Envisioning Futures: How the Modality of AI Recommendations Impacts Conversation Flow in AR-enhanced Dialogue

Steeven Villa
LMU Munich
Germany
villa@posthci.com

Yannick Weiss
LMU Munich
Munich, Germany
yannick.weiss@ifi.lmu.de

Karin Lu
LMU Munich
Germany
m.lu@campus.lmu.de

Moritz Ziarko
LMU Munich
Germany
moritz@ziarko.de

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@lmu.de

Jasmin Niess
University of Oslo
Oslo, Norway
jasminni@uio.no

ABSTRACT

The use of AI is becoming more common among the population every day; the use of generative AI, such as LLMs, empowers individuals by supporting daily life tasks. Yet, the user interaction with AI models is mostly constrained to chatbot interactions. However, we envision that in the near future, individuals will be able to integrate the use of these technologies into their daily activities without refocusing their attention. Consequently, we explores the impact of such integration on individuals' conversations.

In detail, this paper investigates how different modes of information presentation (visual vs. auditory) and triggers for AI action (mechanical vs. ocular) influence conversational dynamics and user experiences. We conducted a mixed-method, within-subjects study with 21 participants using a Discourse Completion Task (DCT) to observe how users develop their discourse in the presence of AI-generated suggestions. Our study examines the effects of presentation modality on response delay, response length, and response similarity to the AI prompt. The results highlight the significance of managing the balance between human and AI input in conversation, revealing insights into user experience factors with AI assistance in face-to-face conversational settings.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Interaction paradigms**; **User studies**.

KEYWORDS

Augmented Reality, Artificial Intelligence, Human-AI collaboration, AR Dialogue

ACM Reference Format:
Steeven Villa, Yannick Weiss, Karin Lu, Moritz Ziarko, Albrecht Schmidt, and Jasmin Niess. 2024. Envisioning Futures: How the Modality of AI Recommendations Impacts Conversation Flow in AR-enhanced Dialogue. In *ICMI '24: ACM International Conference on Multimodal Interaction, November 4–5, San Jose, Costa Rica.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3678957.3685731

ICMI '24, November 4–5, San Jose, Costa Rica

ACM ISBN 979-8-4007-0462-8/24/11.
https://doi.org/10.1145/3678957.3685731

1 INTRODUCTION

Traditionally, people would request AI support using chat-based interfaces or talk explicitly to the AI agent, requiring users to redirect their attention toward the agent when requiring assistance [7]. In the near future, these interactions will be implicitly integrated into people's daily lives [29], allowing them to maintain their focus on a primary task, such as conversations [41] while exchanging information with such AI models.

Yet, conversational contexts are unique in the sense that they demand real-time processing of social cues, nonverbal communication, and rapid information exchange [1, 55], making attention redirection to the AI agent especially detrimental to the quality of the interaction [35, 36]. This raises interesting questions on how to transition AI assistance from chatbot-based interactions to an implicit context within human-to-human conversations, especially in terms of display modalities and how to request AI support, as well as the potential impacts on conversational flow and user experience.

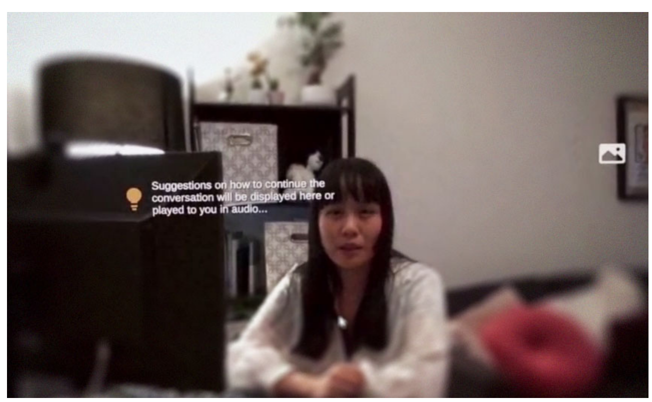


Figure 1: Interface for visual modality retrieval: AI suggestions are rendered the near-peripheral vision of participants, the UI element on the right is a placeholder for pictures in future implementations.

Mode and bandwidth of presentation have been explored in the context of conversations before, showing that smaller batches of information are perceived as more positive and that it is indeed possible to provide users with secondary information without the conversational partner noticing it [31]. It has also been shown that interlocutors can receive information in near-peripheral vision optimally [22] and that people generally prefer visual notifications over auditory ones in collaborative AR [11, 42].

Yet, in the case of AI suggestions, the information has a higher contextual coherence than notifications or a predefined set of words, [11, 22, 31]. Hence, it is more likely that they can indeed influence how the interlocutor thinks about a situation given the plausibility of the suggestion. Considering this, modality can play an additional role in how individuals further integrate these suggestions in their discourse and how they feel in control of the conversation [9]. As we continue to utilize the potential of AI in augmented reality environments, a critical challenge emerges in optimizing the interplay between the ease of information assimilation and the obtrusiveness of AI-generated suggestions. Consequently, there remains a pressing need to explore and build an understanding of how these elements can be balanced to ensure they enhance information processing without detracting from primary tasks such as a good conversation flow.

To shed light on these questions, we explore a face-to-face conversational scenario with AI support and how the presentation modality impacts the conversation. We provided participants with a system that continuously processes the conversation context and offers relevant support to continue the conversation upon user request. With this research, we aim to deepen our understanding of human-AI interaction in the context of face-to-face interactions.

More specifically, we conducted a within-subjects study involving 21 participants with four conditions that varied both the way participants would request support from the system (action triggering) and how the system would display the information back to the participants (retrieval modality) in a Discourse Completion Task (DCT) [2, 44], followed by an open conversation condition [16] to allow participants to use the system without any conversational constraint. Additionally, we conducted interviews to complement the quantitative results and derive insights into participants' experiences and the strategies used during the interaction.

This research contributes (1) An analysis of how AI assistance in conversations impacts conversational flow as measured by response similarity, length, and delay (2) an exploration of user perceptions and experiences of AI-supported face-to-face conversations, and (3) a system designed for this specific context.

2 RELATED WORK

This section reviews relevant research in three key areas: AR-enhanced reality, work on technology-assistet conversation, and research on understanding Human-AI collaboration.

2.1 AR-Enhanced Reality And Subtle Interfaces

The integration of digital technologies to support human cognition is a longstanding theme within HCI, dating back to the early days of computing [12, 13]. From assisting with complex calculations to facilitating note-taking on personal computers, researchers have explored various avenues of technological support [17, 43, 48, 49, 53]. This trend has witnessed a significant acceleration in recent years, driven by the emergence of AI systems capable of comprehending and responding to natural language [3, 38].

A classic example of augmented cognition is SenseCam, introduced by Hodges et al. [20] a system to assist individuals with amnesia, effectively externalizing the memory function and allowing information retrieval on demand [21, 28, 33]. Similarly, Perla et al. [34] developed an augmented reality (AR) system to store step-by-step instructions, allowing users to concentrate on the specific, detailed steps required.

Furthermore, Kimura et al. [25] introduced SilentSpeller, a device designed for silent speech text entry using EEG. Similarly, using silent interfaces, Kapur et al. [24] presented 'AlterEgo,' a system capable of receiving user input, querying a search engine, and retrieving information.

The growing interest in leveraging technology to enhance human cognition highlights the potential for seamless human-AI integration [9, 10]. However, recent discussions, as noted by Chignell et al. [9], reveal a gap in understanding the human perspective on this integration within HCI.

2.2 Technology-Assisted Conversation

While technology has revolutionized long-distance communication, its integration with in-person interactions also presents intriguing research opportunities. Exploring the intertwining of these domains, Zisk and Dalton [56] proposed "dual-purpose speech," where conversational context fuels technology interaction, so the system would capture the context of the speech not specifically addressed to the system input and adapt based on that additional information. Contextual information has further served as input for user support, with Kane and Morris [23] using object context to suggest relevant words for individuals with Amyotrophic Lateral Sclerosis (ALS) and Fontana de Vargas et al. [14] generating context-specific suggestions for users with Autistic Spectrum Disorder (ASD) based on pictures.

Another longstanding area of inquiry concerns the role of information delivery modalities in technology-assisted conversations. Studies by Ofek et al. [31] and Cai et al. [8] explored the effectiveness of visual and auditory modalities, uncovering user preferences for small, visual chunks presented during conversational pauses. While visual delivery is ofter preferred, individual preferences for auditory feedback remain variable [8, 31].

2.3 Understanding Human-AI Collaboration From The Human Standpoint

Collaboration between humans and AI has consistently demonstrated superior outcomes compared to individual AI or human actions [10, 15]. However, human-AI collaboration's psychological and social aspects remain relatively under-explored [9, 50]. This is particularly true in the context of recent AI advancements and the ever-expanding body of knowledge we have accumulated about these systems. This prompts us to question how human behavior, information integration, and the lived experience of combined human-AI intelligence are influenced and shaped.

Recent findings have revealed that the role of humans in human-AI collaboration is more intricate than simply the combined performance of both agents [10]. For instance, Kosch et al. [27] demonstrated that the mere presence of an AI system can enhance an individual's performance in a word puzzle task, a phenomenon they referred to as the 'placebo effect of AI.' Moreover, Villa et al. [47] explored this phenomenon's behavioral and physiological impacts and discovered that individuals with high expectations of AI system performance tend to make riskier decisions.

These examples underscore the need to understand more in-depth the human-AI relationship from a human-centric perspective [26]. In order to elucidate how individuals experience human-AI collaboration, we explore how humans experience AI support in a face-to-face conversational context.

3 SYSTEM SPECIFICATION & APPARATUS

Here, we describe the key mechanisms of the system and its implementation in detail. We developed a system that supports users in conversational situations. The system continuously collects contextual information about the conversation via an omnidirectional microphone. When the user requires support, they can trigger to *offload* the current context of the conversation and start processing the information and *retrieving* the processed information. The system was tuned to suggest ways to continue the conversation based on the current state of the conversation.

3.1 Implementation

The hardware configuration of the system comprises a Video-see-through Augmented Reality (AR) device (Meta Quest Pro by Meta, Menlo Park, The USA) for visual rendering and inputs, non-occluding earpods (Sony Linkbuds, Sony, Tokyo, Japan) for auditory rendering, and an omnidirectional microphone (Senheiser SP20, Senheiser, Wedemark, Germany) for capturing audio input from all individuals present in the room. The system incorporates the Cognitive Services API by Microsoft (Redmond, The USA) for tasks related to text-to-speech and speech-to-text processing. Furthermore, it leverages the capabilities of GPT-3.5 by OpenAI (San Francisco, USA) to process the user's offloaded tasks. We describe all these components in detail below.

3.1.1 Action Triggers. The system incorporates two methods: mechanical, and ocular. The Mechanical Trigger requires the user to press a button while in the Ocular Trigger, the user has to look up right (information processing ocular movement [39]). The first method uses a button on the VR controller. The second uses the integrated eye-tracking feature of the headset.

3.1.2 Retrieval Modalities. The retrieved information was displayed visually and auditorily. In line with previous work, [8] we hypothesized that the visual retrieval offers high information density and precision but can split attention from the conversation. In contrast, auditory retrieval potentially maintains engagement but has a limited bandwidth. The former was implemented with a floating text in the near-peripheral vision of the user [22]. While the later used text-to-speech and rendered the output through headphones. These headphones allow audio rendering without blocking out the surrounding environment.

3.1.3 Conversation Context Processing: We implemented the information offloading process by continuously having the system record contextual information from ongoing conversations through active listening and encoding this into text using speech-to-text. Only when the user decides to offload this contextual information to the AI using an action trigger described in subsubsection 3.1.1, it is actually processed. This approach is designed to mitigate potential bandwidth constraints. For example, in contrast with [24], the user does not have to communicate the full prompt to the system, as the system has preemptively recorded the context.

3.1.4 Task Solving: In the present system, the task-solving stage is executed utilizing the OpenAI API. Subsequently, the accumulated contextual information is used as a prompt, accompanied by customized instructions, to generate alternatives for continuing the conversation. The output information is then filtered to avoid wrong-formatted responses and finally retrieved by the user.

3.1.5 Result Retrieval: Once the offloaded information has been processed, the outcomes of this operation are automatically returned to the user one of the two retrieval modalities.

4 USER STUDY

We conducted a within-subjects laboratory experiment to investigate the user experience and conversation dynamics when receiving AI support. We manipulated two factors: Action Trigger with the levels: Mechanical and Ocular, and Retrieval Modality with the levels Visual and Auditory. The order of the conditions was counterbalanced using Latin Square, and the order of the scenarios in the DCT was randomized. Building on prior research, we adopted an exploratory approach, as directly comparing to conventional baselines can introduce multiple confounding factors. Our primary aim was to deepen our understanding of how the modality of AI recommendations impacts conversation flow in AR-enhanced dialogue, thereby providing valuable insights for future work.

4.1 Discourse Completion Task (DCT)

The Discourse Completion Task (DCT) originates from pragmatics [44]. This method has been employed for research and evaluation purposes [2]. It involves using scenarios to prompt individuals to respond in writing or speech, enabling the collection of diverse and comparable cross-linguistic data [16]. We employed DCT with validated scenarios extracted from previous work that have been shown to be appropriate for discoursive analysis [30, 37, 40].

The experimenter verbally presented a randomly selected scenario to the participant, concluding with a question. The participant is then required to continue the conversation, embodying the character in the scenario. Participants had to decide whether to trigger the AI support or not. Each participant received a total of 12 DCT scenarios in a randomized way (see supplementary material).

4.2 Data Collection

This study aimed to investigate the influence of AI support on face-to-face conversation dynamics, user perceptions, and overall experience. We employed a mixed-method approach, including four controlled conditions and an open conversation scenario.

The initial four conditions served the purpose of familiarizing participants with the system configurations, eliciting initial insights on potential conversational integration during a normal setting, and capturing conversational behaviors under AI assistance. This prepared the participants for the open conversation phase, where they freely utilized the system in a natural dialogue. Additionally, we collected quantitative data on pre-existing perceptions of performance-enhancing technologies and per-block questionnaires to capture participants' evolving perspectives throughout the study. To complement this data, we conducted semi-structured interviews with the participants.

4.2.1 Conversation Dynamics Data. Using the DCT, we recorded the prompts suggested by the system and the responses provided by the participants. Afterward, we transcribed and cleaned all the responses for further processing. From these responses, we calculated response similarity, response length, and response delay.

Response Delay was defined as the time between the experimenter finishing their prompt and the participant starting their verbal response. This measure captured two key processes: (1) the system's technical processing time (recording, transcription, analysis, display) and (2) the participant's cognitive processing time (understanding, formulating, and initiating response). *Response Length* was operationalized as the total number of characters included in the participant response after cleaning transcription artifacts. *Response Similarity* we converted both participant responses and system prompts into numerical representations (embeddings). These capture the semantic meaning of each sentence. We then calculated the cosine similarity between these embeddings, resulting in a score between 0 (no similarity) and 1 (perfect similarity). This allowed us to quantify how closely aligned the participant's response was to the intended meaning of the system prompt.

4.2.2 Questionnaires: Task Load: We administered the NASA-TLX task load questionnaire) [19] to assess potential variations in task load resulting from the system configurations. The NASA-TLX evaluates task load across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. The NASA-TLX is a well-established and validated instrument for measuring task load across different situations [18]. *The Sense of Agency Scale (SoA)* [45] assesses an individual's overall beliefs regarding their sense of agency, which is the feeling of being in control and the initiator of their actions. The scale comprises two interrelated factors: Sense of Positive Agency (SoPA) and Sense of Negative Agency (SoNA). *Society's Attitudes Towards Human Augmentation and Performance Enhancement Technologies (SHAPE) Scale* is a standardized tool to measure attitudes towards performance-enhancing technologies [51]. It consists of two factors: Social Threat and Agency, which measure the perceived societal threat of an augmentation device, and the user's sense of ownership over their actions when using such technology. For evaluating system usability, we employed the *System Usability Scale (SUS)* developed by Brooke [5]. This questionnaire is a well-established and validated tool for assessing the subjective usability of technological systems [6].

4.2.3 Semi-Structured Interviews. All audio recordings from the interviews were transcribed verbatim and imported into the Atlas.ti analysis software. We applied open coding combined with pragmatic thematic analysis, as described by Blandford et al. [4]. As a first step, we familiarized ourselves with the data. Data familiarization involved multiple readings of the material to gain a comprehensive understanding. Then, two researchers coded a representative sample of 25% of the material using open coding in line with Blandford et al. [4]. Next, an initial coding tree was established through iterative discussion. The remaining transcripts were split between the two researchers and coded individually. A final discussion session was conducted to structure the coding tree after the material was coded. This was followed by a final discussion session to construct and refine themes based on our material [4].

4.3 Participants

We recruited participants through the university's mailing lists and our extended networks. We recruited a total of N = 21 participants, from which 8 identified as female and 13 as male. The average age of our participants was twenty-six years (M = 26.85, SD = 4.57). Participants were compensated 6 euros/30 min for participating in the study. The study was approved by an ethics committee.

4.4 Procedure

Informed consent was obtained from participants after providing detailed study information. A demographic questionnaire and the SHAPE scale were then administered. Following this, participants engaged with the prototype, exploring all system configurations. The initial four conditions systematically combined the two Action Triggers and two Retrieval Modalities. Within each block, participants provided feedback through the System Usability Scale (SUS), the Sense of Agency Scale (SoA), and the NASA Task-Load Index (NASA-TLX) after each configuration. The fifth condition involved an open conversation scenario where participants could freely choose their preferred system configuration while collaboratively planning a trip. This condition was always presented last to maximize the ecological validity of the results. Counterbalancing of the first four conditions was ensured using a Latin Square design.

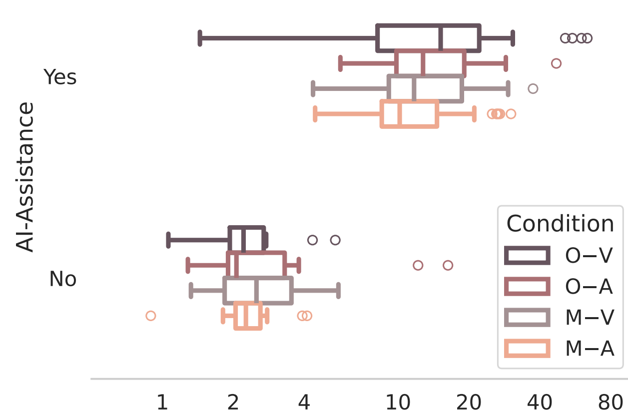


Figure 2: Response Delay (In seconds): All AI assistance conditions presented slower response times given the system processing requirements; however, within the AI assistance context, condition O-V presented significantly higher delays.

Table 1: Participant Demographics

Participant	Age	Gender	Education	Ocupation	Participant	Age	Gender	Education	Ocupation
P1	23	Female	Some Secondary Education	Student	P12	23	Male	Some University but no degree	Student
P2	23	Male	Some University but no degree	Employed full-time	P13	29	Male	Graduate or professional degree	Employed full-time
P3	35	Male	Graduate or professional degree	Student	P14	35	Male	Graduate or professional degree	Student
P4	23	Female	University Bachelors degree	Student	P15	28	Male	University Bachelors degree	Student
P5	29	Male	University Bachelors degree	Employed part-time	P16	27	Male	University Bachelors degree	Student
P6	26	Male	University Bachelors degree	Student	P17	30	Male	Graduate or professional degree	Employed full-time
P7	22	Male	University Bachelors degree	Student	P18	36	Male	Graduate or professional degree	Employed full-time
P8	21	Female	Some University but no degree	Employed part-time	P19	32	Female	Graduate or professional degree	Employed full-time
P9	26	Female	University Bachelors degree	Student	P20	25	Male	Completed Primary	Student
P10	23	Female	University Bachelors degree	Student	P21	23	Female	Completed Secondary	Employed part-time
P11	25	Female	University Bachelors degree	Student					

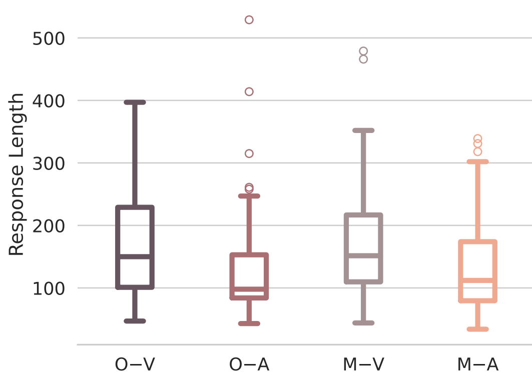


Figure 3: Conversation Flow plots for Response Length (In number of characters): the conditions with visual retrieval evidenced significantly longer responses

Subsequently, participants completed a second administration of the SHAPE scale and participated in a semi-structured interview, providing valuable post-experiment data.

5 RESULTS

In this section, we present the results of our user study. First, we share insights drawn from the Discourse Completion Task and questionnaires. Then, we detail the main themes that we identified based on the semi-structured interviews. For the sake of brevity, we abbreviate the combination of Action Triggers and Retrieval Modalities as M-A for mechanical trigger and auditory retrieval, M-V for mechanical trigger and visual retrieval, O-A For ocular trigger and auditory retrieval, and O-M for ocular trigger and visual retrieval.

5.1 Discourse Data

The data analysis for the DCT revealed non-normal distributions. To address this and ensure the robustness of our statistical tests, we employed aligned rank transformations (ART) on the data. Subsequently, we conducted two-way ANOVAs or Linear Mixed Effects Models (LME) where appropriate, followed by post-hoc tests for significant effects.

5.1.1 Response Delay. For response delay, the ANOVA conducted on aligned rank transformed data, utilizing Wald F tests with Kenward-Roger degrees of freedom, revealed a significant effect of Condition on Response Delay ($F(3, 160.83) = 3.25$, $p = 0.023$). Post-hoc comparisons, adjusted with the Bonferroni method for multiple comparisons, showed specific Condition differences. Notably, the O-V vs. M-A comparison revealed a significant effect on Response Delay ($est. = 24.97$, $SE = 8.75$, $df = 162$, $t.ratio = 2.85$, $p = 0.029$ adjusted). However, no other pairwise comparisons (O-V vs. O-A, O-V vs. M-V, O-A vs. M-V, O-A vs. M-A, M-V vs. M-A) reached significance after Bonferroni correction.

5.1.2 Response Length. In the case of response length, the ANOVA revealed a significant scenario interaction effect on response length ($F(3, 52) = 7.63$, $p < 0.001$, $\eta^2 = 0.30$). Post-hoc comparisons, adjusted for multiple comparisons with Bonferroni correction, identified significant differences in response lengths between scenarios: notably, O-V vs. M-V ($est = 32.951$, $p = 0.008$), O-A vs. M-V ($est. = -29.31$, $p = 0.021$), and M-V vs. M-A ($est. = 32.55$, p = 0.005).

5.1.3 Response Similarity. To understand the impact of Condition on response similarity, we used LMEs. The LME model showed an effect of Condition on Response Similarity ($F(3, 163) = 6.1$, $p < .001$). Post-hoc analyses with Bonferroni correction pinpointed specific differences: notably, scenarios M-A and O-V differed significantly ($est. = -29.09$, $p = .023$), as did M-V and O-A ($est. = 29.28$, $p = .017$) and O-A and O-V ($est. = -34.45$, $p = .004$).

5.2 Questionnaire Data

For the questionnaire data, we excluded 2 participants due to incomplete data. In all the questionnaires, we encountered non-normally distributed data. To address this non-normality and ensure the robustness of our statistical analyses, we applied aligned rank transformations (ART) to the data. Subsequently, we conducted two-way

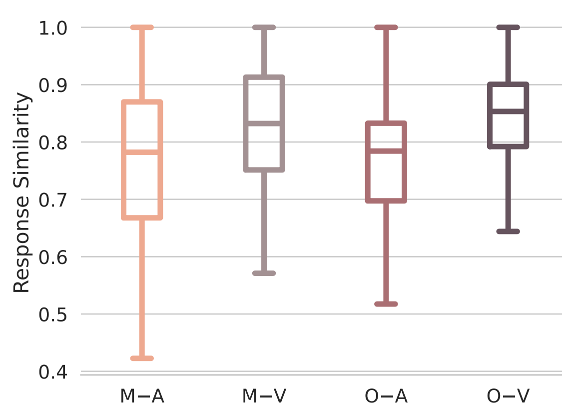


Figure 4: Conversation Flow plots for Response Similarity: Participants tended to follow more closely the system prompts in their responses when these prompts were presented Visually rather than Auditorily.

ANOVA tests when applicable, followed by post-hoc tests when significant effects were detected.

Statistical analyses revealed no significant differences in **Task Load** among the four conditions ($F(3, 54) = 0.68, p = 0.56$) or in the **Sense of Agency scale** ($F(3, 54) = 0.55, p = 0.64$). However, a significant effect was found for the **System Usability Scale** ($F(3, 54) = 4.1606, p = 0.01$). Post-hoc tests with Bonferroni correction indicated that mecano-visual differed significantly from O-A ($est. = 12.84, SE = 4.51, df = 54, t - ratio = 2.847, p = 0.03$) and O-V ($est. = 13.73, SE = 4.51, df = 54, t - ratio = 3.04, p = 0.02$) conditions (see Figure 5). No other pairwise differences reached statistical significance ($p > 0.05$).

In the pre-and post-interaction measurements of the SHAPE scale, we did not find statistically significant differences. However, regarding the Social Threat factor, 66% (N=14) of participants reported reduced or similar perceptions of augmentation technologies as a social threat. Meanwhile, 57% (N=11) indicated a decrease in the perception of individuals using augmentation technologies as being in control of their actions, whereas 21% (N=4) reported no change in this subscale (see Figure 6).

5.3 Interview Findings

Based on our qualitative inquiry, we identified three themes: (a) information moderation, integration, and balance; (b) action trigger; and (c) retrieval modality.

On a general level, our findings showcase how such a technology can support or impact the flow of a conversation, but also how individuals can adapt the use of the technology to the situation and which critical points an ideal system should address, such as correct timing and transparency to avoid disturbing the conversation. We identified mixed opinions about the interaction. While some participants felt like 'tools of the system,' others viewed the system as merely a tool they control, highlighting the variability in user experiences with cognitive augmentation. This indicates that perceived agency can vary significantly among users of a similar system. Our findings showed that the system appears to be helpful in challenging social contexts. There were two cases where the participants reported having social anxiety, the system being of extreme help given their condition, and the difficulty of finding the right words in a conversation, allowing for a continuous discussion. This illustrates that the system could potentially have

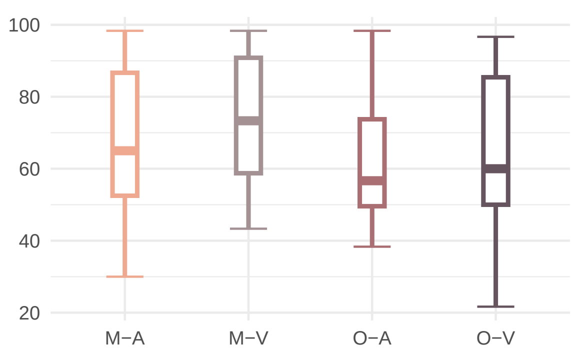
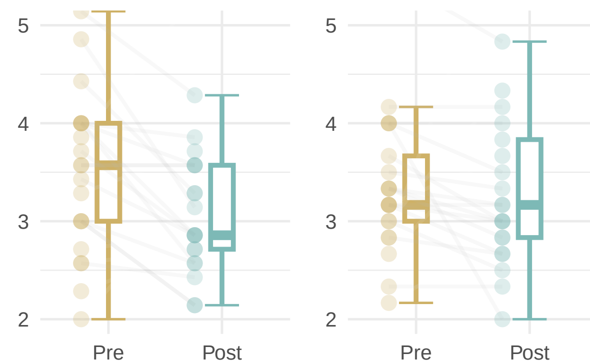


Figure 5: System Usability Score: the O-A and O-V presented a lower usability score.



(a) SHAPE: Social Threat **(b) SHAPE: Agency**

Figure 6: SHAPE Scale: 66% of the participants changed their opinions about performance-enhancing technologies after interaction with the system

assistive properties in certain scenarios. For improved readability, we slightly altered some statements (e.g. grammatical corrections), ensuring words and sentiment were maintained.

5.3.1 *Information Moderation, Integration, and Balance.*

This theme refers to the amount of information that the system provides, the amount of information that the users integrate into their answers, and the amount of information that each agent (i.e. human and AI) contributes to the outcome, this theme has three levels:

- **Information Moderation:** Delivering an appropriate amount of information from the system to the user, avoiding both insufficient and excessive information that could lead to suboptimal user experiences.
- **Information Integration:** Describes the process of how individuals incorporate the system's prompts into their own responses.
- **Information Balance:** Balance between the amount of information and the amount of information missing (knowledge of the person/ the amount of info that the system delivers).

We observed that participants had mixed opinions regarding the amount of information that should be delivered. On the one hand, some participants highlighted the importance of having a high amount of information from which they could select and later integrate into their discourse, often correlated to the preference for the visual presentation of the information. Other participants mentioned that they would prefer a low amount of information, so there would be a lower cognitive workload on the integration process, often correlated with the preference for auditory feedback or under-use of the visual modality. The latter also use the information mostly as a trigger for themselves to elaborate on their responses.

For information integration, the participants used diverse strategies to merge the system outputs with their own thoughts. some read the prompts, some paraphrased them, and others just used specific keywords and integrated them into their responses. Some participants reported integrating information when they found it to be more natural to them or when it was relevant, on the other hand, information that was not in line with the participant's mental model was rejected.

> *So when I read some words I would never use, I would just trash it, would say, no, that's not something I would use. And if something's interesting, I would kind of think about, how I would formulate that sentence (...) and if I would use it exactly like it's suggested, I can maybe use from one word to the full sentence. Maybe I would adjust some words, depending on whether I personally would use the whole sentence.* - P15

In line with that, participants would feel in control as long as the information retrieved is coherent and plausible, but once the information breaks with these expectations, their perceived agency would be reduced.

The last level of this theme is information balance. Our findings indicate that the level of knowledge that people possess themselves determines (to some degree) if and how they offload the information. If they possessed a moderate amount of knowledge, they often integrated the additional information provided by the AI.

Other participants reported that, in the case of extensive personal knowledge, (where they had more information than the system) they would not use the system as this did not feel natural.

> *I wouldn't say I would use it for more personal conversations, it's not natural enough. It's not natural enough to say this to your friends. You wouldn't be the kind of person to say something like this blindly.* - P5

This showcases some of the strategies participants applied to integrate the system suggestions into their conversations, highlighting the relevance of the coherence of the suggestions and the contextual awareness that the system should have in order for the participants to consider the provided information. Further, the findings illustrate the need for the system to adjust to the user in terms of the amount of information delivered to keep the cognitive workload low while providing them with alternatives to select from and integrate the information.

5.3.2 Action Triggers. This theme refers to the experiences that participants voiced regarding the action triggers. For instance, individuals generally prefer using the button as a trigger mechanism, but for social situations, a more discreet or subtle trigger method would be preferable. The levels of this theme are the assessed triggers:

- **Mechanical**: Referring to button press trigger.
- **Ocular**: Referring to eye-tracking-based trigger.

Although we identified mixed opinions in this regard, one of the most recurrent preferred combinations was the mechanical trigger (button) in combination with the visual retrieval modality. Often linked to the concept of autonomy. For example, the mechanical trigger was described as providing more power to the user.

> *I have a preference for pressing the button because I think I have more control over when I want to have a suggestion.* - P1

Our findings illustrate an interesting tension. Particularly, the mechanical trigger was sometimes perceived as more discreet and sometimes as more explicit. Furthermore, if the mechanical trigger was perceived as explicit, the advantage that the use of the system could be disclosed to the conversation partner was often discussed in connection with this.

> *So the button, I think, has the benefit of being very explicit. There you have to intentionally trigger the button. It also has the benefit of conveying to the other person that you're now using this system, which I would say is a good idea (...)* - P17

A highlight of the ocular triggering method was that participants were able to perform hand gestures since there was no need to hold an additional device to trigger the system.

This elucidates the context where different types of triggering methods would be applicable, for example, using explicit methods when the person wants to inform the interlocutor about the presence of the system and using a mechanical method in stages where the user reports low agency. In contrast, implicit methods might be used when the conversation flow is prioritized over the sense of agency.

5.3.3 Retrieval Modality. This theme refers to the comments regarding the retrieval modality. The preference for one or the other retrieval modality was highly dependent on personal preferences. The two levels of this theme were the retrieval modalitys:

- **Auditory**: Referring to the text-to-speech method.
- **Visual**: Referring to the AR, text-based method.

While some participants felt overwhelmed with two streams of audio information (one from the system and one from their human conversation partner), others pointed out that it would be too challenging for them to read the additional information and listen to the conversation at the same time.

> *So, the audio thing obviously has the problem that I'm talking to you, which makes me want to listen to you, and that thing also is talking. It's a bit complicated to have two persons [i.e. the conversation partner and the system] talking to you at the same time, so the information coming in on the same channel makes it harder to follow that.* - P18

The other aspect participants discussed in this context was the amount of information presented. For example, some participants preferred visual information but emphasized that they struggled with the amount of text to go through. Whereas others emphasized that they could just read the text to the other person.

While other participants highlighted that audio retrieval is more effective, given that they do not have to spend time on reading and that it allows them to multitask and helps them trigger conversations.

> *I would say the audio felt nicer to me, I would say so. Because we don't really need the entire sentences. So just to start the conversation, maybe a single sentence would be enough.* - P7

The presentation method also strongly influenced the participant's sense of agency over the process. A frequent comment was that during the auditory presentation, participants had less agency given that the retrieval modality was time-locking (during the information presentation, the participant had to focus on this information exclusively), while the visual presentation allowed for more control over the process given that they were able to decide when and which parts of the retrieved information to engage with.

6 DISCUSSION

We studied how the retrieval modality and action trigger impact user experience and conversational dynamics in AR-enhanced dialogue. For this, we developed a system capable of processing the contextual information of the conversation, processing the information to continue the conversation, and retrieving the information either visually or auditorily. Then, we conducted a user study using the discourse completion task (DCT), a well-known method used in pragmatics to study individuals' discourse. We then interviewed participants to gain insights into their perception of the interaction. We inferred discourse metrics from the participant responses to the DCT and conducted thematic analysis on the interview data. In this section, we synthesize the insights gained through this process and report them in two parts, one focusing on quantitative and one focusing on qualitative insights.

6.1 Quantitative Insights

Our study revealed that the introduction of AI-powered conversation support influences conversational flow. Notably, concerning *conversation pace*, Conversations may experience delays due to limitations in AI model processing, including limited buffer, processing power, and delays. These factors can slow down the flow of conversation. When the AI is delayed, it can cause problems for the conversation, especially if the information is relevant for continuing the conversation. The emergence of smaller, fine-tuned LLMs that run on low-power devices can potentially benefit real-time AI conversation assistance systems [54]. This trend aligns with requirements for real-world deployments, such as wearability and low processing delay [52].

Additionally, we found that participants gave significantly longer responses when using visual prompts than when using auditory ones. Our qualitative findings suggest this is because they were able to formulate their responses while processing the visual prompts at the same time. In contrast, auditory retrieval potentially requires greater cognitive effort to map the information onto an internal representation, significantly impacting both response length and adherence to the original prompt. This is further supported by the observed lower similarity between participant responses and the original prompt when presented auditorily. However, this connection has to be carefully investigated, as these measures do not account for intentionally paraphrased sentences, as this was one strategy reported by some participants. Therefore, the modality can significantly influence user engagement and response characteristics in AI-assisted conversations. In the following sections of the discussion, we touch on the qualitative insights regarding these metrics.

6.2 Qualitative Insights

Retrieval Modality. Regarding the retrieval modality, both channels were seen as suitable to parallelize, yet auditory has the inherent disadvantage that information can overlap with the interlocutor speech. In this sense, it would be perhaps beneficial to adapt the intensity of the feedback to be subtle and not disrupt the interaction [24]. On the other hand, visual retrieval, although observed as a natural way to multi-task, was overwhelming for some participants due to the amount of information.

Autonomy, as Driving Factor of Human-AI interaction. We found that when participants were put into the context of sharing cognitive tasks with the AI system, the first topic that emerged was the autonomy of humans when exposed so closely to an external cognitive agent. We argue that this concern extends to AI systems in general, as shown in recent literature [32, 46]. Yet, this seems more crucial in the context where the AI is constantly in contact with the user while having their own thoughts and interacting in natural situations such as conversations. In line with previous work, we found that some participants felt controlled by a system, especially in situations where they blindly followed the recommendations of the system.

Effects of Action Triggers on Autonomy. Participants attributed more autonomy to the more explicit mechanical trigger, as they felt a connection between their motor action of pressing a button and the information processing action of the AI system. Yet, there was a split on how and when to use each method. Participants with an inherent high agency during the interaction were comfortable using the ocular method. At the same time, some reported a loss of agency because they triggered the system unintentionally, given their natural ocular reactions. Also, the explicit trigger was seen as a way to disclose the use of the system by participants who felt it was necessary; otherwise, they would be "cheating".

Human-AI Autonomy Fluctuation. Another phenomenon that we observed was that some participants started the conversation relying on the system at the beginning of the conversation, when they felt insecure about the topic, and then took over the conversation when they gained confidence. This suggests that autonomy fluctuates depending on the user's cognitive states and personality, among others, in addition to the system design.

Notably, some participants reported having less agency and having a break in the interaction whenever the recommendations were not relevant or implausible. This suggests that in the case of cognitive agency, it is necessary to maintain semantic coherence. Presumably, designing a system that is aware of the user's knowledge and that can deliver information in coherence to the user's mental model can help maintain a high sense of agency on the user. This is supported by a repetitive comment from the participants when describing the information integration in their own words: *"Is it something I would say?"*.

Information Balance. Being overwhelmed by the amount of information relates to the information balance between the user's own knowledge and the system's contribution. We found that some participants relied on the system to compensate for their own lack of knowledge on specific topics and preferred to integrate more information in such cases, yet, very often, they commented that when it comes to conversations about personal matters, they will not offload any information since this information is easily accessible for them.

6.3 Limitations and Future Work

While we do think that our work provides valuable insights, we recognize that our study is prone to certain limitations. Firstly, our study did not cover the topic of privacy. It is worth noting that participants also did not raise this issue during our inquiry. However, we recognize that privacy is a significant and ongoing concern in the realm of human-AI interaction that should be considered in future designs. Additionally, while our study featured two action triggers, the broad technological landscape of HCI offers a wide array of alternatives for this specific aspect. Nevertheless, we are confident that the action triggers we selected provided a sufficiently diverse experience to engage participants, thereby enriching our overall investigation.Furthermore, our study examined the two main modalities people commonly use for information retrieval. However, we acknowledge that there may be additional—and potentially more effective—sensory channels and encoding methods yet to be explored. Importantly, our current research did not specifically look into the experiences of individuals with auditory or visual impairments interacting with such a system.

7 CONCLUSION

In this study, we explored how human-AI integration affects conversational behaviors, designing an AI system that delivers relevant information in response to user triggers. Involving 21 participants, our research identified key themes such as information moderation, integration, balance, the role of triggering, and retrieval methods. These findings provide insights for future human-AI integration development, stressing the importance of human-centric systems that preserve user autonomy and adapt to individual needs and contexts. This work can provide a foundation for researchers looking to explore enhancements in human cognitive capabilities through technology.

REFERENCES

[1] Michael Argyle. 1976. Non-verbal Communication and Language. *Royal Institute of Philosophy Supplements* 10 (March 1976), 63–78. https://doi.org/10.1017/S0080443600011079

[2] K Billmyer. 2000. Investigating instrument-based pragmatic variability: effects of enhancing discourse completion tests. *Applied Linguistics* 21, 4 (Dec. 2000), 517–552. https://doi.org/10.1093/applin/21.4.517

[3] Marcel Binz and Eric Schulz. 2023. Turning large language models into cognitive models. http://arxiv.org/abs/2306.03917 arXiv:2306.03917 [cs].

[4] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Ensuring Quality in Qualitative Research. In *Qualitative HCI Research: Going Behind the Scenes*, Ann Blandford, Dominic Furniss, and Stephann Makri (Eds.). Springer International Publishing, Cham, 87–96. https://doi.org/10.1007/978-3-031-02217-3_8

[5] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval in Industry* 189 (11 1995).

[6] John Brooke. 2013. SUS: A Retrospective. *J. Usability Studies* 8, 2 (feb 2013), 29–40.

[7] Jeffrey A. Brooks, Vineet Tiruvadi, Alice Baird, Panagiotis Tzirakis, Haoqi Li, Chris Gagne, Moses Oh, and Alan Cowen. 2023. Emotion Expression Estimates to Measure and Improve Multimodal Social-Affective Interactions. In *International Cconference on Multimodal Interaction*. ACM, Paris France, 353–358. https://doi.org/10.1145/3610661.3616129

[8] Runze Cai, Nuwan Nanayakkarawasam Peru Kandage Janaka, Shengdong Zhao, and Minghui Sun. 2023. ParaGlassMenu: Towards Social-Friendly Subtle Interactions in Conversations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–21. https://doi.org/10.1145/3544548.3581065

[9] Mark Chignell, Lu Wang, Atefeh Zare, and Jamy Li. 2023. The Evolution of HCI and Human Factors: Integrating Human and Artificial Intelligence. *ACM Transactions on Computer-Human Interaction* 30, 2 (April 2023), 1–30. https://doi.org/10.1145/3557891

[10] Mu-Huan Chung, Mark Chignell, Lu Wang, Alexandra Jovicic, and Abhay Raman. 2020. Interactive Machine Learning for Data Exfiltration Detection: Active Learning with Human Expertise. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 280–287. https://doi.org/10.1109/SMC42975.2020.9282831 ISSN: 2577-1655.

[11] Marina Cidota, Stephan Lukosch, Dragos Datcu, and Heide Lukosch. 2016. Workspace Awareness in Collaborative AR using HMDs: A User Study Comparing Audio and Visual Notifications. In *Proceedings of the 7th Augmented Human International Conference 2016*. ACM, Geneva Switzerland, 1–8. https://doi.org/10.1145/2875194.2875204

[12] Richard F. Dillon. 1983. Human factors in user-computer interaction: An introduction. *Behavior Research Methods & Instrumentation* 15, 2 (March 1983), 195–199. https://doi.org/10.3758/BF03203548

[13] Geraldine Fitzpatrick. 2018. A Short History of Human Computer Interaction: A People-Centred Perspective. In *Proceedings of the 2018 ACM SIGUCCS Annual Conference (SIGUCCS '18)*. Association for Computing Machinery, New York, NY, USA, 3. https://doi.org/10.1145/3235715.3242569

[14] Mauricio Fontana de Vargas, Jiamin Dai, and Karyn Moffatt. 2022. AAC with Automated Vocabulary from Photographs: Insights from School and Speech-Language Therapy Settings. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3517428.3544805

[15] Andreas Fügener, Jörn Grahl, Alok Gupta, and Wolfgang Ketter. 2022. Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. *Information Systems Research* 33, 2 (2022), 678–696. https://doi.org/10.1287/isre.2021.1079 arXiv:https://doi.org/10.1287/isre.2021.1079

[16] A. Golato. 2003. Studying Compliment Responses: A Comparison of DCTs and Recordings of Naturally Occurring Talk. *Applied Linguistics* 24, 1 (March 2003), 90–121. https://doi.org/10.1093/applin/24.1.90

[17] Richard H. R. Harper. 2019. The Role of HCI in the Age of AI. *International Journal of Human–Computer Interaction* 35, 15 (Sept. 2019), 1331–1344. https://doi.org/10.1080/10447318.2019.1631527 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2019.1631527.

[18] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[19] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[20] S. Hodges, E. Berry, K. Wood, Ken Woodberry, and Steve Hodges. 2011. SenseCam: A wearable Camera that Stimulates and Rehabilitates Autobiographical Memory. *Memory* 19, 7 (Jan. 2011), 685–696. https://www.microsoft.com/en-us/research/publication/sensecam-a-wearable-camera-that-stimulates-and-rehabilitates-autobiographical-memory/

[21] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: A Retrospective Memory Aid. In *UbiComp 2006: Ubiquitous Computing (Lecture Notes in Computer Science)*, Paul Dourish and Adrian Friday (Eds.). Springer, Berlin, Heidelberg, 177–193. https://doi.org/10.1007/11853565_11

[22] Nuwan Janaka, Chloe Haigh, Hyeongcheol Kim, Shan Zhang, and Shengdong Zhao. 2022. Paracentral and near-peripheral visualizations: Towards attention-maintaining secondary information presentation on OHMDs during in-person social interactions. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–14. https://doi.org/10.1145/3491102.3502127

[23] 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*. ACM, Edinburgh United Kingdom, 129–134. https://doi.org/10.1145/3064663.3064762

[24] Arnav Kapur, Shreyas Kapur, and Pattie Maes. 2018. AlterEgo: A Personalized Wearable Silent Speech Interface. In *23rd International Conference on Intelligent User Interfaces*. ACM, Tokyo Japan, 43–53. https://doi.org/10.1145/3172944.3172977

[25] Naoki Kimura, Tan Gemicioglu, Jonathan Womack, Richard Li, Yuhui Zhao, Abdelkareem Bedri, Zixiong Su, Alex Olwal, Jun Rekimoto, and Thad Starner. 2022. SilentSpeller: Towards mobile, hands-free, silent speech text entry using electropalatography. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3491102.3502015

[26] Agnes Mercedes Kloft, Robin Welsch, Thomas Kosch, and Steeven Villa. 2024. "AI enhances our performance, I have no doubt this one will do the same": The Placebo effect is robust to negative descriptions of AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–24. https://doi.org/10.1145/3613904.3642633

[27] Thomas Kosch, Robin Welsch, Lewis Chuang, and Albrecht Schmidt. 2023. The Placebo Effect of Artificial Intelligence in Human–Computer Interaction. *ACM Transactions on Computer-Human Interaction* 29, 6 (Jan. 2023), 56:1–56:32. https://doi.org/10.1145/3529225

[28] Catherine Loveday and Martin A. Conway. 2011. Using SenseCam with an amnesic patient: Accessing inaccessible everyday memories. *Memory* 19, 7 (Oct. 2011), 697–704. https://doi.org/10.1080/09658211.2011.610803 Publisher: Routledge _eprint: https://doi.org/10.1080/09658211.2011.610803.

[29] Florian Floyd Mueller, Pedro Lopes, Paul Strohmeier, Wendy Ju, Caitlyn Seim, Martin Weigel, Suranga Nanayakkara, Marianna Obrist, Zhuying Li, Joseph Delfa, Jun Nishida, Elizabeth M. Gerber, Dag Svanaes, Jonathan Grudin, Stefan Greuter, Kai Kunze, Thomas Erickson, Steven Greenspan, Masahiko Inami, Joe Marshall, Harald Reiterer, Katrin Wolf, Jochen Meyer, Thecla Schiphorst, Dakuo Wang, and Pattie Maes. 2020. Next Steps for Human-Computer Integration. 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, 1–15. https://doi.org/10.1145/3313831.3376242

[30] Iman Nikpoo and Hamid Allami. 2024. Exploring (Im)Politeness Strategies in Indirect Reports in Persian across Genders. *Education Research International* 2024 (Jan. 2024), e4642380. https://doi.org/10.1155/2024/4642380 Publisher: Hindawi.

[31] Eyal Ofek, Shamsi T. Iqbal, and Karin Strauss. 2013. Reducing disruption from subtle information delivery during a conversation: mode and bandwidth investigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 3111–3120. https://doi.org/10.1145/2470654.2466425

[32] Marine Pagliari, Valérian Chambon, and Bruno Berberian. 2022. What is new with Artificial Intelligence? Human–agent interactions through the lens of social agency. *Frontiers in Psychology* 13 (2022). https://doi.org/10.3389/fpsyg.2022.954444

[33] Katalin Pauly-Takacs, Chris J. A. Moulin, and Edward J. Estlin. 2011. SenseCam as a rehabilitation tool in a child with anterograde amnesia. *Memory* 19, 7 (Oct. 2011), 705–712. https://doi.org/10.1080/09658211.2010.494046 Publisher: Routledge _eprint: https://doi.org/10.1080/09658211.2010.494046.

[34] Ramakrishna Perla, Gaurav Gupta, Ramya Hebbalaguppe, and Ehtesham Hassan. 2016. InspectAR: An Augmented Reality Inspection Framework for Industry. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 355–356. https://doi.org/10.1109/ISMAR-Adjunct.2016.0119

[35] Henning Pohl, Andreea Muresan, and Kasper Hornbæk. 2019. Charting Subtle Interaction in the HCI Literature. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–15. https://doi.org/10.1145/3290605.3300648

[36] Navin Raj Prabhu, Chirag Raman, and Hayley Hung. 2020. Defining and Quantifying Conversation Quality in Spontaneous Interactions. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*. ACM, Virtual Event Netherlands, 196–205. https://doi.org/10.1145/3395035.3425966

[37] Carsten Roever. 2006. Validation of a web-based test of ESL pragmalinguistics. *Language Testing - LANG TEST* 23 (April 2006), 229–256. https://doi.org/10.1191/0265532206lt329oa

[38] Steven I. Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D. Weisz. 2023. The Programmer's Assistant: Conversational Interaction with a Large Language Model for Software Development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)*. Association for Computing Machinery, New York, NY, USA, 491–514. https://doi.org/10.1145/3581641.3584037

[39] Carola Salvi and Edward M. Bowden. 2016. Looking for Creativity: Where Do We Look When We Look for New Ideas? *Frontiers in Psychology* 7 (2016). https://doi.org/10.3389/fpsyg.2016.00161

[40] Ghiat Sara. 2022. Integrating Dynamic Assessment in Oral Courses to Develop EFL Students' Pragmatic Competence The Case of Third Year Students at the Department of English Batna2 University. Doctorat thesis, (2022) Université de Batna 2. (July 2022). http://dspace.univ-batna2.dz/handle/123456789/187

[41] Albrecht Schmidt. 2017. Augmenting Human Intellect and Amplifying Perception and Cognition. *IEEE Pervasive Computing* 16, 1 (Jan. 2017), 6–10. https://doi.org/10.1109/MPRV.2017.8 Conference Name: IEEE Pervasive Computing.

[42] Gabrielle Almeida de Souza, Laura Amaya Torres, Vinicius Stein Dani, David Steeven Villa, Abel Ticona Larico, Anderson Maciel, and Luciana Nedel. 2018. Evaluation of Visual, Auditory and Vibro-Tactile Alerts in Supervised Interfaces. In *2018 20th Symposium on Virtual and Augmented Reality (SVR)*. 163–169. https://doi.org/10.1109/SVR.2018.00033

[43] M. Stefik, D. G. Bobrow, S. Lanning, D. Tatar, and G. Foster. 1986. WYSIWIS revised: early experiences with multi-user interfaces. In *Proceedings of the 1986 ACM conference on Computer-supported cooperative work (CSCW '86)*. Association for Computing Machinery, New York, NY, USA, 276–290. https://doi.org/10.1145/637069.637107

[44] Emma Sweeney and Zhu Hua. 2015. Discourse Completion Tasks. In *Research Methods in Intercultural Communication*. John Wiley & Sons, Ltd, 212–222. https://doi.org/10.1002/9781119166283.ch14 Section: 14 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119166283.ch14.

[45] Adam Tapal, Ela Oren, Reuven Dar, and Baruch Eitam. 2017. The Sense of Agency Scale: A Measure of Consciously Perceived Control over One's Mind, Body, and the Immediate Environment. *Frontiers in Psychology* 8 (2017). https://doi.org/10.3389/fpsyg.2017.01552

[46] Quentin Vantrepotte, Bruno Berberian, Marine Pagliari, and Valérian Chambon. 2022. Leveraging human agency to improve confidence and acceptability in human-machine interactions. *Cognition* 222 (May 2022), 105020. https://doi.org/10.1016/j.cognition.2022.105020

[47] Steeven Villa, Thomas Kosch, Felix Grelka, Albrecht Schmidt, and Robin Welsch. 2023. The placebo effect of human augmentation: Anticipating cognitive augmentation increases risk-taking behavior. *Computers in Human Behavior* 146 (Sept. 2023), 107787. https://doi.org/10.1016/j.chb.2023.107787

[48] Steeven Villa, Thuy Duong Nguyen, Benjamin Tag, Tonja-Katrin Machulla, Albrecht Schmidt, and Jasmin Niess. 2023. Towards a Haptic Taxonomy of Emotions: Exploring Vibrotactile Stimulation in the Dorsal Region. In *Proceedings of the 2023 ACM International Symposium on Wearable Computers (ISWC '23)*. Association for Computing Machinery, New York, NY, USA, 50–54. https://doi.org/10.1145/3594738.3611363

[49] Steeven Villa, Jasmin Niess, Bettina Eska, Albrecht Schmidt, and Tonja-Katrin Machulla. 2021. Assisting Motor Skill Transfer for Dance StudentsUsing Wearable Feedback. In *Proceedings of the 2021 ACM International Symposium on Wearable Computers (ISWC '21)*. Association for Computing Machinery, New York, NY, USA, 38–42. https://doi.org/10.1145/3460421.3478817

[50] Steeven Villa, Jasmin Niess, Takuro Nakao, Jonathan Lazar, Albrecht Schmidt, and Tonja-Katrin Machulla. 2023. Understanding Perception of Human Augmentation: A Mixed-Method Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3544548.3581485

[51] Steeven Villa, Jasmin Niess, Albrecht Schmidt, and Robin Welsch. 2023. Society's Attitudes Towards Human Augmentation and Performance Enhancement Technologies (SHAPE) Scale. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. (IMWUT)* 7, 3 (2023). https://doi.org/10.1145/3610915

[52] Yawen Wu and Jingtong Hu. 2022. Towards Independent On-device Artificial Intelligence. In *2022 IEEE COMPUTER SOCIETY ANNUAL SYMPOSIUM ON VLSI (ISVLSI 2022)*. IEEE, New York, 351–351. https://doi.org/10.1109/ISVLSI54635.2022.00077 Num Pages: 1 Series Title: IEEE Computer Society Annual Symposium on VLSI Proceedings Web of Science ID: WOS:000886230500064.

[53] Wei Xu, Marvin J. Dainoff, Liezhong Ge, and Zaifeng Gao. 2023. Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI. *International Journal of Human–Computer Interaction* 39, 3 (Feb. 2023), 494–518. https://doi.org/10.1080/10447318.2022.2041900 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2022.2041900.

[54] Le Ye, Zhixuan Wang, Ying Liu, Peiyu Chen, Heyi Li, Hao Zhang, Meng Wu, Wei He, Linxiao Shen, Yihan Zhang, Zhichao Tan, Yangyuan Wang, and Ru Huang. 2021. The Challenges and Emerging Technologies for Low-Power Artificial Intelligence IoT Systems. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I-REGULAR PAPERS* 68, 12 (2021), 4821–4834. https://doi.org/10.1109/TCSI.2021.3095622 Num Pages: 14 Place: Piscataway Publisher: Ieee-Inst Electrical Electronics Engineers Inc Web of Science ID: WOS:000724482800004.

[55] Zhou Yu. 2015. Attention and Engagement Aware Multimodal Conversational Systems. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, Seattle Washington USA, 593–597. https://doi.org/10.1145/2818346.2823309

[56] Alyssa Hillary Zisk and Elizabeth Dalton. 2019. Augmentative and Alternative Communication for Speaking Autistic Adults: Overview and Recommendations. *Autism in Adulthood* 1, 2 (June 2019), 93–100. https://doi.org/10.1089/aut.2018.0007

A DISCOURSE COMPLETION TASK SCENARIOS

R1 You are a university student. You need to get a book from the library to finish your assignment on time. The library is closed and there is only one person you know who has the book you need, one of your lecturers. On the way to his office you meet him in the hallway. What do you say to him?

R2 You need to run a few errands downtown which you think may take you an hour. You go to your manager's office at work with whom you get on well and ask him to cover for you. What do you say to him?

R3 You have been a secretary of a company for some time now. You go to the desk of a new trainee and ask him to answer the telephone while you leave for a few minutes to attend to another urgent matter. What do you say to him?

R4 You are driving your car with a friend. You both must get to X street. Your friend had a map with directions which he had given to you just before leaving the house. You are now lost and do not remember where the map is. You suddenly see a pedestrian at the end of the road and suggest that your friend ask for directions from the pedestrian. What do you say to your friend?

R5 You do not have a car. You ask a neighbor whom you do not know very well to help you move some things out of your apartment with his car. You do not have anyone else to ask since everyone you know appears to be on holiday and you have no money either to hire someone who can help or to arrange transport. You see your neighbor in the lobby and go to ask him for help. What do you say to him?

R6 Your car has just broken down and you need to pick up your spouse from the airport urgently. There is no other means of getting there but by car. You go to your manager's office at work, with whom you get on well, and ask him to borrow his car. What do you say to him?

R7 You have been put in charge of a very important project at work. Your colleague has already booked a ticket to go on holiday. You realize you will need all members of staff to finish the project on time so that you ask him to stay. What do you say to him?

R8 You have been put in charge of a project at work. You go to the desk of a colleague and ask him to type a few letters for you. What do you say to him?

R9 A friend of yours has a house in the countryside. You want to go on holiday somewhere to relax for a week. You know nobody is going to be in the house for at least two weeks. You meet your friend in a pub and seek permission from him to stay in his country house for a week to relax. What do you say to him?

R10 You are on a bus with your child. Although there are plenty of seats on the bus, there are not any two-seater seats that are available. You ask a passenger who is sitting on his own on a two-seater to change seats with you so that you can sit next to your child. What do you say to him?

R11 You have received a lot of house bills which are due for payment. You do not have any money. You can not ask your friends for money since you have a reputation of never paying back. The company where you work will not give you an advance salary because you had already asked for one. You desperately need to pay these bills otherwise you will not have any electricity, gas or telephone service. You go to the office of the recently appointed manager at work and ask him for the money. What do you say to him?

R12 You have been working for a company for sometime now. One of the new trainees has brought his brand new laptop to work. You ask him to use it for a while. What do you say to him?

A CONVERSATION TASK INSTRUCTIONS

Experimenter Text:

> In this task you can prompt the AI as you feel like but keep in mind that if you want to prompt on something I just said, it takes around half a second for everything to be transcribed.
>
> In this task, we will be planning an exciting trip together. We have around 5 minutes for this brief planning session. Imagine that the decisions we make will shape our real adventure. By the end of this exercise, we should have some ideas about the major aspects of the trip.
>
> 1. let's choose an Exciting Destination:
> - Consider each other's interests and preferences.
> - Come up with a list of potential destinations that you both find appealing.
>
> 2. Select Possible Travel Dates:
> - List down a few sets of dates when the trip could take place.
> - Consider the best time of the year for the chosen destination.
>
> 3. Create an Itinerary:
> - Plan a rough itinerary for the trip, including activities, sightseeing, and relaxation time.
> - Decide on the duration of the trip, keeping in mind your schedules and the availability of time.
>
> 4. Decide on Accommodations:
> - Discuss the type of accommodation you prefer (hotel, Airbnb, etc.).
> - Research and shortlist potential places to stay.
>
> 5. Budget and Expenses:
> - Set a budget for the trip and agree on how expenses will be managed.
> - Allocate funds for different aspects of the trip, such as travel, accommodation, food, and activities.
>
> 6. Activities and Attractions:
> - Brainstorm and decide on specific activities or attractions you both want to experience.
> - Consider any must-visit places, events, or experiences.
>
> 7. Packing and Essentials:

- Discuss what essential items to pack for the trip.
- Consider any specific requirements based on the chosen destination and activities.

Try to be as detailed as possible within the allotted time, and most importantly, have fun planning this exciting adventure together! Remember to be imaginative and adventurous in your decision-making. Enjoy the process!

A INTERVIEW QUESTIONS

A **Opening:** What's your impression of the technology?

B **Follow-up**:

- Please elaborate a bit on your experience with the trigger methods used in the study (looking upward or pressing the button)? How did they affect the way you use the system?
- What were your impressions of the different display modalities (Text and audio) employed in the study? Did you find one modality more effective than the other in supporting you?
- In terms of agency (ownership over your responses), how did you perceive your control and autonomy during the conversation USING the system?
- Did you feel empowered to guide the conversation, or did you rely heavily on the system's recommendations?
- How did you feel about the guidance of the system? (effective guidance of the flow of the conversation)
- Could you share any specific situations where you found the system particularly useful?
- Could you share any specific situations where you found the system particularly challenging to use?
- How would you describe the flow and the dynamics of the conversation when incorporating the system's recommendations?
- How did you decide when to accept or reject the system's recommendations?
- Can you provide examples of strategies you employed to integrate the system's suggestions into your responses?

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009