CoExplorer: Generative AI Powered 2D and 3D Adaptive Interfaces to Support Intentionality in Video Meetings

Gun Woo (Warren) Park
warren@dgp.toronto.edu
Microsoft Research
Cambridge, United Kingdom
University of Toronto
Toronto, Canada

Payod Panda*
payod.panda@microsoft.com
Microsoft Research
Cambridge, United Kingdom

Lev Tankelevitch
lev.tankelevitch@microsoft.com
Microsoft Research
Cambridge, United Kingdom

Sean Rintel
serintel@microsoft.com
Microsoft Research
Cambridge, United Kingdom

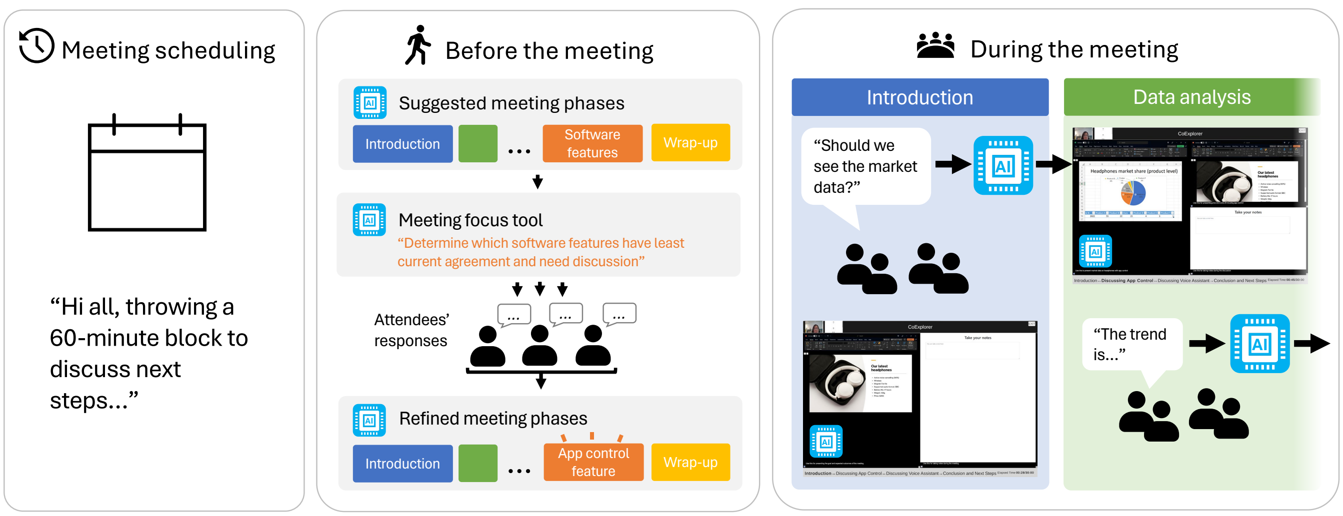

Figure 1: CoExplorer uses Generative AI (GenAI) to simplify the process of achieving meeting objectives. Before the meeting, a GenAI system analyzes an email meeting description and then generates proposed meeting phases as well as a tool for all attendees to identify discussion focus, from which the phases are further refined. In the meeting, the system monitors talk and facilitates transitions between phases, generating applications and optimizing the UI layout to suit each meeting phase.

ABSTRACT

Current online meeting technologies lack holistic support for reducing the effort of planning and running meetings. We present CoExplorer2D and CoExplorerVR, generative AI (GenAI)-driven technology probes for exploring the significant transformative potential of GenAI to augment these aspects of meetings. In each system, before the meeting, these systems generate tools that allow synthesis and ranking of attendees' key issues for discussion, and likely phases that a meeting would require to cover these issues. During the meeting, these systems use speech recognition to generate 2D or VR window layouts with appropriate applications and files for each phase, and recognize the attendees' progress through the meeting's phases. We argue that these probes show the potential of GenAI to contribute to reducing the effort required for planning and running meetings, providing participants with a more engaging and effective meeting experiences.

*Corresponding author

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3650797

CCS CONCEPTS

• **Human-centered computing** → *Collaborative and social computing systems and tools.*

KEYWORDS

meetings, effectiveness, goals, planning, facilitation, videoconferencing, virtual reality, adaptive user interface, windowing, speech recognition, intent recognition, design, generative AI

ACM Reference Format:

Gun Woo (Warren) Park, Payod Panda, Lev Tankelevitch, and Sean Rintel. 2024. CoExplorer: Generative AI Powered 2D and 3D Adaptive Interfaces to Support Intentionality in Video Meetings. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3613905.3650797

1 INTRODUCTION

Productive meetings require effort before, during, and afterwards [64, 73]. However, many organizations neglect formal training for meetings [3], and meeting systems themselves do not provide holistic support for reducing the effort of meeting planning and execution. Generative AI (GenAI) could be transformative in reducing the effort needed to conduct effective meetings. Existing commercial videoconferencing platforms such as Zoom Companion AI [45] and Microsoft Teams Meeting Copilot [55] are already incorporating GenAI capabilities to improve meetings. Nonetheless, we posit that the scope of GenAI's benefits to meeting effectiveness extends far beyond current implementations.

This paper introduces CoExplorer, a technology probe designed to foreground intentionality in distributed meeting systems and challenge classic conceptions of the meeting stage. A shared GenAI infrastructure underpins two interfaces: CoExplorer2D for traditional videoconferencing and CoExplorerVR for virtual reality meetings. CoExplorer uses natural language from a meeting invitation to (a) predict potential meeting objectives and structure, (b) craft a tool to maintain focus on the presumed meeting objectives, and (c) recommend a tailored set of spatially configured applications for distinct meeting phases. CoExplorer autonomously refines the phase outline (i.e., meeting structure) by assessing the context of ongoing tasks, modulating the interactive content. CoExplorer2D assumes a need to fit into traditional 2D meeting interfaces, but uses a dynamic windowing system to adapt content to meeting phases. CoExplorerVR looks forward to 3D virtual reality meetings, and probes new design prospects for the dynamic display of content incorporated into the spatial context. The contributions of this paper are as follows:

- We introduce CoExplorer2D and CoExplorerVR, provocative designs for GenAI-driven meeting systems that adapt the working space in line with meeting goals and contextual activities, and reflect on key opportunities and challenges of adaptive meeting interfaces that support intentionality.
- We provide example prompts that could facilitate design research on adaptive GenAI-driven meeting interfaces.

2 RELATED WORK

Effectiveness of Meetings: Effort is required when planning for effective meetings [17, 33, 58, 77]. Setting goals requires significant time and effort [8, 10, 31]. Lack of understanding of appropriate tasks for meetings is also common [9, 52, 54, 69], and time pressures often lead to inadequate meeting preparation [26, 32]. Running and participating in effective meetings also require effort. Agendas are valuable in structuring meetings [10, 19]. Managing the agenda while allowing flexibility for new ideas and disagreements is important [42, 43, 73, 86]. Determining which topics should be discussed during the meeting versus asynchronously resolved beforehand can reduce agenda length and create space for productive discussion [30, 57, 69]. Additionally, managing a meeting involves shepherding its phases [2], which may align with or go beyond agenda items, and ensuring participants are focused on relevant materials at the appropriate time [23]. Post-meeting effort involves capturing action items or next steps [41, 59, 61], and then acting on them, but if a meeting lacks clear goals or is poorly conducted, post-meeting effort is increased [73]. The effort required for meetings applies to both in-person and online meetings. However, online meetings introduce additional effort.

2D and VR Meetings: The ubiquity of 2D videoconferencing is evidenced by platforms like Microsoft Teams and Zoom. The meeting stage of most commercial systems has been largely unchanged since the 1990s, generally consisting of a grid of people or strip of people adjacent to a single pane of content. The COVID-19 pandemic showed that constant exposure to this stage is not optimized for meeting effectiveness and is also fatiguing [7, 9, 47, 72]. 2D virtual worlds (e.g. Gather [1]) introduce persistent spatial environments that hold some promises for curbing fatigue by virtue of variety of moving into and out of meetings. However, once actually in a meeting, they revert to traditional A/V elements, they make no direct changes to planning or running the meeting, and little change to factors of fatigue [46, 50, 66]. The use of 3D Virtual Reality (VR) meetings is slowly increasing (e.g. in Meta Horizon Workrooms [53], Microsoft Mesh [56]). These VR meetings improve on the sense of person and spatial presence of traditional 2D stages, potentially enhancing overall conversational flow as well as fluid movement between the full meeting and smaller huddles [48, 60, 62, 79]. However, the extent to which VR can be designed to enhance meeting effectiveness is not yet well understood.

Adaptive User Interfaces: Adaptive UIs morph based on user requirements, either during initial setup or throughout their operation, streamlining user efficiency and reducing cognitive strain [4, 12, 27, 28, 51, 78, 85]. Solutions like adaptive windowing systems facilitate immediate access to necessary tools relevant to the task at hand [36, 67, 80, 83, 84]. However, within meetings' specific phases and task-related content, the adoption of adaptive strategies remains scarce. Genre analysis suggests that meeting types could inform tailored UI design, but issues of scale indicated a need for generative AI's capabilities [5, 16].

The spatial environment in 3D VR (and Augmented Reality/physical reality) contexts provides specific contextual cues based on the user's position. For example, standing next to a cooktop implies that this area is intended for interactions related to heating a pot or pan. Although there has been some limited exploration on using these contextual cues to inform the adaptation of placement of VR interactional and informational elements [15, 70], little research has examined how the system can automatically adapt these spatial placements as the meeting progresses.

textbfUsing AI to Prototype Interfaces for Improving Meeting Effectiveness: AI offers tools to streamline the various stages of meetings, including pre-meeting scheduling [17] and agenda item voting [10, 29, 30], decision-making support during the meeting [22, 39], and post-meeting tasks such as summary generation and action item tracking [6, 41, 59, 61, 74] (although we do not focus on post-meeting issues in this paper). Prototyping with LLMs such as GPT-3 simplifies the development of decision-making algorithms, benefiting from their capabilities to process natural language with minimal data [11]. LLMs combined with proactive Voice User Interfaces (VUIs) can provide real-time support, aligning meeting dialogue with predefined objectives [21, 75, 87, 89]. Despite the non-deterministic nature of LLM outcomes, exemplified by ChatGPT's unpredictable predictions, their application in commercial systems suggests a promising avenue for AI enhancement in meeting contexts [14, 25, 35, 40, 49, 68, 88].

In this exploration, we focus on how to design system capabilities that *contribute* to meeting effectiveness. These include reducing struggles of meeting *practice* such as improving pre-meeting articulation of meeting goals and getting attendee buy-in to the focus for agenda items, as well reducing struggles with meeting *technology*, such as making the the right resources available at the right time and in the right place for all attendees for different meeting phases. More direct evaluation on the effect these capabilities have on meeting effectiveness remains for future work.

3 DESIGNING COEXPLORER

Through internal discussion based on our own work experience and prior research, from a large initial set of scenarios we chose one **meeting scenario** that (a) would be familiar for our knowledge worker participants (and, at a high level, knowledge workers more generally), and that (b) would benefit from GenAI-augmented help with planning and running a meeting. We landed on meetings of cross-functional product teams in the technology industry, which often face struggles of coordination [38], especially when planning a course of action about which they have different opinions and stakes, and for which the resources are scattered across different storage locations and apps.

We developed a fictional scenario centered on the need for effective decision-making. In this scenario, a team's current product, the "Strata Headphones 2", is lagging behind competitors in market share, and so the team needs to decide on the features necessary to increase market share of the forthcoming "Strata Headphones 3". The team has files such as the list of design, hardware, and software features, their cost, competitor product information, and the current specification sheet. We used this scenario to formulate three design strategies.

3.1 Design Approaches

Design Approach 1: Incorporate collective feedback to shape meeting objectives. An agenda is typically set by the person organizing the meeting, but should also reflect the priorities of other attendees [82], yet current videoconferencing systems do not easily allow for collaborative crafting of meeting objectives. Participants' contributions can be integrated through mechanisms like voting systems to create or fine-tune the agenda [30]. This design approach explores the potential of a GenAI system to produce a tool that can assimilate varying viewpoints, focusing the conversation on reconciling differences.

Design Approach 2: Clarify underlying needs and available resources. The clear definition of objectives, agendas, preparatory materials, and task-related resources is an essential aspect of efficient meeting preparation and facilitation, notably in virtual settings [18, 44]. Yet, due to time constraints, both meeting planners and participants may cut corners in laying out pre-meeting preparations [76]. During the meeting, while *explicit* items on the agenda direct the discourse, the meeting actually often transitions through *implicit* phases. These phases may align with agenda points but can also encompass several points or be subdivided into finer details. Moreover, essential files and applications are often linked to each action-oriented phase. The implicit nature of these phases makes accessing appropriate resources at the opportune moment burdensome. This design approach explores the challenges and potential benefits of GenAI in identifying implicit phases as they arise during the meeting, and using them to drive a specific arrangement of resources and the meeting's intended progression.

Design Approach 3: Manage the system through a Human-on-the-Loop (HOTL) method. HOTL characterizes human-machine interactions in which the automated system mainly allows humans to *abort* the machine's decisions [63]. This grants automated systems more independence and limits the number of prompts to users, which is compatible with a meeting context where participants often have lower capacity to process information outside the interpersonal interactions of meetings [65]. Introducing HOTL in meeting systems has been impeded by inadequate predictive performance in natural language models, but with the launch of GPT3.5, there has been noticeable improvement in accuracy across a spectrum of fields [14, 40]. Nonetheless, to foster and sustain the trust of users, HOTL systems must clear a higher threshold than the often-preferred human-*in*-the-loop methodology in Human-Computer Interaction [24, 68, 88] [63]. This design approach explores the difficulties and possibilities of building trust with users through HOTL, and how this might influence user perceptions of both efficiency and adaptability.

3.2 Overall implementation:

We implemented CoExplorer using Unity and GPT-3.5. For the real-time video communication (only available for CoExplorer2D), LiveKit was utilized. This choice was made to provide as much cross-compatibility as possible. CoExplorer2D runs on standard Windows computers, while CoExplorerVR runs on Meta Quest 2. See Appendix A for versions of example prompts used with GPT-3.5 to power CoExplorer. In the sections below we detail aspects of the CoExplorer2D and CoExplorer3D systems.

3.3 How CoExplorer2D Facilitates Meetings

3.3.1 Formulating the Initial Phases of the Meeting. The meeting organizer distributes a meeting invitation, and as the meeting time approaches, for each attendee CoExplorer2D outlines the meeting's aim (Figure 3 (A), top left) and describes the basis of that aim (Figure 3A), bottom left). It also suggests phases anticipated for the meeting (Figure 3A), left, shown as segmented bars). Details for each segment include the name of the phase,the expected duration, and the pertinent activities for that phase. Below we expand on how the system implements the design strategies above. The strategies

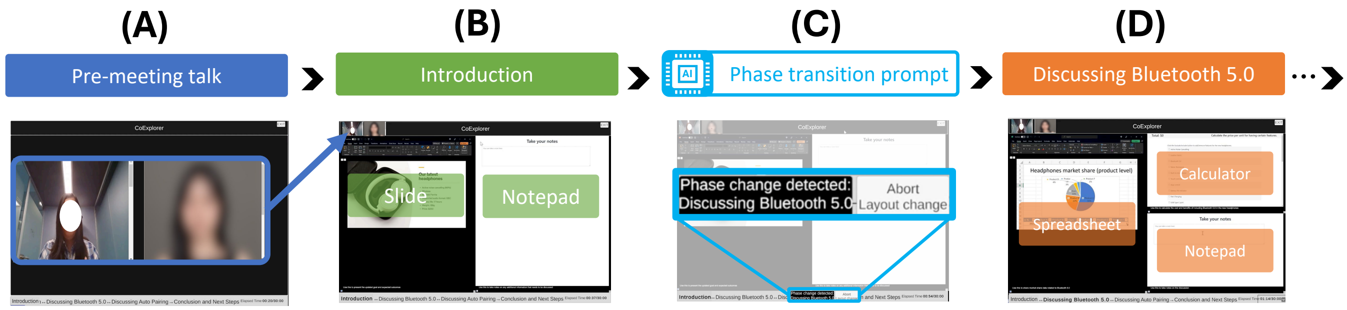

Figure 2: CoExplorer2D listens for meeting attendees transitioning between phases, then notifies the attendees that a change has been detected and modifies the layout to accommodate the new phase.

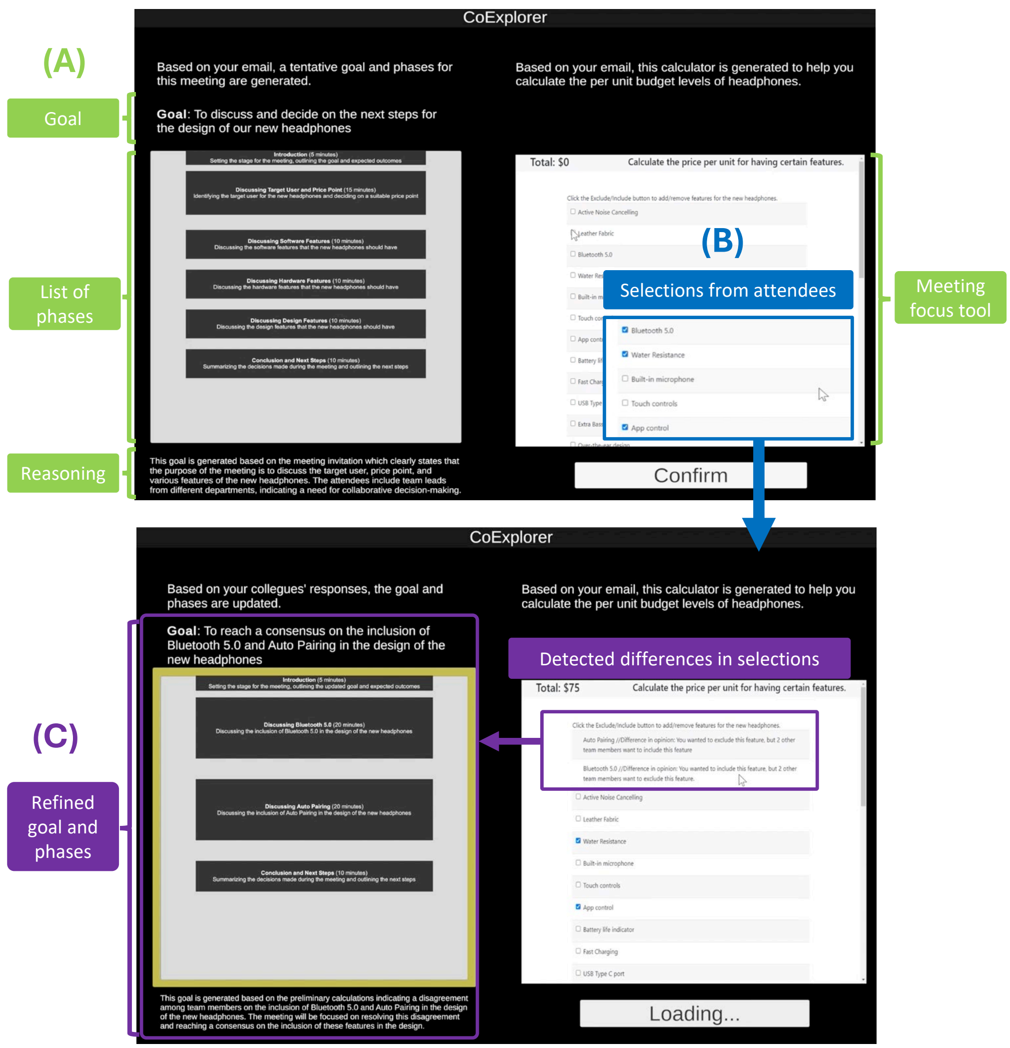

Figure 3: (A) Sequence of phases and tool for determining meeting focus. (B) Meeting attendees employ the focus tool to select preferred features. (C) CoExplorer uses the aggregated preferences to adjust the meeting's objective and phases.

are not implemented in the order presented, and one strategy might be implemented multiple times.

Implementing Design Approach 2: Inspired by the successful use of chaining in Large Language Models, we posited that, given a brief meeting invitation text, GenAI could identify both the purpose of the meeting, and a tailored list of applications for each phase. We emphasize our focus is on exploring the potential applications enabled by GenAI's capabilities, and thus formal evaluation of this capability is out of scope for this paper. GenAI is tasked with elaborating on the provided invitation text and generating a list of phases of the meeting (see Appendix A), complete with titles and descriptions. We then guide GenAI to structure these details into a coherent list that includes explanations for the choices it has made, crafting a narrative that can both inform the CoExplorer2D system and persuade users of the validity of the decisions. To achieve this, we initialise GenAI requests with system prompts, bolstered by examples, to ensure the output is clear and actionable. This preliminary information generated by GenAI is fixed at the beginning and does not undergo refinement during the meeting, intentionally designed to preclude any confusion that might arise from changes during the meeting.

Implementing Design Approach 1: CoExplorer2D also primes participants to consider their own requirements and those of the team by generating a discussion initiator. In this scenario, the discussion initiator takes the form of a Meeting Focus Tool that lets meeting participants assess the implications of including specific features in a product from their role perspective and voice their preferences on key features. This tool is displayed on the right-hand side of Figure 3B as CoExplorer2D creates it. Once all preferences are communicated, as seen in Figure 3B, CoExplorer2D synthesizes these varied responses to refine both the meeting's objectives and phases. CoExplorer2D presents the revised information as exhibited in Figure 3C, which narrows the team's focal points for discussion to areas with the most divergent views. Areas of divergence are chosen for discussion because live meetings are most suited to dynamic discussions, including productive conflict, while areas of agreement can be handled asynchronously or set aside for later [81]. In this scenario, the GenAI system proposed that the tool take the form of a feature ranking aid (Figure 5). GenAI's versatility would also enable a meeting focus tool to manifest in a range of alternative ways, such as a chatbot or a visualization.

During our experiments, we observed that GenAI is indeed proficient at devising a tool for this need. To enable a later evaluation, we opted to employ a pre-generated version of the tool rather than generating it on-the-fly. This decision was made with the intention of ensuring a consistent experience for all users. Upon utilizing the pre-generated tool, we noted certain aspects of the user interface, such as the color scheme used for highlighting selected buttons, could be misinterpreted. Consequently, we implemented manual adjustments to the pre-generated version to refine the prototype for better clarity and functionality.

3.3.2 Dynamic Window Management. Upon examination of the updated phases, the team commences the meeting. CoExplorer2D curates the necessary documents and applications for each phase, generating an ideal layout for the display. Initially, participants engage in a social introduction phase, where their video feeds are maximized (Figure 2A). As the meeting progresses from casual conversation to its formal agenda, CoExplorer2D detects the shift to a new phase, specifically the project introduction. The project manager outlines the problem being addressed.

Implementing Design Approach 2: To process spoken dialogue, we segmented speech into discrete utterances based on pauses. For each utterance, a transcript was produced using the Microsoft Azure Speech API. These transcripts were then provided to GenAI, which determined the pre-identified phase of the meeting the utterance pertained to. If GenAI's prediction indicated a new phase that differed from CoExplorer2D's current phase, the user interface was updated to reflect the new phase, shifting to the window layout associated with it. This layout switch was based on a pregenerated list of window configurations established at the start of the meeting. A small latency was introduced during this process, but since CoExplorer strives to proactively take action, rather than react to user prompts, this latency was not especially noticeable and not directly relevant to the tasks. We expect the latency to decrease as LLMs improve (e.g. see Mixtral 8x7B Instruct[1]).

In organizing the layout of multiple windows, we employed a tiling approach. This decision was influenced by our experimentation with GenAI's ability to generate window sizes and positions. While GPT-4 was adept at creating well-fitted window layouts, GPT-3.5 — the version available to us during prototyping — fell short in this capacity. Consequently, a freeform window layout was deemed unsuitable. We established that a tiling window layout optimally utilized the available screen space within CoExplorer. Tiling thus became the chosen method, ensuring a high degree of screen real-estate utilization across the interface in our prototype.

Implementing Design Approach 3: At this juncture, CoExplorer2D asks all participants on whether they want to halt the transition (Figure 2C). If no objections occur, CoExplorer2D adapts the screen layout to fit this more work-focused phase: video sizes are reduced, with a PowerPoint presentation occupying the left side of the display, and a collaborative notepad on the right (Figure 2B).

In the meeting storyline, the hardware engineer begins discussing the importance of Bluetooth 5.0 for the headphones, picked up on by a software engineer who points out the challenges in supporting this feature through software. Once more, CoExplorer2D senses a phase shift (requesting a confirmation on whether to proceed with the change (Figure 2C). With no opposition, CoExplorer2D transitions to a “Discussing Bluetooth 5.0” phase (Figure 2D). The PowerPoint is replaced with an Excel sheet on the left side, a calculator for the Meeting Focus Tool atop the right side, and the notepad downsized to the bottom right, guiding participants back to the highlighted contentious topic.

A full meeting would proceed this way to a decision and discussion of next steps. Due to time limitations, we did not complete the full meeting scenario in CoExplorer2D. Current Generative AI systems such as Microsoft Copilot are able to detect and outline action items from a transcript, and thus a future CoExplorer-like system could use this in conjunction with adapative windowing to place action items under the video of the relevant person.

[1]https://artificialanalysis.ai/models/mixtral-8x7b-instruct

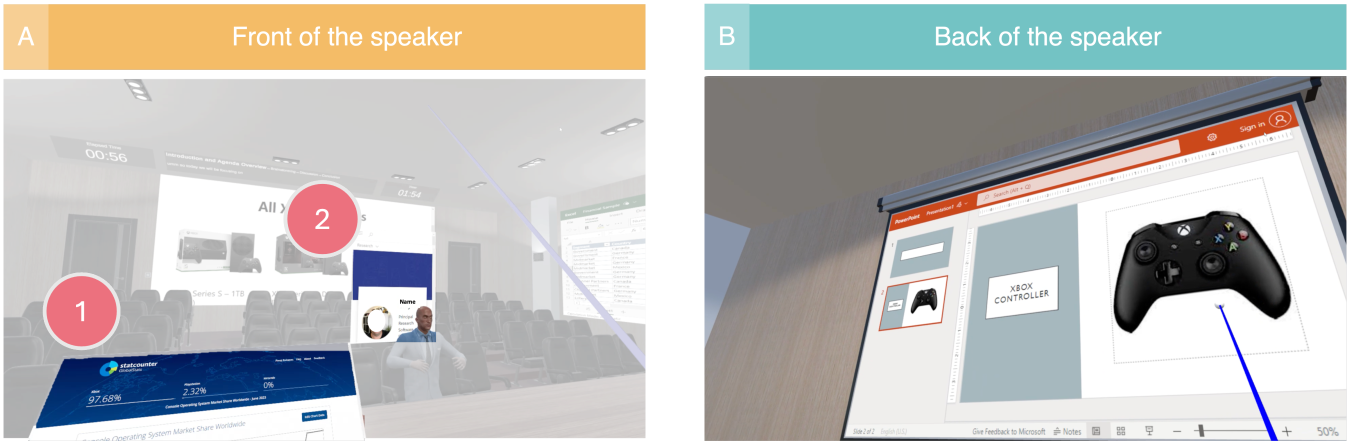

Figure 4: Views using CoExplorerVR. (A) Presenter's view of the audience; (B) Audience's view of the presenter's content.

3.4 Enhanced Spatial Interaction in CoExplorerVR

CoExplorerVR integrates spatial context into its support for seamless meeting facilitation, enhancing the interactions in a 3D virtual environment as depicted in the meeting life cycle (Figure 2). Spatial context refers to the significance attributed to specific locations within the 3D scene [15, 70]. The richer context means that window placement in CoExplorerVR needs to integrate with semantic relevance to the virtual environment more than window positions do in CoExplorer2D. In contrast to the CoExplorer2D, where GenAI optimized window positioning and sizing on a single flat plane, CoExplorerVR expands this functionality into the third dimension, allocating virtual windows to appear around a room in meaningful spatial arrangements.

Visibility: Visibility in a 3D space is a nuanced aspect, as certain objects or windows may only be visible from specific perspectives. A window placed on the main speaker's table (Figure 4A-1) is primarily intended for personal viewing, given the constraints of perspective and field of view. Conversely, a large display positioned behind the speaker (Figure 4B), is intended to be easily visible to the audience.

Object-associated context: In a 3D setting, objects can carry inherent meanings based on their function or placement. CoExplorerVR capitalizes on this by associating contextual information with objects, such as situating a digital contact card at the site of an audience member's chair (Figure 4A-2).

Controllability: The interactivity of objects and information panels in VR is critical. The location of the contact card allows for easy access, yet it also occludes the panel it overlaps, posing a challenge for the ray-based pointer typically used in VR for selection and control (Figure 4A-2). This aspect of spatial controllability must be carefully managed to ensure user comfort and efficiency within the virtual environment.

By incorporating these contextual parameters, the GenAI receives a comprehensive system prompt enabling intelligent window placement that respects the spatial context (given as set of tiles with associated description of the context) and aligns with meeting progression, in parallel to CoExplorer2D's approach.

4 FUTURE WORK

Future Development: GenAI drastically cuts the time and data requirements for developing systems like CoExplorer by eliminating the usual steps of data collection, model training, and testing—instead needing only a few examples. However, validating the outputs from GenAI remains critical for future development. Key challenges include the following:

- **False Positives:** GenAI tends to generate non-null, extensive-change responses due to its sensitivity, leading to inappropriate output. This is exacerbated by training on human-evaluated datasets with inherent errors [13].
- **Processing Time:** Real-time utterance processing is impeded by the slower speeds of Large Language Models (LLMs) compared to current speech-to-text systems.
- **2D/3D Windowing:** Both 2D and 3D display systems offer unique potential for GenAI to enhance user experiences. 2D interfaces benefit from the flexibility in window placement and management, albeit presenting the challenge of handling overlaps and layout dynamics. Our tests found GPT-3.5 struggling with screen space optimization, implying a need for more sophisticated solutions. For 3D displays, adapting to physical contexts for window allocation is critical, requiring further progress in multimodal models.
- **SDK Transition Hurdles:** LiveKit enables real-time video call implementation, but its lack of support for Unity clients led to workarounds. This is particularly problematic for real-time communication due to the critical need for low-latency performance. A more flexible SDK could mitigate these compatibility issues.

Future Evaluation: Our plan is to evaluate CoExplorer2D using a technology probe-based approach [37, 71], which will provide essential insights into the unique opportunities and challenges presented by GenAI-supported meeting systems. Further evaluation of how the ideas in these technology probes might directly impact meeting effectiveness should be conducted using methods similar to that in past work, which includes surveys and telemetry [20, 34].

ACKNOWLEDGMENTS

We acknowledge the use of ChatGPT by OpenAI on the implementation of the prototype, as well as proofreading the text.

REFERENCES

[1] 2023. Gather | Virtual HQ for Remote Teams. https://www.gather.town/
[2] Wendi L. Adair and Jeanne M. Brett. 2005. The Negotiation Dance: Time, Culture, and Behavioral Sequences in Negotiation. *Organization Science* 16, 1 (2005), 33–51. http://www.jstor.org/stable/25145947
[3] Ferah Aksoy-Burkert and Cornelius J. König. 2015. Meeting Training: A Suggestion. In *The Cambridge Handbook of Meeting Science*, Joseph A. Allen, Nale Lehmann-Willenbrock, and Steven G. Rogelberg (Eds.). Cambridge University Press, Cambridge, 69–90. https://doi.org/10.1017/CBO9781107589735.005
[4] Khalid Al-Omar and Dimitrios Rigas. 2009. A user performance evaluation of personalised menus. In *2009 Second International Conference on the Applications of Digital Information and Web Technologies*. IEEE, London, United Kingdom, 104–109. https://doi.org/10.1109/ICADIWT.2009.5273847
[5] Pedro Antunes and Carlos J. Costa. 2003. From Genre Analysis to the Design of Meetingware. In *Proceedings of the 2003 ACM International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '03)*. Association for Computing Machinery, New York, NY, USA, 302–310. https://doi.org/10.1145/958160.958209
[6] 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. https://doi.org/10.48550/arXiv.2307.15793 arXiv:2307.15793 [cs].
[7] Jeremy N. Bailenson. 2021. Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue. *Technology, Mind, and Behavior* 2, 1 (Feb. 2021). https://doi.org/10.1037/tmb0000030
[8] Henning Bang, Synne L. Fuglesang, Mariann R. Ovesen, and Dag Erik Eilertsen. 2010. Effectiveness in top management group meetings: The role of goal clarity, focused communication, and learning behavior: Effectiveness in top management meetings. *Scandinavian Journal of Psychology* 51, 3 (Jan. 2010), 253–261. https://doi.org/10.1111/j.1467-9450.2009.00769.x
[9] Rachel Bergmann, Sean Rintel, Nancy Baym, Advait Sarkar, Damian Borowiec, Priscilla Wong, and Abigail Sellen. 2022. Meeting (the) Pandemic: Videoconferencing Fatigue and Evolving Tensions of Sociality in Enterprise Video Meetings During COVID-19. *Computer Supported Cooperative Work (CSCW)* (Nov. 2022). https://doi.org/10.1007/s10606-022-09451-6
[10] Ana Cristina Bicharra Garcia, John Kunz, and Martin Fischer. 2004. Cutting to the Chase: Improving Meeting Effectiveness by Focusing on the Agenda. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (Chicago, Illinois, USA) *(CSCW '04)*. Association for Computing Machinery, New York, NY, USA, 346–349. https://doi.org/10.1145/1031607.1031664
[11] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. http://arxiv.org/abs/2005.14165 arXiv:2005.14165 [cs].
[12] Peter Brusilovsky. 1996. Methods and techniques of adaptive hypermedia. (1996).
[13] Stephen Casper, Xander Davies, Claudia Shi, Thomas Krendl Gilbert, Jérémy Scheurer, Javier Rando, Rachel Freedman, Tomasz Korbak, David Lindner, Pedro Freire, et al. 2023. Open problems and fundamental limitations of reinforcement learning from human feedback. *arXiv preprint arXiv:2307.15217* (2023).
[14] Lingjiao Chen, Matei Zaharia, and James Zou. 2023. How is ChatGPT's behavior changing over time? http://arxiv.org/abs/2307.09009 arXiv:2307.09009 [cs].
[15] Yi Fei Cheng, Christoph Gebhardt, and Christian Holz. 2023. InteractionAdapt: Interaction-driven Workspace Adaptation for Situated Virtual Reality Environments. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–14.
[16] W.J. Chimiak, R. Rainer, and J. Cook. 1995. An adaptive multi-disciplinary telemedicine system. In *Proceedings of the Fourth International Conference on Image Management and Communication (IMAC 95)*. 244–249. https://doi.org/10.1109/IMAC.1995.532593
[17] Andy Chun, Hon Wai, and Rebecca Y. M. Wong. 2003. Optimizing agent-based meeting scheduling through preference estimation. *Engineering Applications of Artificial Intelligence* 16, 7 (Oct. 2003), 727–743. https://doi.org/10.1016/j.engappai.2003.09.009
[18] Melissa A. Cohen, Steven G. Rogelberg, Joseph A. Allen, and Alexandra Luong. 2011. Meeting design characteristics and attendee perceptions of staff/team meeting quality. *Group Dynamics: Theory, Research, and Practice* 15, 1 (March 2011), 90–104. https://doi.org/10.1037/a0021549
[19] Ross Cutler, Yasaman Hosseinkashi, Jamie Pool, Senja Filipi, Robert Aichner, Yuan Tu, and Johannes Gehrke. 2021. Meeting Effectiveness and Inclusiveness in Remote Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–29. https://doi.org/10.1145/3449247
[20] Ross Cutler, Yasaman Hosseinkashi, Jamie Pool, Senja Filipi, Robert Aichner, Yuan Tu, and Johannes Gehrke. 2021. Meeting effectiveness and inclusiveness in remote collaboration. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–29.
[21] Ritwik Dasgupta. 2018. *Voice User Interface Design: Moving from GUI to Mixed Modal Interaction*. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4125-7
[22] Gert-Jan De Vreede, Robert M. Davison, and Robert O. Briggs. 2003. How a silver bullet may lose its shine. *Commun. ACM* 46, 8 (Aug. 2003), 96–101. https://doi.org/10.1145/859670.859676
[23] Arnulf Deppermann, Reinhold Schmitt, and Lorenza Mondada. 2010. Agenda and emergence: Contingent and planned activities in a meeting. *Journal of Pragmatics* 42, 6 (June 2010), 1700–1718. https://doi.org/10.1016/j.pragma.2009.10.006
[24] Rui Dong, Zhicheng Huang, Ian Iong Lam, Yan Chen, and Xinyu Wang. 2022. WebRobot: Web Robotic Process Automation using Interactive Programming-by-Demonstration. http://arxiv.org/abs/2203.09993 arXiv:2203.09993 [cs].
[25] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 278–288. https://doi.org/10.1145/3025453.3025739 event-place: Denver, Colorado, USA.
[26] Sayed M. Elsayed-Elkhouly, Harold Lazarus, and Volville Forsythe. 1997. Why is a third of your time wasted in meetings? *Journal of Management Development* 16, 9 (Jan. 1997), 672–676. https://doi.org/10.1108/02621719710190185 Publisher: MCB UP Ltd.
[27] Leah Findlater, Karyn Moffatt, Joanna McGrenere, and Jessica Dawson. 2009. Ephemeral adaptation: the use of gradual onset to improve menu selection performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Boston MA USA, 1655–1664. https://doi.org/10.1145/1518701.1518956
[28] Krzysztof Z. Gajos, Katherine Everitt, Desney S. Tan, Mary Czerwinski, and Daniel S. Weld. 2008. Predictability and accuracy in adaptive user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Florence Italy, 1271–1274. https://doi.org/10.1145/1357054.1357252
[29] Ana Cristina Bicharra Garcia, John Kunz, and Martin Fischer. 2003. Meeting Details: Methods to instrument meetings and use agenda voting to make them more effective. In *meeting of the Center for Integrated Facility Engineering, Stanford (no. TR147)*.
[30] Ana Cristina Bicharra Garcia, John Kunz, and Martin Fischer. 2005. Voting on the agenda: the key to social efficient meetings. *International Journal of Project Management* 23, 1 (Jan. 2005), 17–24. https://doi.org/10.1016/j.ijproman.2004.05.003
[31] Jennifer L. Geimer, Desmond J. Leach, Justin A. DeSimone, Steven G. Rogelberg, and Peter B. Warr. 2015. Meetings at work: Perceived effectiveness and recommended improvements. *Journal of Business Research* 68, 9 (Sept. 2015), 2015–2026. https://doi.org/10.1016/j.jbusres.2015.02.015
[32] Paul Geyer. 2020. Do directors prepare well for board meetings? *Governance Directions* 72, 6 (Aug. 2020), 272–276. https://doi.org/10.3316/informit.263991955412630 Publisher: Governance Institute of Australia.
[33] Michael Gibbs, Friederike Mengel, and Christoph Siemroth. 2021. Work from Home & Productivity: Evidence from Personnel & Analytics Data on IT Professionals. https://doi.org/10.2139/ssrn.3843197
[34] Yasaman Hosseinkashi, Jamie Pool, Lev Tankelevitch, Ross Cutler, and Chinmaya Madan. 2023. Meeting effectiveness and inclusiveness: large-scale measurement, identification of key features, and prediction in real-world remote meetings. *arXiv preprint arXiv:2304.00652* (2023).
[35] Steven Houben, Jo Vermeulen, Kris Luyten, and Karin Coninx. 2012. Co-activity manager: integrating activity-based collaboration into the desktop interface. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. ACM, Capri Island Italy, 398–401. https://doi.org/10.1145/2254556.2254633
[36] Donghan Hu and Sang Won Lee. 2020. Scrapbook: Screenshot-based Bookmark for Effective Curation of Digital Resources. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*. ACM, Virtual Event USA, 46–48. https://doi.org/10.1145/3379350.3416181
[37] Hilary Hutchinson, Wendy Mackay, Bosse Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, Björn Eiderbäck, Sinna Lindquist, and Yngve Sundblad. 2003. Technology Probes: Inspiring Design for and with Families. *NEW HORIZONS* 5 (2003).
[38] Avan R. Jassawalla and Hemant C. Sashittal. 2000. Cross-Functional Dynamics in New Product Development. *Research-Technology Management* 43, 1 (2000), 46–49. https://doi.org/10.1080/08956308.2000.11671331
[39] Starr Roxanne Hiltz Jerry Fjermestad. 2000. Group Support Systems: A Descriptive Evaluation of Case and Field Studies. *Journal of Management Information*

Systems 17, 3 (2000), 115–159. https://doi.org/10.1080/07421222.2000.11045657

[40] Douglas Johnson, Rachel Goodman, J Patrinely, Cosby Stone, Eli Zimmerman, Rebecca Donald, Sam Chang, Sean Berkowitz, Avni Finn, Eiman Jahangir, Elizabeth Scoville, Tyler Reese, Debra Friedman, Julie Bastarache, Yuri Van Der Heijden, Jordan Wright, Nicholas Carter, Matthew Alexander, Jennifer Choe, Cody Chastain, John Zic, Sara Horst, Isik Turker, Rajiv Agarwal, Evan Osmundson, Kamran Idrees, Colleen Kiernan, Chandrasekhar Padmanabhan, Christina Bailey, Cameron Schlegel, Lola Chambless, Mike Gibson, Travis Osterman, and Lee Wheless. 2023. *Assessing the Accuracy and Reliability of AI-Generated Medical Responses: An Evaluation of the Chat-GPT Model*. preprint. In Review. https://doi.org/10.21203/rs.3.rs-2566942/v1

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

[42] David J. Kocsis, Gert-Jan de Vreede, and Robert O. Briggs. 2015. *Designing and Executing Effective Meetings with Codified Best Facilitation Practices*. Cambridge University Press, 483–503. https://doi.org/10.1017/CBO9781107589735.021

[43] Kiron Koshy, Alison Liu, Katharine Whitehurst, Buket Gundogan, and Yasser Al Omran. 2017. How to hold an effective meeting. *International Journal of Surgery Oncology* 2, 5 (June 2017), e22–e22. https://doi.org/10.1097/IJ9.0000000000000022

[44] Liana Kreamer, George Stock, and Steven Rogelberg. 2021. Optimizing Virtual Team Meetings: Attendee and Leader Perspectives. *American Journal of Health Promotion* 35, 5 (June 2021), 744–747. https://doi.org/10.1177/08901171211007955e

[45] Lacretia Taylor. 2023. Zoom introduces Zoom AI Companion — available at no additional cost with paid Zoom user accounts. https://news.zoom.us/zoom-ai-companion/

[46] Celine Latulipe and Amy De Jaeger. 2022. Comparing Student Experiences of Collaborative Learning in Synchronous CS1 Classes in Gather.Town vs. Zoom. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2022)*. Association for Computing Machinery, New York, NY, USA, 411–417. https://doi.org/10.1145/3478431.3499383

[47] Sara R. Berzenski Lauren E. Knox and Stefanie A. Drew. 2023. Measuring Zoom Fatigue in College Students: Development and Validation of the Meeting Fatigue Scale for Videoconferencing (MFS-V) and the Meeting Fatigue Scale for In-Person (MFS-I). *Media Psychology* 0, 0 (2023), 1–33. https://doi.org/10.1080/15213269.2023.2204529

[48] Duc Anh Le, Blair Maclntyre, and Jessica Outlaw. 2020. Enhancing the Experience of Virtual Conferences in Social Virtual Environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 485–494. https://doi.org/10.1109/VRW50115.2020.00101

[49] Jiannan Li, Maurício Sousa, Karthik Mahadevan, Bryan Wang, Paula Akemi Aoyagui, Nicole Yu, Angela Yang, Ravin Balakrishnan, Anthony Tang, and Tovi Grossman. 2023. Stargazer: An Interactive Camera Robot for Capturing How-To Videos Based on Subtle Instructor Cues. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–16. https://doi.org/10.1145/3544548.3580896

[50] Chung Kwan Lo and Yanjie Song. 2023. A Scoping Review of Empirical Studies in Gather.town. In *2023 11th International Conference on Information and Education Technology (ICIET)*. 1–5. https://doi.org/10.1109/ICIET56899.2023.10111430

[51] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2013. Patina: dynamic heatmaps for visualizing application usage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 3227–3236. https://doi.org/10.1145/2470654.2466442

[52] Matt Martin. 2020. The State of Meetings in 2020 | Clockwise. https://www.getclockwise.com/blog/the-state-of-meetings-in-2020

[53] Meta. 2024. Horizon Workrooms virtual office and meetings. https://forwork.meta.com/horizon-workrooms/

[54] Microsoft. [n. d.]. In the Office, It's All About Moments That Matter. https://www.microsoft.com/en-us/worklab/in-the-office-it-is-all-about-moments-that-matter

[55] Microsoft. 2023. Get started with Copilot in Microsoft Teams meetings - Microsoft Support. https://products.support.services.microsoft.com/en-us/office/get-started-with-copilot-in-microsoft-teams-meetings-0bf9dd3c-96f7-44e2-8bb8-790bedf066b1

[56] Microsoft. 2024. Introducing Microsoft Mesh | Connect like never before. https://www.microsoft.com/en-us/microsoft-teams/microsoft-mesh

[57] Shaila M. Miranda and Robert P. Bostrom. 1993. The Impact of Group Support Systems on Group Conflict and Conflict Management. *Journal of Management Information Systems* 10, 3 (1993), 63–95. https://doi.org/10.1080/07421222.1993.11518011

[58] Lillio Mok, Lu Sun, Shilad Sen, and Bahareh Sarrafzadeh. 2023. Challenging but Connective: Large-Scale Characteristics of Synchronous Collaboration Across Time Zones. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 611, 17 pages. https://doi.org/10.1145/3544548.3581141

[59] Thomas P. Moran, Leysia Palen, Steve Harrison, Patrick Chiu, Don Kimber, Scott Minneman, William van Melle, and Polle Zellweger. 1997. "I'll get that off the audio": a case study of salvaging multimedia meeting records. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI '97)*. Association for Computing Machinery, New York, NY, USA, 202–209. https://doi.org/10.1145/258549.258704

[60] Catarina Moreira, Francisco PM Simões, Mark JW Lee, Ezequiel R Zorzal, Robert W Lindeman, João Madeiras Pereira, Kyle Johnsen, and Joaquim Jorge. 2022. Toward VR in VR: Assessing Engagement and Social Interaction in a Virtual Conference. *IEEE Access* 11 (2022), 1906–1922.

[61] William Morgan, Pi-Chuan Chang, Surabhi Gupta, and Jason M. Brenier. 2009. Automatically detecting action items in audio meeting recordings. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue (SigDIAL '06)*. Association for Computational Linguistics, USA, 96–103.

[62] Miriam Mulders and Raphael Zender. 2021. An Academic Conference in Virtual Reality?-Evaluation of a SocialVR Conference. In *2021 7th International Conference of the Immersive Learning Research Network (iLRN)*. 1–6. https://doi.org/10.23919/iLRN52045.2021.9459319

[63] Saeid Nahavandi. 2017. Trusted Autonomy Between Humans and Robots: Toward Human-on-the-Loop in Robotics and Autonomous Systems. *IEEE Systems, Man, and Cybernetics Magazine* 3, 1 (Jan. 2017), 10–17. https://doi.org/10.1109/MSMC.2016.2623867

[64] Isabelle Odermatt, Cornelius J. König, and Martin Kleinmann. 2015. Meeting Preparation and Design Characteristics. In *The Cambridge Handbook of Meeting Science*, Joseph A. Allen, Nale Lehmann-Willenbrock, and Steven G. Rogelberg (Eds.). Cambridge University Press, Cambridge, 49–68. https://doi.org/10.1017/CBO9781107589735.004

[65] Eyal Ofek, Shamsi T Iqbal, and Karin Strauss. 2013. Reducing Disruption from Subtle Information Delivery during a Conversation: Mode and Bandwidth Investigation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013).

[66] Pedro R. Palos-Sanchez, Pedro Baena-Luna, and Daniel Silva-O'Connor. 2023. Exploring employees' beliefs regarding the potential benefits of virtual worlds for group cohesion: gather town. *Multimedia Tools and Applications* 82, 16 (July 2023), 24943–24965. https://doi.org/10.1007/s11042-022-14308-7

[67] Seonwook Park, Christoph Gebhardt, Roman Rädle, Anna Feit, Hana Vrzakova, Niraj Dayama, Hui-Shyong Yeo, Clemens Klokmose, Aaron Quigley, Antti Oulasvirta, and Otmar Hilliges. 2018. AdaM: Adapting Multi-User Interfaces for Collaborative Environments in Real-Time. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3173574.3173758 arXiv:1803.01166 [cs].

[68] Kevin Pu, Rainey Fu, Rui Dong, Xinyu Wang, Yan Chen, and Tovi Grossman. 2022. SemanticOn: Specifying Content-Based Semantic Conditions for Web Automation Programs. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. ACM, Bend OR USA, 1–16. https://doi.org/10.1145/3526113.3545691

[69] Qatalog. [n. d.]. Killing Time Report | language.work. https://language.work/research/killing-time-at-work/

[70] Xun Qian, Fengming He, Xiyun Hu, Tianyi Wang, Ananya Ipsita, and Karthik Ramani. 2022. Scalar: Authoring semantically adaptive augmented reality experiences in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[71] Leon Reicherts, Gun Woo Park, and Yvonne Rogers. 2022. Extending Chatbots to Probe Users: Enhancing Complex Decision-Making Through Probing Conversations. In *Proceedings of the 4th Conference on Conversational User Interfaces*. ACM, Glasgow United Kingdom, 1–10. https://doi.org/10.1145/3543829.3543832

[72] René Riedl. 2022. On the stress potential of videoconferencing: definition and root causes of Zoom fatigue. *Electronic Markets* 32, 1 (March 2022), 153–177. https://doi.org/10.1007/s12525-021-00501-3

[73] Steven G. Rogelberg. 2019. *The Surprising Science of Meetings: How You Can Lead Your Team to Peak Performance*. Oxford University Press, New York.

[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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 160 (jan 2018), 24 pages. https://doi.org/10.1145/3161186

[75] Maria Schmidt, Wolfgang Minker, and Steffen Werner. 2020. USER ACCEPTANCE OF PROACTIVE VOICE ASSISTANT BEHAVIOR. *Konferenz Elektronische Sprachsignalverarbeitung* (2020).

[76] Ava Elizabeth Scott, Lev Tankelevitch, and Sean Rintel. 2024. Mental Models of Meeting Goals: Supporting Intentionality in Meeting Technologies. (2024). https://doi.org/10.1145/3613904.3642670

[77] Sandip Sen and Edmund H. Durfee. 1998. A Formal Study of Distributed Meeting Scheduling. *Group Decision and Negotiation* 7, 3 (May 1998), 265–289. https://doi.org/10.1023/A:1008639617029

[78] Anil Shankar, Sushil J. Louis, Sergiu Dascalu, Linda J. Hayes, and Ramona Houmanfar. 2007. User-context for adaptive user interfaces. In *Proceedings of the*

12th international conference on Intelligent user interfaces*. ACM, Honolulu Hawaii USA, 321–324. https://doi.org/10.1145/1216295.1216357

[79] Simardeep Singh, Sylvie Dijkstra-Soudarissanane, and Simon Gunkel. 2022. Engagement and Quality of Experience in Remote Business Meetings: A Social VR Study. In *Proceedings of the 1st Workshop on Interactive EXtended Reality* (Lisboa, Portugal) *(IXR '22)*. Association for Computing Machinery, New York, NY, USA, 77–82. https://doi.org/10.1145/3552483.3556457

[80] Greg Smith, Patrick Baudisch, George Robertson, Mary Czerwinski, Brian Meyers, Daniel Robbins, and Donna Andrews. 2003. GroupBar: The TaskBar evolved. (2003).

[81] Barbara J. Streibel. 2003. *The manager's guide to effective meetings.* McGraw-Hill, New York.

[82] Jan Svennevig. 2012. The agenda as resource for topic introduction in workplace meetings. *Discourse Studies* 14, 1 (Feb. 2012), 53–66. https://doi.org/10.1177/1461445611427204

[83] Craig Tashman. 2006. WindowScape: a task oriented window manager. In *Proceedings of the 19th annual ACM symposium on User interface software and technology.* ACM, Montreux Switzerland, 77–80. https://doi.org/10.1145/1166253.1166266

[84] Craig Tashman and W. Keith Edwards. 2012. WindowScape: Lessons learned from a task-centric window manager. *ACM Transactions on Computer-Human Interaction* 19, 1 (March 2012), 1–33. https://doi.org/10.1145/2147783.2147791

[85] Kashyap Todi, Gilles Bailly, Luis Leiva, and Antti Oulasvirta. 2021. Adapting User Interfaces with Model-based Reinforcement Learning. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* ACM, Yokohama Japan, 1–13. https://doi.org/10.1145/3411764.3445497

[86] Ambreen Usmani, Rehana Rehman, Shazia Babar, and Azam Afzal. 2012. Impact of structured meetings on the learning of faculty members. 26, 03 (2012).

[87] Jing Wei, Tilman Dingler, and Vassilis Kostakos. 2021. Understanding User Perceptions of Proactive Smart Speakers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (Dec. 2021), 1–28. https://doi.org/10.1145/3494965

[88] Haijun Xia, Tony Wang, Aditya Gunturu, Peiling Jiang, William Duan, and Xiaoshuo Yao. 2023. CrossTalk: Intelligent Substrates for Language-Oriented Interaction in Video-Based Communication and Collaboration. https://doi.org/10.1145/3586183.3606773 arXiv:2308.03311 [cs].

[89] Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Voelkel, Johannes Schoening, Rainer Malaka, and Yvonne Rogers. 2022. Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma. In *Proceedings of the 4th Conference on Conversational User Interfaces.* ACM, Glasgow United Kingdom, 1–14. https://doi.org/10.1145/3543829.3543834

A APPENDIX

We attach versions of the example prompts for GPT 3.5-Turbo that we used for implementating CoExplorer. CoExplorer generated prompts automatically based on the templates. During this, CoExplorer optimized the prompts as per the user inputs, and prompted the language model multiple times to obtain the results in the correct format. The prompts given is one instantiation of output prompts that the CoExplorer was producing.

A.1 Phase Generation

A.1.1 Example system prompt.

```
[No prose]
[Output only JSON]
Do not write normal text.
You are a JSON generator which converts meeting agenda text into a more descriptive agenda description.
You always need to have an introduction phase at the beginning.
```

A.1.2 Example user prompt.

```
Please break down the following meeting agenda that someone has sent in email into meeting phases that we would need to go through.
We have a 60 minute meeting scheduled.
Based on the given information, give a goal of the meeting (goal), as well as the explanation on why you chose the goal (exp).
And also give the phase definition in a list (pi).
Each phase definition should include: (1) Phase title (pt), (2) Phase description (pd) which should include a sub goal of a phase, (3) Behaviors to be encouraged (be), (4) Behaviors to be discouraged (bd), (5) priority (p), (6) amount of time allocation (t), (7) direction (d) (i.e., is it an iterative phase or directional phase).
Please only respond in JSON with each element needed as a key within a phase,
i.e., if we have two phases,
[{"pt":"xxx", "pd":"xxx",
"be":["xxx","yyy"],"bd":["xxx","yyy"],
"p":"high","t":2,"d":"iterative"},
{"pt":"yyy","pd":"zzz","be":["ttt","kk"],
"bd":["lll","mmm"],
"p":"low","t":8,"d":"directional"}].
So the overall JSON to export is {"goal":"xxx","pi":
[<phase definitions>],"exp":"xxx"}.
Explanation should start by saying "this goal
is generated..." or similar.
Please use the full 60 minutes.
Here is the meeting invitation:
```

(The email invitation would be attached.)

Refinement could be done using the same script, attaching the refinement scenario. This script can be generated through code, and does not have to be detailed.

A.2 Layout Generation

A.2.1 Example system prompt.

```
[No prose]
[Output only JSON]
Do not write normal text.
You are a helpful assistant who creates screen layout that has appropriate apps that are most helpful for users to complete the task successfully.
Respond only in JSON following the format.
Example format:
[{"PhaseTitle":"xxx","timer":n,"programList":
[{"name":"yyy,"description":"zzz"},
{"name:"kk","description":"lll"}]},
{"PhaseTitle":"xxx","timer":n,"programList":
[{"name":"yyy,"description":"zzz"},
{"name:"kk","description":"lll"}]}].
Strictly follow this format.
n is integer, and programList.name should either be a name of a program in a program list given or a URL.
programList.description is where you should put
an extremely concise reason why you chose that program.
e.g., Use this for presenting agenda;
Use this for viewing relevant budget data
```

A.2.2 Example user prompt.

I will give you the list of phases in a meeting in JSON format.
Each phase in JSON is defined with several keywords.
"pt" represents phase title, "pd" represents phase description, "be" represents behaviours to be encouraged, "bd" represents the behaviours to be discouraged, "p" represents priority (high, medium, low"), "t" represents recommended duration of time for the phase, and "d" represents directionality (directional i.e., cannot be returned, and should be preceeded by a certain phase or iterative i.e., can be transitioned to this whenever).
You need to generate what kind of programs are needed for helping goals of each phase (defined by the description) to be met the most efficiently.
You can generate a list of 1-5 program name/URL, and the sequence of generation will affect where they are being placed, and size.
Therefore, you need to be sensible about ordering so that important programs can be shown with the bigger sizes.
The rule is as follows:
If you have one program on the list, that would be full screen. If you have two programs, it would be one on the left half (first program on the list), and one on the right half.
If you have three programs, the right-hand-side panel will split in half, creating two small panels at the top and the bottom.
Four programs mean the left-hand-side panel will also be split. Five programs mean two equally sized panel at the top, and three equally sized panel at the bottom.
The ordering in the list will be used to place program to panels in a clockwise ordering (top left panel is the first panel).
Here is the list of programs available, and if the program that you want is not listed, please generate a URL for the program that you need instead of the program name.
Please feel free to give Bing Search URL with the search term filled, and generate at least one URL:

(List of programs could take arbitrary formats.)

A.3 Meeting Focus Tool

A.3.1 *Example system prompt for a specialized calculator generation.*

You need to generate a HTML page, and only
HTML+CSS+JavaScript based page as a response which allows me to calculate the total value to have the a list of features for a given scenario.
Each feature needs to have an "include" button (green tick) or "exclude" button (red cross), and you need to calculate the total at the end, when the submit is clicked.
You should not show any prices including itemised ones before this.
You need to generate at least 30 features (features should be unique and descriptive. No "Feature 26" or something like that), and incorporate that into the page, those are relevant for the scenario.
Assign random prices for each feature.
The HTML page needs to incorporate all the features that you generated embedded (i.e., no "..." or "many more features here" etc.).
No prose.
No add more features here etc.
You need to list all the features on the HTML.

A.3.2 *Example user prompt.*

A designer, software engineer, hardware engineer, PM, marketing expert, and a researcher are gathering to think about what features that a new headphone product that they release might have.
It could be electronics feature such as active noise cancelling (95\%) or it could be about materials (leather etc.).

Rather than listing explicit features, this user prompt may occasionally result in *"<!–Here will be inserted many more features–>"*. Such cases can be detected, and additional user prompt could be given to force the system to be explicit by pointing out that it has not followed instruction. The prompt could take a form as follows:

It does not show 30+ features

The system would then generate a web page as in Figure 5:

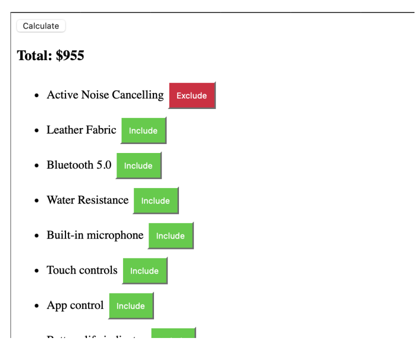

Figure 5: Example raw web app generated by GPT 3.5