

ReliveMeet: Interactive Visual Storyboards and Chat-Based Guide for Engaging Experiences Reliving Missed Meetings

Erzhen Hu
 Microsoft Research
 Redmond, WA, USA
 University of Virginia
 Charlottesville, VA, USA
 eh2qs@virginia.edu

Nicolai Marquardt
 Microsoft Research
 Redmond, WA, USA
 nicmarquardt@microsoft.com

John Tang
 Microsoft Research
 Redmond, WA, USA
 johntang@microsoft.com

Asta Roseway
 Microsoft Research
 Redmond, WA, USA
 astar@microsoft.com

Sasa Junuzovic
 Microsoft Research
 Redmond, WA, USA
 sasa.junuzovic@microsoft.com

Nathalie Riche
 Microsoft Research
 Redmond, WA, USA
 nath@microsoft.com

Ken Hinckley
 Microsoft Research
 Redmond, WA, USA
 kenh@microsoft.com

Kori Inkpen
 Microsoft Research
 Redmond, WA, USA
 kori@microsoft.com

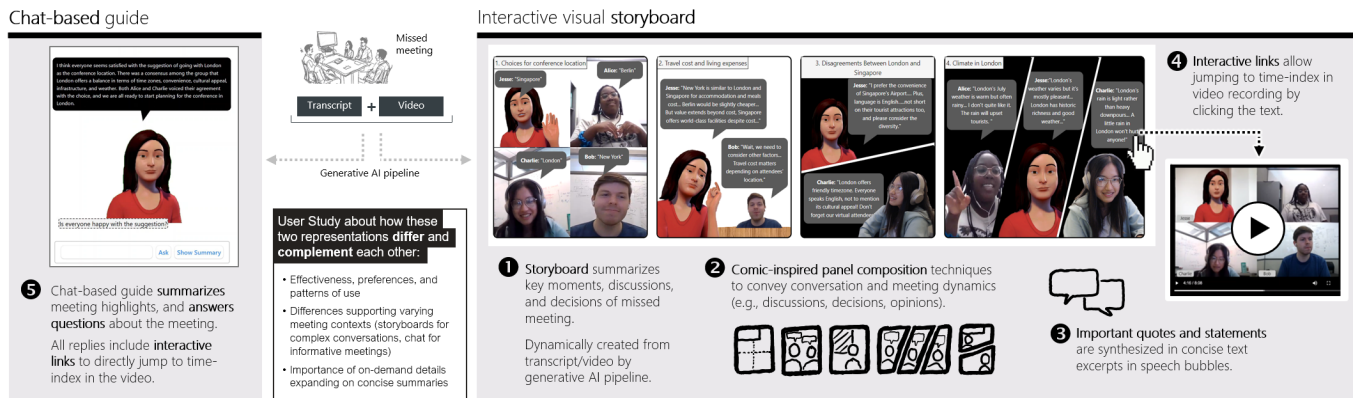


Figure 1: Interactive storyboards provide an engaging visual recap of key moments of a meeting (1), with important quotes and panel composition emphasizing the meeting dynamic (2,3), and direct links to the video (4). The complementary personalized chat-based guide affords asking direct follow-up questions about the meeting (5).

Abstract

Meetings allow sharing knowledge, forming connections, and making decisions. However, we might miss important meetings due to conflicting schedules, and options for catching up are limited and often lack active involvement. To support re-engagement with missed meetings beyond linear consumption of recordings and text summaries, we designed and studied two complementary approaches in our system ReliveMeet: visual storyboards and a chat-based guide. The comic-inspired interactive visual storyboards use frames concisely representing key moments, add relevant quotes, and use panel composition techniques which we describe in a design space. The chat-based guide complements this experience, for asking direct questions in a personalized chat about past events. Frames and chat are interactive, with direct navigation to key moments in the video.

We studied both approaches (12 participants), with findings about their distinct affordances, complementary patterns of use, context-dependent strengths (storyboards for complex conversations, chat for informative meetings), and importance of on-demand details expanding on concise summaries.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI; Empirical studies in ubiquitous and mobile computing.**

Keywords

video-conferencing, interactive meeting recap, storyboards, generative AI

ACM Reference Format:

Erzhen Hu, Nicolai Marquardt, John Tang, Asta Roseway, Sasa Junuzovic, Nathalie Riche, Ken Hinckley, and Kori Inkpen. 2026. ReliveMeet: Interactive Visual Storyboards and Chat-Based Guide for Engaging Experiences Reliving Missed Meetings. In *Graphics Interface 2026 (GI'26)*. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3491102.0000000>



This work is licensed under a Creative Commons Attribution 4.0 International License. *Graphics Interface 2026 (GI'26)*, Waterloo, Canada
 © 2026 Copyright held by the owner/author(s).
 ACM ISBN 978-1-4503-9157-0/00/00
<https://doi.org/10.1145/3491102.0000000>

1 Introduction

In the dynamic environment of modern workplaces, active participation and contribution in meetings are pivotal to effective decision-making, knowledge-sharing, and relationship-building. However, the reality of conflicting schedules, double-booked meetings, or personal commitments like vacations can lead to individuals missing important meetings. The infrequent and insufficient review of recorded meetings [8] compared to live participation may exacerbate inefficiencies and productivity losses. This highlights a gap in post-meeting information transmission and absorption, where passive reviews of recorded meetings fail to adequately engage absent individuals. A key issue with missing live meetings is the loss of the feeling of involvement, which is integral to the collaborative process. Recorded meetings, while useful, often fail to replicate this sense of active participation and connection with the content and other participants. For individuals to stay effectively informed and contribute to ongoing projects, it's essential not only to understand the decisions and discussions that took place but also to feel a part of those processes.

Current AI-augmented meeting tools attempt to bridge this in two main ways. Commercial systems such as Microsoft Teams Intelligent Recap [37] and Otter.AI [40], representative of a growing ecosystem of AI meeting assistants, combine AI-generated notes, topic chapters and timelines, speaker-attributed transcripts, video anchors, and chat-based question answering over a recording. Research has explored behavioral and turn-taking visualizations [45, 55] and automated agents that take notes or contribute on the user's behalf [31, 40]. These approaches have made it easier to skim and query recordings. The dynamic of the conversation (the way participants take turns, agree, disagree, interrupt, and build a decision together) is, however, typically distilled into bullet points or surfaced indirectly through search, rather than being made visually inspectable as a structure that users can revisit.

We contribute ReliveMeet, a meeting recapitulation tool designed to provide an interactive experience to relive missed meetings. ReliveMeet incorporates a comic-inspired visual storyboard that allows meeting participants to interactively review key moments from missed meetings (including contributions from all the active speakers, *i.e.*, human attendees and autonomous agents if enabled) all presented concisely in 5–6 key storyboard frames. Each storyboard frame emphasizes interactional structure through succinct quotes, speaker attribution, and visual cues that highlight turn-taking patterns and conversational dynamics. Frames are also directly linked to the original meeting video, enabling users to fluidly transition between summarized and detailed views. Complementing the visual storyboards, ReliveMeet provides a chat-based guide that allows users to query the meeting after the fact. Through conversational interaction, users can ask targeted questions about decisions, rationales, or points of relevance to their current work, enabling deeper exploration beyond what is captured in the visual summary alone.

To gain insights into the different ways people might use *Storyboards* and the *Chat-based Guide* to review past meetings, we conducted a qualitative study of ReliveMeet with 12 study participants. Rather than comparing the two interfaces against each other, our goal was to characterize their distinct affordances and understand how each supports different aspects of meeting review. For the

scope of the study, we focused on a scenario in which users review a meeting where an autonomous agent attended on their behalf. This scenario allowed us to study the distinct and complementary uses of the storyboards and the chat-based guide.

In summary, we contribute the following:

- The design of ReliveMeet, providing a comic-inspired *storyboard-first visual representation of meeting interaction* that explicitly encodes turn-taking patterns, speakers, key quotes, and frame variations for mutual exchange, disagreement, and interruption, coupled with a complementary chat-based guide. Rather than positioning the storyboard as a replacement for summaries or chat-based recap, we explore it as a complementary visual substrate for re-engaging with missed meetings.
- Findings from a user study ($N = 12$) characterizing how people use visual storyboards and the Chat-based Guide to relive meetings, including their distinct affordances, patterns of complementary use, context-dependent preferences, and the importance of on-demand detail to expand concise summaries.

By integrating multi-modal information into an interactive storyboard format, ReliveMeet aims to shift meeting catch-up from a passive consumption task to an active exploration process. Users can first grasp the high-level narrative of a meeting through storyboards, selectively dive into original video segments, and engage in conversational inquiry to resolve open questions or surface personally relevant details.

2 Related Work

ReliveMeet builds on research in (1) meeting capture and post-hoc review for non-attendees, including recent AI-augmented meeting tools, and (2) comic-based visual representations for communication and learning.

2.1 Recapping Missed Meetings

Reviewing recorded meetings serves multiple purposes, such as enabling individuals to catch up on missed meetings [39, 49], reinforcing participants' memory [24], and supporting social and educational goals. However, raw recordings are time-intensive to navigate and poorly convey interactional context, motivating interfaces that help users locate salient moments and reconstruct what happened.

Early systems integrated multiple meeting streams (e.g., audio, video, slides) with explicit indexing and sensor- or analysis-driven navigation [18, 33, 48, 56]. Examples include FiloChat [58], LiteMinutes [14], TeamSpace [19], and the Carnegie Mellon Meeting Recorder [7], alongside techniques for indexing and navigating video conferences [15, 29]. Subsequent work used visualization to support recall of meeting context and conversational structure [7, 47], such as MeetingVis [47], which depicts changes in communication patterns, and MeetingCoach [45], which surfaces behavioral and contextual cues to support reflection. More recently, Chen *et al.* [13] applied LLMs to generate *dialogue maps* of meeting conversations in real time (visual node-link structures reflecting ideational rather than chronological order), finding that users preferred this non-linear representation over traditional transcripts. Where MeetMap addresses the synchronous, in-meeting case, ReliveMeet addresses the complementary post-meeting case via storyboards and chat.

Several systems explicitly address the needs of non-attendees [19, 39, 58]. Nathan *et al.* [39], for example, enabled sharing attendee annotations during remote meetings. Related, Asthana *et al.* [3] designed and evaluated an LLM-powered recap system with two complementary textual representations (salient *highlights* and a hierarchical *minutes* view), finding both valuable in distinct contexts: highlights supported quick orientation, while minutes enabled deep review. Time Travel Proxy [53] takes a different approach: non-attendees pre-record a message for attendees to view, after which the system supports recap and response through video threads. *Dittos* [31] introduced personalized embodied agents that participate in meetings on behalf of unavailable users. ReliveMeet addresses a complementary *post-meeting* problem: once such an agent has attended a meeting, how can the absent user effectively review and re-engage with what happened? Wang *et al.* [57] recently surveyed information workers characterizing how post-meeting artifacts are used, identifying five core uses including *onboarding non-attendees* and *group sensemaking*. They argue that effective meeting artifacts should let users continue to transform and *engage* with meeting information rather than passively access it.

More recently, commercial systems (e.g., Microsoft Teams [37], Otter.AI [40]) support AI-generated notes, topic detection, and autonomous agent participation in meetings. HCI research explored the role of intelligent personal assistants and conversational agents in inferring context, providing proactive services [52], and representing individuals in different settings [9, 12, 16, 46], including document retrieval in meetings and action item management in professional environments [36]. In contrast to tools centered on playback or text summaries, ReliveMeet focuses on *post-meeting re-engagement*: it summarizes meeting dynamics in a compact visual narrative while enabling open-ended, on-demand exploration via a conversational guide.

2.2 Comic-Based Visual Representation in Computer-Mediated Communication

Data comics have proven effective in communicating complex information [5], enhancing understanding, engagement, and reducing cognitive load [4, 27]. The visual language of comics has been used for data storytelling [35] and introducing programming concepts [50]. Bach *et al.* [6] further articulate a pattern language for sequencing data-comic panels, situating each panel within a deliberate narrative role rather than as a standalone image. ReliveMeet adopts this visual grammar to transform meeting videos and transcripts into a multi-level storyboard that foregrounds salient moments while preserving access to source detail.

Comics have also been used directly as an interface for communication. Comic Chat [30] represents participants as comic characters and automatically generates panels with gestures and expressions to depict conversational flow. This work demonstrates how comics can encode not only *what* was said but also aspects of conversational performance. We extend this idea from synchronous text chat to asynchronous review of recorded video meetings, using storyboard panels to externalize turn-taking, speaker attribution, and quotable moments, with links back to the underlying recording.

Meeting platforms increasingly support mixed-modality participation (audio/video alongside avatars [34]), including commercial

systems such as Microsoft Teams and Gather Town. Empirical comparisons of representation modes suggest that audio, video, and avatars can yield similar outcomes in some settings, while raising concerns about professionalism and trust depending on design and context [23, 26]. Exploring mixed-modality settings, Panda *et al.* [41] investigated the impact of visual representation on co-presence in meetings. By leveraging insights from data storytelling, information visualization, and comic-based communication, we aim to create an effective and engaging way to review and recap agent-attended meetings.

3 ReliveMeet

We designed ReliveMeet with the goal of providing an interactive and engaging experience to relive past meetings. To ground the system design, we first articulate the scope of AI-enabled meeting recap, focusing on the roles of participants and their differing forms of presence. We then describe the design of ReliveMeet, with particular emphasis on two core components: *Storyboards*, which unpack a design space of visual composition and interactivity, and the *dialogue with Chat-based Guide*, which supports query-driven reflection on meeting content.

3.1 Scope of AI-Enabled Recap of Meetings

3.1.1 Role and Presence of Participants. Meeting recaps must account for heterogeneous forms of participation and presence. We categorize meeting participants into three groups based on their presence and contribution during the meeting: (1) **Human attendees**, who are present in the meeting, ranging from minimal presence (muted and cameras off) to full presence (avatars or live video), with varying levels of contribution from passive to highly proactive; (2) **Proactive Non-attendees**, who are not physically present but contribute via autonomous agents acting on their behalf, enabling timely input despite absence (e.g., [31]); (3) **Passive Non-attendees**, who are either absent or represented by AI note-takers (muted and cameras off), characterized by minimal or no presence and contribution.

This classification captures diverse modes of engagement and contribution, and highlights differences in participants' prior involvement, awareness, and perceived presence. These factors may shape how individuals later review and interpret meeting records.

3.1.2 Varied Knowledge of Meeting Content. Before engaging with a meeting recap, the level of awareness and understanding of meeting content differs among the three participant groups, outlined as follows: (1) **Human Attendees** may possess comprehensive or partial understanding of the meeting, shaped by attention, multitasking, and degree of involvement, and are fully aware of their own contributions; (2) **Non-Attendees via Autonomous Agents** may know the instructions given to their agents but lack visibility into how those instructions were executed, how contributions unfolded, and how others responded; (3) **Passive Non-Attendees** typically lack firsthand knowledge and rely entirely on post-meeting artifacts.

These different levels of pre-meeting knowledge underscore distinct recap needs between attendees and non-attendees, motivating systems that adapt to varying degrees of prior knowledge and support user-driven curation of post-meeting experiences.

3.1.3 Availability for Spontaneous Catch-Up. During and after meeting review, participants across all three groups may seek additional clarification or context, especially non-attendees who lacked opportunities to engage in real-time interaction. Prior work shows that post-meeting catch-up commonly occurs through email, messaging, or ad hoc conversations [53]. This highlights the importance of recap systems that not only summarize content but also support spontaneous, lightweight follow-up and inquiry beyond the meeting itself.

3.2 ReliveMeet Overview

The overarching design goal of ReliveMeet is to provide an engaging meeting recap experience that combines multi-modal visuals with interactive exploration. To achieve this, ReliveMeet integrates two complementary components (Figure 1): (1) the **Storyboard**, which offers a visually structured representation of conversation flow, key quotes, active speakers, and meeting dynamics; and (2) the **Chat-based Guide**, which presents an essential meeting summary and supports in-depth, query-driven interaction.

These components embody two *complementary* interaction paradigms for revisiting past meetings: exploration-based interaction through the Storyboard and query-based interaction through the Chat-based Guide. They can effectively work together, with each addressing different user needs on meeting context and personal preference. Users may minimize the Chat-based Guide to focus on a full-screen Storyboard view. Meeting recordings are represented using a grid-based video layout that includes both human attendees and autonomous-agent proxies, ensuring consistent visual treatment across participant types.

3.3 Recap of Meetings in Visual Storyboards

To effectively convey conversational dynamics and participant responses, we designed a visual storyboard technique that captures key meeting moments. The Storyboard presents a sequence of frames, each corresponding to a salient topic discussed within a specific time segment, arranged in chronological order (Figure 1, Figure 2a). Users can scroll through the frames to contextualize discussion flow and speaker interactions.

We chose the storyboard format because comics provide an established visual grammar for representing temporally ordered multi-party interaction in a single glance [5, 6, 17, 35]. Three properties of this grammar are particularly well-suited to meeting recap. First, panels and panel-to-panel transitions support *closure* [35]: readers actively reconstruct narrative continuity between discrete moments, matching how users want to skim past small talk and reconnect the key episodes of a meeting without watching it end-to-end. Second, the spatial co-location of speakers, speech bubbles, and reactions within and across panels externalizes *turn-taking dynamics* (who responded to whom, who was interrupted, who agreed) that are difficult to recover from a linear transcript, topic chapter list, or bullet-point summary. Third, the format affords *multi-level reading*: headlines and panel composition convey the gist at a glance, while quotes and time-linked video provide drill-down. The storyboard is a representation choice optimized for conveying conversational structure, complementing the chat-based guide, which is optimized for targeted information retrieval.

3.3.1 Storyboard Components. The Storyboard consists of five primary components (Figure 2):

- **Frame Pattern:**
The frame patterns (Figure 2a) encode conversational structures such as mutual exchange, disagreement, or interruption, reflecting turn-taking dynamics. Each frame comprises a headline, character(s), speech bubbles with key quotes, and an associated (hidden) time-coded video segment. The order of frames represents the chronological order of the meeting's time segments.
- **Headlines:**
The headlines summarize the high-level topic addressed within each frame, providing contextual grounding for the associated dialogue.
- **Characters:**
Speaker vignettes capture representative moments within a frame. Characters may include both speakers and audience, incorporating non-verbal cues such as facial expressions or gestures, as explored in prior work [38]. Expanding on prior work that has considered facial expressions as meeting tones in post-meeting visualizations (e.g., [45]), representing such cues supports a sense of "reading the room," conveying meeting atmosphere beyond textual summaries alone.
- **Speech Bubbles:**
Each speech bubble (Figure 2b) contains a speaker name and a set of distilled quotes that concisely represent the speaker's original statements (implementation details in Appendix A). Quote length varies from keywords to full sentences, depending on context. For instance, given a headline about a decision-making meeting for conference location choices, a short term like "Singapore" can sufficiently represent the person's choice, as opposed to a longer sentence such as "I prefer Singapore." Clicking a quote navigates users to the corresponding video moment. Within a bubble, quotes follow temporal order, but across bubbles, spatial arrangement reflects interactional turn-taking rather than strict chronology.
- **On-Demand Video Clips:**
The Storyboard prioritizes narrative structure by hiding video clips by default. Selecting a frame or headline jumps to the start of the corresponding topic, while clicking individual quotes reveals finer-grained video segments linked to more specific moments related to the textual information.

Prior work on comic-style communication and speech bubbles explored avatar-attached dialogue [30] and accessible captioning [42]. In contrast, our work focuses on condensing multi-level quotes within speech bubbles to support meeting recap and drill-down from original transcripts, with quotes representing succinct speaker statements rather than direct captioning and without optimizing bubble style or real-time avatar adaptability. Additionally, the storyboard format captures turn-taking patterns, speaker emphasis, and conversation flow that are often lost in text-based meeting summaries or existing visualization approaches, which can increase mental load and reduce engagement [45, 51].

3.4 Design Space of Visual Recap Storyboards

Informing the construction of visual storyboards for meeting recaps, we articulate a design space along two dimensions: (1) **Composition**, encompassing static frame structure, weighting, and spatial

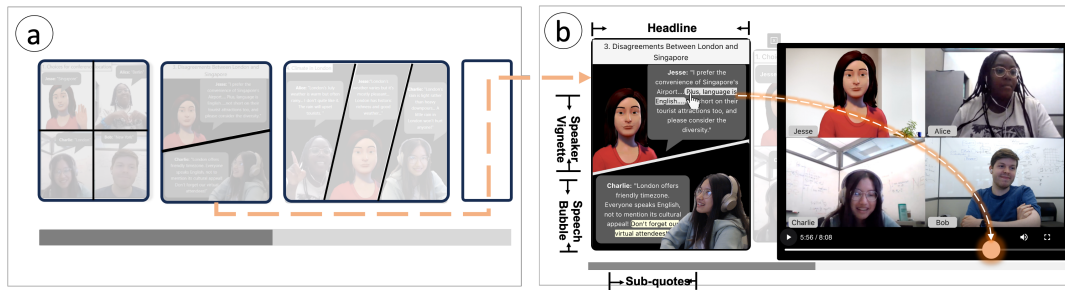


Figure 2: Storyboard Components: a) Variations of storyboard frames based on the conversation type. b) Components in frames: headlines, speech bubbles, and speaker vignettes.

organization; and (2) **Interactivity**, capturing dynamic interactions within and across frames that support narrative continuity, character persistence, shared task spaces, and storyboard layouts.

3.4.1 Composition: Fixed In-Frame Generation.

C1 Variations:

Frame variations represent fundamental conversational units and transitions, enabling depiction of distinct turn-taking patterns. Different frame patterns craft various conversational structures and transitional patterns within a group meeting, delineating the temporal dynamics and allowing for smooth transitions between different conversation states.

C2 Intensity:

Frame significance can be weighted (e.g., by adjusting frame size) in response to participation level or interaction density. The intensity can be adjusted through interactive elements, such as expanding a grid to show more granularity.

C3 Continuity (Moderator or Agenda-guided):

The nature of turn-taking in conversations varies based on the presence or absence of moderation. Structured turn-taking, overseen by a chairperson or moderator, contrasts with more dynamic exchanges in informal settings. The meeting agenda plays a critical role in guiding interactions and dictating the flow from opening remarks to subsequent discussion topics.

C4 Style:

Frame styles can range from straightforward projections of recorded video with photo-realistic human snapshots to diverse graphical representations using different sketch genres. Previous work [41] explored avatar styles and digital representations with varying visual fidelity for mixed-modality video meetings. The style can extend beyond avatars, encompassing elements like speech bubbles or the overall framing aesthetic.

C5 Spatiality:

We distinguish no spatiality (NS), partial spatiality (NS+S), and full spatiality (FS). In remote meetings, spatiality is initially limited, with participants displayed as talking heads. However, recent features in applications like Microsoft Teams' 'together mode'¹, and research prototypes like MirrorBlender [20] and OpenMic [21] offer 2D virtual frames of reference or immersive

co-located spaces. Perspectives [54] seamlessly integrate remote and co-located persons into a virtual and immersive co-located space, demonstrating that spatiality, the sense of space or location in a virtual environment, is not restricted by physicality, the actual physical presence, or configuration of individuals. The chosen level affects the meeting's *establishing shot* [2], influencing focus and context. Maintaining physicality can provide a sense of place and refer to specific gestures or workspaces. Tools like Owl 360° cameras² can dynamically focus on speakers while maintaining context, establishing temporal priorities in real-time meetings and potentially aiding post-meeting keyframe selection.

3.4.2 *Interactivity*. Users' dynamic interactions within and between visual frames can create an immersive narrative experience. This captures different ways characters and shared task spaces *inhabit* the frame, crafting the look and feel of the storyboard. Different discourse transitions can be correlated to support storyline continuity, enhancing narrative flow through techniques like time-lapse videos and *out-of-frame patterns*.

I1/2 Weight (Interactive Elements In-frame/Out-of-frame):

Interaction can vary how the details are shown grid by grid. Interactive elements such as objects and characters can emerge from the frame (move in and out) based on specific triggers.

I3 Binding:

Two different frames can be bound or aligned to emphasize continuity between storyboard panels and certain actions between characters. Connections between frames enhance understanding and flow, treating frames as interconnected parts of a storyline rather than independent conversations. For example, if different participants are responsible for leading different agenda items, the current speaker might conclude their part and then handover to the next person. At the end of discussing one agenda item, the moderator might summarize the key points or decisions made and then introduce the next item.

I4 Transformation:

Objects and characters can undergo meaningful transformations within the frame, transitioning from photorealistic snapshots to digital twins. This includes exploring digital formats of physical

¹Together Mode: <https://www.microsoft.com/en-us/microsoft-teams/teams-together-mode>

²<https://owllabs.com/>

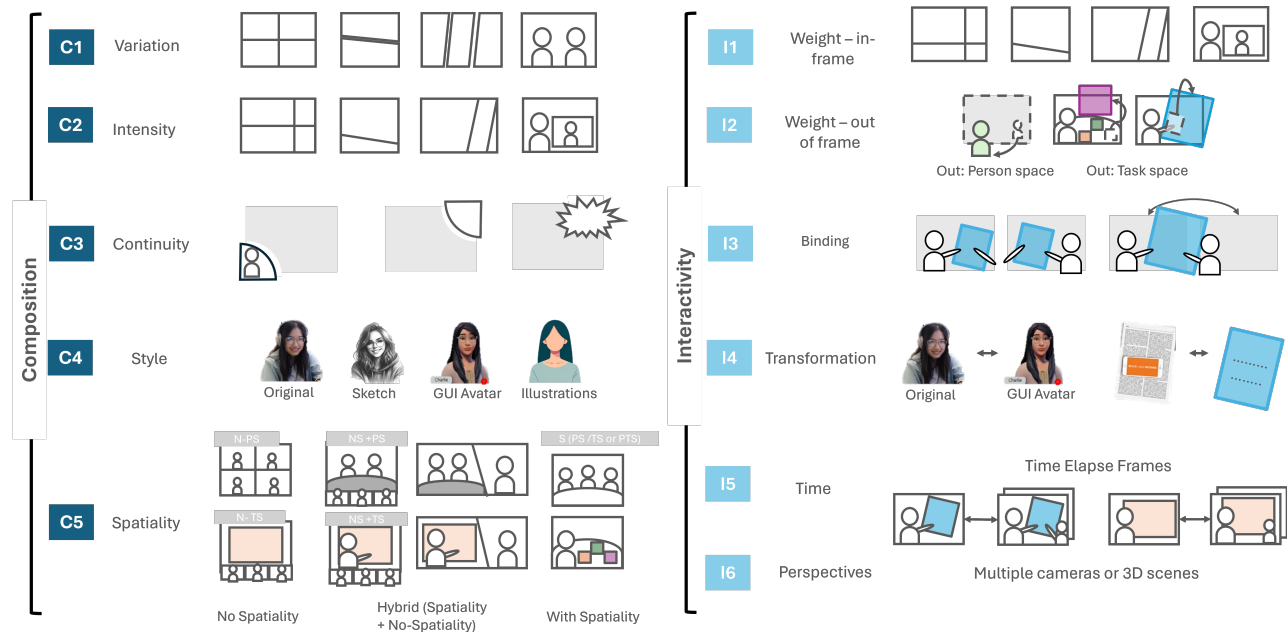


Figure 3: Design Space of interactive storyboards: composition and interactivity.

items (e.g., [22, 25]), such as handwriting recognition or text extraction and augmentation (e.g., [32]), and leveraging prior work (e.g., [28]) on object removal and video inpainting techniques.

15 Time:

User-controlled moment selection and determining the number of frames included in time-lapse videos can contribute to temporal exploration within the frame.

16 Perspectives:

Utilizing multiple cameras or 3D reconstructed views, users can access different perspectives continuously or discretely within the same time frame, enhancing the interactive storytelling experience.

Two additional strategies allow on-demand reveal of more fine-grained information in the storyboard and speech bubbles, which are important for scaling this approach to longer meetings. The goal is to emphasize the narrative and improve the readability and coherence of the storyboard. (1) *Expandable Speech Bubbles*: Speech bubbles can represent one sentence of a certain time period, and this sentence can be shown in longer sentences by varying the token numbers to achieve more details like a caption. The user can choose to expand certain statements/quotes rather than directly clicking to watch videos. (2) *Nested storyboard Layouts*: Nested storyboards present a hierarchical view from broad overviews to detailed discussions, facilitating a comprehensive understanding of meeting themes and specific topics. Upon clicking a coarse-grained pane, the storyboard will be represented in a more fine-grained manner within the topics.

3.5 Storyboard Continuity Editing

Treating frames as parts of a single storyline (rather than as independent thumbnails) required attention to continuity conventions from comics and film [2, 6, 35, 44]. In ReliveMeet, we (i) preserve *speaker*

identity across adjacent panels via a consistent character vignette (extracted from videos) and label [2], so that a participant appearing in two consecutive frames reads as the same person continuing rather than re-introducing themselves; (ii) use *panel adjacency* to imply temporal continuity, so consecutive panels correspond to consecutive time segments of the meeting [6]; (iii) reserve *frame-pattern changes* (C1, Figure 3), such as shifting from a mutual-exchange to a disagreement layout, to mark genuine shifts in conversational structure rather than visual variation for its own sake; and (iv) use *frame weighting* (C2) to reflect participation intensity rather than chronological duration alone. We deliberately avoid spatial co-locations within a frame that would imply face-to-face interactions that did not occur in the underlying meeting. Other storyboard continuity editing techniques could complement these strategies, such as the *180-degree rule* for maintaining consistent spatial relationships between characters, *eye line match* for cross-panel linking through matching line of sight, and *strategic camera/viewpoint angles* and shots [44].

3.6 Chat-based Guide

In addition to visual exploration, ReliveMeet provides a chat-based guide as an alternative and complementary recap modality. Prior work [1, 43] explored chat-based animated virtual humans or avatars for various contexts that augment the conversation. Embodied conversational agents (e.g., [11]) were found to leverage users' knowledge of face-to-face conversation, making interactions more natural and effective. We introduce a chat-based guide in ReliveMeet that allows users to engage in a dialog about the meeting, addressing the challenge of contextualizing and accessing information from AI-generated meeting summaries. The Chat-based Guide serves as a personalized meeting agent, guiding users through the meeting

experience through a more personalized perspective. The conversation with the Chat-based Guide starts with key moments from the meeting summarized in a speech bubble, allowing users to quickly grasp the essence of the meeting. Users can then interact with the Chat-based Guide in several ways: They can type queries to ask free-form questions about the meeting, and by clicking on bullet points or keywords within any of the replies, they can navigate directly to specific moments in the video. In response to entered queries, the Chat-based Guide provides detailed answers based on the user's questions, along with links to the most relevant video moments. By providing a personalized, interactive, and context-aware interface, we aim to enhance users' ability to extract meaningful insights from their meetings.

4 Study: Storyboards and Chat-based Guide

The goal of our study was to characterize the distinct affordances of ReliveMeet's two interaction modalities (visual storyboards and more conventional chat-based meeting recap chat), and understand how each supports different aspects of meeting review. Rather than comparing the two interfaces against each other, our overarching goal was to identify their complementary strengths and understand how users engage with each depending on meeting context and their personal needs. We focused on a representative scenario in which a user dispatches an autonomous agent to attend a meeting on their behalf, and later revisits the meeting through ReliveMeet.

Design Probes. As the study apparatus, we use two interfaces of the ReliveMeet system for video-based meeting recap: the Storyboard view that provides a generalized summary, and the Chat-based Guide, which supports interactive exploration (both interfaces shown in Figure 1 and Figure 2). Storyboard content was pre-generated to ensure consistency across participants, allowing all users to review the same summarized representation of the meeting. In contrast, Chat-based Guide responses varied based on participants' queries, reflecting realistic human-LLM interaction and allowing individual differences in exploration and sensemaking to emerge.

Participants. We recruited 12 participants (6 females, 6 males; aged 20-54) from a large technology company. All had prior experience reviewing missed work meetings using tools such as Microsoft Teams. The in-person study approved by IRB was conducted via a web browser; sessions were recorded in audio and video and lasted 60 minutes, with participants receiving a \$50 gift card.

Participants reported missing meetings for a variety of reasons, including scheduling conflicts (9/12), personal reasons such as vacation (5/12), and time zone differences (4/12). Additional reasons included preferring recordings (3/12) or perceiving meetings as non-critical or outside work hours (2/12 each). All participants were familiar with LLMs, with usage varying (median = 3, IQR = 2) and trust levels ranging from 2-4 (median = 3.5, IQR = 1) on a 5-point scale.

Procedure and Tasks. The study began with a background survey assessing participants' experience, strategies, and challenges in reviewing missed meetings. The experimenter then introduced the concept of deploying a fully autonomous meeting agent and explained the motivation for post-meeting recap. Participants were instructed to imagine that one avatar-based attendee in a recorded

meeting represented their agent, while the remaining three attendees were their colleagues.

For the study materials, we used two LLM-generated simulated meeting transcripts to ensure consistent conversational flow (Appendix B provides optional details about the procedure for generating the transcripts and producing the corresponding meeting videos). This allowed participants to review the meeting as if the agent had attended on their behalf. Participants then completed tasks with the two interfaces in a counterbalanced order. We chose generated transcripts over recordings of real meetings to ensure that all participants reviewed identical content, isolating interface differences from content differences, but it is important to acknowledge that real meetings are typically more complex than the transcripts used in these scenarios.

Tutorial Task (15 min). Before using each interface, participants watched two short tutorial videos. The experimenter then guided them through key features of Storyboard and Chat-based Guide using example search questions, asking participants to perform basic actions such as locating a speaker's discussion or identifying the meeting conclusion.

Summary and Recall Task (20 min). For each interface, participants were given three minutes to review a 7-minute meeting video and informed that they would later summarize the content. The constrained review time was chosen to simulate realistic time pressure, where users who miss meetings might not have time to watch the full meeting recording. Participants then answered three content-related questions, freely using the interface to search for or recall information.

Survey and Interviews (25 min). After completing tasks with each interface, participants completed a survey assessing usefulness, usability, and confidence in understanding and navigating the meeting content using 5-point Likert scales, followed by a short summary quiz and three multiple-choice questions. The study concluded with a post-study interview focused on participants' strategies, feedback, and suggestions. All click-level interactions, as well as queries and responses from Chat-based Guide, were logged for analysis.

5 Findings

General Feedback. A one-way ANOVA ($\alpha = .05$) was used to examine whether participants' confidence differed when revisiting meetings using Storyboard (SB) and Chat-based Guide (CG), as a way to understand the distinct strengths of each modality rather than rank them. Confidence in understanding the meeting content (Medians: SB = 4.5, CG = 4.0; IQRs: SB = 1.0, CG = 2.0; $p = .095$) and in explaining it to others (Medians: SB = 4.5, CG = 4.0; IQRs: SB = 1.0, CG = 2.0; $p = .096$) did not differ by interface. In contrast, participants reported significantly greater comfort navigating to specific moments with Storyboard than with Chat-based Guide (Medians: SB = 5.0, CG = 3.5; IQRs: SB = 1.0, CG = 2.0; $p = .023^*$). Confidence in understanding individual speakers' points showed no significant difference (Medians: SB = 4.0, CG = 3.0; IQRs: SB = 1.25, CG = 1.25; $p = .121$), though three participants reported higher confidence with Storyboard. A summary of the survey results across both interfaces is shown in Figure 4.

When asked about their preferences for Chat-based Guide and Storyboard in different scenarios, participants generally favored

Storyboard for more complex conversations, such as group meetings and stand-up meetings. P3 stated, "It's good for complexity... I think [with] 2-3 or more people I probably would like the storyboard more." P5 added, "...stand up speech person would have a main topic and other people contributing to it and so it makes more sense for topics to be aligned and see how other people reacted to a specific topic that's been discussed." In contrast, Chat-based Guide was preferred for more informative sessions where information extraction was the primary goal. P5 explained, "I think the [Chat-based Guide] helps when it's more of, you aren't that involved. There's not a lot of shifts across topics, and you're kind of observing. So things like things are more informative, Let's say an information session, a lecture, a conversation between a host and a speaker." P11 highlighted the importance of Chat-based Guide for comprehending critical meetings, saying, "If it's a more critical like something I'm trying to grasping very much understand, I would prefer the other one (Chat-based Guide)."

These findings underscore that the two modalities serve different needs depending on meeting context, rather than one being universally preferred over the other. The qualitative pattern suggests that Storyboard was favored for meetings with *higher conversational complexity* (multi-party discussions, stand-ups, and decision-making sessions with shifting viewpoints), where the spatial layout of frames helps externalize who said what and how the conversation moved. Chat-based Guide was favored for *information-dense, lower-interaction* settings such as lectures, one-on-one updates, or critical meetings where users wanted to ask precise, drill-down questions rather than scan for structure. This complementary pattern echoes Asthana *et al.* [3], who similarly observed two LLM-powered recap representations (both through text: highlights and hierarchical minutes) serving different recap needs. Our findings extend that pattern from text-based formats to a visual storyboard paired with a conversational guide. The findings align with our goal of understanding how each interface supports different recap strategies, and suggest that the most effective recap experience may involve offering both modalities together, with the system suggesting (and the user choosing) which to use based on meeting type. This would allow users to fluidly move between them based on their needs and context.

Storyboard. Overall, 8 out of 12 participants expressed a general preference for Storyboard over Chat-based Guide, though as we describe below, preferences were context-dependent. Participants valued the storyboard's ease of understanding (8/12), navigation, association of visuals with text information, structure, and intuitive flow as key factors.

Participants rated Storyboard as easy to understand (Median = 5, IQR = 0.25) and easy to navigate to important moments (Median = 5, IQR = 0). The integration of visuals, especially facial expressions, with text improved contextual understanding. Participants also valued Storyboard's comic-inspired presentation, describing it as more engaging and easier to read than a traditional meeting, with some noting that they did not need to open the video. Most participants (8/12) highlighted the structured and intuitive flow as supporting their understanding of conversational progression.

However, a few participants (3/12) raised concerns about the transparency of the Storyboard design. P1 wanted to see more of the in-between interactions, stating, "...Why don't I see, Alice and Charlie

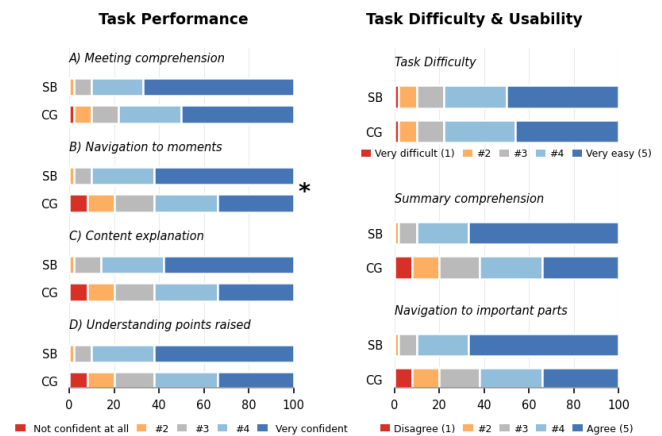


Figure 4: Summary of survey results for Storyboard (SB) versus Chat-based guide (CG)

talk... They must have said something too... I can see that there's something in between here which is important." P3 added, "It splits things up to make it more accessible, but at the same time, it's almost like less transparent..." Some participants (2/12) expressed a desire for more in-depth information from the Storyboard. P2 suggested a feature to view a longer summary, saying, "... the summary and you know the things that you cut out....so I kind of wanted the ability to almost hover over it and see. Maybe I just prefer text... Hover over it and see more." Furthermore, determining the sequence of events or conversations was challenging for some participants (2/12). P2 struggled with identifying who first initiated a particular topic, mentioning, "...some of the ordering matter, right? For here, I couldn't really tell who first suggested it... I think I did, but then the description kind of confused..."

Chat-based Guide. Four out of 12 participants expressed a general preference for the Chat-based Guide, highlighting distinct strengths complementary to the Storyboard. Participants appreciated the ability to ask specific questions about the meeting and activities, receive responses with timestamps, and the spatial layout of the Chat-based Guide interface.

The querying capability of Chat-based Guide was well-received, e.g., P2 appreciated being able to directly query their agent about the meeting. P6 compared Chat-based Guide to the #topics feature in Microsoft Teams but preferred the context provided in the Chat-based Guide responses, saying, "It's really similar to just clicking on the event recording and then having the stream with the #topics. But it drills down to a little more context and semantics of the, you know, jumping around with the topics." P9 likened the interface to using "control F" in transcripts but appreciated the ability to ask questions, stating, "This seemed like a standard transcript and then, going from it. However, having the ability to ask as opposed to control F in the transcripts was good."

Two participants preferred the high-level summary and simplicity of Chat-based Guide, compared to Storyboard. One noted, "...I just like the spatial orientation of things staying in the same place. Being able to go back and forth, with the summary, and the chat." Another added, "So I liked that the summary had all of the key points like right

in one spot rather than having to move all the way through to find all the key points And then it would it was really easy to click on it and go to the video and get the rest of the context for it."

Question Preferences. Most participants expressed interest in asking clarifying questions related to the automated summary or low-key questions they might ask a particular colleague. Several participants mentioned wanting to learn about aspects beyond the meeting transcript, such as emotions, the meeting's vibe, and more complex topics. Two participants preferred having everything included in the summary, negating the need for asking questions.

Affordance of interactions with Storyboard and Chat-based Guide: Navigation Patterns and Strategies. Participants exhibited diverse strategies for navigating and reviewing the meeting content using the Storyboard and Chat-based Guide interfaces. With the Storyboard, most participants began by scanning the headlines to get an overview of the meeting structure, and then dove deeper into specific moments of interest. Some participants read through the Storyboard sequentially like a comic, while others jumped to the end to see the final decision before scrolling back to review the details. When using Chat-based Guide, participants typically started by skimming the automated summary and clicking on key points to jump to the corresponding video segments. They then used the querying feature to ask specific questions and clarify their understanding of the meeting content.

Participants appreciated the high-level overview provided by both interfaces, but also valued the ability to drill down into specific moments and interactions. For example, P9 noted: "*I went and looked into what I was saying ... Is that something I agree with or should I go back and respond to them? Because my autonomous agent was basically speaking for me.*" The click-level interaction logs revealed that participants spent more time watching the video when using Chat-based Guide compared to the Storyboard, although this difference was not statistically significant.

6 Discussion

We investigated how multi-modal visual representations and interactive agents can support more engaging and flexible ways of revisiting missed meetings. The particular study helped characterizing the unique affordances of each modality and understand how they serve different user needs. Our findings highlight complementary strengths of visual, exploration-based recaps and conversational, query-based interaction. These strengths together have the potential to support richer more adaptable meeting recap experiences.

Affordances of Storyboard and Chat-based Guide. The storyboards support an immersive, narrative-oriented recap experience by combining character vignettes, speech bubbles, and video excerpts within a structured, panel-based layout. This design enables users to reconstruct meeting flow through visual storytelling, supporting both sequential reading and selective exploration. Participants often used the Storyboard to first grasp high-level themes and conversational structure, and then drilled down into specific moments of interest.

In contrast, interaction with the Chat-based Guide encouraged more targeted inquiry. Rather than exploring the meeting holistically, users tended to ask focused questions to clarify automated

summaries, understand individual speakers' positions, or probe aspects such as tone, reactions, and implicit context. Together, these findings suggest that visual exploration and conversational querying afford distinct but complementary sensemaking strategies.

Tailoring Recapitulation Interfaces to Meeting Contexts. The preference for Storyboard in complex conversations and Chat-based Guide in information-focused meetings suggests that recap tools should be adaptable, offering multiple, context-aware interfaces that align with meeting characteristics and user needs. Systems could offer multiple recap modes and recommend appropriate interfaces based on meeting characteristics such as number of participants, conversational complexity, or decision criticality.

Balancing Transparency and Summarization in Meeting Recaps. Although participants appreciated the concise summaries offered by Chat-based Guide and the structured overviews provided by Storyboard, they also emphasized the importance of transparency and access to omitted details. This highlights a central design tension in meeting recaps: summarization reduces cognitive load, but excessive abstraction can obscure important context. Future recap systems should support gradual levels of detail, allowing users to move seamlessly from high-level outcomes to underlying evidence. Design strategies such as expandable summaries, contextual annotations (e.g., tooltips), and agent-mediated follow-up queries could help users balance efficiency with thoroughness, enabling both quick catch-up and deeper reflection and additional details when needed.

Limitations of Study. Our study used two design probes, Storyboard and Chat-based Guide, to isolate and understand the contributions of exploration-based and query-based interaction. While this approach provided focused insights into each component, we did not conduct a comparative evaluation against existing meeting recap or transcription tools. This decision reflects the fundamentally different context of agent-attended meetings, but future work could explore how agent-based recaps complement or extend conventional tools in real-world workflows.

Multi-Modal Inference and Comprehensive Meeting Dynamics. The current Chat-based Guide relies primarily on transcript-based LLM querying and does not incorporate visual cues such as facial expressions, gestures, or interaction timing. Integrating these signals could enable richer inference about meeting dynamics, such as engagement, agreement, or tension. Additionally, variability in LLM-generated responses introduces a degree of unpredictability, which mirrors real-world use of generative systems but warrants further study to understand its impact on trust and interpretation.

Closer Coupling of Storyboard and Chat-Based Guide. Our study did not further examine a tighter integration between Storyboard and Chat-based Guide. In the current implementation, Chat-based Guide returns textual responses without explicitly referencing storyboard elements. Future systems could more tightly bind these modalities (for example, allowing Chat-based Guide to point to relevant storyboard panels or to respond with storyboard-level summaries), thereby supporting fluid transitions between visual exploration and conversational querying. More broadly, future work should investigate how different combinations of storyboard-based and chat-based interaction align with different meeting types and user goals. In

addition, our current design emphasizes conversational dynamics and interpersonal space; incorporating shared task artifacts and specialized task contexts into the storyboard may further shape how users interpret and revisit meetings.

7 Conclusion

This paper presented ReliveMeet, a system for reliving missed meetings through AI-powered storyboards and an interactive, chat-based dialog guide. Our study suggests that multi-modal recaps can enable more engaging catch-up experiences: participants used Storyboards to quickly grasp key moments and conversational structure, while Chat-based Guide supported deeper, personalized inquiry when additional context was needed. Together, these complementary interaction modes highlight opportunities for meeting recap tools that integrate visual exploration with query-driven drill-down to help users revisit past meetings more effectively.

References

- [1] Annalena Aicher, Klaus Weber, Elisabeth André, Wolfgang Minker, and Stefan Ultes. 2023. The Influence of Avatar Interfaces on Argumentative Dialogues. In *Proc. of IVA '23* (Würzburg, Germany). ACM, New York, NY, USA, Article 24, 8 pages. <https://doi.org/10.1145/3570945.3607343>
- [2] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. 2015. Understanding Data Videos: Looking at Narrative Visualization through the Cinematography Lens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 1459–1468. <https://doi.org/10.1145/2702123.2702431>
- [3] Sumit Asthana, Sagi Hilleli, Pengcheng He, and Aaron Halfaker. 2025. Summaries, Highlights, and Action Items: Design, Implementation and Evaluation of an LLM-powered Meeting Recap System. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW176 (May 2025), 29 pages. <https://doi.org/10.1145/3711074>
- [4] Benjamin Bach, Natalie Kerracher, Kyle Wm. Hall, Sheelagh Carpendale, Jessie Kennedy, and Nathalie Henry Riche. 2016. Telling Stories about Dynamic Networks with Graph Comics. In *Proc. of CHI '16* (San Jose, California, USA). ACM, New York, NY, USA, 3670–3682. <https://doi.org/10.1145/2858036.2858387>
- [5] Benjamin Bach, Nathalie Henry Riche, Sheelagh Carpendale, and Hanspeter Pfister. 2017. The emerging genre of data comics. *IEEE computer graphics and applications* 37, 3 (2017), 6–13. <https://doi.org/10.1109/MCG.2017.33>
- [6] Benjamin Bach, Zezhong Wang, Matteo Farinella, Dave Murray-Rust, and Nathalie Henry Riche. 2018. Design Patterns for Data Comics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3173612>
- [7] Satanjeev Banerjee, Jason Cohen, Thomas Quisel, Arthur Chan, Yash Patodia, Ziad Al Bawab, Rong Zhang, Alan Black, Roxana Sarbu, Alexander Rudnicki, et al. 2004. Creating multi-modal, user-centric records of meetings with the Carnegie Mellon meeting recorder architecture. (2004).
- [8] Satanjeev Banerjee, Carolyn Rosé, and Alexander I Rudnicki. 2005. The necessity of a meeting recording and playback system, and the benefit of topic-level annotations to meeting browsing. In *Proc. of INTERACT 2005, IFIP*. Springer, 643–656. https://doi.org/10.1007/11555261_52
- [9] Dan Bohus and Eric Horvitz. 2010. Computational models for multiparty turn taking. *Technical Report. Microsoft Research Technical Report MSR-TR 2010-115* (2010).
- [10] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [11] Justine Cassell. 2001. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine* 22, 4 (2001), 67–67.
- [12] Justine Cassell, Tim Bickmore, Lee Campbell, Hannes Vilhjalmsson, Hao Yan, et al. 2000. Human conversation as a system framework: Designing embodied conversational agents. *Embodied conversational agents* (2000), 29–63.
- [13] Xinyue Chen, Nathan Yap, Xinyi Lu, Aylin Gulal, and Xu Wang. 2025. MeetMap: Real-Time Collaborative Dialogue Mapping with LLMs in Online Meetings. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW132 (May 2025), 35 pages. <https://doi.org/10.1145/3711030>
- [14] Patrick Chiu, John Boreczky, Andreas Girgensohn, and Don Kimber. 2001. LiteMinutes: an Internet-based system for multimedia meeting minutes. In *Proc. of International conference on World Wide Web*. 140–149. <https://doi.org/10.1145/371920.371971>
- [15] Patrick Ehlen, Matthew Purver, John Niekrasz, Kari Lee, and Stanley Peters. 2008. Meeting Adjourned: Off-Line Learning Interfaces for Automatic Meeting Understanding. In *Proc. of IUI '08* (Gran Canaria, Spain). ACM, New York, NY, USA, 276–284. <https://doi.org/10.1145/1378773.1378810>
- [16] Camille G Endacott and Paul M Leonardi. 2022. Artificial intelligence and impression management: Consequences of autonomous conversational agents communicating on one's behalf. *Human Communication Research* 48, 3 (2022), 462–490. <https://doi.org/10.1093/hcr/hqac009>
- [17] Ana Figueiras. 2014. Narrative visualization: A case study of how to incorporate narrative elements in existing visualizations. In *2014 18th International Conference on Information Visualisation*. IEEE, 46–52. <https://doi.org/10.1109/IV.2014.79>
- [18] David Geerts, Ishan Vaishnavi, Rufael Mekuria, Oskar van Deventer, and Pablo Cesar. 2011. Are We in Sync? Synchronization Requirements for Watching Online Video Together. In *Proc. of CHI '11* (Vancouver, BC, Canada). ACM, New York, NY, USA, 311–314. <https://doi.org/10.1145/1978942.1978986>
- [19] Werner Geyer, Heather Richter, and Gregory D Abowd. 2005. Towards a smarter meeting record—capture and access of meetings revisited. *Multimedia Tools and Applications* 27 (2005), 393–410. <https://doi.org/10.1007/s11042-005-3815-0>
- [20] Jens Emil Grønbaek, Banu Saatçi, Carla F Griggio, and Clemens Nylandstedt Klok-mose. 2021. MirrorBlender: Supporting hybrid meetings with a malleable video-conferencing system. In *Proc. of CHI '21*. 1–13. <https://doi.org/10.1145/3411764.3445698>
- [21] Erzhen Hu, Jens Emil Sloth Grønbaek, Austin Houck, and Seongkook Heo. 2023. OpenMic: Utilizing Proxemic Metaphors for Conversational Floor Transitions in Multiparty Video Meetings. In *Proc. of CHI '23*. 1–17. <https://doi.org/10.1145/3544548.3581013>
- [22] Erzhen Hu, Jens Emil Sloth Grønbaek, Wen Ying, Ruofei Du, and Seongkook Heo. 2023. ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Things in Video Meetings. In *Proc. of CHI '23* (Hamburg, Germany). ACM, New York, NY, USA, Article 365, 22 pages. <https://doi.org/10.1145/3544548.3581148>
- [23] Kori M. Inkpen and Mara Sedlins. 2011. Me and My Avatar: Exploring Users' Comfort with Avatars for Workplace Communication. In *Proc. of CSCW '11* (Hangzhou, China). ACM, New York, NY, USA, 383–386. <https://doi.org/10.1145/1958824.1958883>
- [24] Murray E Jennex and Lorne Olfinan. 2002. Organizational memory/knowledge effects on productivity, a longitudinal study. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. IEEE, 1029–1038. <https://doi.org/10.1109/HICSS.2002.994053>
- [25] Sasa Junuzovic, Kori Inkpen, Tom Blank, and Anoop Gupta. 2012. IllumiShare: sharing any surface. In *Proc. of CHI '12*. 1919–1928. <https://doi.org/10.1145/2207676.2208333>
- [26] Sasa Junuzovic, Kori Inkpen, John Tang, Mara Sedlins, and Kristie Fisher. 2012. To See or Not to See: A Study Comparing Four-Way Avatar, Video, and Audio Conferencing for Work. In *Proc. of GROUP '12* (Sanibel Island, Florida, USA). ACM, New York, NY, USA, 31–34. <https://doi.org/10.1145/2389176.2389181>
- [27] DaYe Kang, Tony Ho, Nicolai Marquardt, Bilge Mutlu, and Andrea Bianchi. 2021. Toonnote: Improving communication in computational notebooks using interactive data comics. In *Proc. of CHI '21*. 1–14. <https://doi.org/10.1145/3411764.3445434>
- [28] Mohamed Kari, Tobias Grosse-Puppenthal, Luis Falconeri Coelho, Andreas Rene Fender, David Bethge, Reinhard Schütte, and Christian Holz. 2021. Transformr: Pose-aware object substitution for composing alternate mixed realities. In *Proc. of ISMAR '21*. IEEE, 69–79. <https://doi.org/10.1109/ISMAR52148.2021.00021>
- [29] R. Kazman, R. Al-Halimi, W. Hunt, and M. Mantei. 1996. Four paradigms for indexing video conferences. *IEEE MultiMedia* 3, 1 (1996), 63–73. <https://doi.org/10.1109/93.486705>
- [30] David Kurlander, Tim Skelly, and David Salesin. 1996. Comic Chat. In *Proc. of SIGGRAPH '96*. ACM, New York, NY, USA, 225–236. <https://doi.org/10.1145/237170.237260>
- [31] Joanne Leong, John Tang, Edward Cutrell, Sasa Junuzovic, Gregory Paul Baribault, and Kori Inkpen. 2024. Dittos: Personalized, Embodied Agents That Participate in Meetings When You Are Unavailable. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 494 (Nov. 2024), 28 pages. <https://doi.org/10.1145/3687033>
- [32] Zhen Li, Michelle Annett, Ken Hinckley, Karan Singh, and Daniel Wigdor. 2019. Holodoc: Enabling mixed reality workspaces that harness physical and digital content. In *Proc. of CHI '19*. 1–14. <https://doi.org/10.1145/3290605.3300917>
- [33] Anna Macaranas, Gina Venolia, Kori Inkpen, and John Tang. 2013. Sharing Experiences over Video: watching video programs together at a distance. In *Proc. of INTERACT 2013, IFIP*. Springer, 73–90. https://doi.org/10.1007/978-3-642-40498-6_5
- [34] Rosa Mikeal Martey and Mia Consalvo. 2011. Performing the looking-glass self: Avatar appearance and group identity in Second Life. *Popular Communication* 9, 3 (2011), 165–180. <https://doi.org/10.1080/15405702.2011.583830>
- [35] Scott McCloud. 1993. *Understanding Comics: The Invisible Art*. Kitchen Sink Press, Northampton, MA. First edition.
- [36] Moira McGregor and John C. Tang. 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. In *Proc. of CSCW '17* (Portland, Oregon, USA). ACM, New York, NY, USA, 2208–2220. <https://doi.org/10.1145/>

- 2998181.2998335
- [37] Microsoft. 2023. Microsoft Teams Intelligent Meeting Recap. <https://techcommunity.microsoft.com/t5/microsoft-teams-blog/intelligent-meeting-recap-in-teams-premium-now-available/ba-p/3832541>. Accessed: 2025-10-10.
- [38] 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 *Proc. of CHI '21* (Yokohama, Japan) (*CHI '21*). ACM, New York, NY, USA, Article 247, 13 pages. <https://doi.org/10.1145/3411764.3445235>
- [39] Mukesh Nathan, Mercan Topkara, Jennifer Lai, Shimei Pan, Steven Wood, Jeff Boston, and Loren Terveen. 2012. In Case You Missed It: Benefits of Attendee-Shared Annotations for Non-Attendees of Remote Meetings. In *Proc. of CSCW '12*. ACM, 339–348. <https://doi.org/10.1145/2145204.2145259>
- [40] Otter.AI. 2023. Otter.AI. <https://otter.ai/>. Accessed: 2025-10-10.
- [41] Payod Panda, Molly Jane Nicholas, Mar Gonzalez-Franco, Kori Inkpen, Eyal Ofek, Ross Cutler, Ken Hinckley, and Jaron Lanier. 2022. AllTogether: Effect of Avatars in Mixed-Modality Conferencing Environments. In *Proc. of CHIWORK '22* (Durham, NH, USA). ACM, Article 8, 10 pages. <https://doi.org/10.1145/3533406.3539658>
- [42] Yi-Hao Peng, Ming-Wei Hsi, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *Proc. of CHI '18* (Montreal QC, Canada). ACM, New York, NY, USA, 1–10. <https://doi.org/10.1145/3173574.3173867>
- [43] Rabindra Ratan, RV Rikard, Celina Wanek, Madison McKinley, Lee Johnson, and Young June Sah. 2016. Introducing avatarification: An experimental examination of how avatars influence student motivation. In *Proc. of HICSS '16*. IEEE, 51–59. <https://doi.org/10.1109/HICSS.2016.15>
- [44] Karel Reisz and Gavin Millar. 1968. *The Technique of Film Editing* (2nd ed.). Focal Press, London and New York.
- [45] Samiha Samrose, Daniel McDuff, Robert Sim, Jina Suh, Kael Rowan, Javier Hernandez, Sean Rintel, Kevin Moynihan, and Mary Czerwinski. 2021. Meetingcoach: An intelligent dashboard for supporting effective & inclusive meetings. In *Proc. of CHI '21*. 1–13. <https://doi.org/10.1145/3411764.3445615>
- [46] Ameneh Shamekhi, Q. Vera Liao, Dakuo Wang, Rachel K. E. Bellamy, and Thomas Erickson. 2018. Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent. In *Proc. of CHI '18* (Montreal QC, Canada). ACM, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3173965>
- [47] Yang Shi, Chris Bryan, Sridatt Bhamidipati, Ying Zhao, Yaoyue 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. <https://doi.org/10.1109/TVCG.2018.2816203>
- [48] Madeline E. Smith, Leo Ascenzi, Yingsi Qin, and Ryan Wetsman. 2018. Designing a Video Co-Watching Web App to Support Interpersonal Relationship Maintenance. In *Proc. of GROUP '18* (Sanibel Island, Florida, USA). ACM, New York, NY, USA, 162–165. <https://doi.org/10.1145/3148330.3154522>
- [49] Eric W Stein and Vladimir Zwass. 1995. Actualizing organizational memory with information systems. *Information systems research* 6, 2 (1995), 85–117. <https://doi.org/10.1287/isre.6.2.85>
- [50] Sangho Suh, Celine Latulipe, Ken Jen Lee, Bernadette Cheng, and Edith Law. 2021. Using Comics to Introduce and Reinforce Programming Concepts in CS1. In *Proc. of SIGCSE '21* (Virtual Event, USA). ACM, New York, NY, USA, 369–375. <https://doi.org/10.1145/3408877.3432465>
- [51] Wei Sun, Yunzhi Li, Feng Tian, Xiangmin Fan, and Hongan Wang. 2019. How presenters perceive and react to audience flow prediction in-situ: An explorative study of live online lectures. *PACM HCI* 3, CSCW (2019), 1–19. <https://doi.org/10.1145/3359264>
- [52] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. 2020. Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 153 (sep 2020), 23 pages. <https://doi.org/10.1145/3369807>
- [53] John Tang, Jennifer Marlow, Aaron Hoff, Asta Roseway, Kori Inkpen, Chen Zhao, and Xiang Cao. 2012. Time Travel Proxy: Using Lightweight Video Recordings to Create Asynchronous, Interactive Meetings. In *Proc. of CHI '12* (Austin, Texas, USA). ACM, 3111–3120. <https://doi.org/10.1145/2207676.2208725>
- [54] John C. Tang, Kori Inkpen, Sasa Junuzovic, Keri Mallari, Andrew D. Wilson, Sean Rintel, Shiraz Cupala, Tony Carbary, Abigail Sellen, and William A.S. Buxton. 2023. Perspectives: Creating Inclusive and Equitable Hybrid Meeting Experiences. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 351 (oct 2023), 25 pages. <https://doi.org/10.1145/3610200>
- [55] Jaime Teevan, Nancy Baym, Jenna Butler, Brent Hecht, Sonia Jaffe, Kate Nowak, Abigail Sellen, Longqi Yang, Marcus Ash, Kagonya Awori, Mia Bruch, Piali Choudhury, Adam Coleman, Scott Counts, Shiraz Cupala, Mary Czerwinski, Ed Doran, Elizabeth Fetterolf, Mar Gonzalez Franco, Kunal Gupta, Aaron L Halfaker, Constance Hadley, Brian Houck, Kori Inkpen, Shamsi Iqbal, Eric Knudsen, Stacey Levine, Siân Lindley, Jennifer Neville, Jacki O'Neill, Rick Pollak, Victor Poznanski, Sean Rintel, Neha Parikh Shah, Siddharth Suri, Adam D. Troy, and Mengting Wan. 2022. *Microsoft New Future of Work Report 2022*. Technical Report MSR-TR-2022-3. Microsoft. <https://www.microsoft.com/en-us/research/publication/microsoft-new-future-of-work-report-2022/>
- [56] Cheng Yao Wang, Mose Sakashita, Upol Ehsan, Jingjin Li, and Andrea Stevenson Won. 2020. Again, Together: Socially Reliving Virtual Reality Experiences When Separated. In *Proc. of CHI '20* (Honolulu, HI, USA). ACM, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376642>
- [57] Ruotong Wang, Lin Qiu, Justin Cranshaw, and Amy X. Zhang. 2024. Meeting Bridges: Designing Information Artifacts that Bridge from Synchronous Meetings to Asynchronous Collaboration. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 35 (April 2024), 29 pages. <https://doi.org/10.1145/3637312>
- [58] Steve Whittaker, Patrick Hyland, and Myrtle Wiley. 1994. Filochat: Handwritten notes provide access to recorded conversations. In *Proc. of CHI '94*. 271–277. <https://doi.org/10.1145/191666.191763>

A Appendix: Technical Implementation

ReliveMeet consists of a web interface, and a back-end signaling server that logs the click-level and query data. The front-end web interface was built using React+Vite, Javascript, HTML, and CSS. The back-end authentication server was built using Node.js. We leverage generative AI and prompting of Large Language Models (LLMs) to process time-indexed meeting transcripts (from video recordings of meetings). To construct our prompts for guiding LLMs to generate desired outputs by providing them with carefully designed input text [10], we employed few-shot prompting, which involves providing the LLM with a small number of example input-output pairs to demonstrate the desired behavior.

The prompts were designed to extract structured information from the meeting recording transcript, such as timestamps, key conversation patterns, and concise summaries of speaker statements (see Figure 5a). This information is returned in a JSON format, which is then parsed and used to populate the user interface elements in ReliveMeet. For example, the "pattern" and "topic" fields in the JSON output are used to select pre-designed frames and generate headlines for the storyboard panes (Figure 5b). This allows users to quickly grasp the main themes and transitions in the conversation. To extract more granular details for each emerging topic, additional prompts were used to obtain the speaker's name, their specific quotes, and the associated timestamps of these quotes (See Figure 6a). The resulting JSON data was parsed and integrated into the UI, populating the speech bubbles and enabling users to jump to specific moments in the video by clicking on the sub-quotes. The timestamps are embedded within the quotes, allowing users to precisely locate the moment in the video where a particular statement was made.

To create the speech bubbles in Chat-based Guide, more granular information is extracted from the video, such as speaker names, verbatim quotes, and their associated timestamps (Figure 6a). The resulting JSON output is then parsed and used to populate the speech bubbles in the Chat-based Guide interface. The inclusion of timestamps allows users to click on a specific quote and jump directly to the corresponding moment in the video, facilitating efficient navigation and review of the meeting content. To maintain a consistent and engaging narrative, Chat-based Guide is designed to respond to user queries using a first-person perspective, even if the user's input contains different pronouns. This is achieved by prompting the LLM to generate responses that align with the first-person perspective used in the meeting summary and storyboard content.

The communication between the user interface and the back-end server is handled by a signaling server that uses HTTP post requests to send user queries to the Azure AI "gpt-3.5-turbo" service and retrieve the generated responses (Figure 6b). This allows for real-time interaction with the LLM and enables users to engage in a dynamic conversation about the contents of the meeting.

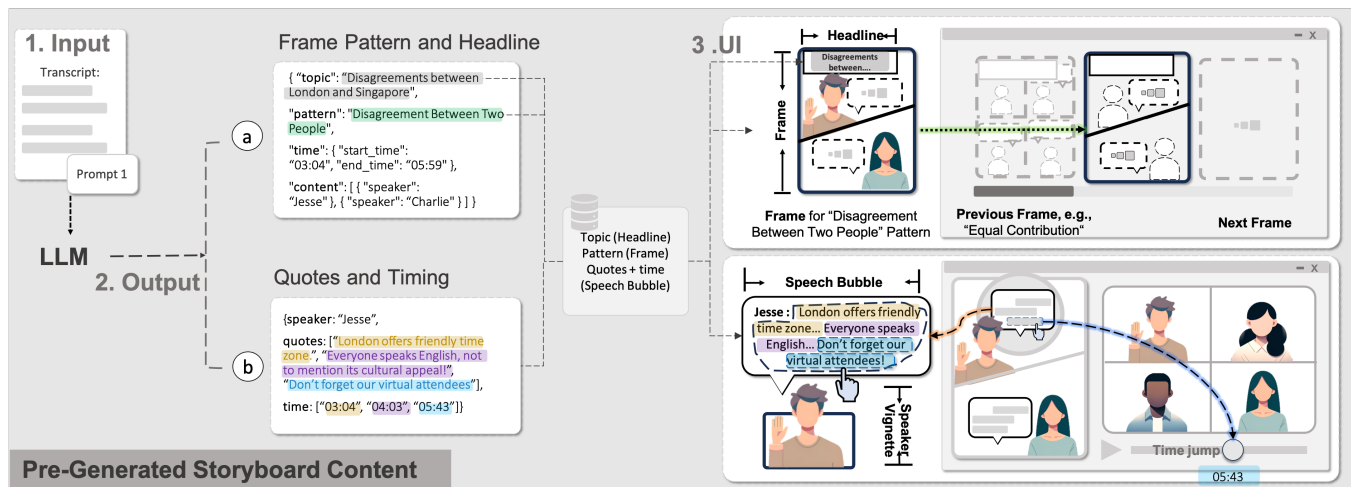


Figure 5: Preloaded Storyboard: The bottom shows the output JSON from LLM. The top illustrates the process of parsing JSON data into the Storyboard user interface for (1) Headline, Frame and (2) Contents.

B Appendix: Additional Information about Study Apparatus and Materials

For the study, two simulated meeting transcripts were created using LLMs to maintain consistency in the conversational flow for the two interfaces (Figure 7).

Each transcript featured various interaction scenarios, including casual discussions, participants sharing preferences, instances of the autonomous agent being mentioned or interrupted by human attendees, a debate involving the autonomous agent and another participant, and a wrap-up session where the autonomous agent interrupts human attendees. The chosen topics for these simulated meetings were: 1) Selecting a city (Berlin, London, Singapore, or New York) to host a conference, and 2) Deciding on post-conference dining options among Italian, Thai, Japanese, or Mediterranean cuisine.

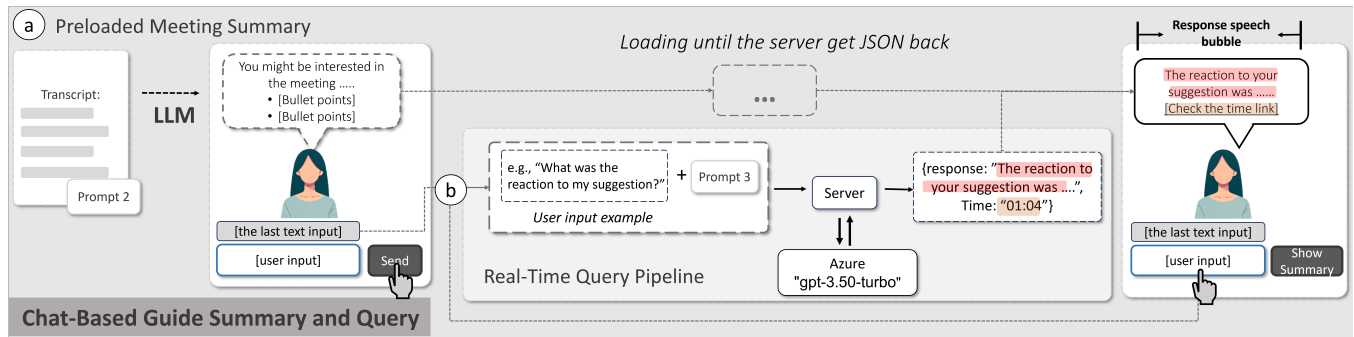


Figure 6: Chat-based Guide Summary and Query (a) depicts how user input text with a prompt is sent to Azure AI, and (b) shows the response JSON coming back from Azure AI and being parsed into the user interface of speech bubble content in real-time.

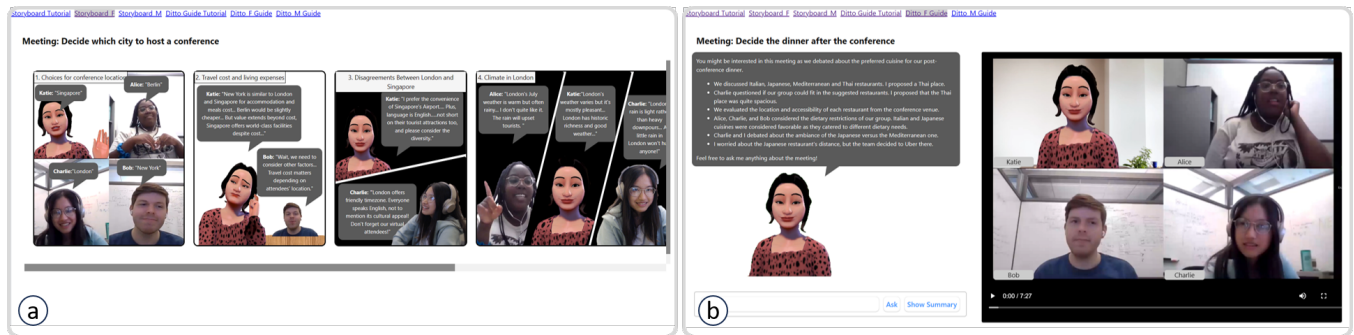


Figure 7: Two Design Probes: (a) Storyboard Interface and (b) the Chat-based Guide.

The meeting scenarios were conducted and recorded using Microsoft Teams, involving four colleagues, with one playing the role of the autonomous agent. The autonomous agent’s representation alternated between female and male avatars based on the participant’s gender, resulting in four distinct 8-minute video recordings. Seven specific avatar behaviors (Dwell, Waving, Thinking, Nodding Head, Talking, Raising Hand, and Showing excitement about a Decision) were integrated into the meeting scenarios, tailored to match the ongoing conversation.

The avatar representations were the same for female and male participants, but the video name tag and speaker label in the Storyboard and Chat-based Guide Tasks were tailored to display the participant’s name. Participants were informed that the avatars in the Storyboard and Chat-based Guide symbolized their agent attending the meeting on their behalf, but did not visually match their appearance or voice for the purpose of the study. The number of participants under each gender group was controlled to counterbalance the effect of the two selected avatar representations.