. Each node representing a publication displays the name of the last author. Graphs are created with the ResearchRabbit tool.

References

1. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int. J. Surg.* **2021**, *88*, 105906. [[CrossRef](#)] [[PubMed](#)]
2. Cole, V.; Boutet, M. ResearchRabbit (product review). *J. Can. Health Libr. Assoc./J. L'Assoc. Bibliothèque Santé Can.* **2023**, *44*, 43–47. [[CrossRef](#)]

3. Johansson, M.; Skantze, G. Opportunities and Obligations to Take Turns in Collaborative Multi-Party Human-Robot Interaction. In Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Prague, Czech Republic, 2–4 September 2015; pp. 305–314. [\[CrossRef\]](#)
4. Skantze, G. Turn-taking in Conversational Systems and Human-Robot Interaction: A Review. *Comput. Speech Lang.* **2021**, *67*, 101178. [\[CrossRef\]](#)
5. Moubayed, S.A.; Skantze, G.; Beskow, J. The furhat back-projected humanoid head–lip reading, gaze and multi-party interaction. *Int. J. Humanoid Robot.* **2013**, *10*, 1350005. [\[CrossRef\]](#)
6. Suzuki, S.; Anuardi, M.N.A.M.; Sripian, P.; Matsuhira, N.; Sugaya, M. Multi-user Robot Impression with a Virtual Agent and Features Modification According to Real-time Emotion from Physiological Signals. In Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, 31 August–4 September 2020; pp. 1006–1012. [\[CrossRef\]](#)
7. Panhwar, M.A.; Saddar, S.; Aftab, A.; Memon, K.A.; Raza, M.O.; ZhongLiang, D.; Noor-ul-Ain. Adaptive Interface for Multi-users and Its Evaluation. *Int. J. Comput. Sci. Netw. Secur.* **2019**, *19*, 198–201.
8. Gu, J.C.; Tao, C.; Ling, Z.H. Who Says What to Whom: A Survey of Multi-Party Conversations. In Proceedings of the IJCAI, Vienna, Austria, 23–29 July 2022; pp. 5486–5493.
9. Traum, D. Issues in Multiparty Dialogues. In *International Workshop on Agent Communication Languages*; Dignum, F., Ed.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 201–211.
10. Lee, G.; An, K.; Yun, S.S.; Choi, J. A simultaneous robot service scheme for Multi-users. In Proceedings of the 2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Goyangi, Republic of Korea, 28–30 October 2015; pp. 373–374.
11. Skantze, G.; Johansson, M.; Beskow, J. Exploring turn-taking cues in multi-party human–robot discussions about objects. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, WA, USA, 9–13 November 2015; pp. 67–74.
12. Louie, W.Y.G.; Vaquero, T.; Nejat, G.; Beck, J.C. An autonomous assistive robot for planning, scheduling and facilitating multi-user activities. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 5292–5298.
13. Dohsaka, K.; Asai, R.; Higashinaka, R.; Minami, Y.; Maeda, E. Effects of conversational agents on activation of communication in thought-evoking multi-party dialogues. *IEICE Trans. Inf. Syst.* **2014**, *97*, 2147–2156. [\[CrossRef\]](#)
14. Snaith, M.; De Franco, D.; Beinema, T.; Den Akker, H.O.; Pease, A. A dialogue game for multi-party goal-setting in health coaching. In Proceedings of the 7th International Conference on Computational Models of Argument, COMMA 2018, Warsaw, Poland, 12–14 September 2018; pp. 337–344.
15. Todo, Y.; Nishimura, R.; Yamamoto, K.; Nakagawa, S. Development and evaluation of spoken dialog systems with one or two agents through two domains. In Proceedings of the International Conference on Text, Speech and Dialogue, Pilsen, Czech Republic, 1–5 September 2013; pp. 185–192.
16. Shibahara, Y.; Yamamoto, K.; Nakagawa, S. Effect of sympathetic relation and unsympathetic relation in multi-agent spoken dialogue system. In Proceedings of the 2016 International Conference on Advanced Informatics: Concepts, Theory and Application (ICAICTA), Penang, Malaysia, 16–19 August 2016; pp. 1–6.
17. Clavel, C.; Cafaro, A.; Campano, S.; Pelachaud, C. Fostering user engagement in face-to-face human-agent interactions: A survey. In *Toward Robotic Socially Believable Behaving Systems-Volume II*; Esposito, A., Jain, L., Eds.; Springer: Cham, Switzerland, 2016; pp. 93–120.
18. Liu, C.; Liu, K.; He, S.; Nie, Z.; Zhao, J. Incorporating Interlocutor-Aware Context into Response Generation on Multi-Party Chatbots. In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), Hong Kong, China, 3–4 November 2019; Bansal, M.; Villavicencio, A., Eds.; Association for Computational Linguistics (ACL): Hong Kong, China, 2019; pp. 718–727. [\[CrossRef\]](#)
19. Ouchi, H.; Tsuboi, Y. Addressee and response selection for multi-party conversation. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 2133–2143. [\[CrossRef\]](#)
20. Ju, D.; Feng, S.; Lv, P.; Wang, D.; Zhang, Y. Learning to Improve Persona Consistency in Multi-party Dialogue Generation via Text Knowledge Enhancement. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 12–17 October 2022; pp. 298–309.
21. Keizer, S.; Kastoris, P.; Foster, M.E.; Deshmukh, A.; Lemon, O. Evaluating a social multi-user interaction model using a Nao robot. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014; pp. 318–322. [\[CrossRef\]](#)
22. Keizer, S.; Foster, M.E.; Gaschler, A.; Giuliani, M.; Isard, A.; Lemon, O. Handling uncertain input in multi-user human–robot interaction. In Proceedings of the The 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014; pp. 312–317. [\[CrossRef\]](#)

23. Ben-Hanania, A.; Goldberg, J.; Cauchard, J.R. Multi-User Control for Domestic Robots with Natural Interfaces. In Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia, Cairo, Egypt, 25–28 November 2018; pp. 433–439. [\[CrossRef\]](#)
24. Divekar, R.R.; Drozdal, J.; Chabot, S.; Zhou, Y.; Su, H.; Chen, Y.; Zhu, H.; Hendler, J.A.; Braasch, J. Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Comput. Assist. Lang. Learn.* **2021**, *35*, 1–29. [\[CrossRef\]](#)
25. Koutsombogera, M.; Deligiannis, M.; Giagkou, M.; Papageorgiou, H. Towards modelling multimodal and multiparty interaction in educational settings. In *Toward Robotic Socially Believable Behaving Systems-Volume II*; Springer International Publishing: Cham, Switzerland, 2016; pp. 165–184.
26. Laine, T.H.; Lee, W.; Moon, J.; Kim, E. Building confidence in the metaverse: Implementation and evaluation of a multi-user virtual reality application for overcoming fear of public speaking. *Int. J. -Hum.-Comput. Stud.* **2025**, *199*, 103487. [\[CrossRef\]](#)
27. Helvert, J.; Rosmalen, P.; Börner, D.; Petukhova, V.; Alexandersson, J. Observing, coaching and reflecting: A multi-modal natural language-based dialogue system in a learning context. In Proceedings of the 11th International Conference on Intelligent Environments, Prague, Czech Republic, 5–17 July 2015; Volume 19; pp. 220–227.
28. Al Moubayed, S.; Lehman, J. Toward better understanding of engagement in multiparty spoken interaction with children. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, WA, USA, 9–13 November 2015; pp. 211–218.
29. Hiraoka, T.; Georgila, K.; Nouri, E.; Traum, D.; Nakamura, S. Reinforcement learning in multi-party trading dialog. In Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Prague, Czech Republic, 2–4 September 2015; pp. 32–41. [\[CrossRef\]](#)
30. Cuayáhuítl, H.; Keizer, S.; Lemon, O. Learning to trade in strategic board games. In *CGW 2015, GIGA 2015. Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2015; Volume 614, pp. 83–95. [\[CrossRef\]](#)
31. Xiao, G.; Georgila, K. A comparison of reinforcement learning methodologies in two-party and three-party negotiation dialogue. In Proceedings of the The Thirty-First International Flairs Conference, Melbourne, FL, USA, 21–23 May 2018.
32. Yuan, J.; Yao, L.; Hao, Z.; Wang, Y. Multi-party Dialogue Games for Dialectical Argumentation. *J. Comput.* **2012**, *7*, 2564–2571. [\[CrossRef\]](#)
33. Nurgaliyev, K.; Di Mauro, D.; Khan, N.; Augusto, J.C. Improved multi-user interaction in a smart environment through a preference-based conflict resolution virtual assistant. In Proceedings of the 2017 International Conference on Intelligent Environments (IE), Seoul, Republic of Korea, 21–25 August 2017; pp. 100–107. [\[CrossRef\]](#)
34. Li, J.; Louie, W.Y.G.; Mohamed, S.; Despond, F.; Nejat, G. A user-study with tangy the bingo facilitating robot and long-term care residents. In Proceedings of the 2016 IEEE international symposium on robotics and intelligent sensors (IRIS), Tokyo, Japan, 17–20 December 2016; pp. 109–115. [\[CrossRef\]](#)
35. Tahir, Y.; Dauwels, J.; Thalmann, D.; Magnenat Thalmann, N. A user study of a humanoid robot as a social mediator for two-person conversations. *Int. J. Soc. Robot.* **2020**, *12*, 1031–1044. [\[CrossRef\]](#)
36. Kim, S.; Eun, J.; Oh, C.; Suh, B.; Lee, J. Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. In Proceedings of the CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 25–30 April 2020; pp. 1–13. [\[CrossRef\]](#)
37. Anbro, S.J.; Housmanfar, R.A.; Thomas, J.; Baxter, K.; Harris, F.C., Jr.; Crosswell, L.H. Behavioral Assessment in Virtual Reality: An Evaluation of Multi-User Simulations in Healthcare Education. *J. Organ. Behav. Manag.* **2022**, *43*, 92–136. [\[CrossRef\]](#)
38. Fan, J.; Beuscher, L.; Newhouse, P.A.; Mion, L.C.; Sarkar, N. A robotic coach architecture for multi-user human–robot interaction (RAMU) with the elderly and cognitively impaired. In Proceedings of the 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 26–31 August 2016; pp. 445–450. [\[CrossRef\]](#)
39. Zhu, L.; Zhang, Z.; Wang, J.; Wang, H.; Wu, H.; Yang, Z. Multi-Party Empathetic Dialogue Generation: A New Task for Dialog Systems. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; pp. 298–307. [\[CrossRef\]](#)
40. Adikari, A.; De Silva, D.; Moraliyage, H.; Alahakoon, D.; Wong, J.; Gancarz, M.; Chackochan, S.; Park, B.; Heo, R.; Leung, Y. Empathic conversational agents for real-time monitoring and co-facilitation of patient-centered healthcare. *Future Gener. Comput. Syst.* **2022**, *126*, 318–329. [\[CrossRef\]](#)
41. Kumar, R.; Chauhan, D.S.; Dias, G.; Ekbal, A. Modelling Personalized Dialogue Generation in Multi-Party Settings. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–6. [\[CrossRef\]](#)
42. Zhang, H.; Chan, Z.; Song, Y.; Zhao, D.; Yan, R. When Less Is More: Using Less Context Information to Generate Better Utterances in Group Conversations. In *Natural Language Processing and Chinese Computing: 7th CCF International Conference, NLPCC 2018, Hohhot, China, August 26–30, 2018, Proceedings, Part I* 7; Zhang, M., Ng, V., Zhao, D., Li, S., Zan, H., Eds.; Springer: Cham, Switzerland, 2018; pp. 76–84. [\[CrossRef\]](#)

43. Gropp, M.; Schmidt, A.; Kleinbauer, T.; Klakow, D. Platon: Dialog Management and Rapid Prototyping for Multilingual Multi-user Dialog Systems. In Proceedings of the International Conference on Text, Speech, and Dialogue, Brno, Czech Republic, 12–16 September 2016; pp. 478–485. [\[CrossRef\]](#)
44. Wang, D.; Hakkani-Tür, D.; Tur, G. Understanding computer-directed utterances in multi-user dialog systems. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8377–8381. [\[CrossRef\]](#)
45. Toxtli, C.; Monroy-Hernández, A.; Cranshaw, J. Understanding chatbot-mediated task management. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; pp. 1–6. [\[CrossRef\]](#)
46. Aylett, M.P.; Romeo, M. 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, Eindhoven, The Netherlands, 19–21 July 2023. [\[CrossRef\]](#)
47. Makri, E.; Spiliotopoulos, D.; Vassilakis, C.; Margaris, D. Human behaviour in multimodal interaction: Main effects of civic action and interpersonal and problem-solving skills. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *11*, 5991–6006. [\[CrossRef\]](#)
48. Makri, E.; Koryzis, D.; Spiliotopoulos, D.; Svolopoulos, V. Metalogue’s Virtual Agent for Negotiation: Its Effects on Learning Experience, Metacognitive and Individual-and-Community-Level Attitudes Pre-and-Post Interaction. In Proceedings of the EDULEARN19, 11th International Conference on Education and New Learning Technologies, Palma, Spain, 1–3 July 2019; IATED Academy: Valencia, Spain, 2019; pp. 1542–1551. [\[CrossRef\]](#)
49. Addlesee, A.; Cherakara, N.; Nelson, N.; Hernandez Garcia, D.; Gunson, N.; Sieińska, W.; Dondrup, C.; Lemon, O. Multi-party Multimodal Conversations Between Patients, Their Companions, and a Social Robot in a Hospital Memory Clinic. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, St. Julians, Malta, 17–22 March 2024; Aletras, N., De Clercq, O., Eds.; ACL: Kerrville, TX, USA, 2024; pp. 62–70.
50. Kondo, Y.; Takemura, K.; Takamatsu, J.; Ogasawara, T. A gesture-centric android system for multi-party human-robot interaction. *J. Hum.-Robot. Interact.* **2013**, *2*, 133–151. [\[CrossRef\]](#)
51. Fan, J.; Bian, D.; Zheng, Z.; Beuscher, L.; Newhouse, P.A.; Mion, L.C.; Sarkar, N. A robotic coach architecture for elder care (ROCARE) based on multi-user engagement models. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *25*, 1153–1163. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Louie, W.Y.G.; Nejat, G. A social robot learning to facilitate an assistive group-based activity from non-expert caregivers. *Int. J. Soc. Robot.* **2020**, *12*, 1159–1176. [\[CrossRef\]](#)
53. Tsamis, G.; Chantziaras, G.; Giakoumis, D.; Kostavelis, I.; Kargakos, A.; Tsakiris, A.; Tzovaras, D. Intuitive and Safe Interaction in Multi-User Human Robot Collaboration Environments through Augmented Reality Displays. In Proceedings of the 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), Vancouver, BC, Canada, 8–12 August 2021; pp. 520–526. [\[CrossRef\]](#)
54. Oertel, C.; Jonell, P.; Kontogiorgos, D.; Mora, K.F.; Odobez, J.M.; Gustafson, J. Towards an engagement-aware attentive artificial listener for multi-party interactions. *Front. Robot. AI* **2021**, *8*, 189. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Kraus, M.; Klein, S.; Wagner, N.; Minker, W.; André, E. A Pilot Study on Multi-Party Conversation Strategies for Group Recommendations. In Proceedings of the 6th ACM Conference on Conversational User Interfaces, Luxembourg, 8–10 July 2024. [\[CrossRef\]](#)
56. Straßmann, C.; Eudenbach, C.; Arntz, A.; Eimler, S.C. “Don’t Judge a Book by its Cover”: Exploring Discriminatory Behavior in Multi-User-Robot Interaction. In Proceedings of the Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, Boulder, CO, USA, 11–15 March 2024; pp. 1023–1027. [\[CrossRef\]](#)
57. Wagner, N.; Kraus, M.; Tonn, T.; Minker, W. Comparing Moderation Strategies in Group Chats with Multi-User Chatbots. In Proceedings of the CUI 2022: 4th Conference on Conversational User Interfaces, Glasgow, UK, 26–28 July 2022. [\[CrossRef\]](#)
58. Li, B.; Lou, R.; Segonds, F.; Merienne, F. Multi-user interface for co-located real-time work with digital mock-up: A way to foster collaboration? *Int. J. Interact. Des. Manuf. (IJIDeM)* **2017**, *11*, 609–621. [\[CrossRef\]](#)
59. Paetzel-Prüsmann, M.; Lehman, J.F.; Gomez, C.J.; Kennedy, J. An Automatic Evaluation Framework for Social Conversations with Robots. In Proceedings of the 2024 International Symposium on Technological Advances in Human-Robot Interaction, Boulder, CO, USA, 9–10 March 2024; pp. 56–64. [\[CrossRef\]](#)
60. Richter, V.; Carlmeyer, B.; Lier, F.; Meyer zu Borgsen, S.; Schlangen, D.; Kummert, F.; Wachsmuth, S.; Wrede, B. Are you talking to me? Improving the Robustness of Dialogue Systems in a Multi Party HRI Scenario by Incorporating Gaze Direction and Lip Movement of Attendees. In Proceedings of the Fourth International Conference on Human Agent Interaction, 4–7 October 2016, Biopolis: Singapore; pp. 43–50. [\[CrossRef\]](#)
61. Alofs, T.; Theune, M.; Swartjes, I. A tabletop interactive storytelling system: Designing for social interaction. *Int. J. Arts Technol.* **2015**, *8*, 188–211. [\[CrossRef\]](#)

62. Canévet, O.; He, W.; Motlicek, P.; Odobez, J.M. The MuMMER Dataset for Robot Perception in Multi-party HRI Scenarios. In Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, 31 August–4 September 2020; pp. 1294–1300. [[CrossRef](#)]
63. Faria, M.; Silva, R.; Alves-Oliveira, P.; Melo, F.S.; Paiva, A. “Me and you together” movement impact in multi-user collaboration tasks. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 2793–2798. [[CrossRef](#)]
64. Müller, A.; Richert, A. No One is an Island—Investigating the Need for Social Robots (and Researchers) to Handle Multi-Party Interactions in Public Spaces. In Proceedings of the 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Busan, Republic of Korea, 28–31 August 2023; pp. 1772–1777. [[CrossRef](#)]
65. Ostrowski, A.K.; Fu, J.; Zygouras, V.; Park, H.W.; Breazeal, C. Speed Dating with Voice User Interfaces: Understanding How Families Interact and Perceive Voice User Interfaces in a Group Setting. *Front. Robot. AI* **2022**, *8*, 730992. [[CrossRef](#)] [[PubMed](#)]
66. Skantze, G.; Al Moubayed, S. IrisTK: A statechart-based toolkit for multi-party face-to-face interaction. In Proceedings of the 14th ACM International Conference on Multimodal Interaction, Santa Monica, CA, USA, 22–26 October, 2012; pp. 69–76. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.