

SMAP: Semantic Route Planning with Map-Grounded Multimodal Alignment

Supplementary Material

A. Limitations

While SMAP shows strong performance across multimodal route planning tasks, several practical considerations suggest natural directions for future work. First, our dataset, though carefully curated and geographically grounded, focuses on major urban regions; expanding to a broader set of cities and terrain types would further test generalization. Second, the current framework adopts a general-purpose routing formulation, and future work could incorporate user-specified planning criteria such as preferred transportation modes or budget-aware constraints to enable more tailored itineraries. Finally, as with most multimodal systems, user-specific contextual information (*e.g.* long-term travel preferences) is not modeled, and incorporating lightweight personalization could further improve real-world usability.

B. Ethical Statement

This work involves ethical considerations related to bias, safety, and responsible use. The training data is constructed from publicly available queries and LLM-generated examples, with no use of private user trajectories or personally identifiable information. It reflects real-world travel behaviors, which may overrepresent popular landmarks and underrepresent marginalized or lesser-known sites; this risks reinforcing existing geographic and cultural biases. While our hybrid data construction aims to improve diversity, the potential for bias amplification remains and warrants careful monitoring.

The system is intended solely as a route suggestion tool for planning purposes and is not designed for real-time navigation or safety-critical decision-making. Users should exercise their own judgment and consult official local information when following any generated route. Finally, the framework must not be used to plan access to prohibited, restricted, or hazardous locations.

C. Dataset Details

The MM-Route dataset comprises 3,000 natural language queries, formed by combining de-duplicated real user queries collected from public travel platforms with additional queries generated using GPT-4o. To broaden geographic and thematic coverage beyond popular destinations, we curate a collection of publicly available travel itineraries and descriptive documents from the open web. Each document typically includes a title and detailed route descriptions covering points of interest, transportation options, and practical tips.

We use these documents as grounding sources for LLM-based query generation with two prompting strategies:

Simple queries (typically fewer than 10 words). For these, we provide only the document title as input to GPT-4o and instruct it to generate concise, high-level requests (*e.g.* “Shanghai one-day tour”). These queries capture broad user intents and often correspond to common search phrases.

Complex queries. For more nuanced and context-dependent requests, we feed both the full title and the entire document body into GPT-4o. The model is prompted to generate detailed, multifaceted questions that require deeper understanding of the content (*e.g.* “Suggest a family-friendly cycling route in Shanghai that avoids busy roads”). These queries often involve constraints, preferences, or implicit reasoning based on the document.

In all cases, we constrain the model to generate queries whose answers are fully supported by information explicitly stated or reasonably inferable from the source document. This ensures factual grounding and minimizes the risk of hallucination. After query generation, we apply rigorous filtering to remove ambiguous, unanswerable, or off-topic queries. The final dataset merges filtered LLM-generated queries with authentic user queries, resulting in 3,000 high-quality, multimodal instances.

Additionally, we provide a dataset demo consisting of 100 annotated examples in the supplementary materials to illustrate the structure and diversity of MM-Route. Upon acceptance of this paper, we will publicly release the full dataset along with the complete training and inference code to ensure reproducibility and support future research by the community.

D. Agreement Between LLM and Human Evaluations

To assess the reliability of using Gemini-2.5-Pro as a proxy evaluator, we randomly sampled 100 model-generated route descriptions and measured the alignment between Gemini’s judgments and human annotators across five evaluation metrics. Agreement measures were selected based on the data type and semantic characteristics of each metric, following standard practices in inter-annotator agreement and correlation analysis.

For CPR, RTPR, and TSPR, the judgments are binary (*e.g.* pass/fail or Model A vs. Model B). Since these variables are nominal with two discrete outcomes, rank-based correlation measures such as Kendall Tau and Spearman Coefficient are inappropriate, as they assume at least

Table 4. The alignment performance between LLM and human evaluators across three metrics (%).

Metric	Agreement Rate	Kendall Tau	Spearman Coefficient
CPR	94.44	–	–
RTPR	99.00	–	–
TSPR	97.00	–	–
ORS	–	71.53	79.81
SHR	–	79.50	82.14

ordinal-scale data. In these cases, the agreement rate, defined as the proportion of samples where Gemini and human annotators provide identical labels, is used as a direct and interpretable measure of concordance.

ORS and SHR involve ordered or quasi-continuous assessments, where rank correlations are suitable. ORS is rated on a 5-point Likert scale (5 = excellent, 1 = poor), reflecting graded route quality. SHR quantifies spatial faithfulness as the ratio of spatial reasoning errors to the total number of spatial assertions. Since the number of assertions varies across routes due to differences in route length and complexity, raw error counts may not directly reflect relative spatial accuracy.

To address this, SHR agreement is evaluated from two complementary perspectives: (1) the absolute number of errors (lower is better), and (2) the absolute number of correct spatial assertions (higher is better). For each perspective, both Kendall Tau and Spearman Coefficient are computed between Gemini and human scores. The mean of the two correlation coefficients is reported as a symmetric, length-invariant measure of agreement, mitigating bias introduced by variable assertion counts.

Across all five metrics, strong alignment is observed between Gemini and human evaluations (see Tab. 4), supporting the use of LLM-as-a-judge as a reliable and scalable proxy for human assessment.

E. Evaluator Robustness Analysis: Gemini vs. GPT-5

To investigate potential biases introduced by using Gemini-2.5-Pro as the automatic evaluator, especially given its training on Gemini-generated data, we performed a comparative study of human preferences between Gemini-2.5-Pro and GPT-5. This experiment aimed to validate whether Gemini’s evaluations more closely align with human judgments.

We utilized the 100 route descriptions from Section D and had both models evaluate them independently across five metrics: CPR, ORS, RTPR, TSPR, and SHR, following identical evaluation guidelines. Human annotators then compared the outputs of the two models for each query and metric, selecting the one that better matched their own judg-

Table 5. Human preference rates (HPR) for Gemini-2.5-Pro vs. GPT-5 across five evaluation metrics (%).

Metric	CPR	ORS	RTPR	TSPR	SHR
HPR	75.28	78.35	86.60	79.38	72.31

ment.

The results, summarized in Table 5, indicate a consistent preference for Gemini’s evaluations over those of GPT-5 across all metrics, with an average preference rate of 78.38%. This finding suggests that Gemini’s superior alignment with human judgments, previously documented in Section D, is not merely an artifact of training data but reflects its enhanced capability as an evaluative tool. These outcomes strongly support the validity of our primary evaluation protocol and mitigate concerns regarding evaluator-induced bias.

F. Impact of Retrieved POI Count

To investigate the sensitivity of the SMAP framework to the number of retrieved POIs and to justify our selection of the 20-POI setting, we conduct an ablation study across different candidate pool sizes. We evaluate both Qwen2.5-VL-7B-HDPO and Qwen2.5-VL-32B-HDPO models with the number of retrieved POIs set to 10, 20, and 40. This decoupling of retrieval from planning allows for a rigorous assessment of the model’s inherent spatial reasoning capabilities and its ability to handle varying levels of information density and noise.

Table 6. Impact of Retrieved POI Count on Model Performance

Count	PSR↑	RDR↑	RTPR↑	TSPR↑	SHR↓	ORS↑
Qwen2.5-VL-7B-HDPO						
10	100	0.837	82.3	70.0	16.3	3.24
20	100	0.802	85.4	70.4	20.4	3.37
40	100	0.740	80.8	69.2	22.6	3.25
Qwen2.5-VL-32B-HDPO						
10	100	0.845	85.9	74.4	13.7	3.76
20	100	0.831	91.0	77.5	14.0	3.89
40	100	0.755	85.9	70.9	15.1	3.69

As presented in Table 6, the 20-POI setting achieves the optimal balance, reaching the highest ORS for both model scales. Our data indicates that typical itineraries usually consist of 5 to 10 POIs. While the 10-POI setting benefits from reduced task complexity and a lower risk of spatial hallucinations, it may fail to retrieve all necessary POIs required by a specific query due to the limited number of candidates. Conversely, the 40-POI setting introduces a significant amount of irrelevant POI noise, which can interfere with the model’s planning process and lead to decreased performance in metrics such as RDR. These find-

ings demonstrate that our current 20-POI setting provides a rigorous benchmark for assessing how MLLMs handle spatial constraints while effectively avoiding hallucinations.

G. Examples of Preference Pair

Fig. 3 shows the input map tile rendered for the query “Qinzhou retro film shooting route”, with retrieved candidate POIs overlaid. Based on this map tile and the corresponding POI metadata, we construct preference pairs for HDPO training. The rejected route exhibits geographic inefficiency, as it contains unnecessary detours and backtracking that violate the principle of path continuity. In contrast, the accepted route follows a direct and walkable sequence that respects spatial adjacency, as verified by cross-referencing the map layout with the POI metadata.

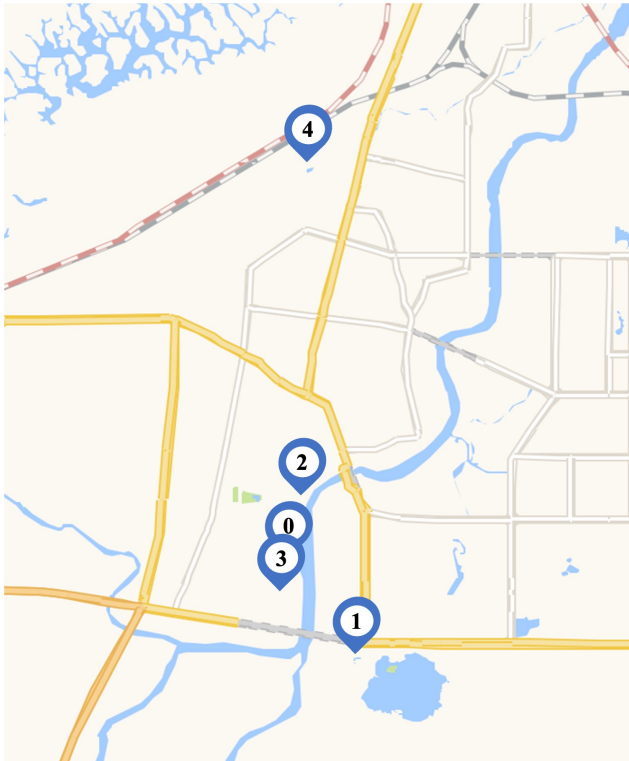


Figure 3. Input map tile rendered for the query “Qinzhou retro film shooting route”.

Query: Qinzhou retro film shooting route

Rejected (draft) response:

```
think:
The user wants a "Qinzhou retro film shooting route".

1. Analyze User Requirements: The key elements are "Qinzhou", "retro",
```

and "film shooting". This means I need to select attractions that are visually appealing, have a historical or nostalgic atmosphere, and are suitable for photography.

2. **Evaluate Candidate Attractions against Requirements:**
 - * `[0]` Wenlan Jiqing Food Court': Tags like 'Photo Check-in' and 'Intangible Cultural Heritage Cuisine' make it highly suitable for a film shoot, especially one with a retro food theme. It's a good fit.
 - * `[1]` Hexie Park - Harmonious Tower Sightseeing Platform': The tag 'Song-style architecture' is a perfect match for a "retro" theme. It's also great for night photography. This is a strong candidate.
 - * `[2]` Qinzhou Song Dynasty City Wall Relic Site Park': The name itself indicates a strong historical and retro theme ('Song Dynasty'). It's a must-visit for this route.
 - * `[3]` Liu Yongfu Former Residence': Liu Yongfu was a historical figure from Qinzhou. A former residence is inherently a historical and retro site, ideal for film shoots. This is another must-visit.
 - * `[4]` Feng Zicai Former Residence': Similar to the Liu Yongfu residence, this is a historical site related to a famous local figure. It fits the theme well.
3. **Select a Subset of Attractions:** All five candidates are thematically appropriate for a "retro film shooting" route. They offer a mix of historical sites, architectural landmarks, and cultural experiences. Therefore, I will include all of them to create a comprehensive itinerary.
4. **Plan a Logical Route using the Map:**

Now I need to arrange the selected attractions (`'0'`, `'1'`, `'2'`, `'3'`, `'4'`) into a logical sequence based on their geographical locations.

 - * Looking at the map, attractions `'0'`, `'3'`, and `'2'` are clustered together in the central area of the city.

* '3' (Liu Yongfu Former Residence) is the westernmost point in this cluster.

* '0' (Wenlan Jiqing Food Court) is located just east of '3'.

* '2' (Qinzhou Song Dynasty City Wall Relic Site Park) is located north of '0' and '3'.

* Attraction '4' (Feng Zicai Former Residence) is located significantly to the north of this central cluster.

* Attraction '1' (Hexie Park) is located to the east of the central cluster.

A logical route would minimize travel distance and avoid backtracking. A path that flows smoothly through the area is preferable.

5. **Construct the Itinerary:**
A good strategy is to start at the northernmost point and work your way south and then east.

* **Start at '[4] Feng Zicai Former Residence':** It's the northernmost attraction, making it a good starting point.

* **Proceed to '[2] Qinzhou Song Dynasty City Wall Relic Site Park':** From '4', travel south to visit '2'. This is a major historical site.

* **Continue to '[3] Liu Yongfu Former Residence':** From '2', move southeast to '3'. This is a short and direct trip.

* **Visit '[0] Wenlan Jiqing Food Court':** From '3', head east to '0'. This is a great spot for a break, lunch, or dinner, and its 'Photo Check-in' tag makes it perfect for the theme.

* **End at '[1] Hexie Park':** From '0', travel northeast to '1'. This provides a nice contrast to the historical sites with its modern architecture and night photography opportunities.

This route '4 -> 2 -> 3 -> 0 -> 1' is geographically efficient. It starts in the north, moves south through the main cluster, and then heads east to conclude the tour. This minimizes backtracking and

creates a smooth flow for the day.

6. **Final Decision:** The planned route is '[4, 2, 3, 0, 1]'.

answer:

[4, 2, 3, 0, 1]



Figure 4. Rejected (draft) route.

Accepted (Refined) response

think:

The user wants a "Qinzhou retro film shooting route". The goal is to select suitable attractions and arrange them into a logical and efficient tour.

- Analyze User Requirements:** The keywords are "Qinzhou," "retro," and "film shooting." This implies a need for locations with historical character, nostalgic ambiance, and strong visual appeal for photography.
- Evaluate Candidate Attractions:**
* '[0] Wenlan Jiqing Food Court': With tags like 'Photo Check-in' and

'Intangible Cultural Heritage Cuisine', this spot is perfect for capturing the retro vibe of local life and food culture.

- * '[1] Hexie Park - Harmonious Tower Sightseeing Platform': The 'Song-style architecture' is a direct fit for the "retro" theme. The 'Night Photography' tag also offers a great opportunity to end the day with dramatic shots.
- * '[2] Qinzhou Song Dynasty City Wall Relic Site Park': The name itself, 'Song Dynasty City Wall', screams history and retro. It's a must-include for historical exploration and photography.
- * '[3] Liu Yongfu Former Residence': Former residences of historical figures are inherently retro and provide an authentic backdrop for film-style shooting.
- * '[4] Feng Zicai Former Residence': Similar to Liu Yongfu's residence, this is another key historical site that perfectly matches the theme.

All five attractions are highly relevant and contribute different facets to the "retro film shooting" theme. Therefore, it's best to include all of them in the itinerary.

3. ****Plan a Logical Route using the Map:**** The next step is to arrange these five attractions into a geographically sensible sequence to minimize travel time and create a smooth flow.
 - * ****Analyze Locations:****
 - * Attraction '4' (Feng Zicai Former Residence) is the northernmost point.
 - * Attractions '2', '0', and '3' form a dense cluster in the central area. Within this cluster, '2' is the most northern, '0' is just south of '2', and '3' is to the southwest of '0'.
 - * Attraction '1' (Hexie Park) is the southernmost point, located to the southeast of the central cluster.
 - * ****Construct the Itinerary:**** A logical route should generally follow a single direction, such as

north to south, to avoid backtracking.

- * ****Start at '[4] Feng Zicai Former Residence':** Being the most northern and somewhat separate attraction, it makes for a perfect starting point.
- * ****Proceed to '[2] Qinzhou Song Dynasty City Wall Relic Site Park':** From '4', travel south to the first stop in the central cluster, '2'.
- * ****Continue to '[0] Wenlan Jiqing Food Court':** From '2', it's a very short and direct trip south to '0'. They are located very close to each other along a north-south line. This is an ideal spot for a lunch break and some vibrant street-style photography.
- * ****Next, visit '[3] Liu Yongfu Former Residence':** From '0', go southwest to '3'. This keeps the travel within the central cluster efficient.
- * ****End at '[1] Hexie Park - Harmonious Tower Sightseeing Platform':** From '3', travel southeast to the final destination, '1'. Its southern location makes it a natural endpoint, and its suitability for night photography provides a perfect conclusion to a day of shooting.

4. ****Final Decision:**** The route '[4, 2, 0, 3, 1]' is the most efficient. It starts in the north, moves progressively south through the central cluster with minimal deviation, and finishes at the southernmost point. This creates a smooth, logical, and time-saving tour.

answer:
[4, 2, 0, 3, 1]

H. Response Examples

Fig. 7 and Fig. 8 show the itineraries and route visualizations generated by the SFT and HDPO models, respectively, for the query "Qinzhou retro film shooting route". The input map tile are shown in Fig. 6. The SFT model produces a geographically inconsistent route that includes hallucinated transitions and inefficient backtracking. In contrast,



Figure 5. Accepted (Refined) route.

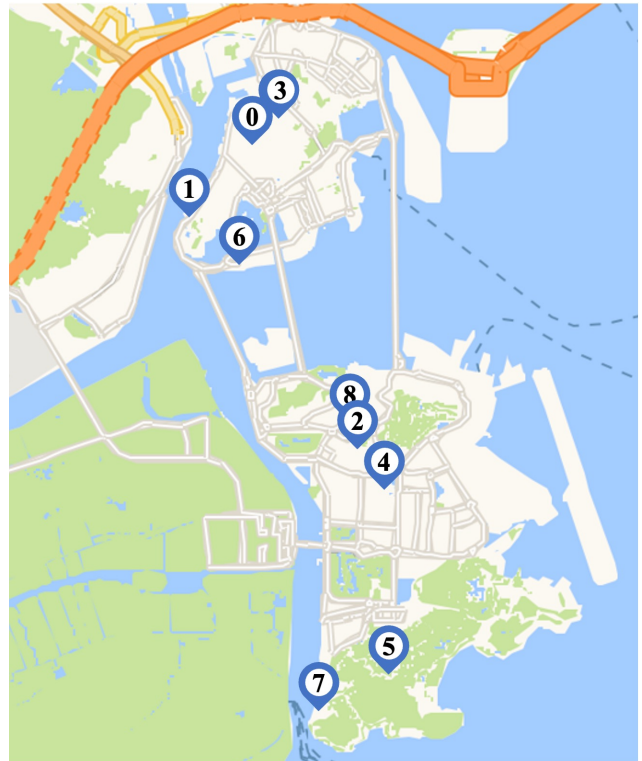


Figure 6. Input map tile rendered for the query “Must-visit itinerary in Macau”.

the HDPO model generates a spatially coherent itinerary with improved path continuity and reduced detours, demonstrating enhanced spatial reasoning capability.

Query: Must-visit itinerary in Macau

SFT model response:

```
think:
The user wants a "Must-visit itinerary in Macau". This is a general request , so I need to create a comprehensive one-day tour that covers the most famous and diverse attractions.

**1. Analyze Candidate Attractions:**
- **[0] Jinsha Site Museum:** A museum on the Macau Peninsula. Good for culture/history.
- **[1] Barra Miu Cin Sightseeing Boat Wharf:** A boat tour wharf on the Macau Peninsula. Offers a unique perspective of the city.
- **[2] Rua do Cunha:** A street in Taipa known for food and shopping. Represents the local lifestyle.
- **[3] Macau Peninsula:** This is a large area, not a single point. It's
```

```

better to select specific attractions within it.
- **[4] OnDakLou Cake Shop...:** A famous egg tart shop inside the Venetian. Represents iconic Macau cuisine.
- **[5] Road WaanDip SekTong Saan:** A hiking spot with a cultural village. More niche than the main tourist spots.
- **[6] Macau Tower:** A major landmark with views and activities like bungee jumping. A definite "must-visit".
- **[7] Road WaanHong Li:** A scenic walkway. Less famous than other options.
- **[8] Taipa:** Another large area, similar to [3]. I'll select specific points within it.

**2. Select Key "Must-Visit" Attractions :**
For a classic Macau tour, I should include:
- A major landmark with great views:
  **[6] Macau Tower**.
```

- A famous historical site: ****[0] Jinsha Site Museum**** (Senado Square area).
- A modern, iconic resort: The Venetian, which can be represented by ****[4] OnDakLou Cake Shop****.
- A traditional Portuguese street: ****[2] Rua do Cunha****.
- A way to see the city from the water: ****[1] Barra Miu Cin Sightseeing Boat Wharf****.

This selection provides a good mix of history, modernity, food, and sightseeing.

****3. Plan the Route based on Geography :****

Now I will arrange these selected attractions ([0], [1], [2], [4], [6]) into a logical sequence using the provided map.

- ****Location Analysis:****
 - ****Macau Peninsula (North):**** [0] Jinsha Site Museum, [1] Barra Miu Cin Sightseeing Boat Wharf, [6] Macau Tower.
 - ****Taipa (South):**** [2] Rua do Cunha, [4] OnDakLou Cake Shop.
 - ****Route Planning:****
- A logical flow would be to travel from north to south to minimize backtracking.
- **Start on the Macau Peninsula:**** Begin the day here.
 - ****[6] Macau Tower:**** It's best to visit in the morning to avoid crowds and get the best light for photos. Duration: ~3 hours.
 - ****[0] Jinsha Site Museum:**** After the tower, explore the nearby historic center. It's a short walk from the Macau Tower area. Duration: ~1 hour.
 - ****[1] Barra Miu Cin Sightseeing Boat Wharf:**** Located near the museum area, this is a perfect afternoon activity to relax and see the city from the water. Duration: ~2 hours.
 - **Cross to Taipa Island:**** Take the Macau Tower or Taipa cable car, or a short taxi ride.
 - ****[4] OnDakLou Cake Shop:**** Start the Taipa exploration with a famous egg tart. It's located in the bustling Venetian complex. Duration: ~1 hour.

- ****[2] Rua do Cunha:**** From the Venetian, it's a short walk to this street. It's a great place to end the day, strolling, shopping, and finding dinner. Duration: ~2 hours.

****4. Final Itinerary and Justification :****

The proposed route is: ****[6] -> [0] -> [1] -> [4] -> [2]****.

- ****Geographical Logic:**** The route starts on the northern peninsula ([6], [0], [1]) and moves south to Taipa ([4], [2]). This is efficient.
- ****Thematic Flow:**** The day progresses from a high-energy landmark ([6]), to quiet historical exploration ([0]), to a relaxing boat tour ([1]), followed by a taste of iconic food ([4]), and ending with a lively street experience ([2]).
- ****Time Management:**** The total duration is approximately 3 + 1 + 2 + 1 + 2 = 9 hours, which is ideal for a full-day tour, leaving some buffer for travel between spots.

This itinerary is well-rounded, covers the most important aspects of Macau, and follows a logical path. I will exclude the other attractions as they are either too time-consuming ([3], [8]), less central to a first-time visitor's experience ([5], [7]), or are covered by the chosen spots ([4] is in [8]).

answer:

[6, 0, 1, 4, 2]

HDPO response:

think:

The user wants a "Must-visit itinerary in Macau". This is a general request, so the goal is to create a classic and comprehensive one-day tour that covers the most famous and diverse attractions, balancing historical sites, modern landmarks, food, and entertainment.

****1. Analyze Candidate Attractions:**** First, I'll evaluate the provided list of attractions to select the most suitable ones for a "must-visit" tour.



Figure 7. The route generate by SFT model.

- "[0] Jinsha Site Museum": Represents the historical and cultural heart of Macau, located on the Macau Peninsula. A great choice for cultural exploration.
- "[1] Barra Miu Cin Sightseeing Boat Wharf": Offers a unique perspective of the city from the water. It's located near the A-Ma Temple, a key historical site on the peninsula.
- "[2] Rua do Cunha": A famous food and souvenir street in Taipa Village. Essential for experiencing local flavors and atmosphere.
- "[3] Macau Peninsula" \& "[8] Taipa": These are large areas, not specific attractions. It's better to select key points within them.
- "[4] OnDakLou Cake Shop and Coffee Room (Venetian Shopping Mall)": Represents the iconic Macau egg tart and is located within the world-famous Venetian Macao, a must-see modern resort.
- "[5] Road WaanDip SekTong Saan" \& "[7] Road WaanHong Li": These are hiking and nature spots in Coloane. While beautiful, they are less

central to a typical first-time, one-day "must-visit" tour which focuses on the main highlights of the Peninsula and Taipa.

- "[6] Macau Tower": An iconic landmark of the Macau skyline, offering breathtaking views and thrilling activities. Definitely a "must-visit".

****2. Select Key Attractions for the Itinerary:****

Based on the analysis, I will select a subset that provides a well-rounded experience.

- For history and culture: "[0] Jinsha Site Museum".
- For iconic landmarks and views: "[6] Macau Tower".
- For unique sightseeing: "[1] Barra Miu Cin Sightseeing Boat Wharf".
- For modern glamour and iconic food: "[4] OnDakLou Cake Shop" (representing the Venetian).
- For local food and shopping: "[2] Rua do Cunha".

This selection ("[0], [1], [2], [4], [6]") covers the essential facets of Macau.

****3. Plan the Route based on Geography and Logic:****

Now, I will arrange these selected attractions into a logical sequence using their locations on the map to create an efficient tour.

- ****Location Analysis:****

- ****Macau Peninsula (North/Central part of the map):**** "[0] Jinsha Site Museum", "[1] Barra Miu Cin Sightseeing Boat Wharf", "[6] Macau Tower".
- ****Taipa/Cotai (South/Central part of the map):**** "[2] Rua do Cunha", "[4] OnDakLou Cake Shop".

- ****Route Planning Strategy:****

The most efficient way to tour Macau is to explore one major area (Macau Peninsula) completely before crossing the bridge to the other (Taipa/Cotai). This minimizes travel time. A common and logical flow is to start with the historical sites on the peninsula in the morning and move to the modern entertainment hub of Taipa/Cotai in the afternoon and

evening.

1. **Morning on the Macau Peninsula:**
 - Start at **[0] Jinsha Site Museum**. It's located in the heart of the historic district, making it an excellent starting point to soak in the city's culture.
 - From there, head south towards the waterfront to **[1] Barra Miu Cin Sightseeing Boat Wharf**. This area is also historically significant (near A-Ma Temple). A boat tour provides a relaxing break and a different view of the city.
 - After the boat tour, proceed to **[6] Macau Tower**. It is located south of the historic center and serves as a perfect final stop on the peninsula, offering panoramic views before heading to Taipa.

2. **Afternoon/Evening in Taipa/Cotai:**
 - Take a taxi or bus from the Macau Tower across the bridge to Taipa. A great first stop is **[4] OnDakLou Cake Shop** inside the Venetian. This allows you to experience the grandeur of the Cotai Strip resorts and taste the famous Portuguese egg tarts.
 - From the Venetian, it's a short and pleasant walk (often via a moving walkway) to **[2] Rua do Cunha** in Old Taipa Village. This is the perfect place to end the day. You can stroll through the lively street, sample more local snacks, shop for souvenirs, and find a great spot for dinner, contrasting the modern glitz of the Venetian with traditional village life.

4. Final Itinerary and Justification:

The resulting itinerary is a logical progression from north to south, and from historical to modern.

- **Final Route:** "[0] -> [1] -> [6] -> [4] -> [2]"
- **Geographical Flow:** The route starts on the Macau Peninsula, moving from the northern historic center ("**[0]**") southward to the waterfront ("**[1]**") and the tower ("**[6]**"). It then crosses to Taipa,

visiting the modern Cotai Strip ("**[4]**") and the adjacent Old Taipa Village ("**[2]**"). This path is highly efficient and avoids backtracking.

- **Thematic Flow:** The day begins with cultural and historical exploration, transitions to scenic views, and concludes with world-class entertainment, food, and shopping.
- **Time Management:** The total estimated duration is $1 + 2 + 3 + 1 + 2 = 9$ hours. This fits perfectly into a full day of touring, with buffer time for travel and meals.

answer:

[0, 1, 6, 4, 2]



Figure 8. The route generate by HDPO model.

I. Prompt

I.1. Prompt for answer generation

```
# Task Description
You are a trip planning expert,
responsible for creating itineraries
for users. Based on our retrieved
```

```

    set of candidate attractions, select
    a subset that fits the user's
    requirements and plan a logical tour
    route.
User's requirements:
{USER QUERY}

# Candidate Attractions
Based on the user's requirements, we
have retrieved some attractions for
you to choose from:
Each attraction includes three pieces of
information: 'name' is the
attraction's name; 'tag' contains
some representative tags for the
attraction; 'duration' is the
recommended visiting time for the
attraction (unit: hours).
{POI INFO}

# Output Format
The output format is: '<think>...</think
> <answer>...</answer>'
Enclose your thinking process within '<
think>' and '</think>'.
Enclose your planned result within '<
answer>' and '</answer>'. The format
should be a list where each element
is the index of an attraction. The
order of attractions in the list
represents the sequence of your
planned route.

# Scatter Plot Distribution
The location distribution of the
attractions is shown in the figure.
Each black dot represents an
attraction, and the number next to
it is its index (starting from 0) in
the candidate attractions list. The
orientation of the map is: North is
up, South is down, West is left,
and East is right.
{MAP IMAGE}

```

I.2. Prompt for answer generation (text-only)

```

# Task Description
You are a trip planning expert,
responsible for creating itineraries
for users. Based on our retrieved
set of candidate attractions, select
a subset that fits the user's
requirements and plan a logical tour
route.
User's requirements:
{USER QUERY}

```

```

# Candidate Attractions
Based on the user's requirements, we
have retrieved some attractions for
you to choose from:
Each attraction includes three pieces of
information: 'name' is the
attraction's name; 'coor' is the
attraction's coordinate; 'tag'
contains some representative tags
for the attraction; 'duration' is
the recommended visiting time for
the attraction (unit: hours).
{POI INFO}

# Output Format
The output format is: '<think>...</think
> <answer>...</answer>'
Enclose your thinking process within '<
think>' and '</think>'.
Enclose your planned result within '<
answer>' and '</answer>'. The format
should be a list where each element
is the index of an attraction. The
order of attractions in the list
represents the sequence of your
planned route.

```

I.3. Prompt for answer generation (fewshot)

```

# Task Description
You are a trip planning expert,
responsible for creating itineraries
for users. Based on our retrieved
set of candidate attractions, select
a subset that fits the user's
requirements and plan a logical tour
route.
User's requirements:
{USER QUERY}

# Candidate Attractions
Based on the user's requirements, we
have retrieved some attractions for
you to choose from:
Each attraction includes three pieces of
information: 'name' is the
attraction's name; 'tag' contains
some representative tags for the
attraction; 'duration' is the
recommended visiting time for the
attraction (unit: hours).
{POI INFO}

# Output Format
The output format is: '<think>...</think
> <answer>...</answer>'
Enclose your thinking process within '<
think>' and '</think>'.

```

Enclose your planned result within '`<answer>`' and '`</answer>`'. The format should be a list where each element is the index of an attraction. The order of attractions in the list represents the sequence of your planned route.

Output example (example format only; content is irrelevant):

```
{EXAMPLE}
```

```
# Scatter Plot Distribution
```

The location distribution of the attractions is shown in the figure. Each black dot represents an attraction, and the number next to it is its index (starting from 0) in the candidate attractions list. The orientation of the map is: North is up, South is down, West is left, and East is right.

```
{MAP IMAGE}
```

The output format is: '`<think>...</think>`' > '`<answer>...</answer>`'

Enclose your thinking process within '`<think>`' and '`</think>`'.

Enclose your planned result within '`<answer>`' and '`</answer>`'. The format should be a list where each element is the index of an attraction. The order of attractions in the list represents the sequence of your planned route.

```
# Existing Planning Content
```

Here is the existing planning content:

```
{ROUTE PLAN}
```

```
# Scatter Plot Distribution
```

The location distribution of the attractions is shown in the figure. Each black dot represents an attraction, and the number next to it is its index (starting from 0) in the candidate attractions list. The orientation of the map is: North is up, South is down, West is left, and East is right.

```
{MAP IMAGE}
```

I.4. Prompt for answer refinement

```
# Task Description
```

You are a trip planning expert, responsible for creating itineraries for users. Based on our retrieved set of candidate attractions, select a subset that fits the user's requirements and plan a logical tour route.

I will also give you a complete plan. Please correct any inaccuracies or inefficiencies within it, particularly regarding directions and route flow, while keeping the core content intact. Present the thinking process without mentioning that it has been modified.

User's requirements:

```
{USER QUERY}
```

```
# Candidate Attractions
```

Based on the user's requirements, we have retrieved some attractions for you to choose from:

Each attraction includes three pieces of information: 'name' is the attraction's name; 'tag' contains some representative tags for the attraction; 'duration' is the recommended visiting time for the attraction (unit: hours).

```
{POI INFO}
```

```
# Output Format
```

I.5. Prompt for CPR evaluation

```
# Task Description
```

You will play the role of a user, providing us with your travel needs and a list of attractions for us to choose from and plan an itinerary that meets your requirements.

You need to comprehensively evaluate the quality of two itineraries we provide, and decide which one you would prefer.

```
# Travel Needs
```

The user's requirements are as follows:

```
{USER QUERY}
```

```
# Candidate Attractions
```

Each attraction includes three pieces of information: 'name' is the attraction's name; 'tag' contains some representative tags for the attraction; 'duration' is the recommended visiting time for the attraction (unit: hours).

The candidate attractions for planning the user's itinerary are as follows:

```
{POI INFO}
```

```
# Candidate itineraries
```

```

The itinerary format is: '<think>...</think> <answer>...</answer>'.
The planning process is enclosed within '<think>' and '</think>'.
The planned route is enclosed within '<answer>' and '</answer>'. The route is a list where each element is the index of an attraction. The order of attractions in the list represents the sequence of the planned route.

## Candidate itinerary A
{ROUTE PLAN A}

## Candidate itinerary B
{ROUTE PLAN B}

# Output Format
Your output should also be in the format of '<think>...</think> <answer>...</answer>'.
Use '<think>' and '</think>' as the start and end tags to output your thought process.
Use '<answer>' and '</answer>' as the start and end tags to output which itinerary you would prefer ('A' or 'B').

# Point Distribution
The distribution of the points is shown in the figure. Each black dot represents an attraction, and the number next to it is its index (starting from 0) in the list of candidate attractions.
{MAP IMAGE}

```

I.6. Prompt for ORS evaluation

```

# Task Description
You will play the role of a user, providing us with your travel needs and a list of attractions for us to choose from and plan an itinerary that meets your requirements.
You need to comprehensively evaluate the itinerary we provide from a user's perspective and give your rating ('A'-'E'):
'A' stands for Excellent. The itinerary's content is entirely correct, the planned route is very smooth, and you will follow the plan for your trip.
'B' stands for Good. The itinerary's content is generally correct, the planned route is mostly smooth, and

```

```

you would most likely follow the plan for your trip.
'C' stands for Average. The quality of the itinerary's content is fair, and the planned route is decent. Even if you don't follow the plan, you would still find it has some reference value.
'D' stands for Poor. The itinerary's content has several errors, and the planned route includes some detours. You would most likely not follow this plan for your trip.
'E' stands for Very Poor. The itinerary's content is full of errors, and the planned route has obvious detours. You will not follow this plan for your trip.

# Travel Needs
The user's requirements are as follows:
{USER QUERY}

# Candidate Attractions
Each attraction includes three pieces of information: 'name' is the attraction's name; 'tag' contains some representative tags for the attraction; 'duration' is the recommended visiting time for the attraction (unit: hours).
The candidate attractions for planning the user's itinerary are as follows:
{POI INFO}

# Planning Result
The itinerary format is: '<think>...</think> <answer>...</answer>'.
Our planning process is enclosed within '<think>' and '</think>'.
Our planned route is enclosed within '<answer>' and '</answer>'. The route is a list where each element is the index of an attraction. The order of attractions in the list represents the sequence of our planned route.
The itinerary we have planned for you is as follows:
{ROUTE PLAN}

# Output Format
Your output should also be in the format of '<think>...</think> <answer>...</answer>'.
Use '<think>' and '</think>' as the start and end tags to output your thought process.

```

```
Use '<answer>' and '</answer>' as the
start and end tags to directly
output your rating ('A'-'E').
```

Point Distribution

```
The distribution of the points is shown
in the figure. Each black dot
represents an attraction, and the
number next to it is its index (
starting from 0) in the list of
candidate attractions.
```

```
{MAP IMAGE}
```

I.7. Prompt for RTPR evaluation

Task Description

```
You will act as a Route Quality Reviewer
, responsible for examining the
quality of itineraries and providing
a result.
```

```
For the given user requirements and
itinerary, determine whether the
theme of the itinerary aligns with
the user's needs, and give a result
('Pass' or 'Fail').
```

```
The routes we provide may not
necessarily be the most efficient,
but as long as the selected POIs are
thematically coherent to user's
query, it is considered reasonable.
Only extremely severe errors (such
as no coffee shop in the route for
the query 'coffee walk') would be
deemed unreasonable.
```

Travel Needs

```
The user's requirements are as follows:
{USER QUERY}
```

Candidate Attractions

```
Each attraction includes three pieces of
information: 'name' is the
attraction's name; 'tag' contains
some representative tags for the
attraction; 'duration' is the
recommended visiting time for the
attraction (unit: hours).
```

```
The candidate attractions for planning
the user's itinerary are as follows:
{POI INFO}
```

Planning Result

```
The itinerary format is: '<think>...</
think> <answer>...</answer>'.
```

```
Our planning process is enclosed within
'<think>' and '</think>'.
```

```
Our planned route is enclosed within '<
answer>' and '</answer>'. The route
```

```
is a list where each element is the
index of an attraction. The order of
attractions in the list represents
the sequence of our planned route.
```

```
The itinerary we have planned for you is
as follows:
```

```
{ROUTE PLAN}
```

Output Format

```
Your output should also be in the format
of '<think>...</think> <answer
>...</answer>'.
```

```
Use '<think>' and '</think>' as the
start and end tags to output your
thought process.
```

```
Use '<answer>' and '</answer>' as the
start and end tags to directly
output your result ('Pass' or 'Fail
').
```

Point Distribution

```
The distribution of the points is shown
in the figure. Each black dot
represents an attraction, and the
number next to it is its index (
starting from 0) in the list of
candidate attractions.
```

```
{MAP IMAGE}
```

I.8. Prompt for TSPR evaluation

Task Description

```
You will act as a Route Quality Reviewer
, responsible for examining the
quality of itineraries and providing
a result.
```

```
For a given user request and planned
itinerary, determine whether the
time schedule is reasonable and give
a result ('Pass' or 'Fail').
```

```
Since the visiting time for POIs is
relatively flexible, you don't need
to strictly fix the duration for
each POI. The routes we provide may
not necessarily be the most
efficient, but as long as the time
arrangement is practical, it is
considered reasonable. Only
extremely severe timing errors (such
as planning to visit locations in
one day that would realistically
take multiple days) would be deemed
unreasonable.
```

Travel Needs

```
The user's requirements are as follows:
{USER QUERY}
```

```

# Candidate Attractions
Each attraction includes three pieces of
information: 'name' is the
attraction's name; 'tag' contains
some representative tags for the
attraction; 'duration' is the
recommended visiting time for the
attraction (unit: hours).
The candidate attractions for planning
the user's itinerary are as follows:
{POI INFO}

# Planning Result
The itinerary format is: '<think>...</
think> <answer>...</answer>'.
Our planning process is enclosed within
'<think>' and '</think>'.
Our planned route is enclosed within '<
answer>' and '</answer>'. The route
is a list where each element is the
index of an attraction. The order of
attractions in the list represents
the sequence of our planned route.
The itinerary we have planned for you is
as follows:
{ROUTE PLAN}

# Output Format
Your output should also be in the format
of '<think>...</think> <answer
>...</answer>'.
Use '<think>' and '</think>' as the
start and end tags to output your
thought process.
Use '<answer>' and '</answer>' as the
start and end tags to directly
output your result ('Pass' or 'Fail
').

# Point Distribution
The distribution of the points is shown
in the figure. Each black dot
represents an attraction, and the
number next to it is its index (
starting from 0) in the list of
candidate attractions.
{MAP IMAGE}

```

I.9. Prompt for SHR evaluation

```

# Role
You are a meticulous geographic
information verifier. Your task is
to verify the accuracy of all '
directional statements' based on a
numbered map and a travel plan text.

# Core Task

```

```

Step 1: Identify all 'directional
statements'
Read through the travel plan text and
identify all sentences or phrases
that contain 'directional statements
'. 'Directional statements' often
include direction words such as ['
East', 'South', 'West', 'North', '
Southeast', 'Northeast', 'Southwest
', 'Northwest'].
Step 2: Verify all 'directional
statements'
For each 'directional statement' you
identified in Step 1, refer to the
map to verify if the description
matches the map information and
record the result.
For clearly incorrect directional
statements, such as 'a westward
direction described as a southward
direction', judge it as 'Incorrect'.
For relatively ambiguous
directional statements, such as 'a
southwest direction described as
west or south' or 'a north direction
described as northeast or northwest
', judge it as 'Ambiguous'. For
clearly correct directional
statements, judge it as 'Correct'.

# Input Information
Map image is as follows:
{MAP IMAGE}

Travel plan text is as follows:
{ROUTE PLAN}

# Output Requirements
Your final answer must be and can only
be a single JSON object, without any
additional explanations or text.
The JSON object must strictly follow
the structure below:
```json
{
 "verifications": [
 {
 "statement": "<The directional
statement being verified, quoted
precisely from the travel plan text
>",
 "judge_result": "<The judgment
result (Correct/Ambiguous/Incorrect)
, indicating whether the statement
matches the map>",
 "correction": "<If judge_result is
not 'Correct', provide a brief
correction; if judge_result is '

```

```
Correct', omit this field>"
}
// For each directional statement,
add a corresponding verification
result object to this list
]
}
```