Universität Zürich[UZH]

Geographisches Institut

UGuideRAG: Intent-Enhanced Retrieval-Augmented Generation with User-Generated Content for Personalized Urban Tourism

GEO 511 Master's Thesis

Author: Jing Tang, 23-743-859

Supervised by: Dr. Inhye Kong

Faculty representative: Prof. Dr. Ross Purves

29.08.2025

Abstract

Citywalk, as an increasingly popular form of urban tourism, emphasizes immersive, diverse, and personalized exploration over conventional sightseeing. These features evolving tourist expectations pose new challenges for intelligent itinerary planning, particularly in capturing the rich experiential attributes of visitor attractions and aligning them with ambiguous and underspecified natural language queries. This thesis proposes UGuideRAG (User-Generated Content-Guided Retrieval-Augmented Generation), a modular framework that leverages user-generated content to construct a comprehensive attraction database, employs large language models for intent-enhanced retrieval and recommendation, and incorporates spatial optimization to ensure coherent itinerary planning. By bridging the gap between partially expressed user goals and the multi-dimensional nature of urban experiences, UGuideRAG enables more insightful and personalized trip recommendations. For walk-centric route planning, UGuideRAG further constructs a scenic pathway database by fusing POI data with geotagged photos to estimate segment-level scenicness using photo density and street interactivity, and integrates this score into a multi-objective route generator that links the candidate attractions while balancing distance, spatial compactness, and accumulated scenic value. Experiments on real-world datasets demonstrate that the proposed framework consistently surpasses existing methods in producing contextually relevant, user-centered, and spatially optimized urban tourism itineraries.

Acknowledgment

I would like to express my sincere gratitude to Dr. Inhye Kong, Prof. Dr. Ross Purves, and Dr. Zhaonan Wang from New York University Shanghai for their invaluable guidance, insightful feedback, and unwavering support throughout this research. Their collective expertise was instrumental in shaping this thesis. It has been a true privilege to learn from them.

Contents

List of Figures

List of Tables

List of Acronyms

UGC	User-Generated Content
VAs/VA	Visitor Attractions
LLMs/LLM	Large Language Models
RAG	Retrieval-Augmented Generation
POIs/POI	Points of Interest
UADC	UGC-based Attraction Database Construction
IER	Intent-Enhanced Retriever
LRR	LLM-based Reranker
CSO	Cluster-aware Spatial Optimization
SPDC	Scenic Pathway Database Construction
MGC	Market-Generated Content
HR	Hit Rate
AM	Average Margin
TD	Total Distance
ST	Spatial Tightness
TSP	Traveling Salesman Problem
PCA	Principal Component Analysis
OSM	OpenStreetMap
DFS	Depth-First Search

1 Introduction

This master's thesis builds upon and extends the author's paper accepted for publication in ACM SIGSPATIAL 2025.

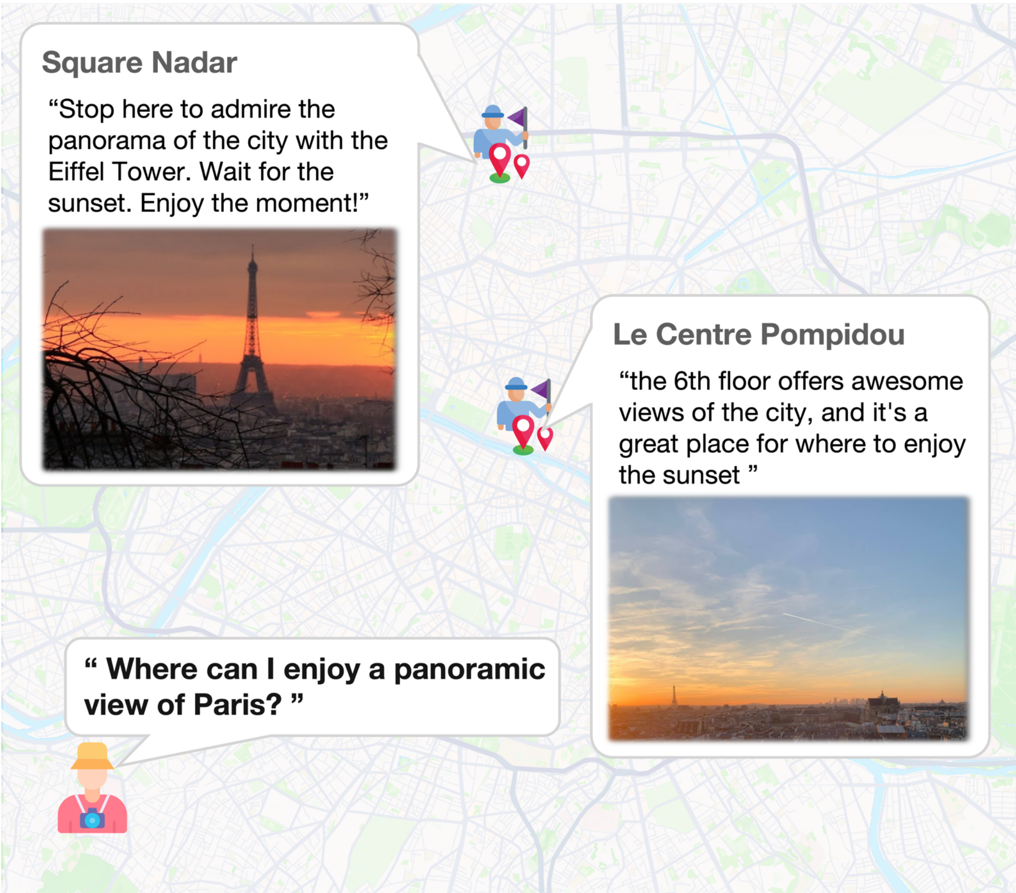


Figure 1.1: *"Where can I enjoy a panoramic view of Paris?"* This figure illustrates how user-generated content can reveal hidden scenic viewpoints that are beyond guidebooks and 2D maps.[1]

With the rapid development of the Internet and communication technologies, social media and user-generated content (UGC) are reshaping the tourism industry (Xiang and Gretzel, 2010). UGC has become a crucial source of information for tourists, supporting activities such as travel planning, destination image construction, and

[1] Photo sources: https://maps.app.goo.gl/fMibardwj3SBE1uM8, https://maps.app.goo.gl/4i48B3rnDuAXVXQf6.

decision-making. Recent statistics show that over 80% of consumers now rely on the internet for travel information, and this percentage continues to grow (Zhang et al., 2019). Compared to marketer-generated content (MGC), UGC is widely perceived by travelers as more reliable, credible, and up-to-date (Li et al., 2023).

Unlike official descriptions, UGC offers authenticity, emotional nuance, and localized insights that help uncover hidden or underrepresented aspects of destinations (Mak, 2017; Marine-Roig and Clavé, 2015; Antonio et al., 2020). As illustrated in Figure 1.1, UGC can reveal detailed and hidden aspects of visitor attractions (VAs) that are often missing from official descriptions or curated travel guides. For instance, while Le Centre Pompidou is widely known for its modern art exhibitions, UGC highlights an alternative facet — its rooftop being appreciated as a scenic viewpoint. At the same time, user reviews surface lesser-known places such as Square Nadar, which offers panoramic views but rarely appears in conventional itineraries. These examples demonstrate how UGC helps uncover both the subtle characteristics of well-known attractions and lesser-known spots in the city.

Building on the rich UGC associated with visitor attractions, tourists can now choose destinations that align more closely with their individual interests and preferences. This accessibility to more detailed VA information has further fueled the demand for personalized travel experiences (Ana and Istudor, 2019). Responding to this growing trend, Citywalk has emerged as a popular form of urban tourism, defined as "a recreational activity including strolling across metropolitan regions to acquire certain experiences while engaging in behaviors that seek diversity" (Wang et al., 2025). Originally developed from the "London Walk" concept, which began as guided tours along predetermined routes aimed at showcasing a city's history, culture, and landscape, Citywalk has gradually evolved into a more flexible form of urban travel (Wu, 2024). This evolution towards immersive urban exploration finds a deeper historical and philosophical antecedent in the 19th-century concept of *flânerie*. The *flâneur*, or the urban stroller, was famously analyzed by Walter Benjamin as a quintessential figure of modernity—a "passionate spectator" who wandered aimlessly through the arcades and streets of Paris, observing the transient tapestry of city life (Benjamin, 2006). Unlike the structured nature of guided tours, *flânerie* emphasizes an unplanned, subjective, and aesthetic engagement with the urban environment, where the act of walking itself becomes a way of reading the city (Tester, 1994).

While the classic *flâneur* was a solitary, almost artistic figure, the modern Citywalk transforms this spirit into a more accessible recreational activity. Unlike traditional travel which often prioritizes visiting well-known landmarks and attractions, Citywalk allows travelers to immerse themselves in the streets and alleys, providing travelers with a deeper connection with the city's history, culture, landscape, and everyday life (Germano, 2023). In addition to Citywalk tourists preferring more personalized travel experiences, Freytag, in his study of repeat visitors to Paris, found that repeat visitors "often neglect or even avoid the iconic sights of mass tourism" and focus on "trying to take part in the everyday life of the local population" (Freytag, 2010a).

As a personalized form of travel, Citywalk allows tourists to explore the city based on personal interests, meaning that itinerary planning is heavily influenced by individual preferences. Unlike traditional itinerary planning, Citywalk tourists focus more on local culture, scenery, architecture, and urban life. VA recommendation not only requires considering popular landmarks but also the specific preferences and activities related to each VA. However, traditional personalized itinerary planning algorithms face challenges in meeting users' diverse needs in real-time. One common approach involves user interaction-based recommendation systems (Savir et al., 2013; Meehan et al., 2013; Lu et al., 2010; Yahi et al., 2015), which categorize and recommend attractions based on VA types—such as museums, landmarks, or natural sites often overlook content differences within the same VA type. As a result, these recommendations often fail to meet the personalized needs of tourists effectively.

Another widely used approach involves location-based social networks (LBSNs)-based systems, which rely on historical user data to suggest VAs based on patterns of similar user behavior (Majid et al., 2015; Chang et al., 2021; Ding and Chen, 2018). While these systems are popular for their ability to identify trends, they are hindered by the cold start problem, where recommendations for new users are limited due to insufficient historical data. Additionally, such systems often produce static recommendations, making it challenging to adapt to real-time changes in user preferences or situational needs.

While user interaction-based (Savir et al., 2013; Meehan et al., 2013; Lu et al., 2010; Yahi et al., 2015) and LBSN-based (Majid et al., 2015; Chang et al., 2021; Ding and Chen, 2018) recommendation systems have gained traction in travel applications, they are fundamentally limited in their ability to leverage the semantic richness embedded in UGC. These systems typically rely on structured input, high-level attraction categories, or behavioral patterns, and often ignore nuanced descriptions, contextual clues, and experiential dimensions conveyed in user narratives.

In response, a line of UGC-based recommendation approaches has emerged to extract attraction features directly from user reviews and digital content. However, earlier UGC-based methods primarily relied on shallow text-mining techniques such as Latent Dirichlet Allocation (LDA) and other statistical approaches, which extract only salient keywords while ignoring contextual and perceptual depth (Liang et al., 2024; Missaoui et al., 2019). As a result, these methods struggle to represent attraction features comprehensively and fail to align with the multifaceted and detailed preferences of travelers.

Recent advances in large language models (LLMs) have introduced new opportunities to overcome these limitations. By processing natural language input, LLMs can bridge the gap between loosely expressed user preferences and semantically rich attraction features derived from UGC. For instance, systems such as *ITINERA* (Tang et al., 2024) leverage LLMs to parse natural language queries into structured subrequirements and retrieve relevant VAs through semantic matching. While these systems represent a major step forward, they still face challenges in handling user inputs that are often ambiguous, incomplete, and highly faceted (Kostric et al., 2024; Keyvan and Huang, 2022; Huang et al., 2025). As a result, the alignment between partially expressed user intent and the complex, multi-dimensional features of VAs remains limited.

Additionally, Citywalk, as a walking-centered travel approach, offers tourists an immersive experience that extends beyond simply visiting VAs. The walking paths connecting these VAs play a crucial role in shaping tourists' overall perception of the city (Chen et al., 2017). However, traditional itinerary planning often prioritizes the shortest routes, neglecting the scenic and experiential quality of these paths (Rahmani et al., 2020; Benouaret and Lenne, 2016; Ding and Chen, 2018). For Citywalk tourists, the journey between attractions is as meaningful as the destinations themselves. While some research has started to address multi-sensory or experiential aspects of urban routes (e.g., recommending beautiful, quiet, or olfactorily pleasant paths (Quercia et al., 2015)), incorporating the scenic route planning connecting different VAs into recommendation systems remains an unresolved problem in many studies.

To address these challenges, this research proposes *UGuideRAG* (User-Generated Content-Guided Retrieval-Augmented Generation), a modular recommendation framework designed for personalized and fine-grained urban tourism. UGuideRAG consists of five components: (1) *UGC-based Attraction Database Construction (UADC)*, which aggregates and structures UGC to enrich VAs with descriptive, experiential, and contextual information that goes beyond official categorizations; (2) *Scenic Pathway Database Construction (SPDC)*, which integrates geotagged photos and point of interest (POI) data to quantify the scenic value of each pathway. (3) *Intent-Enhanced Retriever (IER)*, which decomposes user queries into structured intents across experiential dimensions using LLMs and retrieves semantically relevant content; (4) *LLM-based Reranker (LRR)*, which scores retrieved candidates based on their semantic relevance to the user query; and (5) *Cluster-aware Spatial Optimization (CSO)*, which constructs personalized and spatially coherent itineraries for urban travel.

My overall contributions are as follows:

1. Grounded in tourism research, this research defines a set of perception-aligned attraction features comprising *landscape and content*, *activities*, and *atmosphere*, and employ LLMs to extract these structured features from unstructured user-generated content, providing the data foundation for personalized recommendations.

2. This research proposes an intent-enhanced RAG architecture, in which the retrieval module is guided by LLM-based decomposition of user queries into structured intents across multiple experiential dimensions. Retrieved candidates are then re-ranked using an LLM based on their alignment with the user query, enhancing semantic precision while supporting more personalized and diverse itinerary generation.

3. To enhance the experiential quality of travel routes, the proposed system incorporates a scenic pathway database that leverages geo-tagged photos and POI data to estimate the scenic value of urban pathways, enabling route planning that prioritizes visually and experientially rich walking segments.

4. This research conducts extensive experiments across multiple cities, demonstrating that UGuideRAG generates personalized and spatially coherent itineraries that outperform existing baselines in urban travel recommendations.

2 Related Work

2.1 Visitor Attractions (VAs) and Core Experience Dimensions

According to Pearce's definition, an attraction is a "named site with a specific human or natural feature which is the focus of visitor and management attention" (Pearce, 1991). In previous research, VAs have been classified into seven main categories (Leask, 2010):

Table 2.1: Categories of Visitor Attractions

VA Categories	Subcategories
Theme Parks/Amusement Parks	Water parks, amusements, themed attractions
Museums & Galleries	Art, cultural, historical, collection-based, virtual, open-air museums
Natural	Gardens, national parks, forests
Animal	Safaris, farms, zoos, aquariums
Visitor Centres	Cultural, industrial, transport-focused
Religious Sites	Religious sites, historical religious buildings
Heritage	Castles, forts, historic houses, visitor centers, monuments, industrial, dark, archaeological, military, music

When visiting a destination, various destination attributes or features contribute to tourists' travel experiences. These attributes, often referred to as pull factors, draw people to a destination (Khoo-Lattimore and Ekiz, 2014; Klenosky, 2002). Within the context of VAs, key pull factors can be categorized into three main groups (Faerber et al., 2021):

- **Physical Environment:** This encompasses the infrastructure and quality of goods and tangibles provided by the VA, largely controllable by VA management (Kouthouris and Alexandris, 2005).

- **Service Quality:** Key dimensions include queuing and crowding (Houston et al., 1998), employees' service quality (Alexandris et al., 2006), general information provided by the VA (Booth, 1998), and the perceived cost-benefit ratio (Matzler et al., 2007; Tomas et al., 2002).

- **VA Core Experience:** This includes VA content and presentation (Faerber et al., 2021), entertainment, fun and emotions, atmosphere (Geissler and Rucks, 2011), novelty (Poulsson and Kale, 2004), and authenticity (Moscardo and Pearce, 1986), all of which define the fundamental visitor experience.

While all three categories of pull factors play a significant role in shaping visitor experiences, this study focuses primarily on the VA core experience, as it directly attracts visitors and influences their travel outcomes. To capture the core features of VAs, this research selects three dimensions of the VA core experience: *landscape and content*, *suitable activities*, and *atmosphere*. These dimensions are designed to encapsulate the essential elements of VA core features and serve as a foundation for evaluating and enhancing visitor attractions.

2.2 Tourism-Related User-Generated Content (UGC)

Tourism-related user-generated content, considered by travelers as a more trustworthy source of information, has significantly transformed how consumers search for and evaluate travel information (Akehurst, 2009). With the advancement of communication technologies, the impact and significance of UGC cannot be overlooked, as "digital platforms are revolutionizing the traditional processes of researching, purchasing, selling, experiencing, and sharing travel" (World Bank, 2018). Beyond its critical role in information search and travel planning, UGC also plays an irreplaceable role in shaping the image of tourist destinations and transforming marketing strategies (Cox et al., 2009).

The rapid growth of tourism-related UGC, particularly online travel reviews (OTRs), has dramatically influenced how tourists access travel information. For instance, TripAdvisor stored 10 million OTRs in 2007 (Gretzel and Yoo, 2008), and this figure has since surpassed 1 billion, with 26 million reviews submitted in 2020 alone, covering more than 8 million tourist resources worldwide.[2] According to a survey conducted by the European Union, 51% of Europeans rely on traditional word-of-mouth (WoM) and 34% on electronic word-of-mouth (eWoM) when making travel decisions (European Commission, Directorate-General for Enterprise and Industry

[2] https://www.tripadvisor.com/powerofreviews.pdf

and TNS Political & Social, 2014). Similarly, in the UK, 40% of international visitors identified WoM and 30% eWoM as key influences (VisitBritain, 2019). In the U.S., a survey of more than 2,000 leisure travelers revealed that eWoM (58.2%) had a greater impact on decision-making compared to WoM (45.6%) (Destination Analysts, 2019).

The rise of tourism websites, social media platforms, and online travel agencies (OTAs) has generated massive amounts of data, including feedback, destination reviews, and traveler experiences (Lim and Rasul, 2022). These data enable travelers to make more personalized decisions about their destinations, meeting the growing demand for customized travel experiences (Yang et al., 2024). However, the sheer volume of UGC poses a significant challenge for manual processing (Abbasi-Moud et al., 2021). Therefore, developing recommendation systems based on UGC can better uncover the features of each VA and accurately align them with travelers' preferences, thereby enhancing the personalization and precision of tourism recommendations.

2.3 VA Features Extraction Based on UGC

UGC plays a key role in describing tourist attractions, providing rich perceptual information for tourists (Munar and Jacobsen, 2013). By mining VA features from UGC, more interest-aligned recommendations can be provided from the perspective of tourists. This approach not only reveals the uniqueness and atmosphere of the attractions but also highlights deficiencies and possible activities, offering tourists a more comprehensive and authentic experience.

Currently, feature extraction of attractions is mostly based on word frequency statistics methods. Term Frequency-Inverse Document Frequency (TF-IDF) is a common statistical method used to measure the importance of terms in a document. The weight of a term increases with its frequency in the document, but decreases as its frequency across the entire corpus increases. Burtch et al. quantified the novelty of reviews by calculating the cosine distance of consumer reviews using the TF-IDF model (Burtch et al., 2022). Mishra et al. used TF-IDF to extract keywords from hotel reviews and used cosine similarity to recommend similar hotels (Mishra and Gupta, 2019). Peng and Huang studied tourist hotspots and attraction features in Beijing by analyzing geotagged photos and tourist-generated tags (Peng and Huang, 2017). In addition to using the TF-IDF method, Abbasi-Moud et al. directly extracted the top five most frequently repeated words from the visitor attraction (Abbasi-Moud et al., 2021).

However, word frequency statistics-based methods have limitations. They cannot understand the deeper semantics and context of words, leading to inaccurate descriptions of POI features and potential misunderstandings. Word frequency methods cannot distinguish between synonyms or handle polysemy and are not effective in highlighting important but infrequent keywords. Additionally, subtle differences between similar VAs are difficult to distinguish using word frequency methods, limiting the effectiveness of personalized recommendations.

2.4 Personalized Recommendation Systems

Current personalized travel itinerary recommendation systems mainly include three research directions:

2.4.1 POI Recommendations Based on User Interactions

Lu et al. developed the Photo2Trip system, which identifies popular tourist areas by collecting geographic photos and recommends POIs based on visit time, travel purpose, and style (Lu et al., 2010). Gavalas et al. designed the Scenic Athens system, which allows users to set preferences for different attractions and recommends POIs combined with walking routes (Gavalas et al., 2017). Yahi et al. developed the Aurigo system, which scores POIs based on popularity, distance, and user preferences, allowing users to iteratively build itineraries by selecting POIs (Yahi et al., 2015). Pantano et al. developed a tourism recommender system using 18 user profile themes and a Support Vector Machine (SVM) model to predict destination ratings, integrating contextual factors like time and weather, and demonstrated improved accuracy and personalization in supporting tourist decision-making (Pantano et al., 2019).

However, user interaction-based recommendation systems tend to broadly categorize attractions and users into generalized groups, which often results in a lack of personalization. This limitation underscores the need for more refined recommendation approaches that can better capture individual user preferences and offer truly customized travel experiences.

2.4.2 POI Recommendations Based on Location-Based Social Networks (LBSNs)

LBSNs allow users to share check-in data at real-world locations, enabling collaborative filtering algorithms to recommend POIs that users might be interested in. Majid et al. proposed a system that combines temporal and spatial context factors to recommend tourist attractions and routes by mining geotagged photos from social media (Majid et al., 2015). Chang et al. designed the MANC multi-attention network model, which improves POI recommendation accuracy by combining users' social relationships and POI features (Chang et al., 2021). Ding & Chen's RecNet system integrates co-visitation, geographic, and category influences in LBSNs to learn user behavior patterns for POI recommendations (Ding and Chen, 2018).

Compared to user interaction-based recommendation systems, LBSN-based systems offer a better prediction of POIs that users may be interested in by leveraging contextual and spatial data. However, they are limited to providing static recommendations based on users' historical data and lack the flexibility to address users' dynamic and evolving preferences in real-time. This highlights the need for more adaptive recommendation systems that can account for changing user demands and contexts.

2.4.3 POI Recommendations Based on Language Models (LMs)

Abbasi-Moud et al. developed a system that extracts tourists' preferences from reviews on tourism social networks and identifies each attraction's features based on user-generated comments (Abbasi-Moud et al., 2021). The system semantically compares users' preferences with the features of attractions to suggest the most matching POIs to the users. Chen et al. developed the TravelAgent system, which recommends personalized itineraries by analyzing user preferences and historical data (Chen et al., 2024). Tang et al. proposed the ITINERA system, which combines LLMs with spatial optimization techniques to generate personalized urban itineraries by parsing user needs through natural language (Tang et al., 2024).

However, the main limitation of existing systems lies in their reliance on simple keyword matching to align user queries with attractions, which may lead to unstable and imprecise matching results. By decomposing attractions and user queries into reasonable feature components, the stability and rationality of the system can be improved, enhancing the overall quality of recommendations.

2.5 RAG in Recommendation Systems

In recent years, LLMs have shown advancement in understanding and processing natural language. However, challenges such as hallucinations (Yao et al., 2023) and inefficiencies in fine-tuning (Du et al., 2023) continue to affect their reliability in real-world applications. One promising solution is Retrieval-Augmented Generation (RAG), which combines external information retrieval with generative modeling to enrich input representations and improve the quality of generated content (Lewis et al., 2020).

RAG has been widely adopted for its strong ability to interpret user needs expressed in natural language (Zhang et al., 2025), and it has demonstrated great potential in modeling user preferences and delivering personalized recommendations (Di Palma, 2023; Fan, 2024; Lu et al., 2021; Yu et al., 2025). For example, Di Palma proposed a simple RAG-based recommendation model that leverages structured knowledge from movie and book datasets to enhance recommendation relevance (Di Palma, 2023). Yu et al. introduced Spatial-RAG, an extension of the RAG framework that integrates both semantic and spatial retrieval to support spatial reasoning tasks, enabling LLMs to generate geographically grounded and contextually relevant responses based on user preferences and real-world spatial constraints (Yu et al., 2025).

The RAG framework typically adopts a dual-module architecture consisting of a retrieval module and a reader module, which jointly improve the relevance and informativeness of generated outputs. However, the effectiveness of the retrieval component is often hindered by ambiguous or underspecified user queries, leading to suboptimal retrieval results and degraded overall performance. Recent research has shown that rewriting and expanding user intent representations within input prompts can significantly enhance RAG's performance by improving retrieval quality and alignment with user needs (Shi et al., 2024; Ma et al., 2023).

This issue is particularly pronounced in tourism recommendation scenarios, where user demands extend beyond simple keywords to include nuanced expectations for experiences, emotional responses, and environmental contexts (Terkenli, 2021). While previous efforts, such as Tang's method of extracting positive and negative query components and computing embedding similarities for POI recommendation, have shown initial success, they often fall short in capturing the full breadth of user expectations (Tang et al., 2024). This results in imprecise POI retrieval and limited recommendation diversity. These challenges highlight the importance of developing methods that better capture and represent implicit user intent to support personalized recommendations in complex, experience-driven domains such as tourism.

2.6 Scenic Route Planning in Urban Tourism

For urban tourism, walkable places are fundamental, as most tourist activities in destinations occur while walking. If a street's walkability is high, tourists can enjoy positive walking experiences (Lo and Lee, 2011). Walkable tourism areas provide comfortable and meaningful experiences by allowing direct interaction with the surroundings. Even in cases where pedestrian pathways lack connectivity or quality, the presence of engaging street activities can still make a place appealing (Ujang and Zakariya, 2015). Freytag's survey of visitors to Heidelberg, Germany, revealed that walking and strolling are key activities, particularly for those spending most of their time in the old town (Freytag, 2010b). Similarly, Shoval et al. analyzed GPS data from tourists in Hong Kong and found that tourists staying in city-center hotels are more inclined to walk to attractions (Shoval et al., 2011). This aligns with the arguments of Bieri and Anton Clavé, who suggest that walkability is not only central to tourist activities but has also become an ideal spatial form in capitalist urban planning. Walkable environments foster tourism consumption and social interaction, creating new economic opportunities (Bieri, 2017; Anton Clavé, 2018).

Previous research on street walkability has primarily focused on residents' daily commutes, measuring walkability through dimensions such as street connectivity, residential dwelling density, land-use mix, safety, convenience, and comfort (Leão and Urbano, 2020; Hajna et al., 2015; Villaveces et al., 2012). However, the factors influencing tourists' walkability experiences may differ. Ujang & Muslim's study of tourism areas in Kuala Lumpur, Malaysia, found that the image of a place influences visitors' walking experiences more than the actual quality of pathways or comfort. Enhancing the attractiveness of buildings and spaces for visual enjoyment and providing cultural, commercial, and recreational activities can effectively improve tourists' experiences (Ujang and Zakariya, 2015). Therefore, the scenic value of streets and their interactive engagement with tourists are critical factors affecting the walking experience.

The scenic quality of routes significantly influences tourists' travel experiences. Research by Eby and Molnar demonstrated that scenic routes play a vital role in route selection (Eby and Molnar, 2002)]. Gavalas et al. designed the Scenic Athens tour planner, which integrates user preferences for scenery, nature, waterfronts, market districts, and architecture when navigating between POIs (Gavalas et al., 2017). Zheng et al. developed the GPSView system, which utilizes geo-tagged photos from Flickr to calculate street visibility values and determine the scenic appeal of road segments (Zheng et al., 2013). Runge applied Google Street View data and convolutional neural networks (CNNs) to classify scenic types along roads and incorporate these into scenic route planning (Runge et al., 2016). Scenic quality is particularly essential for Citywalk route planning, as walking not only allows tourists to enjoy urban scenery but also to explore the city's history and culture in depth. These elements collectively form core travel experiences. Improving the scenic quality of routes is thus crucial to enhancing the overall Citywalk experience.

In addition to scenic quality, the interactive engagement between streets and tourists is equally important. Milias et al. surveyed 403 participants in Frankfurt, Germany, and found that the attractiveness of streets is significantly influenced by houses, architecture, and shops (Milias et al., 2023). Chang et al. analyzed tourist mobility data and examined how various POI types affect the walkability of streets in Daejeon, South Korea. The study revealed that beyond cultural heritage sites and parks, POIs such as local markets, bakeries, cafes, restaurants, bookstores, flower shops, and spas showed the highest connectivity strength in the network, indicating their popularity and high visitor numbers (Chang et al., 2023). Anton Clavé introduced two attractiveness indicators in the Washington metropolitan area WalkUP analysis: the Sightseeing Density Index (measuring the concentration of museums, memorials, gardens, and historical sites) and the Entertainment Density Index (measuring the density of amusement attractions, sports arenas, performing arts venues, and top restaurants). These indices were used to observe the differences in tourism characteristics and dynamics across WalkUPs (Anton Clavé, 2016).

While tourism recommendation systems have traditionally focused on selecting attractions, they often overlook the planning of routes between these sites. Since walking constitutes a vital element of tourists' travel experiences, integrating street attractiveness into route planning and balancing factors such as the shortest distance and high attractiveness can significantly enhance the overall travel experience.

3 Problem Formulation

We define the personalized urban itinerary recommendation task as a two-stage problem that integrates semantic relevance, spatial coherence, and experiential enrichment.

Let $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ denote the set of all available VAs in a given city. Each attraction $v_i \in \mathcal{V}$ is associated with experiential features primarily derived from UGC, such as reviews and photos.

Given a natural language user query q, the first objective is to generate a personalized and spatially coherent one-day itinerary:

$$\mathcal{V}_{\text{order}} = [v_{o_1}, v_{o_2}, \ldots, v_{o_M}]$$

where $v_{o_i} \in \mathcal{V}$, and $M \geq n_{\min}$ ensures the itinerary is sufficiently informative for a full-day urban experience. This sequence is optimized to align with the semantics of q while maintaining reasonable travel efficiency and clustering.

The second objective is to construct a scenic-aware walking route $\mathcal{P}$ that connects the selected attractions in $\mathcal{V}_{\text{order}}$ through pedestrian-friendly paths that maximize scenic value:

$$\mathcal{R} = \{\ell_1, \ell_2, \ldots, \ell_K\}, \quad \ell_k \in \mathcal{L}$$

where $\mathcal{L}$ denotes the set of walkable road segments in the street network. Each segment ℓ_k is assigned a scenic score $SS(\ell_k)$ based on the density and orientation of geotagged photos and proximity to experiential POIs.

The final goal is to generate a walking itinerary that:

- selects and orders attractions that are semantically aligned with q,
- ensures spatial walkability and clustering, and
- connects these attractions via routes that maximize the overall scenic experience without excessive distance overhead.

This composite task introduces several research questions:

RQ1: How can we extract structured, perception-aligned features from noisy, unstructured UGC for each $v_i \in \mathcal{V}$?

RQ2: Given a free-form user query q, how can we retrieve and rank a candidate subset $\mathcal{V}_{\text{top-}k} \subset \mathcal{V}$ that best matches the user's intent?

RQ3: How can we select and order a final subset $\mathcal{V}_{\text{order}}$ that is both semantically relevant and spatially coherent?

RQ4: How can we plan a scenic-aware walking route $\mathcal{R}$ through the selected attractions that enriches the overall experiential quality?

4 Data and Methods

4.1 Research Area

To evaluate the proposed framework in realistic settings, this study focuses on Paris and Rome–Vatican. As two of the world's most visited cities—the Paris Île-de-France region welcomed 44 million tourists in 2022 (Comité Régional du Tourisme Paris Île-de-France, 2023), while Rome recorded over 35 million tourist presences in 2023 (Comune di Roma, 2024)—they generate a massive volume of the user-generated content essential for the proposed data-driven approach. Furthermore, these cities are representative of high-density, heritage-rich tourism environments that pose both experiential and spatial challenges for personalized trip planning.

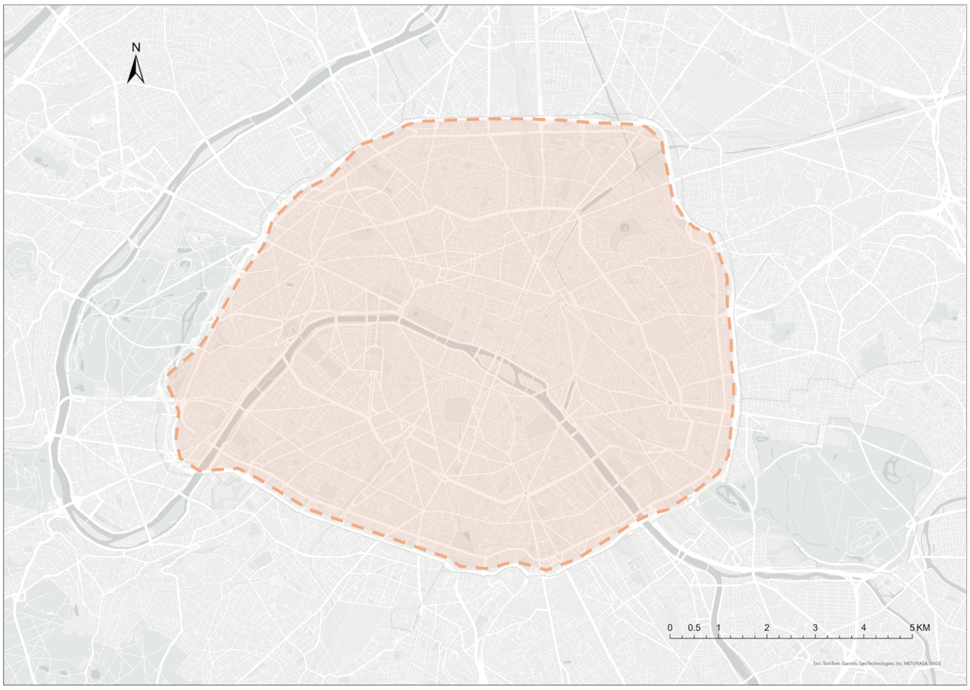


Figure 4.1: Research Area in Paris

Paris, the capital of France, welcomed approximately 37 million visitors in 2023, ranking first globally in international tourist arrivals[3]. Known as the "City of Light,"

[3] https://parisjetaime.com/eng/convention/article/tourism-in-paris-key-figures-a1749

Paris offers an extraordinary mix of iconic landmarks (such as the Eiffel Tower, Louvre Museum, and Notre-Dame Cathedral), scenic urban landscapes along the Seine River, and a vibrant cultural life. Its well-preserved historic core, abundance of museums, and walkable neighborhoods make it an ideal testbed for modeling diverse visitor experiences and optimizing itineraries within a dense urban fabric.

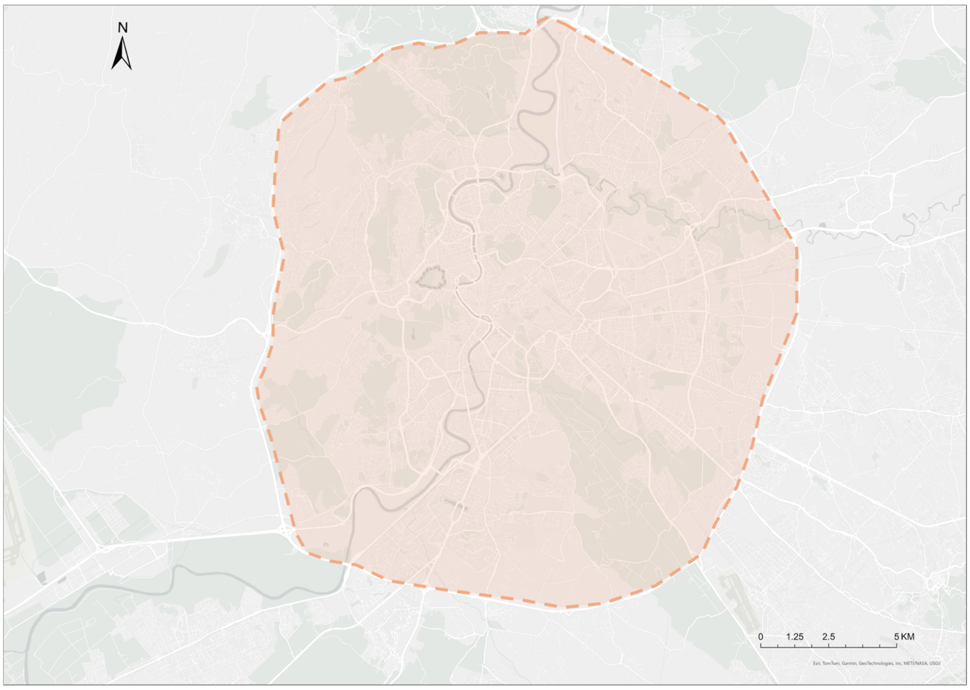


Figure 4.2: Research Area in Rome-Vatican

Rome–Vatican, the capital of Italy and the seat of the Catholic Church, attracted around 35 million visitors in 2023,[4] including nearly 6.8 million visitors to the Vatican Museums,[5] and both the Historic Centre of Rome (with Properties of the Holy See) and Vatican City are inscribed on the UNESCO World Heritage List.[6] The compact layout of historical attractions, coupled with high visitor volume and layered cultural significance, makes the Rome–Vatican area a highly representative case for evaluating semantic retrieval and walkable route optimization.

Together, these two cities provide complementary contexts for testing the proposed system: both are globally recognized cultural capitals with complex attraction networks, rich user-generated content, and significant practical demand for intelligent, intent-aware tourism planning tools.

[4] ANSA, "Turismo: a Roma record di presenze, 35 milioni di pernottamenti," Nov. 30, 2023.
[5] The Art Newspaper, "The 100 most popular art museums in the world—2023," Mar. 26, 2024.
[6] UNESCO WHC, "Historic Centre of Rome . . . " (List 91); "Vatican City" (List 286).

4.2 Data Source

Our framework integrates multiple real-world urban data sources to support attraction representation, spatial reasoning, and experiential evaluation. These datasets span user-generated content, spatial infrastructure, and open geoinformation layers, covering two cities: Paris and the Rome–Vatican region.

(1) Attraction Data and User Reviews Visitor attractions were aggregated from three widely used platforms: Google Maps[7], TripAdvisor[8], and OpenStreetMap (OSM)[9]. This multi-source approach ensured broad coverage of both popular and lesser-known sites.

User-generated reviews for each VA were collected exclusively from Google Maps using Selenium[10], with reviews sorted by relevance to prioritize informative and detailed content. These reviews contain user ratings, narratives, and emotional expressions, offering valuable insights into tourist perceptions, satisfaction levels, and site-specific experiential features. The extracted textual content was further processed to construct a semantically rich attraction database for personalized recommendation.

(2) Points of Interest (POIs) To model the walkability and experiential richness of urban routes, this study collected pedestrian-relevant POIs from OSM. Each POI includes location, type to enable interactive scenic score computation.

(3) Geotagged Photos This study utilized geotagged photos from Flickr [11], one of the largest crowd-sourced photo-sharing platforms, to capture public visual perception of urban environments. Flickr provides rich metadata for each photo, including precise GPS coordinates, timestamps, and user ID, making it a valuable resource for studying spatial and temporal patterns of tourist activity. The platform is particularly well-suited for tourism-related analyses, as it attracts users who often document visits to scenic or culturally significant places. In this work, geotagged photos serve as a proxy for perceived scenic interest, enabling us to estimate visual appeal at a fine-grained urban scale.

[7] https://www.google.com/maps
[8] https://www.tripadvisor.com/
[9] https://www.openstreetmap.org/
[10] https://www.selenium.dev/
[11] https://www.flickr.com/

(4) Street Network Data Pedestrian-accessible street network data was extracted from OSM using the OSMnx library[12]. This includes street geometry, topology, and walkable connectivity. The network structure was used for routing with integrating spatial indicators such as geotagged photos and POI counts.

In line with the framework illustrated in Figure 4.3, these datasets serve two primary functions. The attraction data and user reviews are the core inputs for the UGC-based Attraction Database Construction (UADC) module. This process creates a semantically rich knowledge base of VAs, enabling the system to match attractions with user query. The remaining datasets—POIs, geotagged photos, and the street network—are integrated within the Scenic Pathway Database Construction (SPDC) module. This component builds a comprehensive routing graph where each street segment is enriched with a scenic score, facilitating the generation of spatially coherent and experientially pleasant itineraries.

4.3 Methods

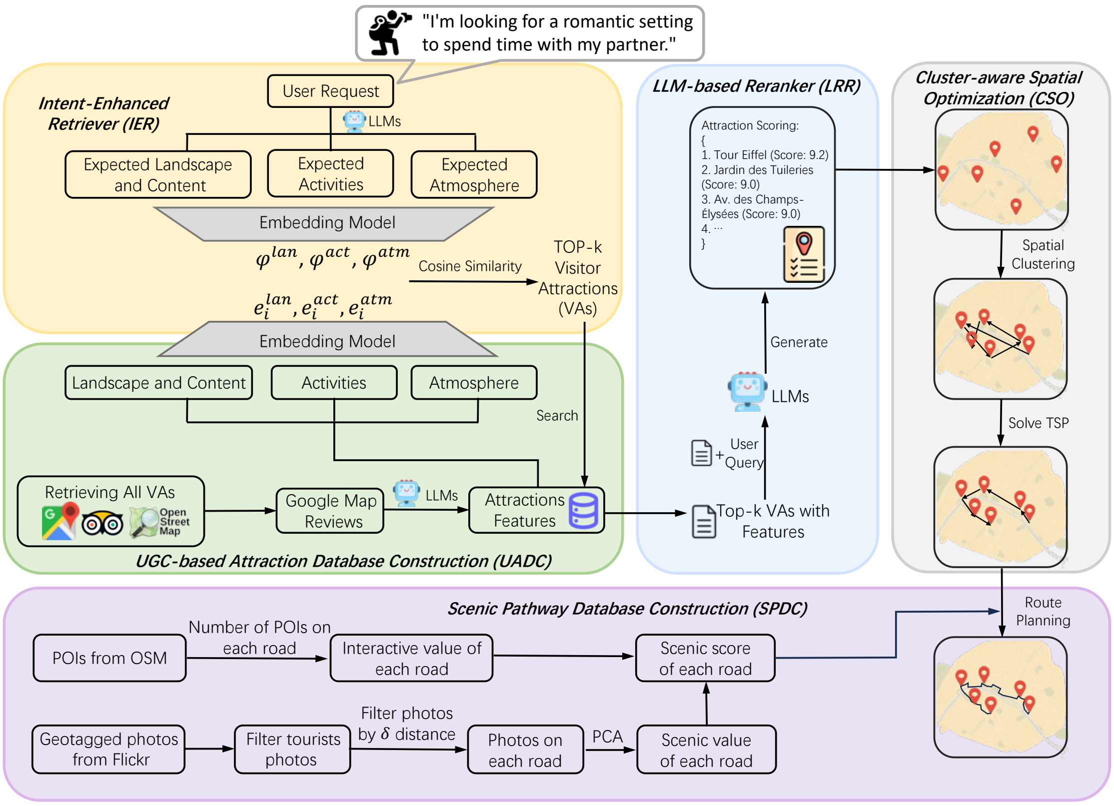


Figure 4.3: Illustration of the proposed *UGuideRAG* framework

[12] https://osmnx.readthedocs.io/en/stable/

4.3.1 UGC-based Attraction Database Construction

To support personalized recommendations, the first step of proposed framework is to construct a structured database that captures the nuanced characteristics of each visitor attraction. This database serves as the foundational knowledge base for all downstream modules, including semantic retrieval, re-ranking, and itinerary construction. Following the experiential framework introduced in Section 2.1, this study extracts these characteristics from user-generated Google Maps reviews, focusing on three key facets: landscape and content, activities, and atmosphere.

Table 4.1: Selected Google Map Reviews for Pont Neuf with Ratings

Rating	Review
5.0	*"Pont Neuf is a beautiful destination to visit in the evening, offering stunning views of the city and the Seine River. As the sun begins to set, the lights of the city come to life, casting a romantic and picturesque ambiance on the bridge. At night, Pont Neuf is illuminated, providing a beautiful backdrop for a romantic stroll or a relaxing evening walk."*
5.0	*"Walking around Paris is one of the best activities one can do when there. This is an amazing sunset spot by the Seine river. Very close to both Notre Dame and Louvre museum. Highly recommend walking around the area and soaking in Paris. Also a great picnic spot near the river."*
4.0	*"Built in 1607 and still look great and solid and probably the most picturesque of all the Parisian bridges. It is made of two spans due to small island in between. This is also where you can go for a boat cruise near the very top of the island. Nice to get views on both sides of the Seine."*

Table 4.2: LLM-Extracted Experiential Features for Pont Neuf

Dimension	Extracted Feature Description
Landscape & Content	Oldest stone bridge in Paris with iconic Seine River views, nearby parks, and historic features like the Henri IV statue. Features scenic vistas of landmarks like Notre Dame and Eiffel Tower.
Activities	Walking, river cruises, photography, sunset viewing, sightseeing landmarks, and boarding Vedettes tour boats.
Atmosphere	Historic yet vibrant, blending romantic charm with lively crowds. Offers peaceful spots for relaxation amid bustling artistic and cultural energy.

For each VA v_i, this study prompts LLMs [13] to analyze its collected reviews R_i, ex-

[13] This study utilized the LLM via the Volcano Engine API, specifically using the `deepseek-r1-250120` model endpoint. This version, indicating a release from January 20, 2025, was the one available during the experiments. The model is multilingual, and all prompts were conducted in English. Access was subject to the platform's standard API usage costs.

tracting structured textual features across the three experiential dimensions: landscape and content f_i^{lan}, activities f_i^{act}, and atmosphere f_i^{atm}, along with a general summary d_i describing the overall character of the site.

To illustrate this process, this study presents an example for the *Pont Neuf* in Paris. Table 4.1 shows selected user reviews, which reflect visitor attention to scenic views, nearby landmarks, and atmospheric qualities. Based on these reviews, LLMs extract structured experiential features summarized in Table 4.2, revealing landscape attributes (e.g., river vistas and historical architecture), common activities (e.g., walking, sunset watching, river cruises), and perceptual atmosphere (e.g., romantic).

All textual features are encoded using an embedding model $\psi(\cdot)$, producing the following embeddings:

$$\begin{aligned} e_i^{\text{lan}} &= \psi(f_i^{\text{lan}}), \quad & e_i^{\text{act}} &= \psi(f_i^{\text{act}}), \\ e_i^{\text{atm}} &= \psi(f_i^{\text{atm}}), \quad & e_i^{\text{des}} &= \psi(d_i) \end{aligned} \tag{4.1}$$

These dimension-specific representations are stored as part of the attraction embedding database:

$$\mathcal{V} = \left\{ \left(e_i^{\text{lan}}, e_i^{\text{act}}, e_i^{\text{atm}}, e_i^{\text{des}} \right) \right\}_{i=1}^{N} \tag{4.2}$$

Prompt for VA Feature Extraction

```
You are an AI travel planning assistant specializing in {city}.

### Task Overview
Your task is to extract the key characteristics of a Paris attraction based **only on visitor reviews**.

For the attraction named "{va_name}", analyze the following reviews: {va_reviews}.

Break down the visitor experience into the following four dimensions:

1. **Landscape and Content**: Describe the physical, cultural, architectural, historical, or informational features of
the attraction as mentioned by visitors.
2. **Suitable Activities**: Identify the main actions or experiences visitors engage in or recommend.
3. **Atmosphere**: Summarize the mood, vibe, or emotional experience reflected in the reviews.
4. **Overall Description**: Provide a concise summary of the overall visitor impression in one paragraph.

Each dimension must be concise (≤50 words) and strictly based on the provided reviews.
The **overall description** must be no more than **100 words**, also based only on the reviews.
---

### Output Format:
Return your analysis in the following JSON format only:

{{
  "landscape and content": "...",
  "suitable activities": "...",
  "atmosphere": "...",
  "overall description": "..."
}}
```

Figure 4.4: Prompt for VA Feature Extraction

4.3.2 Scenic Pathway Database Construction

To generate itineraries that are not only efficient but also experientially rich, the proposed framework requires a detailed understanding of the urban environment between VAs. This section details the construction of a scenic pathway database, which is designed to evaluate and score individual road segments based on their visual appeal and interactive potential. This database enables the final routing module to connect visitor attractions and other points of interest using pathways that are both walkable and aesthetically pleasing. The construction process involves two main stages: cleaning the raw road network and calculating a scenic score for each segment.

4.3.2.1 Road Network Cleaning

Road networks extracted from OpenStreetMap often contain disconnected components due to incomplete mapping, topological noise, or data fragmentation. As a result, the raw graph may consist of multiple disconnected subgraphs that do not represent a single cohesive transport network.

However, for most downstream tasks such as routing, accessibility analysis, or urban planning, this study requires a fully connected road network—i.e., a graph in which any node is reachable from any other node.

To identify such connectivity, this study applys a standard depth-first search (DFS) traversal on the road graph to detect all connected components. The DFS procedure is defined in Algorithm 1, and is used as a subroutine to recursively explore all nodes belonging to the same component. After identifying all connected components, only the largest one is retained. This ensures that the resulting road network is a fully connected subgraph suitable for spatial analysis.

Algorithm 1 Depth-First Search (DFS) for Graph Traversal

Require: Graph $G = (V, E)$, starting node v, empty set *component*
Ensure: Set *component* containing all nodes reachable from v

```
mark v as visited
add v to component
for all neighbors u of v in G do
    if u not visited then
        DFS(G, u, component) {Recursively visit unvisited neighbors}
    end if
end for=0
```

4.3.2.2 Scenic Score Calculation

The number of photos distributed along roads can indicate scenic quality (Zheng et al., 2013). However, high photo density does not necessarily equate to beautiful scenery. To accurately evaluate road scenic value, this study utilizes tourist-uploaded geotagged data and applies Principal Component Analysis (PCA) to compute a scenic score.

(1) Selecting Tourist Photos Tourist-uploaded photos serve as a basis for evaluating scenic quality, but it's essential to distinguish them from photos taken by local residents. Tourists typically focus on capturing scenic views, whereas locals more often upload daily-life content.

To identify tourist photos, the temporal distribution of photos uploaded by a single user is analyzed. If the timespan of a user's photo uploads is less than one month, the photos are classified as tourist photos; otherwise, they are considered local photos.

(2) Scenic Value Calculation While the number of photos along a road (N_i^{pho}) is a key indicator, relying solely on this metric is insufficient due to biases like the popularity of landmark buildings. Hence, the geographic distribution direction of the photos is also considered.

For each road segment, photos within a distance δ are associated with that segment. Let the coordinates matrix of geotagged photos for point of interest C_i be M_i, with mean $\bar{i} = E(M_i)$. The covariance matrix is defined as:

$$\Sigma = E\left\{(M_i - \bar{i})(M_i - \bar{i})^T\right\}$$

Let λ_1, λ_2 be the eigenvalues and d_1, d_2 the corresponding eigenvectors of Σ. The eigenvector d corresponding to the largest eigenvalue λ_1 is chosen as the principal component. The angle α between d and the road segment direction vector d_r is calculated. The scenic value S_i for the road segment is computed as:

$$SV_i = N_i^{\text{pho}} \cdot (\cos(\alpha) \cdot \lambda_1 + \sin(\alpha) \cdot \lambda_2)$$

(3) Interactive Value Calculation Using the OSMnx library, all POI data in Paris is retrieved and filtered for types affecting tourists' walking experience (as defined

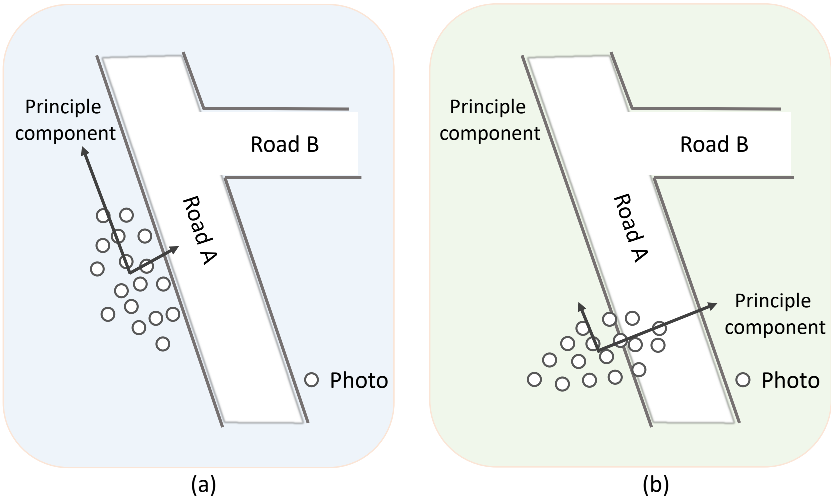


Figure 4.5: Examples of POIs on the roadside. (a) shows a POI visible from a nearby road, with photo distribution aligned along the road. (b) shows a POI with poor visibility, and its photo distribution does not align well with the road. (Source: adapted from Zheng et al. (2013))

in Section 2.6): local markets, bakeries, cafes, restaurants, bookstores, flower shops, and gift shops.

If a POI is within 30 meters of a road segment, it is associated with that segment. The number of POIs on each segment is denoted as N_i^{poi}. The final scenic score SS_i for each road segment is calculated by a weighted sum of the scenic value and POI count:

$$SS_i = w_1 SV_i + w_2 N_i^{\text{poi}}$$

where w_1 and w_2 are the weights assigned to the scenic value and interactive value, respectively.

4.3.3 Intent-Enhanced Retriever

To effectively retrieve personalized recommendations, unstructured natural language queries must first be translated into a structured, machine-processible format. This transformation is crucial, as tourist queries are often ambiguous, incomplete, or composed of multiple semantic facets. To address this fundamental challenge, my proposed retrieval module performs intent decomposition and structured semantic alignment to systematically deconstruct and understand the user's underlying needs. Leveraging the reasoning capabilities of LLMs, each user query q is parsed into three intent components corresponding to core dimensions of attraction experience:

expected *landscape and content* (r^{lan}), *activities* (r^{act}), and *atmosphere* (r^{atm}).

Prompt for Intent-Enhanced Retriever

```
Hello, you are now a travel analysis expert specializing in {city}. Your task is to decompose the user's travel query into
multiple independent experiential requirements based on the following three dimensions:

1. **Landscape and Content**: Includes tangible and intangible visual, physical, natural, man-made, and
informational elements that define the attraction's environment. This can include natural scenery (e.g., mountains,
rivers, beaches), architectural features, cultural or historical elements, artworks, and designed spaces.
2. **Activities**: Refers to specific actions or engagements the user intends to undertake, such as walking,
sightseeing, dining, learning, photography, or attending events.
3. **Atmosphere**: Refers to the mood, tone, or emotional/sensory experience the user is seeking, such as
romantic, peaceful, lively, historic, or adventurous.
---
### Output Format:

You should return a list where each item is a dictionary representing an **independent requirement**, with the
following key-value pairs:
- **expected landscape and content**: Describe what kind of natural or built environments, scenery, or
informational features the user wants to experience. If relevant, include people or cultural references.
- **expected activities**: Describe what specific actions or experiences the user wants to engage in. Include any
associated people or contexts if mentioned.
- **expected atmosphere**: Describe the mood, tone, or emotional quality the user is looking for.
**Do not include any explanations or code. Only return the list.**
The format should be exactly like this:
[
  {{
    "expected landscape and content": "...",
    "expected activities": "...",
    "expected atmosphere": "..."
  }},
  ...
]
---
### User Input
{user_input}
---
### Task Overview
Your goal is to analyze and break down the **user input** into multiple independent experiential requirements along
the three dimensions defined above. Be precise, grounded, and consistent with the definitions.
Now return your output in the required format.
```

Figure 4.6: Prompt for Intent-Enhanced Retriever

Each intent component $r^d \in \{r^{\text{lan}}, r^{\text{act}}, r^{\text{atm}}\}$ is then projected into the embedding space using the same embedding model $\psi(\cdot)$ employed for VA feature representation, yielding separate query embeddings for each experiential dimension $\varphi^d = \psi(r^d)$. Correspondingly, each VA $v_i \in \mathcal{V}$ is represented by a tuple of embeddings $\{e_i^{\text{lan}}, e_i^{\text{act}}, e_i^{\text{atm}}\}$, which encode its semantic profile across the three experiential dimensions.

This study computes the cosine similarity for each dimension d as:

$$\cos(\varphi^d, e_i^d) = \frac{\varphi^d \cdot e_i^d}{\|\varphi^d\| \cdot \|e_i^d\|} \tag{4.3}$$

Using this, the composite relevance score for each candidate VA is defined as:

$$\begin{aligned}\text{Score}_i = & \; w_{\text{lan}} \cdot \cos(\varphi^{\text{lan}}, e_i^{\text{lan}}) \\ & + w_{\text{act}} \cdot \cos(\varphi^{\text{act}}, e_i^{\text{act}}) \\ & + w_{\text{atm}} \cdot \cos(\varphi^{\text{atm}}, e_i^{\text{atm}}) \end{aligned} \tag{4.4}$$

Here, $w_{\text{lan}}, w_{\text{act}}, w_{\text{atm}} \in [0, 1]$ control the relative contribution of each experiential dimension.

By performing this dimension-aware matching, the retriever is able to more robustly align user intent with semantically rich and structurally organized attraction profiles—thereby improving recall for nuanced or under-specified queries. The final ranked list is obtained by computing Score_i for all $i \in \{1, \ldots, N\}$, and selecting the top-k candidates:

$$\mathcal{V}_{\text{top-}k} = \text{Top-}k \left(\{\text{Score}_i\}_{i=1}^{N} \right) \tag{4.5}$$

The resulting set $\mathcal{V}_{\text{top-}k} \subset \mathcal{V}$ serves as the input to the subsequent re-ranking stage, where contextual reasoning is applied via LLMs to refine semantic alignment and preference fit.

4.3.4 LLM-based Re-ranking of Retrieved VAs

While embedding-based retrieval provides a coarse-grained semantic alignment between structured user intent and candidate attractions, it lacks the capacity to perform fine-grained contextual reasoning. To address this, this study introduces a second-stage re-ranking module that leverages the inference capabilities of LLMs to evaluate each retrieved candidate in the full context of the original query.

Given the Top-k retrieved VAs $\mathcal{V}_{\text{top-}k}$, this study constructs a natural language prompt for each candidate $v_i \in \mathcal{V}_{\text{top-}k}$ that integrates: (1) the user's original query q; and (2) the structured attribute descriptions of v_i, including its landscape and content features f_i^{lan}, activities f_i^{act}, and atmosphere f_i^{atm}. These prompts are passed to the LLM, which performs context-aware semantic matching between the user query and each candidate's experiential attributes.

Formally, the LLM outputs a contextual alignment score $s_i^{\text{LLM}} \in [0, 10]$, representing the degree to which the candidate satisfies the user's latent preferences as expressed in natural language. This re-scoring process enables reasoning over implicit user goals, complex lexical variations, and nuanced feature combinations that are often poorly represented in fixed vector spaces.

The re-ranked list $\mathcal{V}_{\text{rerank}}$ is obtained by sorting the candidates in descending order of s_i^{LLM}. This re-ranking stage enhances the semantic fidelity and personalization of the final recommendation results, bridging the gap between discrete feature embeddings and holistic user intent understanding.

Prompt for LLM-based Reranker

```
You are an AI travel planning assistant specializing in {city}.
Your task is to assign a **suitability score** (from 0 to 10) to each of the candidate attractions based on the user's query.
### Scoring Guidelines
For each attraction, evaluate and assign a **total score** between 0 and 10, considering:
1. **Content Relevance (0–10)**: How well the attraction matches the user's desired themes, activities, and atmosphere.
2. **Negative Filtering**: Strongly penalize attractions containing user-prohibited or mismatched elements.
3. **Do NOT consider coordinates or spatial information.**
---
### Input Data
【User Query】
{user_input}
【Candidate Attractions】
Each attraction is represented as a dictionary with the following fields:
{
 "id": "1",
 "name": "Attraction Name",
 "landscape and content": "Description of physical landscape and cultural/historical content",
 "activities": "Available or typical activities for visitors",
 "atmosphere": "General vibe, ambiance, or emotional tone of the place"
}
All candidate attractions are included in the following list:
{attractions_list}
---
### Output Format
Return a **JSON object** where each key is an attraction ID and the value is a float score between 0 and 10 (inclusive). For example:
{{
 "1": 8.5,
 "2": 3.0,
 "3": 0.0
}}
---
### Output Requirements
- Only return the JSON object.
- Do NOT explain your reasoning.
- Do NOT include rankings, coordinates, or any formatting other than valid JSON.
- Scores must be float numbers between 0 and 10 (inclusive).

Begin scoring now.
```

Figure 4.7: Prompt for LLM-based Reranker

4.3.5 Cluster-Aware Spatial Optimization

To ensure that the recommended VAs form a spatially coherent and walkable itinerary, this study introduces a two-step cluster-aware optimization process. The first step selects geographically compact VA groups via spatial clustering, while the second step optimizes the visiting order of selected VAs using a genetic algorithm to minimize travel distance.

Algorithm 2 Spatial Clustering for VA Selection

Require: Sorted list of VAs $\mathcal{V}_{\text{rerank}} = \{v_1, v_2, \ldots, v_k\}$ by LLM score, distance threshold τ, minimum total VAs n_{min}, minimum cluster size n_{cmin}
Ensure: Candidate VAs list $\mathcal{V}_{\text{c}}$

$\mathcal{C} \leftarrow [\,]$ {Initialize empty list of clusters}
$\mathcal{V}_{\text{rerank}} \leftarrow \emptyset$
for $i = 1$ to k **do**
  $v \leftarrow v_i$
  assigned $\leftarrow$ **false**
  for all $C_j \in \mathcal{C}$ **do**
    if $\exists v' \in C_j$ such that $\text{dist}(v, v') < \tau$ **then**
      $C_j \leftarrow C_j \cup \{v\}$; assigned $\leftarrow$ **true**
      break
    end if
  end for
  if not assigned **then**
    $\mathcal{C} \leftarrow \mathcal{C} \cup \{\{v\}\}$ {Create new cluster}
  end if
  $\mathcal{V}_{\text{c}} \leftarrow \bigcup\{C \in \mathcal{C} : |C| \geq n_{cmin}\}$
  if $|\mathcal{V}_{\text{c}}| \geq n_{min}$ **then**
    break
  end if
end for
return $\mathcal{V}_{\text{c}}$ =0

Step 1: Spatial Clustering for VA Selection. Given the Top-k candidate attractions ranked by semantic relevance, this study applys an incremental clustering algorithm that evaluates each VA based on its proximity to existing clusters. A new VA is assigned to a cluster if it lies within a specified distance threshold τ of any member in that cluster; otherwise, a new cluster is created. Clusters with fewer than n_{cmin} members are discarded. The process continues until the number of VAs in valid clusters exceeds a minimum threshold n_{min}. This approach ensures that only sufficiently dense and spatially compact clusters are retained, supporting walk-friendly itineraries. The clustering procedure is outlined in Algorithm 1.

Step 2: Genetic Algorithm for VA Ordering. To determine an optimal visiting order among the selected candidate VAs, this study employs a genetic algorithm that minimizes the path length between locations. The population is initialized with random permutations of the VA list. In each generation, individuals are evaluated

Algorithm 3 Genetic Algorithm for VA Ordering

Require: Candidate VAs $\mathcal{V}_c$, distance matrix $\mathcal{D}$
Ensure: Ordered list of candidate VAs $\mathcal{V}_{\text{order}}$

$\mathcal{P} \leftarrow \{P_1, P_2, \ldots, P_g\}$ {Initialize population}
$t \leftarrow 0$ {Initialize the generation count}
while $t < t_{\max}$ **do**
  for $i = 1$ to g **do**
    fitness(P) {Calculate the fitness score for each P_i in $\mathcal{P}$}
  end for
  for $j = 1$ to $\frac{g}{2}$ **do**
    $P_a, P_b \leftarrow$ selection() {Select two parent routes P_a, P_b based on their fitness}
    $C_a, C_b \leftarrow$ crossover(P_a, P_b) {Crossover between parents to generate children}
    $C_a, C_b \leftarrow$ mutation(C_a, C_b) {Apply mutation to children to introduce variability}
    $P_{\text{new}} \leftarrow$ add(C_a, C_b) {Add C_a, C_b to a new population P_{new}}
  end for
  $t \leftarrow t + 1$ {Increment generation count}
end while
$\mathcal{V}_{\text{order}} \leftarrow P_{\text{best}}$ {Return the best route P_{best} based on the highest fitness score}
=0

using a fitness function based on the total distance traveled. Selection, crossover, and mutation operations are applied iteratively to evolve better route candidates. The algorithm terminates after a fixed number of generations, and the best individual is returned as the optimized sequence. The complete procedure is shown in Algorithm 2.

Step 3: A* Algorithm with Scenic-Aware Cost Function. To generate walking routes that are both efficient and visually pleasant, this study extends the A* search algorithm by incorporating road segment scenic scores into the cost function. Each edge cost is adjusted based on the visual attractiveness of the path, encouraging the algorithm to favor scenic routes without excessively increasing total distance.

This study define the scenic-aware traversal cost between two connected nodes u and v as:

$$\text{cost}(u, v) = \frac{\text{length}(u, v)}{\text{scenic\_score}(u, v)^{\alpha} + \epsilon} \tag{4.6}$$

where:

- length(u, v) is the physical distance (e.g., in meters) of the road segment,

- scenic_score$(u, v) \in [1, 5]$ is the rescaled visual quality score of the segment,
- α is a tunable parameter controlling scenic influence ($\alpha > 0$),
- $\epsilon > 0$ is a small constant to prevent division by zero.

This function allows the algorithm to prioritize scenic paths by reducing the effective cost of high-quality segments, while preserving route realism. Setting $\alpha = 0$ recovers standard shortest-path routing, while increasing α emphasizes scenic preference.

Algorithm 4 Scenic-Aware A* Algorithm

Require: Graph $G = (V, E)$ with scenic score $ss(e)$ on edges, start s, goal g
Ensure: Path from s to g minimizing scenic-adjusted cost

$open\_set \leftarrow \{s\}$
$came\_from[v] \leftarrow$ None for all $v \in V$
$g\_score[v] \leftarrow \infty$, $g\_score[s] \leftarrow 0$
$f\_score[v] \leftarrow \infty$, $f\_score[s] \leftarrow h(s, g)$
while $open\_set \neq \emptyset$ **do**
  $current \leftarrow$ node in $open\_set$ with lowest f_score
  if $current = g$ **then**
    return `reconstruct_path`$(came\_from, current)$
  end if
  Remove $current$ from $open\_set$
  for each neighbor n of $current$ **do**
    $L \leftarrow$ length$(current, n)$
    $ss \leftarrow$ scenic_score$(current, n)$
    $w \leftarrow \dfrac{L}{sc^{\alpha} + \epsilon}$ {scenic-adjusted edge cost}
    $tentative\_g \leftarrow g\_score[current] + w$
    if $tentative\_g < g\_score[n]$ **then**
      $came\_from[n] \leftarrow current$
      $g\_score[n] \leftarrow tentative\_g$
      $f\_score[n] \leftarrow g\_score[n] + h(n, g)$
      if $n \notin open\_set$ **then**
        Add n to $open\_set$
      end if
    end if
  end for
end while
return failure =0

This extension enables flexible multi-objective routing, balancing distance efficiency with user-perceived route quality. When applied to urban tourism, it encourages walkable paths that pass through scenic streets, riversides, or landmark-dense areas—enhancing the experiential value of itineraries while maintaining route practicality.

5 Experiments and Results

5.1 Experiments Setting

5.1.1 VA Distribution in Experimental Cities

To evaluate the spatial characteristics of curated datasets, this study analyze the distributions of VAs in the two experimental cities. For each city, this study collected a set of geolocated VAs—981 in Paris and 867 in Rome–Vatican. These VAs were augmented with user-generated reviews to extract descriptive and perceptual features, forming the foundation of semantic attraction database.

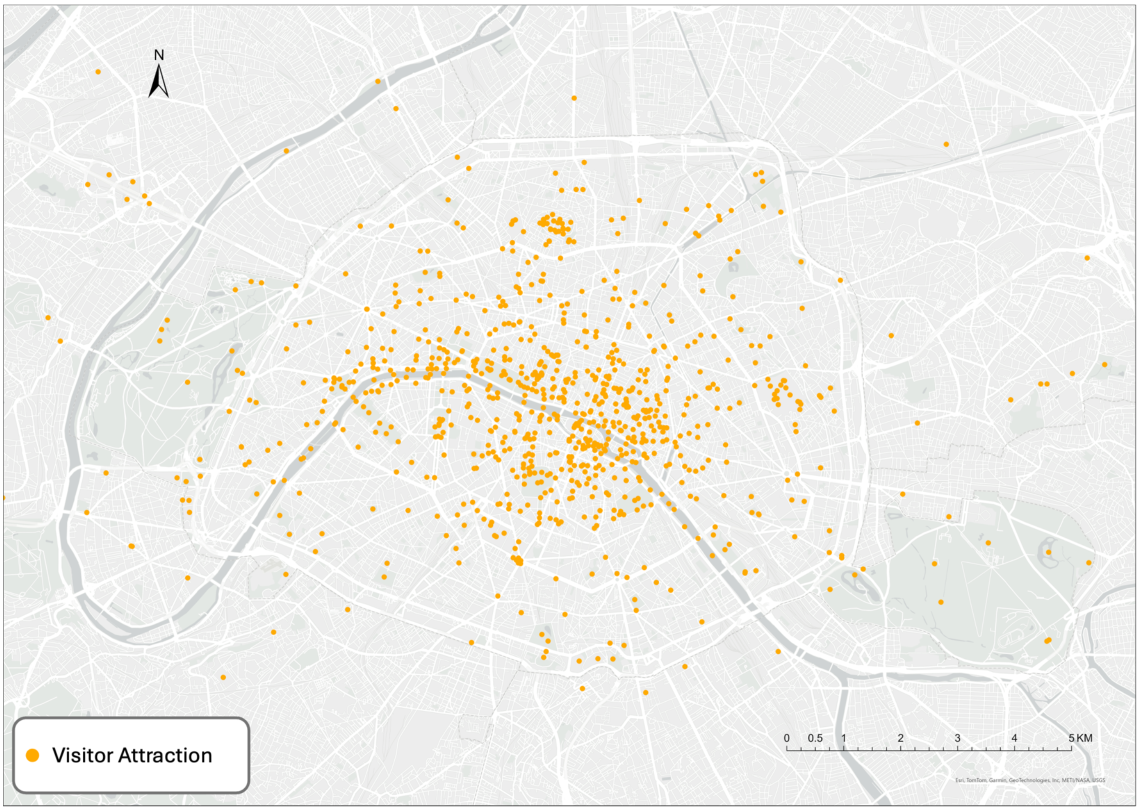


Figure 5.1: Spatial distribution of visitor attractions in Paris.

Figure 5.1 illustrates the spatial distribution of VAs in Paris. As observed, the highest density of attractions is located in central Paris, particularly around the 1st to 6th arrondissements. This includes well-known neighborhoods such as Le Marais, Île de la Cité, and Saint-Germain-des-Prés. Notably, density also extends westward along the Seine and eastward toward Bastille and Nation, suggesting a walkable cultural corridor. Peripheral areas such as Boulogne-Billancourt or Montreuil exhibit sparse coverage, highlighting the centralization of tourist interest zones.

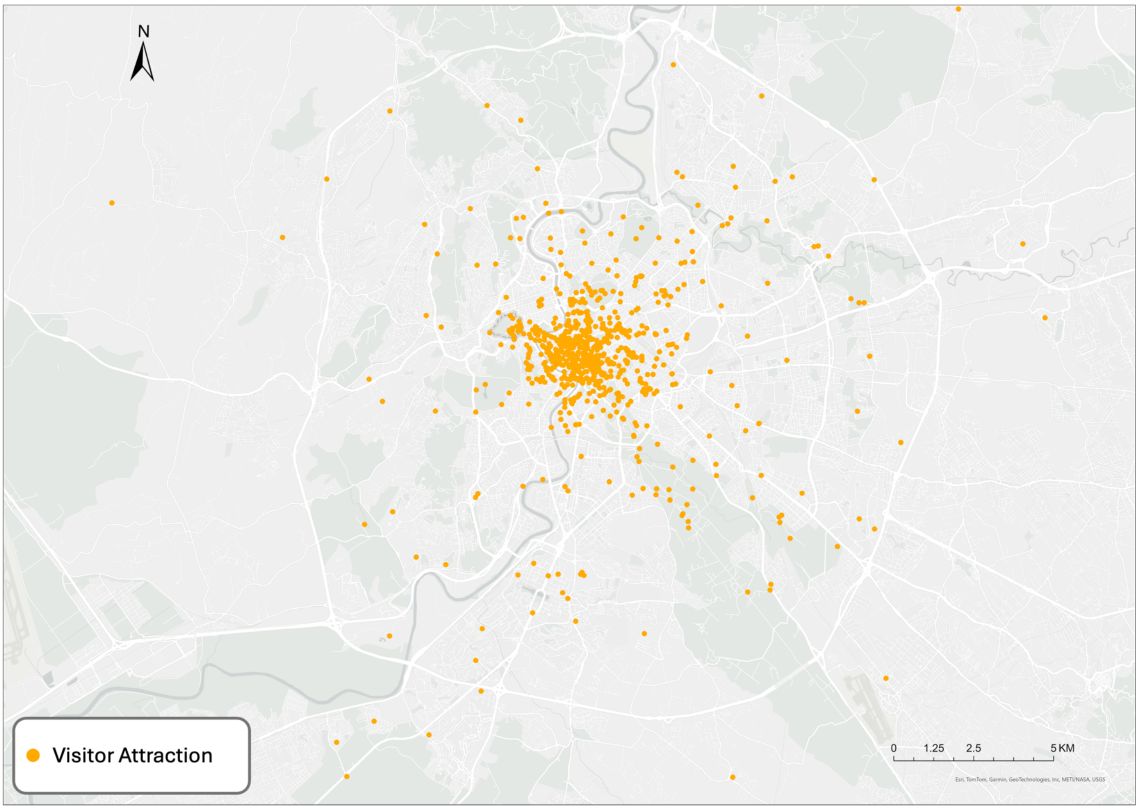


Figure 5.2: Spatial distribution of visitor attractions in Rome–Vatican.

Figure 5.2 presents the spatial distribution of VAs in Rome–Vatican. Similar to Paris, Rome's visitor attractions are densely concentrated in the historical city center, with prominent clusters around the Colosseum, Roman Forum, Trevi Fountain, and Vatican City. The overall spatial spread appears more compact than that of Paris, with a pronounced density drop beyond the central districts.

While both cities exhibit strong central clustering, Paris demonstrates a broader east–west distribution aligned with the Seine River, facilitating attraction connectivity across a larger urban span. In contrast, Rome's denser and more compact clustering emphasizes its ancient core, suggesting a more spatially constrained tourist experience. These distinctions reflect differences in historical development, urban form, and the spatial logic underlying cultural accessibility in each city.

5.1.2 User Queries Generation

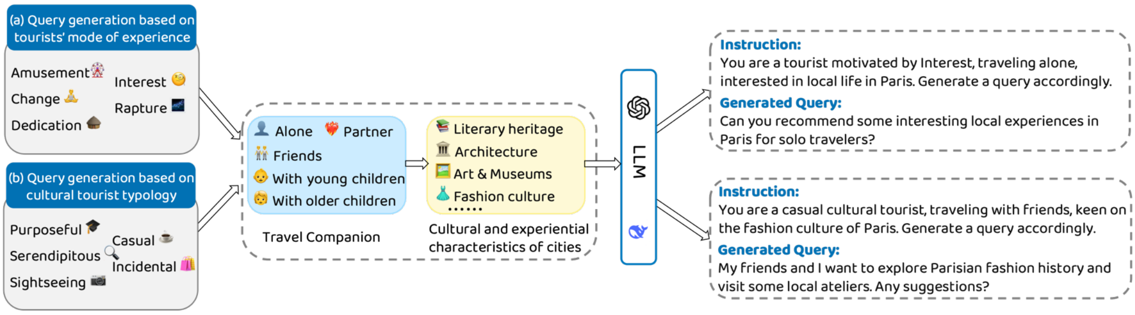


Figure 5.3: An overview of query generation based on tourists' motivation types, travel companions, and urban characteristics. Two generation pathways are illustrated: (a) experience-based typology and (b) cultural tourist typology. These factors are composed into LLM prompts to simulate diverse queries.

To simulate diverse user intentions grounded in psychological, cultural, and social motivations, this study adopt a multidimensional framework for generating natural language queries. This framework supports two complementary generation strategies, each rooted in established tourism typologies and contextualized through travel companionship and urban cultural profiles.

First, this study draw upon Elands and Lengkeek's refinement of Cohen's tourist experience theory (Elands and Lengkeek, 2000; Cohen, 1979). Their work provides a detailed typology of *modes of experience* in tourism, outlining a spectrum of motivations—*Amusement*, *Change*, *Interest*, *Rapture*, and *Dedication*—that represent distinct experiential orientations. The Table 5.1 shows a more detailed breakdown and examples of these types. These motivational profiles were systematically paired with travel companion contexts (e.g., *alone*, *with a partner*, *with young or older children*, or *with friends without children*) and grounded in the cultural and experiential characteristics of real-world cities to simulate general leisure and meaning-seeking user queries in urban environments.

Table 5.1: Modes of Experience (Elands and Lengkeek, 2000)

Mode of Experience	**Underlying Characteristics**	**Items**
Amusement	Fun	For me, having a nice time on vacation means drinking coffee or a beer with the neighbors.
		I like to go to places that attract many tourists and are nice and busy.
	Centre-values	I like to eat Dutch food on vacation.

Mode of Experience	**Underlying Characteristics**	**Items (continued)**
		I like to hear Dutch spoken when I'm on vacation.
	Temporality	I like to go on vacation, but I also like it to go home again.
Change	Escape	I go on vacation to get out of the daily grind.
		I have such a stressful job that I need to escape once in a while / because of the pressure of my daily activities, I have to go out once in a while.
Relaxation	Rest	The most important thing in my vacation is relaxation / I go on vacation for a good rest and relaxation.
	Idleness	To me, vacation means being idle, sunbathing and doing nothing.
	Recover	I need vacation to recharge my batteries.
		I have to go on vacation at least once a year to recover.
		It takes me the first days of a vacation to unwind and forget about my job or housework.
	Context matters	I don't care where I go on vacation, I just have to get away.
Interest	Search for interesting vistas and stories	I always visit a church, castle or historic city centre when I'm on vacation.
		On vacation I don't feel like visiting a church, castle or historic city centre (-).
	Cultural activities	I like to go to local cultural activities.
	Stimulation of imagination	I always read the information boards at tourist sites.
		I always take a travel guide and a map of the area with me on vacation.
		When I'm on vacation, I go first to the local tourist office for specific information about the area.
	Variation	On vacation I want to see new and various things all the time.

Mode of Experience	Underlying Characteristics	Items (continued)
		I like to choose a different vacation destination each year.
Rapture	Self-discovery	When I'm on vacation I like to be alone in the great outdoors for hours on end.
		During my vacation I finally find time for myself.
	Crossing borders	On vacation I like sporty challenges and surprises.
		I have no objections to primitive conditions when I'm on vacation.
		I like active vacations doing strenuous things such as long treks and cycle tours.
	Unexpected	On vacation I like it the most when, beforehand, I have no idea where I will go.
		On vacation I like to be confronted with new experiences and surprises.
Dedication	Quest for authenticity	Once an area starts getting touristy I don't go back.
		My first choice is exotic vacation destinations.
		On vacation I search for wilderness and original landscapes where I won't meet anybody.
	Merge	I am not satisfied with just seeing local cultures and their habits. I would rather be part of it.
		For me vacation means totally immersing myself in other cultures / on vacation I immerse myself totally in another culture.
	Appropriation and devotion	I rather go to the same area because I feel bonded to it.
		The area where I always go on vacation, I really consider as my place.
		I visit .... (fill in name destination) because ... plays an important role in my life.
	Timeless	I would like to live in ... / If I could I would like to live in my vacation place.

Second, recognizing that cultural tourism constitutes a significant subset of urban tourism, this study incorporate McKercher 's typology of cultural tourists, seeing Table 5.2 (McKercher and Du Cros, 2003). This model distinguishes five types of cultural tourists—*Purposeful*, *Serendipitous*, *Sightseeing*, *Casual*, and *incidental*—based on the centrality of cultural motivations and the depth of cultural engagement. These types were similarly paired with travel companion profiles and enriched with city-specific cultural images to guide LLMs in simulating user queries reflecting diverse forms of culturally oriented intent.

Table 5.2: Types of Cultural Tourism (McKercher and Du Cros, 2003)

Types	**Description**
Purposeful Cultural Tourist	The purposeful cultural tourist (high centrality / deep experience). Learning about the other's culture or heritage is a major reason for visiting a destination, and this type of cultural tourist has a deep cultural experience.
Serendipitous Cultural Tourist	The serendipitous cultural tourist (low centrality / deep experience). Cultural tourism plays little or no role in the decision to visit a destination, but while there, this tourist visits cultural attractions and ends up having a deep experience.
Sightseeing Cultural Tourist	The sightseeing cultural tourist (high centrality / shallow experience). Learning about the other's culture or heritage is a major reason for visiting a destination, but this type of tourist has a more shallow, entertainment-oriented experience.
Casual Cultural Tourist	The casual cultural tourist (modest centrality / shallow experience). Cultural tourism reasons play a limited role in the decision to visit a destination and this type of cultural tourist engages the destination in a shallow manner.
Incidental Cultural Tourist	The incidental cultural tourist (low centrality / shallow experience). Cultural tourism plays no meaningful role in the destination decision-making process. However, while at the destination, the person will participate in cultural tourism activities, having a shallow experience. Incidental cultural tourists prefer visiting easy-to-consume, low-involvement, well-known, entertainment-oriented, mass tourism cultural attractions.

The user query generation process is illustrated in Figure 5.3, which summarizes how motivational typologies, cultural intent categories, travel context, and cultural and experiential characteristics of cities were combined to construct a semantically diverse and realistic set of user queries. In generated queries, 35.3% of the Paris queries and 36.1% of the Rome-Vantican queries were derived from the cultural tourist typology, with the remainder grounded in the mode of experience framework.

5.2 Evaluation Metrics

this study adopt a combination of semantic and spatial metrics to evaluate the relevance, efficiency, and spatial coherence of each generated itinerary. Let $\mathcal{V}_{\text{order}} = \{v_1, v_2, \ldots, v_N\}$ denote the ordered set of selected attractions in a given itinerary, and let $\mathbf{q}$ denote the user query. The Euclidean distance between two attractions v_i and v_j is denoted by $d(v_i, v_j)$.

Hit Rate (HR) Hit Rate measures the proportion of attractions in the itinerary that are semantically relevant to the user query $\mathbf{q}$:

$$\text{HR} = \frac{1}{|\mathcal{V}_{\text{order}}|} \sum_{v \in \mathcal{V}_{\text{order}}} I\big[\text{Relevant}(v, \mathbf{q}) = 1\big], \tag{5.1}$$

where $I[\cdot]$ is the indicator function. Relevance is assessed via LLM judgement and verified through human annotation.

Average Margin (AM) Average Margin measures the difference in total Euclidean distance between the generated itinerary and the optimal Traveling Salesman Problem (TSP) solution over the same set of attractions:

$$\text{AM} = D(\mathcal{V}_{\text{order}}) - D^*(\mathcal{V}_{\text{order}}), \tag{5.2}$$

where $D(\cdot)$ denotes the total distance of the visiting order, and $D^*(\cdot)$ is the optimal TSP distance over the same set.

Travel Distance (TD) Travel Distance is the total Euclidean distance incurred when visiting attractions in the recommended order:

$$\text{TD} = \sum_{i=1}^{N-1} d(v_i, v_{i+1}). \tag{5.3}$$

Spatial Tightness (ST) Spatial Tightness measures how spatially clustered the selected attractions are, regardless of their visiting order:

$$\text{ST} = \frac{1}{N} \sum_{i=1}^{N} \min_{j \neq i} d(v_i, v_j). \tag{5.4}$$

Scenic Route Gain Ratio (SGR) Scenic Route Gain Ratio quantifies the scenic value gained per unit of walking distance, offering a normalized measure of experiential enrichment in route planning. Let S_{route} denote the total scenic score accumulated along the generated route, and D_{route} the corresponding walking distance. Then:

$$\text{SGR} = \frac{S_{\text{route}}}{D_{\text{route}} + \epsilon} \tag{5.5}$$

where ϵ is a small constant to avoid division by zero. A higher SGR reflects a more scenic route per unit of effort, balancing scenic richness with walking cost. This formulation allows for clear comparison in ablation studies, even when scenic-aware planning is disabled, and avoids the scale mismatch between score and distance in difference-based metrics.

5.3 Results

5.3.1 Scenic Score Visualization over Road Network

To ensure topological consistency in scenic scoring pipeline, this study first apply DFS to extract the largest connected component from the OSM road network in each city. This step is critical for enabling shortest-path calculations and ensuring every node is reachable from any other, as required for subsequent route-based analysis.

Figure 5.4, Figure 5.5, Figure 5.6 and Figure 5.7 show the road networks of Paris and Rome before and after cleaning. In both cities, this study observe that the raw networks contain fragmented road segments or small disconnected subgraphs, especially at the periphery or in suburban areas. After applying DFS-based filtering, only the largest connected subgraph is retained, resulting in a continuous, navigable network suitable for scenic evaluation.

Based on the cleaned networks, this study visualize the computed scenic scores at the segment level for both cities. As shown in Figure 5.8 and Figure 5.9, each road segment is assigned a score based on photo density, alignment of photo distribution with road geometry, and proximity to interactive POIs. The scores are mapped using a continuous color gradient from blue (low scenic value) to red (high scenic value).

In Paris, the spatial distribution of scenic scores is strongly center-weighted. The

Figure 5.4: Raw road network of Paris before connectivity cleaning.

Figure 5.5: Cleaned connected road network of Paris after DFS-based filtering.

Figure 5.6: Raw road network of Rome before connectivity cleaning.

Figure 5.7: Cleaned connected road network of Rome after DFS-based filtering.

Figure 5.8: Scenic Score Distribution in Paris.

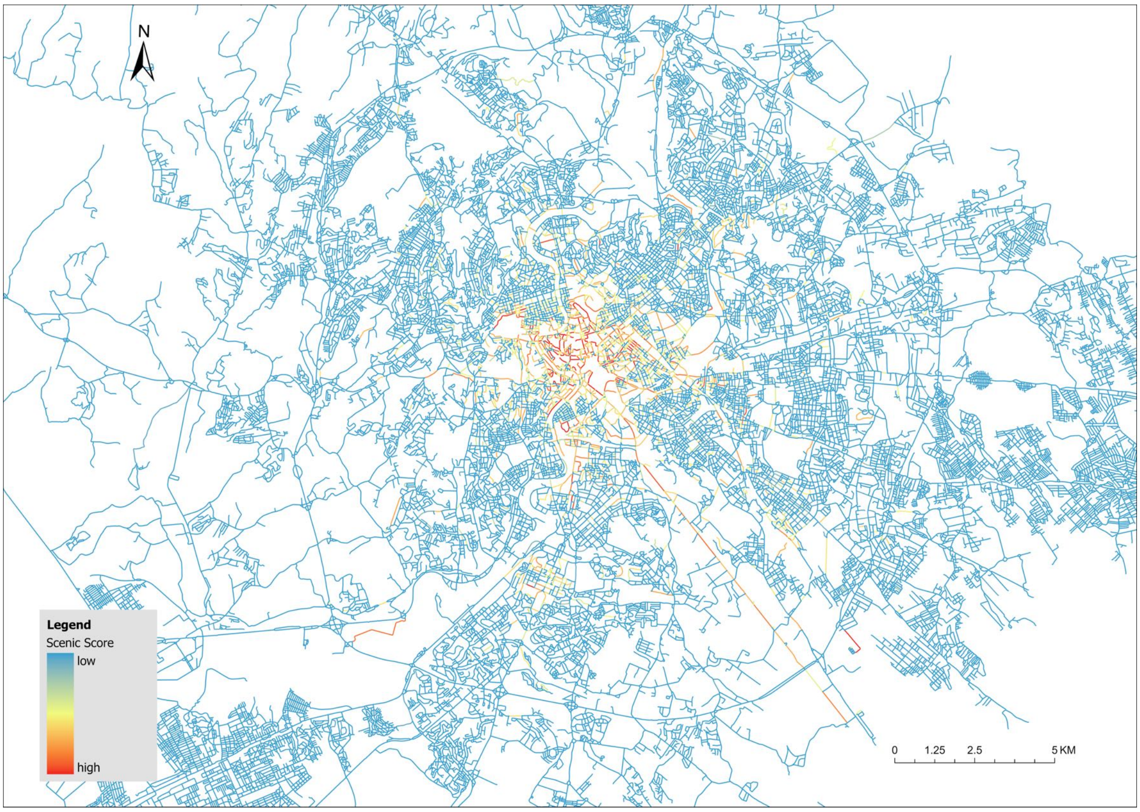


Figure 5.9: Scenic Score Distribution in Rome-Vatican.

highest-value segments cluster in the historic core—the 1st to 8th arrondissements — including the *Louvre–Tuileries–Palais-Royal* axis, *Île de la Cité / Île Saint-Louis*, *Opéra–Madeleine–Saint-Honoré*, *Saint-Germain–Musée d'Orsay*, *Invalides*, and the *Champs-Élysées–Concorde* corridor. A secondary hotspot occurs on and around *Montmartre*, where elevated viewpoints and dense heritage streets (e.g., around *Sacré-Cœur* and *Place du Tertre*) yield consistently high scores.

In addition, the Seine riverfront emerges as a continuous "scenic ribbon": quays on both banks and bridge approaches exhibit above-average values, reflecting waterfront vistas and landmark density. Scores decline toward peripheral districts, where residential grids dominate, producing a clear center-to-edge gradient. These patterns suggest that scenic-aware routing will naturally favor riverfront corridors and central heritage streets, with optional detours to Montmartre when elevation and viewpoints are desired.

In Rome–Vatican, the spatial pattern of scenic scores is highly concentrated. The highest values cluster tightly in the historic core and within *Vatican City*, and a continuous ribbon of elevated scores runs along the *Tiber* riverfront. Outside these hotspots, scores drop off quickly: most streets in the broader urban area register low scenic values—not because they are inherently unwalkable, but because they offer fewer landmarks, viewpoints, and cultural cues that attract tourists. This produces a pronounced core–periphery gradient in which a relatively small central zone accounts for most high-scenic segments. For routing, scenic-aware paths will naturally remain in the historic center or track the river, while excursions into outer districts tend to yield limited scenic gain.

This spatial differentiation highlights the ability of the scenic score framework to capture not only aesthetic and experiential quality but also the underlying spatial structure of tourism intensity.

5.3.2 Overall Results

Table 5.3: Comparison between *UGuideRAG* and *ITINERA* on the Paris and Rome–Vatican datasets.

City	**Method**	**HR** (%)↑	**AM** (m)↓	**TD** (m)↓	**ST** (m)↓	**SGR** ↑	
						w/o SP	**with SP**
Paris	ITINERA	42.3	**455.8**	**5816.3**	441.4	4.3	4.5
	UGuideRAG	**78.5**	651.2	6781.4	**301.8**	5.0	**5.2**
Rome–Vatican	ITINERA	33.8	**446.0**	5367.3	413.5	4.6	4.7
	UGuideRAG	**72.7**	582.4	**5227.1**	**231.8**	4.8	**5.0**

this study evaluate the full UGuideRAG framework against ITINERA[14] (Tang et al., 2024), a recent LLM-based itinerary recommender whose spatial stage relies on a density-based spatial clustering method. Table 5.3 presents a comparison of the two systems in the Paris and Rome-Vatican datasets.

In terms of semantic alignment, UGuideRAG achieves significantly higher HR in both cities—78.5% in Paris and 72.7% in Rome-Vatican—compared to ITINERA (42.3% and 33.8%, respectively). These results indicate a stronger match between user query and the recommended VAs.

For spatial metrics, both systems achieve nearly identical AM values, suggesting comparable efficiency in visiting order relative to the optimal TSP baseline. Despite variations across cities, UGuideRAG maintains TD values around 6000 meters, which translates to a feasible walking distance for a day itinerary, ensuring practical usability for urban tourists. Additionally, UGuideRAG consistently achieves low ST values across both cities, indicating that the recommended attractions are geographically well-clustered and exhibit strong walkable connectivity. Notably, enabling scenic route planning (SP) consistently improves scenic gain for both methods: the SGR with SP exceeds its counterpart without SP for each method, while the actual route distance does not noticeably increase.

Taken together, these findings demonstrate that UGuideRAG delivers substantially improved semantic relevance while maintaining competitive spatial performance. This highlights its potential to improve user satisfaction through context-aware itinerary recommendations without imposing additional travel burden.

5.4 Ablation Study

To assess the individual contributions of each module in the UGuideRAG framework, this study conduct an ablation study on both the Paris and Rome-Vatican datasets (Table 5.4). this study examine four ablated variants: (1) without intent decomposition and UGC-derived VA features (w/o Intent Decomposition & UGC), (2) without intent decomposition (w/o Intent Decomposition), (3) without the LLM-based reranker (w/o LRR), and (4) without cluster-aware spatial optimization (w/o CSO), keeping all other components intact. In addition, this study also compare against two modified variants of the ITINERA. Since ITINERA's original density-based clustering is not well-suited for high-density VA regions such as Paris and Rome-Vatican, this study re-implement ITINERA using UGuideRAG's

[14] Results are obtained using the authors' original implementation.

clustering strategy: (5)ITINERA with UGuideRAG's CSO only, and (6)ITINERA with UGuideRAG's CSO and LRR.

Table 5.4: Performance comparison across different ablation settings on Paris and Rome–Vatican datasets.

City	Method	HR↑ (%)	AM↓ (m)	TD↓ (m)	ST↓ (m)	SGR (%)↑	
						w/o SP	**with SP**
Paris	**UGuideRAG (Full)**	78.5	651.2	6781.4	**301.8**	5.0	5.2
	w/o Intent Decomposition & UGC	52.0	805.7	6752.1	358.5	5.1	5.3
	w/o Intent Decomposition	66.9	**539.2**	6370.7	329.3	5.5	5.7
	w/o LRR	66.4	658.9	6249.9	310.7	4.8	4.9
	w/o CSO	**80.1**	16424.4	36592.7	1535.7	7.2	**7.7**
	ITINERA (w/ UGuideRAG's CSO, w/o LRR)	59.0	651.4	**6001.1**	310.2	5.1	5.3
	ITINERA (w/ UGuideRAG's CSO and LRR)	72.6	627.0	6423.6	312.0	5.4	5.7
Rome–Vatican	**UGuideRAG (Full)**	**72.7**	582.4	5227.1	231.8	4.8	5.0
	w/o Intent Decomposition & UGC	52.7	**369.6**	**3990.1**	**231.0**	5.0	5.0
	w/o Intent Decomposition	64.0	668.3	5476.3	255.5	5.0	5.3
	w/o LRR	63.4	563.3	4736.5	242.9	5.3	5.5
	w/o CSO	72.1	12035.4	21998.2	815.6	5.7	**6.2**
	ITINERA (w/ UGuideRAG's CSO, w/o LRR)	56.7	512.6	5702.8	267.5	5.2	5.4
	ITINERA (w/ UGuideRAG's CSO and LRR)	65.5	669.8	5064.2	259.2	5.1	5.2

The w/o Intent Decomposition & UGC variant relies on each VA's Wikipedia [15] summary for attraction matching, without leveraging structured user intent and UGC-derived VA features. It operates on a reduced attraction pool due to the limited availability of Wikipedia descriptions (454 VAs for Paris and 467 for Rome). This setting yields the lowest HR across both cities, with a notable performance drop compared to the w/o Intent Decomposition variant. These results further highlight the foundational importance of extracting rich experiential VA features from UGC for personalized, fine-grained recommendations.

When the LLM-based reranker module is removed, the system experiences a noticeable drop in recommendation performance, highlighting the importance of fine-grained contextual ranking. While intent decomposition ensures that retrieval broadly aligns with user intent, removing the LLM-based reranker limits the system's ability to distinguish fine-grained semantics beyond what embeddings can represent.

Removing the intent decomposition module leads to a significant drop in HR, as the system fails to infer user intent from multi-faceted, ambiguous, or implicit queries. This degrades retrieval quality and limits the effectiveness of downstream LLM-based reranking.

Although HR remains high, removing the CSO module results in severe degradation of spatial metrics. In particular, AM increases by over 25×, while TD and ST increase by approximately 5×. This indicates that although the selected attractions are semantically relevant, they are spatially scattered and inefficiently ordered.

[15] https://en.wikipedia.org/

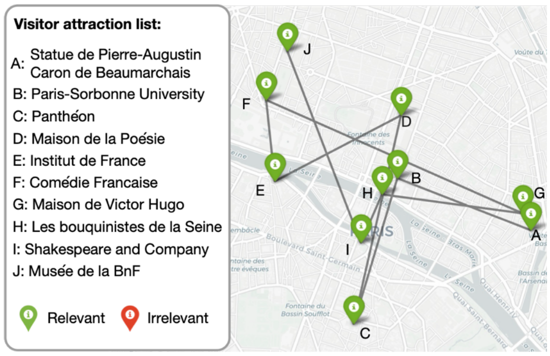
Therefore, CSO is essential for producing spatially coherent and walkable itineraries.

5.5 Case Study

To further demonstrate the effectiveness of proposed framework, this study present two case studies: one comparing recommendation results across methods for a themed query, and another showcasing UGC-based detection of semantic spatial features such as panoramic viewpoints.

5.5.1 Performance Comparison under a Cultural Theme Scenario

(a) **UGuideRAG (Full)**



(b) **UGuideRAG w/o Cluster-aware Spatial Optimization**

(c) **UGuideRAG w/o LLM-based Reranker**

(d) **ITINERA**

Figure 5.10: Case study comparison of recommended attractions across methods for the query *"I'm interested in French literature"*.

To further demonstrate the effectiveness of the framework, this study present a case study based on the user query: *"I'm interested in French literature. What places do you recommend?"* this study compare the outputs of four systems previously introduced in the ablation study: the full **UGuideRAG**, its variants **w/o**

Cluster-aware Spatial Optimization and **w/o LLM Reranker**, and the baseline **ITINERA**.

Figure 5.10 shows the recommended itineraries generated by each method. The selected VAs are labeled alphabetically (A–K), with names listed in each subfigure's legend. Detailed descriptions of all VAs will be included in the supplementary materials.

The full UGuideRAG framework produces the most semantically aligned and diverse itinerary. It identifies a rich mix of attractions closely related to the theme of French literature, including iconic author residences such as *Maison de Victor Hugo* and *Lauzun Hotel* (associated with Baudelaire), as well as sculptures and monuments dedicated to French playwrights, including *Fontaine Molière* and the *Statue de Beaumarchais*. The itinerary also features cultural landmarks like the *Musée de la BnF*, the *Institut de France*, and the historic literary theater *Comédie Française*, along with experiential VAs such as the riverside secondhand book market *Les bouquinistes de la Seine* and the renowned bookstore *Shakespeare and Company.* These results highlight UGuideRAG's strength in identifying attractions related to French literary culture, including historic author residences, public monuments, national literary institutions, and reader-focused VAs such as secondhand book markets and independent bookstores. The resulting itinerary combines well-established landmarks with immersive experiences, offering a coherent and multifaceted exploration of the literary landscape of the city.

Removing the CSO module does not significantly alter the set of selected attractions but results in a disorganized and spatially scattered itinerary. The absence of spatial coherence highlights CSO's essential role in optimizing the visit order and improving overall travel feasibility without sacrificing semantic alignment.

The variant without the LRR still benefits from intent decomposition and successfully retrieves many relevant sites, including *Victor Hugo's house*, *Shakespeare and Company*, *Fontaine Molière*, *Statue de Corneille*, *Comédie Française*, and the *Musée de la BnF*. However, it also includes more marginally relevant or thematically ambiguous places such as the *Musée du Barreau de Paris* and the *Musée Carnavalet*, reflecting a lack of contextual nuance. Despite this, its output is notably more on-topic and diverse than *ItiNera*, suggesting that even in the absence of reranking, structured intent modeling significantly improves semantic relevance in retrieval.

ITINERA, which extracts sub-requirements directly from the user query without explicit intent reasoning, yields the least thematically aligned list. It does include clearly literary venues—such as *Maison de la Poésie*, *Statue de Corneille*, *Institut de*

France, and *Palais Garnier*—but many recommendations are only weakly related to literature or off-theme, including *Remains of the Batille*, *Musée Rodin*, *Square Samuel Rousseau*, *Louvre Pyramid*, *Domaine National du Palais-Royal*, *Musée Banksy – Paris*. This outcome highlights a limitation of direct query parsing: despite a clearly stated literary intent, the system often returns attractions that are superficially relevant but thematically misaligned.

This case illustrates the importance of UGuideRAG's intent decomposition strategy as a key enabling component. By structuring user queries into experiential dimensions—landscape and content, activities, and atmosphere—the system establishes a meaningful foundation for subsequent semantic alignment. However, this potential is fully realized through the addition of the LLM-based reranker module, which enables deep contextual understanding and nuanced evaluation of candidate attractions based on the user's full intent. Together, these components allow UGuideRAG to generate personalized itineraries that are semantically aligned, experientially coherent, and spatially optimized. This results in a more interpretable, engaging and meaningful travel experience.

5.5.2 Panoramic View Recognition through UGC

In contrast to traditional spatial modeling that relies on elevation data or visibility analysis, this approach leverages UGC to detect panoramic viewpoints based on how visitors describe their experiences. UGC supplies attraction-level semantics that 2D maps do not encode, revealing *where* unique views are actually experienced and which landmarks are inter-visible from a given vantage.

Figure 5.11 shows a set of VAs identified through UGC as offering "panoramic views." These include well-known elevated landmarks such as *Eiffel Tower*, *Montmartre*, and *Arc de riomphe*, as well as lesser-known overlooks like *Square Nadar* that emerge in reviews as quiet spots suitable for full-city vistas. In the lower image of Figure 5.11, an author photograph from *Montmartre* clearly shows three visually prominent landmarks predicted by the UGC signals—*Tour Saint-Jacques*[16], *Panthéon*[17], and *Centre Pompidou*—thereby validating that UGC recovers viewpoint attributes beyond what 2D cartography alone provides.

Importantly, the UGC signal also reveals opportunities invisible to elevation-based heuristics. For example, *Parc André Citroën* lies on flat terrain yet is repeatedly

[16] https://maps.app.goo.gl/ciqUykjPBS4ikPTn8
[17] https://maps.app.goo.gl/UTu8edFuUBoUbtE99

tagged with "aerial views" because its tethered helium balloon provides sweeping citywide vistas. By mining natural-language cues in reviews and descriptions—such as "wide views," "city skyline," or "great lookout point"—this framework surfaces landscape-level semantics, captures how users perceive and interact with space, and enables more human-centered, viewpoint-aware retrieval and recommendation.

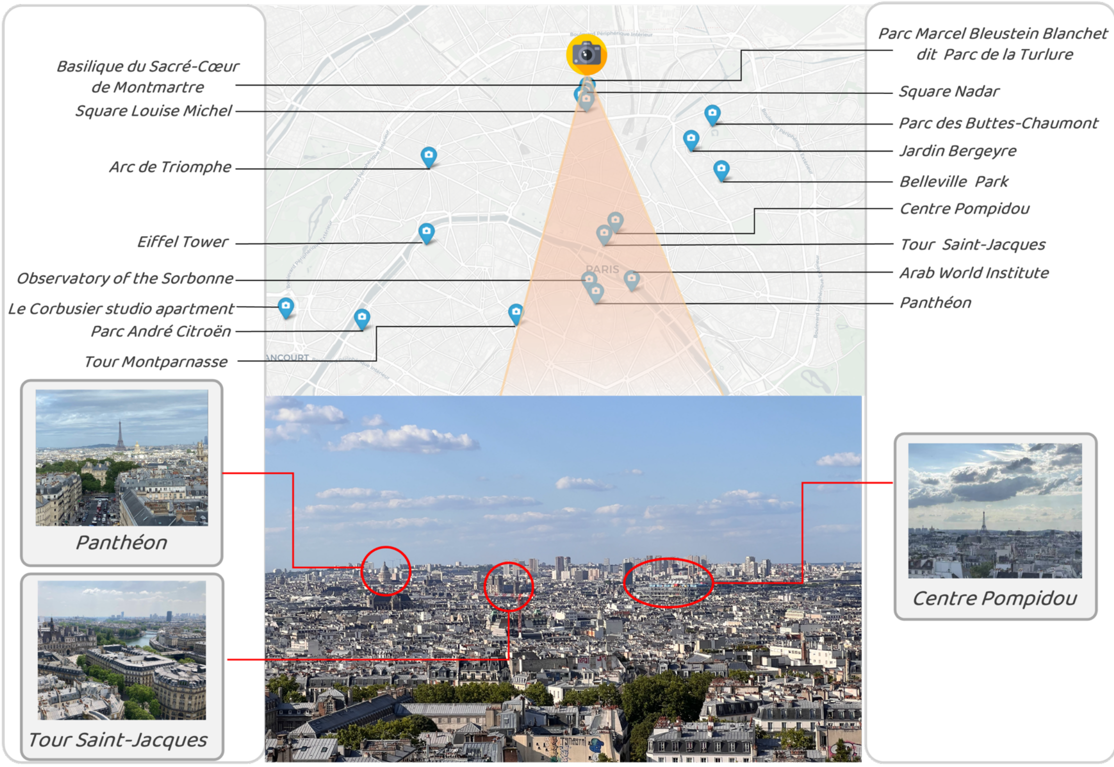


Figure 5.11: UGC-surfaced panoramic viewpoints in *Paris*. **Top:** VAs whose reviews explicitly mention panoramic-view cues; the shaded sector marks the approximate viewing direction and field of view from the indicated camera location (used to illustrate inter-visibility). **Bottom:** author photograph from *Montmartre* in which three predicted landmarks—*Panthéon*, *Tour Saint-Jacques*, and *Centre Pompidou*—are clearly visible.

6 Discussion

6.1 Interpretation of Overall Results

The effectiveness of UGuideRAG demonstrates a valuable approach for addressing the persistent challenges that have limited traditional tourism recommendation systems. Prior methodologies, while valuable, have struggled to meet the nuanced demands of modern travelers, particularly those engaging in immersive urban exploration like Citywalks. For instance, user interaction-based systems that rely on broad categories fail to deliver truly personalized suggestions (Savir et al., 2013; Meehan et al., 2013; Lu et al., 2010; Yahi et al., 2015), while LBSN-based approaches are constrained by the cold-start problem and an inability to adapt to real-time needs (Majid et al., 2015; Chang et al., 2021; Ding and Chen, 2018). Even early UGC-based methods using shallow text-mining could only capture surface-level keywords, failing to grasp the deep contextual and perceptual depth that defines a travel experience (Liang et al., 2024; Missaoui et al., 2019).

UGuideRAG's success stems from its core contribution: the integration of deep semantic relevance with critical spatial feasibility. The overall experimental results confirm the power of this synthesis. In a direct comparison, UGuideRAG not only achieves a far superior semantic alignment than the ITINERA baseline—with a Hit Rate of 78.5% to 42.3% in Paris and 72.7% to 33.8% in Rome-Vatican—but also produces itineraries with significantly greater geographical coherence, evidenced by its much lower Spatial Tightness scores. This dual success is crucial, as superior semantic matching alone is insufficient for creating an effective travel plan. By confronting the inherent "aspatial" nature of many conventional recommenders, which deliver disconnected "interest islands" (Bao et al., 2012) and can exacerbate "popularity bias" (Nguyen and Tong, 2022), UGuideRAG provides a solution that is both highly relevant and realistically executable.

This human-centric design is actualized through a powerful, two-layered spatial strategy that prioritizes the quality of the travel experience. At the macro-level, spatial clustering transforms a mere list of destinations into a holistic proposal for regional exploration. This fosters the continuous narrative and immersion necessary for developing a true "sense of place," allowing the traveler to engage deeply with the urban fabric rather than simply collecting sights (Freytag, 2010b; Shoval et al., 2011; Cohen, 1979). This regional focus is then refined at the micro-level by scenic-aware routing, which optimizes for experiential quality over mere efficiency. By guiding users along paths rich in aesthetic and sensory input, the journey itself is transformed into an act of discovery, a key determinant of a positive walking experience (Mehta, 2008; Ewing and Clemente, 2005). Ultimately, this dual strategy of clustering and scenic routing creates itineraries that are not just semantically and logistically sound, but are fundamentally designed to foster an immersive and meaningful connection with the city (Gavalas et al., 2014).

6.2 Interpretation of Ablation Study

The ablation study provides a granular and insightful deconstruction of the architecture responsible for UGuideRAG's superior semantic matching. The results reveal that its success is not attributable to a single monolithic component, but rather to the powerful synergy of three critical, sequential modules: an LLM-powered engine for extracting deep, multi-faceted attraction features; an intent-enhanced retriever for high-precision recall; and an LLM-based reranker for final, nuanced contextual refinement. At the feature extraction layer, the deployment of LLMs constitutes a paradigmatic shift from the lexical-level "keyword matching" of traditional models to a more profound "conceptual alignment." While foundational methods like TF-IDF can identify salient nouns, they are fundamentally limited by their bag-of-words assumption, rendering them incapable of understanding the syntax, context, and nuance that are abundant in UGC (Jurafsky and Martin, 2023; Xiang et al., 2017). In contrast, LLMs leverage their vast pre-training to parse UGC for underlying concepts, emotions, and implied contexts, distilling unstructured collective intelligence into rich, structured profiles for each attraction. The foundational importance of this deep feature extraction is illustrated by the results: removing this component (w/o Intent Decomposition & UGC) led to a catastrophic decline in performance, causing the Hit Rate to plummet to the lowest recorded levels of 52.0% in Paris and 52.7% in Rome-Vatican, a drop of over 25 and 20 percentage points respectively from the full model.

Following feature extraction, the quality of the retrieval stage proves to be the primary determinant of the overall performance of the RAG system. This study's findings robustly confirm the "garbage in, garbage out" principle in this context:

if the retrieved context is irrelevant or poorly aligned with user intent, the downstream modules, no matter how powerful, are set up for failure (Shi et al., 2023). The ablation experiments quantitatively demonstrate this critical dependency. By disabling only the intent decomposition module (w/o Intent Decomposition), which forces the retriever to work with raw, ambiguous queries, the Hit Rate experienced a drastic decline from 78.5% to 66.9% in Paris and from 72.7% to 64.0% in Rome-Vatican. This highlights that an intent reasoning step is not a luxury but a necessity. UGuideRAG's intent decomposition module acts as a crucial pre-retrieval reasoning engine, deconstructing a vague user request into a set of clear, actionable sub-queries. This process aligns with best practices in information retrieval that aim to maximize query specificity before database interaction (Ma et al., 2023), underscoring the critical insight that the system's bottleneck often lies not in the final generation or ranking, but in the intelligence of the retriever and its ability to accurately decipher true user intent (Yu et al., 2022).

Finally, acting as a fine-grained arbiter of relevance, the LLM-based reranker provides the crucial last step of contextual refinement. This two-stage retrieval-and-ranking architecture is a well-established and highly effective paradigm in modern search and recommendation systems (Covington et al., 2016). While the retriever's job is to efficiently sift through a massive database to recall a broad set of potentially relevant candidates (optimizing for recall), the reranker performs a more computationally expensive but far more sophisticated analysis on this smaller set (optimizing for precision). It leverages the full contextual reasoning power of an LLM to perform a holistic evaluation, assessing the nuanced interplay between all facets of the user's query and the detailed profile of each candidate attraction. Unlike the geometric logic of embedding similarity, the reranker can adjudicate complex trade-offs and correctly identify "near miss" candidates that might be thematically close but experientially wrong. The importance of this stage is confirmed by the experiments, where removing the reranker (w/o LRR) caused a significant drop in Hit Rate to 66.4% in Paris and 63.4% in Rome-Vatican. This demonstrates that for the complex, multi-faceted queries typical of travel planning, a LLM-based reranking mechanism is essential for polishing the candidate set and achieving the highest degree of semantic alignment.

The analysis of the ITINERA variants further illustrates these dynamics. By replacing ITINERA's original CSO method with that of UGuideRAG, ITINERA achieves clear improvements in the semantic relevance of recommended attractions, along with reduced spatial tightness values. This indicates that UGuideRAG's spatial optimization strategy is particularly effective in attraction-dense urban settings such as Paris and Rome-Vatican, as it preserves walking feasibility while ensuring closer semantic alignment between the recommended VAs and the user query.

At the same time, the ablation results reveal that without the LRR module, UGuideRAG consistently outperforms ITINERA in retrieval accuracy, with Hit Rates of 66.4% and 63.4% in Paris and Rome-Vatican, compared to 59.0% and 56.7% for ITINERA. This demonstrates the importance of decomposing user queries into multiple dimensions and performing intent modeling, which substantially enhances retrieval precision. When the LRR module is introduced, both ITINERA (with UGuideRAG's CSO and LRR) and UGuideRAG experience significant gains in Hit Rate relative to their w/o-LRR counterparts—72.6% vs. 59.0% in Paris and 65.5% vs. 56.7% in Rome-Vatican for ITINERA, and 78.5% vs. 66.4% in Paris and 72.7% vs. 63.4% in Rome-Vatican for UGuideRAG. However, the relative performance gap between the two methods remains comparable to that observed without reranking, indicating that while the LRR module delivers a major absolute boost, UGuideRAG's core advantage originates from its intent reasoning at the retrieval stage.

Together, these results demonstrate that the effectiveness of UGuideRAG arises from the complementary contributions of all its modules. UGC-derived VA features provide rich semantic signals, intent decomposition ensures accurate retrieval, and the LLM-based reranker refines results with fine-grained contextual reasoning. Meanwhile, the CSO module achieves a balance between semantic relevance and travel burden, producing coherent and walkable itineraries. Each component is indispensable, and only their integration delivers recommendations that are both semantically aligned and practically feasible.

6.3 Implications

The findings of this study offer several key implications for the design and application of intelligent systems in the travel domain. Firstly, for the field of Recommender Systems, this work signals the necessity of evolving from recommending discrete items to architecting complete, holistic experiences. The results demonstrate that a model's success in a complex domain like tourism depends on its ability to synthesize semantic relevance with spatial coherence and experiential quality. This suggests a new design paradigm where logistical and experiential factors are treated as integral components of the core optimization process, not as secondary filters. This provides a clear path for moving beyond the limitations of "aspatial" and popularity-biased models towards systems that deliver genuinely practical and enriching "experience packages."

Secondly, for the field of applied NLP and RAG systems, this study has a critical methodological implication: the primacy of intent modeling. The detailed analysis of the system's components reveals that while every module is important, the effectiveness of powerful downstream components, such as the LLM-based reranker, is fundamentally capped by the quality of the initial retrieval. The most significant performance gains were unlocked by the intent-enhanced retriever. This implies that future research and development in RAG-based recommenders should place a emphasis on robust query rewriting and decomposition techniques, as this is the foundational stage that enables the full potential of the entire architecture.

Finally, the capabilities demonstrated by UGuideRAG have direct implications for the tourism industry. The ability to process ambiguous, natural-language queries and match them with deep, experiential features extracted from UGC enables a new level of personalization. This allows travel platforms and destination marketing organizations to cater to the long tail of niche travel interests, moving beyond generic suggestions. By generating practical, walkable, and scenic itineraries, this technology provides a powerful tool to directly enhance the quality of the on-the-ground travel experience. It offers a new modality for promoting destinations, not as static lists of attractions, but as dynamic, interconnected experiences that can be customized to each traveler's unique desires.

6.4 Limitations

Despite its promising results, this study is subject to several limitations that warrant careful consideration and frame the agenda for future research. The system's performance is intrinsically and heavily tied to the capabilities and potential flaws of the underlying LLM. This dependency introduces a series of risks. First, the issue of factual hallucination is particularly pernicious in a travel context, as a fabricated detail can directly lead to a negative real-world user experience and damage trust (Yao et al., 2023). Second, the "black box" nature of these models presents a significant challenge for interpretability, a field where Explainable AI is becoming increasingly critical for user adoption (Zhang and Chen, 2020). Third, LLMs are trained on vast internet corpora that often reflect and amplify existing societal biases, which could lead to the underrepresentation of non-dominant cultures or viewpoints (Bender et al., 2021).

Beyond the model-centric limitations, the data pipeline of this study introduces a significant ethical dimension concerning the use of public UGC. While user reviews from platforms like Google Maps are publicly accessible, they are not explicitly anonymized and can contain personally identifiable information, whether directly (e.g., usernames) or indirectly through the content of the review itself. A critical ethical question arises regarding consent. Users who post reviews consent to a platform's terms of service, which typically involves sharing their opinion with other travelers on that platform. It is highly questionable whether this implicit consent extends to their data being systematically scraped, aggregated, and used as input for a third-party LLM for an entirely different purpose, a practice known as secondary data use (Zimmer, 2010). This lack of explicit consent for a new processing purpose presents a notable ethical and legal challenge, particularly in jurisdictions with strong data protection laws like the GDPR[18]. Furthermore, the process of feeding this data into a commercial black-box LLM API raises additional concerns about data privacy, ownership, and security, as the ultimate storage and potential secondary use of this data by the LLM provider are often opaque.

Another major constraint is that the current framework operates on a static representation of the urban environment. It does not yet incorporate the dynamic, real-time variables that are critical for robust, real-world itinerary planning. Key factors such as the typical duration of stay at each attraction, current and forecasted weather conditions, and seasonal changes—like reduced daylight in winter or holiday crowds—are not considered. Furthermore, the model omits other crucial logistics, such as attraction opening hours or public transit delays (Chen et al., 2017). This gap between a theoretically optimal plan and a practically executable one can lead to user frustration.

6.5 Future Work

Building on the foundation of this research, several promising and crucial avenues for future work emerge to address the aforementioned limitations and advance the field of intelligent travel planning. A primary objective should be to enhance the system's robustness and trustworthiness. To counter LLM hallucinations and static data limitations, future iterations should integrate the RAG framework with verified, dynamic knowledge sources. This could involve cross-referencing generated outputs with structured knowledge graphs or real-time APIs (e.g., Google Maps API for live opening hours and transit data), creating a fact-checking layer within the recommendation pipeline.

A second critical direction is to evolve the system towards a truly dynamic, context-

[18] https://gdpr-info.eu/

aware recommendation engine. This involves moving beyond static planning to a model that can ingest real-time data streams from various sources (weather, traffic, social media events). By framing the task as a dynamic, multi-objective optimization problem, such a system could offer truly adaptive and resilient recommendations, capable of re-planning an itinerary on the fly in response to unforeseen events, like a sudden rainstorm or a user's change of mood (Adomavicius and Tuzhilin, 2011).

Another key direction is the incorporation of multimodal feature extraction. User preferences and destination attributes are conveyed powerfully through visual media. Future systems should leverage joint text-image embedding models (e.g., CLIP) to analyze photos and videos from UGC, deriving richer features such as a location's "scenic beauty," "architectural style," or "vibrancy." This would enable more holistic understanding and novel interaction modalities, such as visual query systems ("find me places that look like this") (Zhang et al., 2019).

Furthermore, a key direction for practical improvement is expanding the system's planning horizon from a single-day itinerary to multi-day itineraries. This expansion introduces significant new challenges, including the need to optimize routes over multiple days, incorporate accommodation planning, model user fatigue, and maintain thematic consistency. Therefore, an advanced system would need to intelligently sequence attractions, create a balanced schedule of activities and rest, and operate effectively over a longer planning timeframe.

Finally, future research must prioritize the development of responsible, fair, and transparent recommendation algorithms. To mitigate overtourism and popularity bias, this involves designing systems that explicitly incorporate metrics for fairness and diversity into the optimization process. Techniques such as re-ranking for fairness, where an initial relevance-based list is adjusted to boost the visibility of high-quality but less-popular options, can be employed (Celis et al., 2017). This could evolve the system into a tool for sustainable tourism management, where it collaborates with city planners to help distribute tourist flow more equitably. This vision of a responsible recommender system, combined with advancements in dynamic adaptation and interpretability, charts a course toward the next generation of intelligent guides: systems that are not just more accurate, but are also more trustworthy, responsive, and conscientious partners in our exploration of the urban world.

7 Conclusion

In an era where urban tourism is increasingly shaped by the demand for authentic, personalized, and immersive experiences, conventional recommendation systems have struggled to keep pace. Most remain locked in the paradigm of item-based suggestions, offering users little more than static lists of attractions. Yet today's digital traveler expects far more: not only a catalog of places to visit, but a coherent and meaningful narrative—an itinerary that resonates with their unique interests and fosters a deeper connection with the cultural and spatial fabric of the city. The central challenge, and the motivation for this thesis, lies in bridging the gap between the nuanced, expressive language of human travel desires and the rigid, categorical logic that dominates existing recommendation engines.

This thesis addresses that gap by proposing and validating UGuideRAG, a framework that shifts the focus from isolated point recommendations to holistic itinerary design. By linking user narratives with spatial and experiential constraints, UGuideRAG demonstrates that it is possible to generate itineraries that are aligned with user intent, spatially coherent, and experientially rich. The approach shows that true personalization requires drawing upon the collective intelligence embedded in user-generated content, where diverse experiences provide signals beyond what curated databases can offer. Advanced RAG techniques serve as the bridge, translating unstructured expressions of travel desires into structured, actionable plans faithful to both user preferences and urban realities.

Ultimately, the contribution of this thesis extends beyond a new algorithm. It presents a blueprint for intelligent travel guides that act as companions in urban exploration. The principles established here, emphasizing intent decomposition, spatial coherence, and experiential quality, form a foundation for systems that are more accurate, adaptive, and responsive to travelers' needs. Future work may integrate dynamic, real-time data and multimodal signals to create even more context-aware systems. What this thesis demonstrates is a clear shift: technology can move beyond presenting information to curating a dialogue between traveler and city, enriching the journey and deepening our connection to the places we visit.

References

Zahra Abbasi-Moud, Hamed Vahdat-Nejad, and Javad Sadri. 2021. Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Systems with Applications*, 167:114324.

Gediminas Adomavicius and Alexander Tuzhilin. 2011. Context-aware recommender systems. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, *Recommender Systems Handbook*, pages 217–253. Springer US.

Gary Akehurst. 2009. User generated content: the use of blogs for tourism organisations and tourism consumers. *Service business*, 3(1):51–61.

Konstantinos Alexandris, Constantinos Kouthouris, and Alexandros Meligdis. 2006. Increasing customers' loyalty in a skiing resort: The contribution of place attachment and service quality. *International Journal of Contemporary Hospitality Management*, 18(5):414–425.

Maria-Irina Ana and Laura-Gabriela Istudor. 2019. The role of social media and user-generated-content in millennials travel behavior. *Management dynamics in the knowledge economy*, 7(1/23):87–104.

Salvador Anton Clavé. 2016. Tourism and urban walkability: An opportunity to rethink destination planning and management. Conference presentation, Tourism Naturally Conference. Accessed November 20, 2016.

Salvador Anton Clavé. 2018. Urban tourism and walkability. In *The future of tourism: Innovation and sustainability*, pages 195–211. Springer.

Nuno Antonio, Marisol B Correia, and Filipa Perdigão Ribeiro. 2020. Exploring user-generated content for improving destination knowledge: The case of two world heritage cities. *Sustainability*, 12(22):9654.

Jie Bao, Yu Zheng, and Mohamed F. Mokbel. 2012. Location-based and preference-aware recommendation using sparse geo-social networking data. In

Proceedings of the 20th International Conference on Advances in Geographic Information Systems (GIS '12), pages 199–208. ACM.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pages 610–623. Association for Computing Machinery.

Walter Benjamin. 2006. *The Writer of Modern Life: Essays on Charles Baudelaire*. Harvard University Press, Cambridge, MA.

I. Benouaret and D. Lenne. 2016. A composite recommendation system for planning tourist visits. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 626–631.

Anja Hälg Bieri. 2017. Walking in the capitalist city: On the socio-economic origins of walkable urbanism. In *The Routledge international handbook of walking*, pages 27–36. Routledge.

Bruce Booth. 1998. Information for visitors to cultural attractions. *Journal of Information Science*, 24(5):291–303.

Gordon Burtch, Qinglai He, Yili Hong, and Dokyun Lee. 2022. How do peer awards motivate creative content? experimental evidence from reddit. *Management Science*, 68(5):3488–3506.

L. Elisa Celis, Anay Mehrotra, and Nisheeth K. Vishnoi. 2017. Ranking with fairness constraints. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *ICML'17*, pages 703–711. PMLR.

L. Chang, W. Chen, J. Huang, C. Bin, and W. Wang. 2021. Exploiting multi-attention network with contextual influence for point-of-interest recommendation. *Applied Intelligence*, 51(4):1904–1917.

Mi Chang, Gi-bbeum Lee, and Ji-Hyun Lee. 2023. Analysis of urban visitor walkability based on mobile data: The case of daejeon, korea. *Cities*, 143:104564.

Aili Chen, Xuyang Ge, Ziquan Fu, Yanghua Xiao, and Jiangjie Chen. 2024. Travelagent: An ai assistant for personalized travel planning. *arXiv preprint arXiv:2409.08069*.

C. Chen, X. Chen, Z. Wang, Y. Wang, and D. Zhang. 2017. Scenicplanner: Planning scenic travel routes leveraging heterogeneous user-generated digital footprints. *Frontiers of Computer Science*, 11(1):61–74.

Erik Cohen. 1979. A phenomenology of tourist experiences. *Sociology*, 13(2):179–201.

Comité Régional du Tourisme Paris Île-de-France. 2023. Bilan de la fréquentation touristique en 2022 et perspectives pour 2023.

Comune di Roma. 2024. Turismo, nel 2023 numeri record a roma: per la prima volta superate le 35 milioni di presenze ufficiali.

Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*, pages 191–198.

Carmen Cox, Stephen Burgess, Carmine Sellitto, and Jeremy Buultjens. 2009. The role of user-generated content in tourists' travel planning behavior. *Journal of hospitality marketing & management*, 18(8):743–764.

Destination Analysts. 2019. The state of the american traveler.

Daniele Di Palma. 2023. Retrieval-augmented recommender system: Enhancing recommender systems with large language models. *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 1369–1373.

R. Ding and Z. Chen. 2018. Recnet: A deep neural network for personalized poi recommendation in location-based social networks. *International Journal of Geographical Information Science*, 32(8):1631–1648.

Zong Du, Tianyu Zhang, and Zihan Zhang. 2023. MoDS: The Mirage of Data Sophistication for Instruction Tuning. *arXiv preprint arXiv:2310.06037*.

David W Eby and Lisa J Molnar. 2002. Importance of scenic byways in route choice: a survey of driving tourists in the united states. *Transportation Research Part A: Policy and Practice*, 36(2):95–106.

Birgit Elands and Jaap Lengkeek. 2000. Typical tourists. *Research into the theoretical and methodological foundations of a typology of tourism and recreation experiences. Mansholt Studies*, 21.

European Commission, Directorate-General for Enterprise and Industry and TNS Political & Social. 2014. Preferences of europeans towards tourism: Final report.

Reid Ewing and Otto Clemente. 2005. Streetscape features related to pedestrian activity. Technical Report E-C082, Transportation Research Board.

Lisa S. Faerber, Johanna Hofmann, Dennis Ahrholdt, and Oliver Schnittka. 2021. When are visitors actually satisfied at visitor attractions? what we know from more than 30 years of research. *Tourism Management*, 84:104284.

Wenqi Fan. 2024. Recommender systems in the era of large language models (llms). *IEEE Transactions on Knowledge and Data Engineering*, pages 1–20.

Tim Freytag. 2010a. Déjà-vu: tourist practices of repeat visitors in the city of paris. *Social Geography*, 5(1):49.

Tim Freytag. 2010b. Visitor activities and inner-city tourist mobility: The case of heidelberg. In *Analysing international city tourism*, pages 213–226. Springer.

Damianos Gavalas, Vlasios Kasapakis, Charalampos Konstantopoulos, Grammati Pantziou, and Nikolaos Vathis. 2017. Scenic route planning for tourists. *Personal and Ubiquitous Computing*, 21(1):137–155.

Damianos Gavalas, Charalampos Konstantopoulos, Konstantinos Mastakas, and Grammati Pantziou. 2014. A survey on algorithmic aspects of trip planning. *Journal of Heuristics*, 20(3):245–285.

Gary L. Geissler and Christina T. Rucks. 2011. The overall theme park experience: A visitor satisfaction tracking study. *Journal of Vacation Marketing*, 17(2):127–138.

Alice Germano. 2023. Citywalk: Embracing urban charms and captivating generation z. https://daxueconsulting.com/citywalk/. Daxue Consulting - Market Research and Consulting China.

Ulrike Gretzel and Kye-Hyoung Yoo. 2008. Use and impact of online travel reviews. In *Information and communication technologies in tourism 2008*, pages 35–46. Springer Vienna.

Samantha Hajna, Nancy A Ross, Anne-Sophie Brazeau, Patrick Bélisle, Lawrence Joseph, and Kaberi Dasgupta. 2015. Associations between neighbourhood walkability and daily steps in adults: a systematic review and meta-analysis. *BMC public health*, 15(1):768.

Mark B. Houston, Lance A. Bettencourt, and Steven Wenger. 1998. The relationship between waiting in a service queue and evaluations of service quality: A field theory perspective. *Psychology and Marketing*, 15(8):735–753.

Jiani Huang, Shijie Wang, Liang-bo Ning, Wenqi Fan, Shuaiqiang Wang, Dawei Yin, and Qing Li. 2025. Towards next-generation recommender systems: A benchmark for personalized recommendation assistant with llms. *arXiv preprint arXiv:2503.09382*.

Dan Jurafsky and James H. Martin. 2023. *Speech and Language Processing*, 3rd edition. Prentice Hall.

Kimiya Keyvan and Jimmy Xiangji Huang. 2022. How to approach ambiguous queries in conversational search: A survey of techniques, approaches, tools, and challenges. *ACM Computing Surveys*, 55(6):1–40.

Catheryn Khoo-Lattimore and Erdogan H. Ekiz. 2014. Power in praise: Exploring online compliments on luxury hotels in malaysia. *Tourism and Hospitality Research*, 14(3):152–159.

David B. Klenosky. 2002. The "pull" of tourism destinations: A means-end investigation. *Journal of Travel Research*, 40(4):396–403.

I. Kostric, K. Balog, and F. Radlinski. 2024. Generating usage-related questions for preference elicitation in conversational recommender systems. *ACM Transactions on Recommender Systems*. Forthcoming.

Constantinos Kouthouris and Konstantinos Alexandris. 2005. Can service quality predict customer satisfaction and behavioral intentions in the sport tourism industry? an application of the servqual model in an outdoors setting. *Journal of Sport & Tourism*, 10(2):101–111.

Ana Luiza Favarão Leão and Mariana Ragassi Urbano. 2020. Street connectivity and walking: An empirical study in londrina-pr. *Semina: Ciências Exatas e Tecnológicas*, 41(1):31–42.

Anna Leask. 2010. Progress in visitor attraction research: Towards more effective management. *Tourism Management*, 31(2):155–166.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Ott, Wen-tau Chen, Alexis Conneau, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Xiaoye Li, Jennifer Francois, and Junehee Kwon. 2023. Investigating consumers' online restaurant selection behaviors using eye-tracking technology and retrospective think-aloud interviews. *International Journal of Hospitality & Tourism Administration*, 24(5):720–752.

Kaibo Liang, Huwei Liu, Man Shan, Junhui Zhao, Xiaolan Li, and Li Zhou. 2024. Enhancing scenic recommendation and tour route personalization in tourism using ugc text mining. *Applied Intelligence*, 54(1):1063–1098.

Weng Marc Lim and Tareq Rasul. 2022. Customer engagement and social media: Revisiting the past to inform the future. *Journal of Business Research*, 148:325–342.

Ada S Lo and Candy YS Lee. 2011. Motivations and perceived value of volunteer tourists from hong kong. *Tourism management*, 32(2):326–334.

Xin Lu, Changhu Wang, Jiang-Ming Yang, Yanwei Pang, and Lei Zhang. 2010. Photo2trip: generating travel routes from geo-tagged photos for trip planning. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 143–152.

Yuxia Lu, Juntao Bao, Yimu Song, Zhaohua Ma, Shanshan Cui, Yiming Wu, and Xiangnan He. 2021. Revcore: Review-augmented conversational recommendation. *arXiv preprint arXiv:2106.00957*.

Xinbei Ma, Yeyun Gong, Pengcheng He, Hai Zhao, and Nan Duan. 2023. Query rewriting in retrieval-augmented large language models. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5303–5315.

A. Majid, L. Chen, H. T. Mirza, I. Hussain, and G. Chen. 2015. A system for mining interesting tourist locations and travel sequences from public geo-tagged photos. *Data & Knowledge Engineering*, 95:66–86.

Athena HN Mak. 2017. Online destination image: Comparing national tourism organisation's and tourists' perspectives. *Tourism management*, 60:280–297.

Estela Marine-Roig and Salvador Anton Clavé. 2015. Tourism analytics with massive user-generated content: A case study of barcelona. *Journal of destination marketing & management*, 4(3):162–172.

Kurt Matzler, Birgit Renzl, and Rita Faullant. 2007. Dimensions of price satisfaction: a replication and extension. *International Journal of Bank Marketing*, 25(6):394–405.

Bob McKercher and Hilary Du Cros. 2003. Testing a cultural tourism typology. *International journal of tourism research*, 5(1):45–58.

Kevin Meehan, Tom Lunney, Kevin Curran, and Aiden McCaughey. 2013. Context-aware intelligent recommendation system for tourism. In *2013 IEEE international conference on pervasive computing and communications workshops (PERCOM workshops)*, pages 328–331. IEEE.

Vinit Mehta. 2008. Walkable streets: pedestrian behavior, perceptions and attitudes. *Urban Design International*, 13:145–167.

Vasileios Milias, Shahin Sharifi Noorian, Alessandro Bozzon, and Achilleas Psyllidis. 2023. Is it safe to be attractive? disentangling the influence of streetscape features on the perceived safety and attractiveness of city streets. *AGILE: GIScience Series*, 4:8.

Abhishek Mishra and Ansh Gupta. 2019. Green hotel servicescape: Attributes and unique experiences. *Current Issues in Tourism*, 22(20):2566–2578.

Sondess Missaoui, Faten Kassem, Marco Viviani, Alessandra Agostini, Rim Faiz, and Gabriella Pasi. 2019. Looker: a mobile, personalized recommender system in the tourism domain based on social media user-generated content. *Personal and Ubiquitous Computing*, 23:181–197.

Gianna M. Moscardo and Philip L. Pearce. 1986. The concept of authenticity in tourist experiences. *The Australian and New Zealand Journal of Sociology*, 22(1):121–132.

Ana María Munar and Jens Kr Steen Jacobsen. 2013. Trust and involvement in tourism social media and web-based travel information sources. *scandinavian Journal of Hospitality and Tourism*, 13(1):1–19.

Thao Thanh Thi Nguyen and Shurong Tong. 2022. The impact of user-generated content on intention to select a travel destination. *Journal of Marketing Analytics*, page 1.

Eleonora Pantano, Constantinos-Vasilios Priporas, Nikolaos Stylos, and Charles Dennis. 2019. Facilitating tourists' decision making through open data analyses: A novel recommender system. *Tourism Management Perspectives*, 31:323–331.

Philip L. Pearce. 1991. Analysing tourist attractions. *Journal of Tourism Studies*, 2(1):46–55.

Xia Peng and Zhou Huang. 2017. A novel popular tourist attraction discovering approach based on geo-tagged social media big data. *ISPRS International Journal of Geo-Information*, 6(7):216.

Svein H.G. Poulsson and Sudhir H. Kale. 2004. The experience economy and commercial experiences. *The Marketing Review*, 4(3):267–277.

Daniele Quercia, Rossano Schifanella, Luca Maria Aiello, and Kate McLean. 2015. Smelly maps: the digital life of urban smellscapes. In *Proceedings of the International AAAI conference on Web and Social Media*, volume 9, pages 327–336.

H. A. Rahmani, M. Aliannejadi, M. Baratchi, and F. Crestani. 2020. Joint geographical and temporal modeling based on matrix factorization for point-of-interest recommendation. In J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, and F. Martins, editors, *Advances in Information Retrieval*, volume 12035 of *Lecture Notes in Computer Science*, pages 205–219. Springer International Publishing.

Nina Runge, Pavel Samsonov, Donald Degraen, and Johannes Schöning. 2016. No more autobahn! scenic route generation using googles street view. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pages 147–151.

Amihai Savir, Ronen Brafman, and Guy Shani. 2013. Recommending improved configurations for complex objects with an application in travel planning. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 391–394.

Weijia Shi, Wen Chen, Yixiao Yao, Rui Zhang, and Dong Yu. 2023. Replug: Retrieval-augmented black-box language models. *Preprint*, arXiv:2301.12652.

Yuxiang Shi, Xiao Zi, Zhaoyu Shi, Hongbo Zhang, Qijiong Wu, and Meng Xu. 2024. Eragent: Enhancing retrieval-augmented language models with improved accuracy, efficiency, and personalization. *arXiv preprint arXiv:2405.06683*.

Noam Shoval, Bob McKercher, Erica Ng, and Amit Birenboim. 2011. Hotel location and tourist activity in cities. *Annals of tourism research*, 38(4):1594–1612.

Yihong Tang, Zhaokai Wang, Ao Qu, Yihao Yan, Zhaofeng Wu, Dingyi Zhuang, Jushi Kai, Kebing Hou, Xiaotong Guo, Han Zheng, et al. 2024. Itinera: Integrating spatial optimization with large language models for open-domain urban itinerary planning. *arXiv preprint arXiv:2402.07204*.

Theano S. Terkenli. 2021. Research advances in tourism-landscape interrelations: An editorial. *Land*, 10(9):944.

Keith Tester, editor. 1994. *The Flâneur*. Routledge, London.

Stephen Tomas, David Scott, and John L. Crompton. 2002. An investigation of the relationships between quality of service performance, benefits sought, satisfaction and future intention to visit among visitors to a zoo. *Managing Leisure*, 7(4):239–250.

Norsidah Ujang and Khalilah Zakariya. 2015. Place attachment and the value of place in the life of the users. *Procedia-Social and Behavioral Sciences*, 168:373–380.

Andrés Villaveces, Luis Alfonso Nieto, Delia Ortega, José Fernando Ríos, John Jairo Medina, María Isabel Gutiérrez, and Daniel Rodríguez. 2012. Pedestrians' perceptions of walkability and safety in relation to the built environment in cali, colombia, 2009–10. *Injury prevention*, 18(5):291–297.

VisitBritain. 2019. Researching and planning.

Zhenbin Wang, Hui Zhang, Sridar Ramachandran, and Suiying Cheng. 2025. Why are individuals tracing travel trends? a case study of city walk in malaysia. *PloS one*, 20(2):e0309493.

World Bank. 2018. *The World Bank Annual Report 2018*. World Bank, Washington, DC.

Dongdong Wu. 2024. City walk in a gap day: potential and opportunities for tourism and leisure. *Tourism Review*, 79(9):1576–1581.

Zheng Xiang, Qiang Du, Yong Ma, and Weng Fan. 2017. A comparative analysis of major online review platforms: Implications for hotel managers and researchers. *International Journal of Hospitality Management*, 58:51–65.

Zheng Xiang and Ulrike Gretzel. 2010. Role of social media in online travel information search. *Tourism management*, 31(2):179–188.

Alexandre Yahi, Antoine Chassang, Louis Raynaud, Hugo Duthil, and Duen Horng Chau. 2015. Aurigo: an interactive tour planner for personalized itineraries. In *Proceedings of the 20th international conference on intelligent user interfaces*, pages 275–285.

Xinran Yang, Liaoniao Zhang, and Zixin Feng. 2024. Personalized tourism recommendations and the e-tourism user experience. *Journal of Travel Research*, 63(5):1183–1200.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. *International Conference on Learning Representations (ICLR).*

Dongjie Yu, Ruocheng Bao, Gengchen Mai, and Ling Zhao. 2025. Spatial-rag: Spatial retrieval augmented generation for real-world spatial reasoning questions. *arXiv preprint arXiv:2502.18470.*

Wenhao Yu, Mohit Ghassemi, Zhiting Dai, and Scott Yih. 2022. Generate rather than retrieve: Large language models are strong context generators. *Preprint*, arXiv:2209.10063.

Kun Zhang, Ye Chen, and Chunlin Li. 2019. Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of beijing. *Tourism Management*, 75:595–608.

Yifan Zhang, Shiyu Qiao, Jiazhen Zhang, Tzu-Heng Lin, Chen Gao, and Yihong Li. 2025. A survey of large language model empowered agents for recommendation and search: Towards next-generation information retrieval. *arXiv preprint arXiv:2503.05659.*

Yongfeng Zhang and Xu Chen. 2020. Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval*, 14(1):1–101.

Yan-Tao Zheng, Shuicheng Yan, Zheng-Jun Zha, Yiqun Li, Xiangdong Zhou, Tat-Seng Chua, and Ramesh Jain. 2013. Gpsview: A scenic driving route planner. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 9(1):1–18.

Michael Zimmer. 2010. "but the data is already public": on the ethics of research in facebook. *Ethics and Information Technology*, 12(4):313–325.

8 Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.
For the purposes of language enhancement, translation, and proofreading, I employed tools including ChatGPT, Gemini, and Grammarly. I assume complete responsibility for the intellectual content, arguments, and final text of this work.

Jing Tang

Jing Tang, 29.08.2025