MeetMap: Real-Time Collaborative Dialogue Mapping with LLMs in Online Meetings

ANONYMOUS AUTHOR(S)

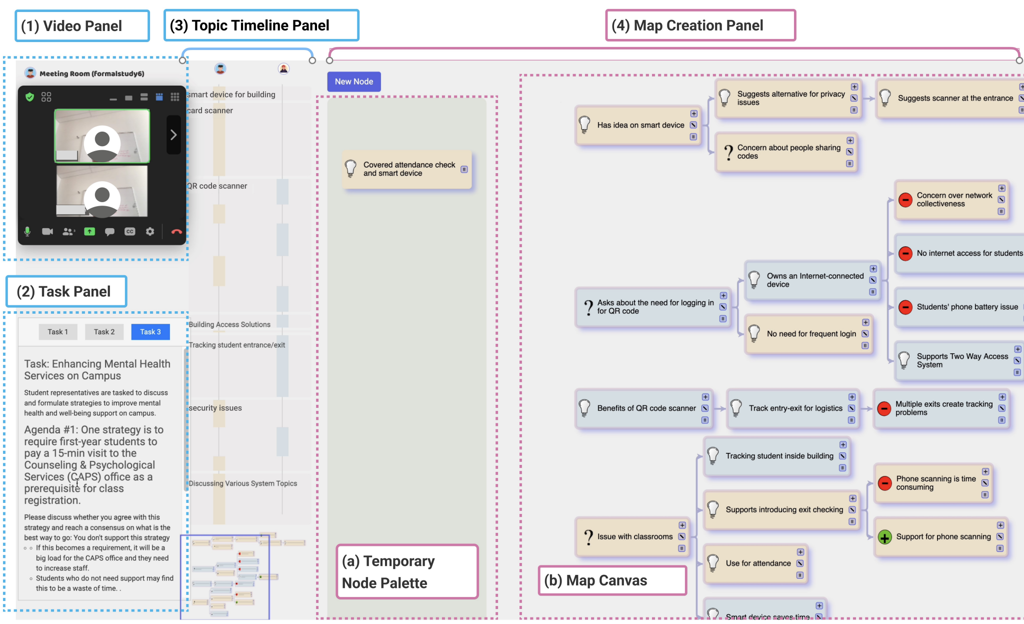

Fig. 1. **MeetMap User Interface**. MeetMap allows users to create dialogue maps collaboratively in real-time during online meetings through (1) a Video Panel and (2) a Task Panel, which displays the meeting agenda; (3) a Topic Timeline Panel, which shows turn exchanges and conversation topics chronologically; (4) a Map Creation Panel, which includes two parts: a) A Temporary Node Palette which presents nodes that summarize users' speech by AI, arranged in chronological order; b) A Map Canvas enable collaboratively work to create and refine a dialogue map using nodes.

Video meeting platforms display conversations linearly through transcripts or summaries. However, ideas during a meeting do not linearly emerge. We leverage LLMs to create dialogue maps in real-time to help people visually structure and connect ideas. Balancing the need to reduce the cognitive load on users during the conversation and give users sufficient control when using AI-generated content, we explore two human-AI collaborative methods. In Human-Map, AI generates summaries of conversations as nodes, and users create dialogue maps with the nodes. In AI-Map, AI produces dialogue maps where users can make edits. We ran a within-subject experiment with ten pairs of users, comparing the two MeetMap variants and a baseline. Users preferred MeetMap to traditional methods for note-taking, which aligned better with their mental models of conversations. Users liked the ease of use for AI-Map due to the low effort demands and appreciated the hands-on opportunity in Human-Map for sense-making. This work informs the future design of AI-assisted tools for real-time cognitive scaffolding in meetings by emphasizing the necessity to balance AI assistance with synchronicity and user agency to enhance collaborative sense-making.

CCS Concepts: • **Human-centered computing** → **Collaborative interaction**; **Collaborative and social computing systems and tools**; *Empirical studies in HCI*.

Additional Key Words and Phrases: Video Meetings, sense-making, Visual Representations, Dialogue Mapping

1 INTRODUCTION

Keeping up with and making sense of the information exchanged in online meetings is crucial yet challenging [1, 7, 65]. Turn exchanges happen in rapid succession and progress chronologically, which does not align with the non-linear way that ideas develop during conversations and how people process and organize information in their minds [17, 76]. The lack of parallel communication channels and body language cues [16, 94], and the temptation for multitasking [13] further the challenge in online video meetings.

To address those challenges, recent work in HCI and CSCW has explored providing real-time support to meeting attendees to make sense of the video meeting in context [4, 16]. This includes providing in-context enrichment materials, including images [57] and participation polls [90] to facilitate discussion. While these aids can enhance active engagement, they fall short in assisting participants to fully comprehend conversations in real-time and foster a shared understanding among team members. ome systems have aimed to bridge this gap; for example, TalkTraces maps discussion content to the meeting agenda in real-time[15], and MeetScript provides real-time transcripts that participants can collaboratively annotate to make sense of meeting content as it unfolds [16]. Although these systems were crafted to aid in real-time comprehension, they also informed the risks of heightened cognitive demands due to the additional content presented instantly. Individuals found it challenging to simultaneously engage in conversation and absorb the verbatim transcripts and agenda visualizations as they unfolded in real-time [15, 16]. Additionally, commercial tools have introduced real-time summaries powered by LLMs [67]. Yet, studies reveal that these AI-generated summaries can be too abstract and difficult to interpret [5], failing to capture the complexity of discussions, particularly the interplay of linear and non-linear information that is critical during discussion process [22, 91].

These findings highlight the need for novel interaction and visualization methods that provide real-time support without imposing high cognitive load and that go beyond overly simplistic textual summaries to capture the full complexity of conversations. Dialogue mapping [20, 21] offers a promising solution by visually representing discussions with real-time summaries using a clear, structured schema. In practice, dialogue mapping is often facilitated by a dedicated facilitator who creates and updates the visual representation of the conversation in real-time by capturing the conversation into 'questions', 'ideas', 'pros', and 'cons', and linking them into a coherent map; meeting attendees can then communicate with the facilitator to modify the dialogue maps [20]. In contrast to existing methods, such as real-time transcripts and overly concise summaries [15, 16], the facilitator creates the dialogue maps gradually as the conversation goes, offering a structured visual format that encapsulates the intricate and non-linear nature of discussions [19, 21]. This incremental mapping method enables attendees to follow the discussion dynamically, alleviating the need to process every spoken word in transcript [16] or endure excessively brief summaries at periodic intervals [5]. However, the requirement of a dedicated facilitator makes this approach impractical for many meetings that do not have such resources [76, 79]. Moreover, employing a dedicated facilitator also takes away the cognitive processes other attendees may engage in to make sense of the content of the meeting actively [46].

Given the potential and limitations of traditional dialogue mapping, We implement and study the idea of "collaborative dialogue mapping" in video meetings, where meeting attendees can collaboratively build dialogue maps to facilitate their real-time understanding of the conversation. This research addresses the question of when meeting attendees are tasked to create dialogue maps collaboratively during video meetings, how would that influence their discussion quality, understanding of the conversation, and meeting experiences compared to business-as-usual meeting setups? To optimize the collaborative dialogue mapping experience, we further explore the role of AI in facilitating this process. Our focus is on finding interaction mechanisms that strike the right balance of AI support during real-time dialogue mapping. This leads us to the core question: when using AI to facilitate real-time collaborative sense-making tasks—such as creating dialogue maps—how much AI assistance is ideal to maintain user engagement in the cognitive process while reducing the burden of manual map creation?

To address the two research questions, we designed and developed MeetMap (Figure 1), an AI-assisted collaborative dialogue mapping system for meeting participants to make sense of the group conversation in real-time during online meetings. The MeetMap user interface contains 1) Video Panel (Figure 1(1)), which connects to a Zoom meeting; 2) Task Panel (Figure 1(1)), which displays the meeting agenda; 3) Map Creation Panel (Figure 1(3)(4)), where users can collaboratively create dialogue maps.

Meanwhile, MeetMap provides AI assistance to facilitate the creation of dialogue maps. Prior work on human-AI interaction has shown that it is critical to give users control over when and how to use the AI outcomes [2, 93], especially for high-stakes creativity tasks [58]. To understand how much assistance from AI is desired by users to support sense-making in meetings, we implemented two variants of MeetMap with different levels of human involvement and AI proactivity: Human-Map and AI-Map. During the meeting, participants' speech is transcribed and summarized into nodes by AI, which appear in the Temporary Node Palette following a chronological order. In Human-Map, AI generates the summary nodes, and users will create the maps themselves. Users can drag and drop the nodes to the Map Canvas to create dialogue maps. In AI-Map, AI generates drafts of dialogue maps, where users can further make edits as they wish.

We performed an evaluation study with 20 participants divided into 10 pairs. Each pair of participants experienced all three conditions, including the two MeetMap variants (Human-Map and AI-Map) and a baseline condition. We adopted a business-as-usual baseline condition in which participants discussed over Zoom while taking notes collaboratively in a Google Doc with AI-generated summaries. In each condition, the pair of participants had a 30-minute discussion to make decisions on a task. Here is a summary of the findings. First, participants found MeetMap (both Human-Map and AI-Map) to be more helpful than the baseline condition in helping them keep track of the meeting content in real-time and facilitating subsequent discussion. Users liked that MeetMap allowed them to collaboratively organize information through dialogue maps without adding more cognitive burden (as shown in 5.1.3). Second, users in Human-Map considered the AI-generated content to be synchronous and succinct so that they could read and act on it in real time. Furthermore, users desired to engage in cognitive and creative thinking processes in Human-Map, even if it required more work. This research guides the future design of AI-assisted tools for real-time cognitive scaffolding during meetings by highlighting the need to balance AI assistance with both synchronicity and user agency to improve collaborative sense-making.

2 RELATED WORK

2.1 Challenges of Keeping up with and Making Sense of Online Meetings

Group meetings are events where participants discuss, negotiate, present, and create materials together in a communicative manner [82]. To make the group meeting effective, it is critical for meeting participants to keep up with the ongoing conversations [31], make sense of the discourse [55, 86], and foster a shared understanding [9, 11, 18].

However, significant challenges arise in real-time comprehension and sense-making of discussions. One primary reason is that there is a lack of visual representation of conversations, making it hard for people to revisit missed content [7, 77]. Moreover, interpreting and making sense of conversations involves an associative memory process, where individuals link scattered topics over time to form a cohesive understanding, while people's working memory and attention spans are limited [87]. The challenges become more severe when the session is running long [45], the meeting is poorly structured [45], and the discussion is back-and-forth[11, 22] with irrelevant side discussions [36]. Additionally, sense-making in meetings is not only an individual behavior but a collective move, wherein meanings are progressively built through collaboration [70]. This process of grounding, or building a shared understanding, is central to successful collaboration [17]. However, building a shared understanding among team members can be extra challenging for virtual or hybrid meetings due to the lack of physical shared space [49, 94], social cues [64], and technological barriers and fatigues [6].

2.2 Technologies to Support Sense-making in Online Group Meetings

To help people keep track of and make sense of synchronous group meetings, researchers have explored various technological interventions. The main focus is post-meeting tools, such as dashboards with visualizations designed to help users review content and reflect on their participation [73, 74]. However, these tools fall short of aiding real-time sense-making. We have witnessed more research in recent years to support real-time sense-making and understanding of meeting content. One line of work visualizes non-verbal cues to mimic the communication experience in face-to-face discussion [4, 52, 64, 77], for instance, amplifying expressions and body language to enhance situational awareness [4, 4, 64]. Another line of work uses real-time dialogue analysis and visualization to improve situated understanding [15, 41, 57]. TalkTraces, specifically, links current discussions with past meeting agendas [15]. Additionally, web-based collaborative whiteboards have been used for sharing screens and discussion resources, creating external memory aids and visual representation of the conversation to reduce cognitive load [94]. However, these approaches often present information passively, missing opportunities for active user engagement in sense-making, which is key for understanding and memory [75]. MeetScript introduced collaborative annotation of live transcripts, showing that actively engaging users to make sense of the transcripts aids comprehension [16]. But the verbosity of full transcripts can be cognitively burdensome [16, 33], and the linear presentation of information is misaligned with the iterative and structural nature of how people understand the conversation [17, 76].

Identifying these gaps, this paper introduces MeetMap to help people make sense of meeting content in situ. First, we aim to actively engage users in the sense-making process, supporting them to collaboratively create an external representation of the discussion. Second, we aim to reduce the unnecessary cognitive load on processing real-time comprehension aids, such as verbatim transcripts. Third, in contrast with the linear representation of the conversation following a chronological order, we aim to provide more flexibility for users to organize their thoughts.

2.3 Collaborative Note-Taking and Concept-Mapping Supports Sense-making in Teams

Collaborative note-taking enables meeting participants to create shared notes [8, 23, 38, 63], which is found to enable real-time clarifications, reduce misinterpretations, and help the team build a shared understanding [26, 43, 72] in comparison to individual note-taking. Moreover, collaborative note-taking reduces the burden on individual team members [34, 43, 61], and improves the quality of the shared notes with diverse viewpoints [30].

Collaborative note-taking methods can be categorized into note-taking following a chronological order and note-taking non-linearly and visually [26, 34, 43, 61, 72]. Studies have shown that taking collaborative notes chronologically enhances engagement and reduces distraction [32] , however, it also can be insufficient to support sense-making for complex information [14]. Taking notes visually and non-linearly, on the other hand, allows for flexibility and better integration of ideas through various forms of external representations like concept maps and diagrams [56, 78].

Prior work has shown that creating concept maps collaboratively as a team improves learning of lecture videos [14, 56] and brainstorming of new ideas [51, 78, 84]. A similar idea, called "Dialogue Mapping", has been explored in n-person group meetings, which uses a concept map to help teams visualize the progression of ideas in real-time, serving as a team facilitation technique[20]. Mapping categorizes conversations into nodes with IBIS notation schemas, including Questions, Ideas, and Pros and Cons arguments, and visually connects these nodes to form a 'map' [21]. Previous dialogue mapping practices often require a dedicated facilitator to create the dialogue map during the meeting [19], limiting the practical use of this approach. Moreover, employing a dedicated facilitator also takes away the cognitive processes people may engage in to make sense of the meeting content actively [46]. Additionally, creating a high-quality dialogue map demands substantial manual effort, making it particularly challenging in synchronous communication scenarios like video meetings, which are already cognitively demanding [27]. Building upon prior work, we aim to integrate collaborative dialogue mapping capabilities in video meeting contexts. More specifically, we aim to redistribute the labor of creating dialogue maps to all meeting participants and use AI to scaffold the process of creating dialogue maps as a group.

2.4 The Use of AI to Support Real-Time Communication

With the recent advancement of large language models (LLMs), researchers have begun to leverage them to support sense-making of texts, e.g., facilitating text comprehension with LLM-generated diagrams [40, 83], and assisting reading and writing [44, 58, 99]. Using LLMs to structure real-time spoken conversations introduces extra challenges [35, 53]. It places a higher emphasis on presenting the essential information to users while they manage to speak, listen, and comprehend the discussion content simultaneously [5, 16].

Prior research has seen initial successes using AI to enhance human-human communication [35, 53]. For example, real-time transcripts and translations help cross-cultural discussion [98], and conversational agents can facilitate communication and collaborative decision-making [47, 48]. However, similar to using AI in other activities that require cognitive engagement, using AI to facilitate in-situ sense-making in conversations may introduce an assistance dilemma [50]. For example, early studies have shown that when one team member is responsible for documenting the content of a meeting, it limits the other attendees' engagement with the content [71]. In contrast, engaging all meeting participants in note-taking reinforces their understanding and retention [43, 60]. Indeed, it is shown that when using AI to support human-human communication, over-reliance on AI may occur [5]. It discourages people from actively engaging in the conversation and may weaken people's skills in regulating and organizing the conversation [48]. Prior work has argued that AI should act as an enabler and not a replacement in human communication [101], users should have more control over when and how to use information provided by AI [57], and that it is critical to provide just the right amount of AI-generated content to users [2, 80]. Our work closely follows the recommendations made in the literature on AI-mediated communication, in which we aim to give users agency and flexibility in organizing their thoughts.

As a summary of previous work, people continue to face challenges in keeping track of and actively making sense of discussions in real-time. Dialogue mapping has been shown to be a successful meeting facilitation technique [20]. However, the requirement of designated facilitators limits the practical use of the approach. In this work, we aim to integrate collaborative dialogue mapping capabilities in video meetings, with the goal of enhancing people's real-time understanding of the conversation. To facilitate the dialogue mapping process, we provide people with AI assistance. We also build upon prior work on AI-mediated communication to ensure that AI provides the right amount of assistance to users.

3 MEETMAP: GENERATING DIALOGUE MAPS AS REAL-TIME COGNITIVE SCAFFOLDS FOR ONLINE MEETINGS

3.1 Design Goals

We propose MeetMap, an AI-assisted collaborative dialogue mapping system for meeting participants to make sense of the group conversation in real-time during the meetings. The design of MeetMap is inspired by prior work on dialogue mapping [20] and collaborative sense-making tools [16, 56, 95]. MeetMap is designed to support collaborative sense-making in a new context that has not been explored before, namely for meeting attendees to develop a shared understanding of the meeting content in situ with the collaborative dialogue map creation. Below, we list the design goals that we derived from the prior work and solutions that we employed.

- **Design Goal 1: Enable people to collaboratively build dialogue maps (D1).** Existing dialogue mapping techniques have been found to enhance real-time understanding by employing a notation schema and visual structure [20, 21]. However, a significant limitation of current dialogue mapping techniques is the need for designated facilitators [66, 102], which might take away the beneficial cognitive processes when people engage in such activities actively [50]. Prior work on collaborative sense-making highlights the importance of active engagement through creating external representations collaboratively rather than passively receiving information [95, 96]. Inspired by these insights, MeetMap provides a shared space where meeting participants can collaboratively create and edit the dialogue maps as the conversation unfolds.
- **Design Goal 2: Help people manage the cognitive load of creating dialogue maps during the conversation (D2).** Creating dialogue maps can be a cognitively demanding process due to the complex visual structure it requires [20, 69]. While encouraging users to collaboratively create dialogue maps, the system should also reduce the extra cognitive effort brought by the collaborative map creation . To achieve this, the design of MeetMap leverages AI to categorize turns and generate short summaries of conversations and enables users to have the capacity to peruse these AI-generated summaries in the creation of dialogue maps.
- **Design Goal 3: Help people easily make sense of the AI assistance when creating dialogue maps (D3)**. Prior work on using AI to support collaborative sense-making informs the need to present information at varying levels of detail, from overarching themes to specific utterances [15, 90]. Prior work also suggests supporting seamless interaction between linear and nonlinear representation to avoid losing context when understanding diagram solely [39]. Taking inspiration from those studies, in the development of MeetMap, we devised several visualization techniques to assist users in interpreting AI-generated content. This includes 1) showing the original transcript of the summaries so that people know where the summaries come from; 2) differentiating the turns by different speakers; 3) organizing the discussion in chronological sequence together with dialogue maps to help participants associate linear and nonlinear information during meetings.
- **Design Goal 4: Identify the right level of AI assistance to provide (D4).** Providing more AI assistance may limit the opportunity for users to engage actively in the sense-making process [28, 54]. Extensive research in human-AI collaboration emphasizes the importance of maintaining user agency [2, 97, 99]. Drawing insights from these studies, we aim to design dialogue mapping systems that seek a desired level of AI assistance with human involvement and control. In the design of MeetMap, we employ two levels of AI assistance. In Human-Map , AI generates short summaries where users will create dialogue maps leveraging the summaries. In AI-Map , AI directly outputs dialogue maps.

3.2 Iterative Design and Development

The goal of MeetMap's design is to develop an AI-assisted collaborative dialogue mapping system that aids meeting participants in understanding conversations in real-time. Nonetheless, the optimal degree of AI assistance and the method of information presentation to reduce cognitive load have yet to be determined. The design and development of MeetMap followed an iterative, user-centered approach, refined through several pilot tests and feedback cycles to optimize usability, cognitive load management, and real-time collaborative mapping with AI assistance. This section provides a detailed explanation of these iterations and the insights gained at each stage.

3.2.1 Overview of the initial design. The initial version of MeetMap was designed to support collaborative dialogue mapping in real-time online meetings, addressing the four core design goals as mentioned in section 3.1.

The core design of MeetMap's features has remained consistent from the initial to the final version. The core function of MeetMap contains a Map Creation Pane, including a Temporary Node Palette with AI-generated summary nodes and a Map Canvas where users collaboratively build a conversation map, as shown in Figure 1. This design enables active engagement in shared mapping **(D1)** while managing cognitive load **(D2)** by generating categorized summaries in near real-time. Visual aids, including color-coded speakers and a Topic Timeline, help users link AI content with the conversation's flow **(D3)**. Additionally, the system provides two levels of AI assistance: in Human-Map , users build the map from AI-generated nodes to structure their ideas, while in AI-Map , the system drafts a structured map for users to refine, offering different levels of guidance **(D4)**. Detailed description is shown in section 3.3. However, the early version of AI-Map initially appended new nodes to create a single, comprehensive map covering the entire conversation—similar to facilitator-led dialogue mapping, where one person continuously adds to a complete conversation map.

3.2.2 Pilot-test and Improvement. We conducted three pilot tests (3 groups of 2) with the early version of MeetMap. During these pilots, participants completed group discussion tasks similar to the formal studies (detailed in Section 4). After each discussion, we interviewed participants for feedback on the system's usability.

Some detailed design decisions are made based on user feedback. **(1) First**, feedback indicated challenges with real-time node updates, as the delayed appearance of nodes interrupted the conversation flow. In response, we implemented a turn-splitting mechanism, as detailed in section 3.3.1, to display AI-generated nodes immediately after each turn, ensuring participants could access summaries in near real-time without waiting for delayed processing. **(2) Second,** the early version of AI-Map automatically generated a large dialogue map for the entire conversation, which some users found overwhelming. Users preferred to receive information gradually to help them digest it. Based on this feedback, we transitioned to a topic-based segmentation, where smaller, digestible maps are generated incrementally, as detailed in section 3.4.2. **(3) Third,** as maps grew larger, users encountered navigation challenges. To address this, we introduced interactive connections between the Map Canvas, Temporary Node Palette, and Topic Timeline, enabling users to seamlessly navigate between linear and non-linear information, as shown in section 3.3.1. Additionally, we incorporated a mini-map at the bottom of the screen, which makes navigation more intuitive.

3.3 System Design

We reported the final system design of MeetMap. An overview of the MeetMap system is shown in Figure 1. MeetMap's front-end interface has three components: 1) A Video Panel for users to have online meetings (Figure 1(1)); 2) A Task Panel to upload meeting agendas (Figure 1(2)); 3) A Map

Creation Panel which contains a Temporary Node Palette (Figure 1(a))and Map Canvas(Figure 1(b)), which enable users to create a shared representation of meeting content in real-time.

3.3.1 Map creation panel. MeetMap offers a Map Creation Panel that allows users to collaboratively and flexibly construct dialogue maps during online meetings with ease **(D1)**, while also being designed to prevent overwhelming users with the additional content supplied by AI **(D2)**.

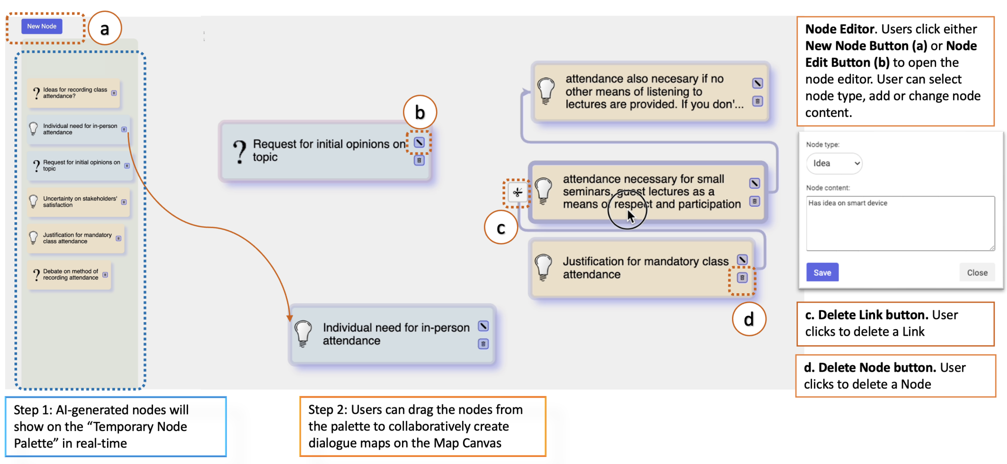

Fig. 2. **Users can collaboratively create dialogue maps in MeetMap.** Users can use the AI-generated nodes to create the map. (1) Nodes are shown in Temporary Node Palette in real-time. (2) Users can drag nodes to the map and create links between the nodes. Using the interaction suite (a-d), users can create/delete nodes (a, d), edit nodes (b), and delete links (b). When creating/editing a node (b), users can specify the node content and node type.

Map Canvas. MeetMap provided a suite of map interaction features, illustrated in Figure 2, to enable users to create dialogue maps collaboratively **(D1)**. As supported by other collaborative concept map authoring tools, e.g., Miro Board, users can collaboratively create, edit, and delete nodes and links on the Map Canvas. These features were provided to meet the basic needs of creating a collaborative dialogue map of group meetings. In addition to editing the node content, users can assign a node category using the "dialogue mapping" notation schema [20].

Temporary Node Palette. To help people manage the cognitive load of creating dialogue maps during the conversation **(D2)**, MeetMap displays AI-generated summaries of users' conversations as nodes on the Temporary Node Palette (Figure 2 (1)) to reduce the human effort of creating notes. MeetMap detects each turn as a single unit to generate the summary. The system recognizes speakers, detects and transcribes turns in real-time using the Azure speech-to-text service [1]. Since participants typically take notes on what has just been discussed, the AI-generated summary should be shown almost synchronously [10]. To mitigate the lag from lengthy monologues, MeetMap employs a 50-word checkpoint to prompt new conversation segments. This threshold, informed by studies indicating an average turn length of 50 words in group meetings [16], does not abruptly end conversations at 50 words but waits for a natural sentence closure. Azure's speech-to-text service detects this natural sentence end. If a sentence concludes after surpassing 50 words, the system initiates a new turn, even without a change of speaker. This strategy prevents both mid-sentence interruptions and reduces delays from extended monologues.

[1]https://azure.microsoft.com/en-us/products/ai-services/speech-to-text

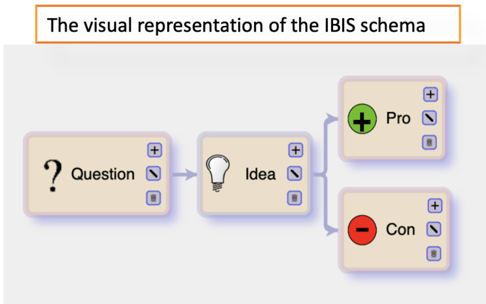

Fig. 3. The visual representation of the IBIS notation schema: We use symbols to visualize the dialogue mapping notation schema, including "Questions, ideas, pros, and cons"

For each detected turn, it is sent to GPT4 to 1) assign a tag based on the IBIS notation schema used in dialogue mapping techniques[20], as shown in Figure 3. The four categories of the IBIS notation schema include **Questions**: problems or issues that need to be addressed; **Ideas**: responses and proposed solutions to issues; Arguments includes **"Pro"** and **"Con"**, as justifications supportive or against a particular idea. 2) generate summary nodes of this turn. The AI will generate one or several summary nodes in one turn if the different part in one turn belongs to various categories. We used the GPT-4 API to categorize turns and generate summaries. The full prompt used in the system can be found in the Appendix. E. The output will appear as one or several nodes in the Temporary Node Palette in real time following a chronological order, as shown in Figure 2 (1). Users can delete any node from the Temporary Node Palette. Our system allows for no minimum turn length, meaning concise responses like "Agreed" or "Yeah!" are evaluated by GPT-4 to assign a dialogue mapping tag with a consistent prompt to maintain tagging uniformity. Non-qualifying turns are excluded and do not produce nodes. This method captures all elements of the dialogue, ensuring that even minor interjections or queries contributing to the conversation's flow are accurately represented using the IBIS notation schema.

Years of experience have shown that IBIS does not introduce too much cognitive overhead but has just enough structure to capture a conversation using only four categories and a limited number of building blocks [19–21, 37]. Although the structure is simple, it effectively promotes asking the right questions and keeping discussions on track [37, 88]. We acknowledge the limitations of the notation schema. In particular, the schema may not cover the whole range of conversational nuances or be appropriate for all meeting categories. However, the primary goal of using this schema in MeetMap is not to assert that it is universally applicable, but to offer it as an example of how AI assistance may aid group understanding through dialogue mapping techniques. MeetMap's other interaction design is agnostic to the notation schema, which allows for the adaptation or replacement of the notation schema if the alternative schema proves more effective in other meeting scenarios. We will further discuss the generalizability of using an AI-generated notation schema to categorize meetings in the discussion.

Interaction between the Map Canvas and the Temporary Node Palette. Different from other tools, the interaction of Map Canvas in MeetMap was specifically designed to reduce the effort of creating dialogue mapping collaboratively during meetings and help users to make use of the AI assistance **(D2)**. The Map Canvas is seamlessly interconnected with the Temporary Node Palette that users can drag the nodes generated by AI in the Temporary Node Palette to create dialogue maps on the Map Canvas.

3.3.2 Visualizing information with different detail levels. MeetMap visualizes conversation information with different granularity to help people easily make sense of the AI-generated content when creating dialogue maps **(D3)**, as illustrated in Figure 4. MeetMap summarizes the conversation into summary nodes in Temporary Node Palette. Users can view the original transcript that AI used to generate the node through double-clicking a node. To help people quickly gain an

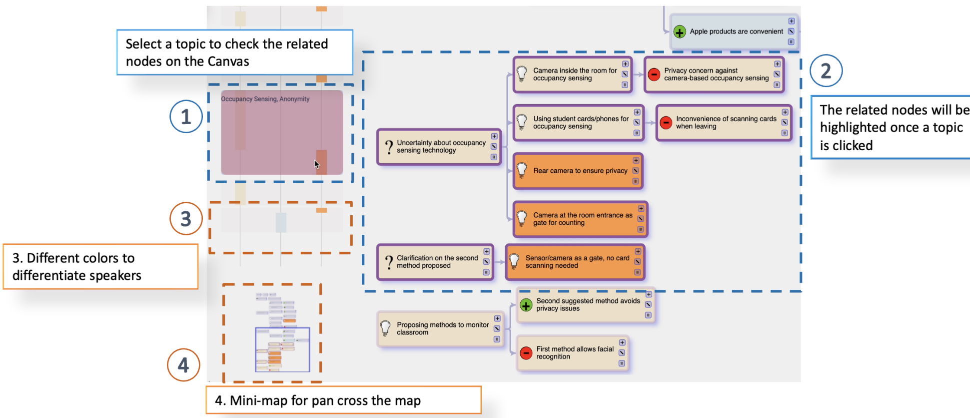

Fig. 4. **We introduced a series of visualizations to improve the usability of MeetMap**: (1) The discussion topics are segmented and labeled in the timeline; (2) When users click a topic in the timeline, the topic block changes colors, and the related nodes will be highlighted on the Map Canvas. (3) Different speakers are represented in different colors in the timeline view; (4) Mini-map for map navigation.

overview of the conversation, MeetMap visualizes the turn-taking and collaborative dynamics through Topic Timeline Panel, with speakers represented in different colors, as shown in Figure 4(3). The AI-generated nodes in the Temporary Node Palette and Map Canvas follow the same color coding for each speaker to visualize individual contributions. To further meet the needs of connecting chronological information with the visual representation of conversation **(D3)**, we support users to map information from the Topic Timeline Panel to the Map Canvas. Users can click on a topic block on the timeline to view the highlighted corresponding nodes on the map canvas (Figure 4(1)(2). Additionally, we added a Mini-map at the bottom of the screen. The mini-map is a scaled-down version of the full map, which shows the overall structure and the color-coding of the full map without details, serving as a map locator for moving large canvases when there are many nodes and connections (Figure 4 (4)).

3.4 Two Variants of MeetMap: Human-Map and AI-Map

Informed by both the prior literature on user agency in human-AI interaction [2, 58] and AI-mediated communication [42] and the pilot study with the early system design, we designed two variants of MeetMap with different levels of user involvement and AI assistance **(D4)** In Human-Map , users see AI-generated nodes and will create dialogue maps themselves. In AI-Map , users see AI-generated small dialogue maps automatically and users can refine them.

3.4.1 Human-Map : AI generates only summary nodes, and humans create the links. Human-Map enables users to create a conversation map that meets their needs. AI will only generate nodes with the dialogue mapping notation schema and present those nodes on the Temporary Node Palette. In this way, AI took on the work — which was shown to be burdensome to users — of continuously noting down the key points in the conversation. Users create links with AI-generated nodes to build a conversational map by themselves. The workflow of Human-Map is shown in Figure 5(left).

3.4.2 AI-Map : AI generates the small dialogue maps based on conversation chunks. In AI-Map , the AI will first generate the nodes and then generate small dialogue maps after identifying a topic

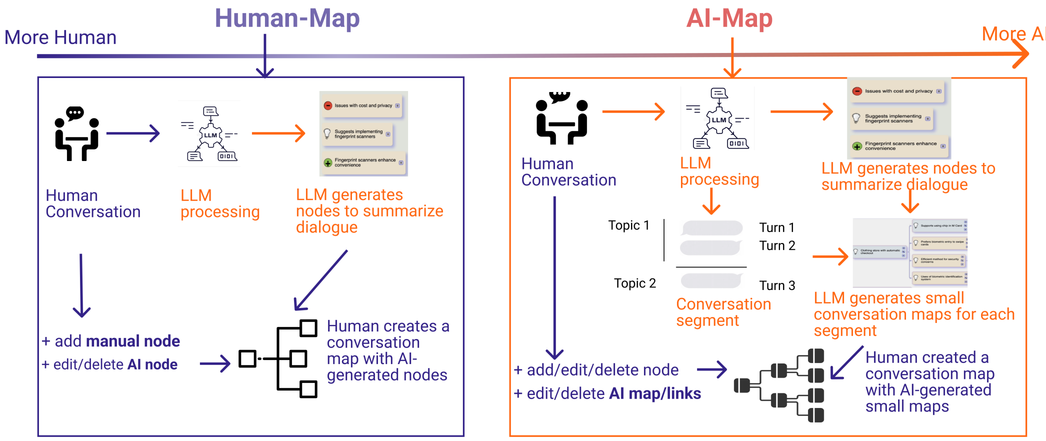

Fig. 5. **Two Variants of the MeetMap System with different levels of human involvement and AI assistance.** In Human-Map , AI only generates the nodes, and users will create the maps by themselves. In AI-Map , AI generates drafts of dialogue maps, where users can further make edits in any way they want.

chunk, as detailed in section 3.4.2. Users can freely edit the AI-generated map in any way they want. Given that the synchronicity of the AI-generated content is critical to users, in AI-Map , we first display the AI-generated nodes in the Temporary Node Palette in real-time, while the system works on creating the dialogue maps in the back-end. The workflow of AI-Map is shown in Figure 5 (right).

Segment the Conversation into Topic Chunks The system takes two consecutive turns as the input. It prompts GPT4 to identify if the new turn is a continuation of the previous turn with the same topic or initiation of a new topic. In AI-Map , the topics are also visualized on the Topic Timeline Panel.

Generate small conversation maps using the topic segment MeetMap then prompts GPT4 to generate a dialogue map for a given topic segment with the nodes in that topic. The system automatically moves the used nodes from the Temporary Node Palette and presents the nodes with links on the Canvas View, shown in Figure 6.

.

3.5 Implementation

For the front-end of the MeetMap, Zoom Web SDK [2] was used to support videoconferencing (as shown in the Figure 1, (1) Video Panel), and go.Js [3] was used to build the Map Creation Panel. We used Django Channels and Django-Redis to handle the real-time updates in MeetMap. We used the Microsoft Azure SpeechSDK [4] to provide transcription. MeetMap collects user audio from the client-side browser and then sends the transcription result to the database. Once turn-taking happens, MeetMap calls the GPT4 API to generate nodes as well as identify the topic in both MeetMap variants. Once a new topic is identified, the GPT4 API will be called again to generate links in the AI-Map condition. More details of the pipeline and prompt to generate the map are shown in the Appendix. E.

[2]https://developers.zoom.us/docs/meeting-sdk/web/
[3]https://gojs.net/latest/index.html
[4]https://azure.microsoft.com/en-us/services/cognitive- services/speech-to-text/

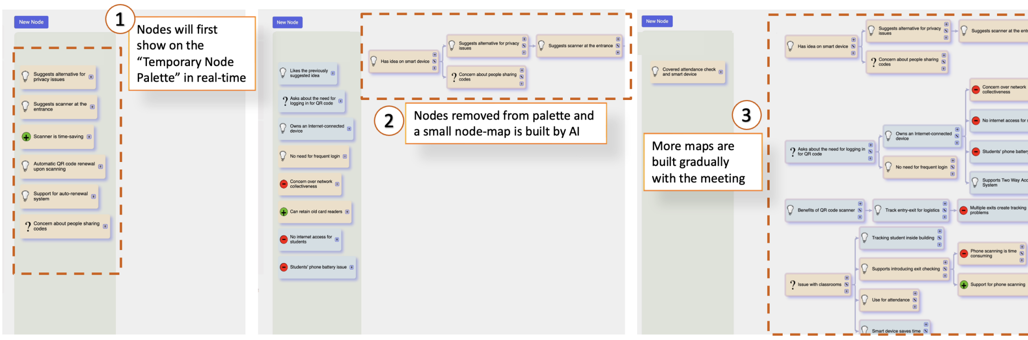

Fig. 6. **AI-Map provides incremental generation of conversation maps so that users perceive the AI-generated content to be in real-time and digestible.** (1) Summary nodes are generated in the Temporary Node Palette in real-time. (2) When a new topic is detected, the nodes in the previous topic are sent to generate a dialogue map. (3) As the conversation progresses, more small dialogue maps are generated.

4 EVALUATION STUDY

To understand how MeetMap supports people keeping up with and making sense of the conversation, we performed an IRB-approved evaluation study with three system setups, including the two MeetMap variants and a business-as-usual Baseline setup. We aim to answer the following research questions:

(1) RQ1: How AI-assisted collaborative dialogue mapping influenced meeting experiences compared to business-as-usual meeting setups?
(2) RQ2: How did the different levels of AI assistance influence users' interaction behaviors and attitudes towards creating dialogue maps during meetings?

4.1 Study procedure

4.1.1 Participant Recruitment. We recruited 20 participants from mailing lists from the University of Michigan and divided them into 10 groups, with 2 participants in each session. The demographic information of the participants is shown in the Appendix D. The selection of dyadic meetings can help assess how the scaffold in MeetMap can help people in meetings that require rapid information exchange and high cognitive demands. All participants were considered to have no hearing or reading difficulties. Each study session lasted for about 150 minutes, and participants were compensated with an hourly rate of $15.

4.1.2 Discussion Tasks. To encourage information exchange and collaboration, we used the jigsaw method to design discussion tasks [3]. Each participant is presented with a unique point of view, and they have to discuss with each other to reach a consensus on their decisions. To avoid additional difficulty in understanding the topic, we designed 3 tasks related to school life, including reevaluating attendance checking in university classes, installing smart devices for university buildings, and enhancing mental health services on campus. We designed 2 discussion agendas for each task to make the discussion concrete and encourage participation. The order of the tasks was counterbalanced for each condition. The tasks used in this study were attached in Appendix C.

4.1.3 Baseline Condition. We designed a Baseline condition that resembled business-as-usual meeting scenarios. We connected Zoom with Otter.ai [5], which is a popular Zoom add-on to provide

[5]https://otter.ai

real-time transcripts and automatic summaries for meetings. The transcripts are sent for a summary generation and presented in the summary panel in Otter.ai [6]. AI presented the summaries as chapters, with several key points and a title for each chapter. The chapter is expandable to present a meeting summary with a two-level hierarchy. After seeing the summary, the user could click on the summary to navigate back to the transcript. To the best of our knowledge, this was the most commonly used service that can provide summaries during meetings, rather than only offering meeting notes as post-meeting reviews. Besides, the hierarchy of the summary and the capability of tracking back the original transcript help digest information. The interval for showing a new summary is around 3 minutes, as set by Otter.ai, which strikes a balance between providing more structured and synthesized information and showing information synchronously.

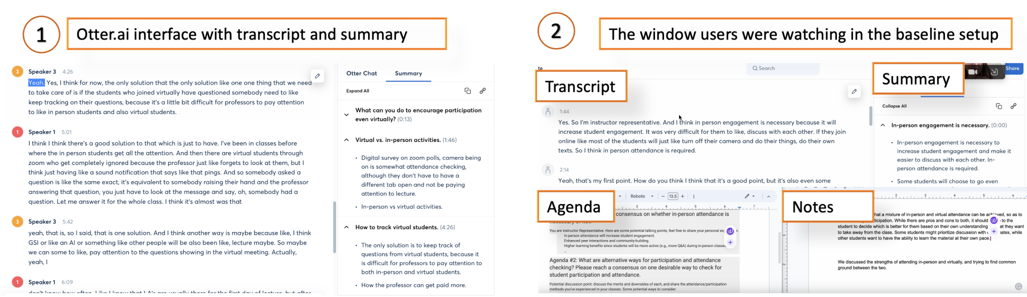

Fig. 7. **The Baseline condition.** (1) ***Otter.ai*** shows real-time transcript (left), and the key points (right) are summarized by AI. (2) In the baseline condition, three windows are opened and arranged for the participants. a) The otter.ai window is on top to show the real-time transcript and AI-generated summaries. b) The task agenda is placed at the bottom left. c) The shared document is on the bottom right.

As shown in Figure 7, the windows of the task and each tool were arranged so that all of them could be read simultaneously. This avoids additional load on the participants when looking up and switching windows during the discussion. The Baseline condition shows how AI was introduced in daily group meetings in a business-as-usual setting, with an editable and retrievable live transcript open and a bullet point meeting summary enabled and presented linearly. We argue the Baseline is comparable to the two MeetMap conditions because all provide AI assistance for creating shared representations with dedicated designs to help make sense of the AI-provided information: MeetMap uses AI-generated nodes, while Otter.ai offers structured summaries, both facilitating information digestion through different levels of granularity and navigation mechanisms. By comparing this Baseline setting to MeetMap, we can see if the AI-assisted dialogue mapping can aid in sense-making during group meetings.

In each session, two participants had a meeting online. Before the session began, we got the participants' verbal consent to record the session. Each participant went through all three conditions. The order of the tasks and conditions was counterbalanced, as shown in Figure 8.

4.1.4 Experimental procedure. In each task, the participants discussed each agenda for 7 minutes, followed by a 3 minute break for recap. They are encouraged to take notes or interact with the map during the discussion. The breaks are introduced since previous studies suggested breaks can help people quickly reflect on the conversation and take time to organize their thoughts[62].

[6]https://help.otter.ai/hc/en-us/articles/5093383818263-Automated-Live-Summary-Overview

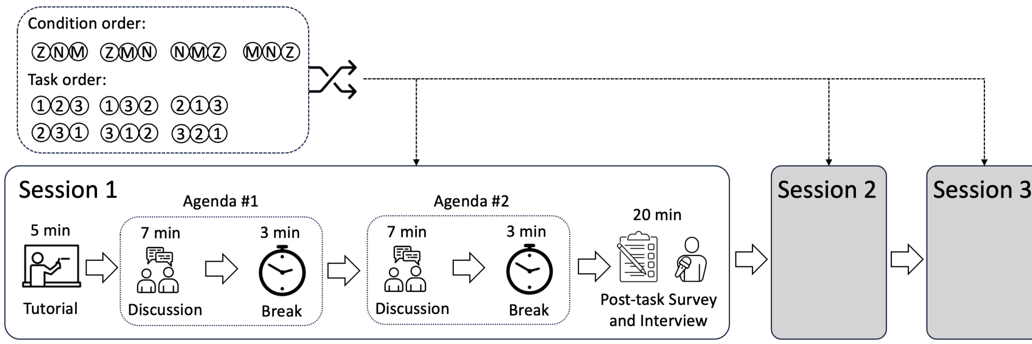

Fig. 8. **Study procedure:** All participants went through the three conditions (M: AI-Map , N: Human-Map , Z: Baseline) and the three tasks. The order of the tasks was counterbalanced. The two MeepMap conditions were always done sequentially, but the order between the two MeepMap conditions and the order between the MeepMap conditions and the baseline condition were also counterbalanced. A tutorial is given before the task. During the task, the participants discussed two agendas, with a break after each agenda. After each session, all participants answered a post-task survey and had an interview.

After each task, the participants answered a 5-Likert scale survey about the usability and usefulness of the system. Participants are also asked to complete the NASA-TLX test after each task, responding to six questions on an unlabeled 21-point scale [12]. Follow-up questions were asked after the survey about their experiences through a 10-minute semi-structured interview after each task. We specifically asked how they perceived and used the AI-generated content during the discussion and during the break. We also asked whether they were satisfied with the AI-generated content, the resulting dialogue maps and notes, and the challenges they experienced. Survey questions and semi-structured interview questions are provided in Appendix A and Appendix B respectively.

4.2 Data Analysis Methods

4.2.1 Log data analysis. We analyzed the note-creation and note-checking behaviors between the two MeetMap conditions and the baseline condition.

The note-creation behaviors included adding, editing, and deleting nodes and maps in the two MeetMap conditions, and creating, editing, deleting notes in Google Docs in the Baseline . For the two MeetMap conditions, the MeetMap system logs user interactions with the nodes and the map. For the Baseline , the edit history of the shared document was used to understand user behavior. Two authors watched the recordings to log what changes each participant made to the shared document. They treated each bullet point added by users as new notes and counted the number of times users deleted or edited notes on Google Docs. Specifically, if a user took notes using short phrases, each short phrase that conveyed a different discussion point was counted as one note-creation activity. If a user wrote a sentence summarizing multiple ideas and reasons, researchers counted each idea and each reason as a note-creation behavior. Additionally, other note-creation behaviors, such as changing the order of notes or editing existing notes, were tracked.

The note-checking behaviors counted the interactions users took to check and read the content. This included locating a node on the map by clicking on the timeline, panning across the map using the mini-map, scrolling the node palette to read more information, and double-clicking the node to check the transcript in the two MeetMap conditions. In the Baseline , note-checking behaviors included users scrolling on the transcript or their collaborative notes, users clicking to expand the summary on otter.ai, and users clicking the summary to jump back to the original transcript.

Although users interacted differently in MeetMap and the Baseline , the note-creation and node-checking behaviors were measured using a comparable granularity for consistent computation.

4.2.2 Survey data analysis. The survey data was evaluated using the Friedman test due to the data not meeting normality assumptions. For each question, we established a null hypothesis that there is no statistically significant difference among the three conditions. To compare the differences between any two conditions, we conducted the Wilcox signed-rank post-hoc test.

4.2.3 Interview analysis. The interviews were transcribed, and two researchers used the Affinity Diagram [59] to analyze the data. In the analysis, two researchers rearranged all quotes iteratively based on emerging affinity to one another through communication and critique. We grouped users' feedback, including how they create notes/dialogue maps with the help of AI both in-situ and post-meeting, why or why not the scaffold of MeetMap is helpful for them, their preferences, and concerns about incorporating AI in synchronous meetings.

4.2.4 Video data analysis. We further conducted a qualitative analysis of the video recordings to observe participants' collaboration behavior and their interactions with the system. The video analysis was used to understand the nuances of interactions and behavior patterns that might not be captured through log data or self-reported measures. Two researchers independently watched the video recordings and wrote memos about how people took notes during the discussion and during the break, how they collaborated with each other, and how they used and edited the AI-generated content. After that, they came together to discuss the memo and used it to complement and explain some findings from log analysis and interview analysis.

5 FINDINGS

The comparison between the two MeetMap variants with the Baseline demonstrates the usability and usefulness of MeetMap in helping people keep track of and understand conversations in real-time (RQ1). Participants valued the flexibility of MeetMap in allowing them to structure the conversation visually. The comparison between Human-Map and AI-Map showed that Human-Map provided more control and agency for users to motivate them to engage in real-time dialogue mapping (RQ2). Additionally, users reported lower tolerance for AI mistakes when they felt they put more effort into creating the map.

5.1 How AI-assisted Collaborative Dialogue Mapping Influence Meeting Experiences in Comparison to Business-as-usual Meeting Setups

We first present results on how MeetMap (Human-Map and AI-Map) helps people make sense of the discussion compared to the Baseline condition.

5.1.1 MeetMap helps individuals keep up with and make sense of the meeting in real-time . Users assessed the effectiveness of MeetMap in helping them keep up with the discussion content in real-time significantly better than that in the baseline ($p = 0.017 < 0.05$, $p = 0.46 < 0.05$). (Figure 9:Q1). Users thought the dialogue map reflected their decision-making process more accurately in both Human-Map ($p = 0.001 < 0.05$) and AI-Map ($p = 0.003 < 0.05$) conditions compared to the shared notes in the baseline condition (Figure 9:Q2). Additionally, users considered the AI-generated summary nodes and dialogue maps to be accurate representations of the conversation in the Human-Map and AI-Map in comparison to the AI-generated summaries of the transcripts in the baseline ($p=0.03 < 0.05$, $p=0.002 < 0.05$ (Figure 9:Q3).

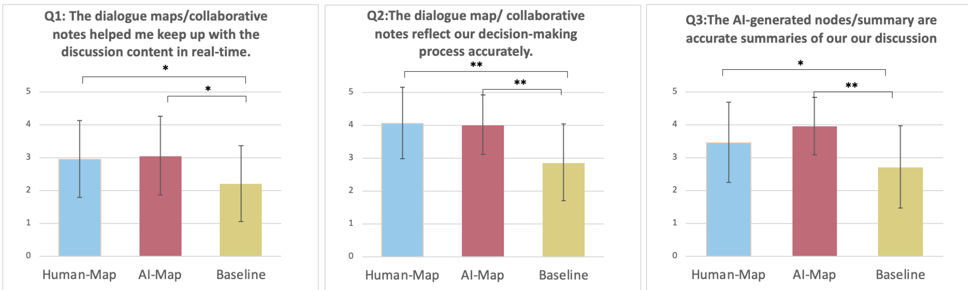

Fig. 9. **Survey questions** 1) Q1: The dialogue maps/collaborative notes helped me keep up with the discussion content in real-time. 2) Q2: The map/the AI transcript and summary reflect our decision-making process accurately. 3) Q3: The AI-generated nodes/summaries are accurate summaries of our discussion. ($* : p < 0.05$; $** : p < 0.01$; $* * * : p < 0.001$)

Qualitative insights revealed that the perceptions were due to the way MeetMap externalizes linear conversations in a structured, visual, and intuitive manner. All users agreed that the visual representation and notation schema facilitated a quick grasp of the content. P2 noted, *"It listed some icons with visual categories to show our opinion and helped me quickly get the main idea and its sub-nodes."* Most users (16/20)) found the design of MeetMap intuitive for connecting ideas over time. As P2 stated, *"We could connect information from the second agenda into the previous agenda (P2 pointed to the agenda topics on the timeline), and synthesize ideas (P2 pointed to the arrows on the Map Canvas)."* Additionally, the right amount of extra information was crucial for in-situ understanding. All users (20 out of 20) believed that a complete transcript provided excessive information in the Baseline condition; In contrast, all users (20 out of 20) preferred the concise content in the AI-generated nodes in MeetMap. P9 observed, *"The nodes took out every point. But every node is really short, so I can quickly glance at it."*

5.1.2 MeetMap fosters team consensus building and facilitates subsequent discussion. Users rated Human-Map ($p = 0.003 < 0.05$) and AI-Map ($p = 0.004 < 0.05$) significantly higher in supporting teams to achieve consensus compared with the Baseline , as shown in Figure 10:Q4).

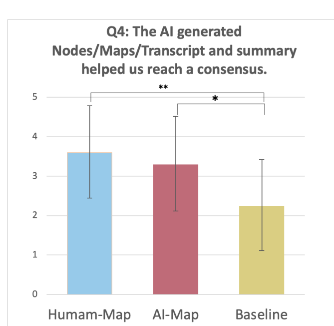

Fig. 10. **Q4: The AI-generated Nodes/Maps/transcript and summary helped us reach a consensus.** Human-Map and AI-Map show significantly higher ratings in helping people reach a consensus. The error bars represent standard deviations. ($* : p < 0.05$; $** : p < 0.01$; $* * * : p < 0.001$)

Some users (8/20) indicated MeetMap helped build shared understanding and facilitated more structured subsequent discussions among teams. The AI-generated nodes in MeetMap were perceived as objective mediators to address misunderstandings. For example, during session three, P6 noticed a node was mistakenly categorized as a "Con" instead of a "Pro" and said, *"Oh, I thought this should be a benefit of this idea rather than a negative one. Maybe we have some misunderstanding, and do we want to discuss it further?"* Users (P7, P10, P11) found it less confrontational to address misunderstandings using AI-generated nodes, *"I didn't feel there was any way to clear up a misunderstanding (in Baseline). I could say that directly, but that felt more combative than correcting a misunderstanding in the other one (in MeetMap).".* Apart from solving misunderstanding, there were 3 of 10 groups of participants who actively created nodes to guide the conversation in Human-Map ; as P2 said, *"I would try to create two top-level nodes, which helped me organize our discussion based on these ideas.".* Some users (7/20) believed that the notation schema in MeetMap helped reflection and enhance less biased discussion. As P17 said, *"In MeetMap, if I found one idea with three pros and one con, I would ask the team to consider whether there are more cons for this idea."*

5.1.3 MeetMap users have more bandwidth to create and read their shared notes during the conversation. We found that Human-Map users had significantly more note-creation behaviors than AI-Map ($p = 0.0001 < 0.001$), who also had more note-creation behaviors than the Baseline ($p = 0.015 < 0.05$), as shown in Figure 11. Besides, users had significantly more interactions to navigate and check the note content in AI-Map than that in the Baseline ($p = 0.035 < 0.05$).

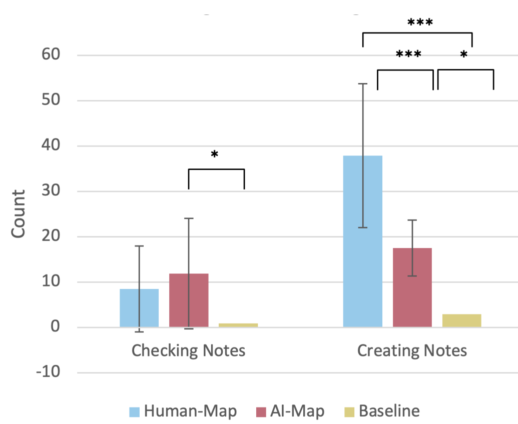

Fig. 11. **Users created more notes with MeetMap.** Users result in a significant increase in note creation and viewing in MeetMap. The error bars represent standard deviations. ($* : p < 0.05$; $** : p < 0.01$; $*** : p < 0.001$)

Creating more notes during the conversation may possibly incur a higher cognitive load on the users. However, based on our NASA-TLX survey, participants did not report a higher task load when comparing two MeetMap conditions and the Baseline , as shown in Figure 15. Despite the similar level of cognitive load in most aspects, users interacted more with the map and created collaborative notes during the two MeetMap conditions. This increased interaction likely enhanced users' ability to understand the conversation in real-time. This conclusion is echoed in the qualitative findings.

Most users (18/20) thought the synchronicity of the immediate summary nodes made taking notes in real-time possible in MeetMap conditions, like P10 said, *"If the nodes are created at the same time that we're having the conversation, my mind is still on the topic being discussed, so it was easier for me to use it."* 7 users especially appreciated the way that AI first generates the nodes and then combines them as maps in AI-Map , which reduced their uncertainty about the AI ability, as

P8 said, *"I know it is working once I see the nodes pop out after I speak, then it generates links, So I don't need to worry about missing anything. "* On the other hand, most users (14/20) were confused by the long and unpredictable waiting time of the AI summary in the baseline condition.

5.1.4 The pair of participants have more balanced contributions to the dialogue maps in MeetMap than Zoom. We analyzed how different group members interacted with the maps under each system variant condition. From Figure 12, it is evident that in both MeetMap conditions, almost every participant actively took notes, showing a trend of more balanced note-taking behavior between the two people in each group. In contrast, under the Zoom condition, each group tended to have one person dominate the collaborative note-taking.

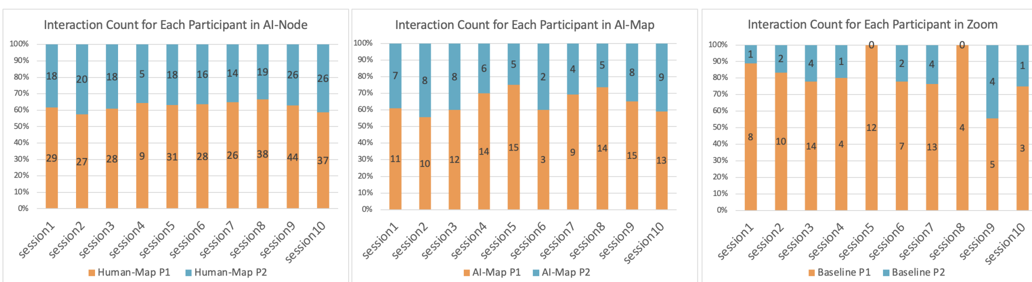

Fig. 12. **The note-taking behaviors for each user in each session.** Human-Map and AI-Map both show a trend of more balanced collaborative note-taking behaviors between the team members.

From the video analysis, we identified collaborative patterns in both MeetMap settings. Users actively engaged in note-taking and spontaneously assumed different roles. In 5 out of 10 groups, one member refined AI-generated content while another added overarching themes for a cohesive structure. All groups worked to merge their notes into a comprehensive map that captured the meeting's overall flow. In contrast, the Baseline sessions showed a different dynamic. In 4 out of 10 groups, one participant took notes while the other watched and provided verbal input. In the remaining 6 groups, both members took notes independently, typically writing in separate sections without combining their content.

Qualitative results explained the reason behind the different collaboration patterns. Users (8/20) were more likely to engage with nodes that reflected their own contributions, as indicated by the color-coded AI-generated nodes in MeetMap. This feature helped users feel ownership over the notes, encouraging active participation. P20 said, *"So I can see those ideas were said by myself, so I am the one to organize those ideas."* This level of natural ownership of the summarized content was less pronounced in the Baseline , where users were unclear about who should take notes and what they should note, as P6 said, *"I was not sure if she would write something "* Besides, users found it easier to refine AI-generated content rather than altering notes created by their peers. As P1 said, *"I didn't feel like I was changing other people's work while working on the AI-generated nodes. "* AI-generated nodes were perceived as a more neutral starting point for grounding, as stated by P13 *"It's like we're all shaping a common ground, using AI's input as blocks and then working toward an integral dialogue map together. It is less personal and more about improving the overall result."*

While we observed a more collaborative trend with MeetMap for generating shared representations, most groups (7/10) collaborated on the map without verbal communication about coordination. This sometimes led to unexpected changes in node content or map structure made by others. For instance, P12 noted, *"It got a bit chaotic when we all started dragging and organizing nodes without a clear strategy."*

5.2 How the Different Levels of AI Assistance Influenced Users' Interaction Behaviors and Attitudes Towards Creating Dialogue Maps During Meetings

In this section, we will examine the user experiences in two variants of MeetMap, Human-Map and AI-Map . Our goal is to contribute to the understanding of how much AI assistance is desirable and useful for people when they work together to create dialogue maps during discussions.

As illustrated in Figure 13, we present detailed counts of dialogue map creation behaviors of the two MeetMap variants on the left. We found that users in Human-Map interacted frequently with features for adding new content to the map, including manually adding nodes and creating new links. Users of AI-Map showed a tendency to be actively involved in reviewing and editing the content generated by the AI. This included modifying the categories of AI-generated nodes, editing the content and relationships of AI-created links, and deleting any unnecessary nodes or links. We presented counts of map-checking behaviors on the right, and the result shows that AI-Map users checked the original transcript of the node frequently by double-clicking the nodes. We'll analyze the reason behind the different usage patterns below.

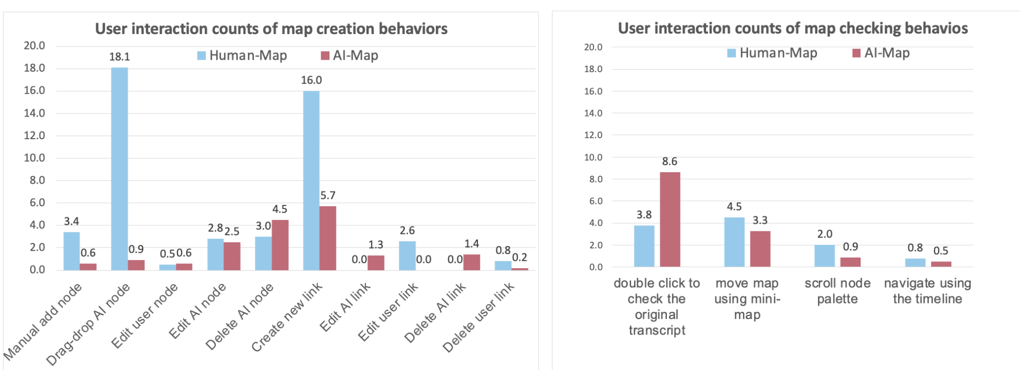

Fig. 13. **The detailed user interaction behaviors in Human-Map and AI-Map .** The average count of dialogue map creation behaviors per person **(left)** shows users in Human-Map frequently interacted with features for adding new content to the map, while users of AI-Map showed a tendency to be actively involved in editing the content generated by the AI. The average count of the dialogue map checking behaviors **(right)** showed that AI-Map users frequently check the original transcript behind an AI-generated node.

5.2.1 Human-Map provided more human agency and control that users were motivated to dialogue map in real-time . In Human-Map , most users (13/20) reported they had more agency and control over using the synchronous AI-generated nodes, as mentioned by P14, *"I saw it was generated here (Temporary Node Palette), and then it was my responsibility for moving it to the map to make it useful."* Conversely, in AI-Map , users preferred to read AI-created maps only and leave the map editing after the discussion rather than modify them in real-time. The reason for this is the delay in displaying AI-generated small maps on the Map Canvas and the uncertainty about when nodes on the Temporary Node Palette will appear as maps on the Map Canvas and what structure AI will use. P12 explained, *"I was not sure which nodes would be added to the map by AI, so I hesitated to add new nodes or modify the AI-generated map during the discussion."*

5.2.2 Users' cognitive effort in Human-Map and AI-Map came from different cognitive processes. Our findings indicate that both systems require active cognitive processing but cater to different user preferences for making sense of information. 6 users (6/20) shared that they needed to put in more cognitive effort in Human-Map to review the content in each node and think about how to organize them than they did in AI-Map , as P16 said, *"I think when the AI tool grouped everything together (AI-Map), it was easier for me just to scan the information and confirm if it was okay and that's it. But in this case (Human-Map), I had to go through the entire discussion again and connect stuff from beginning to end, so it took me longer to think, compare, and remember."* In contrast, 8 users reported that AI-Map simplifies the initial organization of content through AI-generated maps, yet users remain critically involved by refining and verifying the AI-generated outputs. P3 said, *"It was more challenging to understand if I didn't create it myself."*

5.2.3 Users wanted to engage with more cognitive and creative thinking processes even with higher effort, which cannot be replaced by AI. Despite the cognitive effort required in Human-Map , many users (12/20) preferred engaging in this intensive cognitive and creative process. These users believed that organizing the information by themselves was helpful for them to make sense of the content and led to better decision-making, even though the process of creating the dialogue maps could be cognitively demanding. P20 said, *"So it actually takes a lot of cognitive load for us to think about how to organize the nodes. But I think personally it's helpful, and we need to do that.* Most users (14/20) commented that they better understood the session and could recall the content more easily afterward. In comparison, some users (5/20) mentioned they put less effort into creating dialogue maps and found recalling the discussion and decision-making process more challenging in AI-Map , as P5 said, *"It was harder for me to remember it since it (AI) adds another logic to explain the conversation."*.

5.2.4 Users had a lower tolerance for AI mistakes when they considered themselves to own the human-AI collaborative output. Users (6/20) shared that the intensive involvement in Human-Map led to heightened scrutiny of AI outputs, as users felt more responsible for the end result. To create the map, users must first understand what AI is trying to summarize on the nodes, which inherently involves more checks, additions, and modifications, as they need to construct meanings with that node further; as P13 said, *"If it's wrong (Human-Map) and if it's not what I'm trying to express, I have to figure out how to use this one, check what it wants to say, make edits on it, or build a new one."* In contrast, most users of AI-Map , despite encountering similar inaccuracies, were more lenient. Only 2 users pointed out inaccurate content in the AI-generated nodes in AI-Map , and others were all quite liberal with the AI-generated links. P16 said, *"Since I need to be responsible for the quality of this map, and it's not the tool (Human-Map). And the other one, it was so easy that I needed to scan it, and I don't mind some errors (AI-Map)".*

5.3 Challenges of using MeetMap to support collaborative sense-making during meetings

While users appreciated MeetMap for its ability to facilitate conversation and capture meeting content, several issues emerged that warrant further consideration:

5.3.1 Notation schema need to be adaptable. Some users (3/20) found that the predefined categories of the dialogue mapping notation schema did not always suit the content of their conversations, limiting their ability to accurately represent the discussion. P15 mentioned, *"It will be good if I can add a type. Or you can provide many templates for users to choose."* Participants (3/20) need a system that can adapt to the unique dynamics of different conversations, with the ability to introduce custom categories or choose from a broader range of templates.

5.3.2 The need for control over AI granularity. Users also desired more control over the granularity of the AI-generated content.For example, some users complained about the repetition of the nodes due to either back-and-forth discussion or the imperfect performance of AI in chunking the conversation. As P9 said, *"there were some nodes which had similar ideas; it will be great if I can tell AI to include just one."* While users liked to see the visuals of the notes, they wanted to have more control over what information to be recorded and what structure to use to organize the notes. As P12 said, *"Different people map things in different ways. It will be cool if we could decide what kind of structure we want to get in this meeting".*

6 DISCUSSION

6.1 Design Implication for AI-assisted Real-time Sense-making during Video Meetings

The study's findings reveal that visual representation in MeetMap, especially the notation schema and the visual structure, enhances participants' ability to keep up with and comprehend the ongoing discussion compared to traditional text-based notes (§5.1.1.). Besides, the design of MeetMap facilitates a seamless transition between linear and visual information representations through interacting with the Temporary Node Palette and the Map Canvas. This implies the future design to support real-time sense-making in meetings could move beyond pure textual information, e.g., transcript or textual summaries [5], to a more structured and visual format, and should establish a clear connection between chronological and semantic elements to aids users in navigating and locating information effectively [40, 100].

Beyond the visual representation, supporting in-situ understanding in a highly collaborative and synchronous environment also requires careful consideration of how much cognitive load the extra information may place on users [16]. Our study shows that both Human-Map and AI-Map manage cognitive load effectively by encouraging more active interaction in creating shared notes for sense-making (§5.1.3). The synchronicity of MeetMap to provide concise summary nodes helped users have the bandwidth to consume the information and also reduced users' uncertainty and frustration compared to the baseline with longer waiting times. In AI-Map , people receive synchronous information through the Temporary Node Palette and are shown delayed as maps, suggesting that a staged presentation of information could be helpful for users to consume the information in real-time while allowing the machine to produce the result in the back-end. This middle-layer design for displaying AI-generated intermediate results in real time could be utilized in other contexts, enabling users to view content instantly while AI processes substantial data volumes, thus reducing potential delays.

Certain limitations in the current visual representations have been identified. Users have expressed a desire for a more personalized notation schema and diverse visual structures of the dialogue map (§5.3.1). Recent work showed the benefits of LLM-generated templates in helping think, organize information, and reflect on creative work [92]. To support sense-making in meetings, more structured templates can be applied to provide flexibility while also performing as a scaffold to ease the creation of shared representation and enhance understanding. As suggested by recent work CoExplorer [68], the structure of the dialogue maps could also be made adaptable to the context of the meeting and change as the discussion evolves to meet people's various needs of externalizing the verbal communication.

6.2 AI-assisted Collaborative Dialogue Mapping for Collaborative Sense-making and Constructive Discussion

Our study shows that AI-assisted collaborative dialogue mapping can effectively and less confrontationally address misunderstandings. Additionally, the AI-generated content provides a neutral basis for discussions, helping to reduce the personal biases often present in human-generated notes (§5.1.2). The AI-generated nodes were also used as guiding nodes to organize the subsequent discussion. Unlike previous approaches that positioned AI as an active mediator in group discussion [29, 47], our findings indicate that when AI-generated content is designed and presented in a non-intrusive manner, it can subtly enhance people's meta-cognitive awareness and guide behavioral changes [85], for example, motivating users to actively monitor the conversation.

Besides, creating collaborative notes with AI can facilitate balanced and inclusive collaboration dynamics (§5.1.4). Participants in our study engaged actively and shared responsibilities evenly. Participants seem to be less resistant to editing notes generated by AI than those created by their peers. This suggests employing AI-generated content as the collaboration basis can potentially promote shared ownership and reduce the perceived intrusion of modifying peers' contributions in other collaborative note-taking scenarios [24]. Additionally, participants still faced challenges in coordinating with each other in MeetMap, including unexpected edits to each other's notes. To address these issues, more explicit mechanisms for guiding collaboration should be designed. For example, displaying the edit history on the maps and incorporating features to accept or reject changes made by others could potentially increase ownership and improve coordination [25]. Building on our findings, future research could further investigate the impact of AI-assisted collaborative dialogue mapping on conversation and collaboration dynamics.

6.3 User Agency and AI Assistance in Sense-making Tasks

Our study suggests that when using AI to assist users with creating a shared representation of meetings, which is a cognitively demanding task, the goal should not be to reduce the cognitive load and manual effort to zero [44, 50]. While AI-generated summaries were well-received, users also valued the control and agency afforded by the Human-Map (§5.2.1), where they engaged more deeply with the content. Users also wanted to engage with more cognitive and creative thinking processes even with higher effort (§5.2.3). This desire suggests that when using AI to assist with cognitive tasks such as meeting understanding, while AI can handle mechanical aspects of note-taking, it should not replace the cognitive and creative processes that contribute significantly to user involvement and meeting outcomes. Consistent with previous research, our findings indicate that AI is most effective when it complements human cognitive functions rather than replacing them [44, 85]. Moreover, our study recommends offering customizable levels of AI assistance to accommodate different user preferences and cognitive styles (§5.2.2), e.g., allowing users to choose between more hands-on involvement in organizing content or focusing on refining AI-generated structures.

Last, our finding highlights the nuanced dynamics of user engagement with AI-generated content and their varying levels of trust and tolerance towards AI errors (§5.2.4). When users actively refined AI-generated content, they demanded higher accuracy and better outcomes. Moreover, users exhibited lower tolerance for AI mistakes when they felt a strong sense of ownership over the collaborative output. An important implication from these observations is that errors in AI-generated content can be tolerable if users are not required to interact extensively with it so that even imperfect AI can enhance deeper thinking and support collaboration despite potential inaccuracies. However, when significant participant involvement is needed, ensuring AI accuracy becomes paramount. In such a scenario, users preferred understanding the rationale behind the AI-generated content to fully assess the AI output. Therefore, AI needs to be introduced transparently so that users can easily understand and trust the origins of any AI-derived content [2, 89].

6.4 Limitations and Future Work

(1) While two proactivity levels of AI were considered in MeetMap, the design does not consider the full spectrum of AI assistance [2]. Future work can comprehensively evaluate the role of

AI in assisting collaborative note-taking and communication, considering a wider range of AI proactivity levels.

(2) We focused on evaluating the user experience in MeetMap through a user study and did not technically evaluate the accuracy of the AI-generated nodes/maps. Future iterations might develop adaptive algorithms, catering to each individual meeting's dynamic, to produce higher-quality dialogue maps.
(3) The current study of MeetMap focused on new teams where team members did not know each other before. This resembles the experience for many in-class project discussions and ad-hoc workplace meetings [81]. The results found in this work would apply to new and non-established teams. Additionally, we recruited the participants of the evaluation study from a mailing list at one university, and we ran the evaluation study with dyad meetings. In future studies, we hope to recruit a more diverse group of participants and examine the scalability of MeetMap in larger group settings and across different types of organizations, industries, and use cases.
(4) We recognize that additional features can add unnecessary burden. We identified the lack of adaptivity to close unnecessary panels in our system as a limitation. Future systems should allow users to minimize extra cognitive load through more adaptive design.
(5) As shown in Figure 11, both note-checking and note-creation behaviors were minimal in the baseline condition, likely due to the high pace and brief nature of the discussion. In such conditions, users may decide not to take notes if they don't perceive immediate benefits. The two MeetMap variants introduced new interaction mechanisms and AI- scaffolding to encourage collaborative output. In contrast, the baseline setup with Otter.ai transcripts and Google Docs did not provide enough motivation for note-taking. A more suitable baseline might include a built-in collaborative note editor with AI-generated text summaries, which could better validate the usefulness of the design of the collaborative dialogue maps.
(6) The current design uses a low-contrast color scheme, which may not be appropriate for all accessibility requirements. In the future, MeetMap will improve accessibility and visual design by rethinking the color scheme and resizing interactive buttons. These adjustments will help make the system more inclusive and user-friendly.

7 CONCLUSION

Traditional video meeting platforms present discussions linearly, either as transcripts or summaries, while conversation ideas often emerge non-linearly. We explored LLM-assisted real-time collaborative dialogue map generation, which visually represents structured and interconnected ideas. To balance reducing user cognitive load and granting user control over AI-generated content, we introduced two human-AI collaboration approaches: Human-Map and AI-Map . In Human-Map , AI summarizes conversations into nodes, with users linking the nodes to shape a dialogue map. AI-Map allows the AI to create small maps first, which users can subsequently refine. We evaluate these methods through a within-subject study involving ten user pairs. Users preferred MeetMap over conventional note-taking strategies since it reflected the conversation process and aligned with how humans organize information. Users liked the ease of use for AI-Map due to the low effort demands and appreciated the hands-on opportunity in Human-Map for sense-making This study provides insights on enhancing human-AI collaboration in note-taking and sense-making during meetings and provides design implications for involving AI in synchronous human communication activities.

REFERENCES

[1] Joseph A. Allen, Tammy Beck, Cliff W. Scott, and Steven G. Rogelberg. [n.d.]. Understanding Workplace Meetings: A Qualitative Taxonomy of Meeting Purposes. 37, 9 ([n. d.]), 791–814. https://doi.org/10.1108/MRR-03-2013-0067

[2] 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.

[3] Elliot Aronson et al. 1978. *The jigsaw classroom.* Sage.

[4] Bon Adriel Aseniero, Marios Constantinides, Sagar Joglekar, Ke Zhou, and Daniele Quercia. 2020. MeetCues: Supporting Online Meetings Experience. *arXiv preprint arXiv:2010.06259* (2020).

[5] Sumit Asthana, Sagih Hilleli, Pengcheng He, and Aaron Halfaker. 2023. Summaries, Highlights, and Action items: Design, implementation and evaluation of an LLM-powered meeting recap system. *arXiv preprint arXiv:2307.15793* (2023).

[6] Jeremy N. Bailenson. [n.d.]. Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue. 2, 1 ([n. d.]). https://doi.org/10.1037/tmb0000030

[7] Satanjeev Banerjee, Carolyn Rose, and Alexander I. Rudnicky. [n.d.]. The Necessity of a Meeting Recording and Playback System, and the Benefit of Topic–Level Annotations to Meeting Browsing. In *Human-Computer Interaction - INTERACT 2005* (Berlin, Heidelberg, 2005) *()*, Maria Francesca Costabile and Fabio Paternò (Eds.). Springer, 643–656. https://doi.org/10.1007/11555261_52

[8] Jerrold E. Barnett, Francis J. di Vesta, and James T. Rogozinski. 1981. What Is Learned in Note Taking. *Journal of Educational Psychology* 73 (1981), 181–192. https://api.semanticscholar.org/CorpusID:143614892

[9] Jeremy P Birnholtz, Thomas A Finholt, Daniel B Horn, and Sung Joo Bae. 2005. Grounding needs: achieving common ground via lightweight chat in large, distributed, ad-hoc groups. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 21–30.

[10] Antje Bothin and Paul Clough. [n.d.]. A User Evaluation Study: Do Participants' Personal Notes Help Us to Summarise Meetings? 21, 2 ([n. d.]), 122–133. https://doi.org/10.1002/kpm.1439

[11] Susan E Brennan. 1998. The grounding problem in conversations with and through computers. *Social and cognitive approaches to interpersonal communication* (1998), 201–225.

[12] Alex Cao, Keshav K Chintamani, Abhilash K Pandya, and R Darin Ellis. 2009. NASA TLX: Software for assessing subjective mental workload. *Behavior research methods* 41 (2009), 113–117.

[13] Hancheng Cao, Chia-Jung Lee, Shamsi Iqbal, Mary Czerwinski, Priscilla Wong, Sean Rintel, Brent Hecht, Jaime Teevan, and Longqi Yang. 2021. Large Scale Analysis of Multitasking Behavior During Remote Meetings. arXiv:2101.11865 [cs.CY]

[14] Yining Cao, Hariharan Subramonyam, and Eytan Adar. 2022. VideoSticker: A Tool for Active Viewing and Visual Note-taking from Videos. In *27th International Conference on Intelligent User Interfaces*. 672–690.

[15] Senthil Chandrasegaran, Chris Bryan, Hidekazu Shidara, Tung-Yen Chuang, and Kwan-Liu Ma. 2019. TalkTraces: real-time capture and visualization of verbal content in meetings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[16] Xinyue Chen, Shuo Li, Shipeng Liu, Robin Fowler, and Xu Wang. 2023. MeetScript: Designing Transcript-based Interactions to Support Active Participation in Group Video Meetings. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–32. https://doi.org/10.1145/3610196

[17] Herbert H Clark and Susan E Brennan. 1991. Grounding in communication. (1991).

[18] Herbert H Clark and Edward F Schaefer. 1989. Contributing to discourse. *Cognitive science* 13, 2 (1989), 259–294.

[19] Jeff Conklin. [n.d.]. Dialog Mapping: Reflections on an Industrial Strength Case Study. In *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, Paul A. Kirschner, Simon J. Buckingham Shum, and Chad S. Carr (Eds.). Springer, 117–136. https://doi.org/10.1007/978-1-4471-0037-9_6

[20] Jeff Conklin. 2005. *Dialogue mapping: Building shared understanding of wicked problems.* John Wiley & Sons, Inc.

[21] Jeff Conklin, Albert Selvin, Simon Buckingham Shum, and Maarten Sierhuis. 2001. Facilitated hypertext for collective sensemaking: 15 years on from gIBIS. In *Proceedings of the 12th ACM conference on Hypertext and Hypermedia*. 123–124.

[22] Gregorio Convertino, Helena M Mentis, Mary Beth Rosson, John M Carroll, Aleksandra Slavkovic, and Craig H Ganoe. 2008. Articulating common ground in cooperative work: content and process. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1637–1646.

[23] Jamie Costley and Mik Fanguy. 2021. Collaborative note-taking affects cognitive load: the interplay of completeness and interaction. *Educational Technology Research and Development* 69 (2021), 655–671.

[24] Matthew Courtney, Jamie Costley, Matthew Baldwin, Kyungmee Lee, and Mik Fanguy. 2022. Individual versus collaborative note-taking: Results of a quasi-experimental study on student note completeness, test performance, and academic writing. *Internet High. Educ.* 55 (2022), 100873. https://api.semanticscholar.org/CorpusID:250229395

[25] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. 2012. Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 1277–1286.
[26] Richard C Davis, Jason A Brotherton, James A Landay, Morgan N Price, and Bill N Schilit. 1998. NotePals: Lightweight Note Taking by the Group, for the Group. *University of California, Berkeley, Computer Science Division* 8 (1998).
[27] Natasha Dhawan, Molly Carnes, Angela Byars-Winston, and Narjust Duma. 2021. Videoconferencing etiquette: promoting gender equity during virtual meetings. *Journal of Women's Health* 30, 4 (2021), 460–465.
[28] Paramveer S Dhillon, Somayeh Molaei, Jiaqi Li, Maximilian Golub, Shaochun Zheng, and Lionel Peter Robert. 2024. Shaping Human-AI Collaboration: Varied Scaffolding Levels in Co-writing with Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.
[29] Hyo Jin Do, Ha-Kyung Kong, Pooja Tetali, Jaewook Lee, and Brian P Bailey. 2023. To Err is AI: Imperfect Interventions and Repair in a Conversational Agent Facilitating Group Chat Discussions. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–23.
[30] Gregory Dyke and Kristine Lund. 2013. Student Strategies for Collaborative Note-Taking and the Influence of FloorControl. (2013).
[31] 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*. 73–81.
[32] Jingchao Fang, Yanhao Wang, Chi-Lan Yang, Ching Liu, and Hao-Chuan Wang. 2022. Understanding the Effects of Structured Note-taking Systems for Video-based Learners in Individual and Social Learning Contexts. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–21.
[33] Jingchao Fang, Yanhao Wang, Chi-Lan Yang, and Hao-Chuan Wang. 2021. NoteCoStruct: Powering Online Learners with Socially Scaffolded Note Taking and Sharing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–5.
[34] Mik Fanguy, Jamie Costley, Matthew Courtney, and Kyungmee Lee. 2023. Analyzing collaborative note-taking behaviors and their relationship with student learning through the collaborative encoding-storage paradigm. *Interactive Learning Environments* (2023), 1–15.
[35] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 2020. AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication* 25, 1 (2020), 89–100.
[36] Huei-Tse Hou and Sheng-Yi Wu. 2011. Analyzing the social knowledge construction behavioral patterns of an online synchronous collaborative discussion instructional activity using an instant messaging tool: A case study. *Computers & Education* 57, 2 (2011), 1459–1468.
[37] Nick Van Isterdael. [n.d.]. Dialogue Mapping for GuideaMaps. ([n. d.]).
[38] Rs Renée Jansen, Daniel Lakens, and Wa Wijnand IJsselsteijn. 2017. An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review* 22 (2017), 223–233. https://api.semanticscholar.org/CorpusID:148791987
[39] Peiling Jiang, Jude Rayan, Steven P. Dow, and Haijun Xia. [n.d.]. *Graphologue: Exploring Large Language Model Responses with Interactive Diagrams*. arXiv:2305.11473 [cs] http://arxiv.org/abs/2305.11473
[40] Peiling Jiang, Jude Rayan, Steven W. Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. *ArXiv* abs/2305.11473 (2023). https://api.semanticscholar.org/CorpusID:258823121
[41] Sasa Junuzovic, Kori Inkpen, Rajesh Hegde, Zhengyou Zhang, John Tang, and Christopher Brooks. 2011. What did I miss? In-meeting review using multimodal accelerated instant replay (AIR) conferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 513–522.
[42] Kowe Kadoma, Marianne Aubin Le Quere, Xiyu Jenny Fu, Christin Munsch, Danaë Metaxa, and Mor Naaman. 2024. The Role of Inclusion, Control, and Ownership in Workplace AI-Mediated Communication. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–10.
[43] Matthew Kam, Jingtao Wang, Alastair Iles, Eric Tse, Jane Chiu, Daniel Glaser, Orna Tarshish, and John Canny. 2005. Livenotes: a system for cooperative and augmented note-taking in lectures. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 531–540.
[44] Hyeonsu B Kang, Sherry Tongshuang Wu, Joseph Chee Chang, and Aniket Kittur. [n.d.]. Synergi: A Mixed-Initiative System for Scholarly Synthesis and Sensemaking. ([n. d.]).
[45] Katherine A Karl, Joy V Peluchette, and Navid Aghakhani. 2021. Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly. *Small Group Research* (2021), 10464964211015286.
[46] Kenneth A Kiewra. 1989. A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review* 1 (1989), 147–172.

[47] 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*. 1–13.
[48] 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.
[49] Nurit Kirshenbaum, Kylie Davidson, Jesse Harden, Chris North, Dylan Kobayashi, Ryan Theriot, Roderick S. Tabalba, Michael L. Rogers, Mahdi Belcaid, Andrew T. Burks, Krishna N. Bharadwaj, Luc Renambot, Andrew E. Johnson, Lance Long, and Jason Leigh. [n.d.]. Traces of Time through Space: Advantages of Creating Complex Canvases in Collaborative Meetings. 5 ([n. d.]), 1–20. Issue ISS. https://doi.org/10.1145/3488552
[50] Kenneth R Koedinger and Vincent Aleven. 2007. Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review* 19 (2007), 239–264.
[51] Reneta D Lansiquot and Candido Cabo. 2015. Concept Mapping Narratives to Promote CSCL and Interdisciplinary Studies. International Society of the Learning Sciences, Inc.[ISLS].
[52] Gilly Leshed, Diego Perez, Jeffrey T. Hancock, Dan Cosley, Jeremy Birnholtz, Soyoung Lee, Poppy L. McLeod, and Geri Gay. [n.d.]. Visualizing Real-Time Language-Based Feedback on Teamwork Behavior in Computer-Mediated Groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston MA USA, 2009-04-04). ACM, 537–546. https://doi.org/10.1145/1518701.1518784
[53] Claire Liang, Julia Proft, Erik Andersen, and Ross A. Knepper. [n.d.]. Implicit Communication of Actionable Information in Human-AI Teams. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow Scotland Uk, 2019-05-02). ACM, 1–13. https://doi.org/10.1145/3290605.3300325
[54] Jingxian Liao, Mrinalini Singh, and Hao-Chuan Wang. 2023. DeepThinkingMap: Collaborative Video Reflection System with Graph-based Summarizing and Commenting. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 369–371.
[55] Hajin Lim. [n.d.]. Designing to Support Sensemaking in Cross-Lingual Computer-Mediated Communication Using NLP Techniques. https://www.proquest.com/docview/2550610767/abstract/5CFA5E0A7FDA45E3PQ/1
[56] Ching Liu, Juho Kim, and Hao-Chuan Wang. 2018. ConceptScape: Collaborative concept mapping for video learning. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.
[57] 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*. 1–20.
[58] Xinyi Lu, Simin Fan, Jessica Houghton, Lu Wang, and Xu Wang. 2023. ReadingQuizMaker: A Human-NLP Collaborative System that Supports Instructors to Design High-Quality Reading Quiz Questions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.
[59] Andrés Lucero. 2015. Using affinity diagrams to evaluate interactive prototypes. In *IFIP Conference on Human-Computer Interaction*. Springer, 231–248.
[60] Narges Mahyar, Ali Sarvghad, and Melanie Tory. 2012. Note-taking in co-located collaborative visual analytics: Analysis of an observational study. *Information Visualization* 11, 3 (2012), 190–204.
[61] Tamas Makany, Jonathan Kemp, and Itiel E Dror. 2009. Optimising the use of note-taking as an external cognitive aid for increasing learning. *British Journal of Educational Technology* 40, 4 (2009), 619–635.
[62] Microsoft. 2023. https://www.microsoft.com/en-us/worklab/work-trend-index/brain-research
[63] Fanguy Mik et al. 2019. The effects of collaborative note-taking in flipped learning contexts. *Journal of Language and Education* 5, 4 (20) (2019), 25–35.
[64] Prasanth Murali, Javier Hernandez, Daniel McDuff, Kael Rowan, Jina Suh, and Mary Czerwinski. 2021. Affectivespotlight: Facilitating the communication of affective responses from audience members during online presentations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.
[65] Gabriel Murray. [n.d.]. Information Processing and Overload in Group Conversation: A Graph-Based Prediction Model. 3, 3 ([n. d.]), 46. Issue 3. https://doi.org/10.3390/mti3030046
[66] Fung Fai Ng. 2008. Dialogue Mapping and Collaborative Learning. In *Handbook of Conversation Design for Instructional Applications*. IGI Global, 403–418.
[67] Otter.ai. 2023. https://help.otter.ai/hc/en-us/articles/5093383818263-Automated-Live-Summary-Overview
[68] Gun Woo 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*. 1638–1657.
[69] Annie Piolat, Thierry Olive, and Ronald T Kellogg. 2005. Cognitive effort during note taking. *Applied cognitive psychology* 19, 3 (2005), 291–312.

[70] Yan Qu and Derek L Hansen. 2008. Building shared understanding in collaborative sensemaking. In *Proceedings of CHI 2008 Sensemaking Workshop*.
[71] Mark Reilly and Haifeng Shen. 2011. GroupNotes: encouraging proactive student engagement in lectures through collaborative note-taking on smartphones. (2011).
[72] Heather Richter, Gregory D Abowd, Werner Geyer, Ludwin Fuchs, Shahrokh Daijavad, and Steven Poltrock. 2001. Integrating meeting capture within a collaborative team environment. In *Ubicomp 2001: Ubiquitous Computing: International Conference Atlanta Georgia, USA, September 30–October 2, 2001 Proceedings 3*. Springer, 123–138.
[73] 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 amp; Inclusive Meetings. In *CHI 2021*. https://www.microsoft.com/en-us/research/publication/meetingcoach-an-intelligent-dashboard-for-supporting-effective-inclusive-meetings/
[74] Samiha Samrose, Ru Zhao, Jeffery White, Vivian Li, Luis Nova, Yichen Lu, Mohammad Rafayet Ali, and Mohammed Ehsan Hoque. 2018. Coco: Collaboration coach for understanding team dynamics during video conferencing. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 4 (2018), 1–24.
[75] Daniel L Schwartz, Jessica M Tsang, and Kristen P Blair. 2016. *The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them.* WW Norton & Company.
[76] Lucas M Seuren, Joseph Wherton, Trisha Greenhalgh, and Sara E Shaw. 2021. Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of pragmatics* 172 (2021), 63–78.
[77] Yang Shi, Chris Bryan, Sridatt Bhamidipati, Ying Zhao, Yaoxue Zhang, and Kwan-Liu Ma. 2018. Meetingvis: Visual narratives to assist in recalling meeting context and content. *IEEE Transactions on Visualization and Computer Graphics* 24, 6 (2018), 1918–1929.
[78] Patrick C Shih, David H Nguyen, Sen H Hirano, David F Redmiles, and Gillian R Hayes. 2009. GroupMind: supporting idea generation through a collaborative mind-mapping tool. In *Proceedings of the 2009 ACM International Conference on Supporting Group Work*. 139–148.
[79] Joongi Shin, Michael A. Hedderich, AndréS Lucero, and Antti Oulasvirta. [n.d.]. Chatbots Facilitating Consensus-Building in Asynchronous Co-Design. In *The 35th Annual ACM Symposium on User Interface Software and Technology* (Bend OR USA, 2022-10-29). ACM, 1–13. https://doi.org/10.1145/3526113.3545671
[80] Seoyun Son, Junyoug Choi, Sunjae Lee, Jean Y Song, and Insik Shin. [n.d.]. It Is Okay to Be Distracted: How Real-time Transcriptions Facilitate Online Meeting with Distraction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg Germany, 2023-04-19). ACM, 1–19. https://doi.org/10.1145/3544548.3580742
[81] Angela EB Stewart, Mary Jean Amon, Nicholas D Duran, and Sidney K D'Mello. 2020. Beyond team makeup: Diversity in teams predicts valued outcomes in Computer-Mediated collaborations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[82] Susan G Straus. 1996. Getting a clue: The effects of communication media and information distribution on participation and performance in computer-mediated and face-to-face groups. *Small group research* 27, 1 (1996), 115–142.
[83] Sangho Suh, Bryan Min, Srishti Palani, and Haijun Xia. 2023. Sensecape: Enabling Multilevel Exploration and Sensemaking with Large Language Models. *ArXiv* abs/2305.11483 (2023). https://api.semanticscholar.org/CorpusID:258822925
[84] Meng Sun, Minhong Wang, Rupert Wegerif, and Jun Peng. 2022. How do students generate ideas together in scientific creativity tasks through computer-based mind mapping? *Computers & Education* 176 (2022), 104359.
[85] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The metacognitive demands and opportunities of generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–24.
[86] Johanna Telenius et al. 2016. Sensemaking in Meetings-Collaborative Construction of Meaning and Decisions through Epistemic Authority. (2016).
[87] Hao-Chuan Wang, Dan Cosley, and Susan R. Fussell. [n.d.]. Idea Expander: Supporting Group Brainstorming with Conversationally Triggered Visual Thinking Stimuli. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA, 2010-02-06) *(CSCW '10)*. Association for Computing Machinery, 103–106. https://doi.org/10.1145/1718918.1718938
[88] Rupert Wegerif, Bruce M. McLaren, Marian Chamrada, Oliver Scheuer, Nasser Mansour, Jan Mikšátko, and Mriga Williams. [n.d.]. Exploring Creative Thinking in Graphically Mediated Synchronous Dialogues. 54, 3 ([n. d.]), 613–621. https://doi.org/10.1016/j.compedu.2009.10.015
[89] Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie J Cai. [n.d.]. PromptChainer: Chaining Large Language Model Prompts through Visual Programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (New Orleans LA USA, 2022-04-27). ACM, 1–10. https://doi.org/10.1145/3491101.3519729

[90] Haijun Xia, Tony Wang, Aditya Gunturu, Peiling Jiang, William Duan, and Xiaoshuo Yao. [n.d.]. CrossTalk: Intelligent Substrates for Language-Oriented Interaction in Video-Based Communication and Collaboration. ([n. d.]).

[91] Ruochen Xu, Chenguang Zhu, and Michael Zeng. 2022. Narrate Dialogues for Better Summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2022*. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 3565–3575. https://doi.org/10.18653/v1/2022.findings-emnlp.261

[92] Xiaotong Xu, Jiayu Yin, Catherine Gu, Jenny Mar, Sydney Zhang, Jane L E, and Steven P Dow. 2024. Jamplate: Exploring LLM-Enhanced Templates for Idea Reflection. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 907–921.

[93] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.

[94] Fangyu Yu, Peng Zhang, Xianghua Ding, Tun Lu, and Ning Gu. [n.d.]. Exploring How Workspace Awareness Cues Affect Distributed Meeting Outcome. 0, 0 ([n. d.]), 1–20. https://doi.org/10.1080/10447318.2022.2064063

[95] Amy X Zhang and Justin Cranshaw. 2018. Making sense of group chat through collaborative tagging and summarization. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–27.

[96] Amy X Zhang, Lea Verou, and David Karger. 2017. Wikum: Bridging discussion forums and wikis using recursive summarization. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2082–2096.

[97] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. 2021. " An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25.

[98] Yongle Zhang, Dennis Asamoah Owusu, Marine Carpuat, and Ge Gao. [n.d.]. Facilitating Global Team Meetings Between Language-Based Subgroups: When and How Can Machine Translation Help? 6 ([n. d.]), 90:1–90:26. Issue CSCW1. https://doi.org/10.1145/3512937

[99] Zheng Zhang, Jie Gao, Ranjodh Singh Dhaliwal, and Toby Jia-Jun Li. [n.d.]. *VISAR: A Human-AI Argumentative Writing Assistant with Visual Programming and Rapid Draft Prototyping*. arXiv:2304.07810 [cs] http://arxiv.org/abs/2304.07810

[100] Zheng Zhang, Weirui Peng, Xinyue Chen, Luke Cao, and Toby Jia-Jun Li. 2024. LADICA: A Large Shared Display Interface for Generative AI Cognitive Assistance in Co-Located Team Collaboration. *arXiv preprint arXiv:2409.13968* (2024).

[101] Chengbo Zheng, Yuheng Wu, Chuhan Shi, Shuai Ma, Jiehui Luo, and Xiaojuan Ma. 2023. Competent but Rigid: Identifying the Gap in Empowering AI to Participate Equally in Group Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[102] Rosa Zubizarreta. 2013. Co-creative dialogue for meeting practical challenges. *OD Practitioner* 45, 1 (2013), 47–53.

A APPENDIX: SURVEY QUESTIONS AND STATISTICAL RESULTS

			AI-Map		Human-Map		Baseline			
	Question		Mean	Std	Mean	Std	Mean	Std	Chi-Square	p-value
Usability questions	Q1	The dialogue maps (AI transcript/summary) helped me keep up with the discussion content in real-time.	2.95	1.27	3.05	1.19	2.2	1.23	7.13	0.02*
	Q2	The dialogue map (AI transcript/summary) reflects our decision-making process accurately.	4.00	0.92	4.05	1.10	3.04	1.30	14.23	0.0008***
	Q3	The AI-generated nodes (otter.ai summary) are accurate summaries of our discussion.	3.95	0.89	3.45	1.23	2.91	1.38	8.67	0.01*
	Q4	The dialogue map (AI transcript/summary) helped us reach a consensus.	3.30	1.22	3.60	1.19	2.57	1.38	12.42	0.002**
	Q5	Creating the dialogue maps (notes) is cognitively challenging.	2.50	1.15	3.15	1.27	2.78	1.44	5.29	0.07
	Q6	I enjoy creating dialogue maps (creating notes with AI summaries) more than taking notes by myself.	3.70	1.30	3.65	1.46	3.61	1.16	0.60	0.74
NASA NLX test	Q7-1	Mental Demand: How mentally demanding was the task?	6.80	4.19	8.85	4.79	8.04	4.17	1.65	0.43
	Q7_2	Physical Demand: How physically demanding was the task?	4.05	3.61	4.85	4.75	4.26	3.39	0.95	0.62
	Q7_3	Temporal Demand: How hurried or rushed was the pace of the task?	4.60	3.17	7.85	6.41	7.43	4.30	3.86	0.14
	Q7_4	Performance: How successful were you in accomplishing what you were asked to do?	13.40	5.15	13.90	6.05	14.17	3.23	1.77	0.41
	Q7_5	Effort: How hard did you have to work to accomplish your level of performance?	6.75	4.24	8.35	4.46	8.43	3.93	1.71	0.42
	Q7_6	Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?	5.55	4.86	6.70	5.40	5.48	5.02	1.85	0.40

Fig. 14. **Descriptive and statistic analysis of the survey questions** ($*: p < 0.05$; $**: p < 0.01$; $***: p < 0.001$)

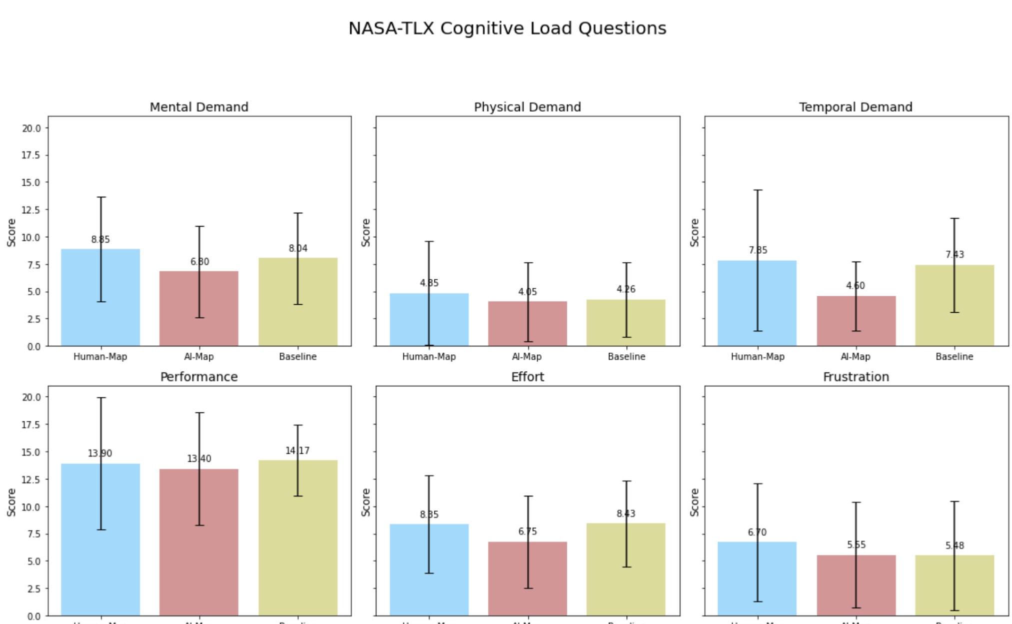

Fig. 15. Cognitive load across the conditions ($*: p < 0.05$; $**: p < 0.01$; $***: p < 0.001$)

B APPENDIX: SEMI-STRUCTURED INTERVIEW QUESTIONS

Questions after Each Condition (Human-Map , AI-Map , Baseline)

(1) Usefulness of AI aids
  (a) Was the AI aids helpful for you? How? Why or why not?
  (b) Were there moments when the AI aids helped you understand the discussion better? Can you tell me more about it?
  (c) How important is it that these AI aids are shown to you in real-time?

(2) Cognitive Load During Discussion with Real-Time Aids
  (a) Did you read the temporary AI aids during the discussion? If yes, how did that influence your discussion?
  (b) Did you refer back to the AI aids during the discussion? If yes, how did that influence your subsequent discussion?

(3) Thoughts on the Break
  (a) How did you distribute your time between creating/reading the AI aids during the discussion and the break?
  (b) Would you prefer to interact with the AI aids during the discussion or during the break? Why?

(4) Use of AI aids in the Reflection Period
  (a) Did you use the AI aids when answering the survey? How did you use them?

(5) Quality of AI aids
  (a) What are your thoughts on the quality of the resulting AI aids?
  (b) What are your thoughts on the quality of the AI-generated nodes (if applicable)?

(6) Challenges
  (a) Can you share some challenges you encountered during the discussion?

Comparison Between the Two Dialogue Mapping Conditions (Human-Map and AI-Map)

(1) Comparison of Cognitive Demands
  (a) Can you share the cognitive demands on you for both dialogue mapping methods (Human-Map and AI-Map)?

(2) Comparison of Quality of Dialogue Maps
  (a) Can you compare the quality of the resulting dialogue maps in both methods?

(3) Trustworthiness of the Methods
  (a) Can you share your thoughts on the trustworthiness of the two dialogue mapping methods?

(4) Preference for Future Meetings
  (a) Which dialogue mapping method (Human-Map or AI-Map) would you prefer for a meeting in the future? Why?

Comparison Between All Three Conditions (Human-Map , AI-Map , and Baseline)

(1) Cognitive Demands Across Conditions
  (a) When you were creating the dialogue maps (Human-Map , AI-Map) versus taking notes yourself (Baseline), which one was more cognitively demanding for you?

(2) Helpfulness for Visualizing Discussion Progress and Decision-Making
  (a) Comparing the resulting dialogue maps (Human-Map , AI-Map) with the AI summaries and transcripts (Baseline), which was more helpful for you to visualize the discussion progress and make decisions?

(3) Quality of Summaries and Dialogue Maps

   (a) Can you comment on the quality of the summaries (Baseline) and the quality of the dialogue maps (Human-Map , AI-Map)?
(4) Preference for Future Meetings
   (a) Which method (Human-Map , AI-Map , or Baseline) would you prefer for a meeting in the future? Why?

C APPENDIX: TASKS FOR USER STUDY

Each task was presented to participants in dyads, with each participant viewing content tailored to their assigned persona. The specific tasks, agendas, and personas for each user are provided below.

C.1 Task 1 - Enhancing Mental Health Services on Campus

Task Description: Student representatives are tasked to discuss and formulate strategies to improve mental health and well-being support on campus.

Agenda 1: One strategy is to require first-year students to pay a 15-min visit to the University Counseling Psychological Services (CAPS) office as a prerequisite for class registration. Please discuss whether you agree with this strategy and reach a consensus on what is the best way to go:

- **Persona 1: You support this strategy** - There is an increase in reported student anxiety, stress, and other mental health issues. But many students don't seek help from CAPS until the problems become serious. - A required visit to CAPS can help students get familiar with the services they provide.

- **Persona 2: You do not support the requirement** - If this becomes a requirement, it will be a big load for the CAPS office and they need to increase staff. - Students who do not need support may find this to be a waste of time.

Agenda 2: Given the diverse student population, what kind of mental health services would you consider essential? Discuss their significance and decide which one should be prioritized. Please propose your own solutions. If you run out of ideas, here are some suggestions.

- **24/7 Helpline with student volunteers:**
  - + Provides instant support during crises.
  - + Can be accessed anonymously.
  - - Needs training for student volunteers to provide counseling services.
  - - Student volunteers may not be as effective as professional counselors.

C.2 Task 2 - Installing Smart Devices for University Buildings

Task Description: Student representatives are tasked to discuss how to install smart devices for a new university building (e.g., facial recognition devices for building entrances, smart lighting systems, etc). Your goal now is to share your perspectives on what smart devices to install and reach a consensus.

Agenda 1: To support easy entrance into buildings, what are some smart techniques that you may consider? Try to generate as many ideas as possible. Please propose your own solutions. If you run out of ideas, here are some suggestions.

- **Persona 1: Budget-focused perspective**

- Scanner at entrance that scans QR code on phone for identification
  - + Students will most likely have phones with them
  - + Cheap and easy (no energy cost)
  - - Insecure, people can share QR codes with anybody
  - - Phones could be out of battery

- **Persona 2: Technology proponent perspective**

- Facial recognition system using computer vision
  - + Students don't need to bring anything with them
  - + Quick and easy
  - - Expensive
  - - AI errors

Agenda 2: To support occupancy sensing, so that students can know which classrooms and spaces are empty from a dashboard, what are some smart devices that you may consider? Try to generate as many ideas as possible. Please propose your own solutions. If you run out of ideas, here are some suggestions.

- **Persona 1: Environmental perspective**
  - Occupancy sensors that detect motions in a room
    - + Cheap
    - + Less privacy concerns
    - - Less information, e.g., don't know how many people
    - - May send errors when there are people but no movements
- **Persona 2: Technology proponent perspective**
  - Facial recognition system using computer vision
    - + Provides more information, e.g., how many people are in a space
    - + Robust to many contexts, classrooms, cafes, etc.
    - - Privacy concerns
    - - Expensive

C.3 Task 3 - Reevaluating Attendance Checking in University Classes

Task Description: Student and faculty representatives are tasked to discuss the merits and concerns about attendance checking in classes at the University. You will each represent a student representative and a faculty representative. Your goal is to reach a consensus on the best way for attendance checking that every stakeholder is happy with.

Agenda 1: Please reach a consensus on whether in-person attendance is necessary or not.

- **Persona 1: Instructor Representative**

- In-person attendance will increase student engagement.
- Enhanced peer interactions and community-building.
- Higher learning benefits since students will be more active (e.g., more QA) during in-person classes.

- **Persona 2: Student Representative**

- Difficulties of transportation between north and central campuses.
- Daytime jobs or other responsibilities that conflict with class timings.
- Enrollment in consecutive or overlapping classes.

Agenda 2: What are alternative ways for participation and attendance checking? Please reach a consensus on one desirable way to check for student participation and attendance. Please propose your own solutions. If you run out of ideas, here are some suggestions.

- **Both personas may consider:**
  - Give students several missed opportunities, e.g., they can miss three in-person classes. Otherwise, it is required.
  - Students can answer a survey within a time window (e.g., within 48 hours), so that they can watch the lecture recording after the lecture time.

D APPENDIX: DEMOGRAPHIC INFORMATION OF PARTICIPANTS

ID	Gender	Age	Ethnicity	School Year	Online Meetings/Week
P1	Male	21 - 30	Asian	Master’s student	5-10
P2	Female	21 - 30	White	Master’s student	>10
P3	Female	21 - 30	Asian	Master’s student	5-10
P4	Male	10 - 20	White	Senior	>10
P5	Female	30 - 40	White	Master’s student	1-4
P6	Female	21 - 30	Asian	Master’s student	1-4
P7	Male	21 - 30	Black	Ph.D. student	5-10
P8	Female	21 - 30	Asian	Junior	1-4
P9	Male	10 - 20	Asian	Sophomore	1-4
P10	Female	21 - 30	White	Master’s student	1-4
P11	Male	21 - 30	Black	Master’s student	>10
P12	Male	21 - 30	White	Master’s student	1-4
P13	Female	21 - 30	Black American	Master’s student	1-4
P14	Male	21 - 30	Black	Master’s student	1-4
P15	Female	21 - 30	Black American	Master’s student	1-4
P16	Male	21 - 30	Black American	Ph.D. student	>10
P17	Male	21 - 30	Middle Eastern	Master’s student	5-10
P18	Female	21 - 30	African American	Senior	1-4
P19	Female	21 - 30	White	Master’s student	1-4
P20	Female	31 - 40	White	Ph.D. student	5-10

Table 1. Demographic Information of Participants

E APPENDIX: LLM PROMPTS

E.1 Topic Segmentation

Your role is a conversational topic analyzer. Given a meeting dialogue, identify and categorize the topics. You’ll get the most recent exchange and an existing topic list. The topic list will be vacant at first. For each new exchange:

- 1. Determine the nature of each exchange in the dialogue: Is it a ’Continuation’ (Tagged as $C-Continuation) of the previous topic, Or is it a ’New Topic’ (Tagged as $N-New)?
- 2. If it is a continuation, modify the last topic to accommodate for the new turn. The updated topic should not be over six words), and the tag should be either $C-Continuation or $N-New. Your output should strictly be in the following JSON format with no additional text or explanations outside of it:

{ "Identified topic": "", "Continuation/New Topic Tag": "", }

E.2 Dialogue Tagging and Summarization

You are a facilitator trying to generate dialogue mapping for the conversation. The user will input one turn, and you will analyze the data following the steps below:

- 1. Assign dialogue mapping tags to dialogue chunks according to the dialogue mapping schema: [$Question], [$Position],[$Pro], [$Con]. It captures Questions (or Issues): What are we trying to solve? Ideas (or Responses): Potential answers or solutions. Arguments: Pros and cons for each idea.

- 2. Assign a dialogue mapping tag like [$Question] and find the corresponding sentences shown in one turn, then summarize the content under six words. Please note that each turn might have one or more tags connected to different sentences, or there is no tag if the turn is about backchanneling, greetings, or talking about some non-argumentive stuff. Please ignore the conversation that is totally off-topic to the decision-making tasks people are working on. Please make sure under each dialogue tag; the summarized content should not be able to divide into smaller pieces. For example, if the summarized content is "discuss two devices, automatic lighting, the cardless entry, "it should not be divided into one question tag with two idea tags, each one indicating one device.

The final output should only be a JSON string, and please do not provide any other text or explanation outside of the JSON format:
{ "dialogueTagArray":[{"Tag": "[$Question1]", "Summary": "Invitation to discuss products", "Quotes": "Does anyone want to talk about their products?"},], }

E.3 Link Identification

You are a facilitator trying to generate dialogue mapping for the conversation. The user will input a list of nodes with dialogue tags. The inputed nodes were assigned with a key and please do not change it. You will analyze the data following the steps below:

- 1. Please identify relationships between nodes in the nodes list, such as [$Positions] answering [$Question], [$Pros] supporting [$Position]. And generate a comprehensive hierarchy node structure for the conversation chunk.
- 2. The link should be in a single direction; this means one node should only be one time as a "from" node, but others can link one node many times.

The final output should only include a JSON data structure of the links like the following, and please do not provide any other text or explanation outside of the JSON format:
{ "linkDataArray": [{ "from": 2, "to": 1, "text": "Support"}] }