A Design Space for Intelligent Dialogue Augmentation

Robin Shing Moon Chan
ETH Zürich
Zürich, Switzerland
robinsmchan@gmail.com

Anne Marx
ETH Zürich
Zürich, Switzerland
anmarx@student.ethz.ch

Alison Kim
Universität Zürich
Zürich, Switzerland
alison.y.kim@berkeley.edu

Mennatallah El-Assady
ETH Zürich
Zürich, Switzerland
melassady@ai.ethz.ch

Abstract

The use of intelligent agents in communication is a growing trend aimed at enhancing the efficiency and quality of interactions. As such, *dialogue augmentation systems*—text processing systems that interactively enhance ongoing written or spoken communication—are gaining significant popularity across domains. While technical limitations had previously inhibited their real-time usage for effective communication augmentation, recent developments in language processing have improved their capabilities to contribute to dialogue as intelligent, emancipated, and proactive agents. While other works on dialogue augmentation focus on evaluating design considerations for specific applications of these systems, we lack a unified understanding of the broader design principles that apply to dialogue more generally. Through a literature review and mixed-methods analysis of 78 existing systems, we iteratively define a comprehensive design space for intelligent dialogue augmentation systems. To further ground our analysis, we interweave Clark's [27] models of dialogue with concepts in human-AI collaboration and discuss trends in the evolving role of dialogue augmentation systems along five dimensions—dialogue context, augmentation context, task, interaction, and model. Based on the identified trends, we discuss concrete challenges for broader adoption, highlighting the need to design *trusted*, *seamless*, and *timely*, and *accessible* augmentations. The design space contributes as a mechanism for researchers to facilitate defining design choices during development, situate their systems in the current landscape of works, and understand opportunities for future research.

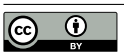
CCS Concepts

• **Human-centered computing** → *Interactive systems and tools.*

Keywords

Speech Processing, Interaction Design, Design Space

ACM Reference Format:
Robin Shing Moon Chan, Anne Marx, Alison Kim, and Mennatallah El-Assady. 2025. A Design Space for Intelligent Dialogue Augmentation. In *30th International Conference on Intelligent User Interfaces (IUI '25), March 24–27, 2025, Cagliari, Italy.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3708359.3712096



IUI '25, Cagliari, Italy

ACM ISBN 979-8-4007-1306-4/25/03
https://doi.org/10.1145/3708359.3712096

1 Introduction

At its most general definition, dialogue refers to spoken or written communication between two or more agents. Beyond the minimal, one-on-one, *dyadic* conversation, dialogue occurs in a vast range of settings. This includes multi-party, *polyadic* structured discussions, informal open brainstorming sessions, or broadcast debates directed at large audiences. However, even if a clear purpose is pre-defined, the nature of practical dialogue is complex and a number of context-dependent issues may arise that lead to the dialogue becoming ineffective. For instance, goal-oriented discussions frequently suffer from procrastination, disruption, and loss of concentration, which inhibit the productivity of discussions [11, 32]. Further, whereas participants in small-group discussions tend to be more active, large-group settings across dialogue settings tend to yield uneven participant engagement and contribution [60, 68, 69]. The unstructured, interwoven nature of such dialogue more generally makes it difficult to organize diverse opinions and requires a high cognitive effort for recollecting or summarizing key findings [72]. Finally, without additional assistive technology, there is a lack of dialogue accessibility for non-native speakers [33] or persons with disabilities to engage in such discussions, such as deaf and hard-of-hearing people [48, 71] or people with speech impairments [128]. Alternative communication technologies beyond face-to-face dialogue, such as instant messaging or video-conferencing, have been shown to further amplify such challenges [22, 30, 39, 59, 122].

This paper discusses *dialogue augmentation systems*—autonomous agents that enhance dialogue to mitigate some of the challenges of natural dialogue. Whereas early works in this area required human supervision due to the lack of performance of transcription and text processing models [33, 91], recent advancements in natural language processing and information retrieval have improved their capabilities and have enabled them to act as increasingly intelligent, emancipated, and proactive agents. As such, a plethora of tools has been developed to autonomously support discussions through automated documentation [65, 84], creativity support [87, 127, 129], decision support [25, 65, 107], mediation [41, 123, 140], or assistive communication [19, 95]. However, the wide range of dialogue settings, application domains, and complexity of roles taken by AI agents make it difficult to attain a holistic understanding of the state and opportunities surrounding this class of systems.

In this paper, we present a *design space* for intelligent dialogue augmentation systems—a taxonomy that characterizes the key axes of system and interaction design. This design space serves three main purposes: (1) to establish shared vocabulary among researchers in the field, (2) to review existing works and facilitate design choices during system development, and (3) to highlight opportunities for future research [81]. While previous studies have extensively explored design practices for conversational agents (CAs) in the dyadic human-CA setting through reviews [79], user studies [2, 26, 131, 134], and interview studies [80, 113, 138]—our focus is on the *augmentation of dialogue itself*. This difference in scope fundamentally changes the nature of the human-AI collaboration, and, consequently, the associated design requirements. While individual studies have empirically developed design practices for specific applications—such as classroom education [6], video-conferencing [87], and chatbot interactions [139]—a comprehensive, unified analysis of the space remains unexplored.

Our contribution is three-fold. First, we systematically review 78 existing studies that develop dialogue augmentation systems. Second, through an iterative, mixed-methods approach, we propose a design space that defines design considerations on five major axes—*dialogue context*, *augmentation target*, *task*, *interaction*, and *model*. We further ground our design space on established frameworks in human-AI collaboration and social interaction studies, namely, Clark's [27] theory of *dialogue as a joint activity*. Finally, to demonstrate the utility of our tool, we present three use case scenarios drawn from requirement analysis studies across various domains, followed by a discussion of the trends, challenges, and opportunities revealed through the design space analysis.

2 Background and Related Work

In this section, we provide background on the frameworks on which we ground our design space and review related literature. These frameworks originate in two differentiable areas of research: social interactive studies that model and dissect dialogue as a *joint activity* and human-AI collaboration.

Dialogue as Joint Activity. A way of modeling dialogue in differing contexts stems from cognitive and social interaction studies and is based on *types of social activities*. Levinson [83] first introduces a categorization of joint activity types with an emphasis on its effect on language use. More relevantly, Clark [27] illuminates discourse *as* such a joint activity and defines "dimensions of variation" that categorize the varieties of dialogue—conversations, lectures, interviews, or letter exchanges—in the framework of activity types. Among the dimensions of variations, *scriptedness*, i.e., the degree to which discourse is planned or structured, *formality*, *cooperativeness*, i.e., working together or adversarially, and *governance*, i.e., the degree to which control is distributed among discussion participants, are defined. As in any joint activity, participants in dialogues are further assumed to take varying *roles*, and participants engage to achieve certain *goals*. Initially, the dimensions of variations serve as a useful tool to categorize the dialogue contexts in which dialogue augmentation systems are used. Ultimately, however, we investigate what roles dialogue augmentation systems can already take and how human collaboration with such systems changes the dialogue dynamics.

Dialogue Augmentation as Human-AI Collaboration. Holter and El-Assady [50] deconstruct the variability space among human-AI systems into three general dimensions: *agency* [34, 90], which refers to which agent holds control during the task-solving process; *interaction* [85, 125], user-system communication and guidance; and *adaptation* [5, 34, 125], which describes which agent adapts its behavior over time in response to the interaction and the feedback received by the other agent. Our work adapts the frameworks of human-AI collaboration to characterize the design space of dialogue augmentation systems along the *agency*, *interaction*, and *adaptation* axes in various dialogue contexts. Several works similarly define system and interaction design axes for specific applications that can be classified as dialogue augmentation. As such, An et al. [6] defines design considerations regarding the user interface and system initiation for language processing systems to enhance classroom teaching. For video conferencing, Liu et al. [87] derive distinctions in the *augmentation timing*, the *augmentation initiation*, and the augmentation system's *input medium*. Further, past studies on conversational agents introduce the differentiation in the aspects of user guidance [79, 134, 139], team configuration [16], as well as chatbot tasks [16, 73]. While offering valuable insights, many of these studies focus on domain-specific design aspects. However, they often lack a holistic perspective that addresses broader design principles applicable across intelligent dialogue augmentation systems, which is the subject of our study.

3 Approach

Figure 1 shows our overall methodology for generating the design space, which we split into a systematic literature review and collaborative design space iterations. The employed open-coding methodology is a common approach to systematically analyze themes and trends in HCI research areas [13, 81, 125]. In the following, we describe this process in more detail.

3.1 Systematic Literature Review

We first conduct a thorough literature review of existing dialogue augmentation systems. To this end, we identify the HCI field as the most relevant field and, therefore, scrape the ACM Digital Library for relevant works. We thereby focus on the top-tier conferences and journals in HCI: CHI, CSCW, CUI, DIS, IUI, UIST, and ToCHI. We further only include full research articles or proceedings, excluding special tracks such as extended abstracts or vision tracks.

Scope and Definitions. As *Dialogue Augmentation* is not a firmly established or commonly used term, we first provide the precise scope of the study. The scope serves as the inclusion criterion for the papers considered in our literature review and is aligned with the objectives of our study. While it serves our specific research focus, it is not intended as a universally applicable definition.

- **Dialogue.** We consider *systems* that augment *dialogue*, which we limit to written or spoken conversational exchange between two or more agents. Following our description in section 2, interlocutors may be passive, and need not necessarily contribute to the discussion. Yet, to clearly distinguish ourselves from monologues, we impose that dialogue participants need to have the opportunity to intervene in the discussion as active agents.

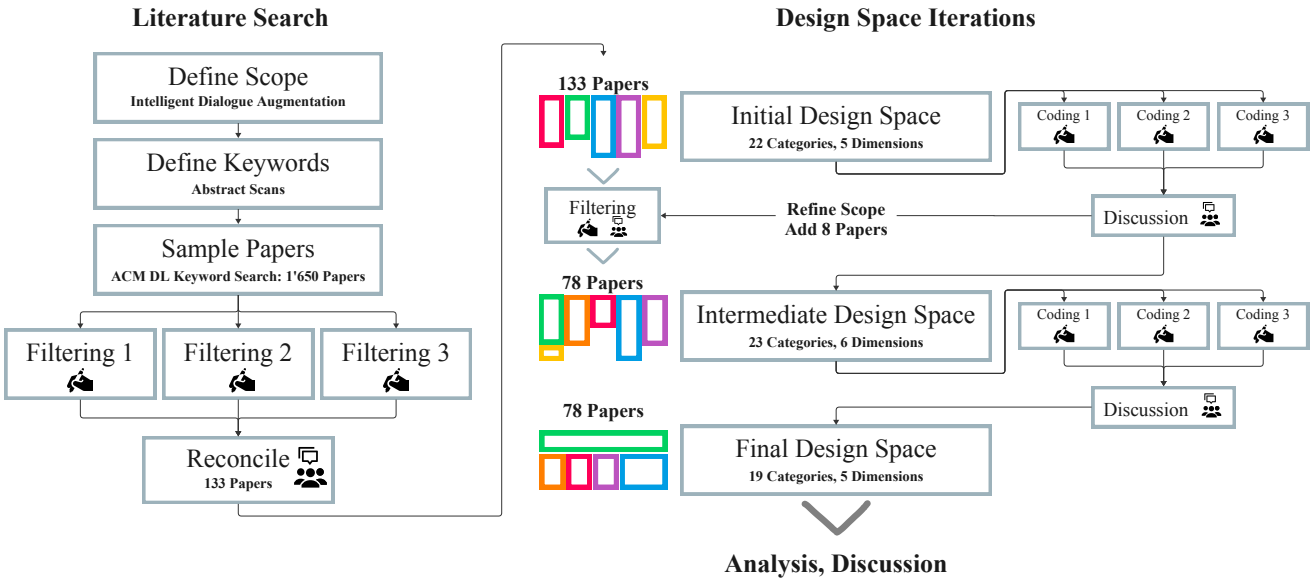


Figure 1: Methodology: The design space creation consists of two phases. Three annotators first thoroughly review and filter through existing works implementing dialogue augmentation systems and then construct our design space through three iterations of paper coding, author discussions, and design space refinement.

- **Augmentation.** By augmentation, we mean enriching a natural dialogue interaction for the intended benefit of at least one of the dialogue participants or the dialogue audience.[1]
- **Intelligent.** We consider a system to be intelligent if it is capable of autonomous decision-making, without further restrictions on the model complexity. This may therefore include rule-based or basic statistical approaches. However, we strictly impose our focus on language and/or audio processing, but include studies that may *additionally* process other inputs, such as visual cues.

In sum, the scope of our study includes studies of *language processing systems* that automatically process and enrich the written or spoken communication between two or more human and/or AI dialogue participants. Such dialogues may, for example, include one-on-one informal exchanges, goal-oriented group messaging, or more asymmetric notions of dialogue, such as interactive presentations. To further cement the scope, we provide an example of the common types of studies that were excluded from the scope during our discussions along with their exclusion criteria in subsection A.2.

Keyword Identification. As a plethora of terms is used to refer to our above-defined scope , we first identify relevant keywords. To this end, we first scan abstracts of the most recent proceedings of all relevant venues[2] and extract common keywords from the title and author keywords of the relevant subset of papers. The set of keywords contains a range of variations and synonyms of "speech", "dialogue", and "chat", consciously yielding a rather extensive range of papers subject to heavy filtering to avoid missing relevant works. The exact search query is listed in A.1 and retrieved 1'650 articles for the relevant venues in September 2024.

Paper Filtering. After discussing the initial scope among all authors, three authors collectively filtered through all retrieved papers to evaluate whether they fit into an initial scope. The initial filtering aimed to retrieve papers with high recall, i.e., to contain any papers that fit into the broader scope but may be removed during tagging as the scope was continually made more specific. The filterers worked independently to avoid agreement bias. To ensure that the set of considered papers was consistent, regular meetings were held to discuss ambiguous cases among the filterers, and, if necessary, the scope was clarified. To further ensure consistency, the filterers had an overlap of 25% of the overall corpus, and, among the overlapping subset, they achieved a satisfying inter-annotator agreement of 92.4%. During a reconciliation discussion, disagreements were discussed and, if necessary, resolved by the lead author, who went through the entire corpus and, therefore, had the best overview to ensure consistency. This process generated an initial corpus of 133 relevant papers.

Design Space Iterations. After filtering, the authors generated an initial design space following a structure commonly used in similar works of interactive systems design spaces. Namely, design space *categories* are grouped into overarching topics, or *dimensions*, and individual dialogue augmentation systems differ in the *category* values, or *codes*. Although *codes* are generally mutually exclusive

[1]We purposefully do not specify that the augmentation needs to happen simultaneously as the dialogue. We further elaborate on this in section 4.
[2]CHI '24, CSCW '23, IUI '24, CUI '24, UIST '23, DIS '24, ToCHI issues from 2024

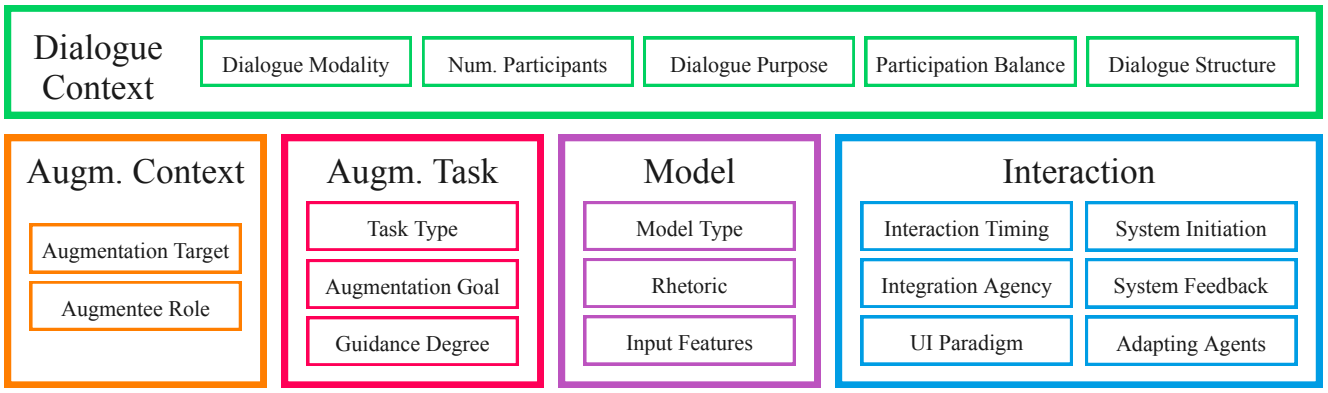


Figure 2: Design Space: We define the design space of intelligent dialogue augmentation systems along 19 categories within five broad dimensions—*dialogue context, augmentation context, augmentation task, model,* and *interaction.*

within a *category*, we did not strictly enforce this as for some reviewed systems, multiple codes applied. This especially holds for the task categories, as systems become increasingly multifunctional. Further, the *categories* are ideally perpendicular, i.e., a system could, hypothetically, take any permutation of codes, independent of the other categories. In practice, trends in system design yielded a pattern of common code combinations in the design space. The initial design space contained 19 categories in the *context*, *interaction*, *task*, and *model* dimensions, which are commonly used axes in design spaces for interactive systems [13, 81, 142]. The design space was then iteratively refined using a mixed-methods approach. First, three authors independently coded a set of papers using the current state of the design space. Through a discussion among all authors, the design space was updated to add new interesting categories and codes or merge redundant ones. Further, during discussions with the group of authors, additional relevant papers came up that did not exactly match the criteria of our keyword search. Such papers were then manually added to the corpus for completeness. Overall, a total of eight papers were added in this way. Additionally, the scope was continuously made more specific upon a closer reading of the corpus. This led to the further exclusion of 53 papers from the initial corpus upon a case-by-case joint author evaluation. This finally left us with 78 total papers that were relevant to the final scope.[3] Based on the discussions and the added papers, the design space was then updated and the annotators recalibrated their tagging. This process was repeated three times, until all authors were satisfied with the design space and all papers had collectively been tagged. Again, to ensure consistency, the annotators for the tagging had an overlap of 10%, for which the inter-annotator agreement was 91.6%. Finally, any disagreements were brought up during discussions and, if necessary, resolved by the lead author, who went through the entire corpus and, therefore, had the most consistent understanding of the scope.

[3]All relevant papers are listed in subsection A.3.

4 Design Space

Through our iterative approach of coding and design space updates, we finally arrive at a design space with five dimensions—*dialogue context*, *augmentation context*, *task*, *interaction*, and *model*—19 categories and 58 codes. The following section gives an overview of the final version of the design space.[4] Note that in the *code* columns of the tables of this chapter, we additionally mention the percentile of papers that fulfill the definition for each code separately. Some papers may not be exclusive to one code per dimension and can fulfill multiple code definitions at once, e.g., the dialogue is performed both in written and spoken modalities, or a paper may not contain information regarding a dimension at all, the sum over all codes within one dimension can be more than or less than 100%.

4.1 Dialogue Context

Dialogue, in our scope, can take many forms, such as informal one-on-one chit-chat, doctor-patient consultations, virtual group breakout brainstorming discussions, instant messaging groups, and large-scale interactive presentations, or debates. Consequently, the requirements and resulting effectiveness of dialogue augmentation systems are often a function of the underlying dialogue setting. To this end, we aim to develop a useful categorization of dialogue context that encodes such differing requirements, which we ground in established theoretical frameworks of practical dialogue described in section 2.

Modality. This category describes the medium in which the underlying dialogue takes place. We coarsely differentiate between written and spoken exchanges as they both have nuanced requirements. For written communication, or messaging, the loss of paralinguistic information and increased distractability [22, 30, 39, 59] make augmentation systems often focus on conveying such paralinguistic cues [10, 124] or supporting structure through moderation [72, 82]. For spoken communication, such content moderation systems need to overcome a larger barrier to equally intervene due to

[4]We acknowledge that the section formatting is adapted from Lee et al. [81] upon their permission, which greatly facilitated the preparation of the current manuscript.

	Code	Definition
Modality	*In which modality does the dialogue take place?*	
	Written (47.4%)	The dialogue occurs in written form, e.g., via instant messaging.
	Spoken (56.4%)	The dialogue occurs in spoken form, e.g., in-person or in a video-conference.
Num. Particip.	*How many agents partake in the dialogue?*	
	One-on-One (42.3%)	A dialogue between two participants.
	Small Group (42.3%)	Dialogue involving fewer than five participants.
	Many-Participants (30.8%)	Dialogue involving more than five participants or a large audience.
Purpose	*Do dialogue participants have practical intent?*	
	Social (51.3%)	The dialogue is focused on informal conversation without a specific goal.
	Task-Oriented (74.4%)	The dialogue is focused on achieving a specific task or goal.
Balance	*Who drives or directs the conversation?*	
	Symmetric (71.8%)	Participants share equal roles in directing the conversation.
	Asymmetric (28.2%)	One participant has more control or influence over the direction of the dialogue.
Structure	*How is the dialogue organized, structurally?*	
	Linear (41.0%)	Dialogue follows a structured, linear flow of interaction.
	Parallel (5.1%)	Dialogue flows in separate, parallel discussions, often involving groups or subgroups.
	Unstructured (65.4%)	Dialogue follows a more free-form, unstructured flow.

Table 1: Dialogue Context: Dimensions, Codes, and Definitions

the added complexity of speech recognition and generation. Therefore, such agents have only recently developed from being post-hoc analyzers [114] to being able to guide spoken dialogue to a similar degree [70, 98, 104]. Still, many interesting aspects of spoken interaction make it subject to a majority of the discussed studies, such as teaching systems that aim to improve a speaker's verbal expression [99, 126, 136] or the listener's understanding by augmenting dialogue with visual channels [87, 97]. For some scenarios, the underlying dialogue may be mixed, such as between interabled interlocutors [103], or settings such as interactive live streams with written viewer feedback [140]. In such cases, the augmentation must occur in a commonly accessible modality.

Num. Participants. This category purely captures the number of interlocutors, spanning from dyadic, one-on-one conversations, small group settings, to many-participant settings, such as classroom settings or forums [86, 143]. The group size can significantly change the dialogue dynamic, the individual's role, and the kind of support deemed most useful [72]. For instance, while participants in small group discourse tend to be more active, participants in large many-participant settings tend to become inactive lurkers as their perceived anonymity grows [60, 72]. Further, participants in smaller groups have been shown to prefer more proactive system engagement, as they fear this interrupts other speakers in a larger setting [87]. However, the majority of discussed studies focus on small-group discussions.

Dialogue Purpose. Following Clark's [27] framework, we further aim to characterize the dialogue setting by the overall *goal* it pursues. As such, we provide a coarse grouping into social and task-oriented dialogues. Whereas socially oriented dialogue may be guided towards supporting relationship-building through icebreaking [89, 94] or building trust [108], most systems aim to tackle task-oriented dialogues, which serve the purpose of completing a specific task, such as decision-making [25, 41], or more creative tasks [129]. We acknowledge that it is common for more natural dialogue to switch between the two [88].

Symmetry. Dialogue can additionally be categorized as either symmetric or asymmetric. In symmetric dialogue, all participants contribute equally and hold similar roles, as seen in open discussions. In contrast, asymmetric dialogue is more uneven or one-sided, as in a lecture, where one participant dominates the exchange. This classification aligns with the concept of *global symmetry* as framed by Pickering and Garrod [102], who describe dialogue as a *joint activity*, and is related to Clark's [27] notion of *governance*, which refers to how control and *roles* are distributed among participants. In symmetric dialogue setups, while the goals of augmentation can vary significantly, fostering inclusivity and equal participation is a common objective [114]. In contrast, asymmetric dialogues often involve predefined roles, such as discussion leaders, where augmentation systems tend to provide facilitation [84] and guidance [82, 93, 136].

	Code	Definition
Target	*Which of the participants does the augmentation target?*	
	Individual (41.0%)	The augmentation is aimed at helping or enhancing a single participant.
	Subgroup (5.1%)	The augmentation is aimed at helping or enhancing a specific subgroup of participants.
	Global (62.8%)	The augmentation is aimed at helping or enhancing all participants in the dialogue.
Role	*What is the role of the augmentation target?*	
	Producer(74.4%)	The augmentation assists a participant in their role as a producer (e.g., improving expression).
	Recipient (85.9%)	The augmentation assists a participant in their role as a recipient (e.g., improving understanding).

Table 2: Augmentation Context: Dimensions, Codes, and Definitions

Dialogue Structure. The structure of a dialogue refers to its overall organization, closely related to the level of *scriptedness* involved. We distinguish between three types: linear, parallel, and unstructured dialogue. Linear dialogue follows a single, sequential flow of communication, often seen in moderated debates, interviews [136], or structured turn-taking exchanges [38]. In contrast, parallel dialogue occurs when the conversation divides into simultaneous subgroups, such as in online classrooms or collaborative work settings [115]. Lastly, unstructured dialogue lacks a clear framework or organization, resembling open, spontaneous discussions where participants freely interact without predefined turns or roles. The augmentation systems for linear dialogue often aim to make interlocutors stick to the predefined structure. On the other hand, augmentations in unstructured discussions focus on managing the fluid nature of the conversation, helping users organize and summarize the diverse range of topics that emerge [36, 52, 133]. Finally, augmentations in the parallel aim to provide an overview across groups by highlighting the subgroup activity [115].

4.2 Augmentation Context

This dimension discusses the varying contexts of the augmentation within a dialogue. Namely, we aim to understand whom the augmentation targets. To that end, we find it to be important to distinguish whether the augmentation is targeted to a subgroup of interlocutors, and whether the persons the augmentation targets have a specific role in the dialogue.

Augmentation Target. This category specifies what subset of dialogue participants are exposed to the augmentation. As such, we differentiate whether the augmentation targets only an individual, such as when there are participants with disabilities in need of support [19, 74, 124], or the augmentation is supposed to only support a specific role within the dialogue, such as moderators, class instructors, or team leaders for guidance [17, 93, 115]. Subgroup augmentation frequently occurs when the roles are assigned group-wise, such as audiences in debates or live streams [41, 140] or students in the educational setting [132]. Finally, the most common paradigm is global exposure, where every dialogue participant or auditor is exposed to the augmentation. Naturally, this ties to the discussion of the usage of shared devices, as the augmentation of individuals or subgroups may require personalized devices or interfaces, rather than shared ones. While personalized devices can offer more adaptive and relevant information to an individual [18, 132], a shared view can foster a common understanding among multiple participants [103].

Augmentee Role. Building on the previous category, we aim to explicitly define the role held by the persons exposed to the augmentation. In this way, we differentiate whether the augmentation helps the producer (i.e., the speaker in voice-based interaction) express themselves better, or helps the recipient (i.e., the listener in voice-based interaction) better understand, or verify what is being said. In the role of enhancing the producer, dialogue augmentation systems may help shortcut to the next actions or topics to discuss [24], guide the producer to use more productive rhetoric [93], support therapeutic speech training [43], or, conversely, adapt peoples' paralanguage to be more accessible to disabled participants [57]. Dialogue augmentation for recipient focuses on better understanding what is being said by simply transcribing speech, highlighting keywords, summarizing main points, or visualizing what is being discussed [64, 65], adding additional context [143], or enhancing written communication with paralinguistic cues [10, 99].

4.3 Task

The task dimension aims to provide a coarse categorization of the overall goal of the augmentation. As such, it is on a conceptually higher level than the concrete tasks that are performed by a language model, such as transcription, classification, or retrieval, as those simply are means to achieve the augmentation goal.

Task Type. As our scope provides a broad set of tasks, we first provide a coarse grouping into documenting, analyzing, and suggesting tasks. Dialogue augmentation systems may have traditionally been as simple as transcription systems documenting the dialogue content for later reference [46, 84] or to build a common understanding among participants. The latter is especially relevant for the inclusion of non-native, disabled, or societally marginalized interlocutors [38, 63, 103]. More developed systems go beyond documentation and actively process and analyze user input and give feedback to a user [76, 114, 119]. Most recently, systems have improved to contribute to the dialogue by directly suggesting how a participant should contribute, such as through starter phrases [24, 128], topic suggestions [94], or by prompting inactive users to engage [86].

	Code	Definition
Task Type	*Overall, does the system display, analyze, or generate/suggest?*	
	Documenting (28.2%)	The system focuses on documenting or recording the dialogue or interaction.
	Analyzing (61.5%)	The system analyzes the dialogue, extracting insights or patterns.
	Suggesting (53.8%)	The system provides suggestions or generates content to assist dialogue participants.
Augmentation Goal	*What is the overall purpose of the augmentation?*	
	Information (29.5%)	The augmentation provides additional information to support the dialogue.
	Accessibility (21.8%)	The augmentation helps make the dialogue accessible to a wider range of participants (e.g., for those with impairments).
	Clarification (12.8%)	The augmentation helps clarify ambiguous or complex points in the dialogue.
	Facilitation (33.3%)	The augmentation aids in the process of memorization or provides scaffolding.
	Guidance: Divergent (20.5%)	The augmentation offers guidance by encouraging the discovery of diverse viewpoints.
	Guidance: Structure (14.1%)	The augmentation helps moderate and structure the dialogue.
	Guidance: Reflection (26.9%)	The augmentation encourages participants to reflect on the dialogue.
	Guidance: Engagement (26.9%)	The augmentation helps keep participants engaged and involved in the dialogue.
Guidance	*To what degree is guidance enforced?*	
	Orienting (38.5%)	The augmentation shows possibly relevant options without an explicit ranking.
	Directing (15.4%)	The augmentation shows ranked relevant options.
	Prescribing (23.1%)	The augmentation shows only the option thought to be optimal for the task.

Table 3: Augmentation Task: Dimensions, Codes, and Definitions

Augmentation Goal. Dialogue augmentation systems are highly diverse in what they aim to achieve overall, and many such systems enhance dialogue in multiple of the following ways. The first purpose is to provide additional information to support dialogue, such as adding contextual paralinguistic information to written messages [7, 10, 47, 124, 140], or retrieving relevant context [133, 143]. Further, augmentation often provides accessibility to the dialogue for participants who can not participate equally, due to disability [103, 128], or being non-native speakers [37]. Given the ambiguity and complex dynamics of dialogue, such systems can also help in the form of clarifications, for example by retrieving images to resolve ambiguities [87], or intervene to establish common ground [128]. Finally, a large group of systems can be grouped as facilitation, which reduces cognitive load during discussion by aiding memorization or providing scaffolding [132]. We further list a range of tasks where the system explicitly provides *guidance* to a user, i.e., directs the user with a specific purpose [50, 139]. Firstly, augmentation systems may encourage discussion participants to discover new viewpoints, i.e., encourage divergence. Similarly, they may also encourage interlocutors to reflect on their contributions, and thereby mediate the discussion [41, 72], or improve their behavior [17]. Dialogue augmentation systems can also guide group discussions to stick to a predefined structure through moderation [54]. Finally, such systems are also frequently used to guide passive users to improve their engagement in discussions through nudging to improve the overall contribution balance of the discussion, and, thus, enable an increase of opinion diversity [72].

Guidance Degree. As described in section 2, theoretical frameworks in human-AI collaboration [23, 50] split the degree of guidance into three categories. Whereas orienting refers to showing possibly relevant options without an explicit ranking, directing guidance adds such a ranking to the provided choices, such as relevance rankings in message recommendations [24]. Finally, prescribing guidance shows only a single option, which is thought to be optimal for the underlying task. Of the three options, orienting and prescribing guidance are the most prevalent in dialogue augmentation systems. Orienting guidance provides neutral suggestions, leaving much of the agency on the human side, whereas prescribing guidance is more rapidly understandable, and may, therefore, disrupt conversation less [19].

4.4 Interaction

In the interaction dimension, we discuss crucial aspects of dialogue augmentation through the lens of human-AI collaboration—user and system feedback, timing, adaptation, and agency.

Interaction Timing. The setting of dialogue augmentation where the AI agent somehow needs to integrate into a flowing discussion highlights the importance of timing. Most commonly, the systems in the scope provide augmentation during the dialogue, which instantaneously lets the participants adapt to model-generated information, clarifications, and suggestions [87, 128, 139]. Few systems exist that generate behavioral analyses for users after the interaction to give feedback to the user for them to adapt in the next interaction [114], or during asynchronous dialogue [67].

	Code	Definition
Timing	*When does the user interact with the system in relation to the dialogue?*	
	During (83.3%)	The user interacts with the system during the dialogue, providing real-time interaction.
	After (23.1%)	The user interacts with the system after the dialogue has concluded for follow-up or analysis.
Adapting Agents	*Which agent is adapting in the interaction?*	
	User (47.4%)	The user adapts their behavior or actions based on the system's responses.
	Both (Co-Adaptive) (52.6%)	Both the user and the system adapt to each other's behavior in a co-adaptive manner.
UI Paradigm	*What user interface is used in the interaction?*	
	Text Interface (10.3%)	The interaction takes place through a text-based interface, such as typing or reading.
	Graphical UI (48.7%)	The interaction takes place through a graphical interface, such as menus, buttons, or windows. This includes web UIs.
	Voice UI (11.5%)	The interaction takes place through voice commands and spoken responses.
	Chat UI (29.5%)	The interaction takes place through a conversational interface, such as a chat window.
	Other (3.8%)	Any other form of user interface that does not fit into the categories above.
System Feedback	*What is the output medium of the system's feedback?*	
	Visualization (42.3%)	The system provides feedback through visual means, such as graphs, charts, or images.
	Text (69.2%)	The system provides feedback through written or printed text.
	Generated Speech (9.0%)	The system provides feedback through generated or synthesized speech.
	Other (10.3%)	The system provides other feedback, such as haptic.
Initiation	*How is system information triggered?*	
	User-Initiated (20.5%)	The user actively triggers the system's information or response when needed.
	System-Initiated (85.9%)	The system automatically initiates information or feedback without user intervention.
Agency	*How is system information integrated into the dialogue?*	
	User agency (75.6%)	The user decides how and when the system's information is integrated into the dialogue.
	System agency (25.6%)	The system takes a more active role in integrating its information into the dialogue, without waiting for the user's intervention.

Table 4: Interaction: Dimensions, Codes, and Definitions

Letting the system provide augmentations during the dialogue may cause distraction, whereas providing feedback afterward may be less useful for some applications. We discuss this more extensively in section 6.

Adaptive Agents. We ask which of the agents is adapting based on the feedback received during the interaction. Holter and El-Assady [50] argue that there is an assumption in the setting of human-AI collaboration that a human agent will generally always learn from the AI. Therefore, we only provide the categorization of whether the system does not have the capacity to learn from the user during the interaction (user-adaptive only) or whether the collaboration is generally co-adaptive, which is more common. Typical co-adaptive systems include early transcription models, where users provide corrections to system transcriptions [91] or chatbots that augment dialogue [132].

User Interface Paradigm. Since we are discussing augmentation systems that engage users beyond automated text input, we also examine how users manually interact with model outputs through the user interface (UI).

Most fundamental interfaces may be text interfaces with support for reading, writing, or editing the system output [84, 91]. Slightly different are chat UIs where the user may engage in turn-taking exchanges with a chatbot, which may yield more flexible and adaptive interaction [132]. Voice UI systems such as smart speakers permit input directly through voice, which may reduce the barrier for interaction [3, 15]. Finally, as a broader setting, we consider dashboards or other graphical user interfaces, which may display modalities of system feedback beyond text. The modality in which users input information into such systems can have an effect on user behavior. For instance, users tend to ask more and different types of questions to an AI agent when they can write out their query, rather than querying in spoken voice [78].

	Code	Definition
Model Type	*What type of model is used for text processing?*	
	Rule-Based (24.4%)	The model relies on a predefined set of rules to process input.
	Statistical ML Model (37.2%)	The model uses statistical machine learning techniques for language processing.
	Deep Neural Model (7.7%)	The model employs deep learning techniques, such as neural networks, to process input.
	LLM-Based (25.6%)	The model is based on a large language model (LLM), leveraging pre-training.
Rhetoric	*Is the system output language adapted to have a specific style?*	
	Affective (6.4%)	The system output is expressed with emotionally nuanced language and/or tone.
	Informational (5.1%)	The system output is expressed to provide information as effectively as possible.
Input Features	*What user input is used for the task?*	
	Text Content Features (75.6%)	The model processes textual input, analyzing the content of the dialogue.
	Acoustic Features (28.2%)	The model uses acoustic features such as prosody, voice direction, and tone for processing.
	Visual Cues (14.1%)	The model utilizes visual cues, such as gaze or gesturing.
	Other (15.4%)	The model tracks other other types of input features recorded during the dialogue.

Table 5: Model: Dimensions, Codes, and Definitions

System Feedback Type. The output medium of the system is related to the discussion of the UI and is a high-cardinality domain, where again, multiple codes may apply per system. We focus on the most predominant, namely, text, generated speech, and visualizations such as images [87, 129], animations [36, 58], or graphs. Visualization may be useful in communicating sentiment [7, 10] or for visualizing numerical feedback [54], whereas generated speech is often used for natural dialogue with voice-assistants [15, 144]. Finally, the other category captures an interesting range of modalities that have been explored, such as haptic feedback for deaf or hard of hearing users [130], or somaesthetic systems [47] to improve co-experience during remote dialogue. When the system feedback type overlaps with the dialogue modality, dialogue augmentation systems have the means to intervene more actively. This may be the case for text and generated speech codes.

Initiation. Independently of the mode of its output, a system is strongly characterized by the way its information is entered into the dialogue. We thereby differentiate between two paradigms. User-initiated means that the user explicitly requests information and thereby triggers the system output—the system is *reactive*. Commonly known examples of fully user-initiated systems are smart speakers or personal assistants [3, 15]. System-initiated refers to when the system automatically provides information without being prompted by a user—the system is *proactive*. A proactive system is especially helpful when the user might not know when they would benefit from augmentation, due to a lack of domain knowledge and/or disability, as in language learning/speech therapy [19, 43]. User-initiated systems are chosen for the fact that people can control their intervention, and are less distracting, whereas system-initiated feedback are highly beneficial for just-in-time feedback [87, 128, 139].

Agency. In the dialogue augmentation context, agency refers to who controls the integration of system output in the discussion. User agency means that the user can decide whether system suggestions are brought up in a dialogue, whereas system agency means such suggestions may autonomously intervene in the dialogue [25, 72]. Li et al. [84] study both types of agency for the use case of documenting clinical notes from a doctor-patient consultation. In the scenario of user agency, a human has the authority to correct clinical notes proposed by the system, whereas in the scenario of system agency, clinical notes are generated solely by the system, which performs worse with regard to user satisfaction. This indicates, that user agency is the preferable design choice when a system doesn't reach satisfactory performance. Other work shows that giving the system the autonomy to intervene may be useful for interrupting conversations or debates when they go emotionally [41, 100] or topically [25] awry, or to more actively promote user inclusion and engagement [72]. Notably, there is a nuanced distinction to initiation: a system may bring up suggestions autonomously (system-initiated feedback), without having the agency intervene in the discussion (user agency), as is common for messaging recommendations in assistive communication [128] or for supporting certain dialogue roles [82].

4.5 Model

As the capabilities of speech, and language processing systems have continued to evolve and enabled the generation of more capable agents, an interesting aspect of dialogue augmentation systems is a consideration of the underlying modeling. Namely, we discuss the model architecture, the modalities processed by the model, and stylistic considerations for generating model output.

Model Type. Our loose definition of *intelligence* permits a wide variety of dialogue augmentation systems and allows us to evaluate the capabilities of such systems over time. Rather simple, rule-based

models were dominant in initial speech processing systems, such as simple word counters for speech training [45]. However, they may still be relevant for systems with functionalities such as measuring voice modulation [19, 47], engagement [72, 115] from acoustic activity, or in cases in which computational power is limited by design. Over time, statistical methods (e.g., linear models, decision trees) [94, 132], or deep neural networks [10, 84] were trained throughout various stages of the system pipeline. However, many such components are increasingly implemented using transformer-based LLMs [70, 74, 87, 124, 128] , as prompting customized conversational LLMs offers a low-resource[5] method to provide a range of tasks in a user-adaptive and personalized manner [132]. We differentiate between LLMs and other DNN architectures to symbolize the increasing importance of this paradigm in dialogue augmentation.

Rhetoric. The influence of the system output on the interaction outcome is commonly evaluated, especially when the model provides emotional, or task-specific guidance. Frequently, the goal of such adaptation is to enhance the user's trust in AI decisions [21, 29], be it through a more affective or informational style. Systems with an affective expressive style aim to be perceived as more empathetic [126], encouraging [114], or supportive [82]. Other systems, however, focus on delivering information concisely [139] or assertively and persuasively [41], which we categorize as informational rhetoric.

Input Features. Finally, we categorize the types of model input data, as the input features encode different aspects of the dialogue. Our study is limited to systems that perform some form of speech-, or language processing. In settings, where either speech is transcribed into text before applying the augmentation [84, 99], or written dialogue, e.g., in the form of chat histories [37] or online forums [143], is processed, we denote the model's input as text content features. Many natural language processing algorithms depend on the input of textual content, e.g., summarization [111], detecting sentiment [7], or generating text [74]. However, communication generally involves a variety of paralinguistic information next to its direct content, such as in the form of acoustic features, or visual cues. For instance, visual cues such as eye-gaze [66], mimics, or deictic gestures [105, 141] may be used to capture user attention, engagement, or non-verbal content. Especially in the domain of assistive technology, paralinguistic information can be crucial to facilitate or even enable communication [57, 66, 71].

5 Design Space Usage Scenarios

In the following, we demonstrate the tangible utility of the design space for designing dialogue augmentation systems in specific, hypothetical application scenarios. To this end, we derive requirements from existing formative or Wizard-of-Oz studies for three different domains and highlight the ways in which using the design space may improve the tool design procedure by more precisely aligning system features with user needs and mitigating the risk of adverse consequences.

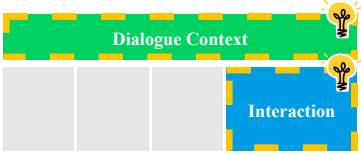
[5]Low-resource, as a user can avoid full model re-training.

5.1 Ideating on an On-Demand Fact-Checking System in Political Debates

Televised political debates are a crucial part of democracy that help citizens form an opinion on parties and candidates. However, it can be challenging for the viewer to identify which factual claims from the candidates can be believed. Suppose, a news channel plans to create an on-demand fact-checking tool [61, 77] for such debates to support the user in assessing the candidates' credibility. After completing research on previous work, the product developers start to design a prototype. While reviewing the design space, they observe that they have already considered numerous aspects in their design, such as the dialogue ***Modality*** (spoken), ***# Participants*** (small group), ***Interaction Timing*** (during), and ***Task Type*** (analyzing). However, during their inspection of the design space, they realize their current prototype will provide prescriptive guidance for clarification, which has been shown to suffer from low user trust (cf. section 4). This could influence the public's opinion on the news channel's transparency and neutrality. Instead, they decide to only highlight discrepancies between the candidates' claims and expert articles and to guide the viewers towards doing more research on their own by linking external knowledge sources to specific keywords used by the candidates (orienting guidance for reflection). In this scenario, the design space helps the designers ideate on different augmentation tasks and follow company policy to mitigate the risk of losing public trust and neutrality.

5.2 A Requirements Survey for a Video Conferencing Augmentation System

In a work meeting, video call participants often take notes during the call to document important information and to remember upcoming tasks. This can increase the mental load of the participants and lead to less efficiency during the meeting. In this example, a group of researchers wants to tackle this problem by implementing a tool that transcribes and summarizes the video conference dialogues [46, 114]. In a pre-study, they plan to extract the video call participants' needs through a survey to ideate on the summarization task of the AI model. To this end, they first consult our design space to understand what dimensions they may ask the user about. First, the design space helps them narrow their scope as they specify the context of the dialogue they want to support through the ***Dialogue Purpose*** as well as the ***Group Size***. In the next, exploratory stage, they discover the design space axis of ***Interaction Timing*** and realize they need to understand requirement trade-offs between the utility and distraction provided by augmentation during or after the dialogue takes place. They further discover the ***Adaptivity*** and ***UI Paradigm*** categories and realize that different people may prefer interactively personalizing summaries with the points most pertinent to them,



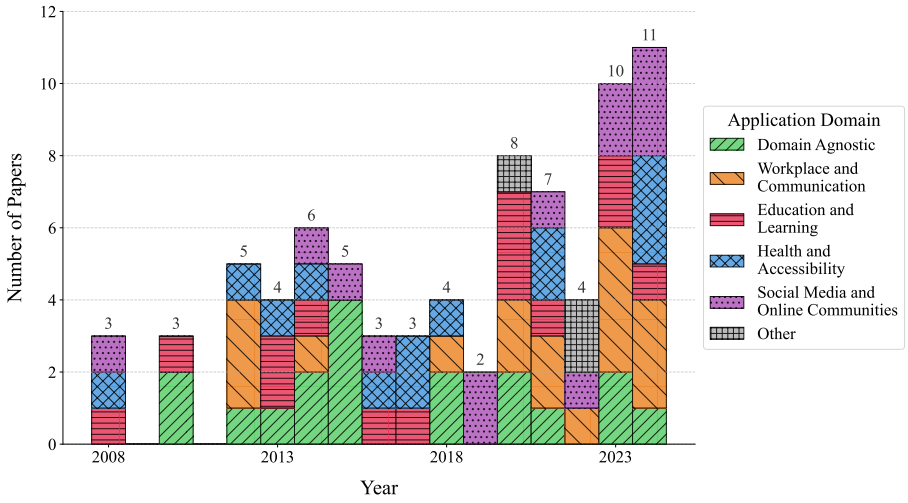
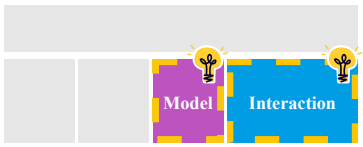


Figure 3: Papers within the project scope, by year and application domain. The works covered by our scope cover many applications across fields and have recently gained significant popularity.

which they include in the survey. In this scenario, the design space inspires additional questions for the survey, possibly leading to fewer iterations in the prototype development and increasing the design's user-friendliness.

5.3 A Dialogue Augmentation Tool for Combatting Implicit Bias in Medical Care

Implicit bias in medical care regarding gender, country of origin, ethnicity, and sexual orientation can harm the patient's health [35, 42]. Furthermore, it can also lead to a loss of trust in the medical provider [28, 44, 112]. Suppose a group of consultants is tasked by a hospital to develop a software prototype to help mitigate implicit bias of healthcare workers [14] during a consultation with a patient. To this end, the consultants ideate metrics on how to measure implicit bias and create mock-ups to visualize the model's results in a dashboard to educate the healthcare workers on their implicit bias after their interaction. However, while refining their interaction design using the design space, they find the ***System Feedback*** category and realize that implementing system-initiated feedback during the interaction could be a crucial feature to create immediate bias awareness before the consultation ends. Further inspection of the ***Input Features*** makes them realize that such bias is frequently not only exerted in the content of speech but also in acoustic features, such as prosody or visual cues, which leads them to add such input channels for their system. Finally, the design space makes the consultants think about the different types of feedback a system could provide in a shared dashboard interface. In this use case, the design space helps to expand the features of the tool, thus making it more efficient for combatting implicit bias in medical consultations.



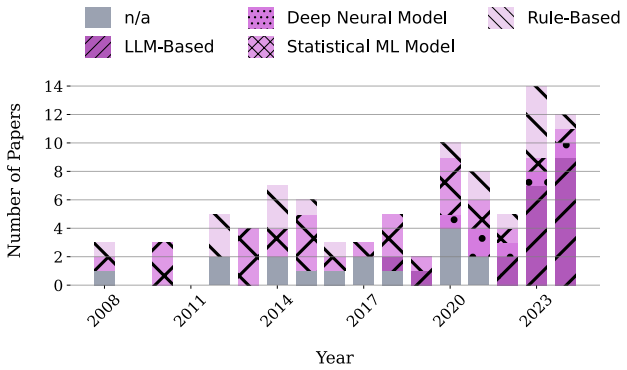


Figure 4: Usage of *Model Type* by year.

6 Discussion

In this section, we analyze the trends, opportunities, and challenges for dialogue augmentation systems. Our main focus thereby lies on the interplay between individual dimensions of the design space and their effect on user perception.

6.1 Trends

Figure 3 shows the overall distribution of the number of papers that fit our scope, grouped by publication year and application domain. It is immediately visible that there is an upward trend in the overall number of works in this domain. Most likely this stems from a combination of the development of more powerful language models from around 2018 [31], and the rise in tools built to support remote co-experience after the COVID-19 pandemic. We further see an increase in domain-specific applications, which promisingly indicates the practical adoption of such systems across domains.

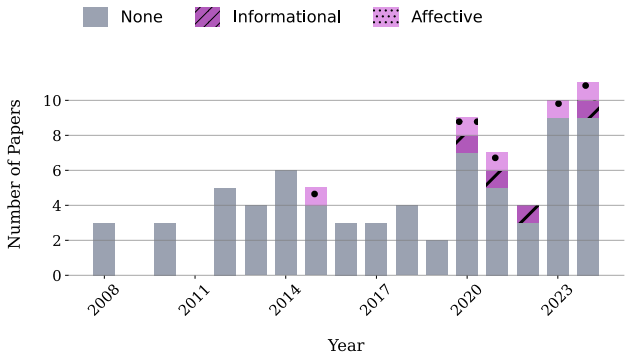


Figure 5: Implementation of a *Rhetoric* style by year.

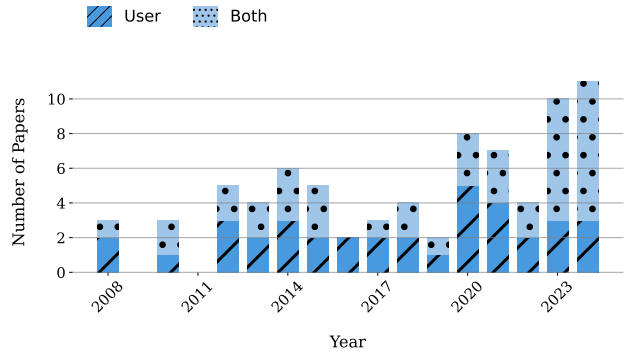


Figure 6: *Adaptive Agents* in systems by year.

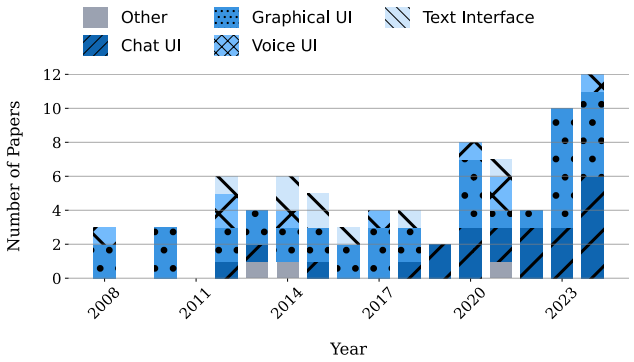


Figure 7: Usage of *UI Paradigms* by year.

***Models*, *Tasks*, *and Roles*.** Research about effectively facilitating dialogue using language processing systems goes back many years. However, due to historically high text processing error rates, many early systems focused on the basic interaction where humans corrected erroneous system outputs [65, 91, 97]. For instance, Munteanu et al. [91] report typical transcription word error rates of 40-45%, which they improve via corrective human interventions. The human's role in validating model output is still crucial in safety-critical applications [84], however, we note that some modern collaborative settings go as far as switching these roles, where now AI agents provide corrective feedback to human agents, e.g., to encourage human self-reflection [93]. This exemplifies how improvements in technical capabilities enable more adaptive and personalized language model interactions, which have generally empowered such systems to take on active supportive roles during dialogues by providing various types of real-time user guidance.This trend can be quantitatively observed in Figure 4, which shows a surge in LLM-based systems as well as other deep neural networks , with a simultaneous rise in co-adaptivity (cf. Figure 6) through interaction paradigms (e.g., Chat UIs, cf. Figure 7) that enable models to adapt more interactively to user inputs. Finally, not only are such models becoming more adaptive during user interaction, but the simplicity of adaptation through prompting in modern LLMs lowers the barrier for developers to generate domain-adapted and personalized models, for which the recent rise in rhetorically adapted models is a good example (cf. Figure 5).

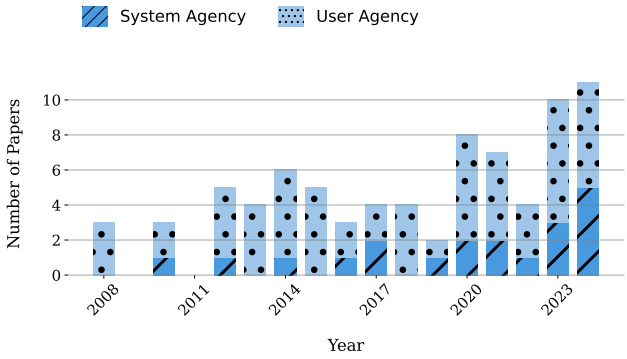


Figure 8: *Agency* assignment by year.

***Agency* and *Initiation*.** Further, we observe the trend that the feedback of such systems generally becomes more system-initiated. Beyond more system-initiated feedback, we also notice a trend in increasingly agentic systems (cf. Figure 8), in which the system can provide guidance or information by directly intervening in the dialogue. By combining these axes, we identify a direction of works that evaluate AI agents as increasingly emancipated and proactive dialogue participants who engage with human agents in a specific role [25, 117] or as equal collaborators [70].

In the framework of *dialogue as a joint activity*, this means that AI agents in such mediation or moderation roles can change the hierarchy of dialogue, as they *redistribute* and *claim* control. The potential positive impacts of increased AI agency and initiation in dialogue augmentation are manifold, such as the effect on social dynamics in regard to speaker balance, topic diversity [72], building interpersonal connection or communication flow [122], the facilitation of communication in assistive settings [57], and efficiency of communication in group settings [72].

6.2 Challenges and Opportunities

However, achieving effective augmentation for dialogue is challenging and some open problems need to be tackled by future research. The overall challenge with designing effective interventions is well-captured by Boyd et al. [19]: "*Interventions are most successful when the people receiving them perceive them to be useful and can tolerate their delivery.*"

Augmentation Distractiveness. The first concern is the distractive effect that augmentations have and the resulting disruption of the conversation flow. This is most prominent in systems that have the agency to interrupt the dialogue, however, some degree of distraction occurs as soon as virtually any information is automatically surfaced to a user in a system-initiated way. Surfacing small artifacts, such as pop-up images, have been described to not interrupt conversation flow too much [87]. However, displaying too much content or longer text passages may make the user break eye contact and lose concentration [84, 99]. This relates to the need to explicitly and carefully design the amount and type of guidance that a user can process without interrupting the dialogue flow, which has been previously explored for chatbot interactions [139]. The severity of the distraction perceived by the users may also depend on the dialogue context, such as the number of participants [87]. Ways of avoiding the disruptiveness of system-initiated feedback may be to develop more easily accessible *user-initiated* feedback mechanisms [139], restructuring dialogues to include conscious breaks for receiving system feedback [114, 128, 132], or to provide system feedback in subtle and private ways, e.g., through the usage of wearables [19, 94]. We encourage future research to explore other non-distractive methods for system-initiated feedback in further settings and use cases.

User Trust and Privacy. The second requirement to perceive an intervention as beneficial is to ensure that its content is perceived to be *useful*. This relates to both a discussion on model accuracy, and user trust. Whereas some users may distrust model-generated feedback from the beginning [114], users may also lose trust upon receiving inaccurate, incomplete, or irrelevant system feedback [84, 144]. Further, some users in AI-mediated communication may even carelessly over-rely on system generations, causing a grey zone for user trust and accountability [49]. As such, further research is needed to improve user trust by equipping system feedback with understandable model explanations or giving back agency to users via information provenance. Finally, data privacy is a frequent concern for users of dialogue augmentation devices such as smart speakers, as many users are uncertain when such devices are recording and what data is being collected [1]. This is especially concerning due to their usage in sensitive environments, such as homes and classrooms [15, 144]. Better educating users about enforced data protection measures, along with developing secure and transparent devices, could help address this issue.

Accessibility. Another aspect of usefulness is accessibility. An intervention may be more likely to be perceived as useful when it is developed with the diverse preferences and needs of the target audience(s), including those who belong to underrepresented groups. Lee et al. [81] refer to this approach as *value-sensitive design*, which can be guided by a comprehensive design space for dialogue augmentation such as the one we propose. For instance, accommodating individuals with cognitive and sensory disabilities might involve tailoring augmentation strategies to such specific user needs. This may include addressing augmentation contexts (e.g., aligning augmentation purpose and structure to support memory or comprehension), tasks (e.g., defining augmentation goals that prioritize accessibility and integrating guidance for clearer interactions), interaction (e.g., adapting timing, UI paradigms, system feedback, initiation mechanisms, and user agency to reduce cognitive load), and model rhetoric (e.g., employing empathetic or plain language to ensure clarity and inclusivity).

6.3 Limitations

We briefly describe some limitations of this study. Firstly, this design space covers the specific scope of intelligent dialogue augmentation systems with a special focus on language processing. As such, the study does not, or only tangentially, cover works done in Wizard-of-Oz studies, or studies describing techniques that could potentially be applied to augment dialogue, for instance. The authors choose the inclusion criteria of the study in an attempt to provide an interesting and useful discussion while aiming to limit the scope in this fast-evolving, productive research area. Further, although much effort was put into the design space to make it relevant to a broad range of dialogue augmentation systems, the ongoing rapid AI advancements may require us to rethink parts of the design space within an uncertain time. Finally, although we ground the design space using established concepts in social interaction studies and human-AI collaboration as well as multiple stages of author discussions, we acknowledge that the design space is by no means exhaustive, or entirely non-subjective. It contains axes that the authors consider relevant to provide an initial characterization of the broader space of dialogue augmentation systems and we encourage further research to expand or improve the design space as dialogue augmentation systems further develop.

7 Conclusion

In this work, we develop a design space for intelligent dialogue augmentation systems on five axes—dialogue context, augmentation context, task, interaction, and model—through a literature study and an iterative, mixed-methods refinement. We further ground our analysis by connecting it to established categorizations of both *dialogue as a joint activity* and *human-AI collaboration*. For each dimension, we identify relevant categories that characterize dialogue augmentation systems, analyze current trends, evaluate the impact of design choices on dialogue dynamics and user perception, and, finally, detect challenges and research opportunities for future dialogue augmentation systems. More concretely, we find that recent improvements in language modeling have led to the development of highly adaptive and personalized user interactions which change the role of AI agents with respect to their initiation and agency. We find that in most dialogue contexts, however, designing *trusted*, *seamless*, and *timely* augmentation is key to the adoption of dialogue augmentation systems for broader use. We hope this foundational work establishes a common ground in this prospering space and inspires further system studies to evaluate effects across the proposed design space axes.

Acknowledgments

The authors thank the anonymous reviewers for their efforts to provide constructive and helpful feedback. We also thank Rita Sevastjanova for providing useful feedback in the early stages of the project and the final draft. Further, we thank Mina Lee and her co-authors for granting permission to re-use their visual formatting of the design space. RC acknowledges support by fyayc. MEA acknowledges support from the SNF grant "Personalized Visual Analytics: Human Preference Elicitation for Ranking-based Multi-Criteria Decision-Support" (project number 200021-231821).

References

[1] Noura Abdi, Kopo M. Ramokapane, and Jose M. Such. 2019. More than Smart Speakers: Security and Privacy Perceptions of Smart Home Personal Assistants. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 451–466. https://www.usenix.org/conference/soups2019/presentation/abdi

[2] Martin Adam, Michael Wessel, and Alexander Benlian. 2020. AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets* 31 (2020), 427 – 445. https://api.semanticscholar.org/CorpusID:216306221

[3] Karan Ahuja, Andy Kong, Mayank Goel, and Chris Harrison. 2020. Direction-of-Voice (DoV) Estimation for Intuitive Speech Interaction with Smart Devices Ecosystems. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 1121–1131. https://doi.org/10.1145/3379337.3415588

[4] Annalena Bea Aicher, Daniel Kornmüller, Wolfgang Minker, and Stefan Ultes. 2023. Self-imposed filter bubble model for argumentative dialogues. In *Proceedings of the 5th international conference on conversational user interfaces*. 1–11.

[5] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[6] Pengcheng An, Kenneth Holstein, Bernice d'Anjou, Berry Eggen, and Saskia Bakker. 2020. The TA Framework: Designing Real-time Teaching Augmentation for K-12 Classrooms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3313831.3376277

[7] Pengcheng An, Jiawen Stefanie Zhu, Zibo Zhang, Yifei Yin, Qingyuan Ma, Che Yan, Linghao Du, and Jian Zhao. 2024. EmoWear: Exploring Emotional Teasers for Voice Message Interaction on Smartwatches. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 279, 16 pages. https://doi.org/10.1145/3613904.3642101

[8] Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and Giulio Jacucci. 2018. Investigating proactive search support in conversations. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 1295–1307.

[9] Omer Anjum, Chak Ho Chan, Tanitpong Lawphongpanich, Yucheng Liang, Tianyi Tang, Shuchen Zhang, Wen-mei Hwu, Jinjun Xiong, and Sanjay Patel. 2020. Vertext: An end-to-end ai powered conversation management system for multi-party chat platforms. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*. 1–6.

[10] Toshiki Aoki, Rintaro Chujo, Katsufumi Matsui, Saemi Choi, and Ari Hautasaari. 2022. EmoBalloon - Conveying Emotional Arousal in Text Chats with Speech Balloons. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 527, 16 pages. https://doi.org/10.1145/3491102.3501920

[11] Riku Arakawa and Hiromu Yakura. 2021. Mindless Attractor: A False-Positive Resistant Intervention for Drawing Attention Using Auditory Perturbation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 99, 15 pages. https://doi.org/10.1145/3411764.3445339

[12] Amos Azaria, Ariella Richardson, and Sarit Kraus. 2015. An agent for deception detection in discussion based environments. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 218–227.

[13] S. Sandra Bae, Clement Zheng, Mary Etta West, Ellen Yi-Luen Do, Samuel Huron, and Danielle Albers Szafir. 2022. Making Data Tangible: A Cross-disciplinary Design Space for Data Physicalization. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 81, 18 pages. https://doi.org/10.1145/3491102.3501939

[14] Emily Bascom, Reggie Casanova-Perez, Kelly Tobar, Manas Satish Bedmutha, Harshini Ramaswamy, Wanda Pratt, Janice Sabin, Brian Wood, Nadir Weibel, and Andrea Hartzler. 2024. Designing Communication Feedback Systems To Reduce Healthcare Providers' Implicit Biases In Patient Encounters. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 452, 12 pages. https://doi.org/10.1145/3613904.3642756

[15] Erin Beneteau, Ashley Boone, Yuxing Wu, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2020. Parenting with Alexa: Exploring the Introduction of Smart Speakers on Family Dynamics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376344

[16] Ivo Benke, Michael Thomas Knierim, and Alexander Maedche. 2020. Chatbot-based emotion management for distributed teams: A participatory design study. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–30.

[17] Ivo Benke, Sebastian Vetter, and Alexander Maedche. 2021. LeadBoSki: A Smart Personal Assistant for Leadership Support in Video-Meetings. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 19–22. https://doi.org/10.1145/3462204.3481764

[18] Adrian Boteanu and Sonia Chernova. 2013. Modeling discussion topics in interactions with a tablet reading primer. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces* (Santa Monica, California, USA) *(IUI '13)*. Association for Computing Machinery, New York, NY, USA, 75–84. https://doi.org/10.1145/2449396.2449409

[19] LouAnne E. Boyd, Alejandro Rangel, Helen Tomimbang, Andrea Conejo-Toledo, Kanika Patel, Monica Tentori, and Gillian R. Hayes. 2016. SayWAT: Augmenting Face-to-Face Conversations for Adults with Autism. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4872–4883. https://doi.org/10.1145/2858036.2858215

[20] Thomas Breideband, Jeffrey Bush, Chelsea Chandler, Michael Chang, Rachel Dickler, Peter Foltz, Ananya Ganesh, Rachel Lieber, William R. Penuel, Jason G. Reitman, John Weatherley, and Sidney D'Mello. 2023. The Community Builder (CoBi): Helping Students to Develop Better Small Group Collaborative Learning Skills. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 376–380. https://doi.org/10.1145/3584931.3607498

[21] Francisco Maria Calisto, João Fernandes, Margarida Morais, Carlos Santiago, João Maria Abrantes, Nuno Nunes, and Jacinto C. Nascimento. 2023. Assertiveness-based Agent Communication for a Personalized Medicine on Medical Imaging Diagnosis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 13, 20 pages. https://doi.org/10.1145/3544548.3580682

[22] Ann Frances Cameron and Jane Webster. 2005. Unintended consequences of emerging communication technologies: Instant Messaging in the workplace. *Computers in Human Behavior* 21, 1 (2005), 85–103. https://doi.org/10.1016/j.chb.2003.12.001

[23] David Ceneda, Theresia Gschwandtner, Thorsten May, Silvia Miksch, Hans-Jörg Schulz, Marc Streit, and Christian Tominski. 2017. Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 111–120. https://doi.org/10.1109/TVCG.2016.2598468

[24] Fanglin Chen, Kewei Xia, Karan Dhabalia, and Jason I. Hong. 2019. MessageOnTap: A Suggestive Interface to Facilitate Messaging-related Tasks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300805

[25] Chun-Wei Chiang, Zhuoran Lu, Zhuoyan Li, and Ming Yin. 2024. Enhancing AI-Assisted Group Decision Making through LLM-Powered Devil's Advocate. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (Greenville, SC, USA) *(IUI '24)*. Association for Computing Machinery, New York, NY, USA, 103–119. https://doi.org/10.1145/3640543.3645199

[26] Yi-Shyuan Chiang, Ruei-Che Chang, Yi-Lin Chuang, Shih-Ya Chou, Hao-Ping Lee, I-Ju Lin, Jian-Hua Jiang Chen, and Yung-Ju Chang. 2020. Exploring the Design Space of User-System Communication for Smart-home Routine Assistants. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376501

[27] Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511620539

[28] Lisa A Cooper, Debra L Roter, Kathryn A Carson, Mary Catherine Beach, Janice A Sabin, Anthony G Greenwald, and Thomas S Inui. 2012. The associations of

clinicians' implicit attitudes about race with medical visit communication and patient ratings of interpersonal care. *American journal of public health* 102, 5 (2012), 979–987.

[29] Samuel Rhys Cox and Wei Tsang Ooi. 2022. Does Chatbot Language Formality Affect Users' Self-Disclosure?. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) *(CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 1, 13 pages. https://doi.org/10.1145/3543829.3543831

[30] Mary Czerwinski and Edward Cutrell. 2000. Instant Messaging and Interruption: Influence of Task Type on Performance. https://api.semanticscholar.org/CorpusID:9439335

[31] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[32] Ann Hill Duin and Ray Archee. 1996. Collaboration via E-mail and Internet Relay Chat: Understanding Time and Technology. *Technical Communication* 43, 4 (1996), 402–412.

[33] Andy Echenique, Naomi Yamashita, Hideaki Kuzuoka, and Ari Hautasaari. 2014. Effects of video and text support on grounding in multilingual multiparty audio conferencing. In *Proceedings of the 5th ACM International Conference on Collaboration across Boundaries: Culture, Distance & Technology* (Kyoto, Japan) *(CABS '14)*. Association for Computing Machinery, New York, NY, USA, 73–81. https://doi.org/10.1145/2631488.2631497

[34] Mennatallah El-Assady and Caterina Moruzzi. 2022. Which biases and reasoning pitfalls do explanations trigger? Decomposing communication processes in human–AI interaction. *IEEE Computer Graphics and Applications* 42, 6 (2022), 11–23.

[35] Chloë FitzGerald and Samia Hurst. 2017. Implicit bias in healthcare professionals: a systematic review. *BMC medical ethics* 18 (2017), 1–18.

[36] Siwei Fu, Jian Zhao, Hao Fei Cheng, Haiyi Zhu, and Jennifer Marlow. 2018. T-Cal: Understanding Team Conversational Data with Calendar-based Visualization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174074

[37] Ge Gao, Bin Xu, David C. Hau, Zheng Yao, Dan Cosley, and Susan R. Fussell. 2015. Two is Better Than One: Improving Multilingual Collaboration by Giving Two Machine Translation Outputs. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 852–863. https://doi.org/10.1145/2675133.2675197

[38] Ge Gao, Naomi Yamashita, Ari MJ Hautasaari, Andy Echenique, and Susan R. Fussell. 2014. Effects of public vs. private automated transcripts on multiparty communication between native and non-native english speakers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 843–852. https://doi.org/10.1145/2556288.2557303

[39] R. Kelly Garrett and James N. Danziger. 2007. IM=Interruption Management? Instant Messaging and Disruption in the Workplace. *J. Comput. Mediat. Commun.* 13 (2007), 23–42. https://api.semanticscholar.org/CorpusID:8013586

[40] Sven Gehring, Markus Löchtefeld, Florian Daiber, Matthias Böhmer, and Antonio Krüger. 2012. Using intelligent natural user interfaces to support sales conversations. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*. 97–100.

[41] Jarod Govers, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2024. AI-Driven Mediation Strategies for Audience Depolarisation in Online Debates. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 803, 18 pages. https://doi.org/10.1145/3613904.3642322

[42] Alexander R Green, Dana R Carney, Daniel J Pallin, Long H Ngo, Kristal L Raymond, Lisa I Iezzoni, and Mahzarin R Banaji. 2007. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of general internal medicine* 22 (2007), 1231–1238.

[43] Joshua Hailpern, Andrew Harris, Reed La Botz, Brianna Birman, and Karrie Karahalios. 2012. Designing visualizations to facilitate multisyllabic speech with children with autism and speech delays. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) *(DIS '12)*. Association for Computing Machinery, New York, NY, USA, 126–135. https://doi.org/10.1145/2317956.2317977

[44] William J Hall, Mimi V Chapman, Kent M Lee, Yesenia M Merino, Tainayah W Thomas, B Keith Payne, Eugenia Eng, Steven H Day, and Tamera Coyne-Beasley. 2015. Implicit racial/ethnic bias among health care professionals and its influence on health care outcomes: a systematic review. *American journal of public health* 105, 12 (2015), e60–e76.

[45] Foad Hamidi and Melanie Baljko. 2014. Rafigh: A living media interface for speech intervention. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1817–1820. https://doi.org/10.1145/2556288.2557402

[46] Ari Hautasaari and Naomi Yamashita. 2014. Do automated transcripts help non-native speakers catch up on missed conversation in audio conferences?. In *Proceedings of the 5th ACM International Conference on Collaboration across Boundaries: Culture, Distance & Technology* (Kyoto, Japan) *(CABS '14)*. Association for Computing Machinery, New York, NY, USA, 65–72. https://doi.org/10.1145/2631488.2631495

[47] Sjoerd Hendriks, Simon Mare, Mafalda Gamboa, and Mehmet Aydın Baytaş. 2021. Azalea: Co-experience in Remote Dialog through Diminished Reality and Somaesthetic Interaction Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 261, 11 pages. https://doi.org/10.1145/3411764.3445052

[48] Marion A. Hersh and Michael A. Johnson. 2003. *Assistive Technology for the Hearing-impaired, Deaf and Deafblind*. Springer-Verlag London, London, England, UK. https://doi.org/10.1007/978-1-4471-0109-3

[49] Jess Hohenstein and Malte Jung. 2020. AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. *Computers in Human Behavior* 106 (2020), 106190. https://doi.org/10.1016/j.chb.2019.106190

[50] Steffen Holter and Mennatallah El-Assady. 2024. Deconstructing Human-AI Collaboration: Agency, Interaction, and Adaptation. *Computer Graphics Forum* 43 (06 2024). https://doi.org/10.1111/cgf.15107

[51] Enamul Hoque and Giuseppe Carenini. 2015. Convisit: Interactive topic modeling for exploring asynchronous online conversations. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. 169–180.

[52] Enamul Hoque and Giuseppe Carenini. 2016. MultiConVis: A Visual Text Analytics System for Exploring a Collection of Online Conversations. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 96–107. https://doi.org/10.1145/2856767.2856782

[53] Enamul Hoque, Shafiq Joty, Luis Marquez, and Giuseppe Carenini. 2017. CQAVis: Visual text analytics for community question answering. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. 161–172.

[54] Reiya Horii, Yurike Chandra, Kai Kunze, and Kouta Minamizawa. 2020. Bubble Visualization Overlay in Online Communication for Increased Speed Awareness and Better Turn Taking. In *Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 59–61. https://doi.org/10.1145/3379350.3416185

[55] Ting-Hao Huang, Joseph Chee Chang, and Jeffrey P Bigham. 2018. Evorus: A crowd-powered conversational assistant built to automate itself over time. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[56] Maggie Hughes and Deb Roy. 2020. Keeper: An online synchronous conversation environment informed by in-person facilitation practices. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing*. 275–279.

[57] Inseok Hwang, Chungkuk Yoo, Chanyou Hwang, Dongsun Yim, Youngki Lee, Chulhong Min, John Kim, and Junehwa Song. 2014. TalkBetter: Family-driven mobile intervention care for children with language delay. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1283–1296. https://doi.org/10.1145/2531602.2531668

[58] Joshua E. Introne and Marcus Drescher. 2013. Analyzing the flow of knowledge in computer mediated teams. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 341–356. https://doi.org/10.1145/2441776.2441816

[59] Shamsi T. Iqbal and Eric Horvitz. 2007. Disruption and recovery of computing tasks: Field study, analysis, and directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 677–686. https://doi.org/10.1145/1240624.1240730

[60] R. Mark Isaac and James M. Walker. 1988. Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism. *The Quarterly Journal of Economics* 103, 1 (1988), 179–199. http://www.jstor.org/stable/1882648

[61] Petar Ivanov, Ivan Koychev, Momchil Hardalov, and Preslav Nakov. 2024. Detecting Check-Worthy Claims in Political Debates, Speeches, and Interviews Using Audio Data. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 12011–12015.

[62] Mahmood Jasim, Pooya Khaloo, Somin Wadhwa, Amy X Zhang, Ali Sarvghad, and Narges Mahyar. 2020. Communityclick: Towards improving inclusivity in town halls. In *Companion Publication of the 2020 Conference on Computer*

Supported Cooperative Work and Social Computing. 37–41.

[63] Mahmood Jasim, Pooya Khaloo, Somin Wadhwa, Amy X. Zhang, Ali Sarvghad, and Narges Mahyar. 2021. CommunityClick: Capturing and Reporting Community Feedback from Town Halls to Improve Inclusivity. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 213 (jan 2021), 32 pages. https://doi.org/10.1145/3432912

[64] Jeesu Jung, Hyein Seo, Sangkeun Jung, Riwoo Chung, Hwijung Ryu, and Du-Seong Chang. 2023. Interactive User Interface for Dialogue Summarization. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 934–957. https://doi.org/10.1145/3581641.3584057

[65] Vaiva Kalnikaitundefined, Patrick Ehlen, and Steve Whittaker. 2012. Markup as you talk: Establishing effective memory cues while still contributing to a meeting. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 349–358. https://doi.org/10.1145/2145204.2145260

[66] Shaun K. Kane and Meredith Ringel Morris. 2017. Let's Talk About X: Combining Image Recognition and Eye Gaze to Support Conversation for People with ALS. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) *(DIS '17)*. Association for Computing Machinery, New York, NY, USA, 129–134. https://doi.org/10.1145/3064663.3064762

[67] Panayu Keelawat. 2023. NBGuru: Generating Explorable Data Science Flowcharts to Facilitate Asynchronous Communication in Interdisciplinary Data Science Teams. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) *(CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 6–11. https://doi.org/10.1145/3584931.3607020

[68] Norbert L. Kerr and Steven E. Bruun. 1981. Ringelmann Revisited: Alternative Explanations for the Social Loafing Effect. *Personality and Social Psychology Bulletin* 7, 2 (1981), 224–231. https://doi.org/10.1177/014616728172007 arXiv:https://doi.org/10.1177/014616728172007

[69] Da-jung Kim and Youn-kyung Lim. 2015. Dwelling Places in KakaoTalk: Understanding the Roles and Meanings of Chatrooms in Mobile Instant Messengers. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 775–784. https://doi.org/10.1145/2675133.2675198

[70] Hanseob Kim, Bin Han, Jieun Kim, Muhammad Firdaus Syawaludin Lubis, Gerard Jounghyun Kim, and Jae-In Hwang. 2024. Engaged and Affective Virtual Agents: Their Impact on Social Presence, Trustworthiness, and Decision-Making in the Group Discussion. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 149, 17 pages. https://doi.org/10.1145/3613904.3642917

[71] JooYeong Kim, SooYeon Ahn, and Jin-Hyuk Hong. 2023. Visible Nuances: A Caption System to Visualize Paralinguistic Speech Cues for Deaf and Hard-of-Hearing Individuals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 54, 15 pages. https://doi.org/10.1145/3544548.3581130

[72] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. 2020. Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376785

[73] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator chatbot for deliberative discussion: Effects of discussion structure and discussant facilitation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26.

[74] Taewan Kim, Seolyeong Bae, Hyun Ah Kim, Su-Woo Lee, Hwajung Hong, Chanmo Yang, and Young-Ho Kim. 2024. MindfulDiary: Harnessing Large Language Model to Support Psychiatric Patients' Journaling. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 701, 20 pages. https://doi.org/10.1145/3613904.3642937

[75] Taewook Kim, Qingyu Guo, Hyeonjae Kim, Wenjie Yang, Meiziniu Li, and Xiaojuan Ma. 2022. Facilitating Continuous Text Messaging in Online Romantic Encounters by Expanded Keywords Enumeration. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*. 3–7.

[76] Joel Kiskola, Thomas Olsson, Heli Väätäjä, Aleksi H. Syrjämäki, Anna Rantasila, Poika Isokoski, Mirja Ilves, and Veikko Surakka. 2021. Applying Critical Voice in Design of User Interfaces for Supporting Self-Reflection and Emotion Regulation in Online News Commenting. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 88, 13 pages. https://doi.org/10.1145/3411764.3445783

[77] Travis Kriplean, Caitlin Bonnar, Alan Borning, Bo Kinney, and Brian Gill. 2014. Integrating on-demand fact-checking with public dialogue. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1188–1199. https://doi.org/10.1145/2531602.2531677

[78] Emily Kuang, Ehsan Jahangirzadeh Soure, Mingming Fan, Jian Zhao, and Kristen Shinohara. 2023. Collaboration with Conversational AI Assistants for UX Evaluation: Questions and How to Ask them (Voice vs. Text). In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 116, 15 pages. https://doi.org/10.1145/3544548.3581247

[79] Raina Langevin, Ross J Lordon, Thi Avrahami, Benjamin R. Cowan, Tad Hirsch, and Gary Hsieh. 2021. Heuristic Evaluation of Conversational Agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 632, 15 pages. https://doi.org/10.1145/3411764.3445312

[80] Sven Laumer, Fabian Tobias Gubler, A. A. Racheva, and Christian Maier. 2019. Use Cases for Conversational Agents: An Interview-based Study. In *Americas Conference on Information Systems*. https://api.semanticscholar.org/CorpusID:199165669

[81] Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L.C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, Antonette Shibani, Disha Shrivastava, Lila Shroff, Agnia Sergeyuk, Jessi Stark, Sarah Sterman, Sitong Wang, Antoine Bosselut, Daniel Buschek, Joseph Chee Chang, Sherol Chen, Max Kreminski, Joonsuk Park, Roy Pea, Eugenia Ha Rim Rho, Zejiang Shen, and Pao Siangliulue. 2024. A Design Space for Intelligent and Interactive Writing Assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 1054, 35 pages. https://doi.org/10.1145/3613904.3642697

[82] Sung-Chul Lee, Jaeyoon Song, Eun-Young Ko, Seongho Park, Jihee Kim, and Juho Kim. 2020. SolutionChat: Real-time Moderator Support for Chat-based Structured Discussion. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376609

[83] Stephen C. Levinson. 1979. Activity types and language. In *Talk at Work: Interaction in Institutional Settings*, Paul Drew and John Heritage (Eds.). Cambridge University Press, Cambridge, 66–100. https://doi.org/10.1515/ling.1979.17.5-6.365

[84] Brenna Li, Noah Crampton, Thomas Yeates, Yu Xia, Xirong Tian, and Khai Truong. 2021. Automating Clinical Documentation with Digital Scribes: Understanding the Impact on Physicians. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 658, 12 pages. https://doi.org/10.1145/3411764.3445172

[85] Haotian Li, Yun Wang, and Huamin Qu. 2024. Where are we so far? understanding data storytelling tools from the perspective of human-ai collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.

[86] Chengzhong Liu, Shixu Zhou, Dingdong Liu, Junze Li, Zeyu Huang, and Xiaojuan Ma. 2023. CoArgue : Fostering Lurkers' Contribution to Collective Arguments in Community-based QA Platforms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 271, 17 pages. https://doi.org/10.1145/3544548.3580932

[87] Xingyu "Bruce" Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Alex Olwal, Peggy Chi, Xiang "Anthony" Chen, and Ruofei Du. 2023. Visual Captions: Augmenting Verbal Communication with On-the-fly Visuals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 108, 20 pages. https://doi.org/10.1145/3544548.3581566

[88] Ye Liu, Stefan Ultes, Wolfgang Minker, and Wolfgang Maier. 2023. Unified Conversational Models with System-Initiated Transitions between Chit-Chat and Task-Oriented Dialogues. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 33, 9 pages. https://doi.org/10.1145/3571884.3597125

[89] Christian Löw, Lukas Moshuber, and Albert Rafetseder. 2020. Grätzelbot: Social Companion Technology for Community Building among University Freshmen. In *Chatbot Research and Design Workshop*. https://api.semanticscholar.org/CorpusID:231777839

[90] Shayan Monadjemi, Mengtian Guo, David Gotz, Roman Garnett, and Alvitta Ottley. 2023. Human–Computer Collaboration for Visual Analytics: an Agent-based Framework. In *Computer Graphics Forum*, Vol. 42. Wiley Online Library, 199–210.

[91] Cosmin Munteanu, Ron Baecker, and Gerald Penn. 2008. Collaborative editing for improved usefulness and usability of transcript-enhanced webcasts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 373–382. https://doi.org/10.1145/1357054.1357117

[92] Yukiko I Nakano and Ryo Ishii. 2010. Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In *Proceedings of the 15th international conference on Intelligent user interfaces*. 139–148.

[93] Tricia J. Ngoon, S Sushil, Angela E.B. Stewart, Ung-Sang Lee, Saranya Venkatraman, Neil Thawani, Prasenjit Mitra, Sherice Clarke, John Zimmerman, and Amy Ogan. 2024. ClassInSight: Designing Conversation Support Tools to Visualize Classroom Discussion for Personalized Teacher Professional Development. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 663, 15 pages. https://doi.org/10.1145/3613904.3642487

[94] Tien T. Nguyen, Duyen T. Nguyen, Shamsi T. Iqbal, and Eyal Ofek. 2015. The Known Stranger: Supporting Conversations between Strangers with Personalized Topic Suggestions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 555–564. https://doi.org/10.1145/2702123.2702411

[95] Alex Olwal, Kevin Balke, Dmitrii Votintcev, Thad Starner, Paula Conn, Bonnie Chinh, and Benoit Corda. 2020. Wearable Subtitles: Augmenting Spoken Communication with Lightweight Eyewear for All-day Captioning. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 1108–1120. https://doi.org/10.1145/3379337.3415817

[96] Mei-Hua Pan, Naomi Yamashita, and Hao-Chuan Wang. 2017. Task rebalancing: Improving multilingual communication with native speakers-generated highlights on automated transcripts. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 310–321.

[97] Yingxin Pan, Danning Jiang, Lin Yao, Michael Picheny, and Yong Qin. 2010. Effects of automated transcription quality on non-native speakers' comprehension in real-time computer-mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1725–1734. https://doi.org/10.1145/1753326.1753584

[98] Gun Woo (Warren) Park, Payod Panda, Lev Tankelevitch, and Sean Rintel. 2024. The CoExplorer Technology Probe: A Generative AI-Powered Adaptive Interface to Support Intentionality in Planning and Running Video Meetings. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1638–1657. https://doi.org/10.1145/3643834.3661507

[99] Yi-Hao Peng, Ming-Wei Hsi, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3173574.3173867

[100] Zhenhui Peng, Taewook Kim, and Xiaojuan Ma. 2019. GremoBot: Exploring Emotion Regulation in Group Chat. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 335–340. https://doi.org/10.1145/3311957.3359472

[101] Lara Piccolo, Azizah C Blackwood, Tracie Farrell, and Martino Mensio. 2021. Agents for fighting misinformation spread on Twitter: design challenges. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–7.

[102] Martin J. Pickering and Simon Garrod. 2021. *Understanding Dialogue: Language Use and Social Interaction*. Cambridge University Press. https://doi.org/10.1017/9781108610728

[103] Anne Marie Piper and James D. Hollan. 2008. Supporting medical conversations between deaf and hearing individuals with tabletop displays. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (San Diego, CA, USA) *(CSCW '08)*. Association for Computing Machinery, New York, NY, USA, 147–156. https://doi.org/10.1145/1460563.1460587

[104] Xun Qian, Feitong Tan, Yinda Zhang, Brian Moreno Collins, David Kim, Alex Olwal, Karthik Ramani, and Ruofei Du. 2024. ChatDirector: Enhancing Video Conferencing with Space-Aware Scene Rendering and Speech-Driven Layout Transition. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 909, 16 pages. https://doi.org/10.1145/3613904.3642110

[105] Francis Quek and Francisco Oliveira. 2013. Enabling the blind to see gestures. *ACM Trans. Comput.-Hum. Interact.* 20, 1, Article 4 (apr 2013), 32 pages. https://doi.org/10.1145/2442106.2442110

[106] Rahul Rajan, Cliff Chen, and Ted Selker. 2012. Considerate Audio MEdiating Oracle (CAMEO) improving human-to-human communications in conference calls. In *Proceedings of the Designing Interactive Systems Conference*. 86–95.

[107] Leon Reicherts, Gun Woo Park, and Yvonne Rogers. 2022. Extending Chatbots to Probe Users: Enhancing Complex Decision-Making Through Probing Conversations. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) *(CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 2, 10 pages. https://doi.org/10.1145/3543829.3543832

[108] Denise M. Rousseau, Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. 1998. Introduction to Special Topic Forum: Not so Different after All: A Cross-Discipline View of Trust. *The Academy of Management Review* 23, 3 (1998), 393–404. http://www.jstor.org/stable/259285

[109] Sherry Ruan, Liwei Jiang, Qianyao Xu, Zhiyuan Liu, Glenn M Davis, Emma Brunskill, and James A Landay. 2021. Englishbot: An ai-powered conversational system for second language learning. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 434–444.

[110] Steve Rubin, Floraine Berthouzoz, Gautham J. Mysore, and Maneesh Agrawala. 2015. Capture-Time Feedback for Recording Scripted Narration. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 191–199. https://doi.org/10.1145/2807442.2807464

[111] John Rudnik, Sharadhi Raghuraj, Mingyi Li, and Robin N Brewer. 2024. CareJournal: A Voice-Based Conversational Agent for Supporting Care Communications. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–22.

[112] Janice A Sabin, Brian A Nosek, Anthony G Greenwald, and Frederick P Rivara. 2009. Physicians' implicit and explicit attitudes about race by MD race, ethnicity, and gender. *Journal of health care for the poor and underserved* 20, 3 (2009), 896–913.

[113] Malak Sadek, Rafael A Calvo, and Celine Mougenot. 2023. Trends, Challenges and Processes in Conversational Agent Design: Exploring Practitioners' Views through Semi-Structured Interviews. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 13, 10 pages. https://doi.org/10.1145/3571884.3597143

[114] Samiha Samrose, Daniel McDuff, Robert Sim, Jina Suh, Kael Rowan, Javier Hernandez, Sean Rintel, Kevin Moynihan, and Mary Czerwinski. 2021. MeetingCoach: An Intelligent Dashboard for Supporting Effective & Inclusive Meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 252, 13 pages. https://doi.org/10.1145/3411764.3445615

[115] Arissa J. Sato, Zefan Sramek, and Koji Yatani. 2023. Groupnamics: Designing an Interface for Overviewing and Managing Parallel Group Discussions in an Online Classroom. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 701, 18 pages. https://doi.org/10.1145/3544548.3581322

[116] Gianluca Schiavo, Alessandro Cappelletti, Eleonora Mencarini, Oliviero Stock, and Massimo Zancanaro. 2014. Overt or subtle? Supporting group conversations with automatically targeted directives. In *Proceedings of the 19th international conference on Intelligent User Interfaces*. 225–234.

[117] Joseph Seering, Michal Luria, Geoff Kaufman, and Jessica Hammer. 2019. Beyond Dyadic Interactions: Considering Chatbots as Community Members. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300680

[118] Nathan Semertzidis, Michaela Vranic-Peters, Josh Andres, Brahmi Dwivedi, Yutika Chandrashekhar Kulwe, Fabio Zambetta, and Florian Floyd Mueller. 2020. Neo-Noumena: Augmenting Emotion Communication. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376599

[119] N. Sadat Shami, Jiang Yang, Laura Panc, Casey Dugan, Tristan Ratchford, Jamie C. Rasmussen, Yannick M. Assogba, Tal Steier, Todd Soule, Stela Lupushor, Werner Geyer, Ido Guy, and Jonathan Ferrar. 2014. Understanding employee social media chatter with enterprise social pulse. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 379–392. https://doi.org/10.1145/2531602.2531650

[120] Chunqi Shi, Donghui Lin, and Toru Ishida. 2013. Agent metaphor for machine translation mediated communication. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. 67–74.

[121] Ryoichi Shibata, Shoya Matsumori, Yosuke Fukuchi, Tomoyuki Maekawa, Mitsuhiko Kimoto, and Michita Imai. 2022. Utilizing core-query for context-sensitive Ad generation based on dialogue. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*. 734–745.

[122] Donghoon Shin, Soomin Kim, Ruoxi Shang, Joonhwan Lee, and Gary Hsieh. 2023. IntroBot: Exploring the Use of Chatbot-assisted Familiarization in Online Collaborative Groups. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for

Computing Machinery, New York, NY, USA, Article 613, 13 pages. https://doi.org/10.1145/3544548.3580930

[123] Joongi Shin, Michael A. Hedderich, AndréS Lucero, and Antti Oulasvirta. 2022. Chatbots Facilitating Consensus-Building in Asynchronous Co-Design. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 78, 13 pages. https://doi.org/10.1145/3526113.3545671

[124] Kellie Yu Hui Sim, Kohleen Tijing Fortuno, and Kenny Tsu Wei Choo. 2024. Towards Understanding Emotions for Engaged Mental Health Conversations. In *Companion Publication of the 2024 ACM Designing Interactive Systems Conference* (IT University of Copenhagen, Denmark) *(DIS '24 Companion)*. Association for Computing Machinery, New York, NY, USA, 176–180. https://doi.org/10.1145/3656156.3663694

[125] Fabian Sperrle, Astrik Veronika Jeitler, Jürgen Bernard, Daniel A Keim, and Mennatallah El-Assady. 2020. Learning and teaching in co-adaptive guidance for mixed-initiative visual analytics. In *Proceedings of the EuroVis Workshop on Visual Analytics (EuroVA)*. 61–65.

[126] Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. 2015. Automated Social Skills Trainer. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta, Georgia, USA) *(IUI '15)*. Association for Computing Machinery, New York, NY, USA, 17–27. https://doi.org/10.1145/2678025.2701368

[127] Edward Tse, Saul Greenberg, Chia Shen, Clifton Forlines, and Ryo Kodama. 2008. Exploring true multi-user multimodal interaction over a digital table. In *Proceedings of the 7th ACM Conference on Designing Interactive Systems* (Cape Town, South Africa) *(DIS '08)*. Association for Computing Machinery, New York, NY, USA, 109–118. https://doi.org/10.1145/1394445.1394457

[128] Stephanie Valencia, Jessica Huynh, Emma Y Jiang, Yufei Wu, Teresa Wan, Zixuan Zheng, Henny Admoni, Jeffrey P Bigham, and Amy Pavel. 2024. COMPA: Using Conversation Context to Achieve Common Ground in AAC. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 915, 18 pages. https://doi.org/10.1145/3613904.3642762

[129] Hao-Chuan Wang, Dan Cosley, and Susan R. Fussell. 2010. Idea expander: Supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA) *(CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 103–106. https://doi.org/10.1145/1718918.1718938

[130] Yiwen Wang, Ziming Li, Pratheep Kumar Chelladurai, Wendy Dannels, Tae Oh, and Roshan L Peiris. 2023. Haptic-Captioning: Using Audio-Haptic Interfaces to Enhance Speaker Indication in Real-Time Captions for Deaf and Hard-of-Hearing Viewers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 781, 14 pages. https://doi.org/10.1145/3544548.3581076

[131] Zhuxiaona Wei and James A. Landay. 2018. Evaluating Speech-Based Smart Devices Using New Usability Heuristics. *IEEE Pervasive Computing* 17, 2 (apr 2018), 84–96. https://doi.org/10.1109/MPRV.2018.022511249

[132] Rainer Winkler, Sebastian Hobert, Antti Salovaara, Matthias Söllner, and Jan Marco Leimeister. 2020. Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376781

[133] Min Wu, Arin Bhowmick, and Joseph Goldberg. 2012. Adding structured data in unstructured web chat conversation. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (Cambridge, Massachusetts, USA) *(UIST '12)*. Association for Computing Machinery, New York, NY, USA, 75–82. https://doi.org/10.1145/2380116.2380128

[134] Meng-Hsin Wu, Su-Fang Yeh, XiJing Chang, and Yung-Ju Chang. 2021. Exploring Users' Preferences for Chatbot's Guidance Type and Timing. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 191–194. https://doi.org/10.1145/3462204.3481756

[135] Haijun Xia, Tony Wang, Aditya Gunturu, Peiling Jiang, William Duan, and Xiaoshuo Yao. 2023. CrossTalk: Intelligent Substrates for Language-Oriented Interaction in Video-Based Communication and Collaboration. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[136] Kuldeep Yadav, Animesh Seemendra, Abhishek Singhania, Sagar Bora, Pratyaksh Dubey, and Varun Aggarwal. 2023. Interviewing the Interviewer: AI-generated Insights to Help Conduct Candidate-centric Interviews. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 723–736. https://doi.org/10.1145/3581641.3584051

[137] Naomi Yamashita, Katsuhiko Kaji, Hideaki Kuzuoka, and Keiji Hirata. 2011. Improving visibility of remote gestures in distributed tabletop collaboration. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (Hangzhou, China) *(CSCW '11)*. Association for Computing Machinery, New York, NY, USA, 95–104. https://doi.org/10.1145/1958824.1958839

[138] Xi Yang, Marco Aurisicchio, and Weston Baxter. 2019. Understanding Affective Experiences with Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300772

[139] Su-Fang Yeh, Meng-Hsin Wu, Tze-Yu Chen, Yen-Chun Lin, XiJing Chang, You-Hsuan Chiang, and Yung-Ju Chang. 2022. How to Guide Task-oriented Chatbot Users, and When: A Mixed-methods Study of Combinations of Chatbot Guidance Types and Timings. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 488, 16 pages. https://doi.org/10.1145/3491102.3501941

[140] Ryan Yen, Li Feng, Brinda Mehra, Ching Christie Pang, Siying Hu, and Zhicong Lu. 2023. StoryChat: Designing a Narrative-Based Viewer Participation Tool for Live Streaming Chatrooms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 795, 18 pages. https://doi.org/10.1145/3544548.3580912

[141] Dongwook Yoon, Nicholas Chen, Bernie Randles, Amy Cheatle, Corinna E. Löckenhoff, Steven J. Jackson, Abigail Sellen, and François Guimbretière. 2016. RichReview++: Deployment of a Collaborative Multi-modal Annotation System for Instructor Feedback and Peer Discussion. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 195–205. https://doi.org/10.1145/2818048.2819951

[142] Qingxiao Zheng, Yiliu Tang, Yiren Liu, Weizi Liu, and Yun Huang. 2022. UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 570, 24 pages. https://doi.org/10.1145/3491102.3501855

[143] Daniel Xiaodan Zhou, Nathan Oostendorp, Michael Hess, and Paul Resni k. 2008. Conversation pivots and double pivots. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1009–1012. https://doi.org/10.1145/1357054.1357209

[144] Tamara Zubatiy, Kayci L Vickers, Niharika Mathur, and Elizabeth D Mynatt. 2021. Empowering Dyads of Older Adults With Mild Cognitive Impairment And Their Care Partners Using Conversational Agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 657, 15 pages. https://doi.org/10.1145/3411764.3445124

A Systematic Literature Review

A.1 Keyword Search Query

When searching the ACM Digital Library, the following query was used. We note our usage of their search syntax containing boolean searches, exact match searches (" "), and wildcard searches (*).

Title:("*speech*") **OR** Keyword:("*speech*")
OR Title:(*speak**) **OR** Keyword:(*speak**)
OR Title:("*spoken*") **OR** Keyword:("*spoken*")
OR Title:(*dialogue*) **OR** Keyword:(*dialogue*)
OR Title:(*chat**) **OR** Keyword:(*chat**)
OR Title:(*conversation**) **OR** Keyword:(*conversation**)
OR Title:(*communicat**) **OR** Keyword:(*communicat**)
OR Title:("*discourse*") **OR** Keyword:("*discourse*")
OR Title:(*discuss**) **OR** Keyword:(*discuss**)
OR Title:(*argument**) **OR** Keyword:(*argument**)
OR Title:(*scaffold**) **OR** Keyword:(*scaffold**)

Paper	Short Paper Summary	Reason for Exclusion
Echenique et al. [33]	The paper investigates how supplemental cues (video or real-time text transcripts) support non-native speakers' participation in multiparty audio conferences.	The system is not actually implemented, only a Wizard-of-Oz study is performed.
Kim et al. [71]	This paper proposes an audio-visualized caption system that automatically visualizes paralinguistic cues into various caption elements, such as thickness, height, font type, and motion.	Captioning of videos is generally not included, as there is no dialogue between the viewer, i.e., the augmentee and the subject of the video.
Rubin et al. [110]	This paper introduces an interface that assists novice users in recording scripted narrations.	Coaching for speech is not included as long as there is no active dialogue.
Semertzidis et al. [118]	The paper introduces a communicative neuroresponsive system that uses brain-computer interfacing and artificial intelligence to read one's emotional states.	The technique may be applicable to dialogue but is not explicitly used for it.
Yamashita et al. [137]	The paper proposes a technique called "remote lag" to alleviate the problems caused by the invisibility of remote gestures.	No text/speech processing, only visual cues are considered.

A.2 Study Scope

In this section, we provide a small subsample of papers in the corpus along with a brief summary and the reasons for their exclusion. We aim for this to help the reader better grasp the scope of the study. Purposefully, we oversample controversial studies that were only marginally excluded through author discussions.

A.3 Corpus

The 78 papers that were considered in the final version of the design space are: [4, 7–10, 12, 15, 17–20, 24, 25, 36–38, 40, 41, 43, 45–47, 51–58, 62, 65–67, 70, 72, 74, 75, 82, 84, 86, 87, 91–94, 96–101, 103–106, 109, 111, 114–116, 119–122, 124, 126, 128, 129, 132, 133, 135, 136, 139–141, 143, 144]