Travel Planning with Large Language Models: A Review and Outlook

Anonymous ACL submission

Abstract

Planning is a critical step in advancing artificial intelligence (AI) systems toward higher levels of intelligence and is one of the core capabilities of autonomous decision-making systems, involving complex processes of understanding, reasoning, and decision-making. Current research on planning with AI mostly focuses on simulated environments. Although significant progress has been made, its application in the real world remains limited due to the unpredictable and complex nature of real-world scenarios. Travel planning, as a practical task, is a prime example of these challenges, involving the coordination of factors such as destination selection, budget constraints, and personalized preferences, while also requiring adaptation to changes in external conditions. This review, based on the key roles of LLMs in travel planning tasks, presents a taxonomy of existing methodologies, categorizing them into three types: planner, reformulator, and knowledge source. Furthermore, it outlines directions for future research. We hope this review will provide valuable background information and guidance for researchers in the field, driving the development of this emerging topic.

Keywords: Large Language Models, Travel Planning, Tourist Trip Design Problem, Natural Language Processing, Agent

1 Introduction

Planning is a critical step in advancing AI systems toward higher levels of intelligence and a core capability of autonomous decision-making systems (Huang et al., 2024a), encompassing complex processes of understanding, reasoning, and decision-making (Long, 2005). In recent years, the development of LLMs has driven a paradigm shift in the AI field (Zhao et al., 2024). These models demonstrate exceptional intelligence in reasoning, tool use, and planning, offering new possibilities for enhancing the planning capabilities of autonomous agents (Dagan et al., 2023). With continuous breakthroughs in LLMs capabilities, researchers have proposed various methods to integrate these models into planning modules, such as task decomposition, plan selection, external modules, reflection, and memory, boosting AI planning to higher levels (Huang et al., 2024b).

However, most current AI planning research remains focused on simulated environments, such as ALFWorld (Shridhar et al., 2020), ScienceWorld (Wang et al., 2022). While these studies have achieved significant progress, they face considerable challenges in real-world applications. The complexity and unpredictability of real-world scenarios far exceed the scope of simulated environments, limiting the broader application of these studies. Consequently, applying planning technologies to real-world tasks, particularly complex scenarios like travel planning, holds significant research value.

Meanwhile, tourism, as a vital component of the global economy, contributed 9.1% to global GDP in 2023, driving economic development through job creation and business opportunities (Herzog et al., 2019; Analytica, 2024). To enhance the travel experience, tourists usually need to plan under multiple constraints, including budget, time, transportation, accommodation, restaurant, and the attractiveness of the destinations (Zheng and Liao, 2019; Rodríguez et al., 2012). However, the overwhelming amount of travel information has led to information overload, making manual travel planning extremely challenging. Users often struggle to identify the best solutions that meet their needs, which requires AI technologies to optimize this process.

Traditional travel planning methods are based on fixed templates and perform poorly when handling unstructured natural language queries (Bhowmick et al., 2012; Zhu et al., 2012). Extracting key information and converting it into structured data is a cumbersome process, and these methods often provide generic solutions that fail to account for users' personalized preferences (de la Rosa et al., 2024). Moreover, traditional planning systems rely on static databases, which limits their ability to update information in real time and respond to user needs (Hsueh and Huang, 2019). They also lack the capacity to handle dynamic changes and complex constraints (Bubeck et al., 2023; Silver et al., 2017).

The emergence of LLMs offers innovative solutions to these challenges. First, LLMs can understand and process queries in natural language, allowing users to describe their needs and constraints directly in natural language (Sumers et al., 2023), greatly simplifying the interaction between user and system. By combining LLMs with traditional constraint-solving techniques, these systems retain the flexibility of natural language processing (NLP) while ensuring the rigor of constraint solvers, thereby delivering end-to-end travel planning solutions (Hao et al., 2024; de la Rosa et al., 2024). Furthermore, LLMs can dynamically retrieve the latest external information, user feedback, and evolving requirements, adjusting plans in real time to meet personalized needs (Ma et al., 2024). With their extensive knowledge base and robust planning capabilities, LLMs can also address complex constraint problems, providing users with more precise and flexible travel planning services (Xie et al., 2024; Miin and Wei, 2024).

Although some research has applied LLMs to travel planning (Xie et al., 2024; Hao et al., 2024; de la Rosa et al., 2024; Zheng et al., 2024; Ma et al., 2024; Miin and Wei, 2024), there has yet to be a systematic review of travel planning solutions in the era of LLMs. Therefore, this paper summarizes the main application scenarios, available datasets, evaluation methods for LLM-powered travel planning, and point out future directions. In Section 2, we analyze different scenarios of travel planning. Then, Section 3 reviews available datasets and evaluation methods. Section 4 provides a detailed overview of the application of LLMs in travel planning. We highlight the opportunities in the era of LLMs in Section 5 and conclude this review in Section 6. The review of traditional travel planning is provided in the Appendix A.

In summary, the main contributions of this review are as follow:

1. We first provide a taxonomy of existing works on LLMs-powered travel planning, which can be categorized into planner, reformulator, and knowledge source, filling a research gap in this field.

2. We propose future research directions for travel planning in the era of LLMs, aiming to expand research horizons and encourage further exploration.

2 Travel Planning Scenarios

In travel planning, different scenarios often correspond to varying travel needs and levels of complexity. To better highlight these distinctions, this section categorizes travel scenarios along two key dimensions: travel type (individual vs. group trip) and travel duration (day tour vs. multi-day tour). This classification helps clarify the basic requirements of each scenario type.

2.1 Travel Type

2.1.1 Individual Trip

Individual trips are a key focus in travel planning research, primarily centered on creating personalized itineraries tailored to a user's preferences. Most traditional research on the Tourist Trip Design Problem (TTDP) has concentrated on individual trip (Ruiz-Meza and Montoya-Torres, 2022; Wörndl et al., 2017; Souffiau et al., 2009; Vansteenwegen et al., 2009), as the planning process only considers the needs of a single user, making it less complex than group travel. Current research on travel planning with LLMs is mostly conducted in this scenario.

2.1.2 Group Trip

In real-world scenarios, tourists may also travel in groups, making it necessary for travel planning to account for the diverse preferences of group members and to find solutions that meet the ends of the entire group. Compared to individual trip, planning for group trip is more complex, as the system must balance the varied needs of members while ensuring fairness. However, research on this problem remains limited (Lim et al., 2016; Anagnostopoulos et al., 2017).

2.2 Travel Duration

2.2.1 Day Tour

A day tour refers to an itinerary that can begin at any time during the day and be completed within the same day, e.g., city tour. These tours require careful consideration of time constraints, information gathering, point-of-interests (POI) selection, route planning, and the personalization of arrangements according to user preferences (Halder et al., 2024).

The key factors in day tour planning can be summarized into two main aspects (Tang et al., 2024): dynamic information adjustment and personalized planning. Due to the flexible start and end times of day tours, dynamic information adjustment is particularly important. The planning process must account for real-time changes in attractions and adapt to unexpected events, such as changes in opening hours or extreme weather conditions. Additionally, personalized planning requires tailoring the itinerary to the specific preferences and time constraints of a user to ensure an optimized experience for them.

2.2.2 Multi-day Tour

Planning a multi-day tour is inherently a complex task as it involves a series of interdependent decisions across various aspects, including destinations, accommodations, transportation, and restaurant arrangements (Xie et al., 2024; Zheng et al., 2024). Compared to single-day trips, multi-day itinerary planning is more challenging, as it requires the careful allocation of daily activities to ensure both coherence and variety, while also considering the travelers' stamina and need for rest. Arranging a multi-day itinerary often entails sequential optimization of locations, taking into account the distances between destinations, transportation conditions, and daily schedules to avoid overexertion or overly tight timelines, therefor providing travelers with a rich and comfortable experience.

Due to the complexity of multi-day travel planning, LLMs struggle to deliver an optimal solution that meets the intricate requirements, and thus their accuracy on this task remains relatively low (Xie et al., 2024; Zheng et al., 2024).

3 Datasets and Evaluation

3.1 Datasets

In the context of travel planning, high-quality datasets specifically designed for LLMs remain scarce. Traditional travel planning datasets often rely on structured data, supporting only limited rules and constraints, which falls short of meeting the complex requirements of practical scenarios. To comprehensively evaluate the actual performance of LLMs in travel planning tasks, it is essential to develop specialized datasets that can encompass multi-level constraints and support natural language interaction. This section introduces an publicly available dataset that is specially designed for travel planning tasks with LLMs: TravelPlanner (Xie et al., 2024)

TravelPlanner (Xie et al., 2024) was constructed by integrating approximately 4 million data points from various open data sources, creating a sandbox environment that supports diverse travel planning tasks. Data sources include the Kaggle Flight dataset, Zomato restaurant dataset, Airbnb accommodations dataset, Google Distance Matrix API (for calculating inter-city distances), and Google Places API (for obtaining POI information). All data have been cleaned and adapted to simulate complex travel scenarios.

It features 1,225 user-generated natural language queries, each incorporating different combinations of constraints and reference plans that cover key elements of a trip such as departure location, destination, and timeframe. For example, "I'd like to travel from Hong Kong to Tokyo from December 8 to 15, 2024. I prefer a more relaxed pace. My budget is $2,000, and I would like a single room." To increase planning complexity, queries are categorized by travel duration (3-day, 5-day, and 7-day trips).Task difficulty is further divided into simple (budget constraints only), medium (budget plus restaurant or lodging requirements), and difficult (multiple constraints such as budget and transportation preferences), thereby testing a model's adaptability and planning abilities under varying constraint combinations.

To ensure data quality and consistency, all queries and corresponding reference plans were meticulously designed by professional annotators, with each plan taking an average of 12 minutes to design. Only plans meeting all predefined constraints were accepted. The data construction process involved multi-stage quality control to guarantee the accuracy and validity of generated plans, providing a reliable benchmark for evaluating LLMs on travel planning tasks.

3.2 Evaluation

To evaluate LLMs for travel planning, we summarize evaluation metrics across two primary dimensions: offline and online assessments. These evaluations comprehensively assessment both the model's efficacy in generating travel plans and user experience.

3.2.1 Offline Evaluation

Offline evaluation focuses on assessing the structural and contextual accuracy of the generated travel plans, emphasizing the model's ability to meet task requirements. Key metrics include:

1. Delivery Rate (Xie et al., 2024): This metric measures the model's ability to successfully generate a complete plan within a predefined step limit, ensuring an efficient planning process that avoids from looping or repeated failures.

2. Final Pass Rate (Xie et al., 2024): This represents the proportion of plans that meet all task-specific constraints, reflecting the model's applicability and the practical feasibility of the generated plans.

3. Exact Match Score (Zheng et al., 2024): By comparing the model's output to the ground-truth plan, this score assesses the level of detail accuracy, quantifying how closely the generated plan aligns with a standard answer.

4. Plan Utility (Li, 2013): This metric aggregates the utility scores of the POIs included in the plan, as an indicator of the quality and relevance of the recommendations in the plan.

3.2.2 Online Evaluation

Online evaluation centers on user interaction, measuring whether the generated plans align with user expectations. Key metrics include:

1. User Satisfaction: Based on user ratings of the generated plans, this metric assesses the practical usefulness and appeal of the plans, providing direct feedback of user acceptance.

2. System Usability and User Experience Questionnaires: Utilizing established questionnaires such as the System Usability Scale (SUS) (Brooke, 1996) and the User Experience Questionnaire (UEQ) (Laugwitz et al., 2008), this approach collects structured feedback on usability and user satisfaction, offering insights into the overall interactive quality.

By summarizing the offline and online evaluation metrics, we establish a comprehensive framework to assess the performance of LLMs on travel planning. It provides a reference for future research and applications in the field, facilitating a balanced focus on both technical effectiveness and user-centered design.

4 How to Apply LLMs to Travel Planning

LLMs, such as GPT-4 and Gemini, have brought about revolutionary changes in NLP and reasoning tasks, demonstrating significant potential in areas like natural language understanding, reasoning, and optimization. (Ge et al., 2024; Team et al., 2023). These models, leveraging the vast knowledge accumulated from extensive public resources and training data, can effectively understand user needs and execute complex instructions. This makes LLMs particularly well-suited for intricate tasks that require broad domain knowledge, such as travel planning. (Valmeekam et al., 2024; Song et al., 2023; Xie et al., 2023a).

Regarding the capabilities of LLMs in planning tasks, three mainstream perspectives exist in the academic community:

1. **Optimists** believe that LLMs not only possess excellent language comprehension abilities but also have the potential to autonomously plan Research has explored the feasibility of applying LLMs for autonomous planning in classic environments such as Blocksworld (Valmeekam et al., 2024), as well as in tasks involving embodied agents (Wang et al., 2023) and web agents (Deng et al., 2024), demonstrating their potential for planning.

2. **Pessimists** are skeptical of the planning capabilities of LLMs, arguing that these models essentially function as advanced translators. They convert reasoning problems embedded in text into symbolic representations, which are then processed by traditional symbolic solvers. (Xie et al., 2023b; Liu et al., 2023; Pan et al., 2023).

3. **Realists** take a more balanced view, asserting that while LLMs cannot independently complete planning tasks, their role extends beyond merely serving as "translators". These models, acting as powerful cognitive assistants, can provide a rich source of knowledge and support planning tasks (Kambhampati et al., 2024). For instance, (Guan et al., 2023) demonstrated that LLMs can act as world models and user preference models in supervised environments, aiding real-world planning tasks and improving planning efficiency.

This diversity of perspectives is also fully reflected in research on travel planning. Based on the three types of attitudes mentioned above, existing methodologies can be categorized into three types according to the roles of LLMs in travel planning tasks: Planner, Reformulator, and Knowledge Source (see Figure 1).

4.1 Planner

The perspective of viewing LLMs as planners reflects the optimistic viewpoint, exploring the potential of LLMs to independently complete planning tasks. As planners, LLMs are tasked with generating a personalized travel itinerary based on natural language inputs from users and external dynamic information. The left side of Figure 1 illustrates the overall framework of such systems, summarizing findings from several studies in recent years (Xie et al., 2024; Tang et al., 2024; Mo et al., 2023; Zheng et al., 2024; Miin and Wei, 2024). Current research mainly distinguishes between two modes: the "Two-Stage Mode" (i.e., tool-use and planning) and the "Sole-Planning Mode". The key difference between them lies in whether the agent utilizes tools to gather information before generating a travel plan.

4.1.1 Two-stage Mode

In this mode, an LLM first employs external tools to gather relevant information, e.g., calling a flight query API to retrieve real-time flight data, and then proceeds to plan based on the collected information. Xie et al. (2024) proposed TravelPlanner, a study that simulates real travel scenarios, creating a sandbox environment with multiple constraints like flights, accommodations, restaurants, and attractions to evaluate the LLM agents' tool-use capabilities as well as their abilities to create reasonable travel plans under various constraints (e.g., budget, time and user preferences). Experiments were conducted using multiple LLMs (e.g., GPT-3.5 and GPT-4) and different planning strategies (e.g., Direct, ZS-CoT (Wei et al., 2022), ReAct (Yao et al., 2022), and Reflexion (Shinn et al., 2024)). The results showed that even the most advanced model GPT-4 only achieved a 0.6% success rate in the two-stage mode, with most issues stemming from errors in tool usage and insufficient information gathering. This highlights the limitations of LLMs in handling complex planning tasks.

4.1.2 Sole-planning Mode

In the solo-planning mode, an LLM generates plans based solely on the available information, without the need for external tools. For example, Tang et al. (2024) applied LLMs to open-domian single-day city itinerary planning. In this study, an LLM generated travel plans that align with user preferences and spatial coherence through reasoning and planning, based on the available information. Mo et al. (2023) explored the ability of LLMs to predict individual travel behavior in a sole-planning mode. They designed prompts that included task descriptions, travel features, and personal attributes, incorporating chain-of-thought and plan-solving strategies. Even without training samples, the predictions made by LLMs were highly competitive, achieving strong accuracy and F1 scores compared to traditional supervised learning methods such as multinomial logistic regression, random forests, and neural networks. Zheng et al. (2024) incorporated intercity flight connectivity information into the LLMs' context to generate travel plans. They found that LLMs could effectively complete travel planning in the sole-planning mode, with

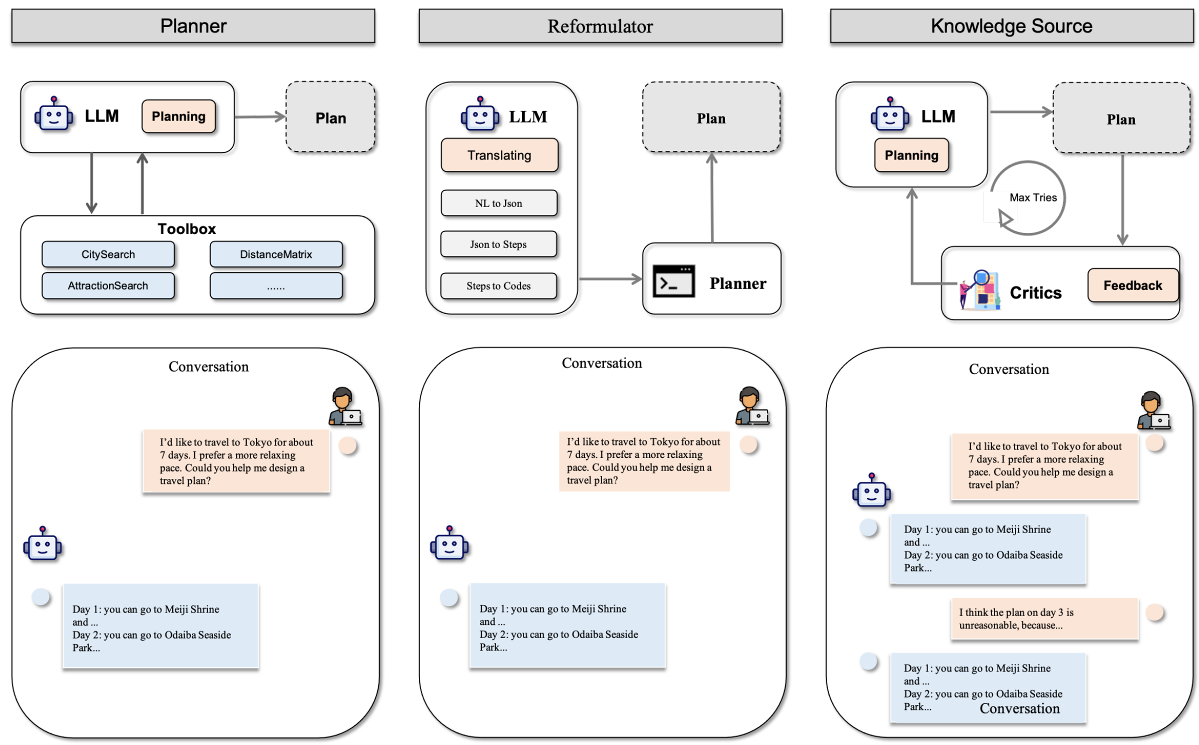

Figure 1: Based on the roles of LLMs in travel planning tasks, existing methodologies can be categorized into three frameworks: Planner, Reformulator, and Knowledge Source.

GPT-4 and Gemini 1.5 Pro achieving 31.1% and 34.8% success rates, respectively. However, as the complexity of the task increased (e.g., involving more cities, people, or days), the performance of LLMs decreased significantly. When it involved 10 cities, the performance of all models dropped below 5%, highlighting the significant gap in current state-of-the-art LLMs' ability to handle natural language-based planning. Miin and Wei (2024) developed a framework incorporating a "human-in-the-loop" feedback mechanism. In this framework, LLMs generate initial prompts, which are iteratively refined with human feedback. The results showed that after one round of human-in-the-loop optimization, the success rate of GPT-4o on the TravelPlanner dataset increased significantly, from 2.78% to 6.67%, demonstrating the potential of human feedback in enhancing the travel planning capabilities of LLMs.

4.2 Reformulator

The view of LLMs as reformulators reflects the pessimist perspective. This line of research leverages LLMs to transform travel planning problems expressed in natural language into structured representations that symbolic solvers, e.g., Planning Domain Description Language (PDDL), then address the complex multi-constraint solving tasks. Previous research (Liu et al., 2023) has demonstrated that LLMs are capable of generating effective PDDL files. The overall framework of such systems is illustrated in the middle of Figure 1. These studies argue that while LLMs excel at parsing human input and facilitating interaction, they are limited in strictly handling all constraints. On the other hand, symbolic solvers are sound and complete when dealing with multi-constraint satisfiability problems but struggle with flexible, general, and sometimes vague natural language demands. Therefore, a framework that combines LLMs with symbolic solvers effectively leverages the strengths of both, overcoming the limitations of LLMs in managing complex constraints and enhancing the efficiency of the entire travel planning process.

Hao et al. (2024) proposed a framework that converts natural language travel planning inputs into Satisfiability Modulo Theories (SMT) problems using an LLM and then solves them with an SMT solver. This framework efficiently solved complex travel planning problems, achieving an impressive 97.0% success rate on the TravelPlanner dataset. When the input query is unsatisfiable, the SMT solver identifies the issue, and the LLM suggests modifications, interacting with the user to refine the query until the constraints are satisfied. Similarly, de la Rosa et al. (2024) introduced the TRIP-PAL framework, which utilizes natural language interactions with users and leverages the broad knowledge of LLMs to identify POIs and user preferences. The user information and travel details are then converted into a data structure that the planner can process. Subsequently, an automated planner generates the optimal travel plan that satisfies the constraints. Experimental results show that TRIP-PAL performs more robustly in complex scenarios compared to using only an LLM for travel planning, particularly when handling more POIs or longer travel durations, effectively maximizing user utility.

4.3 Knowledge Source

The perspective of viewing LLMs as Knowledge source reflects the realists stance, suggesting that LLMs not only possess excellent natural language processing capabilities but also leverage their extensive open-world and domain-specific knowledge to provide knowledge for complex tasks, thereby assisting in planning tasks. The right side of Figure 1 illustrates the overall framework of such a system. In this model, LLMs generate an initial plan, which is critiqued by external critics. These critics may be either the user or another model. Based on the feedback, LLMs subsequently adjust and optimize the plan, iterating through multiple rounds of interaction to ultimately generate a high-quality travel plan.

Ma et al. (2024) proposed the ExploreLLM system, which uses LLMs as a knowledge source. In this system, LLMs decompose users' travel planning needs into multiple subtasks, such as determining dates and duration, booking hotels and flights, etc. For each subtask, LLMs generate multiple alternative options for users to evaluate and select via an interactive interface that can express their preferences. Once all subtasks are completed, ExploreLLM generates a comprehensive travel plan. Experimental results show that ExploreLLM greatly improved planning efficiency and user satisfaction in complex planning tasks.

Kambhampati et al. (2024) further extended this approach with the LLM-Modulo framework, enabling iterative interactions between LLMs and external critics (Gundawar et al., 2024). Within this framework, LLMs first generate an initial travel plan based on contextual information (e.g., flights and hotels). Critics then evaluate the plan, and if it doesn't meet the requirements, LLMs iteratively refine it based on the feedback, generating candidate solutions until consensus or the maximum number of iterations is reached. Experimental results on the TravelPlanner dataset show that LLM-Modulo achieved a final success rate of 20.6%, a 4.6-fold improvement over using LLMs alone, highlighting its effectiveness in handling complex planning tasks.

5 Future Directions

5.1 Datasets for Travel Planning

Traditional itinerary planning typically relies on structured data and predefined rules, whereas LLMs can understand vague requirements and adapt to dynamic contexts through natural language interaction, enabling more flexible planning capabilities. This shift has brought about a demand for new datasets to support LLMs in handling interactive tasks under complex, multi-constraint travel scenarios. However, there is a significant lack of open-source datasets for LLMs-powered travel planning, with TravelPlanner being the only one available benchmark.

While TravelPlanner provides crucial support for evaluating LLMs' performance in travel planning, it has certain limitations, particularly in assessing these models' ability to handle unsatisfiable queries. TravelPlanner is lack of example scenarios where the initial user query cannot be satisfied, which limits the comprehensive evaluation of LLMs' interactive planning capabilities. Specifically, TravelPlanner does not verify LLMs' ability to identify the reasons for unsatisfiability, nor their competence to adjust and optimize plans based on user feedback to better meet user preferences. To bridge this gap, Hao et al. (2024) modified 12 constraints in the TravelPlanner dataset to create unsatisfiable scenarios, and developed an international travel dataset that contains 39 unsatisfiable queries to explore LLMs' repair capabilities when handling such queries. However, this dataset has not been released.

Future research should build more diverse datasets that cover a wide range of travel scenarios and various types of user constraints, especially those that involve complex situations with unsatisfiable queries. Such datasets will not only support the evaluation of LLMs' performance in terms of adaptability and interactive planning but also advance the development of LLMs in intelligent user-centric travel planning systems as well as promoting the applications of LLMs in real-world scenarios.

5.2 Human-in-the-Loop (HITL)

The degree of human supervision, feedback, and intervention for an LLM-based agent during task execution can be viewed as a continuum. At one end of the spectrum, users have full control and validation over all outputs; in the middle, users intervene only when errors occur; and at the other end, the agent can autonomously complete all aspects of the task, including complex causal reasoning. Research by Xie et al. (2024) on the TravelPlanner dataset shows that the success rate of GPT-4 in fully autonomous planning is only 0.6%, revealing the limitations of LLMs in handling complex travel planning tasks and their difficulty in independently completing plans under intricate scenarios.

In certain situations, incorporating human input to enhance AI inference is particularly effective, especially for tasks involving ethical considerations, creative tasks, or ambiguous situations (Durante et al., 2024). Humans can provide critical guidance to the agent, correct errors, and supplement insights that the agent might struggle to infer (Kapoor et al., 2024). For example, Shi et al. (2024) found that simple user feedback improved GPT-4's performance in complex programming tasks from 0% to over 86%, transforming it from nearly ineffective to nearly perfect. Therefore introducing a HITL mechanism to travel planning tasks may also significantly enhance task success rates.

5.3 Multi-agents for Group Travel Planning

In practice, tourists may travel in groups, but research on group travel planning remains relatively limited. Compared to individual travel planning, it needs to accommodate the diverse needs of multiple members that can increase the task complexity.

The application of LLM-based agents in autonomous travel planning represents a promising direction for future research. Intelligent agents are AI-driven systems capable of integrating external knowledge, multimodal inputs, and human feedback, enabling them to autonomously execute complex tasks (Xi et al., 2023). As noted by Zaharia et al. (2024), "AI agents could be the most influential AI trend of 2024, with the potential to maximize AI efficiency in unprecedented ways." Furthermore, AI agents are regarded as a key avenue toward achieving Artificial General Intelligence (AGI) (Durante et al., 2024). With their abilities in contextual understanding, human-like text generation, and complex reasoning, LLM-powered agents facilitate more engaging and smooth interactions between users and travel planning systems, offering them a better experience than traditional methods.

Recent studies show that multi-agent negotiation frameworks based on LLMs perform well in group recommendation tasks (Ji and Ma, 2023; Alves et al., 2023), which may shed light on group travel planning. In a multi-agent system, each agent represents a member, negotiating with each other to generate an itinerary that meets the requirements of the majority. Future research could explore LLM-driven multi-agent frameworks to simulate member preferences and reconcile conflicts.

6 Conclusion

In this paper, we provide a comprehensive review of the application of LLMs in travel planning, and discuss their potential in delivering personalized and efficient solutions. Based on the key roles of LLMs in travel planning tasks, we present a taxonomy of existing methodologies, categorizing them into three types: Planner, Reformulator, and Knowledge Source. In the meantime, we summarize the main application scenarios, available datasets, and evaluation methods for LLMs-powered travel planning, and point out future directions. We believe that, with the continuous advancement of AI, LLMs-powered travel planning can offer more practical and efficient solutions for the tourism industry. We hope that this review can provide valuable background information and guidance for practitioners in the field to advance its development.

Limitations

Nevertheless, this paper presents several limitations. Firstly, our discussion is limited to travel planning and does not extend to other aspects such as transportation planning during the trip. Secondly, available open-source datasets for travel planning with LLMs are limited, leading Section 3 to list only a single dataset.

References

RA Abbaspour and F Samadzadegan. 2009. Itinerary planning in multimodal urban transportation network. *Journal of Applied Sciences*, 9(10):1898–1906.

Patrícia Alves, André Martins, Paulo Novais, and Goreti Marreiros. 2023. Improving group recommendations using personality, dynamic clustering and multi-agent microservices. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 1165–1168.

Aris Anagnostopoulos, Reem Atassi, Luca Becchetti, Adriano Fazzone, and Fabrizio Silvestri. 2017. Tour recommendation for groups. *Data Mining and Knowledge Discovery*, 31(5):1157–1188.

Oxford Analytica. 2024. Tourism in lebanon will take years to recover. *Emerald Expert Briefings*, (oxan-es).

Plaban K Bhowmick, Soumyajit Dey, Abinash Samantaray, Debnath Mukherjee, and Prateep Misra. 2012. A temporal constraint based planning approach for city tour and travel plan generation. In *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, pages 1–6. IEEE.

Igo Ramalho Brilhante, Jose Antonio Macedo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. 2015. On planning sightseeing tours with tripbuilder. *Information Processing & Management*, 51(2):1–15.

J Brooke. 1996. Sus: A quick and dirty usability scale. *Usability Evaluation in Industry*.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Buru Chang, Yonggyu Park, Donghyeon Park, Seongsoon Kim, and Jaewoo Kang. 2018. Content-aware hierarchical point-of-interest embedding model for successive poi recommendation. In *IJCAI*, volume 20, pages 3301–3307.

Gautier Dagan, Frank Keller, and Alex Lascarides. 2023. Dynamic planning with a llm. *arXiv preprint arXiv:2308.06391*.

Tomas de la Rosa, Sriram Gopalakrishnan, Alberto Pozanco, Zhen Zeng, and Daniel Borrajo. 2024. Trip-pal: Travel planning with guarantees by combining large language models and automated planners. *arXiv preprint arXiv:2406.10196*.

Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Sam Stevens, Boshi Wang, Huan Sun, and Yu Su. 2024. Mind2web: Towards a generalist agent for the web. *Advances in Neural Information Processing Systems*, 36.

Z Durante, Q Huang, N Wake, R Gong, JS Park, B Sarkar, R Taori, Y Noda, D Terzopoulos, Y Choi, et al. 2024. Agent ai: surveying the horizons of multimodal interaction. arxiv.

Fedor V Fomin and Andrzej Lingas. 2002. Approximation algorithms for time-dependent orienteering. *Information Processing Letters*, 83(2):57–62.

Liping Gao, Chao Chen, Feng Chu, Chengwu Liao, Hongyu Huang, and Yasha Wang. 2023. Moop: An efficient utility-rich route planning framework over two-fold time-dependent road networks. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(5):1554–1570.

Ines Gasmi, Makram Soui, Khaoula Barhoumi, and Mourad Abed. 2024. Recommendation rules to personalize itineraries for tourists in an unfamiliar city. *Applied Soft Computing*, 150:111084.

Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, and Grammati Pantziou. 2014. A survey on algorithmic approaches for solving tourist trip design problems. *Journal of Heuristics*, 20(3):291–328.

Yingqiang Ge, Wenyue Hua, Kai Mei, Juntao Tan, Shuyuan Xu, Zelong Li, Yongfeng Zhang, et al. 2024. Openagi: When llm meets domain experts. *Advances in Neural Information Processing Systems*, 36.

Lin Guan, Karthik Valmeekam, Sarath Sreedharan, and Subbarao Kambhampati. 2023. Leveraging pre-trained large language models to construct and utilize world models for model-based task planning. *Advances in Neural Information Processing Systems*, 36:79081–79094.

Atharva Gundawar, Mudit Verma, Lin Guan, Karthik Valmeekam, Siddhant Bhambri, and Subbarao Kambhampati. 2024. Robust planning with llm-modulo framework: Case study in travel planning. *arXiv preprint arXiv:2405.20625*.

Sajal Halder, Kwan Hui Lim, Jeffrey Chan, and Xiuzhen Zhang. 2024. A survey on personalized itinerary recommendation: From optimisation to deep learning. *Applied Soft Computing*, 152:111200.

Yilun Hao, Yongchao Chen, Yang Zhang, and Chuchu Fan. 2024. Large language models can plan your travels rigorously with formal verification tools. *arXiv preprint arXiv:2404.11891*.

Daniel Herzog, Linus W Dietz, and Wolfgang Wörndl. 2019. Tourist trip recommendations–foundations, state of the art and challenges. *Personalized Human-Computer Interaction*, 6:159–182.

Yu-Ling Hsueh and Hong-Min Huang. 2019. Personalized itinerary recommendation with time constraints using gps datasets. *Knowledge and information systems*, 60(1):523–544.

Xu Huang, Weiwen Liu, Xiaolong Chen, Xingmei Wang, Defu Lian, Yasheng Wang, Ruiming Tang, and Enhong Chen. 2024a. Wese: Weak exploration to strong exploitation for llm agents. *arXiv preprint arXiv:2404.07456*.

Xu Huang, Weiwen Liu, Xiaolong Chen, Xingmei Wang, Hao Wang, Defu Lian, Yasheng Wang, Ruiming Tang, and Enhong Chen. 2024b. Understanding the planning of llm agents: A survey. *arXiv preprint arXiv:2402.02716*.

Pu Ji and Xiaoyu Ma. 2023. A fuzzy intelligent group recommender method in sparse-data environments based on multi-agent negotiation. *Expert Systems with Applications*, 213:119294.

Jose Luis Jorro-Aragoneses, Maria Belén Díaz Agudo, and Juan Antonio Recio García. 2017. Madrid live: a context-aware recommender systems of leisure plans. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 796–801. IEEE.

Subbarao Kambhampati, Karthik Valmeekam, Lin Guan, Mudit Verma, Kaya Stechly, Siddhant Bhambri, Lucas Saldyt, and Anil Murthy. 2024. Llms can't plan, but can help planning in llm-modulo frameworks. *arXiv preprint arXiv:2402.01817*.

Marisa G Kantor and Moshe B Rosenwein. 1992. The orienteering problem with time windows. *Journal of the Operational Research Society*, 43(6):629–635.

Sayash Kapoor, Benedikt Stroebl, Zachary S Siegel, Nitya Nadgir, and Arvind Narayanan. 2024. Ai agents that matter. *arXiv preprint arXiv:2407.01502*.

Joanna Karbowska-Chilinska and Kacper Chociej. 2019. Optimization of multistage tourist route for electric vehicle. In *Artificial Intelligence and Algorithms in Intelligent Systems: Proceedings of 7th Computer Science On-line Conference 2018, Volume 2 7*, pages 186–196. Springer.

Aristea Kontogianni and Efthimios Alepis. 2020. Smart tourism: State of the art and literature review for the last six years. *Array*, 6:100020.

Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings 4*, pages 63–76. Springer.

Xun Li. 2013. Multi-day and multi-stay travel planning using geo-tagged photos. In *Proceedings of the second ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information*, pages 1–8.

Kwan Hui Lim, Jeffrey Chan, Christopher Leckie, and Shanika Karunasekera. 2016. Towards next generation touring: Personalized group tours. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 26, pages 412–420.

Kwan Hui Lim, Jeffrey Chan, Christopher Leckie, and Shanika Karunasekera. 2018. Personalized trip recommendation for tourists based on user interests, points of interest visit durations and visit recency. *Knowledge and Information Systems*, 54:375–406.

Bo Liu, Yuqian Jiang, Xiaohan Zhang, Qiang Liu, Shiqi Zhang, Joydeep Biswas, and Peter Stone. 2023. Llm+ p: Empowering large language models with optimal planning proficiency. *arXiv preprint arXiv:2304.11477*.

Derek Long. 2005. Automated planning: Theory and practice. *Assembly Automation*, 25(2).

Xiao Ma, Swaroop Mishra, Ariel Liu, Sophie Ying Su, Jilin Chen, Chinmay Kulkarni, Heng-Tze Cheng, Quoc Le, and Ed Chi. 2024. Beyond chatbots: Explorellm for structured thoughts and personalized model responses. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–12.

Annabelle Miin and Timothy Wei. 2024. Smart language agents in real-world planning. *arXiv preprint arXiv:2407.19667*.

Baichuan Mo, Hanyong Xu, Dingyi Zhuang, Ruoyun Ma, Xiaotong Guo, and Jinhua Zhao. 2023. Large language models for travel behavior prediction. *arXiv preprint arXiv:2312.00819*.

Liangming Pan, Alon Albalak, Xinyi Wang, and William Yang Wang. 2023. Logic-lm: Empowering large language models with symbolic solvers for faithful logical reasoning. *arXiv preprint arXiv:2305.12295*.

Guangyao Pang, Xiaoming Wang, Fei Hao, Liang Wang, and Xinyan Wang. 2020. Efficient point-of-interest recommendation with hierarchical attention mechanism. *Applied Soft Computing*, 96:106536.

Beatriz Rodríguez, Julián Molina, Fátima Pérez, and Rafael Caballero. 2012. Interactive design of personalised tourism routes. *Tourism Management*, 33(4):926–940.

José Ruiz-Meza and Jairo R Montoya-Torres. 2022. A systematic literature review for the tourist trip design problem: Extensions, solution techniques and future research lines. *Operations Research Perspectives*, 9:100228.

Quan Shi, Michael Tang, Karthik Narasimhan, and Shunyu Yao. 2024. Can language models solve olympiad programming? *arXiv preprint arXiv:2404.10952*.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2024. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36.

Mohit Shridhar, Xingdi Yuan, Marc-Alexandre Côté, Yonatan Bisk, Adam Trischler, and Matthew Hausknecht. 2020. Alfworld: Aligning text and embodied environments for interactive learning. *arXiv preprint arXiv:2010.03768*.

David Silver, Hado Hasselt, Matteo Hessel, Tom Schaul, Arthur Guez, Tim Harley, Gabriel Dulac-Arnold, David Reichert, Neil Rabinowitz, Andre Barreto, et al. 2017. The predictron: End-to-end learning and planning. In *International Conference on Machine Learning*, pages 3191–3199. PMLR.

Chan Hee Song, Jiaman Wu, Clayton Washington, Brian M Sadler, Wei-Lun Chao, and Yu Su. 2023. Llm-planner: Few-shot grounded planning for embodied agents with large language models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2998–3009.

Wouter Souffiau, Joris Maervoet, Pieter Vansteenwegen, Greet Vanden Berghe, and Dirk Van Oudheusden. 2009. A mobile tourist decision support system for small footprint devices. In *Bio-Inspired Systems: Computational and Ambient Intelligence: 10th International Work-Conference on Artificial Neural Networks, IWANN 2009, Salamanca, Spain, June 10-12, 2009. Proceedings, Part I 10*, pages 1248–1255. Springer.

Theodore R Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L Griffiths. 2023. Cognitive architectures for language agents. *arXiv preprint arXiv:2309.02427*.

Yihong Tang, Zhaokai Wang, Ao Qu, Yihao Yan, Kebing Hou, Dingyi Zhuang, Xiaotong Guo, Jinhua Zhao, Zhan Zhao, and Wei Ma. 2024. Synergizing spatial optimization with large language models for open-domain urban itinerary planning. *arXiv preprint arXiv:2402.07204*.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Alan Toledo, Maria-Cristina Riff, and Bertrand Neveu. 2019. A hyper-heuristic for the orienteering problem with hotel selection. *IEEE access*, 8:1303–1313.

Karthik Valmeekam, Matthew Marquez, Alberto Olmo, Sarath Sreedharan, and Subbarao Kambhampati. 2024. Planbench: An extensible benchmark for evaluating large language models on planning and reasoning about change. *Advances in Neural Information Processing Systems*, 36.

Pieter Vansteenwegen, Wouter Souffriau, Greet Vanden Berghe, and Dirk Van Oudheusden. 2009. Iterated local search for the team orienteering problem with time windows. *Computers & Operations Research*, 36(12):3281–3290.

Pieter Vansteenwegen, Wouter Souffriau, Greet Vanden Berghe, and Dirk Van Oudheusden. 2011. The city trip planner: an expert system for tourists. *Expert Systems with Applications*, 38(6):6540–6546.

Nikolaos Vathis, Charalampos Konstantopoulos, Grammati Pantziou, and Damianos Gavalas. 2023. The vacation planning problem: A multilevel clustering-based metaheuristic approach. *Computers & Operations Research*, 150:106083.

Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2023. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*.

Ruoyao Wang, Peter Jansen, Marc-Alexandre Côté, and Prithviraj Ammanabrolu. 2022. Scienceworld: Is your agent smarter than a 5th grader? *arXiv preprint arXiv:2203.07540*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Wolfgang Wörndl, Alexander Hefele, and Daniel Herzog. 2017. Recommending a sequence of interesting places for tourist trips. *Information Technology & Tourism*, 17:31–54.

Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. 2023. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*.

Jian Xie, Kai Zhang, Jiangjie Chen, Tinghui Zhu, Renze Lou, Yuandong Tian, Yanghua Xiao, and Yu Su. 2024. Travelplanner: A benchmark for real-world planning with language agents. *arXiv preprint arXiv:2402.01622*.

Tianbao Xie, Fan Zhou, Zhoujun Cheng, Peng Shi, Luoxuan Weng, Yitao Liu, Toh Jing Hua, Junning Zhao, Qian Liu, Che Liu, et al. 2023a. Openagents: An open platform for language agents in the wild. *arXiv preprint arXiv:2310.10634*.

Yaqi Xie, Chen Yu, Tongyao Zhu, Jinbin Bai, Ze Gong, and Harold Soh. 2023b. Translating natural language to planning goals with large-language models. *arXiv preprint arXiv:2302.05128*.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*.

Matei Zaharia, Omar Khattab, Lingjiao Chen, Jared Quincy Davis, Heather Miller, Chris Potts, James Zou, Michael Carbin, Jonathan Frankle, Naveen Rao, et al. 2024. The shift from models to compound ai systems. *Berkeley Artificial Intelligence Research Lab. Available online at: https://bair. berkeley. edu/blog/2024/02/18/compound-ai-systems/(accessed February 27, 2024)*.

Zirui Zhao, Wee Sun Lee, and David Hsu. 2024. Large language models as commonsense knowledge for large-scale task planning. *Advances in Neural Information Processing Systems*, 36.

Huaixiu Steven Zheng, Swaroop Mishra, Hugh Zhang, Xinyun Chen, Minmin Chen, Azade Nova, Le Hou, Heng-Tze Cheng, Quoc V Le, Ed H Chi, et al. 2024. Natural plan: Benchmarking llms on natural language planning. *arXiv preprint arXiv:2406.04520*.

Weimin Zheng and Zhixue Liao. 2019. Using a heuristic approach to design personalized tour routes for heterogeneous tourist groups. *Tourism Management*, 72:313–325.

Yu Zheng, Lizhu Zhang, Zhengxin Ma, Xing Xie, and Wei-Ying Ma. 2011. Recommending friends and locations based on individual location history. *ACM Transactions on the Web (TWEB)*, 5(1):1–44.

Chenbo Zhu, Jian-Qiang Hu, Fengchun Wang, Yifan Xu, and Rongzeng Cao. 2012. On the tour planning problem. *Annals of Operations Research*, 192:67–86.

References

RA Abbaspour and F Samadzadegan. 2009. Itinerary planning in multimodal urban transportation network. *Journal of Applied Sciences*, 9(10):1898–1906.

Patrícia Alves, André Martins, Paulo Novais, and Goreti Marreiros. 2023. Improving group recommendations using personality, dynamic clustering and multi-agent microservices. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 1165–1168.

Aris Anagnostopoulos, Reem Atassi, Luca Becchetti, Adriano Fazzone, and Fabrizio Silvestri. 2017. Tour recommendation for groups. *Data Mining and Knowledge Discovery*, 31(5):1157–1188.

Oxford Analytica. 2024. Tourism in lebanon will take years to recover. *Emerald Expert Briefings*, (oxan-es).

Plaban K Bhowmick, Soumyajit Dey, Abinash Samantaray, Debnath Mukherjee, and Prateep Misra. 2012. A temporal constraint based planning approach for city tour and travel plan generation. In *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, pages 1–6. IEEE.

Igo Ramalho Brilhante, Jose Antonio Macedo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. 2015. On planning sightseeing tours with tripbuilder. *Information Processing & Management*, 51(2):1–15.

J Brooke. 1996. Sus: A quick and dirty usability scale. *Usability Evaluation in Industry*.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Buru Chang, Yonggyu Park, Donghyeon Park, Seongsoon Kim, and Jaewoo Kang. 2018. Content-aware hierarchical point-of-interest embedding model for successive poi recommendation. In *IJCAI*, volume 20, pages 3301–3307.

Gautier Dagan, Frank Keller, and Alex Lascarides. 2023. Dynamic planning with a llm. *arXiv preprint arXiv:2308.06391*.

Tomas de la Rosa, Sriram Gopalakrishnan, Alberto Pozanco, Zhen Zeng, and Daniel Borrajo. 2024. Trip-pal: Travel planning with guarantees by combining large language models and automated planners. *arXiv preprint arXiv:2406.10196*.

Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Sam Stevens, Boshi Wang, Huan Sun, and Yu Su.

2024. Mind2web: Towards a generalist agent for the web. *Advances in Neural Information Processing Systems*, 36.

Z Durante, Q Huang, N Wake, R Gong, JS Park, B Sarkar, R Taori, Y Noda, D Terzopoulos, Y Choi, et al. 2024. Agent ai: surveying the horizons of multimodal interaction. arxiv.

Fedor V Fomin and Andrzej Lingas. 2002. Approximation algorithms for time-dependent orienteering. *Information Processing Letters*, 83(2):57–62.

Liping Gao, Chao Chen, Feng Chu, Chengwu Liao, Hongyu Huang, and Yasha Wang. 2023. Moop: An efficient utility-rich route planning framework over two-fold time-dependent road networks. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(5):1554–1570.

Ines Gasmi, Makram Soui, Khaoula Barhoumi, and Mourad Abed. 2024. Recommendation rules to personalize itineraries for tourists in an unfamiliar city. *Applied Soft Computing*, 150:111084.

Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, and Grammati Pantziou. 2014. A survey on algorithmic approaches for solving tourist trip design problems. *Journal of Heuristics*, 20(3):291–328.

Yingqiang Ge, Wenyue Hua, Kai Mei, Juntao Tan, Shuyuan Xu, Zelong Li, Yongfeng Zhang, et al. 2024. Openagi: When llm meets domain experts. *Advances in Neural Information Processing Systems*, 36.

Lin Guan, Karthik Valmeekam, Sarath Sreedharan, and Subbarao Kambhampati. 2023. Leveraging pre-trained large language models to construct and utilize world models for model-based task planning. *Advances in Neural Information Processing Systems*, 36:79081–79094.

Atharva Gundawar, Mudit Verma, Lin Guan, Karthik Valmeekam, Siddhant Bhambri, and Subbarao Kambhampati. 2024. Robust planning with llm-modulo framework: Case study in travel planning. *arXiv preprint arXiv:2405.20625*.

Sajal Halder, Kwan Hui Lim, Jeffrey Chan, and Xiuzhen Zhang. 2024. A survey on personalized itinerary recommendation: From optimisation to deep learning. *Applied Soft Computing*, 152:111200.

Yilun Hao, Yongchao Chen, Yang Zhang, and Chuchu Fan. 2024. Large language models can plan your travels rigorously with formal verification tools. *arXiv preprint arXiv:2404.11891*.

Daniel Herzog, Linus W Dietz, and Wolfgang Wörndl. 2019. Tourist trip recommendations–foundations, state of the art and challenges. *Personalized Human-Computer Interaction*, 6:159–182.

Yu-Ling Hsueh and Hong-Min Huang. 2019. Personalized itinerary recommendation with time constraints using gps datasets. *Knowledge and information systems*, 60(1):523–544.

Xu Huang, Weiwen Liu, Xiaolong Chen, Xingmei Wang, Defu Lian, Yasheng Wang, Ruiming Tang, and Enhong Chen. 2024a. Wese: Weak exploration to strong exploitation for llm agents. *arXiv preprint arXiv:2404.07456*.

Xu Huang, Weiwen Liu, Xiaolong Chen, Xingmei Wang, Hao Wang, Defu Lian, Yasheng Wang, Ruiming Tang, and Enhong Chen. 2024b. Understanding the planning of llm agents: A survey. *arXiv preprint arXiv:2402.02716*.

Pu Ji and Xiaoyu Ma. 2023. A fuzzy intelligent group recommender method in sparse-data environments based on multi-agent negotiation. *Expert Systems with Applications*, 213:119294.

Jose Luis Jorro-Aragoneses, Maria Belén Díaz Agudo, and Juan Antonio Recio García. 2017. Madrid live: a context-aware recommender systems of leisure plans. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 796–801. IEEE.

Subbarao Kambhampati, Karthik Valmeekam, Lin Guan, Mudit Verma, Kaya Stechly, Siddhant Bhambri, Lucas Saldyt, and Anil Murthy. 2024. Llms can't plan, but can help planning in llm-modulo frameworks. *arXiv preprint arXiv:2402.01817*.

Marisa G Kantor and Moshe B Rosenwein. 1992. The orienteering problem with time windows. *Journal of the Operational Research Society*, 43(6):629–635.

Sayash Kapoor, Benedikt Stroebl, Zachary S Siegel, Nitya Nadgir, and Arvind Narayanan. 2024. Ai agents that matter. *arXiv preprint arXiv:2407.01502*.

Joanna Karbowska-Chilinska and Kacper Chociej. 2019. Optimization of multistage tourist route for electric vehicle. In *Artificial Intelligence and Algorithms in Intelligent Systems: Proceedings of 7th Computer Science On-line Conference 2018, Volume 2 7*, pages 186–196. Springer.

Aristea Kontogianni and Efthimios Alepis. 2020. Smart tourism: State of the art and literature review for the last six years. *Array*, 6:100020.

Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings 4*, pages 63–76. Springer.

Xun Li. 2013. Multi-day and multi-stay travel planning using geo-tagged photos. In *Proceedings of the second ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information*, pages 1–8.

Kwan Hui Lim, Jeffrey Chan, Christopher Leckie, and Shanika Karunasekera. 2016. Towards next generation touring: Personalized group tours. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 26, pages 412–420.

Kwan Hui Lim, Jeffrey Chan, Christopher Leckie, and Shanika Karunasekera. 2018. Personalized trip recommendation for tourists based on user interests, points of interest visit durations and visit recency. *Knowledge and Information Systems*, 54:375–406.

Bo Liu, Yuqian Jiang, Xiaohan Zhang, Qiang Liu, Shiqi Zhang, Joydeep Biswas, and Peter Stone. 2023. Llm+ p: Empowering large language models with optimal planning proficiency. *arXiv preprint arXiv:2304.11477*.

Derek Long. 2005. Automated planning: Theory and practice. *Assembly Automation*, 25(2).

Xiao Ma, Swaroop Mishra, Ariel Liu, Sophie Ying Su, Jilin Chen, Chinmay Kulkarni, Heng-Tze Cheng, Quoc Le, and Ed Chi. 2024. Beyond chatbots: Explorellm for structured thoughts and personalized model responses. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–12.

Annabelle Miin and Timothy Wei. 2024. Smart language agents in real-world planning. *arXiv preprint arXiv:2407.19667*.

Baichuan Mo, Hanyong Xu, Dingyi Zhuang, Ruoyun Ma, Xiaotong Guo, and Jinhua Zhao. 2023. Large language models for travel behavior prediction. *arXiv preprint arXiv:2312.00819*.

Liangming Pan, Alon Albalak, Xinyi Wang, and William Yang Wang. 2023. Logic-lm: Empowering large language models with symbolic solvers for faithful logical reasoning. *arXiv preprint arXiv:2305.12295*.

Guangyao Pang, Xiaoming Wang, Fei Hao, Liang Wang, and Xinyan Wang. 2020. Efficient point-of-interest recommendation with hierarchical attention mechanism. *Applied Soft Computing*, 96:106536.

Beatriz Rodríguez, Julián Molina, Fátima Pérez, and Rafael Caballero. 2012. Interactive design of personalised tourism routes. *Tourism Management*, 33(4):926–940.

José Ruiz-Meza and Jairo R Montoya-Torres. 2022. A systematic literature review for the tourist trip design problem: Extensions, solution techniques and future research lines. *Operations Research Perspectives*, 9:100228.

Quan Shi, Michael Tang, Karthik Narasimhan, and Shunyu Yao. 2024. Can language models solve olympiad programming? *arXiv preprint arXiv:2404.10952*.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2024. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36.

Mohit Shridhar, Xingdi Yuan, Marc-Alexandre Côté, Yonatan Bisk, Adam Trischler, and Matthew Hausknecht. 2020. Alfworld: Aligning text and embodied environments for interactive learning. *arXiv preprint arXiv:2010.03768*.

David Silver, Hado Hasselt, Matteo Hessel, Tom Schaul, Arthur Guez, Tim Harley, Gabriel Dulac-Arnold, David Reichert, Neil Rabinowitz, Andre Barreto, et al. 2017. The predictron: End-to-end learning and planning. In *International Conference on Machine Learning*, pages 3191–3199. PMLR.

Chan Hee Song, Jiaman Wu, Clayton Washington, Brian M Sadler, Wei-Lun Chao, and Yu Su. 2023. Llm-planner: Few-shot grounded planning for embodied agents with large language models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2998–3009.

Wouter Souffiau, Joris Maervoet, Pieter Vansteenwegen, Greet Vanden Berghe, and Dirk Van Oudheusden. 2009. A mobile tourist decision support system for small footprint devices. In *Bio-Inspired Systems: Computational and Ambient Intelligence: 10th International Work-Conference on Artificial Neural Networks, IWANN 2009, Salamanca, Spain, June 10-12, 2009. Proceedings, Part I 10*, pages 1248–1255. Springer.

Theodore R Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L Griffiths. 2023. Cognitive architectures for language agents. *arXiv preprint arXiv:2309.02427*.

Yihong Tang, Zhaokai Wang, Ao Qu, Yihao Yan, Kebing Hou, Dingyi Zhuang, Xiaotong Guo, Jinhua Zhao, Zhan Zhao, and Wei Ma. 2024. Synergizing spatial optimization with large language models for open-domain urban itinerary planning. *arXiv preprint arXiv:2402.07204*.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Alan Toledo, Maria-Cristina Riff, and Bertrand Neveu. 2019. A hyper-heuristic for the orienteering problem with hotel selection. *IEEE access*, 8:1303–1313.

Karthik Valmeekam, Matthew Marquez, Alberto Olmo, Sarath Sreedharan, and Subbarao Kambhampati. 2024. Planbench: An extensible benchmark for evaluating large language models on planning and reasoning about change. *Advances in Neural Information Processing Systems*, 36.

Pieter Vansteenwegen, Wouter Souffriau, Greet Vanden Berghe, and Dirk Van Oudheusden. 2009. Iterated local search for the team orienteering problem with time windows. *Computers & Operations Research*, 36(12):3281–3290.

Pieter Vansteenwegen, Wouter Souffriau, Greet Vanden Berghe, and Dirk Van Oudheusden. 2011. The city trip planner: an expert system for tourists. *Expert Systems with Applications*, 38(6):6540–6546.

Nikolaos Vathis, Charalampos Konstantopoulos, Grammati Pantziou, and Damianos Gavalas. 2023. The vacation planning problem: A multi-level clustering-based metaheuristic approach. *Computers & Operations Research*, 150:106083.

Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2023. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*.

Ruoyao Wang, Peter Jansen, Marc-Alexandre Côté, and Prithviraj Ammanabrolu. 2022. Scienceworld: Is your agent smarter than a 5th grader? *arXiv preprint arXiv:2203.07540*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Wolfgang Wörndl, Alexander Hefele, and Daniel Herzog. 2017. Recommending a sequence of interesting places for tourist trips. *Information Technology & Tourism*, 17:31–54.

Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. 2023. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*.

Jian Xie, Kai Zhang, Jiangjie Chen, Tinghui Zhu, Renze Lou, Yuandong Tian, Yanghua Xiao, and Yu Su. 2024. Travelplanner: A benchmark for real-world planning with language agents. *arXiv preprint arXiv:2402.01622*.

Tianbao Xie, Fan Zhou, Zhoujun Cheng, Peng Shi, Luoxuan Weng, Yitao Liu, Toh Jing Hua, Junning Zhao, Qian Liu, Che Liu, et al. 2023a. Openagents: An open platform for language agents in the wild. *arXiv preprint arXiv:2310.10634*.

Yaqi Xie, Chen Yu, Tongyao Zhu, Jinbin Bai, Ze Gong, and Harold Soh. 2023b. Translating natural language to planning goals with large-language models. *arXiv preprint arXiv:2302.05128*.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*.

Matei Zaharia, Omar Khattab, Lingjiao Chen, Jared Quincy Davis, Heather Miller, Chris Potts, James Zou, Michael Carbin, Jonathan Frankle, Naveen Rao, et al. 2024. The shift from models to compound ai systems. *Berkeley Artificial Intelligence Research Lab. Available online at: https://bair.berkeley. edu/blog/2024/02/18/compound-ai-systems/(accessed February 27, 2024)*.

Zirui Zhao, Wee Sun Lee, and David Hsu. 2024. Large language models as commonsense knowledge for large-scale task planning. *Advances in Neural Information Processing Systems*, 36.

Huaixiu Steven Zheng, Swaroop Mishra, Hugh Zhang, Xinyun Chen, Minmin Chen, Azade Nova, Le Hou, Heng-Tze Cheng, Quoc V Le, Ed H Chi, et al. 2024. Natural plan: Benchmarking llms on natural language planning. *arXiv preprint arXiv:2406.04520*.

Weimin Zheng and Zhixue Liao. 2019. Using a heuristic approach to design personalized tour routes for heterogeneous tourist groups. *Tourism Management*, 72:313–325.

Yu Zheng, Lizhu Zhang, Zhengxin Ma, Xing Xie, and Wei-Ying Ma. 2011. Recommending friends and locations based on individual location history. *ACM Transactions on the Web (TWEB)*, 5(1):1–44.

Chenbo Zhu, Jian-Qiang Hu, Fengchun Wang, Yifan Xu, and Rongzeng Cao. 2012. On the tour planning problem. *Annals of Operations Research*, 192:67–86.

A Appendix

Traditional TTDP solutions are generally classified into two categories: recommender systems and operations research methods. Recommender systems frame TTDP as a recommendation problem, such as POI recommendation or travel plan recommendation. They typically leverage common techniques from the recommendation domain, including collaborative filtering and deep learning, to address it (Halder et al., 2024). The solutions leverage common operations research techniques, including exact algorithms, heuristic algorithms, and metaheuristic algorithms. The solutions leverage common operations research techniques, including exact algorithms, heuristic algorithms, and metaheuristic algorithms. (Ruiz-Meza and Montoya-Torres, 2022).

Substantial research has been conducted in the field of traditional travel planning. For example, Ruiz-Meza and Montoya-Torres (2022) developed a taxonomy of existing TTDP research based on the type of optimization objectives (single-objective vs. multi-objective) and conducted a comprehensive analysis of TTDP modeling approaches from an operations research perspective, covering major OP variants and their associated solution techniques. Similarly, Gavalas et al. (2014) discussed the models, algorithms, and methodologies for the Tourist Route Design Problem, gradually extending the basic OP from an operations research perspective.

Additionally, Herzog et al. (2019) reviewed TTDP research from the perspective of itinerary recommender systems, highlighting advancements in recommendation techniques, data analysis, and user interfaces. Halder et al. (2024) Reviewed the entire process of itinerary recommendation, covering data processing and evaluation methods, as well as algorithms tailored for individual tourists and tourist groups. Meanwhile, Kontogianni and Alepis (2020) summarized key concepts in the field of smart tourism, including social media, context awareness, and the Internet of Things.

A.1 Recommender systems

In the context of recommender systems, travel planning can be further divided into two categories: POI recommendation and itinerary recommendation. POI recommendation aims to suggest a series of attractions to users, while itinerary recommendation integrates multiple POIs into a comprehensive trip plan, considering time, distance, and other constraints (Halder et al., 2024). Research on POI recommendation has been extensively explored. For instance, Jorro-Aragoneses et al. (2017) introduced the Madrid Live context-aware recommender system, which combines user preferences, location, and weather factors to recommend tourist and leisure activities in Madrid. Zheng et al. (2011) designed a personalized Geographic Information System (GIS) that predicts user preferences for unvisited locations by analyzing the user's location history and the location data of similar tourists. The CAPE model (Chang et al., 2018) recommends POIs based on users' check-in records and textual information about POIs but does not account for personalized preferences. In contrast, Pang et al. (2020) proposed a POI recommendation method based on a hierarchical attention mechanism to improve recommendation accuracy.

Significant progress has also been made in the area of itinerary recommendation. Lim et al. (2018) proposed the PersTour system, which combines POI visit times with user interest preferences to provide personalized itinerary recommendations. Their study demonstrates that time-based interests play a more significant role than frequency-based interests in trip planning. Brilhante et al. (2015) introduced the TripBuilder algorithm, which models the travel recommendation problem as a Generalized Maximum Coverage (GMC) problem, aiming to optimize POI popularity and user preferences within the user's available time. Gasmi et al. (2024) used multi-objective evolutionary algorithms (such as NSGA-II, SPEA2, and IBEA) to generate personalized itinerary recommendations, aiming to balance POI popularity and user interests for tourists unfamiliar with a city. Comparative studies based on Flickr datasets from different cities showed that NSGA-II performed particularly well in providing personalized itinerary recommendations that meet tourists' needs. Vansteenwegen et al. (2011) developed the City Trip Planner, a web application for planning multi-day trips, which generates recommended itineraries based on the opening and closing times of each POI.

A.2 Operations research methods

Operations research methods typically model the TTDP as an OP or one of its variants (Ruiz-Meza and Montoya-Torres, 2022). The OP involves selecting from multiple candidate POIs, considering both the score of each POI and time constraints, with the goal of planning the optimal route that maximizes the total score of the visited POIs within the available time budget (Souffiau et al., 2009; Vansteenwegen et al., 2009).

Based on Rodríguez et al. (2012); Ruiz-Meza and Montoya-Torres (2022), when modeling TTDP as an OP, the objective is to select a subset of POIs from a set of locations p_i (where $i \in \{1, 2, ..., N\}$, and N is the total number of locations), and maximize the sum of the POI scores s_i within a time budget $T_{\max}$. The objective function can be expressed as:

$$\max \sum_{i=2}^{N-1} \sum_{j=2}^{N} p_i x_{ij} \tag{1}$$

where x_{ij} is a binary variable, $x_{ij} = 1$ if POI i is visited from POI j, and $x_{ij} = 0$ otherwise.

The route must start at the origin p_1 and end at the destination p_N, with each location being visited only once. Meanwhile, the total travel time must not exceed the time budget $T_{\max}$. Additionally, route coherence must be ensured: if a POI i is visited, it must be reached from another node, and a next node must be visited from it. No sub-tours are allowed in the path, subject to the following constraints:

1. Start and end constraints:

$$\sum_{j=2}^{N} x_{1j} = 1, \quad \sum_{i=1}^{N-1} x_{iN} = 1 \tag{2}$$

2. Single visit constraint:

$$\sum_{i=1}^{N-1} x_{ij} \le 1, \quad j = 2, ..., N \tag{3}$$

3. Total time constraint:

$$\sum_{i=1}^{N-1} \sum_{j=2}^{N} t_{ij} x_{ij} \le T_{\max} \tag{4}$$

4. Route coherence constraint:

$$\sum_{i=1}^{N-1} x_{im} = \sum_{j=2}^{N} x_{mj}, \quad \forall m = 2, ..., N-1 \tag{5}$$

5. Sub-tour elimination constraint:

$$2 \le u_i \le N, \quad \forall i = 2, ..., N \tag{6}$$

$$u_i - u_j + 1 \le (N-1)(1 - x_{ij}), \quad \forall i, j = 2, ..., N \tag{7}$$

where t_{ij} denotes the travel time from POI i to POI j, $T_{\max}$ represents the maximum travel time budget, and u_i represents the position of node i in the path.

Through these formulas, the OP-based travel planning problem is defined as an optimization problem, with the objective of selecting appropriate POIs to maximize the score within the time budget and generating an optimal travel route. As the number of POIs increases, the complexity of this problem grows, making it an NP-hard problem. Therefore, in practical applications, heuristic algorithms are often used to solve the problem within a reasonable time frame, providing feasible route planning solutions.

Significant research has been conducted on the TTDP based on the OP. Wörndl et al. (2017) developed an algorithm for planning short city trips on foot, modeling the problem as an OP and solving it using a variant of Dijkstra's algorithm. Karbowska-Chilinska and Chociej (2019) applied a greedy heuristic to the multi-stage electric vehicle TTDP, replacing hotels with electric charging stations to address the specific needs of electric vehicles. Toledo et al. (2019) They proposed a hyper-heuristic algorithm for the OP with hotel selection, combining heuristics like insertion, 2-opt, and hotel improvements to improve solution quality. Abbaspour and Samadzadegan (2009) applied a Genetic Algorithm (GA) to solve the TTDP, emphasizing time constraints and multi-modal transportation.

Recently, researchers have shifted their focus to various OP variants. One such example is the Orienteering Problem with Time Windows (OPTW), where each location can only be visited within a designated time window, often corresponding to the attraction's opening hours. (Kantor and Rosenwein, 1992). The Time-Dependent Orienteering Problem (TDOP) assumes that the travel time between two locations varies depending on the departure time from the first location. (Fomin and Lingas, 2002). This extension is valuable for modeling itinerary recommendations, particularly when accounting for the impact of different modes of transportation. Fomin and Lingas (2002) provided a $(2+\epsilon)$-approximation algorithm to solve the TDOP problem. Gao et al. (2023) studied route planning in large-scale urban networks, focusing on time and utility variations. Vathis et al. (2023) combined multi-level clustering with dynamic programming to define and solve a geographically constrained travel planning problem (VPP).

However, traditional methods rely on rigid templates and struggle with unstructured natural language queries. They often provide generic solutions that don't account for personalized preferences and are cumbersome in extracting and structuring data. Travel planning requires dynamic POI management, real-time updates, and adaptability to unforeseen events, which static systems cannot handle. Additionally, conventional systems lack human-like cognitive abilities, making them inflexible in addressing complex, open-domain problems and diverse constraints such as time, budget, and accessibility.