Proceedings of CONF-MLA 2025 Symposium: Applied Artificial Intelligence Research
DOI: 10.54254/2755-2721/2025.BJ25286

Design and Improvement of a Large Language Model Travel Planning System

Zhongsheng Sun

Shanghai Experimental Foreign Language School, Shanghai, China
wuxie050817@gmail.com

Abstract. With the rapid advancement and widespread adoption of large language models (LLMs), an increasing number of individuals have incorporated these models into their daily routines. To address persistent challenges in travel planning, this paper proposes an LLM-powered specialized travel planning system. By processing concise yet essential user inputs, the system can generate comprehensive travel plans tailored to individual requirements. This study designs user input templates and evaluates the response quality of the large language models Deepseek and Kimi. The responses are assessed and compared using predefined metrics to establish a preliminary system framework and identify limitations, such as insufficient information input, the computational complexity of evaluation metrics, and inefficient and deficient information output. This paper provides potential solutions that include API (application programming interface) integration, enhanced user-system interaction, optimized prompt engineering, and implementation of advanced algorithms. This paper systematically reviews the system design, identifies prevailing challenges, and outlines future development trajectories. It validates the feasibility and potential of the Large Language Model Travel Planning System while establishing a comprehensive framework for LLM-based planning research, serving as a valuable reference for researchers in related fields.

Keywords: Large Language Model, travel planning, API, prompts for LLMs

1. Introduction

In recent years, LLMs have achieved rapid and remarkable advancements. These models now possess web search capabilities and sophisticated reasoning abilities, enabling them to access real-time online data and demonstrate exceptional logical deduction skills [1,2]. Consequently, LLMs have significantly enhanced their capacity for complex problem-solving, multi-source information integration, and real-time information enquiry, substantially improving the quality and reliability of their responses. This technological progress has led to growing public trust in LLMs and their frequent application in daily life. Simultaneously, travel planning remains a persistent challenge for many individuals, as it requires abundant time and effort to research unfamiliar destinations while considering multiple interrelated factors [3]. This paper proposes a system design that leverages LLMs to assist in creating comprehensive travel itineraries. Users simply need to input essential parameters including destination, departure location, departure time, travel duration, group

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.BJ25286

composition, and personal preferences. The system will then generate a complete travel plan consisting of detailed scheduling and information on transportation, hotel, and weather, among others [4]. In addition, the navigation and reservations of hotel or tourist attractions can be used through embedded API endpoints in itineraries.

This study designs a standardized information input template for users and evaluates qualities of responses across different LLMs. The generated itineraries will be assessed and compared using specifically designed evaluation metrics. Establishing a foundational framework for travel planning systems through the above-mentioned steps. However, our experiments revealed several limitations: input information insufficiency for LLMs, and challenges in objectively evaluating qualities of plans due to subjective user preferences.

To address these issues, we propose potential solutions including integration of Application Programming Interface (API) for enhanced data access, user feedback mechanisms for metric refinement, designing effective prompts, and adopting more suitable algorithms or processing frameworks. This paper systematically reviews the system architecture, current challenges, and future research directions, aiming to provide a comprehensive structural framework and valuable references for researchers working on LLM-based planning systems across related domains.

2. Concept of large language model travel planning system

2.1. Information input

The users input information on place of departure, destination, time of departure and duration, composition of travelers, and personal preference. Concurrently, large language models (LLMs) leverage internet connectivity to retrieve real-time information from online sources—a capability already demonstrated by existing LLM implementations. This user-defined input establishes planning guidelines, contextual constraints, and operational boundaries, enabling LLMs to perform targeted extraction of relevant data from digital networks with enhanced precision [5].

2.2. Information output

The information output must include four core elements: itinerary arrangement, resource management, personal preferences, and risk prediction. The itinerary arrangement encompasses time allocation (must be precise to the minute) and destination selection. Resource management represents the control of budgets and coordination of transportation, and transportation should include available flight options. Personal preferences enable tourism personalisation and better alignment with travelers' needs, which includes individual interests and special requirements. Personal interests may manifest as thematic preferences for journeys, such as cultural exploration, natural scenery, family-oriented interactions, or culinary experiences. Special requirements could involve demands for accessibility facilities or dietary restrictions based on religious beliefs. Risk prediction comprises contingency plans and safety advisories. Contingency planning entails developing backup solutions for emergencies such as sudden weather changes or flight delays, along with medical emergency information. Safety advisories include security reminders for destinations and health precautions. These components collectively establish a comprehensive framework that ensures systematic organization, personalisation, and risk mitigation throughout the tourism planning process.

DOI: 10.54254/2755-2721/2025.BJ25286

3. Advances in large language model travel planning system

3.1. Design of information input

First, design a basic information input sentence structure containing fundamental information. The information of the place of departure, destination, time of departure and duration, composition of travelers, and personal preference. For example :“ Hello, I want to travel to Chongqing next Tuesday, please make a detailed travel plan for me. We plan to travel for three days, starting from Xuhui District of Shanghai. I will go with my sister, my mother, my father and my father's colleague's family. I like to learn about the culture and history. We hope the pace of the journey will be relaxed and pleasant. ” Then, we can append additional conditions to this base sentence, such as: “ The requirements are as follows: 1. the daily itinerary should have a specific time arrangement. 2. please take into account the weather during the trip, the impact of the weather to adjust the travel plan. 3. I plan to go to Chongqing by plane, please calculate the flight from Shanghai to Chongqing next Tuesday morning. 4. choose a hotel near Jiefangbei for me, the hotel price should be around 500 yuan per night. 5. the itinerary includes restaurants in Chongqing that serve Chongqing special food. ”

3.2. Comparative analysis of different LLMs

The designed information input sentence structures are applied to test and compare the LLMs Deepseek and Kimi. The results are presented in table 1.

Table 1. Test Deepseek and Kimi with two designed information input sentence structures

	Deepseek	Kimi
Q1: Hello, I want to travel to Chongqing next Tuesday, please make a detailed travel plan for me. We plan to travel for three days, starting from Xuhui District of Shanghai. I will go with my sister, my mother, my father and my father's colleague's family. I like to learn about the culture and history. We hope the pace of the journey will be relaxed and pleasant.	The timing and transportation arrangements are not detailed enough.	Timing arrangements is not detailed enough.
Q2: Q1+The requirements are as follows: 1. the daily itinerary should have a specific time arrangement. 2. please take into account the weather during the trip, the impact of the weather to adjust the travel plan. 3. I plan to go to Chongqing by plane, please calculate the flight from Shanghai to Chongqing next Tuesday morning. 4. choose a hotel near Jiefangbei, hotel price should be around 500 yuan per night. 5. the itinerary includes restaurants in Chongqing that serve Chongqing special food.	Satisfy all requirements.	Restaurants recommended and information of weather is not in schedule. Flights are not adequate. Date is incorrect, Kimi doesn’t use calendar in 2025, it used calendar in 2024 to answer my question. This prevents it from determining dates based on weekday information.

To obtain additional evaluation data for Deepseek and Kimi, various tourist destinations were tested using the Q2 sentence structure. Completeness is a metric used to evaluate the thoroughness of travel plans generated by LLMs. Responses are categorized into seven components: date, flight, hotel, restaurant, tourist attraction, weather, and transportation. The completeness score is calculated

DOI: 10.54254/2755-2721/2025.BJ25286

as the proportion of correctly provided information components.The results are presented in table 2. Average completeness of planning from Deepseek is 85.7%, and average completeness of planning from Kimi is 54.28%. Overall, Deepseek has higher completeness than Kimi, so Deepseek demonstrates enhanced suitability over Kimi for Large Language Model Travel Planning System.

Table 2. Completeness of planning for different destinations generated by LLMs

	Beijing	Guanzhou	HongKong	Singapore	London
Deepseek	85.7%	85.7%	85.7%	85.7%	85.7%
Kimi	57.1%	42.9%	42.9%	57.1%	71.4%

While this completeness assessment remains rudimentary, large language models demand more sophisticated evaluation metrics encompassing performance, robustness, and alignment [1].

In addition, there are several shortages for information input and information output. The information input consists of user-provided information and information acquired by LLMs through the internet. Regarding user input, due to their unfamiliarity with the characteristics of destinations, users often exhibit hesitation in specifying personal preferences. To address this issue, LLMs should provide concise overviews of the destinations suggested by users to assist them in making more satisfactory choices, this improvement method that will be elaborated on in subsequent sections. As for the information acquired by LLMs from the internet, current LLMs can only obtain data from websites. In cases of insufficient information, LLMs may fabricate information to mislead users, this phenomenon observed in both Deepseek and Kimi, such as suggesting non-existent flights in travel itineraries. While current travel plans generated by LLMs still exhibit non-trivial processing latency and suboptimal performance in temporal scheduling and route optimization. These limitations can be addressed through the implementation of other frameworks, models, or algorithms.

3.3. Design of evaluation metrics

To enhance user understanding of complex travel planning, this study proposes four evaluation metrics: schedule intensity, budget range, content diversity and demographic suitability.

Schedule intensity refers to the compactness of travel itineraries, which varies across demographic segments. Younger travelers often prefer high schedule intensity characterized by multiple daily attractions, exhibitions, and recreational facilities with limited breaks. Conversely, older demographics typically favor low schedule intensity—featuring single daily attractions and performances—to accommodate physical stamina requirements and ensure adequate rest periods.

The Budget Range indicates the approximate expenditure range for the entire travel process. This metric is calculated by aggregating costs across key components: hotel accommodations, restaurant expenses, attraction fees, and flight tickets—all obtainable from official sources. Detailed transportation costs (e.g., taxis, and shared bicycles) are included through per-use calculations. Discretionary expenses such as tips, souvenirs and snacks are excluded due to their user-dependent variability and unpredictability.

Content Diversity reflects the breadth of domains and interests incorporated in travel itineraries. Low content diversity plans tend to be monotonous, focusing primarily on a small number of kinds of activities, like only visiting historical sites. In contrast, high content diversity itineraries integrate multifaceted experiences spanning historical sites, artistic engagements (concerts, art exhibitions), and urban exploration (commercial districts), thereby achieving balanced thematic coverage.

DOI: 10.54254/2755-2721/2025.BJ25286

The Demographic Suitability metric evaluates target user groups based on schedule intensity and activity types. High-intensity itineraries are recommended for adolescents and young adults, while those incorporating adult-oriented venues like bars or casinos are flagged as inappropriate for minors.

4. Conclusion

However, there are problems in the calculation methodology for evaluation metrics. Certain evaluation metrics, particularly Schedule Intensity and Content Diversity, primarily rely on subjective user judgments, which pose computational challenges for large language models. In addition, the budget range metric necessitates adequate information from online sources, yet current data acquisition remains insufficient.

References

[1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).

[2] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

[3] Chen, A., Ge, X., Fu, Z., Xiao, Y., & Chen, J. (2024). TravelAgent: An AI assistant for personalized travel planning. arXiv preprint arXiv: 2409.08069.

[4] Turnip, F. F., & Turnip, A. (2020, June). Development of Online Ticket Booking Application for Ferry Crossing Website Based in Toba Lake Area. In 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT) (pp. 381-385). IEEE.

[5] Shao, J. J., Yang, X. W., Zhang, B. W., Chen, B., Wei, W. D., Guo, L. Z., & Li, Y. F. (2024). ChinaTravel: A Real-World Benchmark for Language Agents in Chinese Travel Planning. arXiv preprint arXiv: 2412.13682.