

RETAIL: Towards Real-world Travel Planning for Large Language Models

Bin Deng^{1†}, Yizhe Feng^{1†}, Zeming Liu^{1‡}, Qing Wei²,
Xiangrong Zhu², Shuai Chen², Yuanfang Guo^{1‡}, Yunhong Wang¹

¹School of Computer Science and Engineering, Beihang University, Beijing, China,

²Meituan Inc., China,

db1124@buaa.edu.cn, sy2306328@buaa.edu.cn, zmliu@buaa.edu.cn

Abstract

Although large language models have enhanced automated travel planning abilities, current systems remain misaligned with real-world scenarios. First, they assume users provide explicit queries, while in reality requirements are often implicit. Second, existing solutions ignore diverse environmental factors and user preferences, limiting the feasibility of plans. Third, systems can only generate plans with basic POI arrangements, failing to provide all-in-one plans with rich details. To mitigate these challenges, we construct a novel dataset **RETAIL**, which supports decision-making for implicit queries while covering explicit queries, both with and without revision needs. It also enables environmental awareness to ensure plan feasibility under real-world scenarios, while incorporating detailed POI information for all-in-one travel plans. Furthermore, we propose a topic-guided multi-agent framework, termed TGMA. Our experiments reveal that even the strongest existing model achieves merely a 1.0% pass rate, indicating real-world travel planning remains extremely challenging. In contrast, TGMA demonstrates substantially improved performance 2.72%, offering promising directions for real-world travel planning.¹

1 Introduction

Early tourism planning systems focused on automating travel plan generation based on user queries (Ferraro and Re, 2014; Lucas et al., 2013), primarily simplifying information search and basic planning processes. Recent advances in large language models (Touvron et al., 2023; Jiang et al., 2023) have enabled more sophisticated approaches that consider user preferences and commonsense knowledge (Xie et al., 2024; Chen et al., 2024; Tang et al., 2024)

However, existing travel planning research remains misaligned with real-world complexities. First, current approaches (Xie et al., 2024; Singh et al., 2024) predominantly rely on the assumption that users have explicit intentions when formulating plans, which limits their practical utility. Second, these methods (Chen et al., 2024; Zhang et al., 2024b) ignore the variability of user needs and environmental conditions, thus lacking the ability to dynamically adjust plans. Third, while existing work focuses primarily on basic information (Xie et al., 2024; Singh et al., 2024), it overlooks critical details such as timing, images, ticket pricing, UGC-enriched room info that are essential for real-world practical plans. As shown on the left side of Figure 1, they typically require users to predefine precise details such as duration, budget, and POI preference. However, this does not align with reality due to the casual and imprecise nature of tourists’ speech. Moreover, they also fail to consider dynamic user needs and complex environmental conditions (e.g., rainy days). Additionally, their plans are oversimplified, only including restaurants, attractions, and accommodations, which ultimately leads to a lack of practical feasibility.

To mitigate these challenges, we constructed a **REal-world Tourism All-in-one Interactive pLanning** dataset, termed **RETAIL**. Specifically, it incorporates two complementary mechanisms to support travel planning. The decision-making support process guides users from basic travel requirements to specific preferences through clarification, acknowledging that initial user queries are often implicit and mitigating the first challenge. The plan revision and environmental awareness process can further adjust plans based on changes in user requirements or perceived environmental conditions, which mitigates the second challenge of lacking environmental awareness and plan revision capabilities. It also has all-in-one travel plans that integrate time arrangements, tickets, rooms, cost and

[†] Equal Contribution.

[‡] Corresponding Author: zmliu@buaa.edu.cn.

¹Code and data are available at <https://github.com/BUAA-IRIP-LLM/RETAIL>.

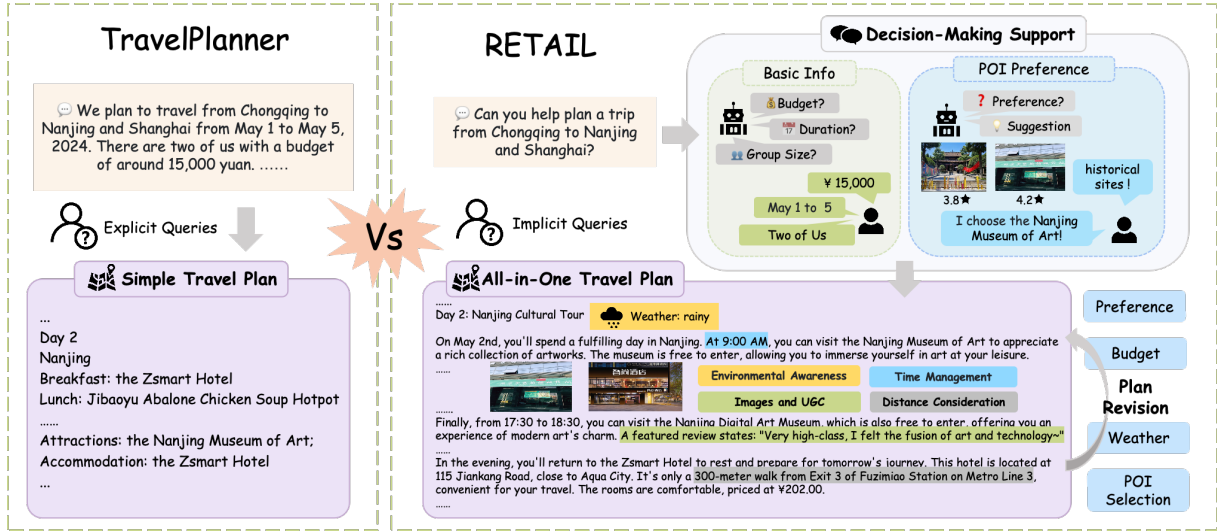


Figure 1: Comparison between TravelPlanner (Xie et al., 2024) and RETAIL, highlighting our UGC-enriched decision-making support for implicit queries, environmental awareness, and all-in-one travel plans with rich details.

etc, ensuring practical feasibility and mitigating the third challenge of lacking comprehensive travel planning. As shown on the right side of Figure 1, RETAIL starts from a implicit query, obtains users’ true intentions through a decision-making support process, and then generates all-in-one travel plans, while allowing plan revision in response to environmental changes.

To improve the ability of real-world travel plan generation, we propose TGMA, a topic-guided multi-agent framework. TGMA features a topic-guided approach and multi-agent collaboration. First, we design a Topic-Guided Interaction Logic that dynamically selects appropriate topics based on dialogue context, enabling natural and expectation-aligned interactions. Second, we develop a multi-agent architecture with three specialized agents: the intent detection agent for extracting accurate user requirements, the overall plan agent for generating structured preliminary plans and the detailed plan agent for producing comprehensive final plans.

This work makes the following contributions:

- **We identify challenges in real-world travel planning:** Existing works assume users provide explicit queries rather than implicit requirements. Travel plans need to adapt to dynamic environmental factors and user needs. All-in-one travel plans require integration of various practical details.
- **To mitigate these challenges, we present a novel dataset RETAIL,** covering plan-

ning from explicit queries and planning after clarifying implicit queries through decision-making support, both with and without revision needs, totaling 10,182 real-world cases. All plans are all-in-one travel itineraries with rich POI information, demonstrating high practical value.

- **We design TGMA, a topic-guided multi-agent framework for travel planning.** Experimental results demonstrate TGMA’s effectiveness in providing coherent decision support and generating all-in-one travel plans that integrate various practical aspects.

2 Related Work

2.1 Traditional Travel Planning

Traditional travel planning systems have evolved significantly over the past decades, from rule-based approaches to sophisticated recommendation engines (Ferraro and Re, 2014). Early systems rely on predefined rules and constraints for basic itinerary generation (Sebastia et al., 2008), later incorporating personalized recommendations (Lucas et al., 2013) and context-aware services (Rodriguez-Sánchez et al., 2013). Various optimization techniques have been developed, from multi-day planning (Chen et al., 2014) to approaches using clustering and deep learning (Huang et al., 2021), as well as weighted mining and ant colony optimization (Yu, 2022; Liang et al., 2021). While mobile and context-aware systems emerge as solutions towards more dynamic and personalized travel expe-

Dataset	Real-world Factors					Tourism KB				Query	Avg. Turns	Total
	DMS.	EA.	AOP.	PR.	DPI	Att.	Res.	Hot.	Wea.			
ITINERA	✗	✗	✗	✗	✗	7,578	0	0	0	Ex.&Im.	1	N/A
TravelPlanner	✗	✗	✗	✗	✗	5,302	9,551	5,047	0	Ex.	1	1,225
Ask-before-Plan	✗	✗	✗	✗	✗	5,302	9,551	5,047	0	Ex.&Im.	2.8	2,000
TravelAgent	✗	✗	✗	✗	✗	N/A	N/A	N/A	N/A	Ex.&Im.	1	20
ChinaTravel	✗	✗	✓	✗	✗	3,413	4,655	4,124	0	Ex.	1	154
RETAIL(ours)	✓	✓	✓	✓	✓	8,229	29,586	22,464	8,784	Ex.&Im.	4.65	10,182

Table 1: Comparison between RETAIL to other datasets, including ITINERA(Tang et al., 2024), TravelPlanner(Xie et al., 2024), Ask-before-Plan(Zhang et al., 2024b), TravelAgent(Chen et al., 2024) and ChinaTravel(Shao et al., 2024). **DMS.** for **UGC-enriched Decision-Making Support**, **EA.** for **Environmental Awareness**, **PR.** for **Plan Revision**, **DPI** stands for **Detailed POI Information, including Images and User-Generated Content like reviews, ratings, and must-visit rankings**, **AOP** for **All-in-One Plan**. **Att., Res., Hot., Wea.** stands for **Attraction, Restaurant, Hotel, Weather**. **Ex., Im.** stands for **Explicit, Implicit**. **N/A** denotes **Not Applicable**, which indicates that no data is provided. "✗" indicates the absence of a dimension, while "✗" indicates that only part of it is included; for example, **Ask-before-Plan** only includes **basic inquiries** without incorporating UGC-enriched recommendations. **Travel Agent** and **ChinaTravel** only incorporate **images** without leveraging UGC content.

periences, the inherent complexity of tourism info and the NP-hard nature of trip design (Gavalas et al., 2014) remain significant challenges.

These systems primarily formulate travel planning as mathematical optimization problems, excelling at computational scheduling. However, they lack decision-making support to help users clarify implicit requirements.

2.2 LLM-based Travel Planning

The emergence of LLMs has revolutionized travel planning through their enhanced capabilities in reasoning (Kojima et al., 2022; Wei et al., 2022), tool use (Schick et al., 2023; Qin et al., 2024), and agent frameworks (Shinn et al., 2023; Yao et al., 2023).

Recent research has explored LLM-based travel planning from multiple perspectives. TravelPlanner (Xie et al., 2024) establishes foundational planning capabilities, further enhanced by TravelPlanner+ (Singh et al., 2024) and ChinaTravel (Shao et al., 2024) for personalization and evaluation. Advanced planning approaches emerge through ItiNera (Tang et al., 2024) and TravelAgent (Chen et al., 2024) for spatial optimization, while TRIPAL (de la Rosa et al., 2024) and TTG (JU et al., 2024) focus on plan feasibility. For interactive planning, APEC-Travel (Jiang et al., 2024) and CEP (Zhang et al., 2024b) explore preference extraction through dialogues, while EVOAGENT (Yuan et al., 2024) employs multi-agent collaboration and human-like reasoning (Xie and Zou, 2024) introduces structured planning strategies. Beyond itinerary generation, LLMs have also

demonstrated potential in mobility analysis in transportation (Zhang et al., 2024c) and climate analysis (Zhang et al., 2024a), providing foundational support for environmental awareness in travel planning.

Existing LLM-based approaches focus on simple trip generation and basic requirement clarification. However, they do not achieve UGC-enriched decision-making support, systematic environmental awareness, or all-in-one travel planning driven by rich POI information that our work enables.

3 RETAIL

To mitigate the limitations in handling implicit requirements, dynamic environmental factors, and all-in-one travel planning, we present RETAIL dataset for real-world travel planning. It supports both direct planning from explicit requirements and clarification through decision-making support, with dynamic plan revision capabilities. A detailed comparison with existing datasets is presented in Table 1.

RETAIL is constructed through three phases: (1) Tourism Knowledge Base Collection; (2) Dataset Construction; and (3) Dataset Annotation.

3.1 Tourism Knowledge Base Collection

To support travel plans with rich POI information, we establish a comprehensive tourism database spanning 24 major Chinese cities, integrating 60,279 POIs with essential real-world planning factors such as detailed POI information, transportation data, and weather data. Detailed examples of

these information can be found in Appendix B.

City Selection. To reflect real-world tourism patterns and visitor flows across China, we select 24 major Chinese cities that represent diverse travel preferences. These destinations include both metropolitan centers (e.g., Beijing, Shanghai, Guangzhou) and cultural hubs (e.g., Chengdu, Xi'an, Xiamen).

POIs Information Collection. To support all-in-one travel plans, we gather nearby POI information around attractions to ensure distance constraints. Each attraction is linked to 3-5 nearby restaurants and hotels, incorporating essential planning factors: attraction information (categories, tickets, reviews), dining options (costs, ratings), and accommodation choices (locations, room types, guest feedback). POIs data were obtained through web crawling, ensuring comprehensive and up-to-date local business information.

Weather and Transportation Collection. To enable environmental awareness, we incorporate weather forecasts and transportation data for travel planning under diverse conditions. The weather data was collected from the **2345 Weather Service Platform**² covering full-year 2024 forecasts, while transportation data were obtained through web crawling from web. These dynamic factors directly impact travel decisions and reflect practical considerations in travel planning.

3.2 Dataset Construction

The Dataset construction consists of three processes: (1) Query Construction, (2) Decision-Making Support, and (3) Plan Generation and Adjustments.

3.2.1 Query Construction

We categorize queries into explicit and implicit types to simulate real-world travel planning scenarios. **For explicit queries**, we select one departure city and 2-4 destination cities, identify 8-10 key attractions with dining and accommodation options, and incorporate weather and transportation information through LLM to construct comprehensive, information-rich queries. **For implicit queries**, based on explicit queries, we develop a 12-field intention slot and strategically remove selected fields to mirror real-world tourist inquiries where preferences are often partially expressed.

3.2.2 Decision-Making Support

For implicit queries commonly encountered in travel planning, we implement a decision-making support mechanism that mirrors natural human consultation processes. Using GPT-4o (OpenAI, 2024), we simulate tourist-agent interactions to reflect how users gradually articulate their travel needs and how agents guide the decision-making process.

Basic Information Clarification. To systematically clarify basic travel requirements, the assistant combines systematic information gathering of essential planning elements (location, target cities, dates, and etc.) with natural conversation management strategies, limiting questions to 1-2 key points per round while employing colloquial language and context-adaptive dialogue progression.

POI Clarification. To address the diverse ways real tourists express their preferences, we implement comprehensive clarification logics considering multiple real-world factors. These logics simulate real-world travel planning scenarios by considering real-world factors like personal interests, weather conditions, POI's feedback. The assistant also provides reviews, ratings and high-quality images, enabling UGC-enriched decision-making through authentic dialogue interactions.

3.2.3 Plan Generation and Revision

To simulate the complexity and adaptability of real-world travel planning, our dataset encompasses both direct generation for explicit queries and dialogue-based planning for implicit queries. To capture the iterative nature of travel planning, we incorporate diverse plan revision cases spanning multiple aspects (dining, transportation, budget, weather-dependent activities). Detailed implementation specifications are elaborated in Section 4.

3.3 Dataset Annotations

To ensure data integrity, we implement rigorous quality control measures across three key aspects: Tourism Knowledge Base Validation, Decision-Making Support Refinement, and Plan Generation and Revision Verification. Detailed information can be found in Appendix C.

Our annotation team comprised 5 professionals (aged 25–40; 3 males, 2 females) with diverse expertise: two held master's degrees in linguistics, two held bachelor's degrees in computer science, and one specialized in psychology. All annotators

²<https://tianqi.2345.com/>

Dataset Statistics					
Dialogue Level		Utterance Level		Token Level	
Total dialogues	10,182	Total utterances	94,668	Total tokens	34,764,374
- Training	6,000	Utterances per dialogue:		Tokens per utterance:	
- Validation	2,182	- Maximum	30	- Maximum	2,569
- Test	2,000	- Average	9.3	- Average	367.22
Multi-turn	4,433	- Minimum	2	- Minimum	3
Single-turn	5,749				

Table 2: Statistics of RETAIL at dialogue, utterance, and token levels.

Expert	Basic Information	Clarification Logic	Presentation Style	Avg. Score
Expert 1	9	8	10	9.00
Expert 2	9	7	9	8.33
Expert 3	8	8	8	8.00
Expert 4	9	10	6	8.33
Expert 5	8	7	9	8.00
Mean	8.6	8.0	8.4	8.33
Std. Dev.	0.55	1.22	1.52	0.47

Table 3: Diversity Assessment by Domain Experts (n=5)

received competitive compensation aligned with industry standards and their respective qualifications, ensuring fairness and motivation.

Tourism Knowledge Base Validation. To establish a robust knowledge base, we conduct comprehensive data cleaning by removing entries with missing key fields (POI names, locations, operating hours), implementing random replacement algorithms for sensitive information (contact details, user identifiers), and validating transportation data through manual field investigations of actual ticket prices. We supplement inter-city transportation data with optimized transfer routes where direct connections are unavailable, ensuring complete coverage of travel options.

Decision-Making Support Refinement. To ensure authentic travel planning interactions, we evaluate dialogues through multiple quality dimensions. We filter conversations with misaligned recommendations, ignored user needs, redundant suggestions, unnatural transitions, mechanical reasoning, inconsistent styles, and insufficient clarifications. Through prompt optimization and dialogue flow enhancement, we refine the dialogue to better mirror natural conversation patterns.

Plan Generation and Revision Verification. To ensure feasibility of travel plans, we filter plans with unreasonable schedules, impractical accommodations, unrealistic transportation, and conflict-

ing activities. Plan revisions balance attraction and transportation adjustments while maintaining feasibility and user preferences, ensuring diverse practical travel plans.

Dataset Quality. Following Shi et al. (2023) and Liu et al. (2020), we conduct human evaluation on 1000 randomly sampled dialogues, where RETAIL achieves an average quality score of 0.86, demonstrating its high quality.

3.4 Dataset Analysis

Table 2 presents the statistics of RETAIL, which contains 10,182 dialogues with a balanced distribution between single-turn (5,749) and multi-turn (4,433) interactions, partitioned into training (6,000), validation (2,182), and test (2,000) sets. The dataset comprises 94,668 utterances and over 34 million tokens in total, representing substantial coverage of travel planning scenarios. For diversity evaluation, we randomly selected 500 samples and 5 experts to assess diversity across three dimensions: basic information, clarification logic, and presentation style, using a rating scale from 0 to 10 (higher scores indicate greater diversity). As shown in the Table 3, the lowest average score reached 8, indicating strong diversity across all aspects. Detailed statistics and examples of RETAIL can be found in Appendix C.

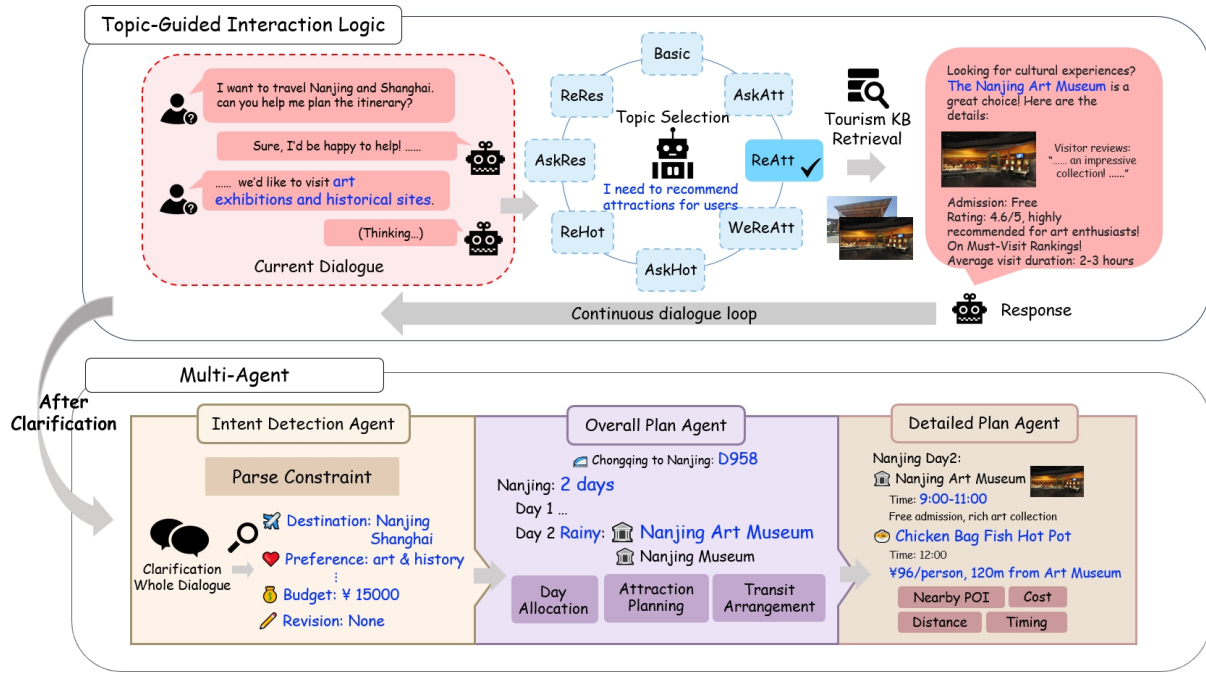


Figure 2: Overview of TGMA framework: the Topic-Guided Interaction Logic handles user queries through topic selection and knowledge retrieval (top), while the Multi-Agent progresses from parsing user requirements to generating high-level plans (including transportation, durations, and attractions), ultimately producing an all-in-one travel plan with specific timing, locations, and other rich details (bottom).

4 TGMA

To mitigate the challenges of real-world travel planning, we propose TGMA framework(as shown in Figure 2), which combines Topic-Guided interaction with Multi-Agent planning. The Topic-Guided Interaction Logic selects appropriate topics based on dialogue context ensuring natural dialogue transitions aligned with user preferences, while the Multi-Agent architecture decomposes complex planning tasks into manageable steps for quality assurance.

4.1 Topic-Guided Interaction Logic

To address the challenge of maintaining coherent multi-turn dialogues in travel consultation, we propose a Topic-Guided Interaction Logic framework. It implements a structured dialogue management approach through an LLM-based topic selection mechanism, where eight interconnected states dynamically guide conversation flow based on dialogue context and user preferences. Our framework integrates dialogue context tracking with a comprehensive tourism knowledge base covering five essential domains (Attraction, Restaurant, Hotel, Transportation, and Weather). This design enables context-aware response generation while maintaining dialogue coherence, allowing the system to

adapt to diverse user requirements through systematic topic transitions and knowledge integration.

4.2 Multi-Agent

To decompose complex travel planning into manageable specialized tasks, we design a multi-agent architecture that orchestrates three agents. This architecture enables progressive refinement of travel plans, from initial requirement analysis to detailed execution planning.

The Intent Detection Agent serves as the foundation by transforming unstructured conversational dialogues into structured representations. Through semantic parsing, it extracts and normalizes key travel parameters including location/time constraints, traveler details, accommodation preferences, and budget specifications, while also capturing users’ modifications and refinements throughout the conversation. This maintains dialogue context and provides standardized inputs for subsequent planning stages.

The Overall Plan Agent leverages the ReAct(Yao et al., 2023) framework to coordinate high-level planning decisions. It systematically processes Day Allocation, Attraction Planning, and Transit Arrangement through an integrated toolset for plan generation, analysis, and information re-

Model	BLEU				METEOR	ROUGE-L	CIDEr
	B-1	B-2	B-3	B-4			
Baselines							
Qwen2.5-7B-Instruct	8.29	3.23	1.44	0.79	17.26	15.34	0.343
Meta-Llama-3-8B-Instruct	1.24	0.45	0.21	0.13	11.21	13.54	0.329
glm-4-9b-chat	7.54	3.90	1.94	1.18	17.09	15.14	0.348
Baichuan2-13B-Chat	9.49	4.41	1.76	0.96	14.74	15.49	0.345
DeepSeek-V3(671B)	1.13	0.69	0.36	0.24	13.19	13.69	0.333
Doubao-Pro	1.77	0.26	0.10	0.00	11.58	13.60	0.328
Baichuan3-Turbo	2.81	1.44	0.79	0.52	16.43	14.21	0.337
Qwen-Plus	7.33	3.99	2.08	1.32	16.81	15.07	0.346
gpt-4-Turbo	4.91	2.67	1.52	1.03	14.18	15.10	0.347
gpt-4o	7.39	3.34	1.63	0.99	15.40	15.31	0.342
Topic-Guided Interaction Logic(ours)							
DeepSeek-V3(671B)	13.43	6.86	4.16	2.91	18.40	15.98	0.354
Qwen2.5-7B-Instruct(SFT)	13.88	6.49	3.32	2.04	21.44	16.33	0.360
gpt-4o	10.96	4.50	2.29	1.36	22.05	16.85	0.343

Table 4: Decision-Making Support Evaluation. ROUGE, BLEU, and METEOR metrics are expressed as percentages (%). Gray and yellow backgrounds denote open-source and closed-source models.

Model	Commonsense Constraint				User Preference Constraint				Final Pass Rate	
	Micro		Macro		Micro		Macro			
	PR.	F1	PR.	F1	PR.	F1	PR.	F1		
<i>Baselines</i>										
Qwen2.5-7B-Instruct	49.38	0.50	0.00	0.32	74.49	0.74	22.87	0.42	0.00	
Meta-Llama-3-8B-Instruct	64.60	0.65	0.00	0.38	68.04	0.68	17.42	0.40	0.00	
glm-4-9b-chat	54.13	0.54	0.17	0.35	77.32	0.77	29.69	0.43	0.10	
Baichuan2-13B-Chat	50.00	0.50	0.00	0.33	60.42	0.75	0.00	0.43	0.00	
DeepSeek-V3(671B)	70.55	0.74	2.92	0.42	79.73	0.80	35.27	0.44	0.81	
Doubao-Pro	62.30	0.62	0.00	0.38	75.94	0.76	26.32	0.43	0.00	
Baichuan3-Turbo	70.72	0.75	0.08	0.39	65.41	0.65	9.62	0.39	0.00	
Qwen-Plus	75.17	0.79	3.84	0.44	79.95	0.79	35.91	0.44	1.00	
gpt-4-Turbo	68.43	0.64	0.35	0.39	77.45	0.79	30.00	0.44	0.20	
gpt-4o	71.84	0.73	0.36	0.42	78.55	0.79	32.90	0.44	0.10	
<i>TGMA(ours)</i>										
DeepSeek-V3(671B)	79.64	0.78	5.73	0.43	82.09	0.84	39.49	0.45	1.38	
Qwen2.5-7B-Instruct(SFT)	81.74	0.77	5.89	0.45	86.40	0.84	42.32	0.46	2.72	
gpt-4o	78.65	0.79	3.49	0.44	80.30	0.81	37.30	0.45	1.07	

Table 5: Travel Plan Evaluation. Pass Rate (PR.) metrics are reported in percentages (%). Gray and yellow backgrounds denote open-source and closed-source models.

trieval. This structured approach ensures that generated plans effectively balance planning objectives with operational constraints, delivering coherent

and executable travel recommendations.

The Detailed Plan Agent finalizes the planning process by transforming high-level outlines into

comprehensive, executable plans. It systematically enriches travel outlines with nearby POI information and practical details, including attraction tickets, accommodation options, and restaurant recommendations. Through structured scheduling and point-of-interest coordination, the agent generates optimized daily arrangements that balance travel efficiency with user comfort while considering practical limitations.

5 Experiment

5.1 Baselines

Following Xie et al. (2024) and Shao et al. (2024), we carefully select a few strong baselines for comparison, including GPT-4o (OpenAI, 2024), GPT-4-Turbo (OpenAI, 2023), DeepSeek-V3 (DeepSeek-AI et al., 2024), and Llama-3 (Room, 2024), as well as Chinese models like Baichuan (Lin et al., 2024), GLM (GLM et al., 2024), and Qwen (Yang et al., 2024).

5.2 Metrics

We assess the system from two aspects: decision-making process’s dialogue capabilities and planning performance. **For dialogue capabilities**, following Shi et al. (2023) and Liu et al. (2022), we employ standard dialogue quality metrics including BLEU-n (Papineni et al., 2002), ROUGE (Lin, 2004) and METEOR (Lavie and Agarwal, 2007) to assess the response generation. **For planning performance**, following Xie et al. (2024), we evaluate plans using commonsense constraint and user preference constraint pass rate, and final pass rate. Detailed definition can be found in Appendix A.1.

5.3 Decision-Making Support Evaluation

The evaluation results in Table 4 demonstrate the effectiveness of our Topic-Guided Interaction Logic framework. When integrated with different LLMs, our framework consistently outperforms their baseline across all metrics. Specifically, with DeepSeek-V3, our framework achieves substantial improvements in BLEU scores, while with GPT-4, it shows significant gains in METEOR (22.05 vs 15.40) and ROUGE-L (16.85 vs 15.31). These improvements indicate that our structured dialogue management approach effectively enhances model performance in travel-specific conversations. Detailed experimental information can be found in Appendix A.3.

5.4 Travel Plan Evaluation

Table 5 demonstrates the superior planning capabilities of TGMA in real world. When integrated with Qwen2.5-7B-Instruct(SFT), our approach achieves the highest final pass rate of 2.72%, significantly outperforming baseline models across both commonsense and user preference constraints. The notable performance gap between our framework and baseline models highlights the effectiveness of our structured planning approach in handling real-world travel planning scenarios. Detailed experimental information can be found in Appendix A.4.

6 Analysis

6.1 RQ1: How Do Different Constraints Impact Travel Plan Final Pass Rate?

To study **how different constraints affect the final pass rate of travel plans** which is essential for system optimization, we use the travel plans to calculate the Pearson correlation coefficients between 14 types of constraints (9 commonsense constraints and 4 user preference constraints) shown in Table 7. Further analysis can be found in A.2.

Figure 3(a) shows Time Interval and POI Validation are the main constraints affecting pass rates, at 0.495 and 0.432 respectively. This suggests plans need better activity scheduling and POI verification. City Coverage at 0.003 has minimal impact, indicating most plans satisfy city coverage requirements.

Figure 3(b) reveals Hotel Type and Required Sites as key preference constraints at 0.733 and 0.696, while Budget and Excluded Sites show lower impact at 0.146 and 0.107. This indicates plans mainly struggle with meeting hotel type and must-visit site requirements.

6.2 RQ2: How Important is Decision-Making Support in Travel Planning?

To quantitatively assess the impact of decision-making support (DMS), we employed the same evaluation metrics outlined in Travel Plan Evaluation (Section 5.4) to compare plans generated with and without DMS. The results, presented in Table 6, reveal significant performance disparities across all metrics. Specifically, plans generated without DMS achieved only half the micro and macro pass rates for commonsense constraints (CC) compared to those with DMS. The gap was even more pronounced for user preference constraints

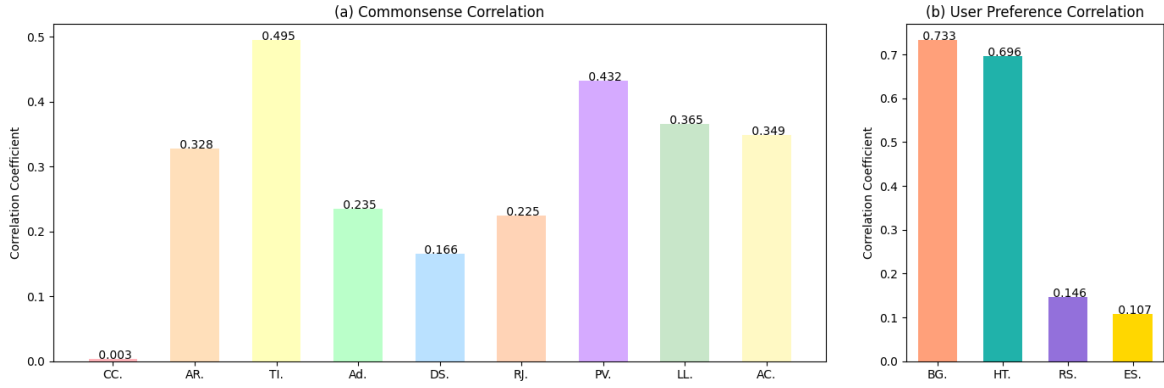


Figure 3: Impact of Different Constraints on Pass Rates. (a) CC: City Coverage, AR: Activity Repetition, TI: Time Interval, Ad: Accommodation, DS: Daily Schedule, RJ: Return Journey, PV: POI Validation, LL: Location Logic, AC: Activity Count. (b) Bg: Budget, HT: Hotel Type, RS: Required Sites, ES: Excluded Sites.

(UPC), where unsupported plans exhibited substantially lower performance in both micro and macro evaluations. Without DMS, models tended to produce generic travel itineraries that failed to capture individual preferences and requirements, resulting in standardized outputs lacking personalization. This limitation led to a zero final pass rate, underscoring that decision-making support is crucial for understanding user-specific needs and creating personalized travel experiences.

DMS	CC		UPC		Final PR (%)
	Micro	Macro	Micro	Macro	
✗	34.19	2.60	31.49	1.17	0.00
✓	78.65	3.49	80.30	37.30	1.07

Table 6: Travel plan evaluation on CC (Commonsense Constraints) and UPC (User Preference Constraints). DMS: Decision-Making Support; PR: Pass Rate.

7 Conclusion

This work identifies three critical limitations in real-world travel planning: insufficient decision-making support for implicit queries, the lack of environmental awareness, and the inability to generate all-in-one plans with rich details. To mitigate these challenges, we first introduce RETAIL covering both explicit and decision-making support for implicit queries. It also features environmental awareness, and rich POI information enabling all-in-one travel plans. Based on RETAIL, we propose TGMA, a topic-guided multi-agent framework for providing coherent and user-aligned decision support and generating all-in-one travel plans that integrate various practical aspects. Through extensive

experiments, we demonstrate TGMA’s effectiveness in real-world travel planning scenarios. We hope our work provides valuable insights and resources for advancing real-world travel planning systems.

Acknowledgments

Thanks for the insightful comments and feedback from the reviewers. This work was supported by the National Natural Science Foundation of China (No. 62406015), and Meituan.

Limitations

RETAIL requires substantial computational resources and poses significant challenges to models’ context processing capabilities, particularly in managing long dialogue contexts and mitigating hallucinations in travel planning. Future work will explore Progressive Generation(Xu et al., 2024) to mitigate these computational constraints and enhance context retention in extended sequences.

Ethics Statement

We make sure that RETAIL is collected in a manner that is consistent with the terms of use of any sources and the intellectual property and privacy rights of the original authors of the texts. All participants provided informed consent and were fairly compensated. Details of data collection procedures are described in Section 3.

References

Aili Chen, Xuyang Ge, Ziquan Fu, Yanghua Xiao, and Jiangjie Chen. 2024. *Travelagent: An AI as-*

- sistant for personalized travel planning. *CoRR*, abs/2409.08069.
- Gang Chen, Sai Wu, Jingbo Zhou, and Anthony K. H. Tung. 2014. [Automatic itinerary planning for traveling services](#). *IEEE Trans. Knowl. Data Eng.*, 26(3):514–527.
- Tomás 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](#). *CoRR*, abs/2406.10196.
- DeepSeek-AI, Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Han Bao, Hanwei Xu, Haocheng Wang, Haowei Zhang, Honghui Ding, Huajian Xin, Huazuo Gao, Hui Li, Hui Qu, J. L. Cai, Jian Liang, Jianzhong Guo, Jiaqi Ni, Jiashi Li, Jiawei Wang, Jin Chen, Jingchang Chen, Jingyang Yuan, Junjie Qiu, Junlong Li, Junxiao Song, Kai Dong, Kai Hu, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Lei Xu, Leyi Xia, Liang Zhao, Litong Wang, Liyue Zhang, Meng Li, Miaojuan Wang, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Mingming Li, Ning Tian, Panpan Huang, Peiyi Wang, Peng Zhang, Qiancheng Wang, Qihao Zhu, Qinyu Chen, Qiusi Du, R. J. Chen, R. L. Jin, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, Runxin Xu, Ruoyu Zhang, Ruyi Chen, S. S. Li, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shaoqing Wu, Shengfeng Ye, Shengfeng Ye, Shirong Ma, Shiyu Wang, Shuang Zhou, Shuiping Yu, Shunfeng Zhou, Shutong Pan, T. Wang, Tao Yun, Tian Pei, Tianyu Sun, W. L. Xiao, and Wangding Zeng. 2024. [Deepseek-v3 technical report](#). *CoRR*, abs/2412.19437.
- Pierluca Ferraro and Giuseppe Lo Re. 2014. [Designing ontology-driven recommender systems for tourism](#). In Salvatore Gaglio and Giuseppe Lo Re, editors, *Advances onto the Internet of Things - How Ontologies Make the Internet of Things Meaningful*, volume 260 of *Advances in Intelligent Systems and Computing*, pages 339–352. Springer.
- Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, and Grammati E. Pantziou. 2014. [A survey on algorithmic approaches for solving tourist trip design problems](#). *J. Heuristics*, 20(3):291–328.
- Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, et al. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*.
- Feiran Huang, Jie Xu, and Jian Weng. 2021. [Multi-task travel route planning with a flexible deep learning framework](#). *IEEE Trans. Intell. Transp. Syst.*, 22(7):3907–3918.
- Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, L  lio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timoth  e Lacroix, and William El Sayed. 2023. [Mistral 7b](#). *CoRR*, abs/2310.06825.
- Song Jiang, Da JU, Andrew Cohen, Sasha Mitts, Aaron Foss, Justine T. Kao, Xian Li, and Yuandong Tian. 2024. [Towards full delegation: Designing ideal agentic behaviors for travel planning](#). *CoRR*, abs/2411.13904.
- Da JU, Song Jiang, Andrew Cohen, Aaron Foss, Sasha Mitts, Arman Zharmagambetov, Brandon Amos, Xian Li, Justine T. Kao, Maryam Fazel-Zarandi, and Yuandong Tian. 2024. [To the globe \(TTG\): towards language-driven guaranteed travel planning](#). *CoRR*, abs/2410.16456.
- Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. [Large language models are zero-shot reasoners](#). In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.
- Alon Lavie and Abhaya Agarwal. 2007. [METEOR: an automatic metric for MT evaluation with high levels of correlation with human judgments](#). In *Proceedings of the Second Workshop on Statistical Machine Translation, WMT@ACL 2007, Prague, Czech Republic, June 23, 2007*, pages 228–231. Association for Computational Linguistics.
- Shengbin Liang, Tongtong Jiao, Wencai Du, and Shengming Qu. 2021. An improved ant colony optimization algorithm based on context for tourism route planning. *PLoS One*, 16(9):e0257317.
- Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.
- Mingan Lin, Fan Yang, Yanjun Shen, Haoze Sun, Tianpeng Li, Tao Zhang, Chenzheng Zhu, Tao Zhang, Miao Zheng, Xu Li, Yijie Zhou, Mingyang Chen, Yanzhao Qin, Youquan Li, Hao Liang, Fei Li, Yadong Li, Mang Wang, Guosheng Dong, Kun Fang, Jianhua Xu, Bin Cui, Wentao Zhang, Zenan Zhou, and Weipeng Chen. 2024. [Baichuan alignment technical report](#). *CoRR*, abs/2410.14940.
- Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020. [Towards conversational recommendation over multi-type dialogs](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1036–1049, Online. Association for Computational Linguistics.

- Zeming Liu, Jun Xu, Zeyang Lei, Haifeng Wang, Zheng-Yu Niu, and Hua Wu. 2022. [Where to go for the holidays: Towards mixed-type dialogs for clarification of user goals](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ACL 2022, Dublin, Ireland, May 22-27, 2022, pages 1024–1034. Association for Computational Linguistics.
- Joel Pinho Lucas, Nuno Luz, María N. Moreno García, Ricardo Anacleto, Ana Maria de Almeida Figueiredo, and Constantino Martins. 2013. [A hybrid recommendation approach for a tourism system](#). *Expert Syst. Appl.*, 40(9):3532–3550.
- OpenAI. 2023. [GPT-4 technical report](#). *CoRR*, abs/2303.08774.
- OpenAI. 2024. Hello gpt-4o. <https://openai.com/index/hello-gpt-4o/>. Accessed: 2025-01-15.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a method for automatic evaluation of machine translation](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA*, pages 311–318. ACL.
- Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, Sihan Zhao, Lauren Hong, Runchu Tian, Ruobing Xie, Jie Zhou, Mark Gerstein, Dahai Li, Zhiyuan Liu, and Maosong Sun. 2024. [Toolllm: Facilitating large language models to master 16000+ real-world apis](#). In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.
- M. C. Rodriguez-Sánchez, Juan Martínez-Romo, Susana Borromeo, and Juan Antonio Hernández Tamames. 2013. [GAT: platform for automatic context-aware mobile services for m-tourism](#). *Expert Syst. Appl.*, 40(10):4154–4163.
- Chat Room. 2024. Llama (llm). *algorithms*, 5(12):40.
- Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. 2023. [Toolformer: Language models can teach themselves to use tools](#). In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.
- Laura Sebastia, Inma Garcia, Eva Onaindia, and Cesar Guzman. 2008. [e-tourism: A tourist recommendation and planning application](#). In *20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2008), November 3-5, 2008, Dayton, Ohio, USA, Volume 2*, pages 89–96. IEEE Computer Society.
- Jie-Jing Shao, Xiao-Wen Yang, Bo-Wen Zhang, Baizhi Chen, Wen-Da Wei, Guohao Cai, Zhenhua Dong, Lan-Zhe Guo, and Yu-feng Li. 2024. [Chinatravel: A real-world benchmark for language agents in chinese travel planning](#). *CoRR*, abs/2412.13682.
- Xiaoming Shi, Zeming Liu, Chuan Wang, Haitao Leng, Kui Xue, Xiaofan Zhang, and Shaoting Zhang. 2023. [Midmed: Towards mixed-type dialogues for medical consultation](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ACL 2023, Toronto, Canada, July 9-14, 2023, pages 8145–8157. Association for Computational Linguistics.
- Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. [Reflexion: language agents with verbal reinforcement learning](#). In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.
- Harmanpreet Singh, Nikhil Verma, Yixiao Wang, Manasa Bharadwaj, Homa Fashandi, Kevin Ferreira, and Chul Lee. 2024. [Personal large language model agents: A case study on tailored travel planning](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: EMNLP 2024 - Industry Track, Miami, Florida, USA, November 12-16, 2024*, pages 486–514. Association for Computational Linguistics.
- Yihong Tang, Zhaokai Wang, Ao Qu, Yihao Yan, Zhaofeng Wu, Dingyi Zhuang, Jushi Kai, Kebing Hou, Xiaotong Guo, Jinhua Zhao, Zhan Zhao, and Wei Ma. 2024. [Itinera: Integrating spatial optimization with large language models for open-domain urban itinerary planning](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: EMNLP 2024 - Industry Track, Miami, Florida, USA, November 12-16, 2024*, pages 1413–1432. Association for Computational Linguistics.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. [Llama: Open and efficient foundation language models](#). *CoRR*, abs/2302.13971.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. [Chain-of-thought prompting elicits reasoning in large language models](#). In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.
- Chengxing Xie and Difan Zou. 2024. [A human-like reasoning framework for multi-phases planning task with large language models](#). *CoRR*, abs/2405.18208.
- 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. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net.

Weiyue Xu, Min Wang, Wengang Zhou, and Houqiang Li. 2024. P-RAG: progressive retrieval augmented generation for planning on embodied everyday task. In *Proceedings of the 32nd ACM International Conference on Multimedia, MM 2024, Melbourne, VIC, Australia, 28 October 2024 - 1 November 2024*, pages 6969–6978. ACM.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. 2024. Qwen2.5 technical report. *CoRR*, abs/2412.15115.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R. Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.

Yang Yu. 2022. Analysis and study on intelligent tourism route planning scheme based on weighted mining algorithm. *Scientific Programming*, 2022(1):5495822.

Siyu Yuan, Kaitao Song, Jiangjie Chen, Xu Tan, Dongsheng Li, and Deqing Yang. 2024. Evoagent: Towards automatic multi-agent generation via evolutionary algorithms. *CoRR*, abs/2406.14228.

Xiyuan Zhang, Ranak Roy Chowdhury, Rajesh K Gupta, and Jingbo Shang. 2024a. Large language models for time series: A survey. *arXiv preprint arXiv:2402.01801*.

Xuan Zhang, Yang Deng, Zifeng Ren, See-Kiong Ng, and Tat-Seng Chua. 2024b. Ask-before-plan: Proactive language agents for real-world planning. In *Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024*, pages 10836–10863. Association for Computational Linguistics.

Zijian Zhang, Yujie Sun, Zepu Wang, Yuqi Nie, Xiaobo Ma, Peng Sun, and Ruolin Li. 2024c. Large language models for mobility in transportation systems: A survey on forecasting tasks. *arXiv preprint arXiv:2405.02357*.

Appendix

A Additional Experiment

A.1 Detailed Evaluation Metrics

To ensure the quality and practicality of generated travel itineraries, we establish a comprehensive evaluation framework comprising two main categories of metrics, as detailed in Table 7. The first category, Commonsense Constraints, consists of ten fundamental metrics that verify the logical coherence and feasibility of travel plans. These metrics range from basic requirements such as city coverage and activity repetition checks, to more sophisticated validations like temporal conflict detection and location logic verification. The second category focuses on User Preferences, incorporating four key metrics to assess how well the generated itineraries align with specific user requirements, including budget constraints and accommodation preferences. Notably, our benchmark introduces several novel metrics (marked in brown) that were not present in previous evaluation frameworks, particularly enhancing the assessment of spatial-temporal consistency and user preference compliance. This expanded set of metrics enables a more rigorous and comprehensive evaluation of travel planning systems.

A.2 Additional Analysis

A.2.1 RQ1: How Do Different Constraints Impact Travel Plan Final Pass Rate?

In Section 6 of the main text, we calculated the Pearson correlation coefficients between 14 constraints and the final pass rate to explore which constraints have a greater impact on the pass rate. Now, we have calculated the Kendall correlation coefficients between them to ensure the comprehensiveness and depth of this part of analysis, and the results are shown in the Figure 4. Through observation and comparison, it can be found that the results of Kendall correlation coefficients and Pearson correlation coefficients show the same trend, which also proves that our analysis results are scientific. We have confirmed that this part of content already has significant optimization implications for subsequent travel planning.

Category	Metric	Description
Commonsense Constraints	City Coverage	Verifies all target cities are included
	Activity Repetition	Detects duplicate activities in itinerary
	Time Interval	Verifies 30-minute minimum gaps between activities
	Accommodation	Ensures lodging is arranged for all nights except last
	Daily Schedule	Validates completion of daily activity planning
	Return Journey	Confirms return trip on final day
	POI Validation	Verifies existence of all points of interest
	Location Logic	Ensures POIs are within designated cities
User Preferences	Activity Count	Verifies minimum 4 activities per city daily (including three meals a day, attractions, and snack shops)
	Budget	Validates total cost against user budget
	Hotel Type	Confirms compliance with hotel preferences
	Required Sites	Verifies inclusion of must-visit attractions
	Excluded Sites	Ensures exclusion of unwanted locations

Table 7: Evaluation metrics for Travel Plan Evaluation. Metrics in black are common to both TravelPlanner (Xie et al., 2024) and our benchmark, while those in brown are unique to our benchmark.

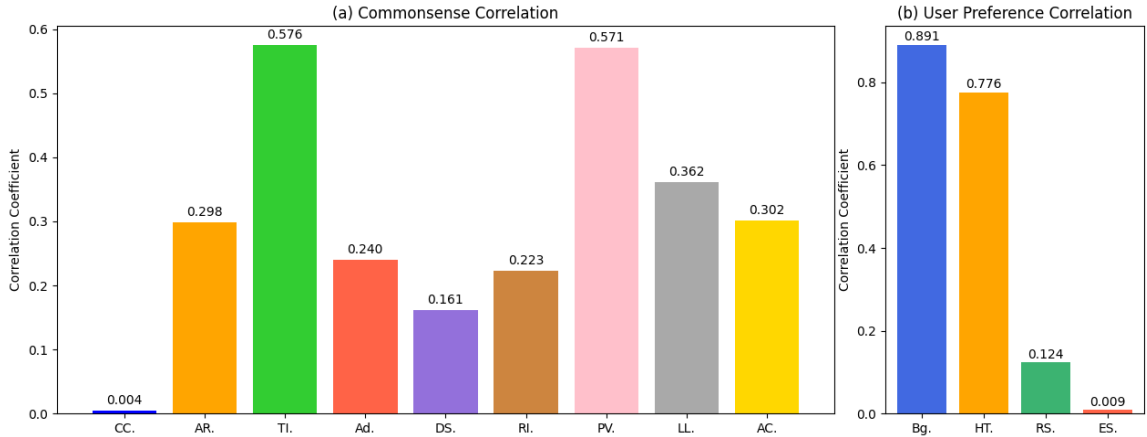


Figure 4: Impact of Different Constraints on Pass Rates. (a) CC: City Coverage, AR: Activity Repetition, TI: Time Interval, Ad: Accommodation, DS: Daily Schedule, RJ: Return Journey, PV: POI Validation, LL: Location Logic, AC: Activity Count. (b) Bg: Budget, HT: Hotel Type, RS: Required Sites, ES: Excluded Sites.

A.2.2 RQ2: How Important is Decision-Making Support in Travel Planning?

To evaluate **the importance of decision-making support**, we further employed **human evaluations** to compare travel plans generated with and without assisted decision-making processes.

As for **human evaluations**, we employed the same team of 5 expert annotators (previously used for dataset annotation). The annotators demonstrated strong agreement during dataset annotation,

with a Cohen’s kappa coefficient exceeding 0.9, ensuring high consistency in their evaluations.

We designed three 5-point metrics for evaluation.

- **Implicit Need Alignment** measures the system’s ability to address unstated requirements, where 5 points indicate full coverage of contextual needs through proactive suggestions, and 1 point reflects complete oversight of implicit preferences.
- **Revision Cost** quantifies plan refinement ef-

fort, where 5 points represent immediate acceptance (0–1 edits required), and 1 point indicates extensive rework (5+ edits needed) to meet basic requirements.

- **Overall Satisfaction** captures holistic user experience, where 5 points reflect perfect alignment with both stated and unstated needs, while 1 point indicates fundamental dissatisfaction with the proposed itinerary.

Table 8 showed statistically significant improvements across all metrics with decision-making support. Implicit need alignment rose from 2.1 to 3.9, with experts noting better recognition of contextual constraints (e.g., mobility requirements). Baseline plans required 2.1 average edits versus 1.2 with support, though three users still needed multiple adjustments for niche preferences. Four participants described the clarification process as "useful for refining initial ideas," while one found it excessive for simple day trips. Time savings averaged 28 minutes per task, with greater benefits for multi-city itineraries.

Setting	Need Align.	Rev. Cost	Satisfaction
w/ DMS	3.9	1.2	4.2
w/o DMS	2.1	2.1	3.1

Table 8: Performance comparison with/without Decision-Making Support (DMS). Need Align.: Implicit Need Alignment; Rev. Cost: Revision Cost; Satisfaction: Overall Satisfaction

A.2.3 RQ3: How Do Multimodal Information and User-Generated Content Enhance Travel Planning?

To evaluate the impact of multimodal information and user-generated content on travel planning quality, we employed the same team of 5 expert annotators (previously used for dataset annotation) to assess four system settings: basic text-only setting, enhanced with user reviews, enhanced with visual content, and enhanced with both elements. The annotators demonstrated strong agreement during dataset annotation, with a Cohen’s kappa coefficient exceeding 0.9, ensuring high consistency in their evaluations.

We designed four evaluation metrics on a 5-point scale.

- **Information Comprehension** measures users’ understanding of travel destinations,

where 5 points indicate complete and clear understanding of destination features, and 1 point reflects confusion or misunderstanding.

- **Decision Efficiency** evaluates the speed and quality of travel decisions, where 5 points represent quick and confident decision-making, and 1 point indicates hesitation and uncertainty.
- **Interaction Experience** assesses the effectiveness of plan modification process, where 5 points reflect smooth and intuitive plan adjustments with clear feedback, and 1 point suggests difficult and confusing modification experience.
- **Overall Satisfaction** captures the comprehensive user experience, where 5 points indicate highly satisfactory planning experience, and 1 point reflects poor user satisfaction.

The experimental results demonstrate that both visual content and user-generated content significantly enhance the travel planning process. While the basic text-only setting achieves moderate performance, the addition of either visual content or user reviews substantially improves user experience across all metrics. The combination of both elements yields the best performance, achieving scores above 4.0 across all dimensions, which validates the importance of incorporating multimodal information in travel planning systems.

Metrics	Basic	+UGC	+Visual	+Both
Info. Comprehension	3.28	3.85	3.88	4.12
Decision Efficiency	3.45	3.92	3.95	4.15
Interaction Exp.	3.52	3.89	3.87	4.08
Overall Satisfaction	3.31	3.88	3.85	4.16

Table 9: Impact of multimodal information and user-generated content on travel planning quality. Basic: text-only setting; +UGC: enhanced with user reviews; +Visual: enhanced with visual content; +Both: enhanced with both elements. All scores range from 0-5, averaged over user ratings.

A.2.4 RQ4: To What Extent Do Environmental Factors Influence Travel Planning?

To evaluate the impact of environmental factors on travel planning, we employed the same team of 5 expert annotators (previously used for dataset

annotation) to assess travel plans generated with and without environmental awareness. Each annotator evaluated 20 different scenarios, resulting in 100 test cases in total. The annotators demonstrated strong agreement during dataset annotation, with a Cohen’s kappa coefficient exceeding 0.9, ensuring high consistency in their evaluations.

We designed three 5-point metrics for evaluation.

- **Weather Adaptability Score** measures plan adaptation to weather conditions, where 5 points indicate comprehensive weather-based adjustments and alternatives, while 1 point reflects minimal weather consideration.
- **Transportation Accessibility Score** evaluates traffic-aware planning, where 5 points represent detailed routing with backup options, and 1 point indicates limited traffic consideration.
- **Environmental Alert Score in Requirement Clarification** assesses proactive environmental guidance, where 5 points represent timely alerts with practical suggestions, and 1 point reflects delayed or missing environmental alerts.

The experimental results demonstrate that environmental awareness significantly enhances travel planning quality. With environmental awareness enabled, all three metrics show notable improvements, particularly in the Environmental Alert Score (from 3.47 to 4.20). These results validate the importance of incorporating environmental factors in travel planning systems.

EA	WAS	TAS	EARC
✗	3.85	3.72	3.47
✓	4.20	4.00	4.20

Table 10: Impact of Environmental Awareness (EA) on travel planning quality. WAS: Weather Adaptability Score; TAS: Transportation Accessibility Score; EARC: Environmental Alert Score in Requirement Clarification. All scores range from 0-5, averaged over 100 test cases.

A.3 Detailed Decision-Making Support Evaluation

As shown in Table 11, our Topic-Guided Interaction Logic approach demonstrates significant improvements in models’ ability to clarify travel requirements. Our enhanced Qwen2.5-7B-Instruct

with SFT achieves a remarkable 252% improvement in BLEU-4 scores from 0.61 to 2.15 in one-shot learning, while maintaining stable performance in two-shot scenarios at 2.25. This consistency contrasts sharply with baseline models’ fluctuations, exemplified by gpt-4-Turbo’s BLEU-4 dropping from 0.85 to 0.35 between one-shot and two-shot settings.

The effectiveness is further validated by improvements across all metrics, with METEOR scores reaching 22.43 for gpt-4o and ROUGE-L achieving 16.95, indicating enhanced semantic understanding. Our method shows strong scalability, enabling smaller models like Qwen2.5-7B-Instruct to outperform larger baselines with a BLEU-1 score of 14.20. The consistent CIDEr score improvements across all enhanced models confirm the overall quality enhancement of generated responses.

These comprehensive improvements suggest our approach effectively addresses the core challenges in travel requirement clarification. The method’s success across both open-source and closed-source models, along with its stability in different shot settings, demonstrates its potential as a robust solution for enhancing travel dialogue systems’ ability to process ambiguous user requests.

A.4 Detailed Travel Plan Evaluation

Tables 12 and 13 demonstrate the significant improvements achieved by TGMA in travel plan generation. Our enhanced Qwen2.5-7B-Instruct with SFT shows remarkable performance, achieving the highest commonsense constraint micro-precision of 83.39% in CoT and 83.05% in one-shot settings, surpassing larger models like gpt-4o. In handling user preferences, our approach pushes the micro-precision to 87.94% and 88.78% respectively, marking substantial gains over baseline performances.

Most significantly, our method addresses the critical challenge of macro-precision scores. While baseline models show macro-precision scores near zero for commonsense constraints, our enhanced models achieve scores of 8.12% and 8.89% across different settings, demonstrating more consistent performance across varied scenarios. The improvement extends to user preference constraints, with macro-precision scores above 43% consistently, indicating better handling of diverse user requirements.

The final pass rate metric clearly demon-

Model	BLEU				METEOR	ROUGE-L	CIDEr
	B-1	B-2	B-3	B-4			
One-shot In-Context Learning							
Baselines							
Qwen2.5-7B-Instruct	9.67	2.78	1.10	0.61	19.91	17.89	0.353
Meta-Llama-3-8B-Instruct	2.10	0.50	0.15	0.06	10.62	13.53	0.328
glm-4-9b-chat	1.11	0.57	0.21	0.10	13.54	13.84	0.333
Baichuan2-13B-Chat	8.82	2.63	0.96	0.53	15.47	16.51	0.350
DeepSeek-V3(671B)	0.35	0.21	0.13	0.10	13.28	14.45	0.332
Doubao-Pro	6.06	1.00	0.38	0.20	12.22	15.04	0.335
Baichuan3-Turbo	2.41	1.15	0.58	0.40	15.91	14.14	0.337
Qwen-Plus	13.37	5.60	2.37	1.35	19.25	16.38	0.350
gpt-4-Turbo	5.30	2.31	1.25	0.85	15.00	15.22	0.341
gpt-4o	9.39	3.30	1.31	0.69	15.22	15.47	0.339
Topic-Guided Interaction Logic(ours)							
DeepSeek-V3(671B)	5.96	2.61	1.16	0.65	16.83	14.95	0.342
Qwen2.5-7B-Instruct(SFT)	14.20	6.70	3.45	2.15	21.80	16.50	0.362
gpt-4o	12.37	4.86	1.98	0.89	22.43	16.95	0.349
Two-shot In-Context Learning							
Baselines							
Qwen2.5-7B-Instruct	7.75	2.21	0.77	0.37	19.56	18.17	0.353
Meta-Llama-3-8B-Instruct	1.37	0.31	0.10	0.05	10.34	13.48	0.328
glm-4-9b-chat	0.70	0.36	0.13	0.07	12.96	13.62	0.331
Baichuan2-13B-Chat	10.16	3.59	1.33	0.69	15.20	15.99	0.349
DeepSeek-V3(671B)	0.29	0.17	0.12	0.09	13.30	13.47	0.336
Doubao-Pro	2.89	0.81	0.34	0.17	14.01	14.15	0.334
Baichuan3-Turbo	1.95	0.88	0.45	0.30	15.85	14.07	0.336
Qwen-Plus	12.30	5.14	2.04	1.12	19.16	16.14	0.347
gpt-4-Turbo	2.47	1.08	0.55	0.35	14.14	14.53	0.337
gpt-4o	9.97	3.99	1.46	0.76	17.07	15.81	0.343
Topic-Guided Interaction Logic(ours)							
DeepSeek-V3(671B)	7.31	3.19	1.67	1.09	17.30	15.30	0.347
Qwen2.5-7B-Instruct(SFT)	14.40	6.90	3.60	2.25	22.00	16.65	0.364
gpt-4o	12.86	4.72	1.98	1.02	22.29	16.76	0.343

Table 11: Decision-Making Support Evaluation under One-shot and Two-shot In-Context Learning. ROUGE, BLEU, and METEOR metrics are expressed as percentages (%). Gray and yellow backgrounds denote open-source and closed-source models.

Model	Commonsense Constraint				User Preference Constraint				Final Pass Rate
	Micro		Macro		Micro		Macro		
	PR.	F1	PR.	F1	PR.	F1	PR.	F1	
CoT									
<i>Baselines</i>									
Qwen2.5-7B-Instruct	47.17	0.49	0.00	0.32	76.33	0.75	26.24	0.43	0.00
Meta-Llama-3-8B-Instruct	63.03	0.63	0.00	0.39	72.41	0.72	24.55	0.42	0.00
glm-4-9b-chat	51.66	0.52	0.21	0.33	77.53	0.77	30.30	0.43	0.10
Baichuan2-13B-Chat	55.00	0.55	0.00	0.35	50.00	0.50	0.00	0.33	0.00
DeepSeek-V3(671B)	67.70	0.71	1.14	0.40	79.49	0.79	36.15	0.44	0.30
Doubao-Pro	54.49	0.52	0.00	0.35	77.45	0.76	31.37	0.43	0.00
Baichuan3-Turbo	70.54	0.71	0.00	0.41	65.94	0.66	10.77	0.39	0.00
Qwen-Plus	68.34	0.70	0.65	0.41	80.18	0.80	37.42	0.44	0.23
gpt-4-Turbo	74.46	0.70	2.14	0.42	77.72	0.77	32.03	0.43	0.87
gpt-4o	77.17	0.77	3.51	0.43	78.99	0.79	33.77	0.44	0.90
<i>TGMA(ours)</i>									
DeepSeek-V3(671B)	79.81	0.78	5.76	0.43	80.02	0.87	42.24	0.45	1.41
Qwen2.5-7B-Instruct(SFT)	83.39	0.79	8.12	0.47	87.94	0.86	43.09	0.46	2.99
gpt-4o	77.78	0.81	4.32	0.45	84.43	0.82	42.67	0.44	2.03

Table 12: Travel Plan Evaluation under Chain-of-Thought(CoT) Setting. Pass Rate (PR.) metrics are reported in percentages (%). Gray and yellow backgrounds denote open-source and closed-source models.

Model	Commonsense Constraint				User Preference Constraint				Final Pass Rate
	Micro		Macro		Micro		Macro		
	PR.	F1	PR.	F1	PR.	F1	PR.	F1	
One-shot In-Context Learning									
<i>Baselines</i>									
Qwen2.5-7B-Instruct	58.26	0.58	0.08	0.37	79.92	0.79	29.95	0.44	0.08
Meta-Llama-3-8B-Instruct	67.81	0.68	0.26	0.40	67.70	0.68	12.52	0.40	0.50
glm-4-9b-chat	71.62	0.72	1.55	0.41	77.97	0.41	33.15	0.44	0.86
Baichuan2-13B-Chat	72.50	0.73	0.00	0.42	58.83	0.58	0.00	0.37	0.00
DeepSeek-V3(671B)	79.47	0.79	4.88	0.44	80.02	0.80	37.52	0.44	1.32
Doubao-Pro	77.85	0.78	2.35	0.43	76.33	0.76	29.74	0.43	0.84
Baichuan3-Turbo	72.76	0.72	1.02	0.42	71.19	0.71	20.14	0.41	0.25
Qwen-Plus	77.64	0.77	3.08	0.44	76.90	0.78	29.95	0.44	0.60
gpt-4-Turbo	74.02	0.74	1.17	0.42	77.26	0.77	30.53	0.43	0.42
gpt-4o	76.64	0.76	2.55	0.43	78.41	0.78	33.24	0.44	0.75
<i>TGMA(ours)</i>									
DeepSeek-V3(671B)	80.02	0.80	6.71	0.43	80.97	0.87	43.90	0.47	2.03
Qwen2.5-7B-Instruct(SFT)	83.05	0.83	8.89	0.62	88.78	0.86	44.29	0.46	4.06
gpt-4o	76.90	0.87	4.49	0.47	89.90	0.83	43.88	0.44	2.31

Table 13: Travel Plan Evaluation under One-shot In-Context Learning Setting. Pass Rate (PR.) metrics are reported in percentages (%). Gray and yellow backgrounds denote open-source and closed-source models.

Setting		BLEU				METEOR	ROUGE-L	CIDEr
KB	Topic	B-1	B-2	B-3	B-4			
✓	✗	9.29	2.58	0.51	0.18	19.89	16.67	0.336
✗	✓	5.42	1.49	0.32	0.12	15.67	15.82	0.335
✗	✗	4.82	1.27	0.35	0.16	15.63	15.90	0.336
✓	✓	10.96	4.50	2.29	1.36	22.05	16.85	0.343

Table 14: Ablation study on Decision-Making Evaluation based on our framework with GPT-4o. ROUGE, BLEU, and METEOR metrics are expressed as percentages (%). ✓/✗ indicates whether the corresponding module is used or not.

Setting		Commonsense Constraint				User Preference Constraint				Final Pass Rate
ID	OP	Micro		Macro		Micro		Macro		
		PR.	F1	PR.	F1	PR.	F1	PR.	F1	
✓	✗	47.36	0.47	3.30	0.28	49.53	0.50	27.74	0.29	0.47
✗	✓	76.70	0.71	6.80	0.44	79.67	0.79	36.80	0.44	0.67
✗	✗	72.19	0.64	7.33	0.43	77.01	0.78	33.60	0.44	0.38
✓	✓	78.65	0.79	33.49	0.44	80.30	0.81	37.30	0.45	1.07

Table 15: Ablation study on Travel Plan Evaluation based on our Multi-Agent framework with GPT-4o. Pass Rate (PR.) and F1 metrics are reported in percentages (%). ID and OP denote Intent Detection and Overall Planning agents respectively.

strates our method’s effectiveness. Our enhanced Qwen2.5-7B-Instruct achieves pass rates of 2.99% and 4.06%, significantly outperforming both open-source and closed-source baselines. This improvement represents success in simultaneously satisfying both commonsense and user preference constraints, a challenging requirement that most baseline models struggle with, as evidenced by their near-zero pass rates. These results validate TGMA as an effective solution for generating practical and user-aligned travel plans.

A.5 Ablation Study

A.5.1 Decision-Making Support

As shown in Table 14, our ablation study reveals the crucial role of both knowledge base and topic-guided components in travel requirement clarification. The full model incorporating both components achieves optimal performance across all metrics, with BLEU-1 reaching 10.96 and BLEU-4 at 1.36, significantly outperforming all ablated variants. The knowledge base component proves particularly vital, as its removal causes BLEU-1 to drop sharply to 5.42, while removing the topic-guided component results in BLEU-1 declining to

9.29.

The synergistic effect of both components is further evidenced by the METEOR scores, where the full model achieves 22.05, compared to 19.89 with only knowledge base and 15.67 with only topic guidance. The model’s performance deteriorates most severely when both components are removed, with BLEU-1 falling to 4.82, demonstrating that neither component alone is sufficient for optimal performance. These results strongly validate our design choice of integrating both knowledge-based reasoning and topic-guided interaction for effective travel requirement clarification.

A.5.2 Travel Plan Generation

As shown in Table 17, our ablation study demonstrates the effectiveness of both Intent Detection (ID) and Overall Planning (OP) agents in travel plan generation. The full model with both modules achieves the best overall performance, with micro-precision reaching 78.65% for commonsense constraints and 80.30% for user preferences, along with a final pass rate of 1.07%. The OP module proves particularly crucial, as its removal causes significant performance degradation across

Model	BLEU				METEOR	ROUGE-L	CIDEr
	B-1	B-2	B-3	B-4			
Pre-fine-tuning Performance							
glm-4-9b-chat	7.54	3.90	1.94	1.18	17.09	15.14	0.348
Qwen2.5-7B-Instruct	8.29	3.23	1.44	0.79	17.26	15.34	0.343
Baichuan2-13B-Chat	9.49	4.41	1.76	0.96	14.74	15.49	0.345
Post-fine-tuning Performance							
glm-4-9b-chat	14.41	7.22	3.79	2.35	18.84	16.57	0.353
Qwen2.5-7B-Instruct	12.89	6.02	3.02	1.79	17.67	16.19	0.360
Baichuan2-13B-Chat	10.28	3.86	1.72	0.92	18.64	15.89	0.347

Table 16: Decision-Making Support Evaluation comparing model performance before and after fine-tuning. ROUGE, BLEU, and METEOR metrics are expressed as percentages (%).

Model	Commonsense Constraint				User Preference Constraint				Final Pass Rate
	Micro		Macro		Micro		Macro		
	PR.	F1	PR.	F1	PR.	F1	PR.	F1	
Pre-fine-tuning Performance									
glm-4-9b-chat	54.13	0.54	0.17	0.35	77.32	0.77	29.69	0.43	0.10
Qwen2.5-7B-Instruct	49.38	0.50	0.00	0.32	74.49	0.74	22.87	0.42	0.00
Baichuan2-13B-Chat	50.00	0.50	0.00	0.33	60.42	0.75	0.00	0.43	0.00
Post-fine-tuning Performance									
glm-4-9b-chat	71.50	0.72	2.20	0.41	75.17	0.42	38.33	0.44	1.00
Qwen2.5-7B-Instruct	73.71	0.73	1.57	0.42	85.55	0.86	56.60	0.46	1.40
Baichuan2-13B-Chat	65.82	0.66	0.95	0.38	71.43	0.72	32.15	0.42	0.72

Table 17: Travel Plan Evaluation comparing model performance before and after fine-tuning. Pass Rate (PR.) metrics are reported in percentages (%).

all metrics, most notably in commonsense constraint micro-precision dropping to 47.36% and user preference micro-precision falling to 49.53%.

Most significantly, the macro-precision scores reveal the complementary nature of these modules. The full model achieves 33.49% macro-precision for commonsense constraints, substantially outperforming variants with single or no modules, which only achieve scores below 8%. This dramatic improvement in macro-precision indicates our complete framework’s superior ability to handle diverse travel planning scenarios consistently. These results validate our design choice of combining

constraint-aware intent detection with controlled planning for robust travel plan generation.

A.6 Fine-tuning Analysis

A.6.1 Decision-Making Support

As shown in Table 16, our Dialog-SFT approach demonstrates significant improvements in models’ ability to clarify travel requirements. The most notable enhancement is observed in glm-4-9b-chat, where BLEU-4 scores increase from 1.18 to 2.35, accompanied by substantial gains across all metrics, with BLEU-1 improving from 7.54 to 14.41 and METEOR rising from 17.09 to 18.84. Similarly,

Qwen2.5-7B-Instruct shows remarkable progress with BLEU-4 more than doubling from 0.79 to 1.79.

Interestingly, smaller models achieve more substantial improvements compared to larger ones, as evidenced by Baichuan2-13B-Chat’s relatively modest gains. This suggests our Dialog-SFT approach effectively enhances model capabilities regardless of model size, particularly benefiting smaller architectures. The consistent improvements in METEOR and ROUGE-L scores across all models indicate enhanced semantic understanding and response coherence, validating the effectiveness of our fine-tuning strategy in improving travel requirement clarification capabilities.

A.6.2 Travel Plan Generation

As shown in Table 17, our fine-tuning approach demonstrates substantial improvements in models’ travel planning capabilities. Qwen2.5-7B-Instruct shows the most remarkable enhancement, with its micro-precision for commonsense constraints increasing from 49.38% to 73.71%, and user preference constraints improving significantly from 74.49% to 85.55%. Most notably, its macro-precision for user preferences sees a dramatic increase from 22.87% to 56.60%, indicating greatly improved consistency across diverse planning scenarios.

The effectiveness of our fine-tuning approach is further validated by the final pass rate improvements. All models progress from near-zero pass rates to achieving meaningful success, with Qwen2.5-7B-Instruct reaching 1.40%. The consistent improvements in both commonsense and user preference constraints across all models, particularly in macro-precision metrics, demonstrate our fine-tuning strategy’s effectiveness in enhancing models’ ability to generate travel plans that satisfy multiple practical constraints simultaneously.

B Tourism Knowledge Base Examples

We present examples from our tourism knowledge base, which covers five main categories: attractions, accommodations, dining, transportation, and weather information. These examples demonstrate the structured information used to support our dialogue system. And we have desensitized the private information.

B.1 City

Our tourism dataset encompasses 24 major cities across China(as shown in Figure 5), representing a diverse geographic and cultural landscape. These cities span from the northern regions (Beijing, Tianjin, Harbin, Shenyang, and Dalian), through the eastern coastal areas (Shanghai, Nanjing, Hangzhou, Suzhou, and Xiamen), to the southern region (Guangzhou, Shenzhen, and Fuzhou). The dataset also covers central China (Wuhan, Changsha, and Nanchang), western China (Chengdu, Xi’an, Lanzhou, and Xining), and southwestern China (Kunming, Guiyang, and Chongqing). This comprehensive coverage ensures our dataset captures the rich diversity of China’s tourism resources, from ancient historical sites to modern urban attractions, and from coastal sceneries to inland landscapes. These cities were selected based on their tourism significance, economic importance, and distinctive cultural characteristics, providing a representative sample of China’s major tourist destinations.



Figure 5: Geographic distribution of the 24 cities in our dataset

B.2 Attraction

Table 18 presents a detailed example of an attraction entry from our tourism database. The entry encompasses comprehensive basic information including name, introduction, visual content, location, and contact details. It also contains visitor-oriented practical information such as recommended visiting times and costs, along with surrounding service details covering nearby restau-

Attraction name: Jiufeng Forest Zoo

Introduction: Wuhan Jiufeng Forest Zoo, nestled in Jiufeng National Forest Park, is just 12km from downtown Wuhan. Set across 53.3 hectares with 85% forest coverage, the zoo houses diverse wildlife in specialized areas including tiger, lion, bear, and monkey gardens. The facility features an animal kindergarten, a parent-child eco-park, and two popular performance venues - the "Dream Theater" and "Bear and Tiger Gathering."

Basic Information:

- Image:
- Address: No. 1 Special, Jiufeng Lion Peak, Hongshan District, Wuhan City, Hubei Province
- Phone: *****
- Geographic Coordinates: *****° East, *****° North
- Type: Attraction-Zoo-Botanical and Zoological Garden

Visitor Information:

- Best Season to Visit: Suitable for all seasons
- Recommended Visit Duration: 3-5 hours
- Must-Visit Index: Must-see attraction
- Average Expense per Person: ** yuan
- Rating: 3.8 stars

Nearby Restaurants:

- Green Tea Restaurant (Wuhan Joy City Branch) - 1.5km away
- Xiaofu Nan Hunan Cuisine (Joy City Branch) - 1.5km away
- Victoria's Style Original Cut Steak Western Restaurant (Joy City Branch) - 1.5km away

Nearby Accommodations:

- Chain Hotels:
 - City Comfort Inn (Wuhan Jiufeng Mountain Forest Park Branch) - 1.0km away
 - Fengyi Fashion Hotel (Wuhan Jiufeng Mountain Forest Park Branch) - 0.9km away
 - City Comfort Inn (Wuhan Optics Valley Science and Technology Convention Center Branch) - 2.1km away
- Upscale Hotels:
 - Yishang Hotel (Wuhan Optics Valley Bio-city Joy City Branch) - 0.9km away
 - Wyndham Garden Wuhan Optics Valley - 1.9km away
 - Yishang Hotel (Wuhan Optics Valley Science and Technology Convention Center Joy City Branch) - 2.2km away

Visitor Reviews:

- The zoo boasts extensive forest coverage and diverse wildlife, with family-friendly attractions including a children's farm and interactive areas.
- Located just 40 minutes from downtown, the zoo offers convenient access and parking. A typical visit takes 2-3 hours, featuring midday animal performances.
- Interactive animal feeding experiences are available for visitors seeking closer encounters with wildlife.

Ticket Information:

- Single Admission Ticket:
 - Adult Ticket: **** yuan
 - College Student Ticket: **** yuan
- Package Deals:
 - Two-Person Ticket: ***** yuan
 - Family Ticket (1 Adult + 1 Child): ** yuan
 - Family Ticket (2 Adults + 1 Child): *** yuan
- Special Attraction Tickets:
 - Forest Secret Realm (Immersive Holographic Interaction): ***** yuan
 - Family Farm + Forest Secret Realm Package: **** yuan
- Combination Offers:
 - Adult Ticket + Forest Secret Realm: ** yuan
 - College Student Ticket + Forest Secret Realm: ** yuan

Table 18: A detailed attraction entry showcasing comprehensive attraction information spanning location details, visitor guidance, nearby POIs (restaurants and accommodations), ticket options, and user feedback for Jiufeng Forest Zoo.

rants and accommodations within 3km. Additionally, the entry includes visitor reviews providing qualitative insights and detailed ticket information with various pricing options. This structured format is consistently applied across our database, ensuring a holistic representation of each attraction through both factual information and user-generated content.

B.3 Restaurant

Table 19 presents a detailed restaurant entry from our tourism database. The entry provides comprehensive information about Xiabu Xiabu, a popular hotpot restaurant in Beijing’s Wangfujing area. It includes basic operational details, location information, and customer reviews. The data structure encompasses both objective information (such as opening hours and pricing) and subjective feedback through customer reviews, offering potential visitors a complete picture of the dining experience.

B.4 Hotel

Table 20 presents a comprehensive hotel entry from our tourism database, showcasing the Shanghai Zunmao Hotel in the Lujiazui Financial District. The entry encompasses detailed information including basic hotel details, facilities, room types with pricing, and guest reviews. This structured format provides potential visitors with both essential operational information and real guest experiences, facilitating informed accommodation decisions.

B.5 Transportation

Table 21 shows a sample train service entry from our tourism database. The entry includes key travel information such as station details, schedule, journey duration, and pricing for high-speed rail service between Hangzhou and Shanghai.

B.6 Weather

Table 22 demonstrates a weather data entry from our tourism database. The entry includes essential meteorological information such as temperature, weather conditions, wind status, and air quality index, providing tourists with crucial environmental data for their travel planning.

C Annotations and Examples of RETAIL

C.1 Detailed Dataset Annotations

C.1.1 Step 1: Tourism Knowledge Base Validation

- **Data Cleaning & Completeness Check:** Thoroughly clean the tourism knowledge base by removing any entries with missing or incomplete key fields, such as POI (Point of Interest) names, locations, operating hours, or accessibility details. Ensure all essential information is standardized and formatted consistently.
- **Privacy Protection & Anonymization:** Handle sensitive data, such as visitor records or booking details, by implementing random replacement algorithms for personal identifiers, contact information, and payment records to comply with data protection regulations.
- **Transportation Data Verification & Optimization:** Validate transportation data by cross-referencing actual ticket prices, schedules, and route availability with official sources. Where direct routes are unavailable, supplement the database with optimized transfer options. Additionally, verify real-time updates for seasonal route changes or service disruptions.

This ensures the knowledge base remains accurate, secure, and user-friendly for travelers.

C.1.2 Step 2: Decision-Making Support Refinement

- **Dialogue Evaluation:** Systematically evaluate conversations across key quality dimensions, filtering out dialogues exhibiting misaligned recommendations, ignored user needs, redundant suggestions, unnatural transitions, mechanical responses, inconsistent styles, or insufficient clarifications.
- **Dialogue Optimization:** Enhance interaction quality by refining prompts and improving dialogue flow to create more natural, human-like travel planning conversations.

This refinement process guarantees the system delivers coherent, personalized travel advice while maintaining conversational naturalness comparable to professional travel agents.

C.1.3 Step 3: Plan Generation and Revision Verification

- **Feasibility Check:** Verify all generated travel plans for practical viability, eliminating those

Introduction: Established in Beijing in 1998, Xiabu Xiabu pioneered counter-style hotpot dining with various soup bases (clear, spicy, sour-spicy, curry) and signature sesame sauce. The restaurant's popularity is reflected in its constant queues, though efficient service ensures short waiting times.

- ### Reviews:

Table 19: A detailed restaurant entry demonstrating comprehensive dining information including basic details, operational hours, dining environment, and customer reviews for Xiabu Xiabu.

Introduction: Located in Lujiazui Financial District, this tech-savvy hotel offers rooms with internet connectivity and on-demand entertainment. Featuring both Chinese and Western restaurants plus a modern bar, the hotel provides comprehensive recreational facilities, conference spaces and banquet halls. Recently renovated floors ensure enhanced guest comfort. Our staff looks forward to serving you!

- Parking: First 3 hours free
- WiFi: Available
- Stars: 4.5 stars

- Superior King Room: *** yuan (Entry-level option)
- Deluxe Twin Room: *** yuan (Mid-range option)
- Executive King Room: **** yuan (Business-class option)
- Business Suite: **** yuan (Premium option)
- Duplex Executive Apartment: **** yuan (Luxury option)

- Manager Xiao Luo delivers exceptional service with great patience and efficiency.
- A well-established hotel near the Stock Exchange in Pudong. Recently renovated and maintains high standards for business conferences.
- Clean, spacious rooms with modern renovations. Front desk staff member Xu Jiahui provides warm and attentive service.

Table 20: A detailed hotel entry illustrating detailed accommodation information spanning location details, facilities, room categories with pricing, and guest reviews for Shanghai Zunmao Hotel.

Origin City: Hangzhou
Destination City: Shanghai
Departure Station: Hangzhou South
Arrival Station: Shanghai Hongqiao
Train Number: G7382
Train Type: G-High Speed Rail
Journey Duration: 1h 13m
Departure Time: 22:30
Arrival Time: 23:43
Ticket Price: **** yuan

Table 21: A transportation database entry showing detailed high-speed rail journey information including route details, timing, duration, and pricing for travel between Hangzhou and Shanghai.

- with unrealistic schedules, incompatible accommodations, unworkable transportation arrangements, or logically conflicting activities.
- **Plan Revision:** During modifications, maintain equilibrium between attraction adjustments and transportation coordination, ensuring all revised plans retain both feasibility and alignment with user preferences.

This verification process guarantees that all recommended travel plans are practical and tailored to individual needs.

C.2 Statistics of RETAIL

Type	Total	Training	Val	Test
Single-turn	3,500	2,100	830	700
Single-turn + Revision	2,500	1,500	119	500
Multi-turn	2,500	1,500	1,167	500
Multi-turn + Revision	1,500	900	66	300

Table 23: **Distribution of Different Planning Cases.** RETAIL is divided into four categories based on conversation turns and revision requirements. The numbers represent the sample size for each split (Total/Training/-Validation/Test).

To systematically evaluate our proposed approach, we construct a comprehensive dataset consisting of both single-turn and multi-turn planning scenarios. As shown in Table 23, the dataset contains 10,000 cases in total, covering four different types of planning scenarios. Specifically, we collect 3,500 single-turn cases and 2,500 single-turn cases with revision requirements, which represent basic planning scenarios and their modifications respectively. For more complex scenarios,

City: Beijing
Date: 2024-01-01
High Temperature: 2° C
Low Temperature: -7° C
Weather: Cloudy
Wind: West wind level 1
Air Quality: 103 (Mild pollution)

Table 22: A weather database entry displaying comprehensive environmental conditions including temperature range, weather status, wind conditions, and air quality index for Beijing on January 1, 2024.

we include 2,500 multi-turn conversation cases and 1,500 multi-turn cases with revision requirements. The dataset is split into training, validation and test sets with approximate ratios of 60%, 20% and 20% respectively, ensuring sufficient samples for model training while maintaining reliable evaluation. This diverse distribution allows us to comprehensively assess the model’s capability in handling various planning scenarios with different complexity levels.

C.3 Case Study on RETAIL

Our travel planning dataset encompasses four distinct interaction patterns, demonstrating various levels of complexity in travel assistance scenarios.

C.3.1 Single-Turn Travel Planning

As shown in Figure 6, the dataset contains examples where comprehensive itineraries are generated from single, detailed user requests. These responses demonstrate the ability to process multiple constraints simultaneously, transforming them into structured daily schedules with precise timing, transportation arrangements, accommodation details, and activity recommendations. The planning maintains both practical feasibility and user preference alignment.

C.3.2 Single-Turn Planning with Revision

As shown in Figure 7, this category showcases the model’s adaptability in real-time plan modification. When presented with new constraints such as weather changes or timing conflicts, the responses demonstrate intelligent adaptation by suggesting appropriate alternative activities while maintaining the original schedule structure and user preferences. The revisions preserve both practical feasibility and the overall travel experience quality.



Hello, I am your intelligent travel planning assistant. How may I assist you?

Hello, I would like to create a travel plan. We plan to depart from Chongqing and head to Nanjing and Shanghai. The departure date is May 1, 2024, and the end date is May 5. There are two of us, and the budget is about 15,000 yuan. We hope to find some cheaper hotels, and we would like to try baozi and hot pot chicken. During the trip, we hope to go to some art exhibitions and historical sites. Please help me plan the detailed itinerary for this trip, thank you!



Five-Day Tour Guide for Nanjing and Shanghai

Day 1: Depart from Chongqing, Arrive in Nanjing

On May 1, 2024, you will depart from Chongqing, take the D958 high-speed train, which departs promptly at 6:30 in the morning, and is expected to arrive in Nanjing at 15:20. The journey will take you through beautiful landscapes, allowing you to enjoy the scenery along the way.

Upon arrival in Nanjing, you will check into the Zsmart Hotel (Nanjing Confucius Temple Himalaya Audiobook Library Branch). This hotel is located at No. 115 Jiankang Road, near Shuiyou City, and is only 300 meters walk from Exit 3 of the Confucius Temple Station on Metro Line 3, making your travel convenient. The rooms are comfortable and priced at ¥202.00.



On May 2, you will spend a fulfilling day in Nanjing. At 9 am, you can go to the Nanjing Art Museum to appreciate a wealth of art works. The museum is free of charge, so you can fully enjoy the influence of art.

![Nanjing Art Museum]

At 12:30 pm, you can enjoy a delicious lunch at Chicken Bag Fish-Abalone Chicken Soup Hot Pot (Changbai Street Branch), which is only 120 meters away from the Art Museum. The average consumption per person is about ¥96.0.

![Chicken Bag Fish-Abalone Chicken Soup Hot Pot]

After lunch, you can go to the **Nanjing Museum**. The adult ticket here is only ¥25.0. You can visit between 14:00 and 15:30 to learn about the history and culture of Nanjing.

![Nanjing Museum]

Then, from 16:00 to 17:00, you can go to Zhangji Big Bowl Belly Noodles to enjoy a bowl of authentic Nanjing noodles. The average consumption per person is only ¥21.0.

![Zhangji Big Bowl Belly Noodles]

Finally, from 17:30 to 18:30, you can visit the Nanjing Digital Art Museum, which is also free of charge, to experience the charm of modern art.

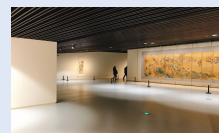
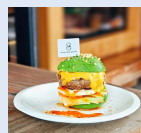
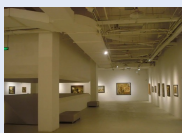
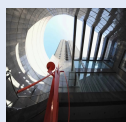
![Nanjing Digital Art Museum]

In the evening, you will return to the Zsmart Hotel to rest and prepare for tomorrow's journey.

Day 3: Nanjing to Shanghai

On May 3, you will take the G7195 high-speed train, departing from Nanjing at 15:00 in the afternoon and expected to arrive in Shanghai at 16:22. The ticket price is ¥101.0.

Upon arrival in Shanghai, you will check into the Home Inn-neo (Shanghai Dapujiao Luban Road Metro Station Branch). The hotel is located at No. 399 Dapu Road. The rooms are comfortable and priced at ¥310.90.



Day 4: Art Tour of Shanghai

On May 4, you will enjoy a rich cultural experience in Shanghai. At 9 am, you can go to the Shanghai Chinese Painting Academy, which is free of charge. Here, you can appreciate the charm of traditional Chinese painting.

![Shanghai Chinese Painting Academy]

At 12:30 pm, you can have lunch at Renheguan (Zhaojiabang Road Branch), not far from the academy. The average consumption per person is about ¥203.0.

![Renheguan]

After lunch, you can go to the Shanghai Oil Painting & Sculpture Institute Art Museum, which is also free of charge. You can appreciate various oil paintings and sculptures from 14:00 to 16:00.

![Shanghai Oil Painting & Sculpture Institute Art Museum]

Next, from 16:30 to 17:30, you can go to CHAR CHAR BISTRO-Just Right Steak Burger (Gubei Branch) for dinner. The average consumption per person is about ¥106.0.

![CHAR CHAR BISTRO]

Finally, from 18:00 to 19:00, you can visit the Haipai Art Museum, which is also free of charge, to appreciate the artistic style of Shanghai.

![Haipai Art Museum]

In the evening, you will return to the Home Inn-neo to rest and prepare for the last day of the journey.

Day 5: Return to Chongqing

On May 5, you will take the D952 high-speed train, departing from Shanghai at 6:45 in the morning and expected to arrive in Chongqing at 17:07. The ticket price is ¥585.0. This concludes your fulfilling and wonderful journey.

We hope that you have gained wonderful memories and experiences from your travels in Nanjing and Shanghai!

Figure 6: **Single-Turn data sample.** User expresses a desire to visit Nanjing and Shanghai, clearly stating multiple requirements such as budget, number of people, and personal preferences. The AI assistant provides a detailed and comprehensive five-day travel plan, covering all aspects of the trip from transportation and activities to meals and accommodation. The plan is highly feasible and well-suited to meet real-world needs. Additionally, the plan takes into account the user's personalized requirements. For instance, since the user mentioned wanting to see some art exhibitions and historical sites, the plan includes visits to the Nanjing Museum of Art and the Nanjing Museum.



Hello, I am your intelligent travel planning assistant. How may I assist you?



Hello, I am planning a trip from Nanjing to Hangzhou and Xi'an from May 1, 2024 to May 5, 2024. I will be traveling alone on this trip, with a budget of about 5,000 yuan. I hope to stay in high-end hotels and enjoy some delicious Cantonese and Shaanxi cuisine. In Hangzhou, I want to go to the Zhejiang Provincial Museum, and in Xi'an, I must go to the White Horse Zen Temple and the Qujiang Dayue City Construction Museum. Please help me make a detailed travel plan.

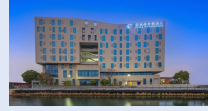
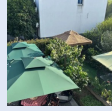


May Day Holiday Travel Guide 2024: Hangzhou and Xi'an
Day 1: Nanjing to Hangzhou (May 1, 2024)

On this sunny day, you will depart from Nanjing and take the G185 high-speed train to the beautiful Hangzhou. The train departs promptly at 15:15. After a journey of about 1 hour and 22 minutes, you will arrive in Hangzhou at 16:37. The ticket price is 85 yuan, making for a relaxed and pleasant journey.

Upon arrival, you will check into the Zprime Hotel (Hangzhou Zhejiang University Zijingang Branch). This is a modern hotel offering comfortable business suites priced at 305 yuan. The hotel is located at No. 806 Yuhangtang Road, convenient for your subsequent itinerary.

![Zprime Hotel]



Day 2: Hangzhou Day Tour (May 2, 2024)

Today's itinerary will take you deep into the culture and cuisine of Hangzhou. At 9 am, you will go to the **Zhejiang Provincial Museum** where you can appreciate a wealth of historical artifacts. The tour takes about an hour and admission is free.

![Zhejiang Provincial Museum]

After the tour, you can walk to the **Hangzhou Xinxin Hotel-1913 Restaurant** for lunch. The restaurant is only 526 meters away from the museum. The average consumption per person is about 121 yuan, and the dishes are fresh and delicious.

![Hangzhou Xinxin Hotel-1913 Restaurant]

After lunch, you will go to the Jinxu Fengshui Cave, a beautiful natural scenic area where you can enjoy the tranquility and beauty of nature. The tour takes about 4 hours and admission is free.

![Jinxu Fengshui Cave]

In the evening, you can go to the Xiyang Village Country Restaurant to taste authentic country dishes. The restaurant is about 1439 meters away from the Jinxu Fengshui Cave. The average consumption per person is about 79 yuan.

![Xiyang Village Country Restaurant]

In the evening, you will stay at the Radisson Seaga Hotel (Xiaoshan Airport Branch). This is a comfortable hotel with a business standard room priced at 320 yuan. It is located at No. 1 Baoshui Avenue, Jingjiang Street.

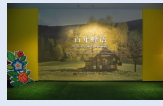
![Radisson Seaga Hotel]

Day 3: Hangzhou to Xi'an (May 3, 2024)

Today, you will take the G2388 high-speed train to Xi'an. The train departs at 09:00 and is expected to arrive at 16:40. The ticket price is 716 yuan. Along the way, you can enjoy the beautiful scenery.

Upon arrival in Xi'an, you will check into the Cophorne Hotel (Xi'an Lintong West Garden Metro Station Branch). This is a comfortable hotel with a cozy twin room priced at 207.96 yuan. It is located at No. 1 Huaqing Road.

![Cophorne Hotel]



Day 4: Xi'an Day Tour (May 4, 2024)

Today's itinerary will take you to appreciate the history and culture of Xi'an. At 9 am, you will go to the White Horse Zen Temple, a tranquil Buddhist temple. The tour takes about an hour and admission is free.

![White Horse Zen Temple]

Next, you will go to the Qujiang Dayue City Construction Museum, where you can learn about the architectural history of Xi'an. The tour takes about 1.5 hours and the ticket price is 60 yuan.

![Qujiang Dayue City Construction Museum]

At noon, you can have lunch at Wei Zhe Yin-Shaanxi Cuisine (Dayue City Branch). The restaurant is only 24 meters away from the museum. The average consumption per person is about 80 yuan.

![Wei Zhe Yin-Shaanxi Cuisine (Dayue City Branch)]

In the afternoon, you will go to the Xi'an Qujiang Art Museum to appreciate modern art works. The tour takes about 1.5 hours and the ticket price is 60 yuan.

![Xi'an Qujiang Art Museum]

In the evening, you will stay at the Xi'an Dayan Pagoda Hilton Home2 Hotel. This is a luxury hotel with a luxury room - king bed priced at 617 yuan. It is located at No. 659 Yunnan First Road, Qujiang New District, only 275 meters away from the Xi'an Qujiang Art Museum.

![Xi'an Dayan Pagoda Hilton Home2 Hotel]

Day 5: Xi'an to Nanjing (May 5, 2024)

On the last day, you will take the G1922 high-speed train back to Nanjing. The train departs at 07:15 and is expected to arrive at 13:17. The ticket price is 542 yuan. End this pleasant trip and return home with wonderful memories.

We hope that you gain a lot from this trip and enjoy every wonderful moment!

The travel arrangements are quite good, but there will be light rain in Hangzhou on the second day. I want to change the Jinxu Fengshui Cave to an indoor activity.



2024 Labor Day Holiday Travel Guide: Hangzhou, Xi'an Trip

Day 1: Nanjing to Hangzhou (May 1, 2024)

In the early morning sunshine, you will depart from Nanjing, take the high-speed train G7611, depart on time at 06:45, and arrive in beautiful Hangzhou at 08:07. The journey only takes 1 hour and 22 minutes, the fare is 85 yuan, easy and convenient.

Upon arrival, you will check into the All Seasons Hotel (Hangzhou West Lake Branch) located by the West Lake. The address is No. 76 Wulin Road, near the Fengqi Road subway station. The room is a superior king room, priced at 515.56 yuan. The hotel's environment is elegant, suitable for relaxation.

![All Seasons Hotel (Hangzhou West Lake Branch)]



Day 2: Hangzhou Tour (May 2, 2024)

Today's itinerary will take you to deeply experience the culture and natural beauty of Hangzhou.

Morning

Zhejiang Provincial Museum(Address: No. 25 Gushan Road;Opening hours: 09:00 - 10:00;Tickets: Free;Here, you can appreciate a wealth of historical relics and artworks, and feel the cultural heritage of Zhejiang.)

![Zhejiang Provincial Museum]

You can visit the Former Residence of Qian Xuesen to learn about the life and achievements of this great scientist. The tour takes approximately one hour, and admission is free.

![The Former Residence of Qian Xuesen]

Lunch

Hangzhou Xinxin Hotel-1913 Restaurant(Address: No. 56-60 Beishan Street (526 meters away from Zhejiang Provincial Museum);Dining time: 13:30 - 15:00;Average consumption: 121 yuan;The dishes here are fresh and delicious, you can taste the authentic Hangzhou cuisine.)

![Hangzhou Xinxin Hotel-1913 Restaurant]

Afternoon

Jiugu Water Restaurant (Baguatan Park Branch)(Address: Next to the north gate of Baguatan Relic Park, Nanfu Road (296 meters away from Yuhuang Mountain);Dining time: 15:30 - 17:00;Average consumption: 95 yuan;Enjoy afternoon tea at the water restaurant, feel the breeze on the lake, and it's extremely comfortable.)

![Jiugu Water Restaurant]

In the evening, you will return to the All Seasons Hotel to enjoy a peaceful night.

(a) Single-Turn-Revise data sample

Day 3: Hangzhou to Xi'an (May 3, 2024)
 Today, you will take the high-speed train G185, depart from Hangzhou at 15:15, and arrive in Xi'an at 16:37. The fare is 85 yuan. Xi'an, a city with a long history, is waiting for your exploration.
 Upon arrival, you will check into the Xi'an Kuan Yu MAX Hotel (Qujiang Datang Everbright City Branch). The address is No. 106, Middle Section of Furong West Road, near the Big Wild Goose Pagoda. The room is a clean and comfortable king room, priced at 289 yuan.

Day 4: Xi'an Tour (May 4, 2024)
 Today's itinerary will take you to appreciate the history and culture of Xi'an.

Morning
 White Horse Zen Temple(Address: White Horse Zen Temple (North of New White Line);Opening hours: 09:00 - 10:30;Tickets: Free;This is one of the earliest Buddhist temples in China, and the quiet environment purifies the mind.)
 ![White Horse Zen Temple]
 Qujiang Joy City Construction Museum(Address: Room F4-L4-09, Xi'an Joy City, No. 66 Ci'en West Road;Opening hours: 11:00 - 12:30;Tickets: Free;This place displays the architectural history and culture of Xi'an, worth a visit.)
 ![Qujiang Joy City Construction Museum]
 Lunch
 Wei Zhe Drink-Shaanxi Cuisine (Joy City Branch)(Address: Roof Garden, 4th Floor, Qujiang Joy City, No. 777 Ci'en Road (24 meters away from Qujiang Joy City Construction Museum);Dining time: 12:30 - 14:00;Average consumption: 80 yuan;Here, you can taste authentic Shaanxi cuisine, satisfying your taste buds.)
 ![Wei Zhe Drink-Shaanxi Cuisine]



Afternoon
 Xi'an Love Museum(Address: Floor -2, momopark Mall, No. 232 Xiaozhai West Road;Opening hours: 14:30 - 16:30;Tickets: Free;This is a place full of romantic atmosphere, suitable for couples and families to visit.)
 ![Xi'an Love Museum]
 In the evening, you will return to the Xi'an Kuan Yu MAX Hotel again to enjoy a comfortable night.

Day 5: Xi'an to Nanjing (May 5, 2024)
 In this pleasant trip, you will take the high-speed train G31 at 20:15, depart from Xi'an, and arrive in Nanjing at 21:24. The fare is 88 yuan. End this wonderful journey and go home with full memories.
 Hope you gain happiness and beauty in this trip, looking forward to seeing you next time!

(b) Continuation of Single-Turn-Revise data sample

Figure 7: **Single-Turn-Revise data sample.** In this case, after the AI assistant provided a complete travel plan, the user mentioned that there would be light rain in Hangzhou on the second day and hoped to switch to some indoor activities. The AI assistant understood this modification request and, in the subsequent revised travel plan, replaced the originally planned visit to the Jinxiu Fengshui Cave with a visit to The Former Residence of Qian Xuesen.

C.3.3 Multi-Turn Travel Planning

These examples in Figure 8 illustrate systematic information gathering through progressive dialogue. The model demonstrates sophisticated conversation management by sequentially clarifying essential details about dates, budget, group composition, accommodation preferences, and activity interests. This methodical approach ensures the final itinerary precisely matches user needs while maintaining practical travel logistics.

C.3.4 Multi-Turn Planning with Revision

As shown in Figure 9, the most complex category combines thorough requirement clarification with dynamic plan adaptation. These interactions begin with systematic information gathering through dialogue, followed by itinerary generation and subsequent refinement based on new constraints. This demonstrates the dataset's capability to handle realistic travel planning scenarios requiring both detailed preference understanding and flexible plan modification.

As shown in Figure 9, the user initially provided only limited information such as the destination and budget. The AI assistant then gradually collected other information through an assisted decision-making process. Once sufficient informa-

tion was gathered, it provided a complete travel plan. However, the user suggested a modification to the first plan, specifically wanting to replace the Panda Pedestrian Street attraction. After responding to this request, the AI assistant provided a second complete plan. This process clearly demonstrates a procedure that simultaneously incorporates assisted decision-making and plan revision.

C.3.5 Bad Case Analysis

We have annotated some bad cases and briefly presented it in Figure 10 and 11. In Figure 10, the itinerary for the second day ends at 15:15 in the afternoon, leaving free time unarranged. Secondly, after returning on the last day, hotel accommodation is still arranged, which is obviously not in line with common sense.

In Figure 11, since there is heavy rain on May 4th, which is not suitable for outdoor activities, the original arrangement to visit Qingyun Peak is inappropriate and should be modified to indoor activities. However, the revised plan to visit Ouyuan Garden is still an outdoor activity.



Hello, I am your intelligent travel planning assistant. How can I assist you?

I want to travel from Shenzhen to Xi'an, can you help me plan the itinerary?



Do you have a start and end date for your planned trip? This way I can better arrange the itinerary for you.

Yes, I plan to go to Xi'an from October 1 to October 5, 2024.



Understood, so what is your approximate budget for this trip? This way I can help you arrange suitable transportation, accommodation, and activities.

The total budget is around 10,000.



How many people are you planning to go with on this trip? This way I can help you better plan your accommodation and itinerary.

It's just me and another adult, no children are going.



Okay, now that I understand your basic situation. For this itinerary, I suggest we first select a place to stay to ensure you are comfortable. So, in terms of hotels, do you have a preferred type of accommodation-on? For example, do you prefer international chain hotels, boutique hotels, or distinctive guesthouses? Or do you have any specific requirements for the hotel, such as facilities, location, budget range, etc.?

I prefer to stay in a more upscale hotel, as long as the location is convenient, I don't have any other special requirements for now.



Okay, I'll note down the upscale hotel. So, let's talk about dining next. There are many famous snacks and delicious restaurants in Xi'an, what type of dining are you more interested in? Do you want to try authentic local snacks, prefer exquisite dishes in upscale restaurants, or have a particular cuisine you want to experience?

I am more interested in Shaanxi cuisine and beef skewers.



It seems that you are interested in Shaanxi cuisine and beef skewers, which is a great choice! Here are a few recommended restaurants for your reference during your trip to Xi'an:

1. Dahai Zi Shaanxi Cuisine (High-tech Branch)

- Location: 2nd Floor, Ziwei Longteng New World, No. 1 Keji Road
- Type: Authentic Shaanxi cuisine, offering authentic Shaanxi flavors
- Average consumption: 60 yuan
- Rating: Rated 50, pretty good! (Out of 50)

This place offers a wide variety of Shaanxi dishes and is a good place to experience local specialties.

2. Halal•Wang Xiaoxing Beef Skewers (Dongyuan West Road Branch)

- Location: Dongyuan West Road
- Type: Special skewers
- Average consumption: 47 yuan
- Rating: Score 47, the taste is very good, especially if you like beef skewers, you must not miss it. It is said that the beef skewers here have a unique flavor and are worth a try!

3. Su Qin Catering-Chang'an Hospitality Dishes (Bell and Drum Tower Branch)

- Location: 4th Floor, Ramada Plaza by Wyndham Xian Bell Tower, No. 79 North Street
- Type: Shaanxi cuisine
- Average consumption: 57 yuan
- Rating: Score 47, their dishes have always been well received.

The location is near the Bell and Drum Tower in the city center, you can start your Bell and Drum Tower tour while tasting delicious food. I hope these restaurants can bring you a delicious experience! Wish you a pleasant trip to Xi'an!



These restaurants all look good, especially the beef skewers shop, I will consider trying it. Thank you for your recommendation!



You're welcome, finding a favorite restaurant is a great pleasure in traveling! Now let's talk about the attractions you want to visit in Xi'an. Are there any particular places you want to go? Such as historical sites, museums, or natural scenery?

I would like to go to the Weiyang Palace of Han Chang'an City and the Xi'an Love Museum.



Xi'an Five-Day Travel Guide

Day 1: Depart from Shenzhen, Arrive in Xi'an

On October 1, 2024, you will depart from Shenzhen, take the high-speed train G822, depart at 8:30 in the morning, and are expected to arrive in Xi'an at 18:31. The journey is nearly ten hours, and the fare is 891 yuan. During this journey, you can enjoy the beautiful scenery along the way, looking forward to the history and culture of Xi'an.

Upon arrival, you will stay at Changning Palace Garden Hotel, an elegant hotel located at No. 1 Jiazi, Yubaotou Village. The room is "Zhongnan Renjia•Scenic Twin Room", priced at 418.89 yuan. The garden view of the hotel will add a sense of tranquility and comfort to your journey.

! [Changning Palace Garden Hotel]

Day 2: Xi'an Cultural Tour

On October 2, you will spend a fulfilling day in Xi'an. At 9 o'clock in the morning, go to the Han Chang'an City Weiyang Palace National Archaeological Site Park. This is a paradise for history lovers. You can feel the glory of ancient Chang'an here. The ticket is 199 yuan (double ticket Han Palace Campsite Package), and the tour time is 2 hours.

! [Han Chang'an City Weiyang Palace National Archaeological Site Park]

At 11:30 noon, you can go to Tang Island Catering (Deshi Square Store) for lunch. The restaurant is about 2042 meters away from the archaeological site park. The average consumption is about 60 yuan. You can taste authentic Xi'an cuisine.

! [Tang Island Catering]



At 2 pm, you will visit the **Xi'an Love Museum**, a place full of romance. You can feel the beauty of love here. The tour time is 2 hours and the admission is free. ! [Xi'an Love Museum]

(a) Multi-Turn data sample

At 4:30 in the evening, you can have dinner at Weizai-Battle Axe Steak (momopark store). The restaurant is only 17 meters away from the museum. The average consumption is about 50 yuan. Enjoy the delicious steak.

!![Weizai-Battle Axe Steak]

At night, you will stay at Chang'an Inn (Xiaozhai Dayan Pagoda Store), which is only 126 meters away from Xiaozhai West Road. The room is "Preferred Standard Double Room", priced at 143.33 yuan.

!![Chang'an Inn]



Day 3: Explore Xi'an's Nature and Wildlife

On October 3, you will continue your adventure in Xi'an. At 9 o'clock in the morning, go to Chang'an Tang Village-Nanbao Ancient Village, a place full of ancient charm. The ticket is 20 yuan, and the tour time is 2 hours.

!![Chang'an Tang Village-Nanbao Ancient Village]

At 11:30 noon, you can have lunch at Tang Village Luyu-Yueshan Farm Restaurant, which is only 549 meters away from the ancient village. The average consumption is about 86 yuan. Enjoy the fresh farm food.

!![Tang Village Luyu-Yueshan Farm Restaurant]

At 2 pm, you will go to Xi'an Qinling Wildlife Park, a place suitable for family fun. The ticket is 100 yuan (including the bus ticket for the Grassland Beasts Area), and the tour time is 3 hours.

!![Xi'an Qinling Wildlife Park]

At 5:30 in the evening, you can have dinner at the Lion Tiger Garden Theme Restaurant, which is located in the zoo. You can enjoy the food while appreciating the surrounding animals.

!![Lion Tiger Garden Theme Restaurant]

At night, you will stay at Xi'an Leleju Guesthouse, which is only 433 meters away from Neiyuan Village. The room is a "Parent-child Room", priced at 138 yuan.

!![Xi'an Leleju Guesthouse]

Day 4: Fun Time at the Water Park

On October 4, you will go to Xi'an Happy Valley-Maya Beach Water Park. Start playing at 9 o'clock in the morning, enjoy the fun of the water park, the play time is 3 hours, and the admission is free.

!![Xi'an Happy Valley-Maya Beach Water Park]

At 12:30 noon, you can have lunch at the beach restaurant in the water park and enjoy the delicious seafood.

At 5:30 in the evening, you can have dinner at the Lion Tiger Garden Theme Restaurant, which is located in the zoo. You can enjoy the food while appreciating the surrounding animals.

!![Lion Tiger Garden Theme Restaurant]

At night, you will stay at Xi'an Leleju Guesthouse, which is only 433 meters away from Neiyuan Village. The room is a "Parent-child Room", priced at 138 yuan.

!![Xi'an Leleju Guesthouse]

Day 4: Fun Time at the Water Park

On October 4, you will go to Xi'an Happy Valley-Maya Beach Water Park. Start playing at 9 o'clock in the morning, enjoy the fun of the water park, the play time is 3 hours, and the admission is free.

!![Xi'an Happy Valley-Maya Beach Water Park]

At 12:30 noon, you can have lunch at the beach restaurant in the water park and enjoy the delicious seafood.



In the afternoon, you will return to the hotel to prepare for the end of the upcoming journey.

At night, you will stay at Vienna International Hotel-Xi'an Fengdong Football Center Store, which is only 2748 meters away from Fengdong Avenue. The room is an "Executive Twin Room", priced at 364.93 yuan.

!![Vienna International Hotel]

Day 5: Farewell to Xi'an, Return to Shenzhen

On October 5, you will end this pleasant trip to Xi'an, take the high-speed train G844, depart at 5:00 pm, and are expected to arrive in Shenzhen at 2:40 am the next day. The fare is 895 yuan. With full memories, look forward to the next trip!

I hope this travel guide can help your trip to Xi'an, wish you a pleasant journey!

(b) Continuation of Multi-Turn data sample

Figure 8: **Multi-Turn data sample.** User initially only mentions wanting to travel to Shenzhen. This is entirely insufficient for generating a detailed travel plan. Therefore, instead of directly generating a plan, the AI assistant guides the user to provide relevant information through an assisted decision-making process. The AI assistant sequentially inquires about the travel dates, budget, number of people, and user preferences for dining and hotels, while also offering its own recommendations. Only after collecting sufficient information does the AI assistant provide a complete travel plan. Clearly, a plan obtained through this process can better adapt to real-world scenarios.



Hello, I am your intelligent travel planning assistant. How can I assist you?

I plan to travel to Chengdu and Lanzhou between May 5 and 9, 2024, and want to stay in high-end hotels. Our budget is about 25,000 yuan. Please give me some travel advice!





May I know which city you are departing from to Chengdu and Lanzhou? This way I can better plan your itinerary.

I am currently in Tianjin.





Okay! How many people are traveling with you on this trip? This way I can help you find suitable accommodation and activity arrangements.

There are two adults and one child on this trip.





Great! Both Chengdu and Lanzhou have a lot of delicious food to taste. Do you have any special expectations or want to try the local food? For example, do you like spicy food or light food? Or is there any particular type of restaurant you want to try, such as hot pot, Sichuan cuisine, or noodles?

We would like to try some local specialty restaurants, such as hot pot. In addition, we are also looking forward to trying Lanzhou's specialty pilaf. If you have any recommended places, let us know.





Okay, it seems that you are very interested in hot pot and specialty pilaf! I have compiled some good restaurants here, hoping to satisfy your taste.

Recommended hot pot in Chengdu

1. Nan Hot Pot (Taikoo Li Main Store)

- Type: Chongqing Hot Pot
- Address: 166 Xia Dong Street, East Street
- Average consumption: ¥80
- Recommended reason: Nan Hot Pot is known for its authentic Chongqing flavor, fresh ingredients, and rich soup base, allowing you to experience the authentic spicy and fresh flavor.

2. Chui Er Ge Hot Pot (Taikoo Li Store)

- Type: Sichuan Hot Pot
- Address: 166 Xia Dong Street, East Street
- Average consumption: ¥55
- Recommended reason: It is famous for its fresh hot pot ingredients and thick soup, very suitable for family gatherings.

Recommended pilaf in Lanzhou

1. Halal Ru Hua Hand-grabbed Restaurant (Lanzhou Main Store)

- Type: Other specialty dishes
- Address: 100 meters north of Western Happy Park, 23 Xiuchuan Road
- Average consumption: ¥93
- Recommended reason: This restaurant is widely praised for its authentic halal pilaf and other halal foods, suitable for experiencing the authentic Lanzhou flavor.



Enjoying delicious food while experiencing local food culture is a great choice! I hope these recommendations will make your trip more enjoyable.

These recommendations sound good, especially those hot pots and Lanzhou's pilaf. I will consider going to these places, thank you for your suggestions!





You're welcome! I'm glad you like the recommendations. Are there any specific attractions you would like to visit while traveling in Chengdu and Lanzhou? For example, are you interested in visiting the Wide and Narrow Alleys, or planning to go to the Giant Panda Breeding Research Base? In addition, Lanzhou has the Yellow River Iron Bridge and Baita Mountain, both are great attractions. Is there any place you are particularly interested in?

We plan to visit the Sancha Lake Water Park, Anren Ancient Town, Financial Expo Center, Panda Pedestrian Street, and Yellow River Building. If there are other must-see attractions, please give us some heartfelt recommendations.





Chengdu and Lanzhou both have many interesting attractions worth visiting. You have already planned to go to some great places, let me recommend a few more to ensure you make the most of your trip!

Chengdu Recommendations:

Guose Tianxiang

- Type: Theme Park
- Address: 1166 Jiangning South Road
- Reason for recommendation: This is a comprehensive theme park that combines extreme experiences, leisure and entertainment, and science education. It has a wealth of entertainment projects, as well as float parades and acrobatic performances, which are very suitable for a full day of fun with children.
- Selected comment: Some people say that its tickets are cost-effective, although the rules are a bit complicated, but overall it is full of fun. Especially the atmosphere during Halloween is very good, it's worth a visit.

Kongzhou Art Museum

- Type: Art Museum
- Address: 694 Wanhong Road
- Reason for recommendation: This is a very artistic art museum that offers free exhibitions and a quiet environment, which is very suitable for enriching leisure time and experiencing the charm of art with children.
- Selected comment: Many people like its quiet environment and fresh painting style, it is a frequent spot for local residents.

Lanzhou Recommendations:

Baita Mountain Park

- Reason for recommendation: Although not mentioned in the original record, this is a paradise for outdoor enthusiasts. You can slowly climb to the top of the mountain along the mountain road and enjoy the beautiful scenery of Lanzhou. It is an excellent choice for relaxation and enjoying nature.
- Selected comment: It is usually a good place recommended by locals and tourists, providing a peaceful contrast to the city.

I hope these recommendations can add a little surprise to your journey, wish you a pleasant trip!

(a) Multi-Turn-Revise data sample

These places sound great, especially Goose Tianxiang and Baita Mountain Park, they should bring us a lot of fun! Thank you for your recommendations, I will consider these places carefully.



Chengdu and Lanzhou Travel Guide
Day 1: Tianjin to Chengdu (May 5, 2024)

In the morning sunshine, you will take the high-speed train G1704 from Tianjin to the beautiful Chengdu. The train departs on time at 06:45 and is expected to arrive at 17:22. The ticket price is 808 yuan. Along the way, you can enjoy the beautiful scenery and look forward to the upcoming Chengdu trip.

Upon arrival in Chengdu, you will stay at "Yunfu Li Yun Su-Xihu Yun Dao Homestay". This homestay is located at No. 1, 88 Linjiang Road, near Shaqu Park. The room is "Yunjian. Lake View Family Room", priced at 1491.11 yuan. The homestay has an elegant environment and charming lake view, making it an ideal place for you to relax.

!Yunfu Li Yun Su-Xihu Yun Dao Homestay]

Day 2: Fun in Chengdu (May 6, 2024)

On your first day in Chengdu, you can enjoy the charm of this city. At 9 o'clock in the morning, go to Sancha Lake Water Park, a water park suitable for the whole family. You can play in the water and enjoy the cool summer time.

!Sancha Lake Water Park]

At 12:30 noon, you can enjoy a delicious lunch at **Wen Yu Fresh Fish Village**, which is only 642 meters away from the water park. The average consumption per person is about 50 yuan, which is definitely worth a try.

!Wen Yu Fresh Fish Village]



At 3 pm, you will go to Anren Ancient Town, a place full of historical charm. You can stroll through the alleys of the ancient town and feel the strong cultural atmosphere.

!Anren Ancient Town]

At 5:30 in the evening, you can have dinner at Liu Xuewang, which is only 42 meters away from Anren Ancient Town. The average consumption per person is about 50 yuan, and the aroma of the hot pot will definitely make you drool.

!Liu Xuewang]

In the evening, you will return to "Yunfu Li Yun Su-Xihu Yun Dao Homestay" and enjoy a peaceful night.

Day 3: Exploring Chengdu (May 7, 2024)

Today's itinerary is equally exciting. At 9 o'clock in the morning, you will go to the Financial Expo Center, a place where modernity and tradition blend. You can feel the economic vitality of Chengdu here.

!Financial Expo Center]

At 12:30 noon, you can have lunch at Chui Er Ge Hot Pot (Taikoo Li Store), which is near the Financial Expo Center. The average consumption per person is about 55 yuan, and the deliciousness of the hot pot is irresistible.

!Chui Er Ge Hot Pot]

At 3 pm, you will go to Panda Pedestrian Street, one of the popular attractions in Chengdu. You can shop and taste delicious food here.

!Panda Pedestrian Street]

At 5:30 in the evening, you can have dinner at Dicos (Panda Base Scenic Area Store), which is only 26 meters away from Panda Pedestrian Street. The average consumption per person is about 41 yuan, and the convenience of fast food allows you to easily enjoy delicious food.

In the evening, you will return to "Yunfu Li Yun Su-Xihu Yun Dao Homestay" again to prepare for tomorrow's journey.



Day 4: Chengdu to Lanzhou (May 8, 2024)

Today, you will take the D168 train from Chengdu to Lanzhou. The train departs at 17:00 and is expected to arrive at 00:19. The ticket price is 213 yuan. Along the way, you can relax and look forward to the beautiful scenery of Lanzhou.

Upon arrival in Lanzhou, you will stay at "Gerei Preferred Hotel (Lanzhou Qilihe Army General Hospital Store)", which is located at 497 Xijin East Road, convenient for your travel.

Day 5: Cultural Tour of Lanzhou (May 9, 2024)

On your first day in Lanzhou, at 9 am, you will go to the Yellow River Building, an excellent place to overlook the Yellow River. The ticket price is 30 yuan, and you can enjoy the magnificent river scenery here.

!Yellow River Building]

At 11:30 noon, you can have lunch at Qingzhen Ru Hua Hand Grasping Restaurant (Lanzhou Main Store), which is only 295 meters away from the Yellow River Building. The average consumption per person is about 93 yuan, and the authentic Lanzhou cuisine will make you feast.

!Qingzhen Ru Hua Hand Grasping Restaurant]

At 1:30 in the afternoon, you will go to Xihu Mosque, an important religious place in Lanzhou, where you can feel the strong religious culture.

!Xihu Mosque]

In the evening, you will take the G1714 train and return to your departure place, ending this unforgettable journey.

I hope this travel guide can help your trip to Chengdu and Lanzhou, wish you a pleasant journey!

There will be moderate rain in Chengdu on May 7th, so it's not suitable to stroll around the Panda Pedestrian Street in the afternoon. I want to change to indoor activities, such as visiting a museum.



Chengdu and Lanzhou Travel Guide

Day 1: Tianjin to Chengdu (May 5, 2024)

In the early morning sunshine, you will take the high-speed train G1704 from Tianjin to the beautiful Chengdu. The train departs on time at 06:45 and is expected to arrive at 17:22. The ticket price is 808 yuan. Along the way, you can enjoy the beautiful scenery and look forward to the upcoming wonderful journey.

Upon arrival in Chengdu, you will stay at the Orange Crystal Hotel (Chengdu High-tech West District Electronic University Store). This is a comfortable hotel with a luxury king room with a view. The price is 589.36 yuan. The hotel is located at No. 2, Building 2, Unit 2, Heshun Road, which is convenient for your subsequent travel.

!Orange Crystal Hotel]

Day 2: Exploring Chengdu (May 6, 2024)

Today's itinerary will take you deep into the charm of Chengdu. At 9 o'clock in the morning, you will go to the Sancha Lake Water Park to enjoy the fresh lakeside air and water fun. The playtime is 3 hours and the admission is free.

!Sancha Lake Water Park]

At 12:30 noon, you can enjoy a delicious lunch at the Wenyu Fresh Fish Village not far from the water park. The average consumption is about 100 yuan.

!Wenyu Fresh Fish Village]

At 3 o'clock in the afternoon, you will go to Anren Ancient Town to feel the history and culture of this ancient town. The playtime is 2 hours and the admission is free.

!Anren Ancient Town]



At 5:30 in the evening, you can have dinner at Liu Xuewang in Anren Town. The average consumption is about 50 yuan, where you can taste authentic Sichuan cuisine.

!Liu Xuewang]

In the evening, you will return to the Orange Crystal Hotel to rest for the night and recharge for tomorrow's itinerary.

Day 3: Chengdu Cultural Tour (May 7, 2024)

Today's itinerary will take you to appreciate the culture and art of Chengdu. At 9 o'clock in the morning, you will go to the Financial Expo Center. The playtime is 3 hours and the admission is free.

!Financial Expo Center]

At 12:30 noon, you can have lunch at Chui Er Ge Hot Pot near the Financial Expo Center. The average consumption is about 55 yuan, where you can experience authentic Sichuan hot pot.

!Chui Er Ge Hot Pot]

(b) Continuation of Multi-Turn-Revise data sample



(c) Continuation of Multi-Turn-Revise data sample

Figure 9: **Multi-Turn-Revise data sample.** User initially provided only limited information such as the destination and budget. The AI assistant then gradually collected other information through an assisted decision-making process. Once sufficient information was gathered, it provided a complete travel plan. However, the user suggested a modification to the first plan, specifically wanting to replace the Panda Pedestrian Street attraction. After responding to this request, the AI assistant provided a second complete plan. This process clearly demonstrates a procedure that simultaneously incorporates assisted decision-making and plan revision.



Figure 10: **Model Failure Case I.** This is a bad case selected from the generated travel plan. For easier observation, we have omitted the parts without errors.



Suzhou Five-Day Travel Guide

.....
Day 4: The Tranquil Beauty of Suzhou

On May 4th, you will visit Jingsi Garden - Qingyun Peak, enjoying the peaceful natural environment. The visiting time is from 9:00 to 11:00, and the ticket is free.

After the tour, you can have a vegetarian lunch at Yunshan Nizhan Tea Vegetarian Cuisine (Yunzhuan Temple Branch). The restaurant is only 143 meters away from the scenic area, offering you a fresh and healthy meal.

.....
Due to heavy rain on May 4th, the revised plan is as follows:

Suzhou Five-Day Travel Guide

.....
Day 4: Exploring Suzhou Gardens and Cuisine On May 4th, your exploration of Suzhou continues. At 9:00 in the morning, you will go to the famous Ouyuan Garden to experience the unique charm of Suzhou gardens.

At 11:15 noon, you can have lunch at Cuihua Restaurant in Zhengda Square, with an average per capita consumption of about 82 yuan.

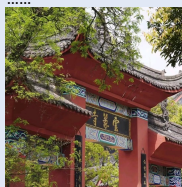


Figure 11: **Model Failure Case II.** This is a bad case selected from the generated travel plan. For easier observation, we have omitted the parts without errors.