Information Research - Vol. 30 No. iConf (2025)

Exploring the role of large language model in collaborative travel planning task

Hikaru Kumamoto and Hideo Joho

DOI: https://doi.org/10.47989/ir30iConf46966

Abstract

Introduction. This study explores how generative AI, specifically ChatGPT, influences collaborative travel planning. Understanding its effect on tasks like trip planning reveals insights into human-AI collaboration, particularly how AI tools support decision-making and streamline information gathering.

Method. Twenty participants (10 pairs) planned a 1-night, 2-day trip under two conditions: (1) using only Google, and (2) using Google with ChatGPT. This within-subject study measured completion time, satisfaction, and plan quality via questionnaires and observation to capture task performance, user behaviour, and collaboration dynamics.

Analysis. Data was analysed using t-tests and Wilcoxon signed-rank tests to compare completion times, satisfaction, and plan quality. Analyses of conversation volume and ChatGPT logs provided insights into AI-assisted collaboration dynamics and interaction patterns.

Results. No significant difference in task completion time was found. However, plans made with ChatGPT were more complete and aligned with requirements. Participants found information more easily with ChatGPT, but satisfaction levels remained similar, suggesting that easier information access did not translate to higher satisfaction.

Conclusions. Generative AI improves collaborative search task quality but does not enhance efficiency or satisfaction. AI tools like ChatGPT are effective for providing structured information and are best used as complementary resources alongside traditional search engines in planning tasks.

Introduction

Navigating the sheer volume of available data has been both a challenge and an opportunity. The ability to collaborate effectively on information-seeking tasks has become essential, particularly for complex undertakings like travel planning, where diverse sources of information must be integrated to accommodate various preferences and constraints (Shar, 2014). Collaborative decision-making presents unique challenges not encountered in individual tasks, such as time inefficiencies due to group deliberation, coordination complexities, and the influence of dominant group members, which can skew outcomes (Kraemer & King, 1988). Understanding these dynamics is crucial for designing effective tools for collaborative settings. The emergence of generative AI models, especially Large Language Models like OpenAI's ChatGPT, is transforming this landscape. These systems provide information, synthesise ideas, and respond conversationally to users' prompts based on patterns identified in extensive pre-existing human-generated data. By addressing the unique challenges of collaborative tasks, this study investigates how generative AI might enhance complex processes like trip planning, where collective decision-making is central.

Generative AI tools like ChatGPT offer unique potential in collaborative contexts. Unlike search engines, which require users to manually sift through and interpret results, ChatGPT engages interactively, providing suggestions, organizing information, and helping users think through problems. This shift from passive search to active collaboration raises possibilities for improving task efficiency, reducing users' cognitive burden, and enhancing the quality of outcomes (Gienapp, 2024). Despite this potential, the impact of AI on collaborative information-seeking, especially in practical settings like travel planning, remains underexplored (Skjuve, et al., 2023). Travel planning, with its multiple stages—from destination selection to booking accommodations—offers an ideal context to examine how AI might influence collaborative processes and outcomes (Imazu, et al., 2011; Morris, 2013).

To explore these dynamics, this study poses the following key research questions:

- How does the use of ChatGPT, impact the efficiency, user satisfaction, and quality of travel plans in collaborative information-seeking tasks compared to traditional search engine use?
- What influence does ChatGPT have on communication dynamics and decision-making during collaborative tasks, and what challenges arise when integrating AI into real-time, dynamic information needs like travel planning?

To address these questions, the study examines the role of ChatGPT in real-world collaborative travel planning tasks. Specifically, it investigates the effect of incorporating ChatGPT into the planning process, comparing it with the traditional use of search engines. While prior research has explored AI's effectiveness in individual information-seeking tasks, little attention has been given to its role in collaborative efforts, where multiple users must coordinate to reach decisions. Travel planning, an everyday activity, reflects the complexity of many collaborative tasks, involving negotiation, compromise, and logistical decision-making. Understanding how AI supports such processes has broader implications for the future of collaboration across various fields.

Methodology

The methodology of this research involved a within-subjects experimental design, allowing participants to complete a travel planning task under two conditions: (1) using only a search engine (condition S), and (2) using both a search engine and ChatGPT (condition C). The purpose was to evaluate task efficiency, satisfaction, and the quality of travel plans produced. The order of conditions was counterbalanced to control for learning effects.

We used a free edition of ChatGPT version 3.5 (OpenAI, 2022), the latest available during the study in August 2023.

Participants

Ten pairs of participants (20 individuals) were recruited for this study. Participants were recruited in pairs and chose partners they felt comfortable collaborating with. This self-selection ensured they did not have difficulty working together, minimizing potential influence of interpersonal dynamics on the results. The participants consisted of undergraduate and postgraduate students from the University of Tsukuba, with ages ranging from 19 to 27 years (mean age = 21.3 years). The sample included 6 women and 14 men, with academic backgrounds spanning informatics, science and engineering, social sciences, humanities, and medicine. Participants were daily users of traditional search engines, ensuring basic familiarity with online search methods such as finding relevant information and comparing sources. Advanced search techniques (e.g., Boolean operators or relevance ranking) were not specifically assessed, as the study aimed to reflect typical user behaviour. Before the experiment, participants were briefed on ChatGPT's interface, capabilities, and general use cases to standardise their understanding of the tool.

Task

The task asked participants to collaboratively plan a two-day, one-night domestic trip. The planning process was divided into four key steps:

- Destination selection: participants chose three tourist destinations they would like to visit.
- Schedule creation: participants organised the trip itinerary, including the sequence of visits to destinations.
- Accommodation booking: participants selected and booked a place to stay.
- Transport booking: participants arranged transportation for the trip, including modes of travel and approximate timings.

Participants were asked to start with the destination selection but could order the remaining steps as they wished. They could not return to a previous procedure, as this would make time recording unreliable.

In condition S (search engine only), participants used Google to gather information and complete the task without AI assistance. In condition C (search engine + ChatGPT), participants used both ChatGPT and Google, with the frequency of ChatGPT use left to their discretion. However, the study did not record the precise duration of ChatGPT use in place of the search engine. Future studies could measure this to better understand time allocation across tools and its impact on performance. For consistency, participants were not allowed to use other AI-based tools or services.

Time	Place	Activity	Transportation	Duration	Price
7:00	City A to City B		Local Train	80 min	840
...					
11:10	Market C	Shopping, lunch		60 min	
...					
17:30	Hotel D	Check-in, break		60 min	3776
...					

Table 1. Template of travel planning task

Procedure

Each experiment session lasted approximately 170 minutes, divided into two phases (condition S and condition C). Participants were given 80 minutes per condition to complete the four steps of travel planning. After each condition, they completed a post-task questionnaire assessing satisfaction with the process and the travel plan produced.

During the experiment, one participant in each pair operated the computer, while the other recorded the travel plan (See Table 1). Participants could switch roles as needed. Importantly, all decisions regarding the travel plan had to be made collaboratively, ensuring the study focused on group dynamics rather than individual decision-making.

Data collection

Data were collected on three key variables:

- Task completion time: the total time participants took to complete the travel planning task was recorded for each condition. Additionally, the time spent on each of the four steps was tracked.
- Satisfaction: after each condition, participants completed a questionnaire using a 5-point Likert scale, rating satisfaction with the planning process, ease of use, and the quality of the travel plan. Questions addressed both overall satisfaction and satisfaction with specific stages of the planning process.
- Travel plan quality: the travel plans were evaluated based on criteria such as completeness (whether the plan covered all essential aspects of the trip), feasibility (whether the schedule and budget were realistic), and attention to detail (whether transportation, accommodation, and destination information were sufficiently detailed).

Additionally, the volume of conversation during the task was recorded and analysed to assess ChatGPT's impact on communication. Conversations were transcribed and measured in terms of the number of words exchanged, both between participants and with ChatGPT.

Ethical considerations

This study was approved by the Ethics Review Committee of Institute of library, information and media science, University of Tsukuba (Approval No. 23-73). All participants provided informed consent before participating in the experiment and were compensated with Amazon gift certificates based on the institution's regulation.

Results

The results of the study are presented in three key areas: task completion time, user satisfaction, and the quality of the travel plans produced. Additionally, we analysed the volume of conversation between participants and ChatGPT's usage in condition C (search engine + ChatGPT). Each of these aspects is compared between the two conditions—condition S (search engine only) and condition C.

Task completion time

Overall task completion time includes the time taken to plan the entire two-day trip. The median task completion time for condition C was 3,553 seconds (59:13 minutes), while condition S had a median time of 3,471 seconds (57:51 minutes). The Wilcoxon rank-sum test indicated that this difference was not statistically significant ($p = 0.50$), suggesting that including ChatGPT did not reduce overall task completion time.

However, when broken down by task procedure, a few notable differences emerged. In the destination selection stage (procedure A), participants in condition C took slightly less time than

those in condition S. The median time for procedure A in condition C was 21.0% of the total time, while in condition S it accounted for 29.9%. This indicates that ChatGPT may have helped participants generate ideas for destinations more quickly. Conversely, the schedule creation stage (procedure B) took slightly longer in condition C (33.1%) compared to condition S (26.0%), though these differences were not statistically significant ($p > 0.05$).

User satisfaction

Participants' satisfaction with the travel planning process was measured using a post-task questionnaire, which included questions on ease of use, satisfaction with the plan produced, and satisfaction with each stage of the planning process. Overall, there were no significant differences in total satisfaction between the two conditions. Participants generally expressed moderate satisfaction in both conditions, with a median satisfaction score of 4 (on a 5-point Likert scale).

However, analysis of individual stages revealed nuanced differences. In the destination selection stage, participants in condition C reported finding it easier to identify relevant information compared to condition S (median satisfaction score of 1.00 vs. 2.00). In contrast, in the accommodation booking stage (procedure C), participants in condition S expressed greater satisfaction with their ability to find accurate and up-to-date information. This difference likely reflects ChatGPT's reliance on static data, limiting its effectiveness for tasks requiring real-time information, such as finding available accommodations.

Despite the overall similarity in satisfaction between the two conditions, open-ended responses from participants highlighted ChatGPT's usefulness for idea generation. One participant remarked, '*ChatGPT helped us brainstorm destinations quickly when we had no initial ideas*,' while another stated, '*the search engine required more effort to manually gather all the information.*'

Quality of travel plans

The quality of the travel plans produced was assessed based on completeness, feasibility, and level of detail. Plans were considered complete if they covered all essential aspects of the trip, including transportation, accommodation, and activities. Feasibility referred to whether the plans were realistic in terms of time and budget, and level of detail was evaluated based on how thoroughly the plans addressed logistical aspects, such as transport schedules and booking details (Imazu, et al., 2011).

Results showed that condition C (search engine + ChatGPT) led to travel plans that were more aligned with the task requirements. Nine of the ten pairs in condition C produced plans that met all task criteria (e.g., stayed within budget, covered all trip details), compared to seven pairs in condition S. Additionally, plans from condition C were often rated as more detailed, especially in terms of destination choices and activities. However, condition S produced more feasible schedules, as participants found it easier to obtain accurate transport and accommodation information using search engines.

For example, one travel plan in condition C proposed a tight schedule including three distant locations in a single day, which may have been unrealistic. In contrast, travel plans in condition S generally reflected more conservative itineraries, with better alignment between travel times and planned activities.

Conversation volume and ChatGPT usage

The volume of conversation between participants was measured to assess how ChatGPT influenced group interaction. In condition C, participants exchanged fewer words compared to condition S, suggesting that ChatGPT may have reduced the need for extended discussions (Joho, et al., 2008). The median number of words exchanged in condition C was approximately 15% lower than in condition S. A Wilcoxon rank-sum test found no statistically significant difference between

the conditions ($p = 0.34$), but the trend suggests the AI tool streamlined some aspects of the decision-making process.

Participants used ChatGPT primarily in the destination selection (procedure A) and schedule creation (Procedure B) stages. The most common type of interaction with ChatGPT was asking for ideas and itinerary suggestions. For instance, participants frequently asked ChatGPT for recommended places to visit in a specific region or to help build a draft schedule. ChatGPT was used less frequently in accommodation and transport booking stages, likely due to the need for real-time data, which ChatGPT could not provide, at the time of study.

Conclusive discussion

This study investigated the impact of generative AI, specifically ChatGPT, on collaborative travel planning tasks by comparing it to traditional search engine use. Collaborative decision-making differs from individual decision-making due to challenges such as reconciling diverse preferences, negotiating compromises, and managing group dynamics, including the influence of dominant members (Kraemer & King, 1988; Shah, 2014). These attributes make collaborative tasks particularly complex and critical for evaluating the potential of AI tools. While no statistically significant differences were found in overall task completion time or satisfaction, the results provide nuanced insights into how AI tools influence collaborative information-seeking behaviour and decision-making. These findings highlight both the potential and limitations of AI integration into real-world tasks, offering several important takeaways for future research and practical applications.

Key findings

The most notable benefit of using ChatGPT in the collaborative planning process was in the destination selection stage, where it helped participants generate ideas quickly. In this phase, participants using ChatGPT took less time to decide on destinations compared to those relying solely on search engines. This suggests that generative AI can serve as an effective tool for brainstorming, particularly in tasks requiring creativity or decision-making when users lack initial direction. Participants noted that ChatGPT helped overcome the initial inertia of decision-making by offering suggestions and guiding the conversation. Participant satisfaction ratings did not indicate concerns about ChatGPT's suggestions being overly leading. This aligns with research suggesting that AI's value lies in its ability to generate ideas and reduce cognitive load in early problem-solving stages (Maaravi, et al., 2021).

However, ChatGPT's advantages did not extend uniformly across all planning stages. In the accommodation and transport booking stages, participants in condition S (search engine only) expressed greater satisfaction and produced more feasible plans. This can be attributed to ChatGPT's lack of real-time data, which limited its ability to provide accurate, up-to-date information on hotel availability, transportation schedules, and other logistics-dependent tasks. In these cases, search engines offered more reliable information, underscoring the continued importance of traditional tools for tasks heavily reliant on real-time data. ChatGPT's structured, system-driven outputs may also not align with the adaptability and creativity humans bring to dynamic contexts.

Despite these insights, several limitations must be acknowledged. First, the sample size was relatively small, with only 10 pairs of participants. This limited the statistical power of the results, potentially explaining the lack of significant differences in task completion times and satisfaction across conditions. A larger sample size could provide more robust data and reveal subtler trends in performance and behaviour.

Additionally, participants were primarily students with limited experience using ChatGPT for travel planning. This familiarity bias may have limited their ability to fully leverage the tool's capabilities, particularly in the task's initial stages. Future studies should include a longer

familiarisation period or recruit participants with greater experience using AI tools, allowing a more accurate assessment of how generative AI can be used effectively in collaborative settings.

Another limitation was the task's time constraint, which many participants found restrictive. The one-hour limit may have prevented participants from fully exploring travel planning options, especially in condition C, where ChatGPT unfamiliarity slowed some processes. Extending task duration in future studies would provide a more realistic reflection of how AI can be integrated into collaborative work, particularly for more complex or fluid tasks like travel planning.

Implications for the iSchool community

For researchers and practitioners in the iSchool community, this study provides valuable insights into the evolving role of AI in information-seeking behaviours. AI tools like ChatGPT can complement, but not replace, traditional search engines in collaborative tasks. Combining AI for idea generation with search engines for real-time data retrieval is a promising approach, particularly in complex tasks like travel planning that require both creativity and logistical accuracy.

The findings also suggest that AI's impact on group communication and dynamics warrants further investigation. While AI may reduce cognitive load and streamline decision-making, it risks diminishing the collaborative aspect of tasks, potentially leading to less robust decision-making. Future work should explore how to balance AI assistance with human interaction to preserve the benefits of collaboration while leveraging AI's strengths.

Conclusion

In conclusion, this study found that while ChatGPT offers distinct advantages in collaborative travel planning, particularly in the ideation phase, it does not enhance all aspects of the task universally. Search engines remain critical for tasks requiring real-time data and detailed logistics. Additionally, while AI may reduce cognitive load and streamline decision-making processes, it changes the nature of collaboration by reducing the need for extended discussions. These findings suggest that generative AI tools are best used as complementary resources, handling early-stage ideation and exploration, while traditional tools support detailed decision-making.

Future research should address the study's limitations by expanding the sample size and including participants with more diverse backgrounds and experience levels to improve the findings' generalisability. Investigating the use of generative AI across a broader range of collaborative tasks—such as project management or academic research—would offer further insights into its applications. To address limitations related to real-time data, future studies could explore integrating AI with live databases or APIs to enhance decision-making accuracy. Lastly, extending task durations in future experiments could allow participants to better engage with AI tools, providing a deeper understanding of leveraging these technologies for long-term, complex collaborative tasks.

Acknowledgements

This work was supported in part by JSPS KAKENHI grant number JP24K03048, Japan. The authors thank the study participants for their valuable contributions and the anonymous reviewers for their constructive feedback, which greatly improved this work.

About the authors

Hikaru Kumamoto was a fourth-year undergraduate student at the College of Knowledge and Library Sciences, School of Informatics, University of Tsukuba, at the time of this work. She received the dean of the School of Informatics award in recognition of her outstanding academic performance.

Hideo Joho is a Professor of Information Retrieval in the Institute of Library, Information and Media Science, University of Tsukuba. His research focuses on understanding human information interactions, building search user interfaces, and developing test collections for evaluation. He can be reached at hideo@slis.tsukuba.ac.jp.

References

Shah, C. (2014). Collaborative information seeking. Journal of the Association for Information Science and Technology, 65(2), 215-236.

Kraemer, K.L. & King, J.L. (1988). Computer-based systems for cooperative work and group decision making. ACM Computing Surveys, 20 (2), 115–146.

OpenAI (2022). Introducing ChatGPT. https://openai.com/index/chatgpt/

Morris, M. R. (2013). Collaborative search revisited. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 1181-1192).

Skjuve, M., Følstad, A., & Brandtzaeg, P. B. (2023). The user experience of ChatGPT: findings from a questionnaire study of early users. In *Proceedings of the 5th international conference on conversational user interfaces* (pp. 1-10).

Gienapp, L., Scells, H., Deckers, N., Bevendorff, J., Wang, S., Kiesel, J., ... & Potthast, M. (2024). Evaluating generative ad hoc information retrieval. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1916-1929).

Imazu, M., Nakayama, S., & Joho, H. (2011). Effect of explicit roles on collaborative search in travel planning task. In *Information Retrieval Technology: 7th Asia Information Retrieval Societies Conference, AIRS 2011, Dubai, United Arab Emirates, December 18-20, 2011. Proceedings 7* (pp. 205-214). Springer Berlin Heidelberg.

Joho, H., Hannah, D., & Jose, J. M. (2008). Comparing collaborative and independent search in a recall-oriented task. In *Proceedings of the second international symposium on Information interaction in context* (pp. 89-96).

Maaravi, Y., Heller, B., Shoham, Y., Mohar, S., & Deutsch, B. (2021). Ideation in the digital age: literature review and integrative model for electronic brainstorming. *Review of Managerial Science*, 15, 1431-1464.