Intent Summary: Supporting Interruption Recovery in Human-LLM Collaboration through Recognition-Based Interface Intervention

Anonymous EMNLP submission

Abstract

Large language models are increasingly embedded in multi-turn collaborative workflows, yet real-world interactions are frequently interrupted, disrupting users' ability to resume tasks effectively. While existing systems ensure technical continuity by preserving conversation history, they often overlook cognitive continuity—users' ability to reconstruct task context and decision intent after interruptions. In this paper, we reconceptualize interruption recovery in human–LLM collaboration as a problem of situational reconstruction rather than mere information retrieval, and introduce Intent Summary (IS), a structured interface intervention that externalizes task progress, user preferences, and pending decisions as compact cognitive anchors to support re-engagement. Through a controlled user study comparing Intent Summary (IS) with traditional linear history interfaces under a 3-minute interruption condition, we find that Intent Summary (IS) significantly reduce resumption lag and perceived workload. More importantly, our results reveal a qualitative shift in users' recovery strategies—from recall-based reconstruction to recognition-based, forward-oriented re-engagement. These findings contribute a conceptual distinction between cognitive and technical continuity, demonstrate the role of structured summary as cognitive scaffolding, and provide design implications for building interruption-resilient human–AI collaborative systems.

1 Introduction

Large language models are increasingly embedded in multi-turn collaborative workflows, supporting complex tasks such as writing and data analysis. In these settings, users often engage in extended interactions that unfold over time rather than in a single session. However, real-world usage is rarely continuous: interruptions—such as attending meetings, switching tasks, or simply taking breaks—are pervasive. When users return, they must reconstruct not only what has been said, but also what they were trying to achieve. This breakdown highlights a critical but underexplored challenge in human–LLM collaboration: how to effectively resume work after interruptions.

Prior work has largely addressed this problem by preserving conversation history, ensuring what we term technical continuity. While access to past interactions is necessary, it is often insufficient for effective resumption. Users must still engage in effortful retrospective processing—scanning lengthy dialogue, recalling prior decisions, and inferring their original intent—processes known to be cognitively demanding in interruption recovery research (Trafton et al., 2003). We argue that what is missing is cognitive continuity: support for users' ability to reconstruct task context and decision intent in a way that aligns with how humans naturally resume interrupted activities.

In this paper, we reconceptualize interruption recovery in human–LLM collaboration as a problem of situational reconstruction rather than mere information access. Drawing on theories of memory and task resumption, we posit that effective recovery depends not on re-reading past interactions, but on re-establishing high-level task structure and intent. To support this process, we introduce Intent Summary (IS), a structured interface intervention that externalizes key elements of the interaction—task progress, user preferences, and pending decisions—as compact cognitive anchors. By shifting the interaction paradigm from recall-based reconstruction to recognition-based re-engagement, IS enables users to resume work more efficiently and with reduced cognitive effort.

This work makes three primary contributions:

- Conceptualizing Cognitive Continuity in Human-LLM Collaboration: We distinguish cognitive continuity from technical continuity in conversational AI systems, arguing that pre-

serving interaction history does not guarantee users' ability to reconstruct task context after interruptions. This distinction reframes interruption recovery as a problem of situational reconstruction rather than information access.

- Introducing Intent Summary as a Design Pattern for Cognitive Anchoring: We propose IS as a structured interface intervention that externalizes task progress, user preferences, and pending decisions into compact cognitive anchors. We articulate how such summary function as a form of cognitive scaffolding that shifts users from effortful recall to recognition-based re-engagement.

- Providing Empirical Evidence on Cognitive Recovery Mechanisms: Through a controlled user study, we demonstrate that structured IS significantly reduce resumption lag and perceived workload. More importantly, our findings reveal a qualitative shift in users' recovery strategies—from retrospective information search to forward-oriented goal recognition—providing new insights into how interface design can support interruption resilience in human-AI collaboration.

2 Background and Related Work

2.1 Cognitive Foundations of Interruption Recovery

Successful task resumption following an interruption depends on the reactivation of previously encoded goals and problem states (Trafton et al., 2003; Altmann and Trafton, 2002). We define cognitive continuity as the user's ability to efficiently reconstruct their mental representation of task context, prior discourse, and decision intent after an interruption. Unlike technical continuity, which merely ensures that the system retains interaction data, cognitive continuity focuses on the preservation and recovery efficiency of the user's mental model.

According to the Memory for Goals model (Altmann and Trafton, 2002), interruptions lead to the decay of goal memory, creating a "problem state bottleneck" that increases resumption lag (Borst et al., 2015). Endsley's theory of situation awareness further suggests that interruptions disrupt the user's perception and projection of task states (Endsley, 1995). When users return from a disturbance, they must not only reposition themselves but also reconstruct where they need to go and why they are making specific decisions. This reconstruction process imposes a significant cognitive load (Sweller, 1988), as users must utilize limited working memory resources to reintegrate prior information (Ericsson and Kintsch, 1995; Miller, 1956).

2.2 Context Management in Conversational AI

Current mainstream LLMs, such as GPT-4, achieve technical continuity through extensive context windows that accommodate tens of thousands of tokens (OpenAI, 2023). The implicit design assumption is that as long as the conversation history is visible, users can independently locate the interruption point and reconstruct the task context. However, this assumption confuses physical availability with psychological accessibility (Ericsson and Kintsch, 1995).

Human working memory is inherently limited, typically processing only five to nine information chunks at a time (Miller, 1956). In conversational interfaces, the lack of visual anchors—such as cursor positions or progress indicators—exacerbates this challenge (Bailey and Konstan, 2006). Users must rely on linear scrolling and text reading to locate information, which is highly inefficient for long-turn interactions. Gnewuch et al. (2023) noted that conversational AI design has long neglected how interfaces can support these cognitive processes, focusing instead on model performance (Gnewuch et al., 2023). Technical completeness, therefore, does not automatically translate into the cognitive accessibility of context.

2.3 Interface Design for Supporting Task Recovery

Prior HCI research has explored various interface interventions to mitigate interruption effects. One line of work focuses on visual cues and progress indicators. Early studies by Myers (Myers, 1985) and Harrison et al. (Harrison et al., 2009) demonstrated that progress feedback improves user patience and subjective experience. In resumption scenarios, highlighting the last work position has been shown to significantly reduce re-entry time (Czerwinski et al., 2004). Similarly, the concept of workspace awareness highlights the importance of making task-relevant activities visible to support continuity (Gutwin and Greenberg, 2002).

Another approach involves information summarization. Structured summary have been proven

to reduce reading time and cognitive effort in information retrieval tasks (Gurevych et al., 2003). Recently, LLMs have demonstrated exceptional capabilities in distilling core intentions from long texts and iteratively refining outputs through self-feedback mechanisms (Zhang et al., 2024; Shinn et al., 2023; Madaan et al., 2023). However, existing research remains limited: most studies focus on avoiding interruptions rather than supporting recovery, and few have explored the unique challenges of human-LLM collaboration. For instance, while the CPG interface (Jeong et al., 2025) visualizes completion progress, it primarily targets user self-efficacy during execution rather than situational re-entry after an interruption.

2.4 Research Gap and Positioning

Based on the existing literature, we identify three critical research gaps:

- Shift in Perspective: Most interruption research focuses on timing and avoidance (Adamczyk and Bailey, 2004), with limited exploration into active system support for efficient recovery.

- Continuity Confusion: Current designs equate history visibility with cognitive recoverability, overlooking the massive cognitive costs of reconstructing context from unstructured, linear dialogue (see Figure 1, left).

- Lack of Specialized Support: While generic tools like progress bars or summary exist, they have not been systematically adapted to the unique "black-box" nature of human-LLM interaction.

This study addresses these gaps by proposing the IS system as a structured recovery intervention (see Figure 1, right). The IS externalizes key elements—task progress, user preferences, and pending issues—into compact cognitive anchors. By shifting the resumption process from effortful recall to recognition-based re-engagement, we aim to minimize resumption costs and support cognitive continuity in complex human-LLM collaboration.

3 Solution Approach

Below we describe our solution approach.

3.1 Design Guidelines

Grounded in our analysis of cognitive disruption during interruption recovery, we derive two core design goals for the IS system.

DG1: Support Cognitive Continuity through Externalized Anchors. Interruptions disrupt goal activation and situational awareness, requiring users to reconstruct task context from memory (Altmann and Trafton, 2002; Endsley, 1995). To mitigate this, the system should externalize key elements of the interaction—such as task progress, user preferences, and pending decisions—into persistent cognitive anchors. This reduces reliance on memory reconstruction and enables rapid re-engagement with the task.

DG2: Reduce Resumption Cost via Structured Information Chunking. Given the limitations of working memory (Miller, 1956), the system should transform unstructured conversational history into compact, structured representations. By organizing information into cognitively manageable chunks, users can recover task context through recognition rather than effortful recall, thereby reducing cognitive load and resumption lag.

To operationalize these goals without disrupting ongoing interaction, we adopt a sidebar-based layout that preserves the primary conversational workspace while presenting structured summaries in a persistent peripheral view.

3.2 System Architecture

As illustrated in Figure 2, the system adopts a dual-agent architecture that decouples task execution from context reconstruction.

Main Collaboration Agent. This agent serves as the primary interaction partner, handling real-time dialogue and task execution. It maintains the complete conversational history and supports the user in accomplishing domain-specific tasks, ensuring technical continuity.

Intent Anchoring Agent. This agent operates in parallel to monitor the interaction flow and generate structured summaries. It continuously analyzes the evolving dialogue to extract high-level task representations, including progress states, user preferences, and unresolved decisions. Upon the user's return from an interruption, the system generates and presents a structured summary panel that externalizes key task elements, supporting rapid reconstruction of task context with minimal cognitive effort.

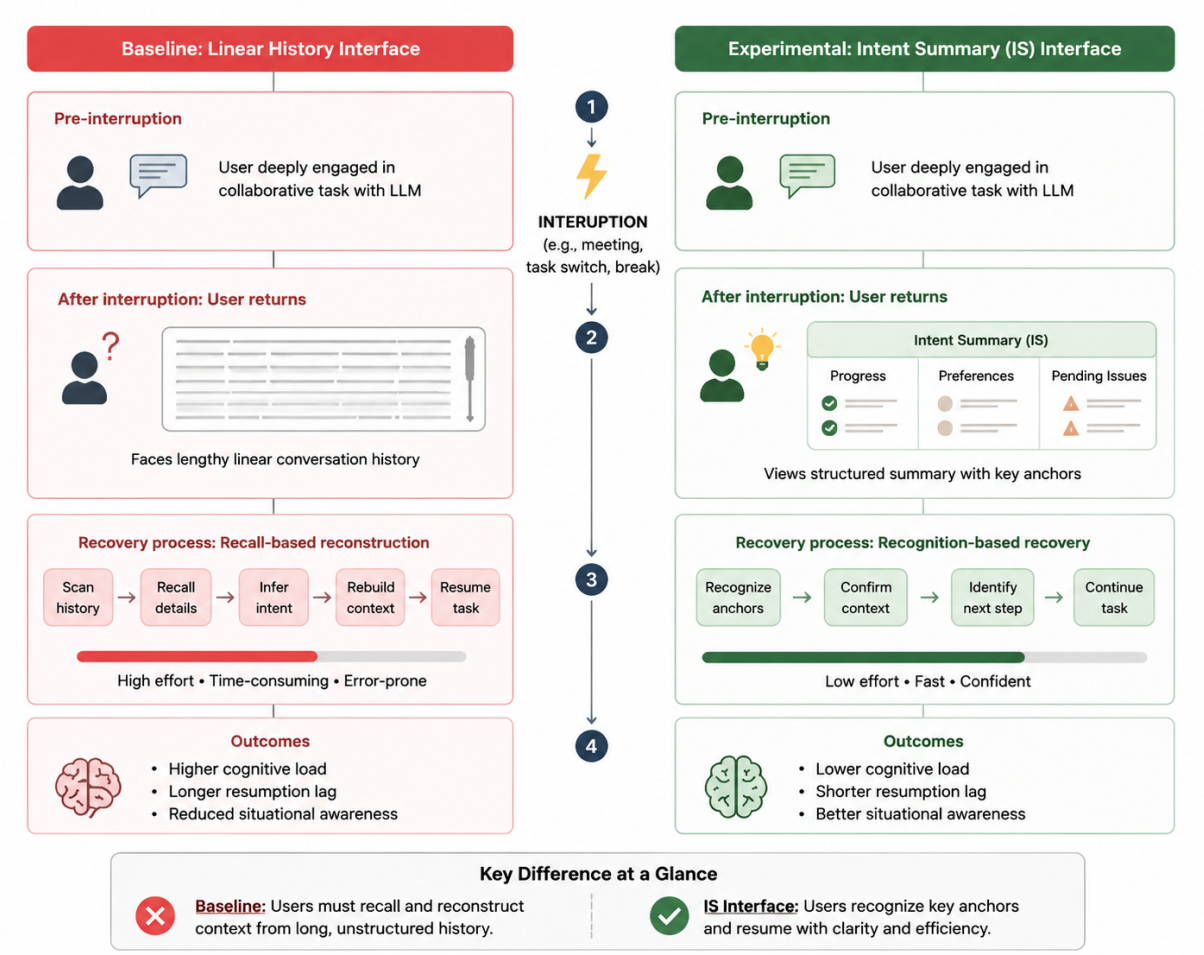

Figure 1: Comparison of technical continuity and cognitive continuity paths in interruption recovery. The figure illustrates how the proposed IS Interface (Right) bridges the memory gap and reduces resumption lag compared to the traditional linear interface (Left).

By separating conversational interaction from contextual abstraction, this architecture enables continuous task support while simultaneously maintaining cognitively efficient representations for interruption recovery.

3.3 User Interface Design

As shown in Figure 3, the user interface adopts a dual-column layout, consisting of a primary conversational workspace on the left and an IS sidebar on the right. This spatial separation decouples raw interaction data from structured contextual information, allowing users to access both detailed history and high-level summaries without interference.

The left panel occupies approximately 70% of the screen and presents a standard conversational interface, preserving the full interaction history and supporting ongoing collaboration with the system. While this area ensures technical continuity, after an interruption it primarily serves as a data repository rather than an efficient entry point for task resumption.

The right sidebar functions as a persistent visual aid, inspired by Trafton et al.'s concept of resumption readiness (Trafton et al., 2003). It is generated by the intent anchoring agent and follows the principle of information chunking, transforming extensive conversational content into structured cognitive anchors. The sidebar consists of three components:

Task Progression Recovery. Located at the top of the sidebar, this component provides a high-level summary of task progress and previously established decisions. By externalizing the current task state, it facilitates the reactivation of goal representations, which is critical for interruption recovery according to the Memory for Goals model (Altmann and Trafton, 2002). This allows users to quickly regain situational awareness without scanning the full conversation history.

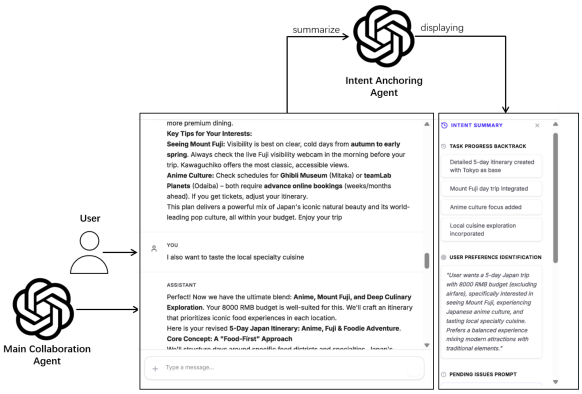

Figure 2: System architecture of the Intent Summary system. The framework decouples task execution and context management through two specialized agents: the Main Collaboration Agent and the Intent Anchoring Agent

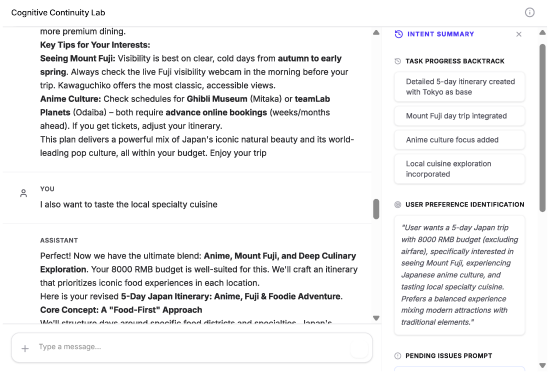

Figure 3: The Intent Summary Interface design. The UI features a dual-column layout with a dynamic collaboration window for primary tasks and a structured sidebar providing cognitive anchors (Progress, Preference, and Pending Issues).

User Preference Identification. The middle section captures both explicit and implicit constraints derived from prior interactions, including user preferences, habits, and decision criteria. By persistently displaying these elements, the system reduces the need to reconstruct decision rationale after interruptions, thereby supporting consistency in task execution. This aligns with prior findings that maintaining relevant contextual cues reduces cognitive load during recovery (Sweller, 1988; Ericsson and Kintsch, 1995).

Pending Decisions Prompt. The bottom component focuses on forward-looking information by highlighting unresolved issues and presenting actionable next steps. By explicitly surfacing pending decisions, the system supports a transition from retrospective context reconstruction to prospective task continuation. This design reduces resumption lag by guiding users directly into meaningful action, minimizing cognitive friction during re-entry.

4 User Study

4.1 Study Design

To evaluate the effectiveness of the IS interface in mitigating resumption costs after interruptions, we conducted a controlled between-subjects user study. Participants were randomly assigned to one of two conditions: (1) a baseline condition, in which users relied solely on conversation history for task recovery, and (2) an IS condition, in which users were supported by the structured IS panel upon resumption. The primary objective of this study was to compare users' efficiency, cognitive load, and subjective experience during task resumption under different recovery support mechanisms.

4.2 Participants

We recruited 30 university students (8 female; mean age = 25, range = 22–27), all of whom reported regular use of LLM-based tools such as ChatGPT. Participants received approximately $8 for completing the task.

4.3 Task and Procedure

Following the experimental instructions detailed in Appendix A.2, participants were required to collaboratively plan a 5-day trip to Japan with a budget of $1,200 using a conversational AI system, ultimately producing a complete travel itinerary.

Before the experiment, participants were given a standardized tutorial introducing the system interface and interaction workflow.

During the task, participants engaged in free-form dialogue with the AI to iteratively refine their travel plan. At a predefined point—when key decisions had been established—the system introduced a controlled interruption.

After the interruption, participants were asked to resume the task and continue decision-making. In the IS condition, a structured summary panel was presented upon resumption, while in the baseline condition, participants relied solely on the conversation history.

Participants were instructed that they could terminate the task at any time and were asked to indicate when they believed the task had been successfully completed. This design simulates an uncon-

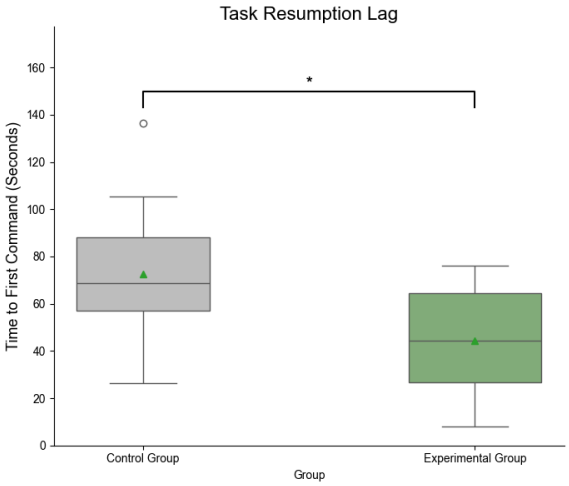

Figure 4: Statistical comparison of Task Resumption Lag between the Control and Experimental groups. The experimental group showed a significant reduction in the time required to send the first valid command after resumption.

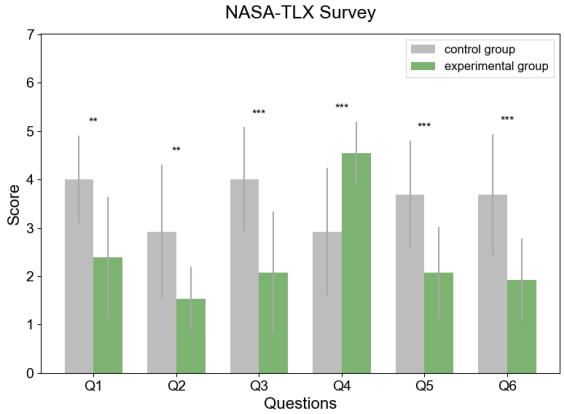

Figure 5: NASA-TLX sub-scale scores across six dimensions. The results indicate that the Intent Summary interface significantly lowered mental demand, effort, and frustration while enhancing perceived performance.

strained, real-world collaborative decision-making process and captures users' subjective confidence in task completion.

4.4 Measures

We collected four types of measures. Task efficiency was assessed using Resumption Lag, defined as the time between the end of the interruption and the participant's first meaningful input. Cognitive load was measured using NASA-TLX(Hart, 1986)(see Table 1). Subjective perception was assessed with three 5-point Likert items evaluating support for identifying task progress, recovering preferences, and determining next steps(see Table 2). We also collected open-ended responses to analyze participants' cognitive strategies during task resumption.

4.5 Implementation

The prototype system used a web-based front-end and back-end architecture to support conversational interaction and automated data collection. Intent Summary was generated from the pre-interruption dialogue and displayed upon resumption.

To ensure experimental control, interruptions were triggered only after participants had completed the four milestones detailed in Appendix A.1.3. This ensured that all participants were interrupted at comparable stages of task completion.

5 Findings

This study comprehensively evaluated the effectiveness of the IS interface in collaborative task interruption recovery through objective behavioral data, subjective cognitive load scales, and qualitative user feedback.

5.1 Task Recovery Efficiency Analysis

As shown in Figure 4, participants in the IS condition exhibited a significantly lower resumption lag (M = 44.25s, SD = 21.58) compared to the baseline (M = 72.66s, SD = 31.67; $t = 2.67$, $p = 0.014$). Beyond mean differences, the data distribution reveals the system's stability: while the baseline group showed a wider interquartile range (IQR) and extreme outliers exceeding 130s—reflecting severe cognitive breaks—the IS condition exhibited a compact distribution with no outliers. This suggests that the structured summary functioned as a robust "cognitive anchor," effectively eliminating the extreme delays typically caused by manually searching through long conversation histories.

5.2 Subjective Cognitive Load Assessment

Due to the severe loss of working memory caused by a 3-minute long interruption, this study employed the NASA-TLX scale to evaluate participants' mental workload during task recovery across six dimensions. Figure 5 shows a comparison of scores across these dimensions, and Welch's t-test results confirm that the experimental group outperformed the baseline group in all dimensions.

In terms of cognitive reconstruction and effort

Table 1: NASA-TLX Sub-scales and Task-Specific Descriptions.

No.	Survey Question	Dimension
1	How intense was the mental effort required to re-establish your train of thought after the interruption?	Mental Demand
2	Although this task is primarily cognitive, please assess the physical strain caused by interacting with the interface?	Physical Demand
3	After resuming the task, did you feel pressured to quickly get back up to speed within a short amount of time?	Temporal Demand
4	Do you feel you were able to continue the travel planning accurately and coherently from where you left off?	Performance
5	How much mental energy did you expend to re-immerse yourself in the "Japan travel planning" mindset?	Effort
6	When faced with the linear history log or summary, did you feel frustrated because you couldn't remember your previous decisions?	Frustration

Table 2: Task-Specific Evaluation Items for Intent Summary Feature.

No.	Survey Question
1	The "pending issues" in the summary clearly reminded me of what to do next.
2	The "user preferences" in the summary accurately reflected my travel style preferences before the interruption.
3	I was immediately able to clearly know exactly where we left off before the interruption.

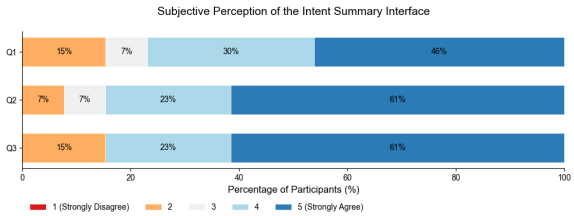

Figure 6: Subjective perception of the Intent Summary Interface. Distribution of 5-point Likert scale ratings for task progress localization, decision consistency, and anchor effectiveness.

expenditure, the experimental group had significantly lower scores on the mental demand and effort dimensions. The baseline group had a mean score of 4.00 when answering the question about "the intensity of thought required to regain focus," whereas the experimental group's mean was only 2.38 ($t = 3.742$, $p = 0.001$). Additionally, the experimental group exerted much less effort in "re-immersing into the mindset of planning a trip to Japan," with the mean dropping from 2.92 to 1.54 ($t = 3.259$, $p = 0.005$). This indicates that the IS successfully took on the context reconstruction work that would otherwise require the user to engage in intensive recall.

In terms of time pressure and performance perception, although the actual interruption time was the same, the experimental group perceived significantly less urgency. The baseline group scored as high as 4 on the "pressure to get back on track quickly," whereas the experimental group scored 2.08 ($p < 0.001$). Correspondingly, the experimental group showed strong confidence in whether they could "continue planning accurately and coherently," with scores rising from 2.92 to 4.54 ($t = -3.945$, $p = 0.001$). This shift from pressure to confidence validates the critical role of the interface design in maintaining the integrity of decision logic.

In the emotional experience and psychological stress dimensions, the experimental group showed the most significant advantage. Participants in the baseline group experienced noticeable frustration and stress when facing traditional linear history logs, as they could not recall previous decisions. Statistics showed that the experimental group's stress levels dropped from a mean of 3.69 to 2.08, and frustration decreased from 3.69 to 1.92, with both $p < 0.001$. This finding demonstrates that the IS, by eliminating memory gaps, provided users with the necessary psychological safety and control when handling complex collaborative tasks.

5.3 User Subjective Perception Analysis

We assessed the subjective acceptance of the IS interface by participants in the experimental group.

The data from the 5-point Likert scale in Figure 6 show that participants gave highly positive feedback about the interface. In all three evaluation items, more than 75% of users responded with "Agree" or "Strongly Agree." In particular, for the ease of re-entering the task (Q2) and confidence in continuing the task (Q3), the percentage of users who "Strongly Agree" reached 61%. These results

Table 3: Themes of participants' initial thoughts during context resumption.

Cognitive Theme	Representative User Quotes
Goal-Oriented Resumption	Optimize the next itinerary; find a hotel with convenient transportation; consider hot springs.
Logic Chain Continuity	I have already finished discussing... I should continue my discussion; where did the previous step go?
Instant Intent Realization	I want to continue the itinerary planning for authentic Japanese food and cherry blossom viewing; what am I going to Japan for?

indicate that the IS not only objectively enhanced efficiency but also provided strong confidence support subjectively.

5.4 Qualitative Findings on Cognitive Recovery Pathways

To reveal the cognitive mechanisms behind the efficiency gains, we analyzed participants' open-ended feedback. As summarized in Table 3, the IS interface fundamentally shifted users' recovery pathways. Rather than engaging in the high-effort process of retrospective reconstruction (scrolling through logs to remember "where I was"), participants in the IS group immediately identified the current task state. Dominant themes include directly extracting to-dos (e.g., "optimize the next itinerary") and recovering logical breakpoints. This shift from memory retrieval to goal identification explains how the IS allows users to bypass linear log-scanning and achieve more coherent decision-making.

6 Discussion and Design Implications

Our findings suggest that Intent Summary (IS) shifts task resumption from effortful recall to recognition-based re-engagement. In conventional conversational interfaces, users must scan lengthy dialogue histories to reconstruct prior decisions and task context. In contrast, IS provides structured semantic cues that directly reactivate task goals, reducing the need for sequential search and supporting faster recovery.

This mechanism explains the significant reduction in Resumption Lag and subjective workload observed in our study. By externalizing task progress, user preferences, and pending decisions, IS serves as a compact cognitive scaffold that helps users quickly regain situational awareness and continue their work with greater confidence.

Our results also suggest that cognitive support should be carefully calibrated. Overly detailed summaries may introduce additional attentional demands. Effective interruption support should therefore emphasize concise, actionable information and be presented at meaningful breakpoints, such as immediately after an interruption.

Based on these findings, we derive three design implications for interruption-resilient conversational systems:

Design for recognition, not recall. Interfaces should provide structured cues that enable immediate identification of task state and next actions;Use event-triggered summarization. Cognitive support should be presented when users are most likely to experience re-entry difficulty, rather than displayed continuously;Represent intent explicitly. Summaries should foreground goals, constraints, and unresolved decisions instead of merely compressing prior dialogue.

7 Conclusion

This work addresses the challenge of situational disorientation following interruptions in large language model–mediated collaboration. We propose the Intent Summary (IS) as a cognitive intervention that transforms lengthy conversational histories into compact semantic anchors, effectively supporting users in reconstructing task context.

Our results demonstrate that IS not only improves task resumption efficiency by nearly 40%, but also fundamentally alters the cognitive pathway of recovery—from effortful recall to rapid recognition. By reducing cognitive load and enabling immediate re-orientation, IS facilitates more seamless continuity in human–AI interaction.

Beyond its empirical contributions, this work highlights a broader shift in interface design: from passive information presentation toward active cognitive support. We argue that future intelligent systems should be designed as cognitive partners that assist users in managing interruptions, maintaining intent, and navigating complex, temporally distributed tasks.

Finally, future work should explore tighter integration between summarization and task modeling. In particular, combining IS with automated subtask decomposition and progress tracking may further enhance users' ability to maintain cognitive continuity across extended interactions.

8 Limitations

This study has several limitations that open avenues for future research.

Initially, the participant sample was relatively homogeneous, consisting primarily of university students with high levels of technical literacy. Prior research suggests that users' mental models and trust in intelligent systems vary significantly across demographic groups (Gnewuch et al., 2023). Future work should include more diverse populations to assess the generalizability of our findings.

Furthermore, our experimental task focused on a structured travel planning scenario. While this allows for controlled evaluation, it does not fully capture the complexity of open-ended or creative tasks, where user intentions are fluid and less explicitly defined. In such contexts, the effectiveness of IS may depend on its ability to accommodate evolving goals and ambiguous task boundaries.

Ultimately, the reliability of automatically generated summaries remains a critical concern. Although large language models are effective at abstraction, they are also prone to hallucination (OpenAI, 2023). Erroneous summaries may mislead users and disrupt decision-making processes. Future systems should therefore incorporate mechanisms for verification, user correction, and transparency, enabling users to calibrate their trust in the system.

Beyond technical limitations, IS may introduce risks related to over-reliance on automated summaries. Incorrect or hallucinated summaries could distort users' understanding of prior decisions and bias subsequent actions. In addition, continuously inferred preference representations may raise privacy and transparency concerns in long-term human–AI collaboration. Future systems should therefore support user verification, editable summaries, and controllable memory persistence to ensure trustworthy cognitive support.

References

Piotr D. Adamczyk and Brian P. Bailey. 2004. If not now, when? In *CHI*, pages 271–278.

Erik M. Altmann and J. Gregory Trafton. 2002. Memory for goals: An activation-based model. *Cognitive Science*, 26(1):39–83.

Brian P. Bailey and Joseph A. Konstan. 2006. On the need for attention-aware systems. *Computers in Human Behavior*, 22(4):685–708.

Jelmer P. Borst, Niels A. Taatgen, and Hedderik van Rijn. 2015. What makes interruptions disruptive? a process-model account of the problem state bottleneck. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2971–2980.

Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A diary study of task switching and interruptions. In *CHI*, pages 175–182.

Mica R. Endsley. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64.

K. Anders Ericsson and Walter Kintsch. 1995. Long-term working memory. *Psychological Review*, 102(2):211–245.

Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2023. The conversational user interface: A new frontier for hci. *Business & Information Systems Engineering*, 65(2):123–128.

Iryna Gurevych, Christoph Müller, and Michael Strube. 2003. The impact of summarization on task performance. In *SIGIR*, pages 379–380.

Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness. *Computer Supported Cooperative Work*, 11(3):411–446.

Chris Harrison, Brian Amento, Sergiy Kuznetsov, and Randy Bell. 2009. Rethinking the progress bar. In *UIST*, pages 109–118.

Sandra G. Hart. 1986. Nasa task load index (tlx).

Yujin Jeong, Joonhwan Lee, and Hyeoncheol Kim. 2025. Cpg: Enhancing self-efficacy in conversational ai. In *CHI*.

Aman Madaan, Niket Tandon, Prakhar Gupta, Abhishek Vashistha, Peter Clark, and Yejin Choi. 2023. Self-refine: Iterative refinement with self-feedback. In *ACL*, pages 4650–4668.

George A. Miller. 1956. The magical number seven, plus or minus two. *Psychological Review*, 63(2):81–97.

Brad A. Myers. 1985. The importance of percent-done progress indicators. In *Proceedings of the SIGCHI Conference*, pages 11–17.

OpenAI. 2023. Gpt-4 technical report. Technical report. ArXiv:2303.08774.

Noah Shinn, Federico Cassano, Edward Smith, Jonathan Mitchell, and Shiry Ginosar. 2023. Reflexion: Language agents with verbal reinforcement learning. In *NeurIPS*.

John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285.

J. Gregory Trafton, Erik M. Altmann, Derek P. Brock, and Farilee E. Mintz. 2003. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58(5):583–603.

Yifan Zhang, Xinyu Chen, Bohan Jin, Shuyuan Wang, Shouling Ji, Wei Wang, and Jiawei Han. 2024. A comprehensive survey of scientific large language models. In *EMNLP*, pages 8783–8817.

A Appendix

A.1 Detailed Instruction Prompts

A.1.1 System Prompt

You are a helpful travel assistant. When users ask for travel suggestions, try to provide detailed responses that:

1. *Introduce at least 2 different destinations.*
2. *Compare the characteristics of these destinations.*
3. *Provide a simple day-by-day itinerary or suggestion.*
4. *Mention estimated accommodation budgets or price ranges.*

A.1.2 User Prompt for Intent Summary Generation

Based on the following conversation history, generate a structured "Intent Summary" to help the user recover after an interruption.

Please return the summary in JSON format with the following structure:

1. *"progress": ["Completed point 1", "Completed point 2"].*
2. *"preferences": "Summary of user preferences and concerns".*
3. *"pendingIssues": "Core conflicts or unresolved issues".*
4. *"suggestedNextSteps": ["Suggested step 1", "Suggested step 2"].*

A.1.3 User Prompt for Interruption Trigger Detection

Analyze the following conversation and determine if the AI has provided a sufficiently detailed travel proposal to justify an interruption for a research study.

The proposal is considered "complete" if it has covered:

1. *Multiple destinations (at least 2).*
2. *Comparison of their features.*
3. *A rough itinerary or day-by-day plan.*
4. *Budget or price estimates for accommodation.*

A.2 User Study Protocol and Task Instructions

In this section, we present the exact experimental instructions and the task configuration provided to the participants in our controlled user study. To ensure ecological validity, the interface and instructions were originally delivered in Chinese (the native language of the participants) and have been translated here for reporting purposes.

A.2.1 Task Briefing and Ethical Considerations

Before commencing the experiment, participants were presented with the following onboarding brief, which includes statements on data privacy and compliance with ethical research standards:

This experiment investigates human-AI collaboration paradigms and takes approximately 10 minutes to complete. You will work alongside an AI assistant to plan a 5-day trip to Japan with a total budget of $1,200 (excluding airfare), culminating in a complete travel itinerary. During the process, a brief interruption task will be introduced, and you will be asked to complete a questionnaire at the end of the session.

The system automatically logs dialogue tokens and interaction telemetry for research purposes only. All data will be strictly anonymized to protect your privacy. Please interact as naturally as you would in your daily AI usage; there are no right or wrong answers.

A.2.2 Collaborative Travel Planning Task Configuration

To maintain sufficient cognitive complexity and prevent superficial interactions, participants were instructed to internalize a simulated scenario where they were planning an imminent personal vacation. The structural constraints and specific delivery objectives of the task were presented to users through the following text guidelines:

Core Travel Constraints The collaborative session required users to plan a solo vacation adhering strictly to the following parameters:

- **Destination:** Japan (with specific regional locations to be negotiated with the AI).
- **Duration:** Exactly 5 days.
- **Budget Limit:** $1,200 / approximately 8,000 RMB (excluding international airfare).
- **Accompaniment:** Solo travel.

Collaborative Objectives Through continuous multi-turn dialogue with the AI assistant, users were required to reach a mutual consensus and explicitly finalize:

1. The specific target city or cross-city route.
2. A granular day-by-day itinerary, distinguishing between morning and afternoon activities (e.g., matching personalized interests such as Mt. Fuji exploration or anime culture tours).
3. Structured accommodation arrangements, detailing the preferred lodging type, geographical neighborhood, and specific hotel names.
4. A clear financial breakdown of budget allocation across activities, food, and lodging.

A.2.3 Interruption and Survey Procedure

To rigorously simulate real-world task suspension and measure cognitive resumption lag, the workflow incorporated a forced asynchronous interruption mechanism defined by the following procedural prompt:

> The system will trigger a mandatory interaction suspension when **key decisions had been established**. The interruption phase will last for precisely **3 minutes**. During this suspension, you are required to perform a distracting cognitive task by watching an external video via the provided link[1] (the video does not need to be watched in full; your re-engagement is governed strictly by the on-screen countdown timer).
>
> Upon the expiration of the countdown, you must **immediately** return to the chat interface to resume your prior planning task. For participants allocated to the treatment group, please actively utilize the **Intent Summary** panel anchored on the right side of the interface to facilitate context reconstruction. Immediately after sending your first post-interruption message, you will be redirected to complete the post-task questionnaire[2] to evaluate your perceived cognitive workload.

[1] https://www.bilibili.com/video/BV1FafbBcEYC/

[2] https://v.wjx.cn/vm/rXWA6a8.aspx