

Competition 5: AI-Powered Intelligent Travel Planning

1. Competition Background

1.1 Industry Trends and Pain Points

In recent years, China's tourism market has continued to expand, with the number of domestic trips exceeding 4.5 billion in 2023. At the same time, the demand for personalized and customized travel has been growing rapidly. However, traditional travel planning tools generally suffer from three major pain points:

(1) Limited understanding of needs: Existing tools rely on fixed templates or simple tag-based filtering, which makes it difficult to deeply interpret complex user intents expressed in natural language (e.g., "I want to take my child to experience intangible cultural heritage, but with a limited budget and away from crowds").

(2) Insufficient dynamic adaptation: They cannot effectively integrate real-time data such as transportation, weather, or crowd control measures, resulting in low reliability of the generated plans.

(3) Lagging knowledge updates: Tourism information changes rapidly (e.g., shifts in popular attractions or policy adjustments), while traditional systems are costly to maintain and often fail to provide comprehensive coverage.

1.2 Academic Research Value

The breakthrough of large language models (LLMs) has brought new opportunities for travel planning, but also poses key challenges:

(1) Multi-source information integration and decision-making: It is necessary to integrate multi-dimensional information such as POI data, user profiles, and spatiotemporal constraints.

(2) Trustworthy and reliable travel plan generation: It is crucial to balance travel experience with feasibility (e.g., route rationality and time costs).

This competition will promote the application of agent-based technologies in complex real-world decision-making scenarios, foster the integration and innovation of large language models with neuro-symbolic systems and agent technologies, and provide an important paradigm for AI applications in vertical domains.

2. Competition Application Scenarios

1.2 Typical Application Scenarios

(1) Deeply Customized Independent Travel

1) User Requirement: "During the National Day holiday, I want to travel with my parents from Shanghai for 5 days, combining slow-paced visits to ancient towns and modern cities. My father has mobility issues and needs minimal walking."

2) Agent Response: Automatically avoid attractions with stairs, design a combined route including ancient towns (e.g., Wuzhen) and urban cultural landmarks (e.g., Shanghai Astronomy Museum), and recommend accessible transportation options.

(2) Immersive Local Cultural Experience

1) User Requirement: “Taking the kids on an educational trip to deeply learn about local history and culture, with a total budget not exceeding 8000 RMB.”

2) Agent Response: Match local museums, historic residences, and cultural sites, automatically calculate discounted ticket + accommodation packages, and dynamically adjust dining standards to control costs.

(3) Special Needs Response

1) User Requirement: “Next week, I have a business trip in Beijing with 3 meetings, and I want to try local specialty restaurants in between while handling emails in the evening.”

2) Agent Response: Use the meeting locations to intelligently cluster hotel and restaurant options, recommend efficient commuting routes, and schedule reservations for quick-access local specialty restaurants.

2.2 Technology-Scenario Mapping

(1) Multi-Source Travel Information Integration: Integrate multi-dimensional information such as attractions, restaurants, accommodations, geographic locations, opening hours, and average costs to generate travel plans.

(2) Neuro-Symbolic Decision Planning: Combine the LLM’s strength in understanding user intent with symbolic reasoning’s ability to ensure trustworthy and reliable decisions, enhancing the agent’s capability to satisfy user requirements while ensuring feasible and dependable travel plans.

Through this competition, participants will directly tackle complex constraints in real-world application scenarios, advancing conversational AI from “information retrieval” to “trustworthy decision-making,” and providing critical technological support for the digital transformation of the tourism industry.

3. Competition Information

This competition is organized by the expert team of the organizing committee. The hosts are Nanjing University and Huawei. The problem setters are: Guo Lanzhe, Li Yufeng, Shao Jiejing, Yang Xiaowen, Zhang Bowen, Han Siyu, Yu Kunyang, and Zhou Zhi.

4. Competition Task

This competition provides participants with a travel sandbox along with an

associated user requirement dataset.

The travel sandbox contains detailed information on travel resources in popular tourist destinations and cities across China.

The user requirement dataset includes synthetic requirements expressed in natural language as well as real requirements collected from surveys, aiming to capture the complex user needs present in actual travel scenarios.

4.1 Training Process

During the training phase, participants will design agent algorithms based on the provided sandbox environment and user requirement data, interacting with the sandbox to obtain travel-related information and ultimately generate travel plans that meet the specified requirements. Participants can fine-tune their algorithms and test models to optimize the planning results. In addition to the provided data, participants are allowed to synthesize their own data or use open-source internet data for model training. However, during the semifinals and finals, they must submit the corresponding data and provide a detailed explanation in the final technical report regarding the data synthesis methods or the sources of any externally annotated data.

4.2 Testing Process

During the testing phase, the system must accurately understand users' travel requirements through natural language interactions and automatically generate travel plans that satisfy all specified constraints. The evaluation of planning performance will use the constraint satisfaction rate as the core metric, with a focus on assessing the system's robustness in key challenges: the accuracy of fine-grained spatiotemporal planning, the ability to handle complex requirements under multiple constraints, and adaptability to individual travel preferences.

The ultimate goal is to build an intelligent travel planning system with high reliability and strong generalization capability in real-world scenarios.

5. Dataset and Data Description

The offline dataset provided in this competition is generated based on the travel sandbox. The preliminary round dataset is intended only for participants to verify and debug their algorithms, while the semi-final and final rounds provide entirely new datasets (with the same data format as the preliminary round). This design simulates the real-world challenge where "models deployed online need to handle new users with different demands," helping to identify algorithms that are more generalizable and reliable. Detailed information about the dataset is as follows:

The travel sandbox covers 10 well-known cities in China: Beijing, Chengdu,

Chongqing, Guangzhou, Hangzhou, Nanjing, Shanghai, Shenzhen, Suzhou, and Wuhan. It includes over 700 airline routes and more than 5,700 train routes between cities. For each city, the sandbox provides basic information for an average of 300+ attractions, 400+ restaurants, and 400+ hotels.

The dataset consists of both synthetic and real user data, with each entry containing a natural language query representing a user's travel request. Sample data can be accessed at: <https://box.nju.edu.cn/d/8941a6d3cddc4c7591e4/>.

6. Performance Metrics Requirements

The main evaluation metrics for this competition are requirement satisfaction rate and travel preference satisfaction. Participants are required to submit a policy model, which will be interactively tested by the evaluation server using the travel sandbox. The final score of the model will be determined based on the constraint satisfaction rate across all test data.

The evaluation metrics include the following items:

6.1.Environmental Constraints: This metric evaluates whether the generated travel plan is consistent with the information provided in the sandbox environment, measuring the feasibility of the plan.

$$EPR - micro = \frac{\sum_{p \in P} \sum_{c \in Env} 1_{passed(c, p)}}{|P| * |Env|}$$

$$EPR - macro = \frac{\sum_{p \in P} \prod_{c \in Env} 1_{passed(c, p)}}{|P|}$$

6.2.Conditional Logic Constraints: This metric evaluates, under the premise of satisfying environmental constraints, how well the generated travel plan meets the user's personalized requirements.

$$C - LPR = \frac{\sum_{p \in P} 1_{passed(Env, p)} \cdot \sum_{c \in C_p} 1_{passed(c, p)}}{\sum_{p \in P} |P|}$$

6.3.Final Constraint Satisfaction Rate: This metric represents the proportion of the generated travel plan that satisfies all environmental constraints and conditional logic constraints.

$$C - LPR = \frac{\sum_{p \in P} 1_{passed(Env, p)} \cdot \sum_{c \in C_p} 1_{passed(c, p)}}{\sum_{p \in P} |P|}$$

6.4. Preference Evaluation

In the TPC competition, we provide three common travel preference indicators:

Daily Average Attractions Visited (DAV): Maximize the number of attractions visited per day. The value is normalized to the range [0, 1] as the score.

$$DAV-score = (DAV - 0) / 4$$

Averaged Transportation Time (ATT): Minimize the average travel time per day. The value is normalized to the range [15, 120] minutes as the score.

$$ATT-score = \max(\min((120 - ATT) / (120 - 15), 1), 0)$$

Daily Dining Recommendations (DDR): Maximize the number of dining recommendations per day. The value is normalized to the range [0, 3] as the score.

$$DDR-score = \min((DDR - 0) / (3 - 0), 1)$$

6.5. Final Score:

Overall Score = 10% * EPR-micro + 10% * EPR-macro + 25% * C-LPR + 40% * FPR + 5% DAV-Score + 5% ATT-Score + 5% DDR-Score

7. Functional Requirements

7.1 Algorithm Design Requirements

Training Method: Model fine-tuning based on the provided data

7.2 Algorithm Optimization

(1) The travel planning assistant interacts with the travel sandbox environment and integrates the obtained information to provide practical and feasible travel plans that are consistent with the sandbox data.

(2) The travel planning assistant should accurately understand users' natural language requests to ensure that the final travel plans precisely meet the users' requirements.

(3) The travel planning assistant should possess strong generalization ability, capable of adapting to new users' novel travel demands.

8. Development Environment

Participants are required to develop using Python. It is recommended that they ensure their development environment is consistent with the provided testing environment to avoid runtime issues caused by environmental differences.

9. Evaluation Criteria

This competition is divided into three stages: preliminary, semi-final, and final:

9.1. Preliminary Stage: The preliminary stage provides a dataset solely for participants to validate and debug their algorithms. The results submitted at this stage will not count toward the final competition score.

9.2. Semi-Final Stage: A new dataset is provided for the semi-final (format consistent with the preliminary stage). Participants submit their algorithm models and technical reports. The organizers verify the algorithm outputs and assign a comprehensive score based on performance metrics as well as the completeness and quality of submitted materials.

9.3. Final Stage: The online final provides a validation dataset (format consistent with the preliminary and semi-final stages). The offline final determines the overall score based on both objective and subjective evaluation, with a weighting of 70% and 30%, respectively (for second- and third-place winners, the 30% subjective component does not apply).

(1) Objective Evaluation: Based on the machine-evaluated scores after standardization.

(2) Subjective Evaluation: Based on the standardized defense score, which considers participants' defense performance as well as the submitted technical proposals and code documentation.

10. Problem-solving approach

10.1. Language Model Prompt Tuning: Use prompt engineering and in-context learning with large language models (LLMs) to reason over natural language user requests, enhancing the travel planning assistant's comprehension of user needs.

10.2. Language Model Fine-Tuning: Further train the LLM to improve its understanding and interpretation of complex natural language queries.

10.3. Program Workflow Optimization: Optimize the program-level workflow to increase the efficiency of the travel planning assistant in generating solutions

Key Technical Challenges to Address:

1) Complex Natural Language Understanding: Design algorithms for understanding diverse and complex user expressions, fine-tune the language model, and improve the extraction of formalized constraints from natural language inputs.

2) Constraint-Based Plan Correction: Using a symbolic verifier, design plan correction algorithms that leverage unsatisfied constraint feedback to guide the language model in adjusting and improving the plan.

3) sNeuro-Symbolic System Efficiency: Address the low solving efficiency of neuro-symbolic systems by designing optimized workflows to accelerate solution

generation.

11. Reference Resources

Travel planning and large language model (LLM) related papers and open-source repositories, for example:

ChinaTravel: A Real-World Benchmark for Language Agents in Chinese Travel Planning

To the Globe (TTG) : Towards Language-Driven Guaranteed Travel Planning

TravelPlanner: A Benchmark for Real-World Planning with Language Agents

ITINERA: Integrating Spatial Optimization with Large Language Models for Open-domain Urban Itinerary Planning

12. Submission Requirements

The competition requires participants to submit their algorithm inference code and model within the provided testing framework. The test server will use the submitted model to interact with the travel sandbox and generate a batch of planning results for the dataset. The final score of the model will be calculated as the average constraint satisfaction rate of this batch of generated plans.

Participants should package all files into a single .zip file following the testing code standards. During development, participants can run the evaluation scripts and interact with the travel sandbox provided by the test framework to verify whether their strategy implementation has any issues.

Resource and Format Constraints:

The inference time for a single data sample must not exceed 5 minutes; exceeding this limit will result in failure.

13. Updates and Q&A

The competition tasks may be updated. To address questions encountered by participants during the contest, a dedicated Q&A group will be established (accessible to participants after official registration).

14. Competition Process and Award Settings

14.1 Registration Phase

Participants register on the official competition website, submit personal or team information, and obtain the download link for the preliminary round dataset.

14.2 Preliminary Round Phase

Participants use the training datasets provided by the organizers to design their algorithmic models and submit results on the test set. The organizers then evaluate the model scores based on these results.

14.3 Provincial Semifinal Phase

After the preliminary round ends, the competition progresses to the semi-final stage, at which point the semi-final dataset download link is made available. Only teams that submitted valid results during the preliminary round are eligible to enter the semi-finals. During this stage, participants use the provided semi-final data to fine-tune and adjust their algorithmic models, and submit their models along with a technical report before the semi-final deadline. The organizers then verify the algorithm's execution results and assign comprehensive scores based on performance evaluation metrics, as well as the completeness and quality of the submitted materials.

14.4 Provincial Semifinal Results Announcement

The semi-final results are published on the official competition website. The number of teams that advance to the semi-finals serves as the base for award calculation. According to the provincial competition's award proportion limits, first, second, and third prizes are awarded (with provincial-level award certificates). During the semi-final award evaluation, any submission whose algorithm performance falls below the baseline reference score provided by the organizers is deemed invalid and will not receive an award. Teams that win first or second prize in the semi-finals advance to the national final.

14.5 National Final Phase

(1) Online Evaluation of the Final: The algorithm models submitted by teams advancing from the semi-finals are evaluated and scored based on the national final validation dataset. The number of teams entering the final serves as the award base. According to the award proportion limits of the national competition, the candidate list for the first prize and the winners of the second and third prizes are determined (with national second- and third-prize certificates issued).

(2) Final Submission: Candidate teams for the national first prize must submit complete technical documents, algorithm code and model files, demonstration videos, and supplementary materials within the specified timeframe. No modifications or additions will be accepted after the submission deadline.

(3) Final Review Stage: A professional review team evaluates the submissions of national first prize candidates, scoring them comprehensively based on performance metrics and the completeness and quality of submitted materials. Reviewers may request explanations from participants if any questions arise during evaluation.

(4) Offline Final Evaluation: Candidate teams for the national first prize submit their finalized technical documents, algorithm code and model files, demonstration

videos, and supplementary materials, and participate in the on-site final review and defense. The final winners and rankings of the national first prize are determined based on this evaluation (teams that do not participate in the on-site review are considered to have forfeited the award). Honor certificates are awarded to national first prize winners.

15. Prize Setting

To encourage participants' engagement, this competition awards cash prizes to the top five teams based on the final competition results:

- (1) Champion: 1st place, prize of ¥5,000 per team
- (2) Runner-up: 2nd place, prize of ¥2,000 per team
- (3) Third Place: 3rd place, prize of ¥1,000 per team
- (4) Excellence Award: 4th–5th place, prize of ¥500 per team

16. Other Instructions

(1) Fairness: Any form of cheating is strictly prohibited, including but not limited to data leakage, overlap between pretraining data and test data, and plagiarism of others' code. Once detected, the participant's qualification will be immediately revoked, and relevant responsibilities will be pursued.

(2) Intellectual Property: Submitted works must be original and must not have won awards or been publicly published in other competitions. The organizers have the right to display and promote the submitted works, but the intellectual property rights remain with the participants.

17. Contact Information

Competition communication QQ group: 768721120

Email: guolz@nju.edu.cn

Registration website: www.aicomp.cn