NAVMIND: FROM OPEN-ENDED NAVIGATION REQUESTS TO GROUNDED PERSONALIZED ROUTES

Anonymous authors
Paper under double-blind review

ABSTRACT

Conventional route planners optimize distance or time but rarely jointly consider visit needs and preferred street-level experience. Planning with large language models (LLMs) alone seems strong in open-ended requests, but may hallucinate places, misjudge street environments, or violate road-network constraints. We present NavMind, an LLM-guided personalized routing framework that grounds visit intent (e.g., a photogenic restaurant serving good steak) and preferred route experience (e.g., lively, visually engaging streets) in separate evidence sources using three modules. (1) Hybrid Point-of-Interest Retrieval-Augmented Generation (Hybrid POI-RAG) encodes POI locations and review-derived semantics to identify POIs that match the visit intent. (2) The perceptual-cost module scores street qualities with a fine-tuned vision-language model, while LLM-as-a-Jury reconciles different LLM estimates of each quality's importance to the requested experience. (3) The route-optimization module casts visit intent as waypoint constraints and preferred street-level experience as edge costs in a generalized traveling salesman problem, yielding a network-valid route. Evaluation on Beijing data spans diverse personas and requests. Qualitative cases illustrate how NavMind adapts waypoints and routes to traveler preferences, generating diverse routes across intentions. At scale, simulations indicate that this adaptation requires only modest detours, with median detour rates below 20% for every persona. In user assessment, the resulting routes score 1.78 points higher than shortest paths on a seven-point scale and are preferred nearly four times as often. Component-level results further support these system-level findings: Hybrid POI-RAG achieves over five times the Precision@5 of semantic-only or spatial-only baselines; street-view scores agree with human ratings across four qualities (Pearson's $r > 0.90$); and Jury profiles align more closely with human judgments than single LLM. These results highlight NavMind's potential for meaningful personalized navigation.

1 INTRODUCTION

Route navigation has evolved from a basic wayfinding utility into core digital infrastructure supporting everyday mobility and urban services (Zheng et al., 2014; Bast et al., 2016; Gavalas et al., 2014; Toth & Vigo, 2014). Yet, a route is not merely a geometric connection, but a continuous sequence of urban spaces that a traveler visually and physically experiences (Yang et al., 2026). Existing navigation engines typically reduce routing to a rigid optimization problem minimizing distance or travel time. Although dynamic factors like traffic and weather are now considered, the underlying objectives remain strictly efficiency-driven. Empirical studies show human mobility is rarely governed by origin-to-destination efficiency alone. Instead, travelers are driven by multifaceted, open-ended intents, often seeking to visit intermediate locations (e.g., cafes) besides final destinations. Furthermore, environmental perceptions substantially shape route choices, leading people to deviate from shortest paths in pursuit of safer, quieter, greener, or more vibrant streets (Quercia et al., 2014; Bongiorno et al., 2021). This gap raises a critical question for next-generation personalized navigation: How can we systematically bridge open-ended human travel intents and deterministic spatial optimization to jointly accommodate flexible semantic visits and experiential preferences on real-world road networks?

To quantify travelers' experiential preferences, pioneering research has leveraged street-view imagery as a proxy for human visual perception. Initiatives like Place Pulse introduced large-scale

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

pairwise judgments of urban scenes, while Streetscore and subsequent deep models mapped those perceptual qualities (e.g., *Beautiful*, *Safe*, *Lively*) across entire cities (Salesses et al., 2013; Naik et al., 2014; Dubey et al., 2016; Wang et al., 2025). While demonstrating that streetscape experiences are quantifiable, these resulting static perceptual maps fundamentally misalign with personalized routing demands. First, these methods predominantly aggregate population-level consensus into isolated spatial units, inherently failing to disentangle nuanced, persona-specific preferences. Second, they output disjointed scalar scores rather than traversable spatial structures. Identifying high-quality environments is insufficient; personalized navigation requires systematically integrating these desirable, yet isolated, perceptual signals into continuous, topologically valid trajectories on road networks.

Furthermore, a holistic perceptual routing system needs to optimize not only the continuous streetscapes between an origin and a destination, but also the alignment of intermediate waypoints to travel intents semantically. Travelers frequently incorporate flexible, context-aware stops—such as a "quiet cafe"—into their routing plans to maximize their overall experiential satisfaction. While existing POI recommendation systems successfully retrieve semantically relevant places, they typically evaluate these locations in a spatial vacuum (Li et al., 2024; Wang et al., 2023; Li & Lim, 2026). They prioritize semantic matching but ignore the spatial detour penalty and the continuity of the overall journey. Conversely, conventional routing algorithms cannot process such open-ended requests; they act as rigid calculators that require exact geographic coordinates and predefined metrics before computing a path (Held & Karp, 1962; Tang et al., 2024; de la Rosa et al., 2024). Consequently, neither static perceptual maps nor spatially blind POI recommendations are sufficient.

The core bottleneck in realizing such a coordinated system lies in computationally translating unstructured human desires into the deterministic parameters demanded by spatial algorithms. Large Language Models (LLMs) overcome this translation barrier by parsing natural language to directly extract flexible POI needs and street-level human perceptions, eliminating the need for numerical weights. However, despite their promise in urban reasoning (Feng et al., 2024; Zhou et al., 2025; Xie et al., 2024; Tang et al., 2024; Chen et al., 2024; de la Rosa et al., 2024; Shao et al., 2025; Yuan et al., 2025), a fundamental challenge remains: describing a plausible journey does not equate to computing a traversable one. Unconstrained text generation frequently hallucinates geographic entities (Yu et al., 2026; Schneider et al., 2025) and lacks topological connectivity, while general vision models may misjudge authentic human perceptions (Moreno-Vera & Poco, 2025; He et al., 2026). Therefore, a robust paradigm must restrict LLMs to structuring open-ended intents, delegating factual POI grounding, perceptual estimation, and exact route construction to deterministic spatial tools.

To overcome these limitations, we propose **NavMind**, a tool-augmented planning framework translating natural-language requests and persona profiles into traversable, perception-aware routes. NavMind separates what a traveler wants to visit (*local visit intent*, including flexible waypoints) from how they want the journey to feel (*global experiential preference*). To fulfil these intents, a hybrid POI-RAG module grounds flexible visits using route-compatible POIs and reviews (Lewis et al., 2020; Yu et al., 2026; Schneider et al., 2025). Concurrently, a perceptual-cost module combines street-view assessments with an LLM jury to convert experiential preferences into persona-conditioned edge costs in the network (Verga et al., 2024; Qian et al., 2026). Finally, an exact grouped-route optimizer determines the visit sequence and reconstructs the trajectory on the road network. This decoupling enables the LLM to flexibly structure requests while ensuring the final route is grounded in urban evidence and physically executable.

Our contributions are:

- We formulate grounded personalized perceptual routing as a language-to-route compilation problem that jointly considers local visit needs, global experiential preferences, and road-network constraints.
- We propose NavMind, a grounded language-to-route framework that integrates hybrid POI-RAG, human-aligned street-view perception, and exact route optimization through explicit intermediate states.
- We evaluate our framework on multi-source Beijing data through component validation, 36,000 simulated routes, and blinded human assessments, demonstrating its capability to translate subjective intents into executable urban trajectories.



Figure 1: NavMind grounds visit needs in POI reviews and route style in street-view qualities, yielding route-compatible stops and a network-valid path (right); distance-based navigation selects nearby but preference-mismatched stops and streets inconsistent with the request (left).

2 Related Work

2.1 LLM-Enabled Travel and Route Planning

LLM-based travel planning has progressed from language-conditioned itinerary generation toward grounded and solver-assisted planning. TravelPlanner, TravelAgent, TP-RAG, and TripCraft interpret open-ended constraints and assemble multi-step itineraries (Xie et al., 2024; Chen et al., 2024; Ni et al., 2025; Chaudhuri et al., 2025), while ITINERA, TRIP-PAL, Personal Travel Solver, and LLMAP couple LLMs with planners or optimization modules to improve feasibility (Tang et al., 2024; de la Rosa et al., 2024; Shao et al., 2025; Yuan et al., 2025). Related studies use LLMs for heuristic search or route generation (Meng et al., 2024; Deng et al., 2024), and map-aware retrieval grounds plans and POI recommendations in geographic evidence (Manvi et al., 2024; Balsebre et al., 2024; Li et al., 2024; Yu et al., 2026; Schneider et al., 2025). These efforts move travel planning from free-form text toward executable decisions, but route quality is still commonly represented through place utility or predefined constraints. The street environment experienced between recommended places therefore remains largely outside the planning objective.

2.2 Street-view Assessment of Urban Perception

Urban theory has long emphasized that people experience cities through sequences of streets and places, from Lynch's *imageability* and Cullen's *serial vision* to Jacobs's *street life* (Lynch, 1960; Cullen, 1961; Jacobs, 1961); transport studies and pedestrian trajectories likewise show that built environments shape travel behavior beyond distance alone (Ewing & Cervero, 2010; Bongiorno et al., 2021). Street-view imagery makes this experience measurable at scale. Place Pulse, Streetscore, and subsequent deep models infer perceptions such as *Beautiful*, *Safe*, *Lively*, and *Wealthy* from urban scenes (Salesses et al., 2013; Naik et al., 2014; Dubey et al., 2016), with later work relating these measures to urban functions and multimodal street understanding (Wang et al., 2025; Zhang et al., 2025). However, general pretrained models may diverge from human ratings (Moreno-Vera & Poco, 2025; He et al., 2026), and most perception studies characterize individual images or spatial distributions rather than how user preferences should accumulate along a continuous route.

2.3 Perception-Aware Navigation on Road Networks

Perception-aware navigation uses street-level qualities to influence route choice. Work on *happy routes* shows that people may prefer beautiful, quiet, or pleasant alternatives even when they are not the shortest (Quercia et al., 2014), while later methods use visual features or walking preferences to recommend pleasant and personalized pedestrian routes (Wakamiya et al., 2019; Shah & Wang, 2023). These studies establish environmental quality as a meaningful routing signal, but typically rely on predefined preference categories, fixed perceptual weights, or manually designed objectives. They rarely infer route priorities from open-ended intent, coordinate them with flexible visit re-

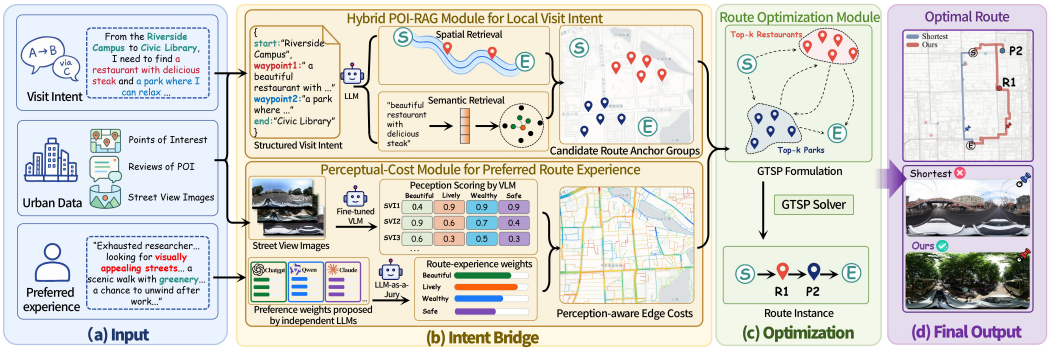

Figure 2: NavMind separates local visit intent and global route style, grounds them in POI reviews and street views through Hybrid POI-RAG and the perceptual-cost module, then jointly selects anchors and reconstructs a network-valid route.

quests, or jointly select suitable POIs under road-network constraints. NavMind addresses this gap by connecting language-derived local visit intent and global route style with grounded POI retrieval, human-aligned street perception, and deterministic grouped-route optimization.

3 Methodology

3.1 Problem Definition

Given a visit intent x^{visit} and a preferred route experience x^{exp}, we optimize a valid route $\pi = (v_0, \ldots, v_T)$, where $(v_{t-1}, v_t) \in E$, and optional intermediate POIs $\mathbf{z}$. The road network is represented by a graph $G = (V, E)$, POI and review evidence $\mathcal{F}$, and street-view evidence $\mathcal{V}$. The visit intent x^{visit} and POI evidence $\mathcal{F}$ define candidate anchor groups $\mathcal{C} = \{C_1, \ldots, C_m\}$. The feasible set $\Omega(\mathcal{C}; G)$ contains all valid route–POI pairs that select one anchor from each group and connect the selected anchors on the road graph. The preferred experience and $\mathcal{V}$ define the personalized routing cost $\mathcal{J}_{x^{\text{exp}}}(\pi; G, \mathcal{V})$, where lower values indicate better alignment. The task is formulated as

$$(\pi^\star, \mathbf{z}^\star) = \arg \min_{(\pi, \mathbf{z}) \in \Omega(\mathcal{C}; G)} \mathcal{J}_{x^{\text{exp}}}(\pi; G, \mathcal{V}). \tag{1}$$

3.2 Framework Overview

To solve Eq. 1 while avoiding LLM spatial hallucinations, NavMind coordinates language reasoning and spatial computation through a comprehensive framework. As illustrated in Figure 2, our NavMind processes three inputs: open-ended visit intents, multi-modal urban data (e.g. POIs, street-view images), and subjective experiential preferences. These inputs enter the Intent Bridge, where the Hybrid POI-RAG Module structures requests via spatial-semantic retrieval and generates Candidate Route Anchor Groups to construct the feasible set Ω. Concurrently, the Perceptual-Cost Module leverages a fine-tuned VLM and an LLM jury to convert text preferences into Perception-aware Edge Costs, formulating the objective $\mathcal{J}_{x^{\text{exp}}}$. Finally, utilizing these anchors (Ω) and edge costs $c_{x^{\text{exp}}}(e)$, the Route Optimization Module formulates a Generalized Traveling Salesman Problem (GTSP). A deterministic solver computes the optimal sequence, reconstructing a perception-aligned, traversable route.

3.3 Hybrid POI-RAG for Local Visit Intent

This module transforms the parsed visit intent x^{visit} into candidates of route-anchor groups $\mathcal{C} = \{C_1, \ldots, C_m\}$ (Figure 2). Requests containing place names are parsed into deterministic waypoints via geocoding. For flexible requests (e.g., 'a photogenic restaurant serving good steak'), semantic retrieval alone often yields geographically unrelated locations. NavMind mitigates this issue by pipelining spatial filtering, semantic retrieval, and listwise reranking detailed below.

Spatial Filtering and Semantic Retrieval. First, we create a buffer region (a corridor) around a reference route connecting the origin, destination, and deterministic waypoints. POIs outside this corridor are left out to prevent impractical detours. Within this filtered subset of POIs, we encode the flexible request into a semantic query vector $\mathbf{q}$, which is then used to fetch the most relevant POI (embeddings of its metadata and reviews) by cosine similarity. From this we obtain top-M candidates that satisfy both spatial and semantic requirements. The mathematical definitions of the corridor and semantic similarity functions are detailed in Appendix A.1.

Evidence-Aware Reranking. Preliminary experiments indicated that spatial-semantic retrieval may overlook review polarity, such as matching "not quiet at all" to "a quiet place," thereby compromising downstream route planning. We thus use an LLM to evaluate and rerank the top-M candidates. The LLM jointly evaluates the raw review text, the similarity score, and the distance between a candidate and the reference route, from which we select the best $K(K \leq M)$ candidates. We retain the top-K candidates as equally viable alternatives for a flexible visit intent, increasing route diversity and reducing the risk of excessive detours from committing to a single POI. The annotation protocol is detailed in Appendix A.1.

3.4 Persona-Conditioned Perceptual Cost Modeling

To realize the routing objective $\mathcal{J}_{x^{\text{exp}}}$, this module turns street-level visual evidence and user preferences into customized edge costs $c_{x^{\text{exp}}}$ on G. It maps street-level visual perception into a weighted graph $G_{x^{\text{exp}}} = (V, E, c_{x^{\text{exp}}})$.

Visual Scoring and Jury Personalization. Urban route experience is inherently subjective, yet personalized routing requires a consistent representation of such preferences. We therefore ground street-level perception in human judgments and use perceptual attributes as an intermediate representation between street-view imagery and routing costs. From Place Pulse 2.0, we adopt four positively oriented dimensions, $\mathcal{D}$={B,S,L,W} (Beautiful, Safe, Lively, and Wealthy), grounded in large-scale pairwise human judgments for reliable street-level perception modeling. Using a LoRA-fine-tuned Vision-Language Model (Qwen2.5-VL-7B) trained on normalized human judgments, we assign an absolute perception score $\hat{y}_d(e) \in [0, 1], d \in \mathcal{D}$ to every road edge e. To determine how much each dimension matters to a specific user, we employ an LLM-as-a-Jury mechanism. Multiple LLM proposers independently suggest dimension weights according to the preferred route experience x^{exp} in the user request, and an external judge LLM selects the most coherent normalized weight vector $\mathbf{w}_{x^{\text{exp}}} \in [0, 1]^4$.The Jury achieves lower overall error against human reference ratings than individual proposers on the evaluated personas (Appendix B.2, Table 6).

Perceptual Cost Function. We convert the positive visual perception into a penalty $\hat{s}_d(e) = 1 - \hat{y}_d(e)$. Following the power-law nature of human sensory perception (Stevens, 1957; Quercia et al., 2014), we apply a cubic transformation $\phi(r) = r^3$ to impose a stronger penalty on route segments that substantially deviate from the user's stylistic preferences. The personalized cost for edge e integrates both physical distance and weighted perceptual penalties:

$$c_{x^{\text{exp}}}(e) = \ell(e) \left[1 + \sum_{d \in \mathcal{D}} w_{x^{\text{exp}},d} \cdot \phi(\hat{s}_d(e)) \right]. \tag{2}$$

where $\ell(e)$ is the physical edge length. This ensures the routing algorithm defaults to the geometric shortest path when perception weights are zero, but smoothly trades off distance for preferred street atmospheres when directed by the persona. The mathematical definitions of perceptual cost modeling are detailed in Appendix A.2.

3.5 Exact Route Optimization via GTSP

Given the perceptual-cost graph $G_{x^{\text{exp}}}$ and the candidate anchor groups $\mathcal{C}$, the final task is to select one representative anchor $z_j \in C_j$ for each flexible request and determine their optimal visiting order. NavMind formulates this process as a Generalized Traveling Salesman Problem (GTSP), a deterministic solver that guarantees global optimality and road network constraints that a LLM-based planner does not always follow.

Metric Closure and Subset Dynamic Programming. To make joint anchor selection and ordering tractable on large urban networks, our method decouples network-level path computation from

combinatorial optimization using a metric closure and subset dynamic programming. Specifically, we construct the closure over the origin, destination, and candidate anchors using pairwise shortest perceptual paths in $G_{x^{\text{exp}}}$. A Held-Karp subset dynamic programming (Held & Karp, 1962) then jointly optimizes anchor selection and ordering. By backtracking, we recover the optimal order of anchors $\mathbf{z}^\star$, and then insert the corresponding precomputed path between them, yielding the complete continuous route $\pi^\star$. Full details and pseudo-code are provided in Appendix A.3.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Study Area and Datasets. Here, we evaluate NavMind by routing experiments in Beijing. The routing graph G is constructed using OpenStreetMap (OSM) data, which is geometrically simplified, topologically repaired, and projected into a metric coordinate system. To instantiate urban functional information $\mathcal{F}$, we process 5,829 Amap POIs[1] alongside their associated user reviews. For urban visual-perceptual information $\mathcal{V}$, we utilize 7,083 geo-referenced Baidu Street View images[2], matching them to corresponding road segments to build the perceptual-cost graph. To ensure authentic evaluation, we curate 12 distinct personas distilled from social media travel guides. Engineered to minimize attribute overlap, the personas are designed to be representative of the local society, spanning diverse real-world needs (e.g., a tourist seeking lively streets versus an efficiency-driven programmer).

Implementation Details. GPT-5 serves as the primary engine for both the intent parser and the POI reranker. To ensure deterministic and reproducible reasoning, the generation temperature across all large language models is strictly set to 0.0. Persona weights are proposed independently by Qwen3-Max, Claude Sonnet 4.5, GPT-5, and GLM-4.6, with Gemini-2.5 Pro acting as the external judge. For street-view perception, we fine-tune the Qwen2.5-VL-7B model using Low-Rank Adaptation (LoRA) with a rank of 16. This process utilizes 20,000 images from the Place Pulse 2.0 dataset, which are split into training and validation sets at a 4:1 ratio to ensure robust calibration across the four perceptual dimensions. The Hybrid POI-RAG module initially retains $M = 10$ spatial-semantic candidates before reranking, ultimately yielding a group of $K = 5$ candidates for each flexible visit. Finally, the GTSP route solver efficiently computes the trajectory by combining Dijkstra's metric closure with an exact subset dynamic programming algorithm. Exhaustive details regarding data processing, model roles, and solver configurations are provided in Appendix B.1.

4.2 GROUNDED ROUTE RESULTS

Perceptual Dimension Adjustments. As shown in Figure 3, isolating a single perceptual dimension dynamically shifts both the trajectory and its functional-zone exposure. Optimizing for *Beautiful* steers routes toward greener corridors, quantitatively supported by massive spikes in natural/tourism exposure (up to +58.8 percentage points). Conversely, *Lively* pulls routes into active storefront environments, surging commercial exposure (+15.9 to +20.6 pp). The *Safe* objective avoids cluttered alleys in favor of wide, legible roads, while *Wealthy* bypasses older and messy streets for modern architecture and polished facades, aligning perfectly with urban affluence.

Persona-Conditioned Generation. Human perceptual needs often involve complex multi-dimensional interplays. Figure 4 demonstrates NavMind's ability to satisfy these intents by balancing competing perceptual weights (radar plots). Given identical constraints, the optimizer tailors paths strictly to the persona. In Case 1, a driver seeking "safe and open" roads traverses broad avenues, while a dancer desiring "alive" streets navigates vibrant street blocks. Case 2 highlights preference-efficiency tradeoffs: the researcher (Lin Jing) accepts detours for tree-canopied streets, whereas the programmer (Zhao Lei) with low perceptual weights follows the shortest path, confirming that NavMind defaults to spatial efficiency unless weights dictate otherwise. Case 3 shows that even with shared safety priorities, NavMind captures subtle perceptual differences, routing a caregiver to active streets and an older resident to shaded community roads. Case 4 illustrates persona-dependent routing, directing a vlogger to photogenic avenues and a resident to neighbor-

[1] https://lbs.amap.com/

[2] https://lbsyun.baidu.com/

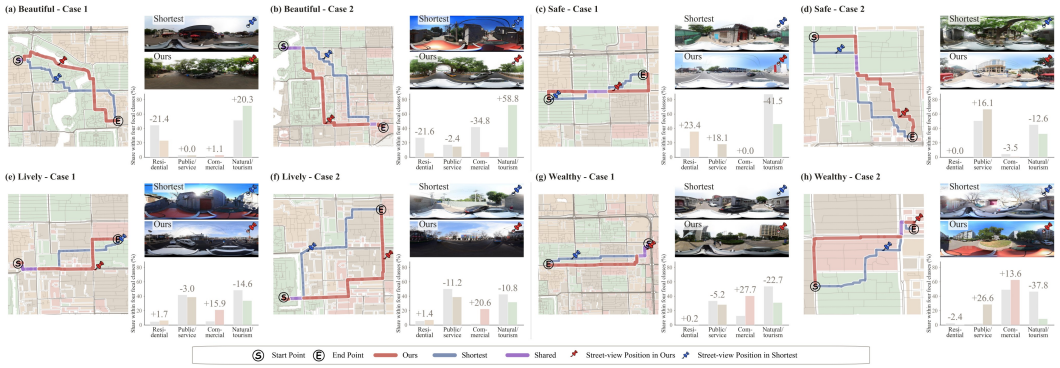

Figure 3: Single-dimension routing for Beautiful, Safe, Lively, and Wealthy: shortest versus perception-optimal paths, street views, and changes in functional-zone exposure (percentage points).

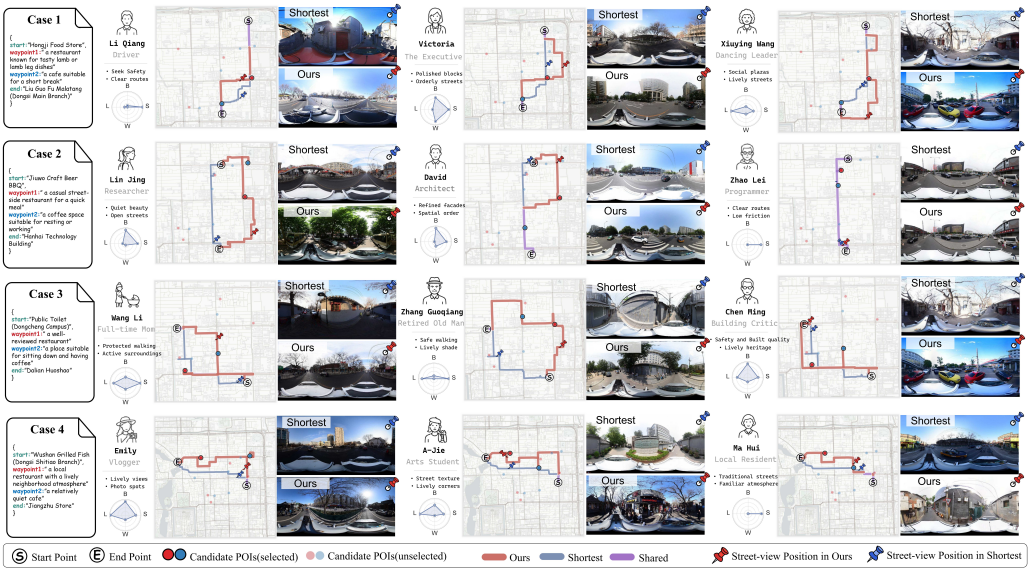

Figure 4: Persona-conditioned routes. Each row compares three personas on the same task; perceptual weights (radar plots) yield distinct executable routes, POIs, and street views.

hood backstreets. Ultimately, these results confirm that NavMind actively constructs personalized, perception-optimal routes based on travel intents and emotional preferences, harmonizing trip personalization with network accessibility.

End-to-end comparison. We compared NavMind with LLMAP (Yuan et al., 2025) and ITINERA (Tang et al., 2024) on 108 route-generation tasks (Table 1) using Intent Completion (IC), Extra Distance (ED), and Win Rate of Preference (WRP) (definitions provided in Appendix B.3). NavMind was preferred over ITINERA and LLMAP in 66.67% and 70.37% of the style comparisons, respectively, consistent with its explicit optimization of perceptual style. NavMind and ITINERA achieved higher IC, as both support fine-grained POI preferences. NavMind also achieved the lowest mean ED, potentially benefiting from spatial information to be incorporated into the route optimization process earlier.

4.3 Large-Scale Routing Simulation

Large-scale route simulations are further conducted to examine whether NavMind produces systematically biased behavior. A total of 4,000 origin–destination (OD) pairs evenly distributed across four shortest-path distance bins were tested; optimizing each pair under four single-perception

Table 1: End-to-end comparison. Win Rate of Preference indicates how often the row method is preferred; 50% indicates equal preference.

Method	IC (%)↑	ED (m)↓	WRP (%)	
			vs. ITINERA	vs. LLMAP
NavMind	**66.20**	**486.4**	66.67	70.37
ITINERA	58.33	1703.8	–	67.86
LLMAP	33.33	1195.8	–	–

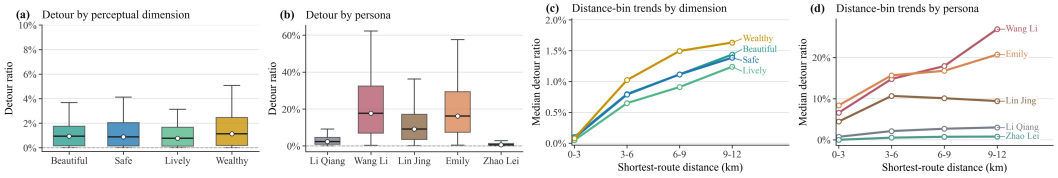

Figure 5: Detour patterns across simulated OD pairs. (a–b) Route-level detour distributions under four single-perception objectives and five persona profiles. (c–d) Median detour ratios across four shortest-path distance bins.

objectives and five persona profiles yields 36,000 simulated routes. The results show that NavMind delivers preference-aligned experiential gains while keeping geographic detour costs within a reasonable range.

Preference-dependent Routing Behavior. Figure 5a and 5b illustrates the navigation behaviors induced by different optimization objectives. Single-perception routes maintain a median detour of approximately 1%, whereas persona-conditioned routes exhibit more pronounced preference-efficiency tradeoffs. When a user's preference closely aligns with the primary functional road network, as in Zhao Lei's profile, NavMind incurs almost no additional detour ($< 1\%$). Only when a user seeks highly specific environmental characteristics, as in Wang Li's profile, does the system expand its route exploration with increased detours (median ratio of 17.67% with fairly large quantiles).

Controllable Detour Costs. While the growth of detour ratio slows down as trip distance increases (Figure 5c and 5d), persona-conditioned routes exhibit non-uniform growth patterns. Zhao Lei and Li Qiang remain relatively stable, whereas Wang Li and Emily, who place greater emphasis on street quality, reached the highest detours in the largest distance bin, at 26.93% and 20.73%, respectively. Longer base paths contain more candidate segments, offering these environment-oriented users more opportunities for route optimization. Across all objectives and personas, the overall median detours remain below 20% (Figure 5a and 5b), while commonly used Google Maps navigation supports detours of up to approximately 50% (Sevtsuk et al., 2021).

4.4 Human Preference Assessment

We further conduct a user study against conventional shortest-path navigation to examine whether personalized perceptual routing can serve as a viable alternative when users value the experience of the route itself. We recruited 36 student participants with urban-related academic backgrounds, spanning undergraduate, master's, and doctoral levels. Each participant was assigned to simulate specific persona profiles (e.g., Student A-Jie, Tourist Emily, Programmer Zhao Lei). Prior to the formal blind test, they underwent a structured training session to thoroughly internalize the stylistic preferences and needs of their assigned roles. They reviewed the task, POIs, and street views for NavMind-generated versus the shortest routes, rated both routes and explained their choices (see Appendix B.4 for details), yielding $N = 324$ valid comparisons.

As shown in Table 2, NavMind received higher ratings than the shortest path (5.26 vs. 3.48) and was selected nearly four times as often (208 vs. 56). Free-text responses further revealed: among comparisons favoring NavMind, 63.5% cited route environments that better matched persona-specific travel goals, while 36.5% cited more appropriate waypoint POIs, with 52 explanations mention-

Table 2: Human assessment (1–7 in Likert Scale). Δ: NavMind minus shortest path; N/T/S: counts preferring NavMind/tie/shortest path.

Group	Shortest	NavMind	Δ	N/T/S
Student (A-Jie)	3.63	**5.26**	+1.63	80/8/20
Tourist (Emily)	3.85	**4.81**	+0.96	56/32/20
Programmer (Zhao Lei)	2.96	**5.70**	+2.74	72/20/16
Overall	3.48	**5.26**	+1.78	208/60/56

Table 3: POI grounding on 127 route-based recommendation tasks (higher is better). Best and second-best results are bolded and underlined.

Method	Precision@5	Hit@5	MRR
Textual RAG	0.003	0.016	0.012
Spatial-only	0.069	0.236	0.132
Semantic-only	0.036	0.181	0.077
Ours (w/o Reranking)	<u>0.280</u>	<u>0.717</u>	<u>0.489</u>
Ours (Full)	**0.359**	**0.795**	**0.583**

ing both. Conversely, although some participants choosing the shortest path still acknowledged NavMind's experiential advantages, 64.3% cited detour, route directness, or fine-grained walking conditions as decisive factors. NavMind delivers personalized travel value that users can clearly perceive. Detailed analysis is given in Appendix B.5.

4.5 Component-Level Effectiveness Evaluation

Effectiveness of Hybrid POI-RAG. We evaluated 127 route-based recommendation queries (e.g., "Where can I get crispy-skinned Peking duck on the way from Peking University to the Palace Museum?"). Annotators ranked the top five POIs by route proximity and Amap reviews; Precision@5 measures the proportion of relevant, route-feasible results among them. Table 3 shows that Textual RAG hallucinates geographic proximity, Spatial-only retrieval ensures feasibility but misses fine-grained intent, while Semantic-only retrieval selects geographically inaccessible POIs.

Our framework addresses this issue by cascading the constraints. In the ablation variant without LLM reranking ("Ours w/o Reranking"), the system already achieves a Precision@5 of 0.280, proving that spatial filtering combined with dense retrieval successfully isolates a feasible candidate pool. However, our method ("Ours Full") further boosts this metric to 0.359. This significant gain confirms that the final LLM listwise reranking is critical for capturing nuanced subjective cues (e.g., distinguishing a "quiet" cafe from a "popular but noisy" one based on review sentiment), which are often overlooked by semantic vector retrieval alone ("Ours w/o Reranking").

Effectiveness of VLM Perceptual Alignment. The perceptual-cost graph requires continuous, absolute scoring, which poses a significant challenge. Existing studies demonstrate that zero-shot Vision-Large Language Models (Zero-shot VLLMs) usually fail at this task, as they either output discrete categorical labels or collapse scores into a narrow, isolated distributions, which is not effective in differentiating street scenes.

We address this issue by employing Qwen2.5-VL-7B fine-tuned via LoRA, which maps raw panoramas to a calibrated $[0, 1]$ continuous scale. Evaluated against the Place Pulse 2.0 dataset, the fine-tuned model successfully mirrors the variance of human perceptual judgments, achieving high Pearson correlations ($r > 0.90$) across all four dimensions, peaking at 0.945 for *Beautiful*. This confirms that parameter-efficient fine-tuning (PEFT) is not merely an incremental optimization, but a necessary transformation to turn a general VLLM into a reliable, high-resolution urban perception evaluator. Additional dimensional metrics and comprehensive visualization results of the distribution comparisons are provided in Appendix B.2.

5 Conclusion

We presented NavMind, a personalized route-navigation framework that translates open-ended visit intents and preferred experiences into grounded routes by combining real-world urban evidence (POIs, reviews, and street views). Across qualitative cases, 36,000 personalized route simulations, a blinded user assessment and component-level evaluations, NavMind produced preference-aligned journeys with controllable detour costs. By establishing a practical interface between human experience and constrained urban optimization, this work can support human-centered navigation and provide experiential evidence for urban planning. Richer preference representations, dynamic city signals, and broader evaluation remain important next steps.

ETHICS STATEMENT

NavMind uses subjective street-level perceptual attributes, including Beautiful, Safe, Lively, and Wealthy, which may reflect cultural or socioeconomic variation in the underlying imagery and human judgments. These attributes are intended as perceptual signals rather than objective characterizations of urban environments or communities. The human evaluation uses role-based judgments under a consistent protocol, and the resulting assessments should not be interpreted as universal travel preferences. Further discussion is provided in Appendix C.

REPRODUCIBILITY STATEMENT

The NavMind framework and route objective are described in Section 3. Implementation details, data preparation, model configurations, and evaluation procedures are provided in Appendix B.1 and Appendix A. The human-assessment protocol and detailed results are reported in Appendices B.4 and B.5. We will release the code, prompts, evaluation scripts, and processed data required to reproduce the experiments.

USE OF LARGE LANGUAGE MODELS

LLMs and VLMs are used as components of NavMind for request interpretation, POI reranking, street-view perception, and preference modeling. During manuscript preparation, LLMs were also used for language polishing and manuscript organization. The authors reviewed and verified the final text, references, and reported results.

REFERENCES

Pasquale Balsebre, Weiming Huang, and Gao Cong. LAMP: A language model on the map. *arXiv preprint arXiv:2403.09059*, 2024. URL https://arxiv.org/abs/2403.09059.

Hannah Bast, Daniel Delling, Andrew Goldberg, Matthias Müller-Hannemann, Thomas Pajor, Peter Sanders, Dorothea Wagner, and Renato F. Werneck. Route planning in transportation networks. In *Algorithm Engineering: Selected Results and Surveys*, volume 9220 of *Lecture Notes in Computer Science*, pp. 19–80. Springer, 2016. doi: 10.1007/978-3-319-49487-6_2.

Geoff Boeing. OSMnx: A python package to work with graph-theoretic OpenStreetMap street networks. *Journal of Open Source Software*, 2(12):215, 2017. doi: 10.21105/joss.00215.

Christian Bongiorno, Yihong Zhou, Ivan Kryven, David Theurel, Alberto Rizzo, Paolo Santi, Joshua Tenenbaum, and Carlo Ratti. Vector-based pedestrian navigation in cities. *Nature Computational Science*, 1(10):678–685, 2021. doi: 10.1038/s43588-021-00130-y.

Soumyabrata Chaudhuri, Pranav Purkar, Ritwik Raghav, Shubhojit Mallick, Manish Gupta, Abhik Jana, and Shreya Ghosh. TripCraft: A benchmark for spatio-temporally fine grained travel planning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 17035–17064, Vienna, Austria, July 2025. Association for Computational Linguistics. doi: 10.18653/v1/2025.acl-long.834. URL https://aclanthology.org/2025.acl-long.834/.

Aili Chen, Xuyang Ge, Ziquan Fu, Yanghua Xiao, and Jiangjie Chen. TravelAgent: An AI assistant for personalized travel planning. *arXiv preprint arXiv:2409.08069*, 2024. URL https://arxiv.org/abs/2409.08069.

Gordon Cullen. *Townscape*. Architectural Press, London, 1961. Reissued as *The Concise Townscape*, Architectural Press / Routledge, 1971, ISBN 9780750620185.

Tomas de la Rosa, Sriram Gopalakrishnan, Alberto Pozanco, Zhen Zeng, and Daniel Borrajo. TRIPPAL: Travel planning with guarantees by combining large language models and automated planners. *arXiv preprint arXiv:2406.10196*, 2024. URL https://arxiv.org/abs/2406.10196.

Hourui Deng, Hongjie Zhang, Jie Ou, Chaosheng Feng, et al. Can LLM be a good path planner based on prompt engineering? mitigating the hallucination for path planning. *arXiv preprint arXiv:2408.13184*, 2024. URL https://arxiv.org/abs/2408.13184.

Amit Dubey, Nikhil Naik, Devi Parikh, Ramesh Raskar, and Cesar Hidalgo. Deep learning the city: Quantifying urban perception at a global scale. In *European Conference on Computer Vision (ECCV)*, pp. 196–212, 2016. doi: 10.1007/978-3-319-46448-0_12.

Reid Ewing and Robert Cervero. Travel and the built environment: A meta-analysis. *Journal of the American Planning Association*, 76(3):265–294, 2010. doi: 10.1080/01944361003766766.

Jie Feng, Jun Zhang, Tianhui Liu, et al. CityBench: Evaluating the capabilities of large language models for urban tasks. *arXiv preprint arXiv:2406.13945*, 2024. URL https://arxiv.org/abs/2406.13945.

Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, and Grammati E. Pantziou. A survey on algorithmic approaches for solving tourist trip design problems. *Journal of Heuristics*, 20(3):291–328, 2014. doi: 10.1007/s10732-014-9242-5.

Jun He, Yi Lin, Zilong Huang, Jiacong Yin, Junyan Ye, Yuchuan Zhou, Weijia Li, and Xiang Zhang. UrbanFeel: A comprehensive benchmark for temporal and perceptual understanding of city scenes through human perspective. In *The Fourteenth International Conference on Learning Representations*, 2026.

Michael Held and Richard M. Karp. A dynamic programming approach to sequencing problems. *Journal of the Society for Industrial and Applied Mathematics*, 10(1):196–210, 1962. doi: 10.1137/0110015.

Jane Jacobs. *The Death and Life of Great American Cities*. Random House, New York, 1961. ISBN 9780679600474.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 9459–9474, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

Kunrong Li and Kwan Hui Lim. RALLM-POI: Retrieval-augmented LLM for zero-shot next POI recommendation with geographical reranking. In Yi Mei, Chao Qian, Qing Bai, Bing Xue, and Sanjeev Khanna (eds.), *PRICAI 2025: Trends in Artificial Intelligence*, volume 16452 of *Lecture Notes in Computer Science*, pp. 210–218, Singapore, 2026. Springer Nature Singapore. ISBN 978-981-95-7072-0. doi: 10.1007/978-981-95-7072-0_14.

Peibo Li, Maarten de Rijke, Hao Xue, Shuang Ao, et al. Large language models for next point-of-interest recommendation. *arXiv preprint arXiv:2404.17591*, 2024. URL https://arxiv.org/abs/2404.17591.

Kevin Lynch. *The Image of the City*. MIT Press, Cambridge, MA, 1960. ISBN 9780262120043.

Rohin Manvi, Samar Khanna, Gengchen Mai, Marshall Burke, David Lobell, and Stefano Ermon. GeoLLM: Extracting geospatial knowledge from large language models. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=TqL2xBwXP3.

Silin Meng, Yiwei Wang, Cheng-Fu Yang, Nanyun Peng, and Kai-Wei Chang. LLM-a*: Large language model enhanced incremental heuristic search on path planning. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 1087–1102, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-emnlp.60. URL https://aclanthology.org/2024.findings-emnlp.60/.

Felipe Moreno-Vera and Jorge Poco. Assessing urban environments with vision-language models: A comparative analysis of ai-generated ratings and human volunteer evaluations. In *International Joint Conference on Neural Networks*, 2025. URL https://files.visualdslab.com/files/papers/2025-UrbanVLM-IJCNN.pdf.

Nikhil Naik, Jade Philipoom, Ramesh Raskar, and Cesar A. Hidalgo. Streetscore—predicting the perceived safety of one million streetscapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 779–785, 2014. URL https://openaccess.thecvf.com/content_cvpr_workshops_2014/W20/html/Naik_Streetscore_-_Predicting_2014_CVPR_paper.html.

Hang Ni, Fan Liu, Xinyu Ma, Lixin Su, Shuaiqiang Wang, Dawei Yin, Hui Xiong, and Hao Liu. TP-RAG: Benchmarking retrieval-augmented large language model agents for spatiotemporal-aware travel planning. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 12392–12418, Suzhou, China, November 2025. Association for Computational Linguistics. doi: 10.18653/v1/2025.emnlp-main.626. URL https://aclanthology.org/2025.emnlp-main.626/.

OpenStreetMap contributors. OpenStreetMap. https://www.openstreetmap.org, 2026. Map data licensed under the Open Database License.

Mengjie Qian, Guangzhi Sun, Mark J. F. Gales, and Kate M. Knill. Who can we trust? LLM-as-a-jury for comparative assessment. *arXiv preprint arXiv:2602.16610*, 2026. URL https://arxiv.org/abs/2602.16610.

Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. The shortest path to happiness: recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, pp. 116–125, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450329545. doi: 10.1145/2631775.2631799. URL https://doi.org/10.1145/2631775.2631799.

Philip Salesses, Katja Schechtner, and Cesar A. Hidalgo. The collaborative image of the city: Mapping the inequality of urban perception. *PLOS ONE*, 8(7):e68400, 2013. doi: 10.1371/journal.pone.0068400.

Nicole R. Schneider, Nandini Ramachandran, Kent O'Sullivan, and Hanan Samet. DistRAG: Towards distance-based spatial reasoning in LLMs. *arXiv preprint arXiv:2506.03424*, 2025. URL https://arxiv.org/abs/2506.03424.

Andres Sevtsuk, Rounaq Basu, Xiao-Jiang Li, and Raul Kalvo. A big data approach to understanding pedestrian route choice preferences: Evidence from San Francisco. *Travel Behaviour and Society*, 25:41–51, oct 2021. doi: 10.1016/j.tbs.2021.05.010. URL https://doi.org/10.1016/j.tbs.2021.05.010.

Urmi Shah and Jia Wang. A personalised pedestrian navigation system (short paper). In *12th International Conference on Geographic Information Science (GIScience 2023)*, 2023. doi: 10.4230/LIPIcs.GIScience.2023.67. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.GIScience.2023.67.

Zijian Shao, Jiancan Wu, Weijian Chen, and Xiang Wang. Personal travel solver: A preference-driven LLM-solver system for travel planning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 27622–27642, 2025. URL https://aclanthology.org/2025.acl-long.1339/.

S. S. Stevens. On the psychophysical law. *Psychological Review*, 64(3):153–181, 1957. doi: 10.1037/h0046162.

Yihong Tang, Zhaokai Wang, Ao Qu, Yihao Yan, Zhaofeng Wu, Dingyi Zhuang, Jushi Kai, Kebing Hou, Xiaotong Guo, Han Zheng, Tiange Luo, Jinhua Zhao, Zhan Zhao, and Wei Ma. ITINERA: Integrating spatial optimization with large language models for open-domain urban itinerary planning. *arXiv preprint arXiv:2402.07204*, 2024. URL https://arxiv.org/abs/2402.07204.

Paolo Toth and Daniele Vigo (eds.). *Vehicle Routing: Problems, Methods, and Applications*. Society for Industrial and Applied Mathematics, 2 edition, 2014. doi: 10.1137/1.9781611973594.

Pat Verga, Sebastian Hofstätter, Sophia Althammer, Yixuan Su, Aleksandra Piktus, Arkady Arkhangorodsky, Minjie Xu, Naomi White, and Patrick Lewis. Replacing judges with juries: Evaluating LLM generations with a panel of diverse models. *arXiv preprint arXiv:2404.18796*, 2024. URL https://arxiv.org/abs/2404.18796.

Shoko Wakamiya, Panote Siriaraya, Yihong Zhang, Yukiko Kawai, Eiji Aramaki, and Adam Jatowt. Pleasant route suggestion based on color and object rates. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pp. 786–789, 2019. doi: 10.1145/3289600.3290611. URL https://dblp.org/rec/conf/wsdm/WakamiyaS0KAJ19.

Xinglei Wang, Meng Fang, Zichao Zeng, and Tao Cheng. Where would i go next? large language models as human mobility predictors. *arXiv preprint arXiv:2308.15197*, 2023. URL https://arxiv.org/abs/2308.15197.

Zixuan Wang, Xiang Zhang, Yuchuan Zhou, Yiyi Jiang, and Haibin Xu. Exploring functional zone-dependent nonlinear associations between objective features and subjective perceptions: A case study in beijing. *International Journal of Applied Earth Observation and Geoinformation*, 142: 104682, 2025. doi: 10.1016/j.jag.2025.104682.

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

Guosheng Yang, Yunlei Su, Minwei Zhao, Chaosu Li, and Cai Wu. Beyond single snapshots: Quantifying multi-scale heterogeneity from street-view imagery and what it reveals about perception. *Cities*, 175:107227, 2026. ISSN 0264-2751. doi: https://doi.org/10.1016/j.cities.2026.107227. URL https://www.sciencedirect.com/science/article/pii/S0264275126004592.

Dazhou Yu, Riyang Bao, Ruiyu Ning, Jinghong Peng, Gengchen Mai, and Liang Zhao. Spatial-RAG: Spatial retrieval augmented generation for real-world geospatial reasoning questions. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 11094–11112, San Diego, California, United States, July 2026. Association for Computational Linguistics. doi: 10.18653/v1/2026.findings-acl.539. URL https://aclanthology.org/2026.findings-acl.539/.

Liangqi Yuan, Dong-Jun Han, Christopher Brinton, and Sabine Brunswicker. LLMAP: LLM-assisted multi-objective route planning with user preferences. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pp. 7866–7894, Suzhou, China, November 2025. Association for Computational Linguistics. doi: 10.18653/v1/2025.findings-emnlp.416. URL https://aclanthology.org/2025.findings-emnlp.416/.

Jiaxin Zhang, Yunqin Li, Tomohiro Fukuda, and Bowen Wang. Urban safety perception assessments via integrating multimodal large language models with street view images. *Cities*, 165:106122, 2025. ISSN 0264-2751. doi: 10.1016/j.cities.2025.106122.

Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology*, 5(3):38:1–38:55, 2014. doi: 10.1145/2629592.

Baichuan Zhou, Haote Yang, Dairong Chen, Junyan Ye, Tianyi Bai, Jinhua Yu, Songyang Zhang, Dahua Lin, Conghui He, and Weijia Li. UrBench: A comprehensive benchmark for evaluating large multimodal models in multi-view urban scenarios. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 10707–10715, 2025. doi: 10.1609/aaai.v39i10.33163. Issue 10.

A Additional Methodological Details

A.1 Request Compilation and Hybrid POI-RAG

The controller compiles the visit intent x^{visit} as

$$\text{Parse}(x^{\text{visit}}) = \big(m_s, m_t, \mathcal{R}^{\text{det}}, \mathcal{R}^{\text{flex}}\big), \tag{3}$$

where m_s and m_t are the origin and destination mentions, $\mathcal{R}^{\text{det}}$ contains deterministic waypoints that name specific places, and $\mathcal{R}^{\text{flex}}$ contains flexible visit requests. Place-name matching and geocoding map the terminal mentions to road-network nodes s and t and map each deterministic waypoint to a POI o_i. If geocoding fails or returns an ambiguous result, the request is returned for disambiguation before route optimization.

For each flexible request r_j, the text encoder produces a query vector $\mathbf{q}_j = f(r_j)$. Each POI document c_o contains the available metadata and cleaned review text for POI o. We first build a reference route

$$\Gamma = \text{RouteRef}\big(s, t, \{o_i\}_{i=1}^{n_{\text{det}}}\big), \tag{4}$$

which connects the origin, destination, and deterministic visits. If the request gives an explicit spatial scope, that scope determines the corridor width δ_j; otherwise, the system uses its default near-route width. The spatial candidate set is

$$\mathcal{P}_j^{\text{sp}} = \big\{o \in \mathcal{P} : \text{dist}(o, \Gamma) \le \delta_j\big\}. \tag{5}$$

This filter limits where a flexible visit may be selected. It does not set a hard upper bound on the length of the final route.

The semantic similarity score is

$$s_j^{\text{sem}}(o) = \cos\big(\mathbf{q}_j, f(c_o)\big), \tag{6}$$

and dense retrieval is restricted to $\mathcal{P}_j^{\text{sp}}$. It returns

$$\mathcal{P}_j^{(M)} = \text{TopM}\big(\mathcal{P}_j^{\text{sp}}, s_j^{\text{sem}}\big). \tag{7}$$

The listwise reranker receives r_j and, for each retrieved candidate, its POI metadata, review evidence, semantic score, distance to Γ, and corridor width. We write this evidence as

$$\begin{aligned} x_j(o) &= \big(o, c_o, s_j^{\text{sem}}(o), \text{dist}(o, \Gamma), \delta_j\big), \\ \rho_j &= \text{LLMRank}\big(r_j, [x_j(o)]_{o \in \mathcal{P}_j^{(M)}}\big), \\ C_j &= \text{TopK}(\rho_j). \end{aligned} \tag{8}$$

We use $M = 10$ and retain $K = 5$ candidates. Each deterministic waypoint forms a singleton candidate anchor group, whereas each flexible visit forms one top-K candidate anchor group. The complete set of candidate anchor groups is

$$\mathcal{C} = \big\{\{o_i\} : 1 \le i \le n_{\text{det}}\big\} \cup \big\{C_j : 1 \le j \le n_{\text{flex}}\big\}. \tag{9}$$

The route optimizer treats every C_j as a retained candidate set. It does not use the reranking position or semantic score within C_j. Hybrid POI-RAG therefore restricts the candidate space, and the route optimizer selects among these candidates using route cost. The solver's exactness guarantee applies to this fixed candidate space, not to semantic satisfaction of the original natural-language request.

POI evaluation and worked example. We evaluate POI retrieval on 127 route-aware requests manually written from sampled Beijing routes and recurring needs reflected in Amap reviews. The requests cover common en-route needs such as buying water or breakfast, finding a quick meal, choosing restaurants by distinctive dishes, and making a brief sightseeing stop. Three annotators, separate from the user-study participants, independently selected five POIs per request using route proximity, POI metadata, and review evidence; discussion resolved disagreements. We report Precision@5, Hit@5, and MRR against the resulting human Top-5 reference sets. Source reviews are not redistributed because of platform restrictions.

Figure 6 follows one evaluation request from the original query to the final POI candidate group, illustrating route-aware filtering, dense retrieval, and evidence-aware reranking.

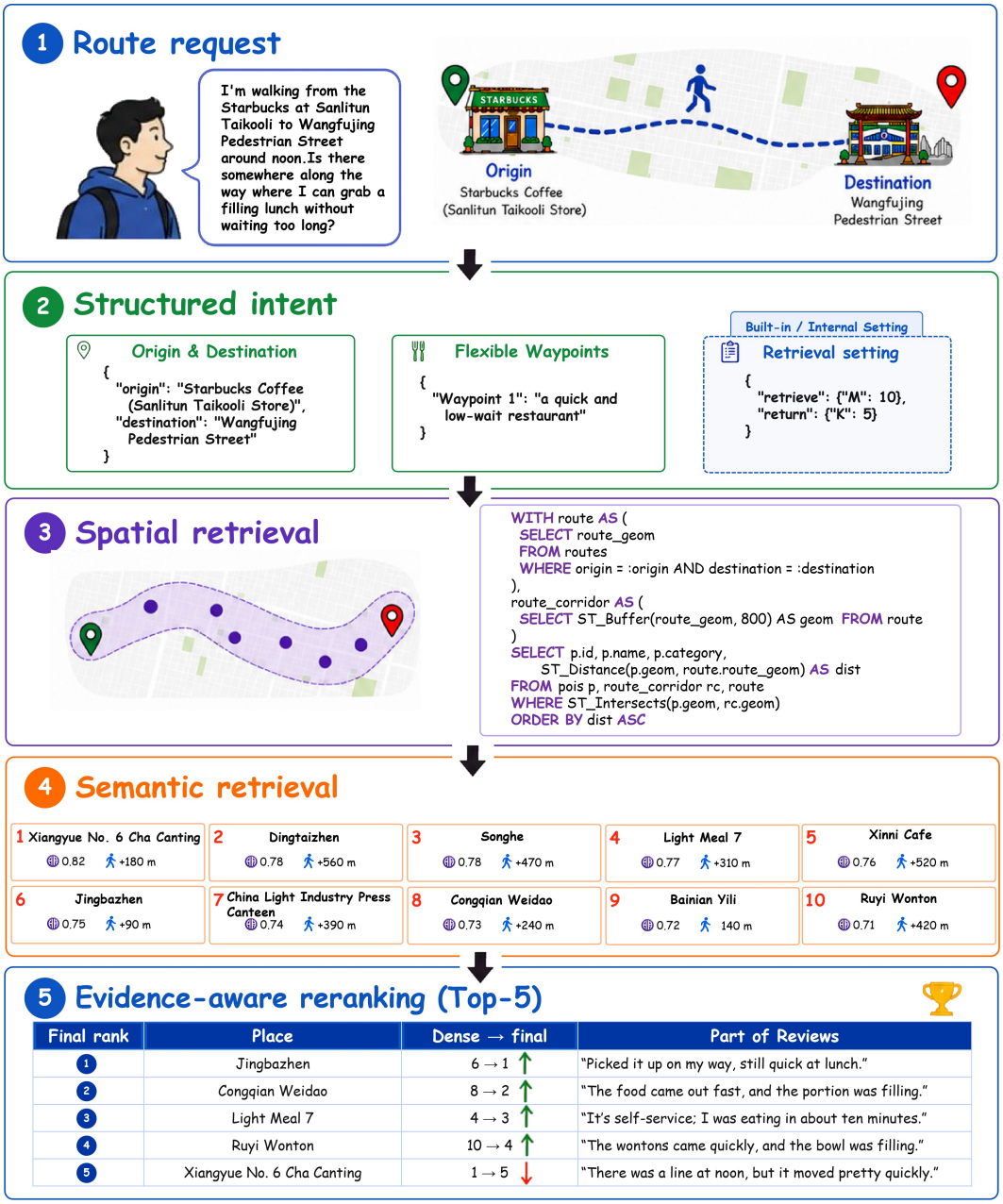

Figure 6: End-to-end example from the POI retrieval evaluation set. Similarity values are cosine similarities; reported distances are measured from the preliminary route.

A.2 PERSONA-CONDITIONED PERCEPTUAL COST MODELING

Place Pulse 2.0 provides pairwise judgments rather than directly comparable absolute labels (Salesses et al., 2013; Dubey et al., 2016). We estimate converged TrueSkill scores and map them to $[0, 1]$ separately for *Beautiful*, *Safe*, *Lively*, and *Wealthy*. After removing overexposed, underexposed, and heavily occluded views, we sample 20,000 images across the target range. We then fine-tune Qwen2.5-VL-7B with LoRA to predict $y_d(I) \in [0, 1]$ for image I and perception dimension $d \in \mathcal{D} = \{B, S, L, W\}$. Evaluation uses a held-out Place Pulse split that is not used for training. It measures agreement with Place Pulse judgments and is not a separate validation study on Beijing street-view imagery.

We next convert the preferred route experience x^{exp} into perception weights using an LLM-as-a-Jury procedure (Verga et al., 2024; Qian et al., 2026). We separately query four proposer models: Qwen3-Max, Claude Sonnet 4.5, GPT-5, and GLM-4.6. For dimension d, proposer $m \in \mathcal{M}$ returns a score $a^{(m)}_{x^{\text{exp}},d} \in [0, 10]$ and a rationale $r^{(m)}_{x^{\text{exp}},d}$. Gemini-2.5 Pro serves as the external judge and selects

$$\begin{aligned} m^\star_d &= \arg\max_{m \in \mathcal{M}} J_d\big(x^{\text{exp}}, a^{(m)}_{x^{\text{exp}},d}, r^{(m)}_{x^{\text{exp}},d}\big), \\ w_{x^{\text{exp}},d} &= \frac{a^{(m^\star_d)}_{x^{\text{exp}},d}}{10}. \end{aligned} \tag{10}$$

Selection is performed separately for each perception dimension. Thus, $\mathbf{w}_{x^{\text{exp}}} \in [0, 1]^4$ combines the dimension-wise scores selected by the judge; it is not a new weight vector generated directly by the judge. Human reference ratings use the same 0–10 scale but are collected independently and used only for component-level evaluation. They are not provided to the judge or the route optimizer.

Let $\mathcal{I}(e)$ be the street-view images associated with road edge e. When $|\mathcal{I}(e)| > 0$, the edge-level perception score is

$$\hat{y}_d(e) = \frac{1}{|\mathcal{I}(e)|} \sum_{I \in \mathcal{I}(e)} y_d(I). \tag{11}$$

For an edge without direct street-view coverage, the system imputes a score from adjacent observed edges. This provides a score for every edge in the study graph but introduces uncertainty in unobserved areas; the current route objective does not model this uncertainty explicitly.

We convert each positive perception score into a disutility and apply the cubic transformation used in Eq. 2. To keep the equations within one column, let $p_{x^{\text{exp}}}(e)$ denote the weighted perception penalty:

$$\begin{aligned} \hat{s}_d(e) &= 1 - \hat{y}_d(e), \\ \phi(r) &= r^3, \\ p_{x^{\text{exp}},d}(e) &= w_{x^{\text{exp}},d}\, \phi\big(\hat{s}_d(e)\big), \\ p_{x^{\text{exp}}}(e) &= \sum_{d \in \mathcal{D}} p_{x^{\text{exp}},d}(e), \\ c_{x^{\text{exp}}}(e) &= \ell(e)\big[1 + p_{x^{\text{exp}}}(e)\big]. \end{aligned} \tag{12}$$

Here, $\ell(e)$ is the physical edge length. This construction yields the perceptual-cost graph $G_{x^{\text{exp}}} = (V, E, c_{x^{\text{exp}}})$. Because all lengths, weights, and penalties are nonnegative, its edge costs are nonnegative.

A.3 EXACT ROUTE OPTIMIZATION VIA OPEN GTSP

The route optimizer receives the perceptual-cost graph $G_{x^{\text{exp}}}$, terminals s and t, and candidate anchor groups $\mathcal{C} = \{C_1, \ldots, C_N\}$. For a candidate POI z, $\text{acc}(z) \in V$ is its attached road node. The key-node set is

$$\mathcal{K} = \{s, t\} \cup \big\{\text{acc}(z) : z \in C_j,\ 1 \le j \le N\big\}. \tag{13}$$

Algorithm 1 Exact Grouped Route Solver

Require: $G_{x^{\exp}}$, terminals s, t, groups $\mathcal{C}$
Ensure: Route $\pi^\star$ and selected anchors $\mathbf{z}^\star$
1: Build the key-node set $\mathcal{K}$
2: **for** each $v \in \mathcal{K}$ **do**
3: Run Dijkstra to obtain closure costs and witness paths
4: **end for**
5: Mark unreachable closure transitions as infeasible
6: Initialize all singleton states $F(\{j\}, j, z)$
7: **for** each subset S in increasing order of size **do**
8: Update every feasible state $F(S, j, z)$
9: **end for**
10: Add t to the best full state and backtrack $\mathbf{z}^\star$
11: Expand the selected closure edges to recover $\pi^\star$

We run Dijkstra's algorithm from each key node on $G_{x^{\exp}}$ to build a shortest-path closure. For $v_i, v_j \in \mathcal{K}$, its cost is

$$\begin{aligned} c_{x^{\exp}}(\pi) &= \sum_{e \in \pi} c_{x^{\exp}}(e), \\ D_{x^{\exp}}(v_i, v_j) &= \min_{\pi : v_i \rightsquigarrow v_j} c_{x^{\exp}}(\pi). \end{aligned} \tag{14}$$

We also store a witness path for each finite closure cost. An unreachable pair is treated as infeasible. In the equations below, a candidate used as an argument of $D_{x^{\exp}}$ denotes its access node, and we write $D_{\exp} = D_{x^{\exp}}$ for brevity. Let σ be a permutation of $\{1, \dots, N\}$. The open grouped-route problem is

$$\begin{aligned} \min_{\substack{\sigma \\ z_j \in C_j}} \quad & D_{\exp}\big(s, z_{\sigma(1)}\big) \\ & + \sum_{k=1}^{N-1} D_{\exp}\big(z_{\sigma(k)}, z_{\sigma(k+1)}\big) \\ & + D_{\exp}\big(z_{\sigma(N)}, t\big). \end{aligned} \tag{15}$$

Let $F(S, j, z)$ be the minimum cost of starting at s, visiting one candidate from every group in S, and ending at $z \in C_j$. The Held–Karp-style recurrence (Held & Karp, 1962) is

$$\begin{aligned} F(\{j\}, j, z) &= D_{\exp}(s, z), \\ F(S, j, z) &= \min_{\substack{i \in S \setminus \{j\} \\ q \in C_i}} \Big\{ F(S \setminus \{j\}, i, q) + D_{\exp}(q, z) \Big\}. \end{aligned} \tag{16}$$

The optimum is obtained by adding the destination cost to every full state:

$$\min_{\substack{1 \le j \le N \\ z \in C_j}} \Big\{ F(\{1, \dots, N\}, j, z) + D_{\exp}(z, t) \Big\}. \tag{17}$$

Backtracking gives the selected candidate anchors and their order. Expanding the stored witness paths then recovers a road-network-valid route on the original graph. When $N = 0$, the method reduces to the shortest path from s to t on $G_{x^{\exp}}$.

Let $Q = |\mathcal{K}|$. Building the shortest-path closure takes

$$\mathcal{O}\big(Q(|E| + |V|) \log |V|\big) \tag{18}$$

time with the stated Dijkstra implementation. In the worst case, the subset dynamic program uses $\mathcal{O}(2^N Q^2)$ transitions. Retrieval limits each flexible group to K candidates, but the solver remains exponential in the number of requested groups. Assuming nonnegative edge costs, nonempty candidate groups, and finite reachability for the selected transitions, the dynamic program returns a global optimum of the open grouped-route problem over the fixed graph, edge costs, and retained candidate groups.



Figure 7: Study area and multi-source data. Panel a shows Beijing's urban core. Panel b shows Amap POIs on the road network. Panel c shows georeferenced Baidu Street View panoramas. Panel d shows the twelve persona profiles used for personalized evaluation.

B Additional Experimental Details and Results

B.1 Data Sources and Implementation Details

Data acquisition. The study area covers Beijing's urban core and combines road-network, POI, review, street-view, perception-supervision, persona, and evaluation data. Table 4 summarizes the sources and their roles.

Table 4: Summary of data sources and evaluation samples used in NavMind.

Data component	Source or construction	Scale	Role in NavMind
Road network	OpenStreetMap, processed with OSMnx	Beijing urban core	Spatial grounding and route optimization
POIs	Amap	5,829 POIs	Geographic entities for deterministic and flexible visits
POI reviews	Amap	184,281 cleaned reviews	Semantic evidence for retrieval and reranking
Street-view imagery	Baidu Street View	7,083 georeferenced panoramas	Road-edge perception scoring
Perception supervision	Place Pulse 2.0	20,000 images; 4:1 train–validation split	Fine-tuning and validation of the perception model
Persona profiles	Manually curated from social-media and lifestyle platforms	12 personas	Persona-conditioned weighting and route evaluation
POI retrieval evaluation	Manually annotated requests	127 route-aware requests	Spatial–semantic retrieval evaluation
Route simulation	Sampled origin–destination pairs	4,000 OD pairs; 36,000 optimized routes	Large-scale routing analysis
Human assessment	Within-subject evaluation	36 participants; 9 cases; 324 paired comparisons	Route satisfaction and preference evaluation

The road graph is derived from OpenStreetMap and processed with OSMnx (OpenStreetMap contributors, 2026; Boeing, 2017). The POI collection contains 5,829 Amap records and 184,281 cleaned reviews. The visual collection contains 7,083 georeferenced Baidu Street View panoramas associated with nearby road segments. Place Pulse 2.0 provides the human-perception supervision used to train the street-view scorer.

For the persona profiles, 12 human annotators collected more than 300 travel guides and route recommendations for the study area from major social-media and lifestyle platforms, including Xiaohongshu, Weibo, Dianping, X, and Instagram. Qualitative thematic analysis was used to extract recurring psychographic traits and spatial preferences while reducing overlap between profiles, yielding the twelve personas shown in Figure 7(d). Figure 12 provides an example of the source material used in this curation process.

Data preprocessing. The OpenStreetMap graph is geometrically simplified, topological artifacts are repaired, and coordinates are projected to metric space. Raw Amap POIs and reviews are cleaned to remove duplicates, invalid coordinates, and empty text; POI documents are embedded offline for semantic retrieval. Baidu panoramas are aligned to nearby road segments, and edges without direct street-view coverage receive estimates propagated from adjacent observed edges. These preprocessing steps produce the road, semantic, and perceptual representations used by the routing pipeline.

Table 5: Model and solver roles in NavMind.

Component	Model or algorithm	Role and configuration
Intent parsing	GPT-5	Structured local intent and route-style extraction
POI retrieval	Spatial filter and dense encoder	Request-conditioned corridor and intermediate $M = 10$
POI reranking	GPT-5	Listwise ranking with reviews, similarity, and route distance
Street perception	Qwen2.5-VL-7B with LoRA	20,000 Place Pulse images and four positive dimensions
Persona proposers	Qwen3-Max, Claude Sonnet 4.5, GPT-5, GLM-4.6	Independent scores and rationales
Persona judge	Gemini-2.5 Pro	Dimension-wise selection from proposer outputs
Route optimization	Dijkstra and Held-Karp dynamic programming	Exact closure and grouped route reconstruction with $K = 5$

Evaluation protocol. Component evaluation covers POI retrieval, street-perception scoring, and persona-weight alignment. POI retrieval is evaluated on 127 route-aware requests with human Top-5 reference sets; street perception is evaluated on the held-out Place Pulse validation split; and persona weights are compared with human reference ratings for the twelve profiles. Route analysis uses 4,000 OD pairs, equally distributed across four shortest-path distance bins, under four single-dimension objectives and five personas, yielding $4{,}000 \times (4 + 5) = 36{,}000$ routes. The human assessment uses nine cases and 36 participants, yielding 324 paired comparisons. The route and participant analyses are descriptive because the design does not support population-level inference.

Each route-aware POI request specifies an origin and destination context and one flexible visit need. Reference candidates are assessed using route proximity, POI metadata, and review evidence. Precision@5 measures the share of relevant candidates in the first five results. Hit@5 indicates whether at least one relevant candidate appears in the first five results. MRR measures the reciprocal rank of the first relevant candidate.

B.2 Perception and Persona Alignment

The fine-tuned VLM obtains Pearson correlations of 0.945, 0.933, 0.921, and 0.905 for Beautiful, Safe, Lively, and Wealthy. The Beautiful MAE is 0.157, and all reported R^2 values exceed 0.81. Figure 8 shows that the predictions retain the target distribution spread. These metrics assess calibration to Place Pulse labels. Application to Baidu Street View introduces domain and capture-condition differences that are not separately quantified in the current study.

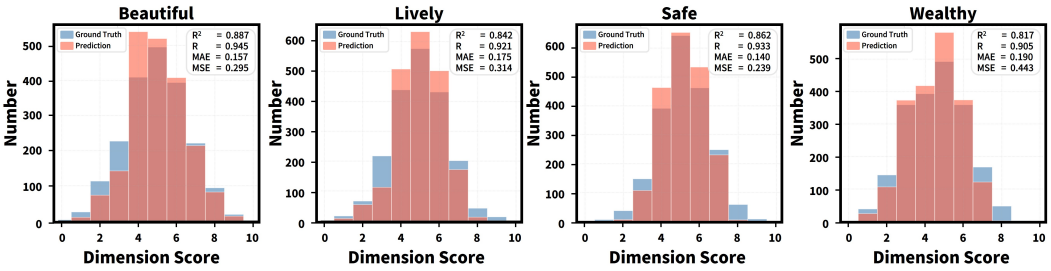

Figure 8: Ground-truth and predicted score distributions for Beautiful, Lively, Safe, and Wealthy on the held-out Place Pulse validation split.

Table 6: Agreement with human persona ratings on a 0 to 10 scale. B, S, L, and W report dimension-wise MAE for Beautiful, Safe, Lively, and Wealthy, respectively.

Method	MAE	Spearman	B	S	L	W
Jury-selected output	**0.950**	**0.948**	**0.850**	**0.283**	1.067	**1.600**
Qwen3-Max	1.029	0.881	1.050	0.400	0.900	1.767
Claude Sonnet 4.5	1.271	0.929	0.933	0.983	1.350	1.817
GPT-5	1.617	0.830	2.550	1.317	**0.750**	1.850
GLM-4.6	2.042	0.864	2.250	2.367	1.050	2.500

The jury-selected output obtains the lowest overall MAE and the highest rank agreement on the evaluated profiles. It also remains close to the four-proposer median after normalization, with an MAE of 0.083 and a Spearman correlation of 0.931. The table labels the result as a jury-selected output because Gemini acts as a selector rather than producing the final dimension values directly. The comparison is limited to the evaluated personas and does not remove prompt or model-version uncertainty.

B.3 End-to-end Comparison

Experimental configuration. The 108 route-generation tasks were constructed following the annotation protocol used in our POI Benchmark (A.1), by combining annotated POI requirements with perceptual-style preferences to form complete routing requests. Each task specified two POI requirements and one perceptual style. All methods used GPT-5 with temperature 0, together with the same POI database and road network. The baselines followed their original POI selection and ordering procedures and connected the selected POIs using shortest-road paths, while NavMind jointly optimized POI selection and perceptual route cost.

Intent Completion. Intent Completion measures how well the selected POIs satisfy the requested visit intents. For task t with N_t visit intents, let $p_{tm,1}, \ldots, p_{tm,N_t}$ denote the POIs selected by method m, and let $R_{t,i}$ denote the Top-K reference set for intent i. We match the selected POIs to the visit intents one-to-one and choose the assignment that maximizes the number of satisfied intents:

$$C_{tm} = \frac{1}{N_t} \max_{\pi \in S_{N_t}} \sum_{i=1}^{N_t} \mathbf{1}\left[p_{tm,\pi(i)} \in R_{t,i}\right].$$

Here, S_{N_t} denotes all possible one-to-one assignments between the selected POIs and the visit intents. Thus, $C_{tm} \in [0, 1]$ represents the fraction of visit intents satisfied for task t. The overall Intent Completion of method m is

$$\mathrm{IC}_m = \frac{100}{T} \sum_{t=1}^{T} C_{tm},$$

where T is the number of evaluated tasks.

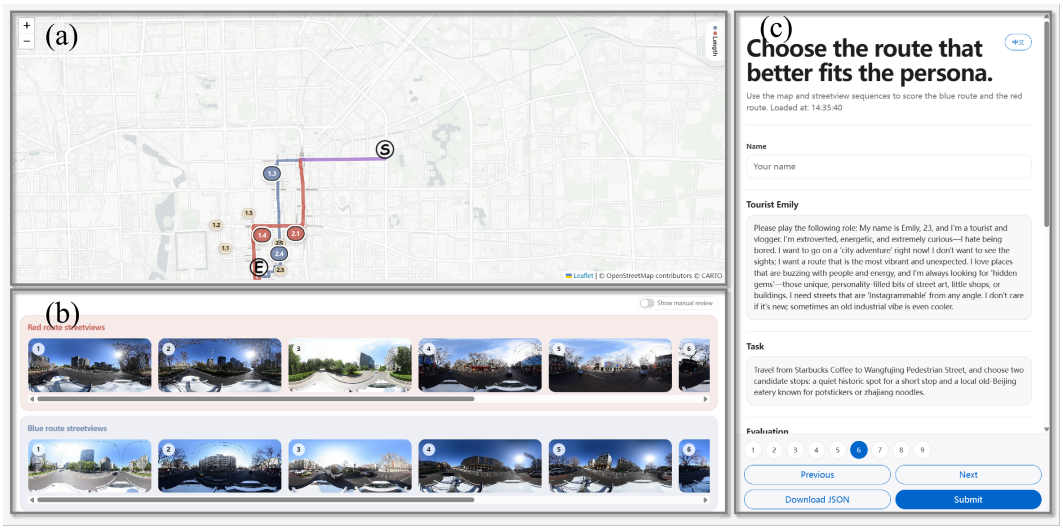

Figure 9: Blinded user-assessment interface with route maps, route-specific street views, persona and task prompts, and rating questions.

Extra Distance. Extra Distance measures the additional travel required by a generated route relative to the shortest-road path between the same origin and destination. Let L_{tm} denote the length of the route produced by method m for task t, and let $d_G(o_t, d_t)$ denote the corresponding shortest-road distance. The task-level Extra Distance is

$$\Delta L_{tm} = L_{tm} - d_G(o_t, d_t).$$

The mean Extra Distance of method m is

$$\mathrm{ED}_m = \frac{1}{T} \sum_{t=1}^{T} \Delta L_{tm}.$$

Win Rate of Preference. Preference Match measures how well the visual appearance of a generated route matches the requested perceptual style. Street-view images sampled along each route were presented anonymously to a separate GPT-5 evaluator together with the requested style. The evaluator was not given method names, POI information, or route distances. Each pair of methods was evaluated twice with reversed presentation order to reduce order bias.

The two judgments were merged into a single pairwise outcome. Consistent preferences produced a win or loss, while two ties produced a draw. For method m compared with method n, Win Rate of Preference is defined as

$$\mathrm{WRP}_{m,n} = 100 \frac{W_{m,n} + 0.5 D_{m,n}}{W_{m,n} + D_{m,n} + L_{m,n}},$$

where $W_{m,n}$, $D_{m,n}$, and $L_{m,n}$ denote the numbers of wins, draws, and losses of method m against method n, respectively. A win contributes 1, a draw contributes 0.5, and a loss contributes 0.

B.4 User Assessment Protocol

Figure 9 presents the evaluation platform. The map shows two anonymized routes, endpoints, and candidate or selected POIs; the two street-view sequences provide visual evidence for the corresponding routes; and the questionnaire presents the persona, task, ratings, route choice, and free-text explanation.

Thirty-six participants evaluated nine cases in a within-subject design, yielding 324 paired comparisons. The three personas were selected because their student, tourist, and programming-related attributes were familiar to the participant group, making the assigned roles easier to adopt. Each persona was paired with tasks containing zero, one, and two flexible waypoints. For each case,

participants inspected the persona, task, routes, street views, and POIs before rating overall route satisfaction, preferred-experience alignment, and POI appropriateness when applicable. Exact route lengths were then revealed for the detour-acceptability rating, followed by the final route preference and free-text decision rationale. Route identities remained hidden during evaluation.

The questionnaire separated perceived route benefits from travel cost. Before route lengths were shown, participants independently rated the overall satisfaction of both routes, compared their alignment with the preferred route experience, and evaluated POI appropriateness in waypoint cases. They then viewed the exact lengths, rated detour acceptability, reported their final route preference, and explained the decision in free text. Table 7 summarizes the questionnaire and response scales.

Table 7: Questionnaire items and response scales. Q1 and Q2 independently rated the blue and red routes; Q4 applied only to waypoint cases, and Q5 was answered after route lengths were revealed.

Score	Q1–Q2: Overall route satisfaction	Q3: Preferred-experience alignment	Q4: POI appropriateness	Q5: Detour acceptability
1	Very dissatisfied	Blue clearly better	Very inappropriate	Completely unacceptable
2	Dissatisfied	Blue better	Inappropriate	Unacceptable
3	Slightly dissatisfied	Blue slightly better	Slightly inappropriate	Slightly unacceptable
4	Neutral	Similar	Neutral	Neutral
5	Slightly satisfied	Red slightly better	Slightly appropriate	Slightly acceptable
6	Satisfied	Red better	Appropriate	Acceptable
7	Very satisfied	Red clearly better	Very appropriate	Completely acceptable
Q6	Final route preference: Blue route / Tie / Red route			
Q7	Decision rationale: Free-text explanation			

The free-text prompt asked why the preferred route was better and why the alternative was worse, including whether the route environment, practical travel conditions, waypoint POIs, and detour matched the assigned persona and task. These explanations form the basis of the attribution analysis below.

B.5 Detailed User Assessment Results

Table 8 reports the detailed case-level results. NavMind received higher mean satisfaction scores in eight of the nine cases and overall (5.26 vs. 3.48; $\Delta = +1.78$), with 208 NavMind choices, 60 ties, and 56 shortest-path choices. C9 showed the largest gain ($+3.67$), whereas C4 was the only negative case (-0.44). The overall means for preferred-experience alignment, POI appropriateness, and detour acceptability were 5.16, 5.35, and 5.20, respectively.

Table 8: Case-level ratings and final route preferences. "WP" is the number of flexible waypoints. Satisfaction is rated on a 1–7 scale. "Exp. align." is comparative (1: shortest path; 4: similar; 7: NavMind). POI appropriateness and detour acceptability use the scales in Table 7. "N/T/S" reports NavMind, tie, and shortest-path choices.

(a) Route satisfaction and detour

Case	Persona	WP	Detour (%)	Shortest satisfaction	NavMind satisfaction	Δ
C1	Student A-Jie	0	6.4	3.22	6.00	+2.78
C2	Student A-Jie	1	8.2	3.67	4.89	+1.22
C3	Student A-Jie	2	22.6	4.00	4.89	+0.89
C4	Tourist Emily	0	16.1	4.22	3.78	-0.44
C5	Tourist Emily	1	11.5	3.56	5.56	+2.00
C6	Tourist Emily	2	19.4	3.78	5.11	+1.33
C7	Programmer Zhao Lei	0	1.3	2.67	5.33	+2.67
C8	Programmer Zhao Lei	1	0.6	4.00	5.89	+1.89
C9	Programmer Zhao Lei	2	0.6	2.22	5.89	+3.67
Overall	All	–	9.6	3.48	5.26	+1.78

(b) Experience ratings and final preference

Case	Exp. align.	POI appropriateness	Detour acceptability	N/T/S
C1	6.11	–	6.22	36/0/0
C2	5.44	5.44	5.22	24/4/8
C3	4.33	5.33	4.33	20/4/12
C4	3.67	–	3.67	8/20/8
C5	5.33	5.78	5.44	28/0/8
C6	5.11	4.89	4.67	20/12/4
C7	5.22	–	5.67	28/0/8
C8	5.33	5.56	5.44	12/20/4
C9	5.89	5.11	6.11	32/0/4
Overall	5.16	5.35	5.20	208/60/56

Table 9: Free-text attribution by final route preference. Cells report count (percentage within the corresponding preference group). "Route environment" indicates better alignment with the preferred experience; "Waypoint POIs" indicates better support for the visit intent; and "Route practicality" covers detour, directness, and fine-grained walking conditions. Codes are non-exclusive.

Final preference	N	Evidence favoring the selected route			Tie rationale
		Route environment	Waypoint POIs	Route practicality	No clear net advantage
NavMind	208	132 (63.5%)	76 (36.5%)	120 (57.7%)	–
Tie	60	–	–	–	56 (93.3%)
Shortest path	56	32 (57.1%)	12 (21.4%)	36 (64.3%)	–

Table 9 summarizes the non-exclusive manual coding of the original Chinese explanations. Among NavMind choices, 63.5% cited better route-environment alignment, 36.5% cited better waypoint support, and 57.7% cited route practicality. Among shortest-path choices, the corresponding shares were 57.1%, 21.4%, and 64.3%, while 56 of the 60 ties reported no clear net advantage. Because one explanation could receive multiple codes, percentages within a row do not sum to 100%.

C Discussion and Limitations

Extensibility. We instantiate the perceptual basis with four established dimensions of urban perception, Beautiful, Safe, Lively, and Wealthy, following prior work on quantifying how people perceive streetscapes (Salesses et al., 2013; Dubey et al., 2016). These dimensions provide a well-grounded basis for learning street-level perception from large-scale human judgments. Our results on both individual attributes and persona-conditioned combinations show that language-derived preferences can be translated into interpretable combinations of measurable street-level qualities. This formulation naturally extends to broader preference descriptions: general preferences can be represented through individual attributes or their combinations, while more specialized street properties can be introduced as additional perceptual signals when suitable supervision is available. For example, sky openness may be partially reflected by existing perceptual dimensions, or modeled explicitly using a dedicated scorer trained with specialized annotations. Since NavMind separates preference interpretation, perceptual estimation, and route construction, such attributes can be incorporated into the same perceptual-cost and routing pipeline. The present cost formulation favors higher attribute values; preferences for lower or target values can be represented by changing the corresponding attribute-to-cost mapping while retaining the same downstream routing procedure.

Practical efficiency and deployment. NavMind's architecture separates preference interpretation, perceptual estimation, and route construction, allowing street-level perceptual scores to be precomputed and reused across users. With perceptual scores cached, online processing handles request interpretation and POI grounding, followed by edge-cost construction and route optimization. New perceptual attributes can likewise be added at the perception layer without redesigning the downstream solver.

Evaluation scope. Our experiments focus on one area of Beijing, one POI and review ecosystem, and one street-view source. The human assessment also evaluates a fixed set of routing cases, while several language components rely on closed-source models whose behavior may vary across versions. Broader validation across cities, data sources, user groups, and dynamic urban conditions would further establish the generality of the framework.

Data and code availability. Amap POI records and user reviews and Baidu Street View panoramas were accessed through provider-controlled services and remain subject to the respective platforms' licensing and redistribution terms. We therefore do not redistribute raw platform-owned records, user-generated review text, or street-view imagery. The implementation code, prompts, evaluation scripts, and derived artifacts needed to reproduce the reported experiments will be released subject to these data restrictions.

D Prompts and Additional Qualitative Materials

Hybrid POI-RAG Module

(a) Visit-intent structuring prompt

```
You are the visit-intent structuring module of a route-navigation. Convert an open-ended route request into a
structured record of the origin, destination, and intermediate visit requirements.

Input:
<ROUTE_REQUEST>

Task:
1. Extract the origin and destination explicitly stated in the request.
2. Extract every intermediate visit requirement mentioned in the request.
3. Classify each intermediate visit as:
  - deterministic, when it refers to a specific named place; or
  - flexible, when multiple POIs could satisfy the requirement.
4. Preserve all explicitly stated visit-specific information, including place names, POI types, functional needs, descriptive
  attributes, and spatial conditions.
5. Use only information stated in the request. Do not infer missing place names, categories, locations, coordinates,
attributes, or spatial constraints.

Return only the following structured record:
{
 "origin": {
  "mention": "..."
  "place_name": "...",
 },
 "destination": {
  "mention": "..."
  "place_name": "...",
 },
 "deterministic_visits": [
  {
   "id":waypoint...,
   "mention": "...",
   "place_name": "...",
   "spatial_constraint": "..."
  }, ...
 ],
 "flexible_visits": [
  {
   "id":waypoint...,
   "mention": "...",
   "poi_type": "...",
   "requirements": ["...", "..."],
   "spatial_constraint": "..."
  },...
 ]
}
```

Hybrid POI-RAG Module

(b) Evidence-aware reranking prompt

```
You are the evidence-aware listwise POI reranker of a route-navigation system. Rank only the supplied candidate POIs
for one flexible visit requirement.

The candidates have already passed route-aware spatial filtering and dense semantic retrieval.

Input:
<FLEXIBLE_VISIT_REQUIREMENT>

<TOP_M_CANDIDATES>

Each candidate contains:
- POI ID;
- POI name and category;
- distance to the reference route;
- semantic-similarity score;
- available POI metadata;
- relevant review excerpts.

Ranking criteria:
1. Prioritize candidates whose metadata and reviews provide direct evidence that the POI satisfies the requested visit
requirement.
2. Consider all explicitly requested functional and descriptive attributes.
3. Lower the rank of candidates whose review evidence is negative, negated, contradictory, or otherwise inconsistent
with the  request.
4. Use semantic similarity as supporting evidence, but do not let  it override explicit metadata or review evidence.
5. When candidates have comparable semantic and review evidence,  prefer the candidate closer to the reference route.
6. Do not introduce, rename, merge, or infer any POI outside the  supplied candidate set.
7. Do not assume attributes that are not supported by the supplied metadata or reviews.

Return at most five candidates, or all candidates if fewer than five are supplied.

Return only the following structured result:
{
 "ranked_pois": [
  {
   "rank": 1,
   "poi_id": "...",
   "justification": "One concise sentence stating how the supplied evidence supports the visit requirement and route
compatibility."
  }
 ]
}
```

Figure 10: Prompt used by the Hybrid POI-RAG module.

Preception Cost Module

You are a professional urban perception and behavioral modeling expert. Your task is to deeply analyze a specific character persona and "role-play" as this character to infer their intrinsic preference weights for different perceptual dimensions when making urban path choices.

Your output MUST be in JSON format.

[PERCEPTION DIMENSIONS]
This is the list of four perceptual dimensions you must evaluate:
1. Beautiful
2. Safe
3. Lively
4. Wealthy

[TASK INSTRUCTION]
You will be given a [ROLE BACKGROUND (Persona)]. You must "become" this character and analyze your own internal needs. Your task is to:

1. **Reflect on Preferences**: Based on your character background and needs (e.g., cautious, seeking inspiration, functional), determine how important each dimension in the [PERCEPTION DIMENSIONS] list is to your path choice.
2. **Assign Weights**: Assign a "decision weight" (0.0 to 1.0) to each dimension. 0.0 means completely irrelevant, 1.0 means top priority. Please use one decimal place.
3. **Provide Rationale**: For each weight, provide a clear, concise "reasoning" that explains *why* this dimension has this level of importance *to you* (the character), linking it directly to your persona description.

[OUTPUT FORMAT]
You must *only* provide a single JSON object strictly following this structure. Do not include any explanatory text outside the JSON.

```
{
  "persona_id": "UNIQUE_PERSONA_ID", // This ID will be provided in the user message.
  "persona_summary": "(A brief summary of the character's core needs)",
  "dimension_weights": [
    {
      "dimension": "Beautiful",
      "weight": 0.0,
      "reasoning": "(Explain why 'Beautiful' has this weight for the character)"
    },
    {
      "dimension": "Safe",
      "weight": 0.0,
      "reasoning": "(Explain why 'Safe' has this weight for the character)"
    },
    {
      "dimension": "Lively",
      "weight": 0.0,
      "reasoning": "(Explain why 'Lively' has this weight for the character)"
    },
    {
      "dimension": "Wealthy",
      "weight": 0.0,
      "reasoning": "(Explain why 'Wealthy' has this weight for the character)"
    }
  ]
}
```

Figure 11: System prompt used by the perception-cost module.

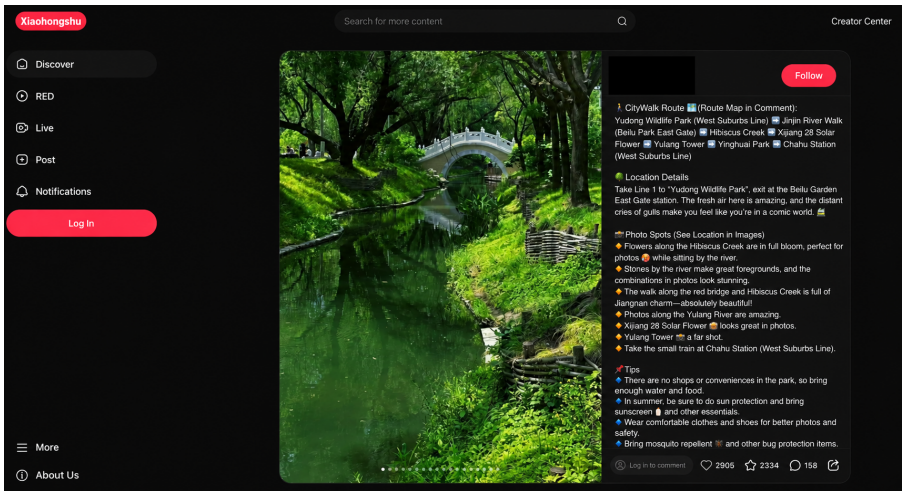

Figure 12: Example of a real-world route recommendation from Xiaohongshu collected during persona curation.



Figure 13: Additional qualitative cases for Li Qiang (top) and Victoria (bottom). Red denotes the shortest-path baseline, blue denotes NavMind, and purple denotes route segments shared by both routes.



Figure 14: Additional qualitative cases for Wang Xiuying (top) and Lin Jing (bottom). Red denotes the shortest-path baseline, blue denotes NavMind, and purple denotes route segments shared by both routes.



Figure 15: Additional qualitative cases for David Chen (top) and Zhao Lei (bottom). Red denotes the shortest-path baseline, blue denotes NavMind, and purple denotes route segments shared by both routes.



Figure 16: Additional qualitative cases for Wang Li (top) and Zhang Guoqiang (bottom). Red denotes the shortest-path baseline, blue denotes NavMind, and purple denotes route segments shared by both routes.



Figure 17: Additional qualitative cases for Chen Ming (top) and Emily (bottom). Red denotes the shortest-path baseline, blue denotes NavMind, and purple denotes route segments shared by both routes.