EchoMind: Supporting Real-time Complex Problem Discussions through Human-AI Collaborative Facilitation

WEIHAO CHEN, Tsinghua University, China and Ministry of Education, China
CHUN YU*, Tsinghua University, China and Ministry of Education, China
YUKUN WANG, Tsinghua University, China
MEIZHU CHEN, Tsinghua University, China
YIPENG XU, Tsinghua University, China
YUANCHUN SHI, Tsinghua University, China, Qinghai University, China, and Ministry of Education, China

Teams often engage in group discussions to leverage collective intelligence when solving complex problems. However, in real-time discussions, such as face-to-face meetings, participants frequently struggle with managing diverse perspectives and structuring content, which can lead to unproductive outcomes like forgetfulness and off-topic conversations. Through a formative study, we explores a human-AI collaborative facilitation approach, where AI assists in establishing a shared knowledge framework to provide a guiding foundation. We present *EchoMind*, a system that visualizes discussion knowledge through real-time issue mapping. EchoMind empowers participants to maintain focus on specific issues, review key ideas or thoughts, and collaboratively expand the discussion. The system leverages large language models (LLMs) to dynamically organize dialogues into nodes based on the current context recorded on the map. Our user study with four teams (N=16) reveals that EchoMind helps clarify discussion objectives, trace knowledge pathways, and enhance overall productivity. We also discuss the design implications for human-AI collaborative facilitation and the potential of shared knowledge visualization to transform group dynamics in future collaborations.

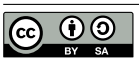
CCS Concepts: • **Human-centered computing → Collaborative and social computing systems and tools**; **Interactive systems and tools**; *Visualization.*

Additional Key Words and Phrases: Group Discussions; Complex Problems; Issue Mapping; Human-AI Collaboration; Large Language Models

ACM Reference Format:
Weihao Chen, Chun Yu, Yukun Wang, Meizhu Chen, Yipeng Xu, and Yuanchun Shi. 2025. EchoMind: Supporting Real-time Complex Problem Discussions through Human-AI Collaborative Facilitation. *Proc. ACM Hum.-Comput. Interact.* 9, 7, Article CSCW406 (November 2025), 38 pages. https://doi.org/10.1145/3757587

*Corresponding author.

Authors' Contact Information: Weihao Chen, Tsinghua University, Department of Computer Science and Technology, Beijing National Research Center for Information Science and Technology, Beijing, China and Ministry of Education, Key Laboratory of Pervasive Computing, Beijing, China, chenwh20@mails.tsinghua.edu.cn; Chun Yu, Tsinghua University, Department of Computer Science and Technology, Beijing National Research Center for Information Science and Technology, Beijing, China and Ministry of Education, Key Laboratory of Pervasive Computing, Beijing, China, chunyu@tsinghua.edu.cn; Yukun Wang, Tsinghua University, Department of Computer Science and Technology, Beijing, China, wang-yk21@mails.tsinghua.edu.cn; Meizhu Chen, Tsinghua University, School of Architecture, Beijing, China, cmz23@mails.tsinghua.edu.cn; Yipeng Xu, Tsinghua University, Department of Computer Science and Technology, Beijing, China, wang-yk21@mails.tsinghua.edu.cn; Yuanchun Shi, Tsinghua University, Department of Computer Science and Technology, Beijing National Research Center for Information Science and Technology, Beijing, China and Qinghai University, Xining, Qinghai, China and Ministry of Education, Key Laboratory of Pervasive Computing, Beijing, China, shiyc@tsinghua.edu.cn.



ACM 2573-0142/2025/11-ARTCSCW406
https://doi.org/10.1145/3757587

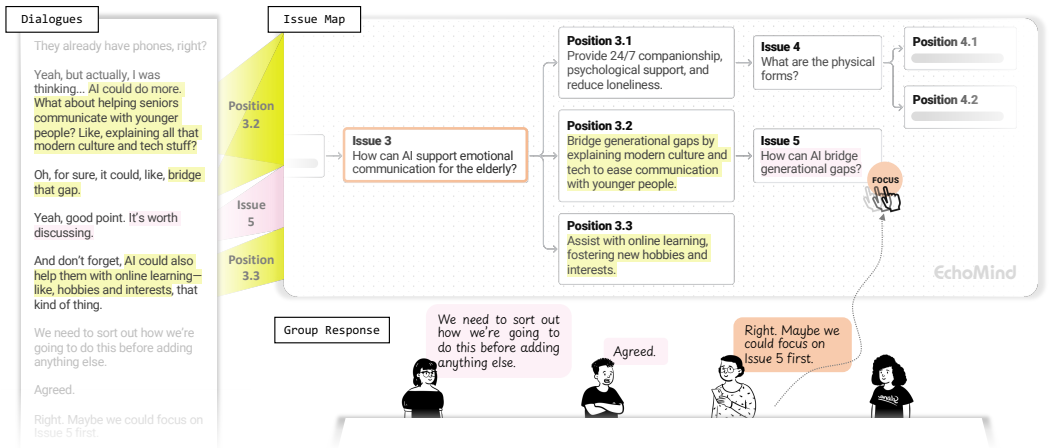


Fig. 1. *EchoMind* supports group discussions by providing a real-time issue map (right) derived from ongoing conversations (left). The system enables the team to focus on specific issues (highlighted with an orange border), review updated positions, and explore deeper issues (below). Busts illustrations from Open Peeps.

1 Introduction

Teams in various fields frequently face complex problems that cannot be solved by individuals alone [35]. These problems arise in diverse contexts, such as startups exploring market opportunities [34, 60], cross-functional groups refining strategies [50], research teams advancing scientific innovation [76], and policymakers addressing societal challenges [20]. Such issues often involve evolving priorities and ambiguous goals – commonly referred to as "wicked problems" [71]. A common approach to addressing these issues is real-time collaborative discussions, where a group of participants contribute their expertise through natural, fast-paced interactions [1, 4] like face-to-face meetings [6] or video conferences [41]. Teams aim to leverage collective intelligence through these dynamic discussions to achieve outcomes greater than individual efforts [18].

Despite its prevalence, managing real-time discussions effectively remains challenging, and poorly facilitated sessions can easily lead to unproductive outcomes [26, 32, 45]. For instance, during a product design meeting, team members may start brainstorming enthusiastically, but without a clear goal or direction, the fast-paced conversation can easily drift off-topic or get sidetracked by unrelated details. Important ideas risk being forgotten in the flow of dialogue, and conflicting viewpoints may cause the discussion to circle around and lose focus. As a result, the team often struggles to synthesize insights and make concrete decisions, ultimately undermining the potential for collective intelligence.

Supporting group discussions and decision-making has been a central focus of research in Computer-Supported Cooperative Work (CSCW) [39, 41, 65, 70, 82, 92, 97]. Prior work has explored using virtual agents as facilitators to guide and structure conversations [7, 27, 28, 51, 61], employing visualizations and interactive elements to enhance remote engagement and coordination [16, 52, 83], and conducting post-hoc analysis of group problem-solving and decision-making behaviors [17, 42, 78, 87, 95]. While these efforts provide valuable insights, they primarily focus on asynchronous or text-based discussions and often rely on predefined agendas or rules to organize conversations. However, real-time synchronous discussions involving complex problem-solving pose unique challenges due to their dynamic and fast-paced nature. Existing studies have not sufficiently addressed these challenges, and more effective facilitation support strategies are required.

This work explores a new model of **human-AI collaborative facilitation**, where AI actively supports human-driven discussions by managing and externalizing shared knowledge. Previous research has emphasized the importance of maintaining shared knowledge in complex problem-solving [72], due to its long-term value for future meetings and extended collaborations [59, 92]. Frameworks like the Issue-Based Information System (IBIS) [53] and the derivative facilitation technique dialogue mapping [18] have demonstrated the effectiveness of structuring and visualizing knowledge to guide discussions [19]. However, these methods heavily rely on manual efforts, placing significant demands on facilitators' expertise [3, 9, 17, 38]. Building on these insights, we aim to explore how AI can contribute to knowledge management within a visual interface to enhance facilitation. Through a formative Wizard-of-Oz study, we identified several design goals for such a system, focusing not only on effective coordination between humans and AI but also on the new challenge of aligning their shared understanding.

Guided by these design goals, we propose *EchoMind*, a collaborative system that generates issue maps based on real-time conversations to provide clarity and guidance throughout complex discussions. EchoMind leverages large language models (LLMs) to dynamically organize evolving thoughts into a structured issue-position map, enhancing shared understanding among participants. We introduced a focus feature that directs participants' attention to specific issue nodes, enabling the system to expand the map at those points. This functionality facilitates alignment among participants and between humans and the system, ensuring a coherent discussion flow. EchoMind also supports user-driven modifications to the map's structure and content, providing cognitive support while preserving flexibility.

In an example use case of EchoMind (see Fig. 1), a team is exploring solutions to a challenging scenario involving elderly care. As the discussion progresses, the team focuses on the issue of how AI can support emotional communication for older adults. EchoMind automatically organizes the relevant conversation into corresponding positions on the issue map and suggests potential new issues for further exploration. The group interacts with EchoMind by proposing new ideas, and under the facilitator's guidance, shifts their focus to deeper issues as the conversation evolves.

To evaluate EchoMind in group discussions, we conducted a user study with four teams of 16 participants engaged in product design discussions. We also compared EchoMind with an automatic summarization document system, similar to existing tools, to better understand its impact. Each team participated in separate sessions, experiencing support from both systems. Results indicate that EchoMind significantly clarified and refined participants' thoughts, enhanced their sense of purpose, and improved discussion quality overall. Notably, the relational structure of the issue map–rather than its specific content–played a crucial role in aiding participants in reviewing knowledge. Additionally, the focus feature facilitated alignment among participants and between humans and the system, while also maintaining a minimal operational burden. These findings inform design implications for human-AI collaborative facilitation, highlighting how effective management of shared understanding can enhance real-time group dynamics and support productive outcomes.

In summary, this work makes the following contributions:

- **Design Goals of Human-AI Collaborative Facilitation**: Through our formative study, we identify challenges in complex problem discussions and summarize the design goals for how AI can facilitate discussions through a visual structure.
- **EchoMind**[1]: We design and implement a collaborative system that leverages large language models (LLMs) to dynamically build issue maps from real-time dialogues, providing a shared knowledge framework that facilitates productive discussions.

[1]The source code is available at https://github.com/atomiechen/EchoMind.

- **Evaluation Results and Design Implications**: We provide empirical evidence from a comparative user study that demonstrates EchoMind's advantages in aligning focus, clarifying knowledge relationships, and improving discussion quality. We then present design implications for future human-AI collaborative facilitation based on these results.

2 Related Work

Effective facilitation is crucial for productive group discussions. This paper focuses on how methods and tools can address common challenges in complex discussions, particularly through the use of visual map structures to enhance facilitation. In this section, we first revisit the principles of group facilitation, then explore tools for managing knowledge in complex discussions, and finally examine the potential of using LLMs for dialogue mapping.

2.1 Group Discussion Facilitation

Facilitating group discussions requires effective methods for addressing complex problems, often referred to as wicked problems [12], which are characterized by uncertainty, conflicting stakeholder interests, and evolving constraints [71]. In general, facilitation strategies involve two key components: structuring content and organizing procedure [9].

Content structuring focuses on providing a clear framework to organize both current and potential discussion topics. Examples include the development of shared mental models, the use of tools like the Issue-Based Information System (IBIS) [53] and concept mapping [66]. These methods aim to help participants navigate complex discussions by breaking down problems into manageable components [38, 40]. Structuring content is critical for mitigating focus drift, which can degrade the quality of information sharing and increase the risk of participants relying on incomplete or inaccurate information, leading to inefficient or erroneous decisions [55, 77, 84].

Organizing procedure builds upon this structured content by introducing multiple phases to guide participants through different aspects of the discussion. Methods like the Nominal Group Technique (NGT) [24, 49] and Six Thinking Hats [23] help participants systematically explore various perspectives, ensuring that important viewpoints are not overlooked. When discussions lack a clear procedural structure, participants must expend additional effort to make sense of scattered or unstructured inputs, resulting in increased cognitive load and diminished capacity for integrating information effectively [81]. While these procedural methods offer professional approaches, they often limit flexibility. In our work, we prioritize content structuring, allowing the facilitator to set and adjust the discussion agenda to ensure adaptability across various domains and types of problems.

Our system draws inspiration from IBIS [53] for mapping dialogues to an explicit diagram. However, a key challenge with these structured strategies is that they require facilitators to be highly skilled in breaking down complex problems, guiding participants in articulating positions, and synthesizing arguments [18]. To address this, our system provides support in these areas, enhancing the facilitator's ability to manage complex discussions and ensure productive outcomes.

2.2 Tools for Visualizing Shared Knowledge

Effective shared understanding and management in group discussions requires visual tools that support the retention of complex ideas. Visualizations provide a structured format that helps participants engage with the evolving dialogue, making it easier to track and integrate information [79].

Manual tools for knowledge management offer various ways to represent information. Text documents are simple and linear but become difficult to navigate as discussions expand, making

it challenging to maintain focus on key relationships [73]. Tables or matrices strengthen the representation of direct relationships but often lead to the neglect of more complex and important connections [73, 80]. Diagrams like mind maps or concept maps allow for hierarchical and associative visualizations, but their effectiveness depends heavily on the facilitator's skill, and manual updates increase cognitive load [62, 85]. The more complex the tool, the higher the cognitive demands on participants, which can disrupt engagement and reduce discussion quality.

In addition to manual tools, there are automated AI tools designed to provide visualizations based on real-time verbal content [44, 64]. For instance, TalkTraces [13] maps discussion topics against a predefined agenda, allowing participants to track how closely the conversation aligns with the intended goals. CrossTalk [94] generates shared panel substrates for video meetings, which serve as dynamic representations of participants' intentions and conversational focus. Existing commercial tools, such as Otter.ai [68], offer segmented summaries and question-answering chatbots based on conversations; however, they still rely on users to manually trigger actions. These tools function primarily as passive recording or assistive tools, limiting their role in discussion facilitation.

In contrast, our system is designed as an active facilitation tool. It combines the strengths of representational tools and AI-driven natural language understanding to collaboratively generate maps with participants for more productive discussions.

2.3 LLMs for Dialog Mapping

Large language models (LLMs) capture statistical relationships in language and have demonstrated impressive performance across a range of natural language processing tasks [8, 89]. They can transform spoken language, which is often more context-dependent than written language, into structured insights [14, 15, 36, 48, 91, 93, 96, 98]. This gives them an advantage over traditional argumentation mining methods, which often require domain-specific tuning and struggle to handle the nuances of conversational language [54, 56].

LLMs are able to identify relationships between entities within the text, which can be used to generate knowledge graphs [10, 88, 99]. For example, systems like Graphologue [46] use inline annotations to construct graphs based on their responses, while GraphRAG [29] leverages LLMs for offline analysis to build entity graphs from documents. However, these tools are not designed for real-time interaction with human users. In our use case, the system operates as a real-time collaborative agent, where the issue map serves as the agent's structured memory. We employ concise prompting strategies by including the current issue path in the prompt to maintain sufficient context without compromising response time.

3 Human-AI Collaborative Facilitation for Real-time Discussions

In group discussions, facilitators play an important role in organizing and guiding conversations toward meaningful outcomes [3, 9, 38]. However, in complex discussions involving multiple perspectives, facilitators face significant challenges. They must process a large volume of conversational information, synchronize participants' understanding, and prevent discussions from straying off-topic [38]. Balancing these tasks requires cycling between advancing the discussion and reflecting on its content to ensure productive outcomes.

To address these challenges, we propose a human-AI collaborative facilitation model that leverages the strengths of both AI and participants in managing discussions. AI excels at recording and summarizing information, providing a computational backbone for maintaining shared knowledge, while human facilitators offer intuition and contextual understanding to guide conversations.

This approach shifts from implicit knowledge stored in human minds to explicit visualization, making information accessible and transparent for all participants, such as through shared documents or meeting notes. However, linear formats like sequential transcripts or condensed summaries

struggle to represent complex issues and may not effectively support real-time facilitation. To overcome these limitations, we propose using explicit map-based structures as a shared framework for organizing and navigating discussions. By visualizing the relationships between key points, we believe this approach has the potential to help facilitators keep the discussion on track while enabling participants to clarify the context.

Therefore, we conducted a formative study to explore how an AI-generated visual map can enhance facilitation, understanding, and engagement during group discussions. This study provides valuable insights for subsequent system design.

3.1 Study Design

We adopted a Wizard-of-Oz design [21], in which human wizards dynamically created visual maps based on ongoing group discussions. We allowed wizards to choose representations that they believed would best aid participants in understanding the key elements of the discussion. This freedom aimed to uncover the varying needs of different users, and we anticipated that this flexibility might result in different map structures.

Participants. We invited research groups because of their experience in engaging in-depth, collaborative problem-solving, which aligns with the type of discussions we aim to study. A total of 7 participants (2 male, 5 female, aged 20 to 30) from two teams took part in the study discussions. Each team included one leader. Two wizards (both male, aged 26), who were external to these teams, were responsible for one discussion each. Both wizards were familiar with the basic operations of drawing maps on the canvas.

Setup. We seated participants in a meeting room equipped with a large screen, while a human wizard sat in another room with a PC. We connected the wizard's PC and the meeting room's large screen to a shared Figma[2] canvas, allowing participants to view real-time content updated by the wizard. The wizard listened to the discussion audio through remote conferencing software, using a single microphone in the meeting room to follow the conversation.

Procedure. Before the discussion, we conducted a brief individual interview with each participant. The discussion lasted about 45 minutes and focused on real topics that the research group needed to address in the near future. After the discussion, we held a group interview with the discussion team, and the wizard participated in a separate individual interview to gather feedback.

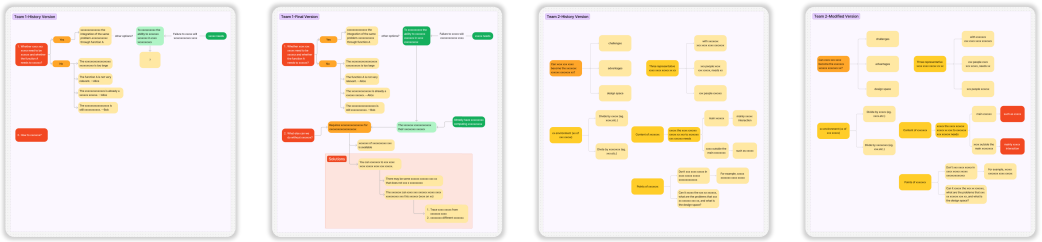
Discussion Task. In each discussion, the group leader acted as the facilitator. We required participants to refer to the AI-generated map on the screen as much as possible during the discussion, and manually modify it when necessary to move the discussion forward. The facilitator verbally instructed the AI to generate or update the map. Only the facilitator could operate the meeting room PC and modify the canvas content.

Wizard Task. We instructed the wizard to update the map as the conversation progressed, reflecting key information such as important topics, relationships, and decisions, based on their understanding of the discussion. To manage the speed of map generation, the wizard could type "processing" on the canvas to indicate they were working on an update. After completing the update, the wizard deleted the message.

3.2 Findings

Our findings highlight that the visualization map proved valuable, particularly in complex scenarios. Participants found that a visual representation of discussion content facilitated better understanding and navigation. During the study, participants exhibited notable behaviors of clarification and alignment when interacting with the map. Facilitators actively utilized the map to steer conversations

[2]https://www.figma.com/



Team 1 Snapshot 1 **Team 1 Snapshot 2** **Team 2 Snapshot 1** **Team 2 Snapshot 2**

Wizard: Custom visualization with vibrant nodes, complex structure, multi-directional arrows, and layered sections.
Participant Feedback: Difficult to understand due to unconventional structure.

Wizard: Oversimplified node content, with significant delays in update.
Participant Feedback: More clarifying explanations were needed, along with a desire for real-time interaction.

Fig. 2. Example snapshots from both teams illustrating different approaches in map visualization and corresponding participant feedback.

and outline subsequent agenda items, indicating that it can significantly enhance the facilitation process by providing clarity and direction rather than merely serving as a static record. Example snapshots and related participant feedback are shown in Fig. 2.

Despite these advantages, generating and using the map presents several challenges. A core issue is that the interaction between humans and AI requires clear protocols to avoid cognitive overload. We summarize the following design goals, with DG1 and DG2 focusing on requirements for AI generation, while DG3 and DG4 address user participation needs.

DG1: Limit node types and impose connection constraints. A key design principle is to minimize cognitive load, as users can easily become overwhelmed by excessive complexity. [74]. In Team 1, the wizard employed a custom method to create the map, but participants often found the complex color schemes and connection structures confusing. The wizard's misunderstanding of the conversation further exacerbated participants' confusion. In contrast, the Team 2 wizard used a simpler structure, which participants found much easier to understand. Therefore, it is necessary to minimize the learning curve for users by limiting the node types to common ones such as issues, topics, and decisions, and applying connection constraints that reflect typical relationships, such as cause-and-effect, correlation, progression, or inclusion.

DG2: Avoid oversimplification of text summaries and maintain context. This design goal is based on the principle of preserving semantic richness while simplifying the visual structure. Some participants in Team 2 reported that the wizard retained too little text content at each node (e.g., only a few keywords), which blurred the interpretation of the visual map and led to ambiguity. They suggested that each node could use a short sentence, which was similar to the node description style in Team 1. Although the structure of the map itself contained rich information, retaining sufficient natural language within each node is essential, particularly for local or secondary details.

DG3: Allow users to clearly anticipate where content changes will occur. This design goal is to facilitate transparency and predictability of the AI system [5]. In this study, the wizard could update the map at any location, which made it difficult for participants to track changes. Participants expressed a preference for AI content updates to occur in designated positions or specified areas, reducing the additional cognitive load of searching for updates.

DG4: Provide users with editorial control to avoid conflicts. Participants may need to modify the content or structure generated by the wizard due to errors or changes in discussion priorities. However, they prefer that subsequent automatic updates do not override these changes.

To ensure user control, the system should allow users to retain primary authority over the content, while AI provides supportive assistance [5].

These design goals balance the complementary strengths of AI and humans, aiming to create a seamless, collaborative environment where both can effectively contribute to complex discussions.

4 EchoMind

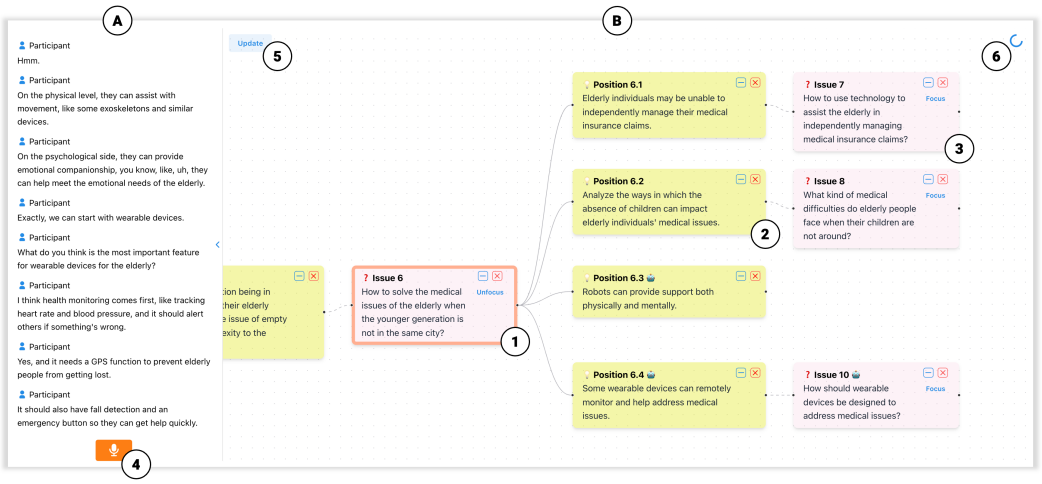


Fig. 3. The interface of EchoMind includes transcription area (A) and issue map area (B). Key interactive components are as follows: issue node (1), position node (2), related issue node (3) connected by a dashed line to the position node, recording switch (4), optional manual update button for user-initiated updates (5), and indicator of updating status (6).

Guided by the formative study and the design goals for human-AI collaborative facilitation, we introduce EchoMind, a system that embodies this collaborative model by generating real-time issue maps to support productive and structured discussions. EchoMind addresses the challenges identified in complex discussions, functioning not merely as a passive recording tool but as an active assistant that fosters engagement and clarity.

As shown in Fig. 3, EchoMind provides a shared visual interface that helps participants track the flow of the discussion and manage structured content with flexibility. To promote clearer communication and quicker consensus, EchoMind employs an issue-position structure with limited node types and connection constraints (DG1). By combining automatic speech recognition (ASR) with the natural language understanding capabilities of large language models (LLMs), the system captures participants' ideas and progressively builds an evolving knowledge structure with sufficient context to be accessible and actionable for all participants (DG2). Additionally, EchoMind explicitly marks the focused issue node to maintain alignment among participants and the system (DG3). This feature enables users to stay oriented within the discussion, while maintaining control over edits and decisions (DG4), ensuring a seamless integration of human-AI collaboration.

In this section, we will elaborate on the design rationale and key features of EchoMind, including the real-time generation pipeline powered by LLMs and the issue map structure. We provide additional technical details in Appendix B.

4.1 Issue Map Structure

EchoMind's knowledge representation is issue-centered, aiming to keep discussions focused on key problems and solutions. We base our approach on the Issue-Based Information System (IBIS) model [53] for several reasons. First, a key challenge in complex problem discussions is balancing structure with semantic completeness, and the IBIS framework provides a validated method for this [11, 30]. Second, IBIS supports a clear visual structure with limited node types and connection constraints, which can be easily customized to meet our design goals. Finally, by replacing manual map creation with AI generation, we overcome the practical limitations of IBIS, which prior studies have discussed in the context of human constraints such as time pressure and oversimplified node content [11, 19].

IBIS traditionally includes three node types: issues, positions, and arguments. But EchoMind simplifies this by using only issue and position nodes, based on findings from our formative study (DG1). This simplification avoids cognitive overload in fast-paced discussions and enables participants to focus on core problems and solutions without the added complexity of argument nodes, which our pilot tests revealed to be unnecessary in real-time settings.

Issue nodes in this system represent core problems and serve as anchors for both human and AI alignment. Issues define the discussion space, guiding participants to focus on specific problems at any given time. Each participant contributes knowledge relevant to the current issue, ensuring that the conversation remains centered and organized around well-defined problem areas.

Position nodes represent potential solutions, ideas, or thoughts related to a specific issue. A single position node can integrate similar perspectives from multiple participants, allowing diverse input to converge on shared solutions or insights. This structure not only helps in capturing varied viewpoints but also facilitates consensus-building around core ideas.

To maintain logical simplicity and prevent uncontrolled expansion, EchoMind does not allow direct connections between issue nodes or between position nodes, similar to the IBIS framework. This encourages step-by-step progression through complex problems, keeping discussions organized and focused.

The map grows from a root issue, evolving into a tree structure that is easier to visualize and facilitates mental model formation. EchoMind prioritizes clarity by avoiding complex structures like node merging or loops, in line with DG1. We will consider such advanced features in future iterations.

4.2 Focus / Unfocus

During our formative study, we observed that AI-generated updates to the map caused participants to spend considerable effort checking where the changes occurred, which disrupted the flow of the discussion (DG3). Even without AI intervention, facilitators often summarize or ask clarifying questions to help participants stay focused and prevent the conversation from becoming too scattered. This observation highlighted the importance of managing attention in real-time collaborative settings.

To address these challenges, we introduced the focus / unfocus operation for issue nodes. This mechanism serves two key purposes:

- **Human-AI Alignment**: The AI needs to understand which part of the map to update, while participants need clarity on where they should expect changes.
- **Participant Alignment**: The focus mechanism helps manage participants' attention by ensuring that everyone is discussing the same issue.

When an issue node is focused, EchoMind only updates the direct positions associated with that issue, as well as any issues linked to those positions. Visually, this means that only the descendant

nodes within two levels of the focused issue will be affected. Only one issue can be focused at a time, and focus can be switched between issues. Once no issue is in focus, the system stops modifying the map.

Facilitators are expected to actively manage focus during discussions. They should explicitly inform participants of the current issue under discussion and navigate the focus on the map accordingly. Forgetting to switch focus may result in AI-generated content appearing in unexpected locations, but it will not be lost. However, this extra effort helps ensure that the AI and participants stay aligned on the same goals.

4.3 Collaborative Interface

EchoMind's interface is designed for human-AI collaborative issue mapping. While the AI is responsible for automatically generating new positions and issues, the system also allows users to manually modify and guide the process. This balance ensures that participants can leverage AI's automation while maintaining direct control over the discussion structure (DG4).

The interface consists of two main sections, as shown in Fig. 3. Left Panel (Transcription Area) displays real-time transcriptions of the ongoing discussion, allowing users to follow along with the conversation. Users can control whether to enable or pause the microphone. Right Panel (Map Area) displays the issue map. The system periodically updates the map when there is a focused issue. Users also have the option to manually trigger map updates via a button, offering additional control over when changes are reflected.

AI's Role in Map Operations. Within the focused issue, the AI can perform the following actions (No deletion is allowed):

(1) Create New Positions: Adds new positions related to the focused issue, based on the discussion content.
(2) Modify Existing Positions: Updates the content of existing positions to reflect new insights.
(3) Create Linked Issues: For each position without associated issues, the AI will try to generate one if the conversation suggests further discussion.

User's Role in Map Operations. Users can modify any part of the map, not limited to the focused issue, as shown in Fig. 4:

(1) Create Nodes: Hover over a parent node to add a child issue or position.
(2) Modify Nodes: Click on a node's text to edit it in place.
(3) Delete Nodes: Click the delete button to remove the node.

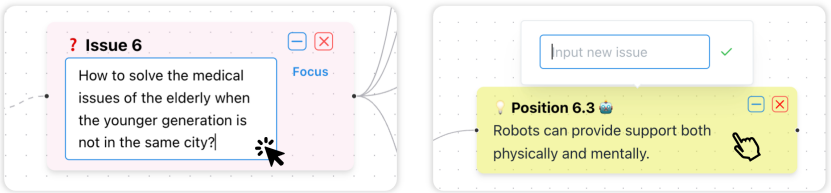


Fig. 4. Map operations of users. Users modify nodes by clicking on them. Hovering over either issue or position nodes allows for the addition of new nodes. Users delete nodes by clicking the red cross in the upper right corner. The robot icon indicates that the node can be updated in real time by the system.

Conflict Resolution in Human-AI Collaboration. We mark all AI-generated nodes by displaying a robot icon on the node. For positions, this also indicates that they can still be updated by the system. Once a user manually modifies a position, or if an issue has been focused on, the

robot icon disappears. This means the nodes have been "taken over" by the user, and they can expect these nodes will not automatically change in the future (DG4).

4.4 Real-time Generation and Modification

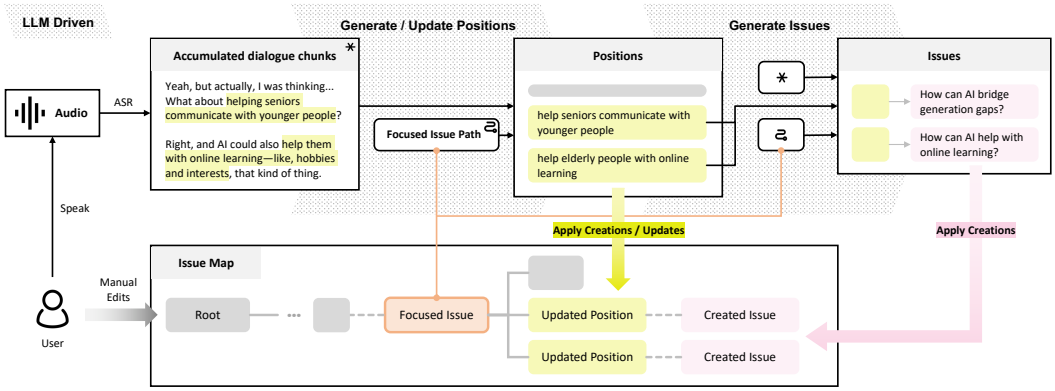


Fig. 5. Real-time generation pipeline of positions and issues.

EchoMind provides real-time support by tracking the dynamic flow of discussions. This process begins with continuous transcription of audio speech by an ASR (Automatic Speech Recognition) model into sentence fragments. Building on these fragments, the system periodically evaluates the discussion state and triggers automatic updates. Specifically, we have designed a generation pipeline for EchoMind, as shown in Fig. 5.

Maintaining Context under the Focused Issue. The system maintains a buffer of dialogue chunks associated with the currently focused issue. When the number of unprocessed words in this buffer exceeds a certain threshold, the system triggers a generation process using these accumulated chunks as input. Upon a focus switch to a different issue, the buffer is cleared, and a new context begins accumulating.

This design addresses two key considerations. First, keeping the LLM prompt concise ensures faster processing and prevents performance degradation in LLMs with longer inputs. Second, maintaining sufficient semantic continuity within the buffer avoids fragmented context that may confuse the model or reduce output quality. By scoping the LLM's input to the buffered dialogue of the current focus, we ensure that the input remains coherent and grounded, while other parts of the discussion—already reflected in the issue map—are excluded from reprocessing. We recognize that this localized scoping may constrain constrain broader coherence and cross-topic referencing—a trade-off we accepted to prioritize predictability and user alignment in fast-paced discussions.

Generating Positions and Issues. Our generation pipeline uses OpenAI's GPT-4o model [67] as the underlying LLM. Once generation is triggered, the system extracts new positions from the accumulated dialogue using prompt templates (see Appendix C.1). This process references existing positions to determine whether to update or create a new position. We also include the focused issue path, which is the path from the root of the issue map to the current issue. This path provides a structured semantic trajectory of the discussion, reflecting how participants have decomposed the problem. We deliberately exclude the entire map structure from the prompt to prevent irrelevant branches from introducing noise or ambiguity, and to keep the prompt length within optimal bounds for real-time inference.

Next, the system extracts new related issues from the accumulated dialogue, representing cases where users directly or implicitly suggest a new topic for discussion. However, this step is limited to newly created or modified positions that do not already have a linked issue. We adopt this conservative strategy to minimize the number of issue nodes on the screen, as we found in our pilot study during development that generating nodes far from the focused issue distracted users and that participants infrequently made decisions about creating new issues during conversations.

Updating the Issue Map. The system adds these updated positions and newly generated issues to the issue map under the currently focused issue node, and skips any modifications that conflict with manual edits. As a result, the system only updates the subgraph defined by the current focus, and no utterance is ever linked to multiple issue paths. In this way, we ensure that the system and the users collaborate in an unambiguous and consistent manner throughout the ongoing discussion.

4.5 Delay and User Experience

The most time-consuming step in the pipeline is the LLM invocation. Based on our pilot testing, we observed that when switching to a new focus, the generation delay for both positions and issues typically starts around 4 seconds. If the focus remains unchanged, and the accumulated dialogue context under that issue exceeds 5,000 Chinese characters (approximately 14 minutes of conversation), the average delay for generating positions can increase to around 14 seconds, while the delay for generating issues remains under 5 seconds on average.

To mitigate the potential disruption caused by such delays, we provide a visual indicator of updating status in the interface (see Fig. 3) to inform participants when the system is processing input. This helps maintain user awareness and prevent confusion during delayed updates.

Notably, processing delays may not be perceived negatively in discussion contexts. Prior research suggests that short waiting periods can enhance perceived control by providing time for reflection or parallel activities [57]. In our setting, participants often do not attend to the interface immediately while speaking, but instead return to it periodically for review or synthesis. From this perspective, the observed latency may fall within a cognitively acceptable range.

5 Evaluation

We conducted a user study to evaluate the effectiveness of EchoMind in supporting group discussions. The study also involved an exploratory comparison between EchoMind and an automatic summarization system, which served as an illustrative example of existing tools, to better understand how AI can provide cognitive support through a visual interface. We made this choice to better understand how participants engage with and adapt to these unfamiliar dynamics, and to surface key design factors and evaluation dimensions that should inform the design of future, more rigorous comparative studies.

We adopted a within-subject design in which each team alternated between using different systems to facilitate their discussions. We selected a consistent product design question across all groups to ensure a moderate level of complexity. This decision allowed each participant to contribute their thoughts while still requiring substantial discussion and reflection.

5.1 Comparison Conditions

We chose an automatic summarization system as the comparison condition, since linear document interfaces are commonly used in discussions and many existing AI meeting tools, such as Otter.ai [68], offer text summarization features alongside document records.

To focus the comparison on discussion facilitation rather than meeting minutes, we implemented this comparison system, referred to as AutoDoc. It uses a rich text editing interface, where the LLM periodically appends key dialogue points (see Appendix C.2) to the end of the document to create a

time-sequenced record. We adopted the format to align with existing tools, rather than introducing a new structure. The system does not delete or modify any content, but users can freely edit or move the AI-generated content, consistent with common practices.

Both EchoMind and AutoDoc are web applications connected to a remote server, and use the same ASR model and interface layout to control for unrelated variables. The left side of the interface displays transcribed text, while the right side displays the map area for EchoMind and the document area for AutoDoc. The primary distinctions between the two setups lie in the form of knowledge representation and the mode of interaction. Further details can be found in the Appendix B.

5.2 Participants

We recruited participants through a call posted on campus social platforms. We did not require participants to have a product design background, as real-world product design often involves interdisciplinary collaboration, and mixed-background teams better reflect the complexity of such discussions. Additionally, we did not disclose the discussion topic (elderly care products) in advance to avoid recruiting domain experts or participants who might have prematurely prepared.

In total, we recruited 16 participants (6 female, 10 male, aged 19 to 26) and organized them into four teams. All participants were native Chinese speakers, with 3 graduate students and 13 undergraduate students. Each participant had prior experience with group discussions. We distributed the 3 participants with specific product design experience across different teams.

5.3 Setup

We held each group discussion in a dedicated meeting room. We provided each team with a laptop for audio recording and for operating the system via keyboard and mouse. Participants accessed the interface through the browser, which was displayed on a 65-inch LED screen (3840 × 2160 pixels) connected to the laptop.

5.4 Task

We asked all teams to assume the role of a startup competition team, with the same task of designing an intelligent product tailored for the elderly. The goal was to develop an initial solution through discussion.

The discussion consisted of two sessions, each lasting 30 minutes. In the first session, the focus was to thoroughly analyze the challenges and needs faced by the elderly in modern society. We instructed participants to identify three key pain points with potential by the end of the discussion. In the second session, the task was on exploring potential solution directions for each of the three identified pain points. By the end of the session, we required participants to develop an initial solution proposal for each problem. Although the two sessions addressed different aspects, both followed a discussion process of divergence and convergence, requiring brainstorming, extensive note-taking, organization, and multiple reviews throughout the discussion.

We selected one participant from each team to take responsibility as the facilitator, based on a consideration of their experience and willingness. The facilitator was required to guide the group discussion and maintain both systems throughout the whole process, ensuring that the entire process was captured, rather than just the final outcomes. Although other participants did not directly operate the system, we expected them to assist the facilitator by monitoring changes in the interface and providing suggestions for revisions.

5.5 Procedure

For each team, the experimenter began by introducing the background of the discussion topic and providing a printed handout with detailed context. We then selected a facilitator and explained their

responsibilities to all participants. The team was instructed to use both EchoMind and AutoDoc, each supporting one discussion session. The order of system usage was counterbalanced across teams [69].

Following this, the experimenter introduced the corresponding system according to the assigned order, allowing participants a 10-minute trial period to familiarize themselves with the system before each session. Once all participants confirmed their familiarity, the session commenced. After each session, we asked participants to complete a questionnaire. Upon completion of both sessions, we conducted a 15-minute interview to gather participants' feedback.

5.6 Measures and Data Analysis

We video-recorded all sessions with participants' consent and employed five types of measures for analysis.

5.6.1 Questionnaire. We collected subjective ratings from all participants. The questionnaire focused on three aspects: discussion quality, perceived system effectiveness and value, and cognitive load. The first two aspects were inspired by prior studies [13, 57, 94]. Cognitive load was measured using the standard NASA-TLX [37] in a 7-point Likert scale, while the other two aspects used a 5-point Likert scale to reduce the burden on participants when completing the questionnaire. We conducted Wilcoxon signed-rank tests for each measure.

5.6.2 Semi-structured interviews. To gather participants' feedback and experiences on the discussion process, we asked several open-ended questions, including: "*Overall, how did the two discussions differ for you?*", "*How did each system support your discussions?*", and "*In what scenarios would you consider using these systems in the future?*" Other questions were included to gather deeper insights, which are not included here for brevity. One of the authors conducted a thematic analysis of the interview results.

5.6.3 Behavioral Measures of Discussion Quality. To provide an objective measure of discussion quality, we defined a Discussion Quality Index (*DQI*), which aggregates key behaviors observed during the discussions. Inspired by relevant research [86], we identified the following behaviors as indicators of productive and high-quality discussions: the introduction of new issues (*CI*), the creation of new positions (*CP*), the summarization and evaluation of prior points (*SE*), referencing or mentioning previous statements (*REF*), and supporting or opposing positions with elaboration (*ATT*). The final DQI score represents the average occurrence rate of these behaviors and serves as a numeric measure of the overall quality of the discussions.

Two annotators (the authors) independently labeled instances of these behaviors in the conversation transcripts, with a first-pass Krippendorff's alpha of 0.585. Any discrepancies between the annotators were resolved through discussion to reach a consensus. The annotators were not explicitly informed of the experimental condition associated with each transcript, but certain content cues (e.g., participants discussing how to take notes within the system) may have implicitly revealed the condition, which was difficult to avoid. However, we believe this does not substantially threaten validity, as the annotations were grounded in observable behaviors, which limited interpretive ambiguity and reduced susceptibility to bias.

5.6.4 Facilitation behavior. We also labeled the participants' facilitation behaviors, including guiding, planning, and monitoring the progress of the discussion. Example behaviors include: "*Let's move on to the next issue.*", "*Do you think this part of the map/text needs to be modified?*", and "*Your point is interesting, but I think we should first focus on the current issue.*"

Two annotators (the authors) independently labeled the behaviors, with a first-pass Cohen's kappa of 0.574. Any discrepancies in labeling were resolved through discussion to reach a consensus.

As in the previous annotation process, annotators were not explicitly informed of each transcript's condition, but it was difficult to prevent them from inferring it. Nevertheless, we believe any potential bias was limited given the clear, objective criteria used for annotation.

5.6.5 Manual edits. We tracked the facilitator's collaboration behavior with the systems, including the frequency of operations (structural modifications) and word count (content modifications). Structural modifications include actions like adding or deleting nodes and focusing/unfocusing in EchoMind, as well as adding, deleting, moving, copying, or pasting lines in AutoDoc. Content modifications capture the total number of characters added or deleted in both systems.

6 Results

In this section, we begin by presenting an exploratory comparative analysis between EchoMind and AutoDoc, drawing both quantitative and qualitative insights. Following that, we provide a deeper analysis of EchoMind, focusing on its performance and user interaction behaviors to reveal specific findings about its impact on group discussions and the role of its unique features.

6.1 Exploratory System Comparison

We present the subjective ratings results for both EchoMind and AutoDoc in Fig. 6, along with statistics from an example discussion featuring two sessions from a team in Fig. 7. A detailed analysis is provided in the following sections.

6.1.1 Discussion quality. The results collectively show that EchoMind led to a more productive discussion outcome. First, annotators' objective evaluation of the discussions revealed that EchoMind had a DQI of 55.57 ($CI = 8.75$, $CP = 25.75$, $SE = 9$, $REF = 4$, $ATT = 8.25$), while AutoDoc had a DQI of 44 ($CI = 6$, $CP = 24.25$, $SE = 7.25$, $REF = 1.25$, $ATT = 5.25$). These scores suggest that EchoMind facilitated higher-quality discussion behaviors compared to AutoDoc.

Further, participants significantly felt they addressed the set topics more effectively ($p < .05, W = 9.0$) and were more satisfied with the overall quality ($p < .01, W = 0.0$) in the sessions using EchoMind. Regarding the outcome, participants perceived a significantly greater depth ($p < .05, W = 19.5$) and breadth ($p < .05, W = 9.0$) in the EchoMind sessions. Notably, P9 emphasized the depth aspect, stating that "*If the topic requires deep exploration, such as continuously branching out, the map is more suitable.*" P16 noted that the map "*helps people remember more content, particularly because its clear hierarchical structure brings key points into focus,*" thereby enhancing the perception of breadth.

The results also show that EchoMind discussion sessions were perceived to be less disorganized ($p < .05, W = 3.5$) and provided a clearer sense of direction ($p < .05, W = 10.0$) compared to AutoDoc sessions, indicating that EchoMind facilitated better guidance and planning during the discussion.

In terms of guidance, the generated map structure "*offered an issue-based logic*", and as the discussion progressed, the generation of related issue nodes helped participants address issues sequentially within the current context (P4). The issue nodes supported flexible transitions between issues in EchoMind. For instance, when participants "*delved deeply into one issue and needed to jump back to the original issue, the map allowed them to continue seamlessly, whereas with AutoDoc, it required scrolling back, and the distant temporal sequence of the text made it difficult to establish new connections*" (P1).

In terms of planning, participants reported that the map structure in EchoMind made it easier to plan the next issues by providing a more intuitive display of progress. For example, P16 noted, "*When an issue has accumulated several positions, it seems like a good time to move on to the next issue, which isn't as clear in AutoDoc.*" Additionally, in EchoMind, participants were more likely to

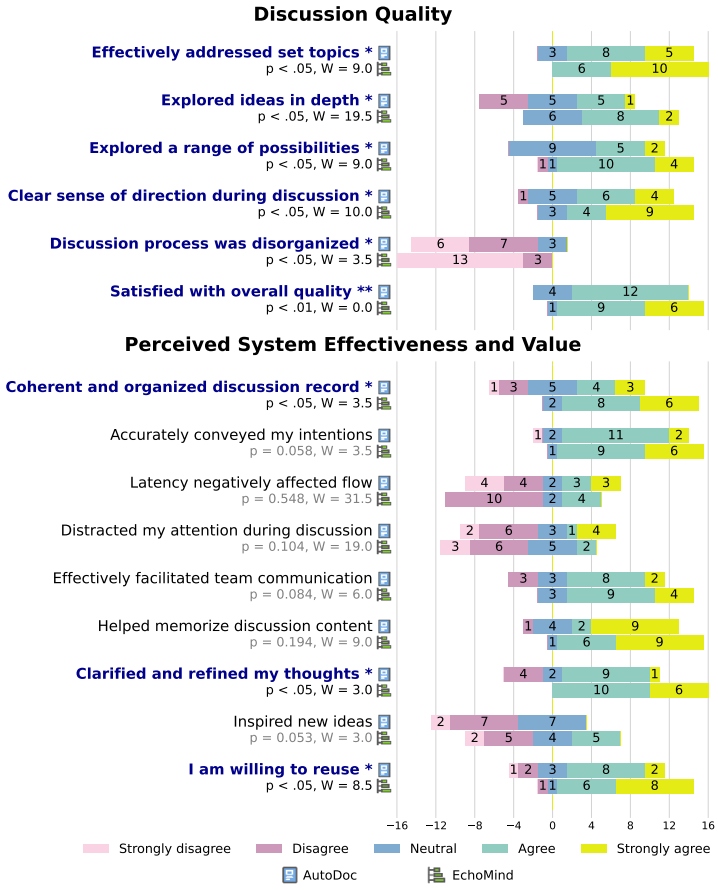


Fig. 6. Subjective ratings of discussion quality, perceived system effectiveness and value. Titles in dark blue and bold indicate significant differences (*p < .05; **p < .01).

remember and prioritize issues that they had mentioned but not yet discussed in depth, whereas in AutoDoc, participants were less likely to bring up such issues again or give them the same attention.

6.1.2 Perceived system effectiveness. In terms of recording, both systems accurately conveyed users' intentions, with no significant differences between them. This suggests that AI systems can effectively meet users' facilitation needs for recording, regardless of whether the information is presented in a map or document format. However, in the reviewing phase, EchoMind showed a significant advantage in providing a coherent and organized discussion record ($p < .05, W = 3.5$), which notably helped participants clarify and refine their thoughts ($p < .05, W = 3.0$).

Interviews suggest that participants' primary need was to **recall relationships rather than detailed content during review**. In terms of revisiting key points for reviewing, AutoDoc "*faithfully recorded all key points, lacked a clear hierarchy, making it harder to find anything during review*" (P5, P6). Participants noted that "*if the discussion was long, it could become quite confusing*" (P5, P6). P14 commented, "*You have to read through the document line by line, whereas the map is more intuitive.*" P4 mentioned that, "*The map's logical structure is very clear, which helps with memory.*" These

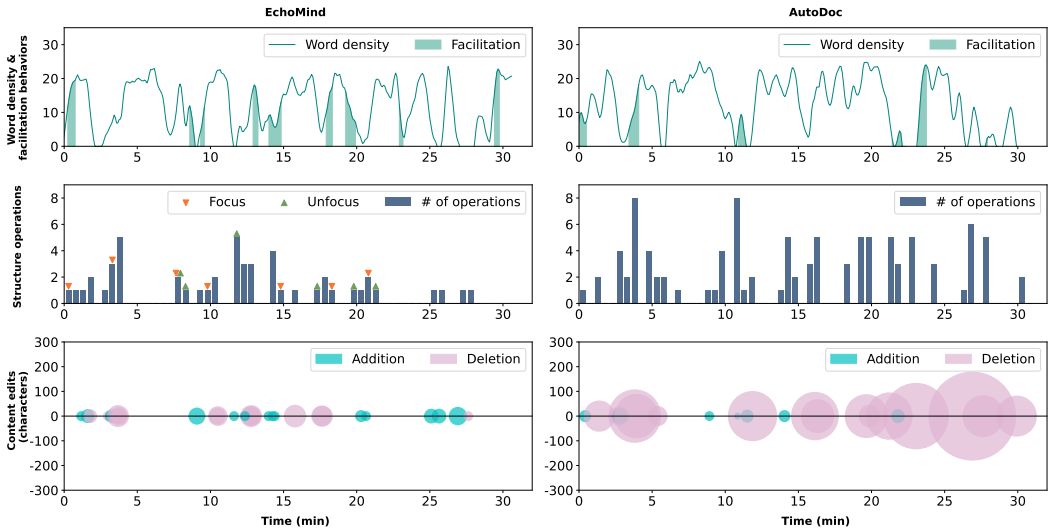


Fig. 7. Facilitation behaviors, structure and content operations of one team across two sessions.

findings show that the map structure effectively supported real-time organization of knowledge paths during review.

Participants found the knowledge paths visualized in EchoMind useful. When using EchoMind, participants focused on "*which points had already been discussed and how they were distributed*" (P6). P12 mentioned, "*At the start of the session, I missed some parts of the discussion while responding to messages, but by looking at the branches on the map and the current issue, I could quickly trace the progress.*" P14 highlighted, "*Seeing what we were just discussing on the map helped me shorten the mental distance to new ideas.*"

6.1.3 Facilitation behavior. The annotators recorded an average of 10.25 facilitation behaviors per session for EchoMind and 8.75 for AutoDoc. In Fig. 7, the facilitation behavior (the first row) in AutoDoc reflects the typical distribution of facilitation actions in discussions, with more activity at the beginning and end and only intermittent, simple facilitation in the middle. In contrast, the EchoMind session exhibited a marked increase in facilitation during the middle phase, while the operations on structure (the second row) show that most focus and unfocus operations also occurred during this middle period. This suggests that the focus feature may encourage facilitators to take a more active role in guiding the entire discussion. Participants' subjective ratings supported this observation, as they reported perceiving a significantly clearer sense of direction during the EchoMind session ($p < .05, W = 10.0$).

6.1.4 Manual Edits on Structure & Content. On average, EchoMind involved 13.75 structural operations per session, with modifications covering 146.5 characters, while AutoDoc involved 17 operations and 624.5 characters. Fig. 7 presents a typical distinction between the two systems in terms of manual edits on structure and content, showing that the facilitator made fewer edits in EchoMind compared to AutoDoc, objectively reflecting the lower operational load of EchoMind. Moreover, NASA-TLX ratings (Table 1) indicate that EchoMind has a significant advantage in terms of temporal demands ($p < .05, W = 4.0$).

The lower modification burden in EchoMind is due to its focus mechanism, which ensures that key point records are issue-relevant and precise. In contrast, AutoDoc records sequentially and

indiscriminately, which "*requires a lot of time to maintain*" (P7). During the user study, facilitators' collaboration strategy with AutoDoc mainly involved "*selecting useful points from AutoDoc's record, pasting them under self-generated titles, and either ignoring the rest of the document or deleting all unselected content*" (P1, P16). This difference was especially noticeable in small group discussions, where the facilitator also played the role of recorder. They noted that "*EchoMind has a lower cognitive load, allowing facilitators to maintain the system and facilitate the group discussion for a longer period*" (P7, P16).

6.1.5 Subjective feedback. Participants expressed a significantly higher willingness to reuse EchoMind ($p < .05, W = 8.5$), as shown in Fig. 6. The majority (14 out of 16) indicated they would use EchoMind in the future. During the interviews, participants discussed potential scenarios for reusing EchoMind. For example, P1 and P15 suggested introducing EchoMind to "*teaching assistants responsible for organizing seminars*", as it could effectively help inexperienced assistants structure coherent seminar discussions among students with diverse backgrounds. Some participants also shared more creative ideas, such as using EchoMind to collaborate with AI for "*organizing travel plans*", "*drafting research proposals*", or "*taking notes during lectures*" (P5, P8). These responses confirmed participants' recognition of EchoMind's ability to help them organize human knowledge.

It is worth noting that documents also have their own applicable scenarios and advantages. Participants mentioned that "*EchoMind is more suitable for discussion processing, helping us better organize the group to achieve established goals*" (P1, P7), while for post-discussion purposes, they preferred "*a more formal, written summary in the form of document*" (P7). Additionally, we report the feedback from a participant with a lower willingness to reuse EchoMind. This participant, as a highly-skilled document user, proposed that documents could more flexibly support brainstorming. This highlights that the success of the facilitator's collaboration with the AI system is strongly influenced by prior experience using non-AI tools. Each system has its own applicable scenarios, strengths, and more appropriate target users.

We also observed an interesting connection between latency and reliability in user perception. Although EchoMind requires fewer words (50 Chinese characters) to trigger the generation of new key points compared to AutoDoc (200 Chinese characters), some participants perceived the latency differently. They had higher expectations for EchoMind, anticipating updates after every brief pause. Some participants noted that the higher quality of key points generated by EchoMind raised their expectations for the display of recorded points or the feedback deemed valuable by the system, which in turn affected their perception of latency.

6.2 In-depth Analysis of EchoMind

In this section, we first analyze the technical performance of the EchoMind pipeline based on data collected from the evaluation study. Next, we investigate how EchoMind's unique focus mechanism influences group discussions, followed by an examination of the behavioral differences between facilitators and non-facilitators. We conclude by discussing the respective roles of the AI generation and visual structure in supporting discussion facilitation, as well as their combined effects.

6.2.1 Post-hoc technical analysis. We demonstrate how well current LLMs perform on the tasks of generating positions and issues, providing insights that can inform future technical designs. Additionally, we ran a simulation to test whether automating focus switching with LLMs could further optimize user interaction.

Position generation. In each EchoMind session, we selected 3 dialogue segments where the user did not perform manual operations, totaling 12 segments and 34.25 minutes. One author manually annotated all key points in these segments, while another author reviewed the annotations for omissions or errors and discussed necessary corrections. To account for linguistic variation, both

authors adopted a clear semantic-level matching criterion: an AI-generated position was considered a correct match if it conveyed the same idea as a manually annotated key point, regardless of exact wording. We acknowledge that this process involves subjective interpretation, and future work could incorporate multiple independent annotators to improve objectivity.

The results showed that the 66 positions generated by EchoMind covered 49 out of the 52 manually annotated key points (94.23%). Additionally, 4 of the 66 generated positions were not part of the annotated key points, representing information that was weakly associated with the discussion. We identified two main reasons for the 3 uncovered points. First, the discussion logic was somewhat disjointed, making it challenging to accurately interpret. For instance, one participant said, "*Using marketing strategies like this to promote a good health supplement.*" Here, the participant was referring to an actual health supplement, not the smart product mentioned earlier by others, but the LLM summarized the position as "*The product can be marketed to the elderly using health supplement marketing strategies.*" The second case involved a lengthy discussion under one issue, where the fast pace of speech caused the dialogue text included in the prompt to become too long, resulting in the LLM failing to focus on the later text to extract a position. This suggests that there is room for optimization in our pipeline's ability to segment and compress dialogue text.

Issue Generation. This task involves generating explicit or implicit issues that may be further discussed, which is more akin to issue suggestion based on the ongoing conversation. We evaluated system-generated issues along two dimensions: semantic relevance and actual usage in the discussion. To assess relevance, we considered a system-generated issue to be relevant if it reflected a meaningful question or sub-topic that was semantically connected to the preceding conversation and represented a plausible direction for further exploration. Two authors independently reviewed all issues generated by EchoMind in the selected segments and resolved any disagreements through discussion. All issues generated in these segments were judged to be relevant under this criterion.

However, whether these issues were actually adopted by participants is a separate matter. We examined a total of 24 focus switches across all sessions and found that 7 of the selected issues came from the system, while the remaining 17 were manually input by users. We identified two main reasons why the system did not provide adequate assistance: first, participants did not mention the issue in the dialogue; second, participants introduced new issues simultaneously when switching focus, and the system did not have time to respond. Currently, the system design relies solely on the accumulated dialogue up to that point and cannot update an issue once it has been generated. This limitation restricts its effectiveness. Future designs could explore the generation of multiple issues and dynamic updates to improve functionality.

Simulating Automatic Focus Switching. We conducted an exploratory test to evaluate whether LLMs could replace users in automatically switching the focus position on the issue map. We reviewed the moments when the system updated the map across all sessions, excluding any instances where user interventions could have influenced the data. This resulted in 24 focus switch records and 126 non-switch records. We prompted the GPT-4o model to predict whether a switch should occur and identify the target issue, based on the accumulated dialogue and issue map at those moments, as detailed in Appendix C.3. Out of 50 predicted switches, 14 were correct, yielding a precision of 28%. Considering all ground truth switch records, the recall was 58.33%. This indicates a high false positive rate and low recall, reflecting a poor match with the actual user intent. When considering only whether a switch occurred, excluding the target issue, precision and recall improved to 36% and 75%, respectively. Yet they still fell short of practical application.

We believe the primary reason for the limited predictive performance is an inherent misalignment between the model and user intent. In many cases, participants' focus switches were driven not solely by the immediate content of the dialogue, but by an implicit sense of agenda progression or group coordination needs—factors that are not directly observable in the text history. A potential

improvement would be to incorporate explicit models of conversational structure or topic hierarchy. For example, tracking high-level discussion phases (e.g., problem framing, solution generation, decision convergence) could provide contextual signals that guide focus prediction. We consider this a promising direction for improving AI alignment in real-time collaborative settings.

6.2.2 Impact of focus mechanism. We further analyzed the reasons behind the 24 focus switches. Of these, 11 were shifts to sub-issues, 9 to sibling issues, and 4 to parent issues. These shifts to more in-depth issues typically indicated the need for further exploration of more specific or thought-provoking topics. For instance, the focus shifted from the "*Market value*" issue to the question of "*What are the main concerns of elderly people and their children as the primary consumer market?*" In another case, the discussion moved from a position about "*A device for detecting emergency situations in elderly people [...]*" to the issue of "*What should be detected?*" The remaining switches primarily involved concluding and transitioning to new topics. For example, after thoroughly discussing a pain point, a participant suggested, "*We should explore the market value of this product,*" leading to a switch to the "*Market value*" issue. Alternatively, a participant or facilitator might propose, "*We can summarize,*" which would lead to a switch to the "*Pain point summary*" issue.

As reported in interviews, the focus feature promotes alignment and provides a sense of direction. For alignment between participants, the focus feature "*helped ensure more consistent issue alignment*" (P5), reminding participants to "*avoid irrelevant information and stay on topic*" (P1, P7, P9), which "*significantly improved efficiency*" (P1). The focus feature also facilitated alignment between participants and the AI. Without it, the AI struggled to identify the specific issue being discussed, which could lead to recording errors or incorrect generation of hierarchical nodes. P4 observed, "*Focus limits the scope of discussion, enabling the AI to organize and generate content within the fixed topic range, producing more structured results and making it easier for us to organize our thoughts.*" Participants also highlighted the implicit value of the focus feature: "*Not every sentence we say is useful. Using focus allows the AI to filter out irrelevant information*" (P3, P7).

For facilitators, the focus feature helped manage content generation at a local level without requiring them to constantly oversee the global discussion. This approach offered a distinct advantage in supporting longer facilitation sessions. For instance, regarding guidance, P13 noted, "*With a document, you have to actively think of a title to guide the next step in the discussion and decide where to input it, while simply focusing on the nodes in the map is more convenient.*" For review, participants mentioned that "*in a document, you need to put extra effort into dividing paragraphs, filtering which parts to keep, and rearranging content*" (P7), whereas "*if the facilitator can effectively use the focus feature, the AI can automatically record and organize content under the relevant issue*" (P9). Overall, participants found that focus and unfocus actions "*align well with logical thinking*" (P3), are "*easy to operate*" (P1), and help reduce the mental load on facilitators during facilitation tasks.

6.2.3 Facilitators and non-facilitators in human-AI collaboration. We analyze the reactions of other group members when the facilitator manages the issue map. The facilitator creates or modifies issues specifically to facilitate the upcoming focus switch. Of these actions, participants collectively discussed and confirmed 20 (83.33%), while the facilitator directly initiated only 4 (16.67%), which other participants tacitly accepted. Notably, these 4 instances did not occur within the same session, suggesting that they were not driven by individual characteristics of the participants. This suggests that focus switching is not solely determined by the facilitator; participants are aware of the switching actions and may even proactively suggest them. For instance, in one discussion, a participant suggested, "*How about we discuss [...]?*" The facilitator and other participants responded positively, and together they defined the content and position of the new issue.

In terms of position management, the facilitator primarily takes control over the AI-generated output. Typical examples include adding formatting to correct AI-generated content (e.g., prepending a summary keyword at the beginning), merging scattered content across different positions into a single position, or adding supplementary details. These actions prevent the AI from updating the content of these nodes. Similar to issues, in most cases, non-facilitators explicitly support these operations, with only a few instances of tacit acceptance. Additionally, as the discussion approaches its conclusion, participants need to consolidate the final solution and will manually input positions instead of relying on the AI. All participants engage in the process of discussing and inputting content during this phase.

6.2.4 Role of AI generation and visual structure. Further analysis of the discussion behaviors revealed that EchoMind's support for group discussion facilitation arises from both the AI pipeline's generation and its unique visual interface. Firstly, the hierarchical structure of the map helps participants understand of the logical relationships between contents. For instance, participants may point out that a new piece of information does not effectively address the current issue and seek to place it in another location. Additionally, the parallel structure of positions under the same issue encourages participants to check whether the content is aligned and to consciously discuss new ideas in parallel with existing content. Participants also make use of the collapse and focus functions of the visual interface to temporarily set aside ideas that do not move the discussion forward, and then move on to the next topic. On the other hand, AI's integration of long conversations reduces participants' cognitive load when reviewing dialogue content. Generally, when participants conclude a topic, they can quickly review the AI-generated summary in under 10 seconds and transition smoothly to the next topic. The comprehensiveness of AI summaries also alleviates participants' concerns about missing important information of the conversation.

In fact, the most common scenario involves the combined effects of AI generation and the visual structure. For example, participants may first assess whether the discussion is complete based on the number of nodes, but they also review the content within each node to identify any potential gaps. In the final stages of the session, participants take a comprehensive view of the map, using it as a framework to quickly locate the areas they need to review and then examine the content of all nodes. These observations suggest that we should not consider the map structure independently from the content of the nodes, as they jointly influence decision-making during the discussion.

7 Discussion

7.1 Design Implications

In this section, we present insights from the user study and offer design implications for future AI-supported facilitation using shared knowledge visualization.

7.1.1 Providing Practical Visual Structure. Our study confirms the value of providing a real-time visual structure during discussions, particularly in helping to organize knowledge paths and users' thoughts, facilitating better guidance and planning, and ultimately leading to productive outcomes. This structure can be effective whether presented as a map or a hierarchical document.

The key to design a visual structure for discussion lies in finding an effective way to represent discussion knowledge, with special consideration of the limited cognitive load capacity in real-time reflections. The structure should not be overly complex, and it must allow participants to naturally and effortlessly classify and organize the knowledge generated during the discussion into appropriate hierarchy and nodes within the structure.

We employed a relatively basic structure in the form of a tree, containing only two types of nodes—issues and positions—to support a wide range of problem-solving discussions. Actually,

different domains have their own familiar problem-solving structures or methodologies. Based on shared domain knowledge, it is possible to provide more practical node types, such as suggestion nodes, conclusion nodes or to-do list nodes, or even generate discussion summary by converting the map into a document. Therefore, determining which additional nodes to include to make the visual structure more practical depends on the scenario and users the AI-supported facilitation system is designed to serve.

7.1.2 Turning Organizing Expertise into Standardized Collaboration Operations. Our user study confirmed that EchoMind interface effectively supported collaboration in group facilitation, particularly through the focus feature, which enhances alignment and maintenance while reducing the facilitator's operational burden for manual edits.

An interesting observation during the user study was that EchoMind functioned somewhat like a tutor by normalizing the facilitator's behavior. This discovery stemmed from an intriguing phenomenon: when facilitators adhered more closely to the recommended system usage guidelines provided before the sessions, their teammates responded more positively to the overall discussion quality and experience. These guidelines included switching issues based on the discussion flow, verbally announcing focus changes when using the focus feature, and consulting others before making node edits.

The results revealed that even the facilitators with less experience in organizing product design discussions received higher evaluations when they followed these guidelines, compared to more experienced facilitators who did not. This suggests that the system's required interactions, such as focus/unfocus and manual edits, inherently promote facilitation expertise and lead to a more inclusive collaboration environment. When facilitators used EchoMind correctly, their behaviors became more consistent and effective, with lower cognitive effort.

In future AI-supported facilitation systems, the principle of turning organizing expertise into standardized collaboration operations can be adopted, with careful attention to reducing the users' operational burden.

7.1.3 Preserving Useful Context for LLMs. The key requirement for integrating LLMs into a real-time facilitation system is that they must be prompted frequently, with high frequency and short inputs, rather than processing large amounts of text at once. The goal is to retain all the necessary reasoning semantics within the limited prompt length. Therefore, we cannot simply input the entire conversation history for the LLMs to process every time. Instead, we need to filter the relevant parts of the dialogue.

The critical factor in preserving all useful reasoning semantics lies in aligning AI's understanding with human cognitive context during the discussion. The human cognitive context during discussions is dynamic, meaning that at different times, the memory on past dialogue shifts. Our research explored a general cognitive model of how humans process information in discussions: the most recent information related to the current issue is retained in detail, but not an unlimited history. Instead, a truncated dialogue history is used, while earlier discussions are summarized as paths and relationships from the root issue to the current issue. The user study confirmed that the strategy we employed achieved relatively fast and accurate results.

Future designs for the cognitive context structure of LLM agents can be more finely tailored to the specific discussion tasks. For example, creating layered summaries of the dialogue history can retain global context while preserving different levels of information detail.

7.2 Ethical and Practical Concerns of Relying on LLMs

This study focuses on the design of elderly care products that remain neutral from a technological ethics standpoint, avoiding politically sensitive topics. Throughout the study, there were no

instances where the LLM actively censored or avoided responding. However, for more general discussion contexts, the inherent censorship of LLMs can indeed have a potential impact on the conversation [33, 58]. For example, the model may filter or alter discussions on controversial topics, especially if the platform or model developers deem the content inappropriate. This raises concerns about the potential limitations on freedom of expression and the scope of discussions.

Furthermore, LLMs may inherit biases from their training data [63, 90]. If the training data over-represent certain demographic groups or perspectives, the LLM might prioritize specific viewpoints [2, 22]. In our case, the discussion task of elderly use of technology could be affected by these biases. For example, the model might misunderstand concepts related to parent-child relationships in the Chinese context, adopting a Western perspective instead, or it might focus on solutions that align with certain cultural norms, rather than considering a broader, global perspective. To address this issue, it is important to consider training or fine-tuning the LLM with data that better aligns with the values and cultural contexts of the target users. A potential solution could be to continuously adapt the model based on the discussion data accumulated from users over time, allowing it to better reflect evolving user preferences and cultural nuances.

Finally, frequent calls to LLMs incur financial and environmental costs, as they require significant computational power [47, 75]. In our study, EchoMind averaged 74.25 API calls per 30-minute session, which, at the rates during the study period, amounted to roughly $1 per session. To ensure more sustainable practices, future developments should consider using smaller, more cost-effective models, or even explore on-device deployment, which may become increasingly feasible as LLM technology continues to evolve [25, 43].

8 Limitations and Future Work

We acknowledge several limitations in our studies. First, regarding the formative study, while research groups are a strong foundation for understanding group dynamics in complex problem discussion, we recognize that different groups may approach complex problems differently. Future work could explore other types of groups to broaden the scope of our findings.

In the evaluation study, the number of groups was relatively small. Although the four groups demonstrated strong common characteristics, increasing the study's scale could further enhance the robustness of both our quantitative and qualitative analysis. Our evaluation focused exclusively on graduate and undergraduate students, whose primary discussion contexts involved coursework, research projects, and startup competitions. While they represent a typical group for problem-solving discussions, the results may differ for other user groups, such as teachers conducting seminars or managers in corporate settings. A broader participant variety could provide a more comprehensive evaluation of the system.

Additionally, the exploratory comparison between EchoMind and AutoDoc was intended as a reference to help understand its impact, rather than a rigorous experimental comparison. Future work could involve more controlled and systematic comparisons between EchoMind and a broader range of conditions—for example, manipulating design variables such as the choice of LLMs, the timing of map updates, the types and granularity of nodes, or the structural constraints of the visual map. It could also explore new evaluation metrics, such as the timing and extent of user intervention in the AI pipeline, or the relationship between textual input and multimodal behaviors like speech-based interaction. These directions would help further unpack the design space of human-AI collaborative facilitation and clarify the causal contributions of different system components.

Finally, our simulation of automatic focus switching revealed a mismatch between LLM predictions and user intent. A key limitation is that the model lacked awareness of higher-level conversation structure or agenda flow. Future work could explore structured modeling of discussion

dynamics to better guide AI attention and improve alignment with the evolving goals and direction of the discussion.

9 Conclusion

In this paper, we introduced EchoMind, a human-AI collaborative system designed to support group discussions by generating real-time issue maps. Through the integration of real-time speech recognition and large language models (LLMs), EchoMind supports participants in structuring and navigating complex discussions. By centering its knowledge representation around issues, the system provides both flexibility and clarity, enabling participants to stay focused on key problems and solutions.

Our user study with 4 teams (16 participants) comparing EchoMind with an automatic summarization document system demonstrated that EchoMind is more effective at helping participants clarify ideas, maintain focus, and improve the overall quality of discussions. The results highlight the value of EchoMind's map structure, which promotes knowledge review through relational connections rather than relying solely on content. Additionally, the focus mechanism enabled efficient alignment between users and the system with minimal operational effort, promoting seamless human-AI collaboration. We also elaborated on the design implications of human-AI collaborative discussion facilitation.

Our work contributes to the design of collaborative AI systems that enhance real-time discussions through human-machine alignment and attention management. It offers new insights into how AI can support structured, goal-oriented group interactions with minimal user intervention, and opens new directions for optimizing human-AI collaboration in dynamic decision-making contexts.

Acknowledgments

This work was supported by the Natural Science Foundation of China under Grant No. 62132010, Beijing Key Lab of Networked Multimedia, and Key Research and Development Program of Ningbo City under Grant No. 2023Z062.

We sincerely thank Dr. Jie Cai for valuable suggestions that helped improve our work and strengthen its contribution to the CSCW community.

References

[1] Joseph A. Allen, Tammy Beck, Cliff W. Scott, and Steven G. Rogelberg. 2014. Understanding workplace meetings. *Management Research Review* 37, 9 (Jan. 2014), 791–814. doi:10.1108/MRR-03-2013-0067 Publisher: Emerald Group Publishing Limited.

[2] Abubakar Abid, Maheen Farooqi, and James Zou. 2021. Persistent Anti-Muslim Bias in Large Language Models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21)*. Association for Computing Machinery, New York, NY, USA, 298–306. doi:10.1145/3461702.3462624 event-place: Virtual Event, USA.

[3] Fran Ackermann. 1996. Participants' perceptions on the role of facilitators using Group Decision Support Systems. *Group Decision and Negotiation* 5, 1 (Jan. 1996), 93–112. doi:10.1007/BF00554239

[4] Joseph A. Allen and Nale Lehmann-Willenbrock. 2023. The key features of workplace meetings: Conceptualizing the why, how, and what of meetings at work. *Organizational Psychology Review* 13, 4 (Nov. 2023), 355–378. doi:10.1177/20413866221129231 Publisher: SAGE Publications.

[5] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. doi:10.1145/3290605.3300233

[6] David Atkin, M. Keith Chen, and Anton Popov. 2022. The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley. doi:10.3386/w30147

[7] Aadesh Bagmar, Kevin Hogan, Dalia Shalaby, and James Purtilo. 2022. Analyzing the Effectiveness of an Extensible Virtual Moderator. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP (2022), 18:1–18:16. doi:10.1145/3492837

[8] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2022. On the Opportunities and Risks of Foundation Models. doi:10.48550/arXiv.2108.07258 arXiv:2108.07258 [cs].

[9] Robert P. Bostrom, Robert Anson, and Vikki K. Clawson. 1993. Group facilitation and group support systems. In *Group Support Systems: New Perspectives*, Leonard M. Jessup and Joseph S. Valacich (Eds.). Vol. 8. Macmillan, 146–168.

[10] Tuan Bui, Oanh Tran, Phuong Nguyen, Bao Ho, Long Nguyen, Thang Bui, and Tho Quan. 2024. Cross-Data Knowledge Graph Construction for LLM-enabled Educational Question-Answering System: A Case Study at HCMUT. In *Proceedings of the 1st ACM Workshop on AI-Powered Q&A Systems for Multimedia (AIQAM '24)*. Association for Computing Machinery, New York, NY, USA, 36–43. doi:10.1145/3643479.3662055

[11] K. C. Burgess Yakemovic and E Jeffery Conklin. 1990. Report on a development project use of an issue-based information system. In *Proceedings of the 1990 ACM Conference on Computer-Supported Cooperative Work (CSCW '90)*. Association for Computing Machinery, New York, NY, USA, 105–118. doi:10.1145/99332.99347 event-place: Los Angeles, California, USA.

[12] Jeffrey Kok Hui Chan and Wei-Ning Xiang. 2022. Fifty years after the wicked-problems conception: its practical and theoretical impacts on planning and design. *Socio-Ecological Practice Research* 4, 1 (March 2022), 1–6. doi:10.1007/s42532-022-00106-w

[13] Senthil Chandrasegaran, Chris Bryan, Hidekazu Shidara, Tung-Yen Chuang, and Kwan-Liu Ma. 2019. TalkTraces: Real-Time Capture and Visualization of Verbal Content in Meetings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300807

[14] Weihao Chen, Yuanchun Shi, Yukun Wang, Weinan Shi, Meizhu Chen, Cheng Gao, Yu Mei, Yeshuang Zhu, Jinchao Zhang, and Chun Yu. 2025. Investigating Context-Aware Collaborative Text Entry on Smartphones using Large Language Models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3706598.3713944

[15] Weihao Chen, Chun Yu, Huadong Wang, Zheng Wang, Lichen Yang, Yukun Wang, Weinan Shi, and Yuanchun Shi. 2023. From Gap to Synergy: Enhancing Contextual Understanding through Human-Machine Collaboration in Personalized Systems. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3586183.3606741

[16] Xinyue Chen, Shuo Li, Shipeng Liu, Robin Fowler, and Xu Wang. 2023. MeetScript: Designing Transcript-based Interactions to Support Active Participation in Group Video Meetings. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2 (2023), 347:1–347:32. doi:10.1145/3610196

[17] Irene-Angelica Chounta, Alexander Nolte, Tobias Hecking, Rosta Farzan, and Thomas Herrmann. 2017. When to say "Enough is Enough!": A Study on the Evolution of Collaboratively Created Process Models. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (2017), 33:1–33:21. doi:10.1145/3134668

[18] Jeff Conklin. 2005. *Dialogue Mapping: Building Shared Understanding of Wicked Problems*. John Wiley & Sons, Inc., USA.

[19] Jeff Conklin and Michael L. Begeman. 1988. gIBIS: a hypertext tool for exploratory policy discussion. *ACM Transactions on Information Systems* 6, 4 (1988), 303–331. doi:10.1145/58566.59297

[20] Leonardo Lucio Custode and Giovanni Iacca. 2022. Interpretable AI for policy-making in pandemics. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '22)*. Association for Computing Machinery, New York, NY, USA, 1763–1769. doi:10.1145/3520304.3533959

[21] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies: why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces (IUI '93)*. Association for Computing Machinery, New York, NY,

USA, 193–200. doi:10.1145/169891.169968

[22] Sunhao Dai, Chen Xu, Shicheng Xu, Liang Pang, Zhenhua Dong, and Jun Xu. 2024. Bias and Unfairness in Information Retrieval Systems: New Challenges in the LLM Era. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*. Association for Computing Machinery, New York, NY, USA, 6437–6447. doi:10.1145/3637528.3671458 event-place: Barcelona, Spain.

[23] E. De Bono. 1985. *Six Thinking Hats: The Power of Focused Thinking*. International Center for Creative Thinking. https://books.google.com/books?id=M6llAAAACAAJ

[24] André L. Delbecq and Andrew H. Van de Ven. 1971. A Group Process Model for Problem Identification and Program Planning. *The Journal of Applied Behavioral Science* 7, 4 (1971), 466–492. doi:10.1177/002188637100700404 _eprint: https://doi.org/10.1177/002188637100700404.

[25] Nobel Dhar, Bobin Deng, Dan Lo, Xiaofeng Wu, Liang Zhao, and Kun Suo. 2024. An Empirical Analysis and Resource Footprint Study of Deploying Large Language Models on Edge Devices. In *Proceedings of the 2024 ACM Southeast Conference (ACMSE '24)*. Association for Computing Machinery, New York, NY, USA, 69–76. doi:10.1145/3603287.3651205 event-place: Marietta, GA, USA.

[26] Michael Diehl and Wolfgang Stroebe. 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology* 53, 3 (1987), 497–509. doi:10.1037/0022-3514.53.3.497 Place: US Publisher: American Psychological Association.

[27] Hyo Jin Do, Ha-Kyung Kong, Jaewook Lee, and Brian P. Bailey. 2022. How Should the Agent Communicate to the Group? Communication Strategies of a Conversational Agent in Group Chat Discussions. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (2022), 387:1–387:23. doi:10.1145/3555112

[28] Hyo Jin Do, Ha-Kyung Kong, Pooja Tetali, Jaewook Lee, and Brian P. Bailey. 2023. To Err is AI: Imperfect Interventions and Repair in a Conversational Agent Facilitating Group Chat Discussions. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1 (2023), 99:1–99:23. doi:10.1145/3579532

[29] Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, and Jonathan Larson. 2024. From Local to Global: A Graph RAG Approach to Query-Focused Summarization. doi:10.48550/arXiv.2404.16130 arXiv:2404.16130 [cs].

[30] Clarence Ellis and Jacques Wainer. 1994. A conceptual model of groupware. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (CSCW '94)*. Association for Computing Machinery, New York, NY, USA, 79–88. doi:10.1145/192844.192878 event-place: Chapel Hill, North Carolina, USA.

[31] Zhifu Gao, Zerui Li, Jiaming Wang, Haoneng Luo, Xian Shi, Mengzhe Chen, Yabin Li, Lingyun Zuo, Zhihao Du, and Shiliang Zhang. 2023. FunASR: A Fundamental End-to-End Speech Recognition Toolkit. 1593–1597. doi:10.21437/Interspeech.2023-1428

[32] Diether Gebert, Sabine Boerner, and Eric Kearney. 2010. Fostering Team Innovation: Why Is It Important to Combine Opposing Action Strategies? *Organization Science* 21, 3 (June 2010), 593–608. doi:10.1287/orsc.1090.0485 Publisher: INFORMS.

[33] David Glukhov, Ilia Shumailov, Yarin Gal, Nicolas Papernot, and Vardan Papyan. 2023. LLM Censorship: A Machine Learning Challenge or a Computer Security Problem? doi:10.48550/arXiv.2307.10719 arXiv:2307.10719 [cs].

[34] Catarina Gralha, Daniela Damian, Anthony I. (Tony) Wasserman, Miguel Goulão, and João Araújo. 2018. The evolution of requirements practices in software startups. In *Proceedings of the 40th International Conference on Software Engineering (ICSE '18)*. Association for Computing Machinery, New York, NY, USA, 823–833. doi:10.1145/3180155.3180158

[35] Vera Hagemann and Annette Kluge. 2017. Complex Problem Solving in Teams: The Impact of Collective Orientation on Team Process Demands. *Frontiers in Psychology* 8 (Sept. 2017). doi:10.3389/fpsyg.2017.01730 Publisher: Frontiers.

[36] Jawad Haqbeen, Sofia Sahab, and Takayuki Ito. 2024. From Interactive Idea Development to Solution Planning: An AI-assisted Discussion Facilitation Timeline for Collaborative Planning. In *Proceedings of the 25th Annual International Conference on Digital Government Research (dg.o '24)*. Association for Computing Machinery, New York, NY, USA, 1026–1029. doi:10.1145/3657054.3659120

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

[38] John Heron. 1999. *The Complete Facilitator's Handbook*. Kogan Page. Google-Books-ID: W2Kev6TqdtoC.

[39] Emily Hindalong, Jordon Johnson, Giuseppe Carenini, and Tamara Munzner. 2022. Abstractions for Visualizing Preferences in Group Decisions. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1 (2022), 49:1–49:44. doi:10.1145/3512896

[40] Stefan Holtel. 2024. How to Crack Complex, Ill-Defined, Nonimmediate Problems by Issue Trees: McKinsey on a Shoestring: Simple Patterns for Root Cause Analysis. In *Proceedings of the 28th European Conference on Pattern Languages of Programs (EuroPLoP '23)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3628034.3628050

[41] Yasaman Hosseinkashi, Lev Tankelevitch, Jamie Pool, Ross Cutler, and Chinmaya Madan. 2024. Meeting Effectiveness and Inclusiveness: Large-scale Measurement, Identification of Key Features, and Prediction in Real-world Remote Meetings. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (2024), 93:1–93:39. doi:10.1145/3637370

[42] Bernd Huber, Stuart Shieber, and Krzysztof Z. Gajos. 2019. Automatically Analyzing Brainstorming Language Behavior with Meeter. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (2019), 30:1–30:17. doi:10.1145/3359132

[43] Erik Johannes Husom, Arda Goknil, Merve Astekin, Lwin Khin Shar, Andre Kåsen, Sagar Sen, Benedikt Andreas Mithassel, and Ahmet Soylu. 2025. Sustainable LLM Inference for Edge AI: Evaluating Quantized LLMs for Energy Efficiency, Output Accuracy, and Inference Latency. doi:10.48550/arXiv.2504.03360 arXiv:2504.03360 [cs].

[44] Shota Imamura, Hirotaka Hiraki, and Jun Rekimoto. 2024. Serendipity Wall: A Discussion Support System Using Real-time Speech Recognition and Large Language Model. In *Proceedings of the Augmented Humans International Conference 2024 (AHs '24)*. Association for Computing Machinery, New York, NY, USA, 237–247. doi:10.1145/3652920.3652931

[45] Irving L. Janis. 1971. Groupthink: The desperate drive for consensus at any cost. *Psychology Today* 5, 6 (Nov. 1971), 43–76.

[46] Peiling Jiang, Jude Rayan, Steven P. Dow, and Haijun Xia. 2023. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–20. doi:10.1145/3586183.3606737 arXiv:2305.11473 [cs].

[47] Peng Jiang, Christian Sonne, Wangliang Li, Fengqi You, and Siming You. 2024. Preventing the Immense Increase in the Life-Cycle Energy and Carbon Footprints of LLM-Powered Intelligent Chatbots. *Engineering* 40 (2024), 202–210. doi:10.1016/j.eng.2024.04.002

[48] Zhihan Jiang, Jinyang Liu, Zhuangbin Chen, Yichen Li, Junjie Huang, Yintong Huo, Pinjia He, Jiazhen Gu, and Michael R. Lyu. 2024. LILAC: Log Parsing using LLMs with Adaptive Parsing Cache. *Proc. ACM Softw. Eng.* 1, FSE (2024), 7:137–7:160. doi:10.1145/3643733

[49] Heather Kanuka, Liam Rourke, and Elaine Laflamme. 2007. The influence of instructional methods on the quality of online discussion. *British Journal of Educational Technology* 38, 2 (2007), 260–271. doi:10.1111/j.1467-8535.2006.00620.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8535.2006.00620.x.

[50] Pranav Khadpe, Lindy Le, Kate Nowak, Shamsi T Iqbal, and Jina Suh. 2024. DISCERN: Designing Decision Support Interfaces to Investigate the Complexities of Workplace Social Decision-Making With Line Managers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. doi:10.1145/3613904.3642685

[51] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator Chatbot for Deliberative Discussion: Effects of Discussion Structure and Discussant Facilitation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (2021), 87:1–87:26. doi:10.1145/3449161

[52] Tae Soo Kim, Nitesh Goyal, Jeongyeon Kim, Juho Kim, and Sungsoo Ray Hong. 2021. Supporting Collaborative Sequencing of Small Groups through Visual Awareness. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1 (2021), 176:1–176:29. doi:10.1145/3449250

[53] Werner Kunz and Horst W. J. Rittel. 1970. *Issues as elements of information systems.* Institute of Urban and Regional Development, University of California, Berkeley. OCLC: 5065959.

[54] John Lawrence and Chris Reed. 2019. Argument Mining: A Survey. *Computational Linguistics* 45, 4 (Dec. 2019), 765–818. doi:10.1162/coli_a_00364 Publisher: MIT Press.

[55] Yinghao Li, Rampi Ramprasad, and Chao Zhang. 2024. A Simple but Effective Approach to Improve Structured Language Model Output for Information Extraction. doi:10.48550/arXiv.2402.13364 arXiv:2402.13364 [cs].

[56] Marco Lippi and Paolo Torroni. 2016. Argumentation Mining: State of the Art and Emerging Trends. *ACM Transactions on Internet Technology* 16, 2 (2016), 10:1–10:25. doi:10.1145/2850417

[57] Yiren Liu, Si Chen, Haocong Cheng, Mengxia Yu, Xiao Ran, Andrew Mo, Yiliu Tang, and Yun Huang. 2024. How AI Processing Delays Foster Creativity: Exploring Research Question Co-Creation with an LLM-based Agent. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–25. doi:10.1145/3613904.3642698

[58] Todor Markov, Chong Zhang, Sandhini Agarwal, Florentine Eloundou Nekoul, Theodore Lee, Steven Adler, Angela Jiang, and Lilian Weng. 2023. A holistic approach to undesired content detection in the real world. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'23/IAAI'23/EAAI'23)*. AAAI Press. doi:10.1609/aaai.v37i12.26752

[59] Lynne M. Markus. 2001. Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success. *Journal of Management Information Systems* (May 2001). https://www.tandfonline.com/doi/abs/10.1080/07421222.2001.11045671 Publisher: Routledge.

[60] Suéllen Martinelli, Tiago Silva da Silva, and Luciana Zaina. 2024. Challenges of UX research and long-term UX: A multiple case study with software startups. In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction (NordiCHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3679318.3685381

[61] Elinor Mizrahi, Noa Danzig, and Goren Gordon. 2022. vRobotator: A Virtual Robot Facilitator of Small Group Discussions for K-12. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (2022), 472:1–472:22. doi:10.1145/3555573

[62] Lisette Munneke, Jerry Andriessen, Gellof Kanselaar, and Paul Kirschner. 2007. Supporting interactive argumentation: Influence of representational tools on discussing a wicked problem. *Computers in Human Behavior* 23, 3 (May 2007), 1072–1088. doi:10.1016/j.chb.2006.10.003

[63] Moin Nadeem, Anna Bethke, and Siva Reddy. 2021. StereoSet: Measuring stereotypical bias in pretrained language models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 5356–5371. doi:10.18653/v1/2021.acl-long.416

[64] Tricia J. Ngoon, https://orcid.org/0000-0002-3749-0897, View Profile, S Sushil, https://orcid.org/0000-0003-3120-7349, View Profile, Angela E.B. Stewart, https://orcid.org/0000-0002-6004-9266, View Profile, Ung-Sang Lee, https://orcid.org/0000-0001-6580-1317, View Profile, Saranya Venkatraman, https://orcid.org/0009-0009-0284-1587, View Profile, Neil Thawani, https://orcid.org/0009-0003-2080-9085, View Profile, Prasenjit Mitra, https://orcid.org/0000-0002-7530-9497, View Profile, Sherice Clarke, https://orcid.org/0000-0003-1151-662X, View Profile, John Zimmerman, https://orcid.org/0000-0001-5299-8157, View Profile, Amy Ogan, https://orcid.org/0000-0003-2671-6149, and View Profile. 2024. ClassInSight: Designing Conversation Support Tools to Visualize Classroom Discussion for Personalized Teacher Professional Development. *Proceedings of the CHI Conference on Human Factors in Computing Systems* (May 2024), 1–15. doi:10.1145/3613904.3642487

[65] Karin Niemantsverdriet and Thomas Erickson. 2017. Recurring Meetings: An Experiential Account of Repeating Meetings in a Large Organization. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (2017), 84:1–84:17. doi:10.1145/3134719

[66] Joseph D. Novak and Alberto J. Cañas. 2006. *The Theory Underlying Concept Maps and How to Construct and Use Them.* research report 2006-01 Rev 2008-01. Florida Institute for Human and Machine Cognition. http://cmap.ihmc.us/Publications/ResearchPapers/TheoryCmaps/TheoryUnderlyingConceptMaps.htm

[67] OpenAI, Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, A. J. Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Mądry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, Alex Nichol, Alex Paino, Alex Renzin, Alex Tachard Passos, Alexander Kirillov, Alexi Christakis, Alexis Conneau, Ali Kamali, Allan Jabri, Allison Moyer, Allison Tam, Amadou Crookes, Amin Tootoochian, Amin Tootoonchian, Ananya Kumar, Andrea Vallone, Andrej Karpathy, Andrew Braunstein, Andrew Cann, Andrew Codispoti, Andrew Galu, Andrew Kondrich, Andrew Tulloch, Andrey Mishchenko, Angela Baek, Angela Jiang, Antoine Pelisse, Antonia Woodford, Anuj Gosalia, Arka Dhar, Ashley Pantuliano, Avi Nayak, Avital Oliver, Barret Zoph, Behrooz Ghorbani, Ben Leimberger, Ben Rossen, Ben Sokolowsky, Ben Wang, Benjamin Zweig, Beth Hoover, Blake Samic, Bob McGrew, Bobby Spero, Bogo Giertler, Bowen Cheng, Brad Lightcap, Brandon Walkin, Brendan Quinn, Brian Guarraci, Brian Hsu, Bright Kellogg, Brydon Eastman, Camillo Lugaresi, Carroll Wainwright, Cary Bassin, Cary Hudson, Casey Chu, Chad Nelson, Chak Li, Chan Jun Shern, Channing Conger, Charlotte Barette, Chelsea Voss, Chen Ding, Cheng Lu, Chong Zhang, Chris Beaumont, Chris Hallacy, Chris Koch, Christian Gibson, Christina Kim, Christine Choi, Christine McLeavey, Christopher Hesse, Claudia Fischer, Clemens Winter, Coley Czarnecki, Colin Jarvis, Colin Wei, Constantin Koumouzelis, Dane Sherburn, Daniel Kappler, Daniel Levin, Daniel Levy, David Carr, David Farhi, David Mely, David Robinson, David Sasaki, Denny Jin, Dev Valladares, Dimitris Tsipras, Doug Li, Duc Phong Nguyen, Duncan Findlay, Edede Oiwoh, Edmund Wong, Ehsan Asdar, Elizabeth Proehl, Elizabeth Yang, Eric Antonow, Eric Kramer, Eric Peterson, Eric Sigler, Eric Wallace, Eugene Brevdo, Evan Mays, Farzad Khorasani, Felipe Petroski Such, Filippo Raso, Francis Zhang, Fred von Lohmann, Freddie Sulit, Gabriel Goh, Gene Oden, Geoff Salmon, Giulio Starace, Greg Brockman, Hadi Salman, Haiming Bao, Haitang Hu, Hannah Wong, Haoyu Wang, Heather Schmidt, Heather Whitney, Heewoo Jun, Hendrik Kirchner, Henrique Ponde de Oliveira Pinto, Hongyu Ren, Huiwen Chang, Hyung Won Chung, Ian Kivlichan, Ian O'Connell, Ian O'Connell, Ian Osband, Ian Silber, Ian Sohl, Ibrahim Okuyucu, Ikai Lan, Ilya Kostrikov, Ilya Sutskever, Ingmar Kanitscheider, Ishaan Gulrajani, Jacob Coxon, Jacob Menick, Jakub Pachocki, James Aung, James Betker, James Crooks, James Lennon, Jamie Kiros, Jan Leike, Jane Park, Jason Kwon, Jason Phang, Jason Teplitz, Jason Wei, Jason Wolfe, Jay Chen, Jeff Harris, Jenia Varavva, Jessica Gan Lee, Jessica Shieh, Ji Lin, Jiahui Yu, Jiayi Weng, Jie Tang, Jieqi Yu, Joanne Jang, Joaquin Quinonero Candela, Joe Beutler, Joe Landers, Joel Parish, Johannes Heidecke, John Schulman, Jonathan Lachman, Jonathan McKay, Jonathan Uesato, Jonathan Ward, Jong Wook Kim, Joost Huizinga, Jordan Sitkin, Jos Kraaijeveld, Josh Gross, Josh Kaplan, Josh Snyder, Joshua Achiam, Joy Jiao, Joyce Lee, Juntang Zhuang, Justyn Harriman, Kai Fricke, Kai Hayashi, Karan Singhal, Katy Shi, Kavin Karthik, Kayla Wood, Kendra Rimbach, Kenny Hsu, Kenny Nguyen, Keren Gu-Lemberg, Kevin Button, Kevin Liu, Kiel Howe, Krithika Muthukumar, Kyle Luther, Lama Ahmad, Larry Kai, Lauren Itow, Lauren Workman, Leher Pathak, Leo Chen, Li Jing, Lia Guy, Liam Fedus, Liang Zhou, Lien Mamitsuka, Lilian Weng, Lindsay McCallum, Lindsey Held, Long Ouyang, Louis Feuvrier, Lu Zhang, Lukas Kondraciuk, Lukasz Kaiser, Luke Hewitt, Luke Metz, Lyric Doshi, Mada Aflak, Maddie Simens, Madelaine Boyd, Madeleine Thompson, Marat Dukhan, Mark Chen, Mark Gray, Mark Hudnall, Marvin Zhang,

Marwan Aljubeh, Mateusz Litwin, Matthew Zeng, Max Johnson, Maya Shetty, Mayank Gupta, Meghan Shah, Mehmet Yatbaz, Meng Jia Yang, Mengchao Zhong, Mia Glaese, Mianna Chen, Michael Janner, Michael Lampe, Michael Petrov, Michael Wu, Michele Wang, Michelle Fradin, Michelle Pokrass, Miguel Castro, Miguel Oom Temudo de Castro, Mikhail Pavlov, Miles Brundage, Miles Wang, Minal Khan, Mira Murati, Mo Bavarian, Molly Lin, Murat Yesildal, Nacho Soto, Natalia Gimelshein, Natalie Cone, Natalie Staudacher, Natalie Summers, Natan LaFontaine, Neil Chowdhury, Nick Ryder, Nick Stathas, Nick Turley, Nik Tezak, Niko Felix, Nithanth Kudige, Nitish Keskar, Noah Deutsch, Noel Bundick, Nora Puckett, Ofir Nachum, Ola Okelola, Oleg Boiko, Oleg Murk, Oliver Jaffe, Olivia Watkins, Olivier Godement, Owen Campbell-Moore, Patrick Chao, Paul McMillan, Pavel Belov, Peng Su, Peter Bak, Peter Bakkum, Peter Deng, Peter Dolan, Peter Hoeschele, Peter Welinder, Phil Tillet, Philip Pronin, Philippe Tillet, Prafulla Dhariwal, Qiming Yuan, Rachel Dias, Rachel Lim, Rahul Arora, Rajan Troll, Randall Lin, Rapha Gontijo Lopes, Raul Puri, Reah Miyara, Reimar Leike, Renaud Gaubert, Reza Zamani, Ricky Wang, Rob Donnelly, Rob Honsby, Rocky Smith, Rohan Sahai, Rohit Ramchandani, Romain Huet, Rory Carmichael, Rowan Zellers, Roy Chen, Ruby Chen, Ruslan Nigmatullin, Ryan Cheu, Saachi Jain, Sam Altman, Sam Schoenholz, Sam Toizer, Samuel Miserendino, Sandhini Agarwal, Sara Culver, Scott Ethersmith, Scott Gray, Sean Grove, Sean Metzger, Shamez Hermani, Shantanu Jain, Shengjia Zhao, Sherwin Wu, Shino Jomoto, Shirong Wu, Shuaiqi, Xia, Sonia Phene, Spencer Papay, Srinivas Narayanan, Steve Coffey, Steve Lee, Stewart Hall, Suchir Balaji, Tal Broda, Tal Stramer, Tao Xu, Tarun Gogineni, Taya Christianson, Ted Sanders, Tejal Patwardhan, Thomas Cunninghman, Thomas Degry, Thomas Dimson, Thomas Raoux, Thomas Shadwell, Tianhao Zheng, Todd Underwood, Todor Markov, Toki Sherbakov, Tom Rubin, Tom Stasi, Tomer Kaftan, Tristan Heywood, Troy Peterson, Tyce Walters, Tyna Eloundou, Valerie Qi, Veit Moeller, Vinnie Monaco, Vishal Kuo, Vlad Fomenko, Wayne Chang, Weiyi Zheng, Wenda Zhou, Wesam Manassra, Will Sheu, Wojciech Zaremba, Yash Patil, Yilei Qian, Yongjik Kim, Youlong Cheng, Yu Zhang, Yuchen He, Yuchen Zhang, Yujia Jin, Yunxing Dai, and Yury Malkov. 2024. GPT-4o System Card. doi:10.48550/arXiv.2410.21276 arXiv:2410.21276 [cs].

[68] Otter.ai. 2024. Otter.ai - AI Meeting Note Taker & Real-time AI Transcription. https://otter.ai/

[69] Alexander Pollatsek and Arnold D. Well. 1995. On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21, 3 (1995), 785–794. doi:10.1037/0278-7393.21.3.785 Place: US Publisher: American Psychological Association.

[70] Sajjadur Rahman, Pao Siangliulue, and Adam Marcus. 2020. MixTAPE: Mixed-initiative Team Action Plan Creation Through Semi-structured Notes, Automatic Task Generation, and Task Classification. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (2020), 169:1–169:26. doi:10.1145/3415240

[71] Horst W. J. Rittel and Melvin M. Webber. 1973. Dilemmas in a general theory of planning. *Policy Sciences* 4, 2 (June 1973), 155–169. doi:10.1007/BF01405730

[72] Jeremy Roschelle and Stephanie D. Teasley. 1995. The Construction of Shared Knowledge in Collaborative Problem Solving. In *Computer Supported Collaborative Learning*, Claire O'Malley (Ed.). Springer, Berlin, Heidelberg, 69–97. doi:10.1007/978-3-642-85098-1_5

[73] Baruch B. Schwarz, Yair Neuman, Julia Gil, and Merav Ilya. 2003. Construction of collective and individual knowledge in argumentative activity. *Journal of the Learning Sciences* 12, 2 (2003), 219–256. doi:10.1207/S15327809JLS1202_3 Place: US Publisher: Lawrence Erlbaum.

[74] Frank M. Shipman and Catherine C. Marshall. 1999. Formality Considered Harmful: Experiences, Emerging Themes, and Directions on the Use of Formal Representations in Interactive Systems. *Computer Supported Cooperative Work (CSCW)* 8, 4 (Dec. 1999), 333–352. doi:10.1023/A:1008716330212

[75] Aditi Singh, Nirmal Prakashbhai Patel, Abul Ehtesham, Saket Kumar, and Tala Talaei Khoei. 2025. A Survey of Sustainability in Large Language Models: Applications, Economics, and Challenges. In *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)*. 00008–00014. doi:10.1109/CCWC62904.2025.10903774

[76] Alison Specht and Kevin Crowston. 2022. Interdisciplinary collaboration from diverse science teams can produce significant outcomes. *PLOS ONE* 17, 11 (Nov. 2022), e0278043. doi:10.1371/journal.pone.0278043 Publisher: Public Library of Science.

[77] Garold Stasser and William Titus. 1985. Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology* 48, 6 (1985), 1467–1478. doi:10.1037/0022-3514.48.6.1467 Place: US Publisher: American Psychological Association.

[78] Angela E.B. Stewart, Hana Vrzakova, Chen Sun, Jade Yonehiro, Cathlyn Adele Stone, Nicholas D. Duran, Valerie Shute, and Sidney K. D'Mello. 2019. I Say, You Say, We Say: Using Spoken Language to Model Socio-Cognitive Processes during Computer-Supported Collaborative Problem Solving. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (2019), 194:1–194:19. doi:10.1145/3359296

[79] Daniel D. Suthers. 2003. Representational Guidance for Collaborative Inquiry. In *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments*, Jerry Andriessen, Michael Baker, and Dan Suthers (Eds.). Springer Netherlands, Dordrecht, 27–46. doi:10.1007/978-94-017-0781-7_2

[80] Daniel D. Suthers and Christopher D. Hundhausen. 2003. An Experimental Study of the Effects of Representational Guidance on Collaborative Learning Processes. *Journal of the Learning Sciences* 12, 2 (April 2003), 183–218. doi:10.1207/S15327809JLS1202_2 Publisher: Routledge.

[81] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12, 2 (April 1988), 257–285. doi:10.1016/0364-0213(88)90023-7

[82] 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 (2023), 351:1–351:25. doi:10.1145/3610200

[83] Sunny Tian, Amy X. Zhang, and David Karger. 2021. A System for Interleaving Discussion and Summarization in Online Collaboration. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3 (2021), 241:1–241:27. doi:10.1145/3432940

[84] R. Scott Tindale and Tatsuya Kameda. 2000. 'Social sharedness' as a unifying theme for information processing in groups. *Group Processes & Intergroup Relations* 3, 2 (2000), 123–140. doi:10.1177/1368430200003002002 Place: US Publisher: Sage Publications.

[85] Jan M. van Bruggen, Henny P. A. Boshuizen, and Paul A. Kirschner. 2003. A Cognitive Framework for Cooperative Problem Solving with Argument Visualization. In *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, Paul A. Kirschner, Simon J. Buckingham Shum, and Chad S. Carr (Eds.). Springer, London, 25–47. doi:10.1007/978-1-4471-0037-9_2

[86] Henrika Adriana Theodora Van der Meijden. 2005. *Knowledge construction through CSCL: Student elaborations in synchronous and three-dimensional learning environments.* sl: sn.

[87] Vishal B. Vatnani and Jordan B. Barlow. 2024. The Effect of Individual Traits on Emerging Roles in Synchronous Computer-Mediated Groups. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (2024), 181:1–181:22. doi:10.1145/3641020

[88] Radhakrishnan Venkatakrishnan, Emrah Tanyildizi, and M. Abdullah Canbaz. 2024. Semantic interlinking of Immigration Data using LLMs for Knowledge Graph Construction. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24)*. Association for Computing Machinery, New York, NY, USA, 605–608. doi:10.1145/3589335.3651557

[89] Aishwarya Vijayan. 2024. A Prompt Engineering Approach for Structured Data Extraction from Unstructured Text Using Conversational LLMs. In *Proceedings of the 2023 6th International Conference on Algorithms, Computing and Artificial Intelligence (ACAI '23)*. Association for Computing Machinery, New York, NY, USA, 183–189. doi:10.1145/3639631.3639663

[90] Yixin Wan, George Pu, Jiao Sun, Aparna Garimella, Kai-Wei Chang, and Nanyun Peng. 2023. "Kelly is a Warm Person, Joseph is a Role Model": Gender Biases in LLM-Generated Reference Letters. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 3730–3748. doi:10.18653/v1/2023.findings-emnlp.243

[91] Junjie Wang, Dan Yang, Binbin Hu, Yue Shen, Wen Zhang, and Jinjie Gu. 2024. Know Your Needs Better: Towards Structured Understanding of Marketer Demands with Analogical Reasoning Augmented LLMs. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)*. Association for Computing Machinery, New York, NY, USA, 5860–5871. doi:10.1145/3637528.3671583

[92] 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 (2024), 35:1–35:29. doi:10.1145/3637312

[93] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Qin Liu, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wensen Cheng, Qi Zhang, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huan, and Tao Gui. 2023. The Rise and Potential of Large Language Model Based Agents: A Survey. doi:10.48550/arXiv.2309.07864 arXiv:2309.07864 [cs].

[94] Haijun Xia, Tony Wang, Aditya Gunturu, Peiling Jiang, William Duan, and Xiaoshuo Yao. 2023. CrossTalk: Intelligent Substrates for Language-Oriented Interaction in Video-Based Communication and Collaboration. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3586183.3606773

[95] Amy X. Zhang and Justin Cranshaw. 2018. Making Sense of Group Chat through Collaborative Tagging and Summarization. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (2018), 196:1–196:27. doi:10.1145/3274465

[96] Chao Zhang, Yuren Mao, Yijiang Fan, Yu Mi, Yunjun Gao, Lu Chen, Dongfang Lou, and Jinshu Lin. 2024. FinSQL: Model-Agnostic LLMs-based Text-to-SQL Framework for Financial Analysis. doi:10.48550/arXiv.2401.10506 arXiv:2401.10506 [cs].

[97] Qianqia (Queenie) Zhang, Soya Park, Michael Muller, and David R. Karger. 2024. "How fancy you are to make us use your fancy tool": Coordinating Individuals' Tool Preference over Group Boundaries. *Proc. ACM Hum.-Comput. Interact.* 8, GROUP (2024), 4:1–4:31. doi:10.1145/3633069

[98] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A Survey of Large Language Models. doi:10.48550/arXiv.2303.18223 arXiv:2303.18223 [cs].

[99] Yuqi Zhu, Xiaohan Wang, Jing Chen, Shuofei Qiao, Yixin Ou, Yunzhi Yao, Shumin Deng, Huajun Chen, and Ningyu Zhang. 2024. LLMs for Knowledge Graph Construction and Reasoning: Recent Capabilities and Future Opportunities. doi:10.48550/arXiv.2305.13168 arXiv:2305.13168 [cs].

A Additional Evaluation Results

Subjective ratings evaluating privacy and content trust are presented in Fig. 8. The results suggest that participants highly understood the necessity of data collection and were comfortable with the content generated by both EchoMind and AutoDoc. The cognitive load results are displayed in Table 1. These findings indicate a significant advantage of EchoMind in terms of reducing temporal demands.

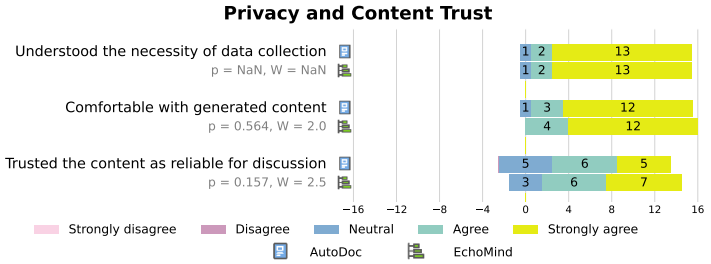


Fig. 8. Subjective ratings of privacy and content trust metrics.

Table 1. Results of NASA-TLX ratings. Higher ratings indicate better performance. ** indicates $p < .01$.

	AutoDoc		**EchoMind**		**Wilcoxon test**	
	Mean	*SD*	*Mean*	*SD*	*W*	*p*
Mental demands	4.19	1.60	4.38	1.02	35.50	0.773
Physical demands	4.00	1.63	4.75	1.39	13.00	0.071
Temporal demands	3.94	1.39	5.25	1.13	4.00	**
Performance	4.75	1.06	5.31	1.08	3.00	0.054
Effort	3.75	1.13	4.00	1.03	13.00	0.476
Frustration	4.25	1.29	5.00	0.97	12.00	0.055

B System Implementation Details

EchoMind and AutoDoc are both built on a web-based frontend and a server backend. The frontend utilizes React Flow (for EchoMind) and TipTap (for AutoDoc) as an interactive editor to build the user interface. The backend runs on a Python server with FastAPI and handles the system's core functions.

Real-time communication between the system's frontend and backend is facilitated via HTTP requests and Socket.IO. The frontend transmits audio data to the backend. The backend utilizes the open-source FunASR [31] speech recognition model to transcribe the audio and subsequently returns the transcription results to the frontend.

Once the transcriptions accumulate to a set threshold, the backend uses OpenAI's GPT-4o model `gpt-4o-2024-05-13` for dialogue analysis, configured with a temperature of 0.2 to ensure high accuracy and consistency in the generated responses. To enhance reliability in production environments, we implement an exponential backoff retry mechanism (maximum 3 attempts) for API requests.

B.1 EchoMind

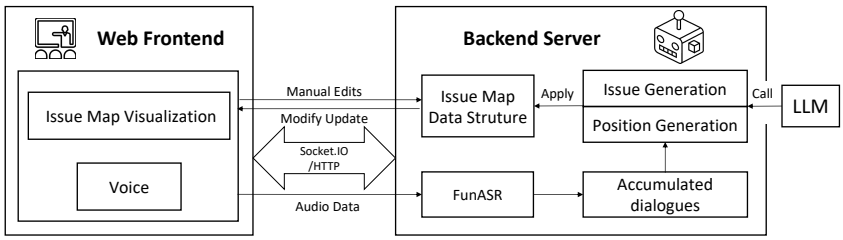


Fig. 9. System overview of EchoMind.

EchoMind is an agent designed to automatically generate positions and issues based on user dialogue, as illustrated in Fig. 9. When the cumulative character count of new dialogue reaches 50 Chinese characters, the system generates a summary of the positions relevant to the ongoing issue. It also identifies derivative issues based on these positions. This approach ensures that the conversation remains focused on the current issue while allowing the exploration of related issues.

B.2 AutoDoc

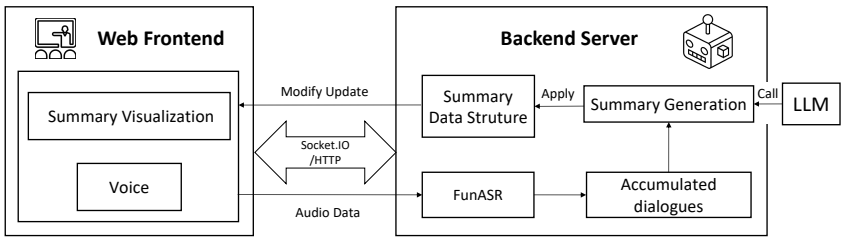


Fig. 10. System overview of AutoDoc.

AutoDoc is an agent designed to automatically generate summaries based on participants' dialogue, as depicted in Fig. 10. When the cumulative word count of newly added dialogue reaches 200 Chinese characters, AutoDoc extracts and highlights key points from the recent additions.

The user interface of AutoDoc is illustrated in Fig. 11. On the left side, the transcription results are displayed, while the document editor is located on the right side. The summary generated by AutoDoc will appear in a toggle format at the bottom of the interface. Users have the option to add, delete, or edit the generated content.

C Prompts

The prompts follow a dialogue format: the `SYSTEM` message defines the task requirements, while the initial `USER` message includes the runtime input and any supplementary notes. Variables enclosed in `%` are placeholders, which are programmatically replaced with actual values during execution.

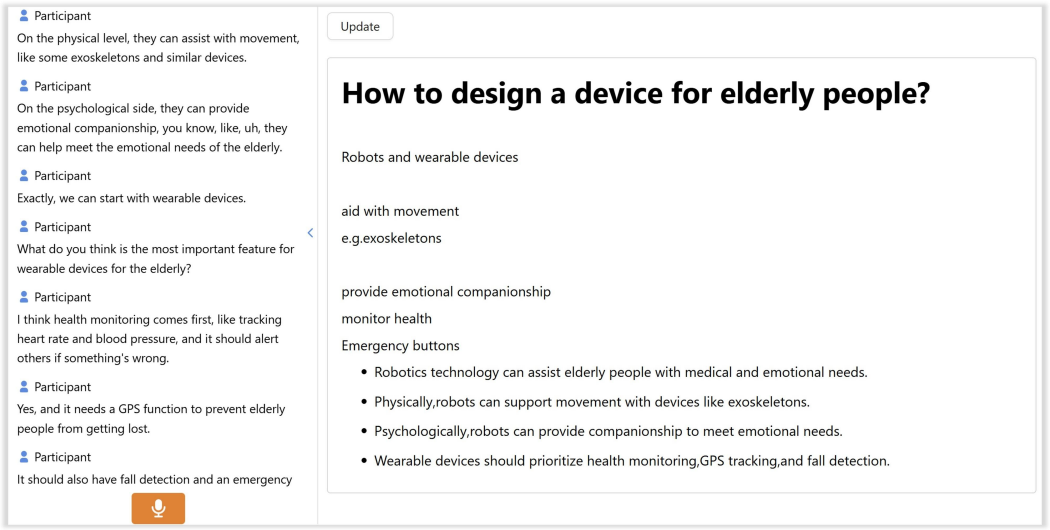


Fig. 11. The interface of AutoDoc.

C.1 Prompts of EchoMind

Prompt 1. Position generation of EchoMind.

```
SYSTEM
You are an AI meeting discussion assistant, adept at understanding discussion contents in a
     structured way to facilitate participants in better organizing their thoughts and
     expressing themselves. You are required to analyze given dialogue text using the IBIS (
     Issue Based Information System) method for issue mapping. You will need to analyze and
     modify existing issue maps and new dialogues.

## Basic Elements

- **Issue**: A question that needs exploration, presented as a concise and formal interrogative
     sentence. It defines the scope of the discussion.
- **Position**: Answers or solutions, which are direct responses to the issue, should remain
     neutral.

## Task Requirements

- When users specify an issue currently under discussion, you will need to add or modify a
     position and its note under that issue based on the existing notes and the current dialogue.

- Capture the core key points, not every detail, and do not reproduce the original text.
- Information must remain faithful to the original dialogue, with no addition of unmentioned
     elements.
- Use Chinese to generate positions and use English to write notes.

## Output Format Example
```

```
// For each position under the current issue, provide the position followed by a brief English
     note. This position content should be written in Chinese while the note should be written
     in English. The note should assess whether the discussion on that position is complete. If
     the generated position is a phrase or the corresponding dialogue is incomplete, note that
     the discussion on this position is not finished.

// Output in XML tags (put 'None' inside the tags if there is no content):

<position_and_note>
${position_full_id} position ${position content}
// ${note content}
...
1.1 position {position 1.1 content}
// {note 1.1 content}
1.4 position {position 1.4 content}
// {note 1.4 content}
...
</position_and_note>
```

USER

```
Ongoing issue chain:
<issue_chain>
%issue_chain%
</issue_chain>

Existing positions under the current issue:
<current_positions>
%current_positions%
</current_positions>

Newly added dialogue during the discussion of this issue:
<dialog>
%dialog%
</dialog>

Read the dialogue, modify or add positions under the current issue to better align with the
     dialogue content. Describe each change in natural language. Then, based on the changes,
     output all updated positions and notes under the current issue.

Note:
- You can only modify the content of modifiable positions; their numbers must remain unchanged.
- For non-modifiable positions marked as 'unmodifiable', both the number and content must remain
      unchanged.
- When adding new positions, use a larger number than the last existing position. For example,
     if the highest number is 2.6, new positions should start from 2.7.
- Deletion of positions is not allowed.
- The content in positions must be **entirely faithful to the original dialogue**; you cannot
     add new information.
- If there are **different opinions** on the same viewpoint, they should be presented within a
     single position, and the note should assess whether further discussion on this issue is
     likely.
- If there are multiple **similar ideas** related to the same viewpoint, they should be listed
     as examples within a single position, ensuring that no important information is omitted.
- If the argument related to a position is mentioned in the dialogue, you must include that
     argument in the position content.
- Do not include the original text in the positions.
- Avoid duplicate content in the positions.
```

```
- Make sure the total number of positions is as small as possible, typically within 5, unless
     the dialog explicitly mentions multiple structured aspects.
```

Prompt 2. Issue generation of EchoMind.

```
SYSTEM
You are an AI meeting discussion assistant, skilled in understanding discussion content in a
     structured manner, and capable of using the IBIS (Issue Based Information System) method to
      map issues. You need to break down issues from the given positions based on the user's
     conversation and the existing issue chain.

## Basic Elements

- Issue: A question that needs to be explored, presented in the form of a concise, formal inquiry
     . It defines the scope of the discussion.
- Position: An answer or solution, directly responding to the issue, and should remain neutral.

## Task Requirements

- The user provides the already discussed structure (issue chain), and the dialogue text (dialog)
     . When you find words in the dialogue that indicate an issue will be discussed, such as "
     let's discuss..." or similar phrases, pay special attention to the content that follows.
     Based on this discussion and your own knowledge, extract up to two issues from the
     positions listed in the positions_list. However, if the word "discuss" does **not** appear
     in the dialogue text, do not make any changes.

- For extracting issues, here's an example: For the position "Determine the internal division of
      labor for the experiment," the extracted issues could be "How to determine the internal
     division of labor for the experiment" or "What specific tasks are involved in the
     experiment."

- Be sure to evaluate the discussion value of the extracted issues. If the issue has already
     been mentioned and answered in the dialogue, discard it and do not output it. Only provide
     issues that have not been fully discussed.

## Output Format Example

Analysis:

${A detailed natural language analysis(use English)}

Sub Issue List:
<sub_issue_list>
${position_full_id} position ${Original position text}
${position_full_id}.${Issue number, starting from 1 within each position} sub_issue ${Content of
      the split issue}
${position_full_id}.${Issue number, starting from 1 within each position} sub_issue ${Content of
      the split issue}
1.1 position position_content
1.1.1 sub_issue sub_issue_content
</sub_issue_list>

USER
Currently Discussed Issue Chain:
<issue_chain>
%issue_chain%
```

```
</issue_chain>

Positions Requiring Issue Addition:
<positions_list>
%positions_list%
</positions_list>

Below is the Latest Dialogue:
<dialog>
%dialog%
</dialog>

Notes:
- Capture the core key points, not every detail, **do not include the original text**.
- Split issues should be in question form, concise, and formal.
- The split issues should be directly related to the position.
- Only generate new issues for a position if words like "discuss" are identified; **otherwise,
    do not make any changes to the corresponding position**.
- If the issue you split has a related conclusion in the dialogue, **do not output** that issue.
     Instead, include the conclusion in your analysis text, and only generate issues that have
    **not yet been concluded** and are **worth discussing**.Each issue you output must be
    carefully evaluated for its **value** before deciding whether to include it.
- Focus only on the dialogue following words like "discuss," and do not generate issues based on
     the dialogue preceding these words.
- If a position has not been discussed, **do not generate** a corresponding issue. Do not invent
     issues; only generate issues when there is a clear "discussion" signal.
- The generated issues must correspond to content explicitly mentioned in the dialogue. Please **
    detail** the issues you provide and **clearly reference** the original statements. If there
     is no clear reference in the original text, **do not output** that issue, and **do not
    suggest** an issue.
- A maximum of **1 issue** should be generated for each position.
- Please use Chinese for your **output**.
- If no issue is generated, output 'None' inside the XML tags.
- You should not change the index or content of the provided position.
```

C.2 Prompt of AutoDoc

Prompt 3. Summary generation of AutoDoc.

```
SYSTEM
You are an AI meeting discussion assistant, adept at understanding discussion contents in a
    structured way to facilitate participants in better organizing their thoughts and
    expressing themselves. You are required to summary the given dialogue text and provide a
    concise and structured summary of the key points discussed in the dialogue.

## Basic Elements

- **Summary point**: A concise sentence or paragraph that captures the core key ideas discussed
    in the dialogue, considering the context.

## Task Requirements

- Capture the core key points, not every detail, and do not reproduce the original text.
- Information must remain faithful to the original dialogue, with no addition of unmentioned
    elements.
- Use Chinese to generate summary points.
```

```
## Output Format Example

// Output in XML tags (put 'None' inside the tags if there is no content):

<summary>
- ${summary content}
- ${summary content 2}
...
</summary>

USER
Newly added dialogue during the discussion of this issue:
<dialog>
%dialog%
</dialog>

Read the dialogue, and summarize the key points discussed in the dialogue.

Note:
- The content must be **entirely faithful to the original dialogue**; you cannot add new
     information.
- Do not include the original text.
- Avoid duplicate content.
- Make sure the total number of summary points is as small as possible, typically within 5,
     unless the dialogue explicitly mentions multiple structured aspects.
```

C.3 Simulation Prompt for Automatic Focus Switch

Prompt 4. Post-hoc simulation of automatic focus switch.

```
SYSTEM
You are an AI meeting discussion assistant, adept at understanding discussion contents in a
     structured way to facilitate participants in better organizing their thoughts and
     expressing themselves. The current conversation has been mapped into an issue map, and you
     need to judge whether a new issue is being discussed based on the current text dialog.


## Output Format Example

// In <switch> you should output true or false. If a new issue needs to be switched to, output
     true; otherwise, output false.

// In <target_issue> you should output the issue number in the format [issue${issue number}] and
      the complete issue content in the format ${complete issue content}. If a new issue needs
     to be switched to, output the number of the new issue and its content; if no switch is
     needed, output the current issue's number and its content.

<switch>
${true | false}
</switch>

<target_issue>
[issue${issue number}] ${complete issue content}
</target_issue>
```

```
USER
Complete issue map:
<issue_map_full>
%issue_map_full%
</issue_map_full>

Currently discussed issue:
<current_focused_issue>
%current_focused_issue%
</current_focused_issue>

The participants' dialogue:
<dialog>
%dialog%
</dialog>

Note:
- Focus on the "trend" in the conversation, i.e., whether the participants have shifted from the
      current issue to a new discussion topic.
- If the conversation gradually diverges from the current issue and starts discussing new
     content, consider it a switch to a new issue.
- You can only choose issues from the issue_map_full; do not create new issues yourself.
- If a switch to a new issue is needed, output true, the issue number and content of the new
     issue; if no switch is needed, output false, the current issue's number and content.
```

Received October 2024; revised April 2025; accepted August 2025