

# Where shall we meet? Group meeting location recommendation via preference-mobility reasoning with MEET-LLM



Shuang Yang<sup>1,4</sup>, Lening Wang<sup>1,2</sup>, Zhiyong Cui<sup>1,2</sup>, Yilong Ren<sup>1,2,3,4</sup>, Liang Xu<sup>1,4</sup>, Mohamed Abouelela<sup>5</sup>, Haiyang Yu<sup>1,2,3,4</sup>, and Aoyong Li<sup>1,2,4</sup>✉

<sup>1</sup> School of Transportation Science and Engineering, Beihang University, Beijing 100191, China

<sup>2</sup> State Key Laboratory of Intelligent Transportation System, Beihang University, Beijing 100191, China

<sup>3</sup> Zhongguancun Laboratory, Beijing 100194, China

<sup>4</sup> Hangzhou Innovation Institute, Beihang University, Hangzhou 311115, China

<sup>5</sup> School of Transportation System Engineering, Technical University of Munich, Munich 80333, Germany

Received: 8 January 2026

Revised: 4 February 2026

Accepted: 10 February 2026

## KEYWORDS

Chain-of-Thought Reasoning, Group Meeting Location Recommendation (GMLR), Intelligent Transportation Systems (ITS), Large Language Models (LLMs)

## ABSTRACT

Large Language Models (LLMs) are propelling Intelligent Transportation Systems (ITS) toward an agentic paradigm, enabling intent understanding, constraint-aware reasoning, and decision-making. This paper focuses on the Group Meeting Location Recommendation (GMLR) problem, which aims to jointly optimize group preferences and mobility costs under real-world transportation network constraints. However, existing methods often overlook real-world geographic context, fail to jointly coordinate group preferences and spatial feasibility, and rely heavily on static, structured data to model group preferences, limiting their applicability to complex and dynamic group scenarios. We propose MEET-LLM, an agentic, LLM-driven, closed-loop framework for GMLR. MEET-LLM comprises three modules: (1) a Preference Profiling Module that constructs structured semantic representations of users and merchants by using LLMs to extract multilevel preferences from natural language reviews, overcoming the limitations of static aggregation; (2) a Spatial Optimization Module that introduces a mobility cost (MC) metric to evaluate group-level travel convenience and account for user-specific travel constraints; and (3) a Chain-of-Thought (CoT) Reasoning Module that coordinates group preferences and mobility costs through multi-turn LLM reasoning to generate interpretable, executable recommendations with map-based navigation guidance. MEET-

Address correspondence to Aoyong Li, [liao Yong@buaa.edu.cn](mailto:liao Yong@buaa.edu.cn)

© The author(s) 2026. The articles published in this open-access journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

LLM jointly coordinates group preferences and spatial feasibility, bridging natural language reasoning with real-world mobility execution. It enables dynamic, structured, and action-ready recommendations in real urban environments, highlighting the potential of LLMs as the cognitive core for agentic mobility coordination in ITS.

## 1 Introduction

Group activities in cities, such as dining, coworking, and family gatherings, often require multiple travelers to converge at a common location. Group Meeting Location Recommendation (GMLR) thus arises as a decision-support problem in Intelligent Transportation Systems (ITS) and urban mobility services, where the system must recommend locations that are simultaneously preference-compatible and transportation-feasible under real-world road-network constraints [1]. Unlike individual recommendation, GMLR involves travelers departing from different origins with heterogeneous preferences and asymmetric accessibility across members, and therefore, the system must mediate collective decisions under unequal mobility burdens. GMLR, therefore, requires coordinating group satisfaction with network-constrained reachability, making it a representative multi-user, multi-constraint mobility coordination task.

Recent advances in Large Language Models (LLMs) provide a natural foundation for such coordination problems. From an ITS perspective, LLMs can serve as a cognitive layer for agentic decision-support, transforming unstructured intent into goal-directed decisions while reasoning over real-world constraints. At the same time, their strong natural language understanding enables preference modeling to move beyond structured ratings toward rich textual signals. Prior work has incorporated LLMs into recommendation frameworks [2, 3], including review understanding, interest inference, and interactive generation [4, 5], and has demonstrated promising results in domains such as movies and music [6].

However, most LLM-based recommender systems remain centered on single-user settings and primarily operate in the semantic space of retrieval and ranking [7, 8]. Even when extended to group recommendation, many approaches still rely on static ratings

or aggregated tags, which limits their ability to extract nuanced individual intents from natural language and to support dynamic preference negotiation in multi-user contexts [9].

These limitations are particularly problematic for GMLR, where the system must go beyond semantic relevance and jointly reconcile preference compatibility with spatial feasibility under real-world road-network constraints. Yet existing GMLR methods rarely integrate preference reasoning with transportation feasibility, and their outputs often remain ranked lists rather than executable mobility decisions. Moreover, even the few studies that incorporate spatial factors typically depend on static ratings or predefined features, making it difficult to capture fine-grained, text-derived preference signals and to deliver personalized recommendations in dynamic group scenarios [10, 11]. This gap motivates a framework that leverages natural language signals for intent and preference modeling while explicitly enforcing transportation feasibility in real urban environments.

To address these challenges, we propose MEET-LLM, a framework for GMLR that serves as decision support in ITS by jointly integrating dynamic group preferences with spatial feasibility to generate executable meeting location recommendations. Centered on a perception, decision, action, and feedback loop, it comprises three key modules:

*Preference Profiling Module (PPM)*: This module leverages LLMs to extract multi-level user preferences and location attributes from natural language reviews, capturing dynamic, diverse, and latent preference signals that are difficult to represent with static, structured aggregation.

*Spatial Optimization Module (SOM)*: This module introduces a mobility cost metric to filter candidate locations based on member-level mobility costs, so that the selected meeting location remains spatially feasible and convenient for all group members in the urban

travel environment.

*Chain-of-Thought (CoT) Reasoning Module (CRM)*: This module performs multi-turn language reasoning to jointly reconcile group preferences and mobility costs, and produces structured outputs including a recommended location, a reasoning explanation, and a route-feasible travel plan with navigation guidance, supporting integration with urban mobility services (Figure 1).

Our contributions are summarized as follows:

- **System-level framework**: We propose MEET-LLM as a framework for GMLR that leverages LLM-based reasoning to integrate preference modeling with routing-based spatial feasibility, enabling meeting location recommendations that are executable in real-world urban travel contexts.
- **Framework workflow**: We design a three-stage workflow for MEET-LLM that builds user and merchant profiles from reviews, applies an initial spatial screening to the candidate set based on geographic locations, and uses CoT-based group reasoning to output structured meeting recommendations with

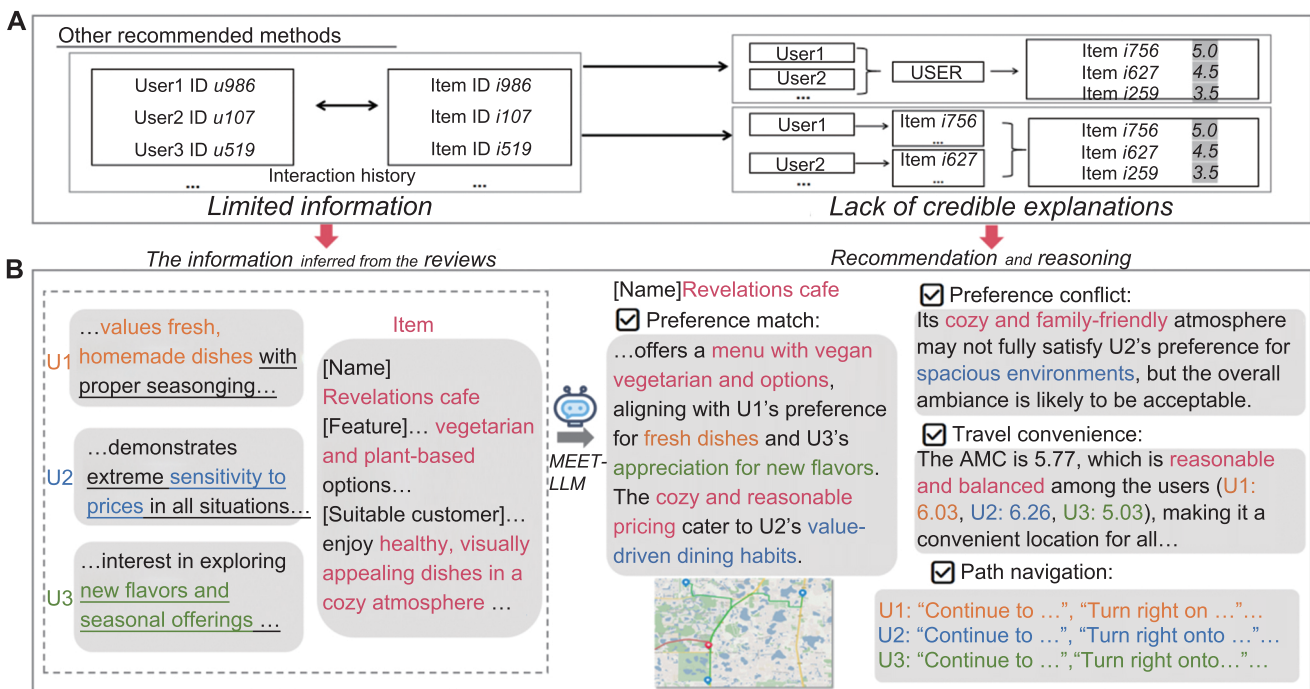
route-feasible guidance, supporting practical deployment in urban travel scenarios.

- **Empirical validation**: We evaluate MEET-LLM on the public Yelp dataset through baseline comparisons, cross-city validation, and module-level ablations with contribution analysis. The results confirm the feasibility of the MEET-LLM framework for the GMLR task.

## 2 Related Work

### 2.1 Group recommendation with geographic factors

Group recommender systems have long relied on preference aggregation strategies to support collective decision-making. Two main strategies exist: score aggregation, which merges individual recommendation results or ranked lists into a unified group list, and preference aggregation, which integrates members' preference information to construct a single group profile and performs recommendations for this virtual user [12, 13, 14]. Building on these aggregation



**Figure 1** Comparison between traditional group recommendation methods and MEET-LLM. (A) Traditional methods typically output recommended meeting locations based on user-item ratings, with limited explanation support. (B) MEET-LLM outputs a ranked set of candidate meeting locations, illustrated here with one example from the recommendation set, together with reasoning explanations on user preferences, preference conflicts, and travel convenience, as well as navigation paths

paradigms, recent research has increasingly adopted deep learning models to better capture complex preference signals and user-item relations in group settings. Specifically, prior work introduces deep learning techniques such as Graph Neural Networks (GNNs), attention mechanisms, and knowledge graphs to model complex user-item interactions [15, 16, 17]. However, these methods often overlook spatial mobility constraints, limiting their applicability to real-world location-based scenarios.

To address the lack of spatial factors in group recommendation, several studies have attempted to incorporate geographic modeling into group recommendation tasks. The context-based geoprocessing framework [18] searches meeting locations on road networks that minimize total group travel cost by combining path planning with heuristic search. Similarly, the optimal sequenced route for group meetup (OSR-G) model [19] applies origin-destination (OD) path optimization to reduce the overall travel distance of group members. While these approaches make progress in geographic optimization and are effective in spatial terms, they generally lack modeling support for user preference signals.

To account for individual differences, the method in [11] incorporates MBTI (a widely used but high-level personality typology) by mapping users into type-based categories and applying spatial clustering to improve fairness in meeting location selection. Beyond personality descriptors, other studies adopt topic-based user modeling approaches such as P-LAG [1]. While these designs introduce some degree of personalization, they rely on static or coarse-grained representations and therefore struggle to capture fine-grained, context-aware preference variations, which limits their adaptability to diverse recommendation needs.

Overall, existing studies suggest that effective GMLR requires jointly considering mobility costs, user preferences, and location attributes to produce recommendations that are both preference-consistent and spatially feasible. To address these issues, we quantify travel costs and jointly model user preferences and merchant attributes, enabling recommendations that balance spatial feasibility, personal preferences, and

group satisfaction.

Unlike traditional Point of Interest (POI) recommendation models focused on individual trajectory prediction [20, 21], as well as recent group POI recommendation models [22, 23, 24], our method targets group decision-making for meeting locations under spatial constraints while aligning collective preferences, making it well-suited for scenarios such as group meetups or collaborative meeting location planning.

## 2.2 LLMs for group recommendation

Recent research has increasingly integrated LLMs into recommender systems, leveraging their strong text understanding and generation capabilities to reformulate recommendation tasks as language modeling problems. Representative work, such as the P5 framework transforms user-item interactions and review texts into unified prompts, enabling recommendation tasks through language modeling [25]. Building on the semantics of reviews, RDRec and HDRec further enhance preference learning by incorporating review content and rationale distillation, enabling LLMs to capture deeper user intent beyond sparse interaction logs [26, 27].

To improve generalization ability, studies such as Zero-Shot and GenRec demonstrate the strong adaptability of LLMs in zero-shot and generative recommendation scenarios [28, 29]. FineRec further proposes extracting attribute-opinion pairs from reviews to construct fine-grained preference graphs for enhanced structural explainability [30], while POD [31] and ReviewRec [32] improve LLM efficiency and personalization through prompt compression and review-driven mechanisms.

Despite these advances in individual recommendation, the application of LLMs to group recommendation remains in its early stages. Some studies attempt to use LLMs to generate group recommendation rankings and rationales based on members' rating lists [33], or extract topic labels from user reviews and learn group preference representations via graph neural networks before generating recommendations [34], or adopt graph-based aggregation, with some recent work beginning to explore fairness concerns [35].

However, these methods often rely on explicit

topic modeling and static structures, making it difficult to capture the heterogeneity of users' semantic preferences, and thus lack the capacity for dynamic negotiation during the recommendation process.

Our work fills this gap by constructing semantic profiles for users and merchants directly from free-form reviews, enabling fine-grained modeling of preferences and implicit intentions. We further adopt CoT prompting to reason over multiple profiles, balancing satisfaction and conflict via multi-turn inference. The resulting recommendations align more closely with both individual intent and group consensus.

## 3 Methodology

### 3.1 MEET-LLM framework

To address the disconnect among multi-preference coordination, spatial feasibility, and real-world executability in GMLR, we propose MEET-LLM, an LLM-powered multi-agent collaborative recommendation framework [36]. MEET-LLM integrates semantic understanding, spatial optimization, and dynamic user modeling to bridge the digital preference space and the physical travel space in urban environments, where feasibility is shaped by geographic context and road-network routing. The framework moves beyond abstract rankings toward executable meeting location decisions that are accompanied by route-feasible plans.

As shown in Figure 2, the framework follows a perception-decision-action-feedback loop. Perception constructs mobility-aware contextual evidence, decision performs conflict-aware preference reasoning under feasibility constraints, action produces executable outputs that can be integrated with navigation services, and feedback updates historical context for continual improvement. For implementation, we instantiate this loop with seven components,  $L$ ,  $C$ ,  $M$ ,  $S$ ,  $D$ ,  $O$ , and  $K$ .

$L$  (*Large Language Model*): Provides semantic understanding throughout the pipeline, supporting review analysis, preference extraction, and recommendation generation. It serves as the core reasoning agent within the perception-inference-feedback loop.

$C$  (*Character*): Builds foundational representations for perception by modeling users' semantic preferences and mobility capabilities, as well as merchant attributes and target customer features. It constructs semantic and reachability representations between users and venues.

$M$  (*Memory*): Stores historical mobility data, preference evolution, and feedback logs of users. It provides contextual support for perception and decision-making, receives updates from the feedback module, and enables long-term adaptive learning of user profiles.

$S$  (*Perception*): Based on user-merchant profiles from  $C$  and historical data from  $M$ , this module integrates user geographic positions and mobility-related

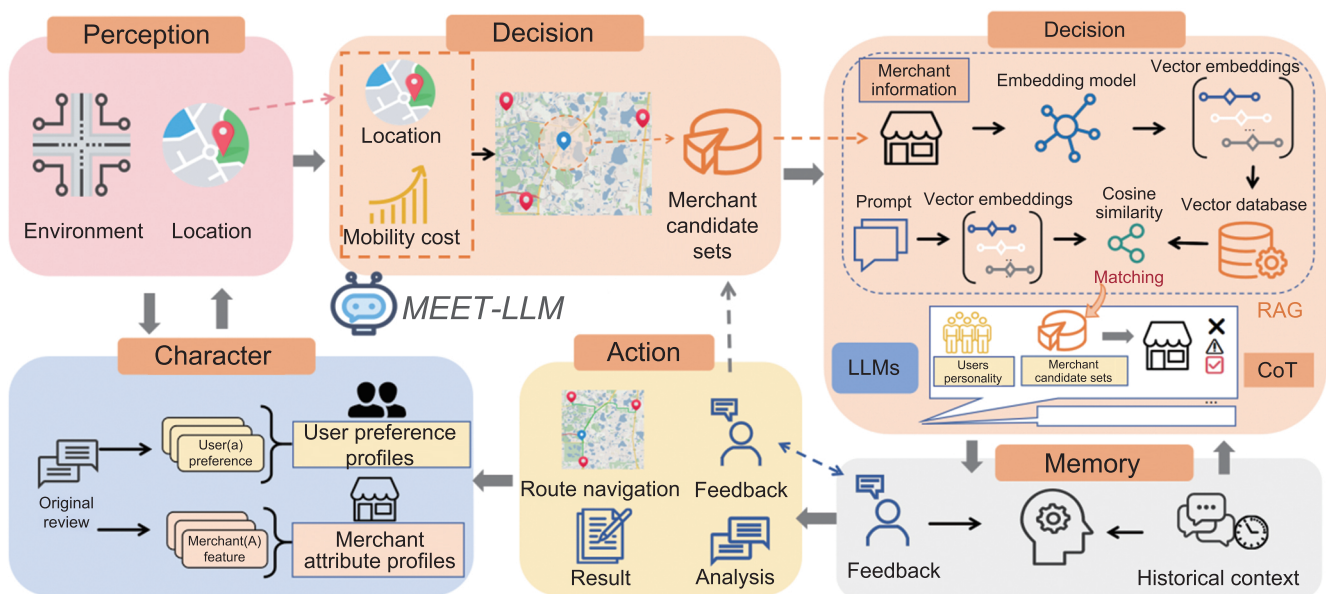


Figure 2 The framework of MEET-LLM

characteristics to construct a multimodal perception state that captures preference-relevant mobility constraints, which is then provided as input to  $D$ .

$D$  (*Decision*): Receives the perception state from  $S$  and the semantic reasoning capability from  $L$ , introduces the group-level mobility cost (MC) to filter out infeasible locations, and performs multi-round CoT reasoning to resolve preference conflicts and generate recommendations that are both preference-compatible and spatially feasible.

$O$  (*Action*): Based on the output from  $D$ , this module evaluates mobility feasibility and transfers the decision to real navigation or coordination services, generating a structured and executable recommendation, where the system outputs a ranked Top- $k$  list ( $k = 3$  by default) of candidate meeting locations, each accompanied by a reasoning explanation and navigation path.

$K$  (*Feedback*): Generates recommendation explanations, gathers feedback on acceptance and interpretability, and triggers re-inference through  $S$  and  $D$  when results are unsatisfactory or when the average semantic match (ASM) score falls below 0.4, including feeding back to  $S$  for adaptive spatial optimization when the number of recommended merchants is insufficient. All feedback is stored in  $M$  for continual optimization.

The entire system forms a closed-loop reasoning chain centered on LLMs—linking perception, decision, action, and feedback—to achieve integrated reasoning across semantic and geographic spaces. The  $L$  module provides semantic capabilities throughout the pipeline;  $C$ ,  $S$ , and  $M$  jointly construct the perception foundation;  $D$  coordinates group preferences and spatial feasibility;  $O$  ensures real-world execution; and  $K$  enables adaptive optimization. Together, they form a dynamic multi-agent recommender system that transforms group recommendation from abstract semantic ranking into executable, collaborative urban-level tasks.

### 3.2 Core framework modules

The MEET-LLM framework aims to recommend meeting locations that are both preference-compatible and convenient for all group members to reach. First, the PPM extracts structured user preference profiles and merchant attribute profiles from historical review data. Second, the SOM computes mobility cost and identifies

a set of spatially filtered candidate merchants. Third, the CRM integrates user preference profiles, merchant attribute profiles, and mobility cost through multi-turn LLM reasoning to generate the final meeting location recommendations.

### 3.3 Preference Profiling Module

Existing group recommendation methods often rely on ratings or predefined structures, which are insufficient to capture nuanced preferences expressed in user-generated text. To enable fine-grained and intent-aware preference modeling, we construct natural-language profiles from reviews. User reviews provide scalable, semantically rich, and behaviorally informative signals that reveal implicit decision cues beyond structured data [37]. Leveraging this, this module employs an LLM with guided CoT prompting to take historical reviews as input and generate expressive user preference profiles and merchant attribute profiles.

Building on prior work such as P5 [25], this module introduces a scene-differentiated strategy to extract fine-grained user preferences across diverse merchant contexts. Using carefully designed prompt templates and guided reasoning, the system distills structured preferences from reviews into a three-part representation: *User Likes*, *User Dislikes*, and *Additional Insights*. Likes and Dislikes summarize explicit, scene-specific positive and negative preferences expressed toward different types of merchants. Additional Insights consolidate implicit signals beyond surface-level statements, including (i) cross-scene and contrastive behavioral patterns (e.g., solitude-seeking in cafés but family-oriented behavior in shopping contexts) and (ii) inferred decision motives and psychological tendencies that help explain why certain attributes matter in a given scene. This dual-layer semantic modeling constructs a more complete user profile by combining surface-level preferences with inferred behavioral patterns and decision motives, thereby enhancing personalization and interpretability in downstream recommendation reasoning.

For each user-merchant interaction, the module extracts a structured user preference profile and a merchant attribute profile from the review text. The

user preference profile summarizes what the user likes, dislikes, and additionally mentions in reviews. The merchant attribute profile consists of two parts: the functional attribute profile and the target audience profile.

To ensure profile completeness and semantic fidelity, the system aggregates each user's historical reviews and performs instance-level reasoning across dual layers. Merchant profiles are similarly derived by aggregating multi-user reviews to identify consistent features and positioning. Detailed prompt texts are provided in Tables S2 and S3 of the Supporting Information (SI).

### 3.4 Spatial Optimization Module

In this module, we assume that users can adopt feasible travel modes given their individual conditions and local infrastructure constraints. Accordingly, mobility cost is treated as an abstraction of travel burden, primarily determined by route distance and physical ability. Specific transportation modes can be selected by users at the final navigation stage.

To model group-level differences in physical mobility and preliminarily filter candidate meeting locations, this module aims to identify a candidate central area that offers relatively low mobility costs and is balanced across group members. The system begins by collecting the geographic coordinates (latitude and longitude) of all group members and calculating their arithmetic mean to determine an initial centroid. This centroid serves as the starting point for filtering candidate meeting locations.

The mobility burden of each user is determined by physical ability and route distance. This is simplified into an individual mobility cost ( $MC_i$ ), as shown in Equation (1):

$$MC_i = b_i \cdot d_i \quad (1)$$

where  $b_i$  is the physical ability weight (set to 2 for mobility-impaired users, 1.5 for elderly users aged  $\geq 65$ , and 1 for average adults);  $d_i$  is the route distance from user  $i$  to the current centroid (computed using GraphHopper).

The specific weight values used in this study are illustrative and chosen to demonstrate the feasibility of the proposed framework. In practice, these parameters can be dynamically adjusted based on contextual

data, regional characteristics, or learned from historical usage patterns.

While real-world mobility burden may involve multiple factors (e.g., weather and transportation availability), the current implementation uses physical ability and travel distance as illustrative components to model mobility cost. The proposed framework is inherently extensible and can readily incorporate additional factors as relevant data become available.

To reflect group heterogeneity, the system computes normalized weights based on the inverse of  $MC_i$ :

$$w_i = \frac{1/MC_i}{\sum_{j=1}^n 1/MC_j} \quad (2)$$

Using these weights, the updated centroid coordinates ( $lat'$ ,  $lng'$ ) are calculated as:

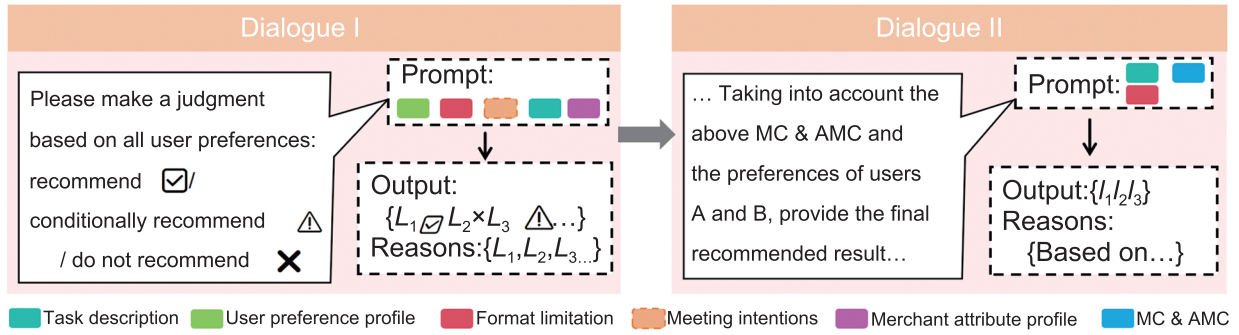
$$lat' = \sum_{i=1}^n w_i \cdot lat_i, \quad lng' = \sum_{i=1}^n w_i \cdot lng_i \quad (3)$$

This iterative process continues until convergence, as shown in Equations (2) and (3). Finally, the system uses the converged centroid as the center and a radius  $R$  to select businesses within this region as candidate meeting locations for downstream recommendation reasoning, with the radius adaptively adjusted through the feedback mechanism when necessary.

### 3.5 CoT Reasoning Module

To generate the final GMLR recommendations, the CRM feeds an LLM with user preference profiles, merchant attribute profiles, and user and merchant coordinates within a retrieval-augmented generation (RAG) structure. Guided by a CoT prompting process (Figure 3), the LLM outputs Top-3 meeting-location recommendations together with explanations and executable navigation routes for group members. LLM-based reasoning is sensitive to prompt design [38]. Excessive context can introduce noise and degrade inference [39]. We therefore use a multi-turn dialogue that decomposes reasoning across rounds while maintaining contextual continuity and retaining key information for multi-source decision integration. Detailed prompt texts are provided in Tables S4 and S5 of the SI.

Leveraging CoT prompting [40], we divide the reasoning process into two structured rounds. Each round



**Figure 3** Flowchart of the CRM

involves the following key prompt components:

*Task Description:* Specifies the round's objective.

*User Preference Profiles:* Loaded into the prompt to support inference.

*Format Constraints:* Guide LLM output format to reduce ambiguity and enforce structured responses.

*Merchant Attribute Profiles:* Filtered merchant sets generated from the SOM are introduced as the decision space.

*MC and average mobility cost (AMC):* The MC and AMC values are integrated to guide spatial feasibility decisions.

*Meeting Intentions:* User goals (e.g., dining and massage) may or may not be specified and are used to refine decision alignment when available.

By composing prompts with these structured components, the LLM performs layered reasoning, generating both final recommendations and interpretable explanations.

### 3.5.1 Preliminary recommendation

Considering that a large merchant candidate set may interfere with the reasoning ability of the LLM, we introduce a candidate merchant subset filtered by the SOM to effectively compress the search space. Subsequently, the system specifies the task goal to the LLM through customized prompt engineering: analyze the user preference profiles and meeting intentions, identify consistent and conflicting features of group preferences, and classify the candidate merchants into three categories—Recommended, where the majority preferences and intentions are largely matched; Conditionally Recommended, where partial alignment exists with minor trade-offs or conflicts; and Not Recommended, where mismatches or strong rejection signals

are present. Each classification is accompanied by a natural language explanation. Dialogue memory ensures contextual continuity for the next round of reasoning. To ensure sufficient preliminary recommendations, the framework incorporates an adaptive spatial expansion mechanism. When the number of recommended merchants falls below a predefined threshold (five in our implementation), the search radius is enlarged, and spatial optimization and CoT reasoning are re-initiated over the expanded candidate set. This iterative refinement continues until a sufficient number of suitable merchants is obtained, improving robustness under non-central or uneven user distributions.

### 3.5.2 MC-aware recommendation

Building on the previous classification, the system incorporates MC to refine final recommendations by balancing group preferences and geographic location. The LLM is prompted to analyze semantic alignment, reconcile preference conflicts, and jointly consider individual MC and AMC values. Since candidate merchants have already passed spatial filtering, this stage prioritizes semantic matching, with mobility-related factors serving as a secondary constraint to improve group acceptance. By default, the system outputs the Top-3 most suitable locations, each accompanied by an explanatory rationale. If a particular merchant significantly aligns with group preferences, the system allows a small subset of users to bear slightly higher MC, but such trade-offs must be explicitly justified in the model's explanation to maintain transparency and user trust. Final recommendations are seamlessly integrated with map navigation and routing systems. In our implementation, we use GraphHopper to generate mobility routes.

## 4 Experimental

### 4.1 Experimental setup

This study evaluates the effectiveness of the proposed MEET-LLM system for GMLR, focusing on group acceptance. Experiments are conducted on the Yelp dataset [41], which provides user IDs, review histories, merchant IDs, and merchant latitude and longitude. We use Lutz as the primary evaluation city and additionally include Wesley Chapel to examine generalization in cross-city scenarios. To ensure data quality and obtain sufficiently stable preference profiles, we retain users with more than 10 historical review interactions. GPT-4o is used as the backbone model due to its strong reasoning capabilities. To ensure consistent output, we set the temperature to 0, and structured prompts with format constraints and multi-turn memory are adopted to enhance consistency. To simulate real-world group decision-making, we generate 10 groups each for sizes 3, 4, and 5. For spatial realism, user locations are sampled from residential areas and transport hubs, with five geographically distributed configurations created per group size. Detailed group composition information is provided in the SI, Section 3.

To benchmark MEET-LLM, we adopt four representative group recommendation strategies as baselines: CF-A, which aggregates predicted scores by averaging across group members [42]; CF-L, which follows the least-misery principle by selecting the minimum score among members [42]; TGR-RA, which implements recommendation aggregation [43]; LLM-SC, which prompts an LLM to apply social choice-based aggregation strategies to produce group recommendations [9]. We further evaluate robustness by replacing the backbone with GPT-3.5-turbo, GPT-4o-mini, and GPT-4.1 under the same prompts.

All baselines use user-based collaborative filtering [44] for score prediction:

$$r_{u_i, x_j} = r_{u_i}^- + \frac{\sum_{u_k \in N} (r_{u_k, x_j} - r_{u_k}^-) \cdot \text{Similarity}(u_i, u_k)}{\sum_{u_k \in N} |\text{Similarity}(u_i, u_k)|} \quad (4)$$

where,  $r_{u_i, x_j}$  denotes the predicted rating of user  $u_i$  for merchant  $x_j$ ;  $r_{u_i}^-$  and  $r_{u_k}^-$  represent the average ratings

of users  $u_i$  and  $u_k$ , respectively;  $r_{u_k, x_j}$  is the rating given by neighbor  $u_k$  to merchant  $x_j$ ; Similarity  $(u_i, u_k)$  indicates the similarity between users  $u_i$  and  $u_k$ , calculated using the weighted Pearson correlation; and  $N$  is the set of neighbors similar to user  $u_i$ .

To evaluate the quality of generated explanations, we conducted a human study based on pairwise comparison. Since research on LLM-based group recommendation is still in its early stage, we compared our structured reasoning approach with two representative single-user explanation generation models: P5 [25], which uses a T5-based model, and XRec [45], which summarizes user reviews using GPT-4o. For each model, we randomly sampled 100 explanation instances for evaluation.

### 4.2 Evaluation strategy

To comprehensively evaluate MEET-LLM, we design a framework using six core metrics across four dimensions.

#### 4.2.1 System performance metrics

**AMC:** Represents the average mobility burden for all group members. Lower values indicate higher practical feasibility.

**ASM:** Measures how well recommended merchants align with user preferences. GPT-4o scores each user-merchant pair on a 1–5 scale, normalized to the  $[0, 1]$  range.

#### 4.2.2 Composite scoring metric

**Composite acceptance score (CAS):** A unified metric combining preference match and mobility cost, defined as:

$$\text{CAS} = \alpha \cdot \text{ASM} + (1 - \alpha) \cdot (1 - \text{NormMC}) \quad (5)$$

As shown in Equation (5), NormMC denotes the mobility cost normalized to the  $[0, 1]$  range using min-max scaling, with lower values indicating better group-level travel convenience. In the GMLR setting, the weighting coefficient  $\alpha$  lies in  $(0, 1)$ . Since preference consistency is typically prioritized over travel burden, to ensure consistency and comparability, we adopt a fixed weight of  $\alpha = 0.6$  in all main experiments to balance semantic preference matching and group-level travel burden. In addition, we further assess the robustness of this choice through a sensitivity analysis in Section 4.3.

Note that NormMC depends on the cost distribution and may vary with group size; therefore, CAS is not intended for strict cross-group-size comparisons.

#### 4.2.3 Human acceptance metrics

We evaluate the practical acceptability of group recommendations using two metrics: Top-3 Full Acceptance (G@3) and Majority Accept Rate (MAR). Each group is simulated by dedicated annotators—one per member—who assess each recommendation from the user's perspective, considering both preference fit and mobility feasibility. The evaluation was conducted on 100 randomly selected group instances.

**G@3:** A case is considered successful if any of the Top-3 recommended meeting locations receives unanimous "Yes" responses from all group members, each making a binary (Yes/No) judgment.

**MAR:** A meeting location is considered accepted if it receives "Yes" from at least  $\left\lceil \frac{n}{2} + 1 \right\rceil$  members in a group of size  $n$ . MAR is the proportion of recommended meeting locations satisfying this condition, reflecting the system's ability to generate broadly agreeable options.

#### 4.2.4 Explanation Quality Evaluation (EQE)

To evaluate the quality of explanations generated by LLM-based methods, we conducted a human evaluation study on Amazon Mechanical Turk (AMT) and recruited three independent annotators [32]. For each group recommendation instance, the annotators were shown a pair of explanations generated by two models and asked to indicate which one was better under six dimensions: clarity assesses whether the explanation

is clearly expressed and easy to understand; reasoning examines whether the explanation reflects the reasoning capability of LLMs by covering key decision factors such as user preferences and merchant attributes; justification evaluates whether the explanation convincingly supports why a merchant was recommended, with arguments that are sound and specific; trustworthiness measures whether the explanation appears reliable and consistent with the provided user and business data; faithfulness assesses which explanation more accurately reflects the actual content or attributes of the recommended merchant; and overall preference captures the annotators' holistic judgment considering all the above aspects.

### 4.3 Sensitivity analysis of $\alpha$

To examine the robustness of the CAS in Equation (5), we vary the weighting coefficient  $\alpha$  and recompute CAS while keeping all other settings unchanged. As shown in Figure 4, we analyze for both 3-person and 5-person groups with  $\alpha$  in  $\{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ . Notably, our method remains consistently competitive and achieves the best CAS in most settings across the tested  $\alpha$  range.

### 4.4 Experimental results

Across all group sizes, MEET-LLM consistently outperforms the baseline models in terms of semantic preference matching, spatial feasibility, and group-level acceptability, as shown in Table 1. Specifically, it achieves the highest scores on CAS, G@3, and MAR, indicating a strong ability to generate recommendations

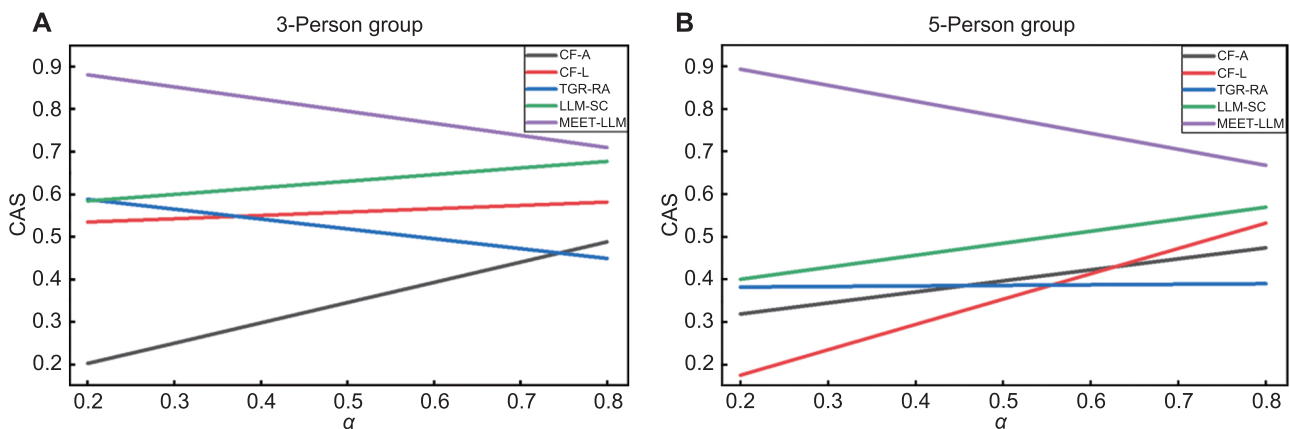


Figure 4 Sensitivity analysis of  $\alpha$  on CAS

**Table 1** Performance comparison of CF-A, CF-L, TGR-RA, LLM-SC, and M-L (MEET-LLM) under different group sizes. Arrows indicate the desired direction: up arrow means higher is better; down arrow means lower is better. Best results in each block are highlighted in bold

Models	ASM $\uparrow$	AMC $\downarrow$	CAS $\uparrow$	G@3 $\uparrow$	MAR $\uparrow$
3-Person group					
CF-A	0.5833	11.2332	0.3927	0.34	0.4533
CF-L	0.5972	8.9178	0.5658	0.48	0.6200
TGR-RA	0.4028	8.2691	0.4954	0.46	0.5133
LLM-SC	0.7083	8.7263	0.6461	0.68	0.7600
M-L	0.6528	6.5636	0.7668	0.92	0.9533
4-Person group					
CF-A	0.5625	8.8240	0.4079	0.28	0.5067
CF-L	0.4896	8.4178	0.4119	0.40	0.5867
TGR-RA	0.4271	7.0591	0.5338	0.32	0.3933
LLM-SC	<b>0.7292</b>	8.4528	0.5515	0.58	0.7067
M-L	0.6563	<b>6.3655</b>	<b>0.7527</b>	<b>0.84</b>	<b>0.8800</b>
5-Person group					
CF-A	0.5250	13.6792	0.4216	0.26	0.3733
CF-L	<b>0.6500</b>	15.9455	0.4123	0.36	0.4733
TGR-RA	0.3917	12.4708	0.3865	0.30	0.3400
LLM-SC	0.6250	12.8542	0.5122	0.54	0.6133
M-L	0.5917	<b>6.1361</b>	<b>0.7420</b>	<b>0.78</b>	<b>0.8667</b>

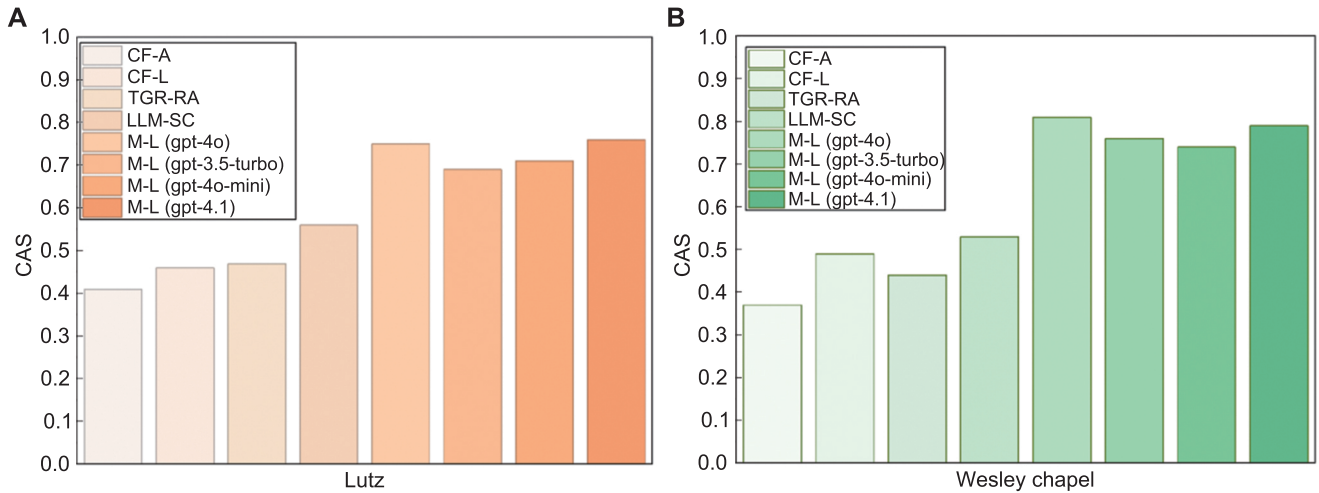
that are not only semantically aligned and geographically feasible but also broadly accepted by group members. G@3 and MAR reflect different dimensions of group acceptability—namely, unanimous agreement and majority consensus—and offer complementary perspectives on how well the model supports group coordination.

Although MEET-LLM does not always achieve the highest ASM, this is mainly due to its explicit consideration of MC. Unlike methods that simply maximize semantic preference alignment, MEET-LLM selects meeting locations that are convenient for all members, leading to significantly lower AMC values. This comprehensive approach has substantially enhanced group satisfaction. Compared with traditional baselines, CF-L improves preference matching by focusing on the least satisfied member, whereas CF-A and TGR-RA show unstable performance across different group settings. Notably, even with large language models, LLM-SC performs worse than MEET-LLM in the location recommendation setting. Ultimately, these baseline approaches lack a joint consideration of overall group

satisfaction and spatial feasibility, resulting in suboptimal performance.

We further examine robustness across LLM backbones and urban contexts. As shown in Figure 5, MEET-LLM maintains consistently strong CAS when switching the backbone model, and the overall advantage over baselines remains stable in both Lutz and Wesley Chapel. These results suggest that the observed gains are not specific to GPT-4o and that the proposed framework generalizes across different LLM backbones and city-level spatial distributions. We further report efficiency statistics of MEET-LLM, including end-to-end runtime and total token consumption across four LLM backbones, as shown in Table 2.

Under the EQE protocol, MEET-LLM consistently outperforms the two baseline explainers across all six criteria. As illustrated in Figure 6, it achieves uniformly higher scores on clarity, reasoning, justification, trustworthiness, faithfulness, and overall preference. The most pronounced gains appear in reasoning and justification, indicating that MEET-LLM generates more



**Figure 5** CAS comparison of CF-A, CF-L, TGR-RA, LLM-SC, and M-L (MEET-LLM) in Lutz (A) and Wesley Chapel (B) across different LLM backbones

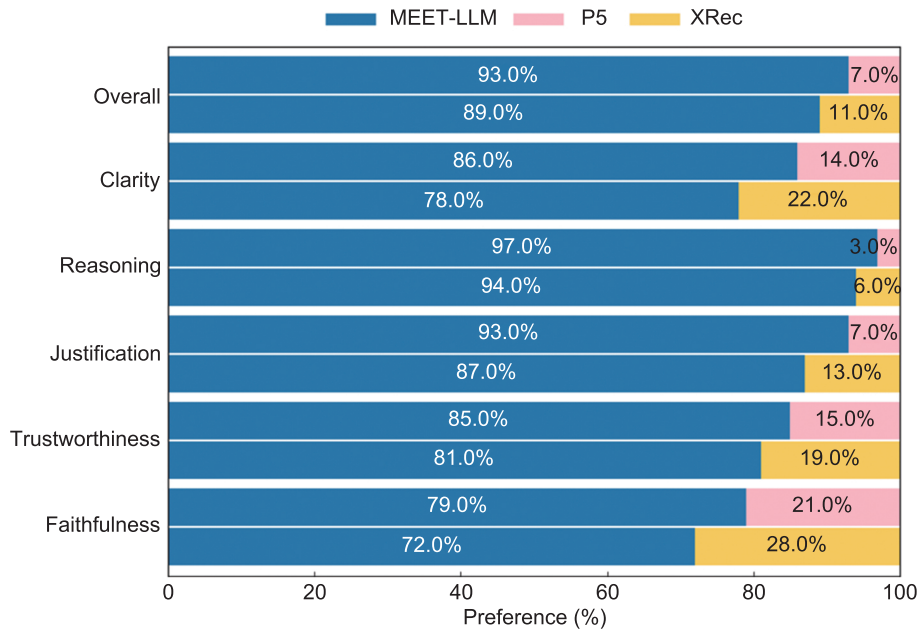
**Table 2** Runtime and token consumption of MEET-LLM under different LLM backbones

Metric	gpt-4o	gpt-4o-mini	gpt-4.1	gpt-3.5-turbo
Tokens	135,357	137,629	144,615	137,905
Time (s)	90	101	106	113

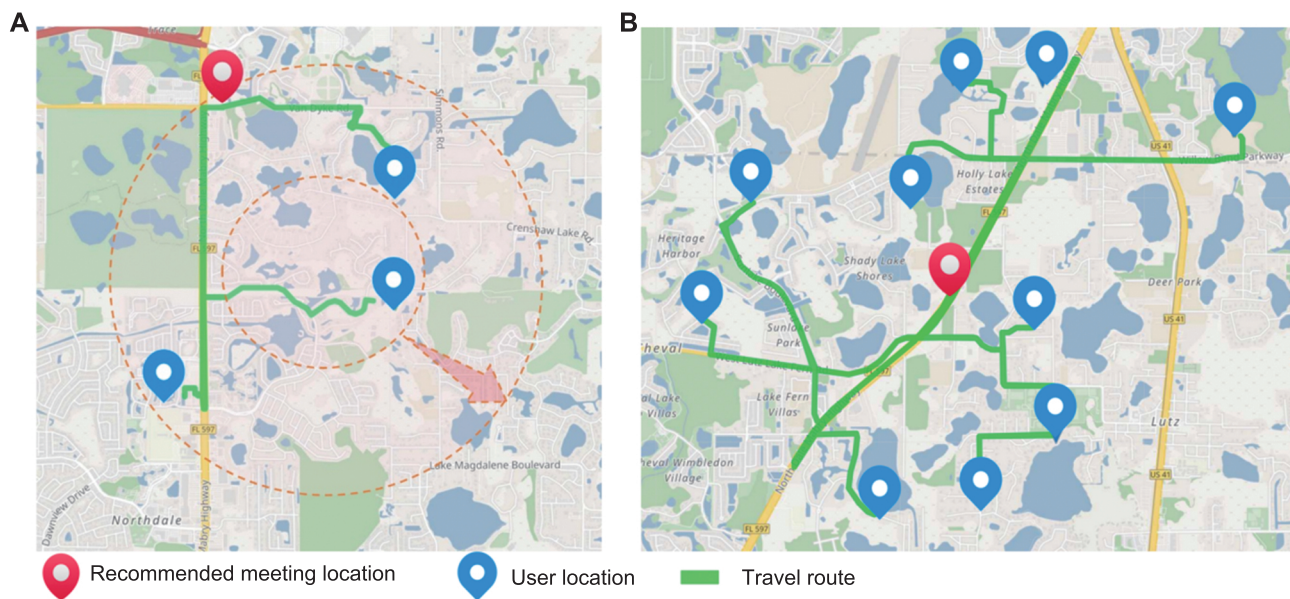
structured and persuasive explanations.

Figure 7A illustrates a representative suburban case in which the initial spatial region yields insufficient

candidate merchants. Through feedback-driven adaptive expansion of the search radius, the framework enables the CRM to identify a sufficient set of suitable merchants for recommendation. This example demonstrates the ability of the proposed framework to dynamically adjust the candidate search region and handle non-central or uneven user distributions. In addition, a recommendation scenario involving a larger group of ten users is provided as an illustrative example (Figure 7B), indicating the applicability of



**Figure 6** Evaluation of pairwise quality comparisons of human-evaluated explanations generated by MEET-LLM and baseline models (p-value < 0.05)



**Figure 7** Adaptive spatial expansion and large-group recommendation in MEET-LLM using a case study in Lutz, Florida, USA. (A) Illustration of adaptive spatial expansion under a suburban user distribution. The inner dashed circle denotes the initial spatial region with insufficient candidate merchants; through feedback-driven refinement, the search radius is expanded to the outer circle to identify suitable merchants. (B) Recommendation results for a larger group of ten users

the framework under increased group size and spatial dispersion.

#### 4.5 Ablation study

The MEET-LLM framework consists of three core modules: PPM, SOM, and CRM. To evaluate the contribution of each module to the overall recommendation pipeline, we conduct a series of ablation experiments.

Each ablation setting removes a critical component, as follows: No-PPM: disables PPM, directly feeds raw review text into CRM; No-SOM: omits the spatial filtering step in SOM and directly passes the unrefined candidate set to the reasoning stage; No-CRM: disables CRM and ranks candidates using a deterministic CAS-based weighted score of ASM and MC; No-MC:

excludes all MC-related prompts from both SOM and CRM; and Only-MC, an MC-only baseline where only SOM is retained. We assess the effects of each ablation on preference, mobility feasibility, and group-level acceptance metrics. Results are summarized in Table 3.

Through the ablation study, we observe that No-PPM directly feeds raw review text into CRM. Compared with MEET-LLM, removing structured preference and attribute profiles leads to consistent drops in CAS, G@3, and MAR. This suggests that profile-based representations provide cleaner and more decision-relevant signals than raw review text, which is often noisy and redundant. Without structured profiles, the reasoning stage becomes less effective at capturing the meeting intention and identifying mutually acceptable

**Table 3** Ablation study results. Up arrow means higher is better; down arrow means lower is better. Best values are in bold

Variant	ASM $\uparrow$	AMC $\downarrow$	CAS $\uparrow$	G@3 $\uparrow$	MAR $\uparrow$
MEET-LLM	0.7375	5.1593	<b>0.7986</b>	<b>0.85</b>	<b>0.90</b>
No-PPM	0.7223	5.2845	0.7112	0.64	0.79
No-SOM	0.7514	6.3648	0.6909	0.73	0.82
No-CRM	0.7465	5.1729	0.7967	0.61	0.75
No-MC	<b>0.8014</b>	10.8406	0.5869	0.19	0.28
Only-MC	0.3722	<b>4.8979</b>	0.5589	0.27	0.35

options, resulting in lower group-level acceptance. The No-SOM setting exposes the model to an unfiltered merchant pool. While this increases flexibility for preference matching, it leads to higher AMC and reduced acceptance. Thanks to the downstream CRM, which still considers mobility, feasibility degradation is partially mitigated. The No-MC variant builds on No-SOM by further removing all mobility cost prompts from CRM. The system thus prioritizes semantic fit alone, achieving the highest ASM but causing AMC to spike and acceptance metrics to drop. This confirms the joint importance of SOM and CRM, particularly CRM, in balancing preferences and feasibility. The Only-MC configuration disables both PPM and CRM, selecting locations purely by travel cost. Although AMC is minimized, semantic preference match and group acceptance drop sharply, highlighting that mobility alone cannot replace the central role of preference reasoning. No-CRM disables CRM and instead applies a deterministic CAS-based weighted scoring rule. However, while this variant achieves a relatively high CAS, its acceptance-oriented metrics (G@3 and MAR) remain clearly below those of MEET-LLM. This suggests that the deterministic rule does not explicitly model the meeting intention, resolve preference conflicts, or provide nuanced trade-offs, which can reduce the perceived acceptability of the final recommendations.

Table 4 summarizes the ablation settings, and Table 5 reports the corresponding relative contribution ratios.

**Table 4** Component removal checklist for ablation settings (check mark indicates the component is removed)

Component	No-PPM	No-SOM	No-CRM	Only-MC
PPM	√	–	–	√
SOM	–	√	–	–
CRM	–	–	√	√

**Table 5** Relative contribution ratios of PPM, SOM, and CRM across CAS, G@3, and MAR

Component	CAS	G@3	MAR
PPM	0.4437	0.3684	0.3235
SOM	0.5467	0.2105	0.2353
CRM	0.0096	0.4211	0.4412

The results show that SOM contributes the most to CAS sensitivity, CRM contributes most strongly to acceptance-oriented metrics such as G@3 and MAR, and PPM provides a stable contribution across all metrics. PPM extracts structured preferences from reviews to support alignment and explanation. SOM filters candidates based on travel feasibility. CRM plays a central role in final decision-making, balancing preference fit and mobility cost. Together, these modules enable recommendations that are both acceptable and feasible in real-world group settings.

## 5 Conclusions

MEET-LLM is a system-level, closed-loop framework for GMLR in ITS, designed to coordinate group preferences with routing-based spatial feasibility to produce executable meeting decisions in real urban travel contexts. The framework comprises three modules: a PPM that derives structured user and merchant profiles from natural-language reviews, a SOM that introduces an MC metric to filter the candidate set under feasibility constraints, and a multi-turn CRM that reconciles preference conflicts and generates structured, actionable outputs, including ranked venues, concise rationales, and route-feasible navigation guidance. Experiments on the Yelp dataset across different group sizes, cross-city settings, and multiple LLM backbones show that MEET-LLM consistently outperforms representative baselines in both feasibility and acceptance, as evidenced by improvements in CAS, G@3, and MAR together with reduced AMC, confirming the effectiveness and robustness of the proposed framework. Overall, MEET-LLM demonstrates the value of integrating LLM-driven preference reasoning with real-world mobility executability, providing a principled system-level pathway that advances GMLR from abstract semantic ranking toward action-ready, spatially grounded mobility coordination.

Several limitations should be acknowledged. First, the current evaluation relies on metrics and normalization schemes that are not strictly comparable across different group sizes, and the CAS trade-off weight  $\alpha$  is fixed for consistency; while sensitivity analysis shows robustness, future work should develop group-size-invariant evaluation protocols and calibrate  $\alpha$  through

user studies or feedback-driven learning. Second, preference profiling quality depends on the availability and reliability of review data, and this revision does not include systematic analyses of profile heterogeneity, sparsity, or noise; more robust profiling methods and dedicated quality metrics remain important directions. Third, LLM-based reasoning may exhibit instability or counterintuitive trade-offs when candidates are near-tied on preference and mobility, motivating future stability-enhancing strategies such as explicit tie-breaking rules and robustness checks. Fourth, the scalability of CoT reasoning with respect to candidate-set size and prompt efficiency is not exhaustively evaluated; controlled scalability studies and cost–quality trade-off analyses are needed. Finally, to ensure consistency and comparability in controlled evaluation, MEET-LLM currently adopts fixed parameter settings; future work will extend the framework with learnable and adaptive components based on logged decisions or acceptance feedback.

### Author contributions

Shuang Yang: Data processing; experiments; characterization; writing original draft. Lening Wang: Data collection; methodology; experiments. Zhiyong Cui: Validation; supervision. Yilong Ren: Validation; supervision. Liang Xu: Data processing; characterization. Mohamed Abouelela: Writing; reviewing & editing. Haiyang Yu: Validation; supervision. Aoyong Li: Conceptualization; methodology; experiments; writing; reviewing & editing.

### Acknowledgments

This research has been supported by the Beijing Natural Science Foundation (Grant No. L251059), the Fundamental Research Funds for the Central Universities, and the National Natural Science Foundation of China (Grant No. T2588101).

### Conflict of Interests Statement

The authors declare that they have no conflict of interest.

### Data Availability Statement

The data that support the findings of this study are

available from the corresponding author upon reasonable request.

### References

1. Y. Qian, Z. Lu, N. Mamoulis, and D. W. Cheung, "P-LAG: Location-Aware Group Recommendation for Passive Users," in *Proceedings of the Advances in Spatial and Temporal Databases*, edited by M. Gertz et al., 242–259, Springer, 2017, [https://doi.org/10.1007/978-3-319-64367-0\\_13](https://doi.org/10.1007/978-3-319-64367-0_13).
2. Z. Zhao, W. Fan, J. Li, et al., "Recommender Systems in the Era of Large Language Models (LLMs)," *IEEE Transactions on Knowledge and Data Engineering* 36, no. 11 (2024): 6889–6907, <https://doi.org/10.1109/TKDE.2024.3392335>.
3. L. Wu, Z. Zheng, Z. Qiu, et al., "A Survey on Large Language Models for Recommendation," *World Wide Web* 27 (2024): 60, <https://doi.org/10.1007/s11280-024-01291-2>.
4. K. Bao, J. Zhang, Y. Zhang, W. Wang, F. Feng, and X. He, "TALLRec: An Effective and Efficient Tuning Framework to Align Large Language Model with Recommendation," in *Proceedings of the 17th ACM Conference on Recommender Systems (2023)*, edited by J. Zhang et al., 1007–1014, Association for Computing Machinery, 2023, <https://doi.org/10.1145/3604915.3608857>.
5. F. Liu, Y. Liu, H. Chen, Z. Cheng, L. Nie, and M. Kankanhalli, "Understanding Before Recommendation: Semantic Aspect-Aware Review Exploitation via Large Language Models," *ACM Transactions on Information Systems* 43, no. 2 (2025): 1–26, <https://doi.org/10.1145/3704999>.
6. J. Zhang, K. Bao, Y. Zhang, W. Wang, F. Feng, and X. He, "Is ChatGPT Fair for Recommendation? Evaluating Fairness in Large Language Model Recommendation," in *Proceedings of the 17th ACM Conference on Recommender Systems (2023)*, edited by J. Zhang et al., 993–999, Association for Computing Machinery, 2023, <https://doi.org/10.1145/3604915.3608860>.
7. Y. He, X. Liu, A. Zhang, Y. Ma, and T. S. Chua, "LLM-2Rec: Large Language Models Are Powerful Embedding Models for Sequential Recommendation," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, edited by L. Antonie et al., 896–907, Association for Computing Machinery, 2025, <https://doi.org/10.1145/3711896.3737029>.
8. X. Wang, X. He, Y. Cao, M. Liu, and T. S. Chua, "KGAT: Knowledge Graph Attention Network for Recommendation," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019)*, edited by A. Teredesai et al., 950–958, Association for Computing Machinery, 2019, <https://doi.org/10.1145/3292500.3330989>.

9. C. Waterschoot, N. Tintarev, and F. Barile, "The Pitfalls of Growing Group Complexity: LLMs and Social Choice-Based Aggregation for Group Recommendations," in *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization (2025)*, 322–330, Association for Computing Machinery, 2025, <https://doi.org/10.1145/3708319.3733659>.
10. M. Li, L. Chen, G. Cong, Y. Gu, and G. Yu, "Efficient Processing of Location-Aware Group Preference Queries," in *Proceedings of the 25th ACM International Conference on Information and Knowledge Management (2016)*, 559–568, Association for Computing Machinery, 2016, <https://doi.org/10.1145/2983323.2983757>.
11. A. Dewan, K. Thapar, N. M. Telkar, P. Sengar, and S. R. Dey, "A Real-Time Adaptive Location-based Recommender System Integrating Personalities," paper presented at 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, April 7–9, 2023, <https://doi.org/10.1109/I2CT57861.2023.10126198>.
12. J. Kumar, B. K. Patra, B. Sahoo, et al., "Group Recommendation Exploiting Characteristics of User-Item and Collaborative Rating of Users," *Multimedia Tools and Applications* 83 (2024): 29289–29309, <https://doi.org/10.1007/s11042-023-16799-4>.
13. S. Birnkammerer, W. Woerndl, and G. Groh. Recommending for Groups in Decentralized Collaborative Filtering. Technical University of Munich, Department of Informatics, Technical Report TUM-I0927, 2009.
14. J. Gorla, N. Lathia, S. Robertson, and J. Wang, "Probabilistic Group Recommendation via Information Matching," in *Proceedings of the 22nd International Conference on World Wide Web (2013)*, 495–504, Association for Computing Machinery, 2013, <https://doi.org/10.1145/2488388.2488432>.
15. X. Wang, X. He, M. Wang, F. Feng, and T. S. Chua, "Neural Graph Collaborative Filtering," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (2019)*, 165–174, Association for Computing Machinery, 2019, <https://doi.org/10.1145/3331184.3331267>.
16. B. Wang and Y. Lu, "Graph Neural Network with Interaction Pattern for Group Recommendation," arXiv preprint, arXiv: 2109.11345, September 21, 2021, <https://doi.org/10.48550/arXiv.2109.11345>.
17. Z. Wang and Y. Li, "Knowledge Graph-Based Graph Neural Network Models for Multi-Perspective Modeling of Group Preferences," *Electronic Commerce Research* 25, no. 4 (2025): 2985–3008, <https://doi.org/10.1007/s10660-023-09771-9>.
18. S. Wang, S. Gao, X. Feng, A. T. Murray, and Y. Zeng, "A Context-Based Geoprocessing Framework for Optimizing Meetup Location of Multiple Moving Objects Along Road Networks," *International Journal of Geographical Information Science* 32, no. 7 (2018): 1368–1390, <https://doi.org/10.1080/13658816.2018.1431838>.
19. B. Chen, H. Zhu, W. Liu, J. Yin, W. C. Lee, and J. Xu, "Querying Optimal Routes for Group Meetup," *Data Science and Engineering* 6, no. 2 (2021): 180–191, <https://doi.org/10.1007/s41019-021-00153-5>.
20. M. Li, K. H. Lim, T. Guo, and J. Liu, "A Transformer-Based Framework for POI-Level Social Post Geolocation," in *Proceedings of the Advances in Information Retrieval*, edited by J. Kamps et al., 588–604, Springer, 2023, [https://doi.org/10.1007/978-3-031-28244-7\\_37](https://doi.org/10.1007/978-3-031-28244-7_37).
21. S. Feng, H. Lyu, F. Li, Z. Sun, and C. Chen, "Where to Move Next: Zero-shot Generalization of LLMs for Next POI Recommendation," paper presented at 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, June 25–27, 2024, <https://doi.org/10.1109/CAI59869.2024.00277>.
22. J. Long, L. Qu, G. Ye, T. Chen, Q. V. H. Nguyen, and H. Yin, "Unleashing the Power of Large Language Models for Group POI Recommendations," arXiv preprint, arXiv: 2411.13415, November 20, 2024, <https://doi.org/10.48550/arXiv.2411.13415>.
23. X. Song, Z. Liu, L. Meng, et al., "Accurate POI Recommendation for Random Groups with Improved Graph Neural Networks and a Multi-Negotiation model," *Scientific Reports* 15 (2025): 7531, <https://doi.org/10.1038/s41598-025-91805-3>.
24. Z. Liu, L. Meng, Q. Z. Sheng, D. Chu, J. Yu, and X. Song, "POI Recommendation for Random Groups Based on Cooperative Graph Neural Networks," *Information Processing & Management* 61, no. 3 (2024): 103676, <https://doi.org/10.1016/j.ipm.2024.103676>.
25. S. Geng, S. Liu, Z. Fu, Y. Ge, and Y. Zhang, "Recommendation as Language Processing (RLP): A Unified Pretrain, Personalized Prompt & Predict Paradigm (P5)," in *Proceedings of the 16th ACM Conference on Recommender Systems*, edited by J. Golbeck et al., 299–315, Association for Computing Machinery, 2022, <https://doi.org/10.1145/3523227.3546767>.
26. X. Wang, J. Cui, Y. Suzuki, and F. Fukumoto, "RDRec: Rationale Distillation for LLM-based Recommendation," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, edited by L. W. Ku et al., 65–74, Association for Computational Linguistics, 2024, <https://doi.org/10.18653/v1/2024.acl-short.6>.
27. L. Zhang, W. Ling, S. Daizhou, and L. Kuang, "HDRRec: Hierarchical Distillation for Enhanced LLM-based Rec-

- ommendation Systems," paper presented at the 2025 IEEE International Conference on Acoustics, Speech and Signal Processing, Hyderabad, India, April 6–11, 2025, <https://doi.org/10.1109/ICASSP49660.2025.10890603>.
28. L. Wang and E. P. Lim, "Zero-Shot Next-Item Recommendation Using Large Pretrained Language Models," arXiv preprint, arXiv: 2304.03153, 6 April, 2023, <https://doi.org/10.48550/arXiv.2304.03153>.
  29. J. Ji, Z. Li, S. Xu, et al., "GenRec: Large Language Model for Generative Recommendation," in *Proceedings of the Advances in Information Retrieval*, edited by N. Goharian et al., 494–502, Springer, 2024, [https://doi.org/10.1007/978-3-031-56063-7\\_42](https://doi.org/10.1007/978-3-031-56063-7_42).
  30. X. Zhang, B. Xu, Y. Wu, Y. Zhong, H. Lin, and F. Ma, "FineRec: Exploring Fine-grained Sequential Recommendation," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1599–1608, Association for Computing Machinery, 2024, <https://doi.org/10.1145/3626772.3657761>.
  31. L. Li, Y. Zhang, and L. Chen, "Prompt Distillation for Efficient LLM-based Recommendation," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 1348–1357, Association for Computing Machinery, 2023, <https://doi.org/10.1145/3583780.3615017>.
  32. J. Kim, H. Kim, H. Cho, et al., "Review-Driven Personalized Preference Reasoning with Large Language Models for Recommendation," in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1697–1706, Association for Computing Machinery, 2025, <https://doi.org/10.1145/3726302.3730055>.
  33. C. Waterschoot, N. Tintarev, and F. Barile, "Consistent Explainers or Unreliable Narrators? Understanding LLM-Generated Group Recommendations," in *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, 539–544, Association for Computing Machinery, 2025, <https://doi.org/10.1145/3705328.3748015>.
  34. S. Feng, Z. Lang, J. He, H. Zhang, W. Chen, and J. Cao, "A Group Recommendation Method Based on Automatically Integrating Members' Preferences via Taking Advantages of LLM," *Information Sciences* 709 (2025): 122067, <https://doi.org/10.1016/j.ins.2025.122067>.
  35. A. Tommasel, "Fairness Matters: A look at LLM-Generated Group Recommendations," in *Proceedings of the 18th ACM Conference on Recommender Systems*, edited by T. D. Noia et al., 993–998, Association for Computing Machinery, 2024, <https://doi.org/10.1145/3640457.3688182>.
  36. S. Kuang, Y. Liu, X. Wang, X. Wu, and Y. Wei, "Harnessing Multimodal Large Language Models for Traffic Knowledge Graph Generation and Decision-Making," *Communications in Transportation Research* 4 (2024): 100146, <https://doi.org/10.1016/j.commtr.2024.100146>.
  37. J. Yu, J. Zhao, L. Miranda-Moreno, and M. Korp, "Modular AI Agents for Transportation Surveys and Interviews: Advancing Engagement, Transparency, and Cost Efficiency," *Communications in Transportation Research* 5 (2025): 100172, <https://doi.org/10.1016/j.commtr.2025.100172>.
  38. X. Wang, M. Fang, Z. Zeng, and T. Cheng, "Where Would I Go Next? Large Language Models as Human Mobility Predictors," arXiv preprint, arXiv: 2308.15197, August 29, 2023, <https://doi.org/10.48550/arXiv.2308.15197>.
  39. Y. Hou, J. Zhang, Z. Lin, et al., "Large Language Models are Zero-Shot Rankers for Recommender Systems," in *Proceedings of Advances in Information Retrieval*, edited by N. Goharian et al., 364–381, Springer, 2024, [https://doi.org/10.1007/978-3-031-56060-6\\_24](https://doi.org/10.1007/978-3-031-56060-6_24).
  40. J. Wei, X. Wang, D. Schuurmans, et al., "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," in *Advances in Neural Information Processing Systems* 35 (2022): 24824–24837, <https://doi.org/10.48550/arXiv.2201.11903>.
  41. Yelp, "Yelp Open Dataset (2023)," accessed August 2, 2025. <https://www.yelp.com/dataset>.
  42. S. Amer-Yahia, S. B. Roy, A. Chawla, G. Das, and C. Yu, "Group Recommendation: Semantics and Efficiency," *Proceedings of the VLDB Endowment* 2, no. 1 (2009): 754–765, <https://doi.org/10.14778/1687627.1687713>.
  43. A. Felfernig, L. Boratto, M. Stettinger, and M. Tkalcic, *Group Recommender Systems* (Springer, 2018).
  44. S. Schiaffino, D. Godoy, J. A. D. Pace, and Y. Demazeau, "A MAS-Based Approach for POI Group Recommendation in LBSN," in *Proceedings of the Advances in Practical Applications of Agents, Multi-Agent Systems, and Trustworthiness. The PAAMS Collection*, edited by Y. Demazeau et al., 238–250, Springer, 2020, [https://doi.org/10.1007/978-3-030-49778-1\\_19](https://doi.org/10.1007/978-3-030-49778-1_19).
  45. Q. Ma, X. Ren, and C. Huang, "XRec: Large Language Models for Explainable Recommendation," in *Findings of the Association for Computational Linguistics: EMNLP 2024*, edited by Y. Al-Onaizan et al., 391–402, Association for Computational Linguistics, 2024, <https://doi.org/10.18653/v1/2024.findings-emnlp.22>.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section.