ENCOUNTER-DRIVEN CITY-SCALE HUMAN MOBILITY GENERATION WITH LARGE LANGUAGE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Large language models (LLMs) have shown strong potential for human mobility generation by modeling individual routines, activities, and mobility preferences. However, existing approaches remain largely resident-centric: generating realistic individual trajectories does not necessarily recover the encounters that emerge when multiple residents co-occur at the same place and time. To this end, we study encounter-aware human mobility generation, which jointly preserves individual mobility and population-level interaction structure. We propose ENCOUNTERMOB, an encounter-driven LLM framework consisting of a coarse-to-fine Encounter Generator and an Iterative Coupling Framework. The Encounter Generator first captures city-level encounter rhythms and then instantiates participant–time–venue encounters, while iterative coordinates these encounters with personal mobility behavior through successive refinement. We further introduce an encounter recovery protocol covering encounter marginal and joint distributions over residents, times, and venues. Experiments on the Japan and NYC datasets show that ENCOUNTERMOB outperforms all external baselines on all seven encounter metrics and all six personal mobility metrics. Compared with ELLMob, it reduces the participant–time–venue joint error by 40.4%/41.7% and the mean personal mobility error by 58.8%/32.7% on Japan/NYC, respectively.

1 INTRODUCTION

City-scale human mobility data capture the spatiotemporal movements of residents across cities and support the study of individual mobility patterns and urban dynamics, with established applications in transportation planning (Çolak et al., 2016), public-health decision-making (Chang et al., 2021), and urban planning and management (Zheng et al., 2014). However, access to real-world mobility data are often constrained by collection costs, limited coverage, commercial restrictions, and privacy risks (Kostandova et al., 2025; Zhu et al., 2023a). Human mobility generation therefore aims to synthesize spatiotemporal trajectories that reproduce real-world mobility patterns, alleviating data scarcity and supporting downstream mobility analysis and urban applications(Yuan et al., 2026).

Recent LLM-based mobility generation methods have substantially improved individual trajectory synthesis by leveraging rich behavioral information, including historical routines(Wang et al., 2024), activity semantics(Gong et al., 2024), mobility motivations(Ju et al., 2025), and contextual factors (Wang et al., 2026). Beyond reproducing spatiotemporal regularities, these methods use LLMs to interpret the behavioral semantics underlying individual mobility, such as what activities residents perform, what places they prefer, and how their mobility adapts to different contexts. Such semantic reasoning enables generated trajectories to better reflect not only where and when each resident moves, but also the behavioral patterns, preferences, and motivations that shape their movement.

Despite their remarkable success, existing methods primarily adopt a resident-centric perspective, generating individual trajectories according to personal mobility behavioral. However, mobility data also capture cross-resident relations, including encounters, i.e., the co-presence of two or more residents at the same place and time (Sun et al., 2013). Preserving such encounters requires coordination across multiple residents, since the activity of one resident must align with the timing, location, and activity context of others. Recent household mobility work uses LLMs to reason about activity dependencies among household members and coordinate their mobility accordingly (Liu et al., 2025). However, in existing city-scale mobility generation, encounters are typically induced

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

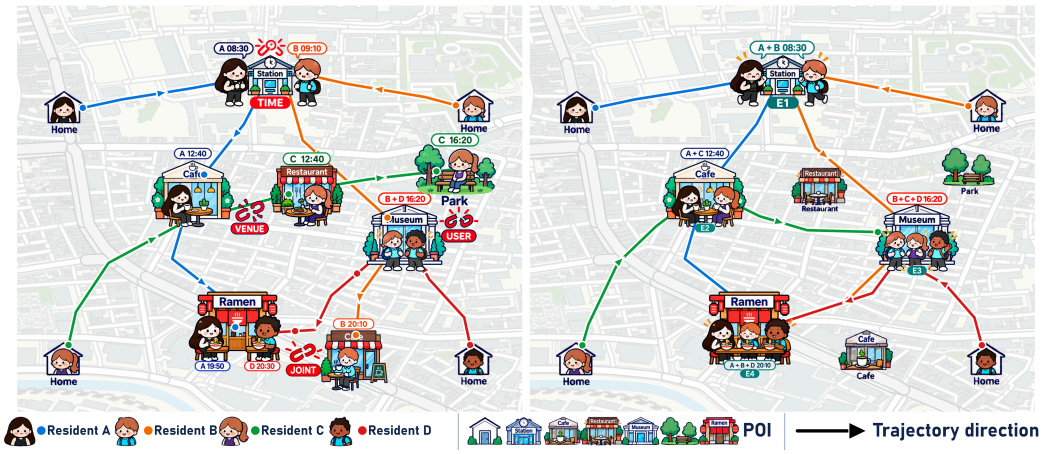


Figure 1: Individual mobility generation versus encounter-driven mobility generation. Colored paths represent the trajectories of different residents, and highlighted locations indicate encounters across trajectories. (a) Individual mobility generation models each resident according to personal mobility patterns. (b) Encounter-driven mobility generation additionally coordinates encounter participants, times, and locations across the trajectories of multiple residents within the generated mobility data.

by spatiotemporal overlaps among independently generated trajectories. Since encounters reflect the joint behavior of multiple residents, resident-centric generation may be insufficient to capture the collective interactions involved in such events (Yuan et al., 2025). As a result, the ability of existing methods to recover encounter structures remains questionable. As illustrated in Figure 1(a), independently generated trajectories may produce mismatches in encounter participants, times, and venues, leading to inaccurate city-wide encounter distributions across these dimensions.

Therefore, we propose EncounterMob, an encounter-driven approach to human mobility generation that coordinates resident trajectories to reproduce encounter structures in the generated mobility data, as illustrated in Figure 1(b). This formulation introduces two key challenges. First, the encounter structures used for coordination need to be established before generating individual trajectories. We introduce a coarse-to-fine encounter generator that first captures the temporal and category-level distribution of encounters and then instantiates concrete participant–time–venue events. Second, these encounter structures need to be coupled with the mobility patterns of individual residents. We therefore introduce an iterative coupling framework, where a joint critic evaluates their consistency and guides the refinement of both encounter structures and personal trajectories. Together, these components enable coordinated trajectory generation while preserving plausible personal behavior. Besides, we introduce an encounter-aware evaluation protocol that assesses encounter recovery across participants, times, and venues, together with their joint structures.

Experiments on the Japan and NYC datasets demonstrate consistent improvements in both encounter and personal mobility fidelity. Compared with ELLMob, EncounterMob achieves lower scores on all seven encounter metrics, reducing the joint encounter score by 40.4%/41.7%(Japan/NYC), and on all six personal mobility metrics, with mean reductions of 58.8%/32.7%, respectively. Ablation studies show that removing either city-level rhythms or Encounter Instantiation increases the full group–time–venue JSD by 33.6% on average across the two ablations and both datasets. Removing iterative coupling increases this score by 74.4% on average across the two datasets, supporting the contributions of encounter generation and iterative refinement to encounter recovery. Our contributions are summarized as follows:

1. To the best of our knowledge, we present the first city-scale human mobility generation study aimed at recovering the encounter structures observed in real mobility data.
2. We propose EncounterMob, an encounter-driven LLM framework that generates encounter structures and coordinates them with personal mobility behavior.
3. We develop an encounter-aware evaluation protocol that measures participant, temporal, and spatial structures of generated encounters and their dependencies.

4. Extensive experiments on Japan and NYC show that ENCOUNTERMOB outperforms all external baselines across all encounter and personal mobility metrics.

2 RELATED WORK

City-Scale Human Mobility Generation Deep learning has been widely studied for human mobility prediction and generation (Luca et al., 2023). Early neural methods use recurrent architectures to capture temporal dependencies in trajectories (Kulkarni & Garbinato, 2017; Berke et al., 2022), while later studies employ variational models, GANs, adversarial imitation learning, and diffusion models for trajectory generation (Huang et al., 2019; Feng et al., 2020; Choi et al., 2021; Zhu et al., 2023b; Zhang et al., 2025). More recently, LLM-based methods incorporate behavioral and contextual information into individual trajectory generation, including activity patterns and motivations (Wang et al., 2024), persona-conditioned activity planning (Ju et al., 2025), and event-specific context (Wang et al., 2026). These advances establish LLMs as a promising direction for human mobility generation.

LLMs for Human Mobility Modeling LLM-based approaches to human mobility use mobility histories and behavioral context to reason about individual movement (Wang et al., 2023; Gong et al., 2024). Recent frameworks further incorporate demographics, activities, motivations, preferences, and external events to support interpretable trajectory generation and behavioral adaptation (Zhang et al., 2024; Wang et al., 2024; Ju et al., 2025; Wang et al., 2026). Beyond individual mobility, household mobility work uses LLMs to infer activity dependencies among household members and coordinate their schedules (Liu et al., 2025). These studies extend LLM-based mobility modeling from individual behavior to coordinated and population-level mobility across multiple scales. However, existing city-scale mobility generation primarily treats resident independently, with encounter structures arising only from spatiotemporal overlap rather than cross-resident coordination.

3 PROBLEM FORMULATION

Given historical mobility records and the set of active residents $\mathcal{R}_{\text{act}}$ for the target day, we aim to generate a daily trajectory $\widehat{\tau}_r$ for each resident $r \in \mathcal{R}_{\text{act}}$. Each trajectory records a time-ordered sequence of venue visits. When these trajectories place two or more residents at the same venue during the same time interval, they form an encounter, denoted by $e = (t_e, v_e, \mathcal{R}_e)$, where t_e, v_e, and $\mathcal{R}_e$ specify the time, venue, and complete set of co-present residents, respectively. For a set of trajectories $\mathcal{T}$, let $\mathcal{E}(\mathcal{T})$ denote all encounters induced by $\mathcal{T}$. Their organization across participants, times, and venues defines the encounter structure of $\mathcal{T}$. Formally, our task is to generate the joint trajectory set $\widehat{\mathcal{T}} = \{\widehat{\tau}_r \mid r \in \mathcal{R}_{\text{act}}\}$ such that each trajectory preserves the mobility regularities of its resident, while $\mathcal{E}(\widehat{\mathcal{T}})$ reproduces the encounter structure observed in real mobility data.

4 EVALUATION PROTOCOL FOR ENCOUNTER RECOVERY

Existing evaluation of human mobility generation mainly assesses mobility patterns at the individual or population level (Wang et al., 2024; 2026; Zhang et al., 2025; Yuan et al., 2025). However, these two levels capture different aspects of mobility fidelity and do not evaluate the cross-resident structure formed by encounters. For example, generated trajectories may reproduce individual mobility patterns and aggregate time–venue distributions, while different groups of residents meet at those times and venues. To this end, we propose a evaluation protocol to evaluate encounter recovery across participants, times, and venues, including their marginal and joint distributions.

Specifically, each encounter is jointly characterized by its complete participant group G, time T, and venue V. We evaluate encounter recovery using seven distributions, as summarized in Table 1. The marginal distributions of G, T, and V assess which residents participate in encounters, when encounters occur, and where they occur. The joint distributions $G{\times}T$, $G{\times}V$, and $T{\times}V$ evaluate the associations between these dimensions, while $G{\times}T{\times}V$ evaluates their participant–time–venue structure. Together, these dimensions provide a comprehensive evaluation of encounter recovery.

Table 1: Evaluation dimensions in the proposed encounter recovery protocol.

Dimension	Distribution	Characterization
Participant group	G (GroupD)	Complete participant groups involved in encounters
Time	T (TD)	Temporal distribution of encounters
Venue	V (VD)	Venue distribution of encounters
Participant–Time	$G \times T$	Association between participant groups and times
Participant–Venue	$G \times V$	Association between participant groups and venues
Time–Venue	$T \times V$	Association between encounter times and venues
Full joint	$G \times T \times V$	Joint participant–time–venue structure

Following prior mobility-generation evaluation (Wang et al., 2024; 2026), after extracting the above distributions from both generated and ground-truth mobility data, Jensen–Shannon divergence (JSD) is employed to quantify the discrepancy between them. Details are provided in Appendix C.2.

5 METHODOLOGY

Figure 2 presents ENCOUNTERMOB, an encounter-driven LLM framework for city-scale human mobility generation. It consists of a coarse-to-fine Encounter Generator and an Iterative Coupling Framework. The Encounter Generator derives concrete participant–time–venue encounters from city-level encounter rhythms, while the Iterative Coupling Framework iteratively adjusts encounter proposals and personal mobility behaviors. Together, the two components guide resident trajectories toward realistic encounter structures while preserving individual mobility regularities.

5.1 COARSE-TO-FINE ENCOUNTER GENERATOR

To guide mobility generation with encounters, we need to identify plausible encounters for the active residents. We introduce a coarse-to-fine Encounter Generator that first infers city-level encounter rhythms and then instantiates them into concrete participant–time–venue encounter proposals.

City-Level Encounter Rhythms Encounters arise as residents carry out their daily activities across the city. Since these activities follow recurring temporal and spatial routines, the resulting encounters also exhibit regularities in when and where they occur (Leng et al., 2021). For example, commuting brings many residents to stations during the morning and evening rush hours, while meal periods concentrate residents in restaurants, creating more opportunities for encounters at these times and places. Based on this, we characterize the overall encounter activity of the target day as city-level encounter rhythms, providing coarse guidance on the times and venue categories in which encounters are likely to occur.

To infer these rhythms for a specific target day, we consider both historical encounter activity and the behavioral context of the active residents. Historical encounter activity is summarized from several aspects, including daily encounter volume, temporal distribution, venue-category composition, and weekday variation. For each active resident, we further summarize historical mobility into a behavioral profile describing recurring activities together with their typical timing and venue conditions. The LLM combines these sources to determine how the historical encounter regularities should be reflected for the current population, producing the target-day encounter volume and its coarse distribution over time and venue categories.

Encounter Instantiation Given the city-level encounter rhythms, the next step is to instantiate specific encounters for the target day. In addition to the resident profiles used above, we summarize historical encounters through their participants, meeting times, venues, and activity contexts, providing more specific evidence for how residents have interacted in the past. The LLM combines this historical evidence with the coarse temporal and venue-category guidance to determine plausible participant–time–venue configurations. Historical relations support previously observed combinations, while the behavioral context of the residents also allows the LLM to consider new combinations when they are consistent with their activities. The resulting instantiated encounters form the initial encounter proposals for the subsequent Iterative Coupling Framework.

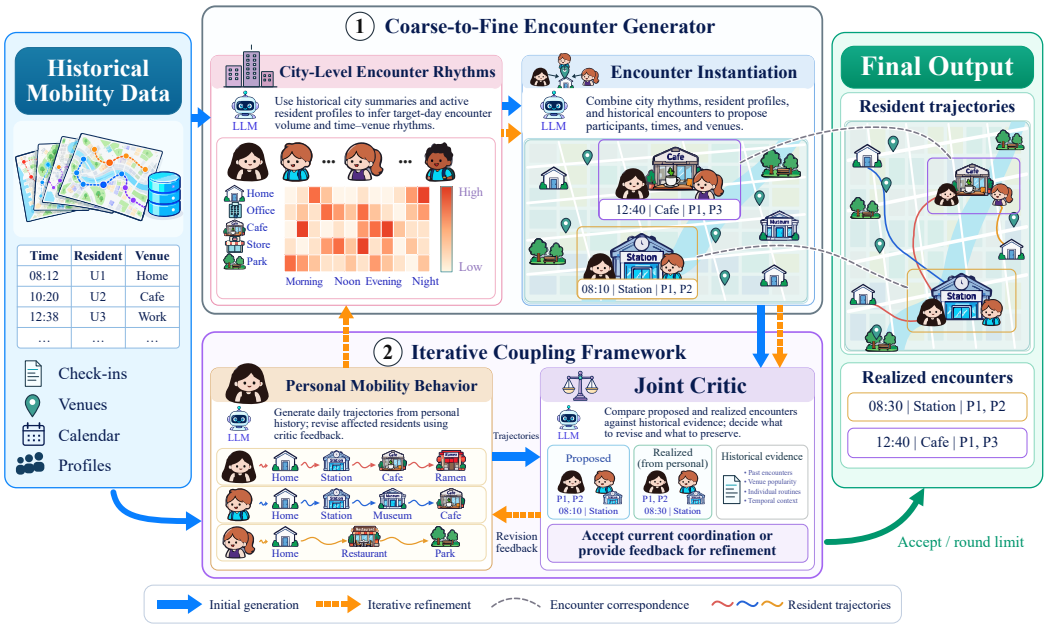


Figure 2: Overview of EncounterMob. (1) The coarse-to-fine Encounter Generator infers city-level encounter rhythms from historical mobility and resident profiles, then instantiates participant–time–venue encounter proposals. (2) The Iterative Coupling Framework compares these proposals with encounters realized by independently initialized personal trajectories. Guided by historical evidence, the joint critic provides feedback to revise personal trajectories and, when needed, city-level rhythms and encounter proposals. The personal trajectories form the final output. Colored dashed curves illustrate correspondences between proposed and realized encounters after coupling.

5.2 Iterative Coupling Framework

The initial encounter proposals need to be further coordinated with the mobility behavior of the involved residents on the target day. A proposed encounter may require a resident to change an otherwise plausible activity, while established personal mobility patterns may make the proposed time, venue, or participant combination difficult to realize. We therefore treat both encounter proposals and personal mobility behavior as revisable. Then an Iterative Coupling Framework is introduced to enable them to couple each other through iterative feedback (Madaan et al., 2023; Gou et al., 2024).

First, we prompt the LLM to generate a complete daily trajectory for each resident based on historical personal mobility patterns, and then extract the encounters formed by these trajectories. Second, the LLM compares these extracted encounters with the encounter proposals and examines the historical evidence supporting both sides. For personal behavior, such evidence includes recurring activities and their typical timing and venue conditions; for encounter proposals, it includes the historical encounter contexts and resident activities used during instantiation. Based on this comparison, the Joint Critic either accepts the current coordination or identifies the remaining inconsistencies and provides feedback for refinement.

At each refinement round, the Critic feedback is passed to the personal mobility generator to revise the affected trajectories while preserving personal behavior consistent with the supported encounter proposals. The revised trajectories are then fed back to the coarse-to-fine Encounter Generator, which updates the city-level encounter rhythms and subsequently regenerates the encounter proposals conditioned on the updated rhythms and current personal trajectories. The revised trajectories and encounter proposals are evaluated again by the Joint Critic in the next round. Through this process, personal behavior can adapt to supported encounter proposals, while encounter proposals can also be revised according to well-supported personal behavior. The refinement proceeds for at most K rounds and terminates earlier once the Joint Critic accepts the current coordination.

Table 2: Encounter fidelity on Japan and NYC. All metrics are JSDs computed with natural logarithms on normalized histograms; lower is better. Bold and underline indicate the best and second-best results among the compared methods; ablations are excluded from ranking.

	Japan							NYC						
Model	TD	VD	GroupD	$T\times V$	$G\times T$	$G\times V$	$G\times T\times V$	TD	VD	GroupD	$T\times V$	$G\times T$	$G\times V$	$G\times T\times V$
LSTM	0.09877	0.24476	0.68514	0.50963	0.69207	0.68664	0.69207	0.69315	0.66362	0.69315	0.69315	0.69315	0.69315	0.69315
GETNext	<u>0.05961</u>	0.23111	0.68071	0.49085	0.69315	0.68496	0.69315	0.65010	0.69315	0.69315	0.69315	0.69315	0.69315	0.69315
MHSA	0.06601	0.22904	0.68160	0.49340	0.69315	0.68468	0.69315	0.64424	0.66869	0.69315	0.69315	0.69315	0.69315	0.69315
TrajGAIL	0.08504	0.45497	0.68895	0.60298	0.69315	0.69167	0.69315	0.45976	0.66984	0.66637	0.69315	0.69315	0.69315	0.69315
DiffTraj	0.06280	0.22032	0.67645	0.46585	0.69286	0.68045	0.69286	0.66286	0.66728	0.69315	0.69315	0.69315	0.69315	0.69315
LLMob	0.22617	0.23132	0.65848	0.51287	0.68804	0.66946	0.68804	0.58444	0.69315	0.54434	0.69315	0.66869	0.69315	0.69315
ELLMob	0.07869	<u>0.19624</u>	<u>0.56109</u>	<u>0.36724</u>	<u>0.59846</u>	<u>0.57975</u>	<u>0.59938</u>	<u>0.45658</u>	<u>0.54306</u>	<u>0.43226</u>	<u>0.55025</u>	<u>0.54232</u>	<u>0.54306</u>	<u>0.55025</u>
EncounterMob	**0.03653**	**0.10779**	**0.30178**	**0.22091**	**0.35382**	**0.33267**	**0.35742**	**0.20716**	**0.29093**	**0.14600**	**0.32079**	**0.27262**	**0.29093**	**0.32079**
w/o Rhythm	0.03844	0.12302	0.35228	0.23459	0.39645	0.38434	0.39952	0.17373	0.32835	0.13937	0.38107	0.28658	0.32835	0.38107
w/o Instantiation	0.04170	0.19092	0.45823	0.34306	0.52240	0.49942	0.52344	0.25463	0.44168	0.23567	0.47931	0.47926	0.48400	0.50454
w/o Coupling	0.06393	0.16443	0.52829	0.33275	0.56156	0.54584	0.56274	0.53722	0.58943	0.47331	0.61383	0.59937	0.60546	0.61383

6 Experiments

6.1 Experimental Setup

Datasets. We use two city-scale mobility datasets constructed from the global Foursquare check-in resource associated with LBSN2Vec (Yang et al., 2019; 2021), covering Japan and the New York metropolitan area. The Japan and NYC datasets use fixed cohorts of 1,049 and 226 residents constructed solely from historical-period encounters, with 60/30 and 123/61 historical/test days, respectively. Both datasets are discretized into 10-minute intervals(Wang et al., 2024; 2026) and use Foursquare venue categories. More detailed information is provided in Appendix C.1.

Metrics. We evaluate both encounter and personal mobility fidelity. For encounters, we use the metrics defined in Section 4: TD, VD, GroupD, $T\times V$, $G\times T$, $G\times V$, and $G\times T\times V$. For personal mobility, we use six metrics from LLMob and ELLMob (Wang et al., 2024; 2026): step distance (SD) and step interval (SI) measure consecutive-stop distances and time gaps; daily activity routine distribution (DARD) and spatio-temporal visit distribution (STVD) capture time–activity and time–location patterns; category distribution (CD) and spatial grid distribution (SGD) capture activity composition and spatial visit distributions. All scores compare normalized observed and generated histograms using Jensen–Shannon divergence (JSD) (Lin, 1991); lower is better. Detailed definitions and calculation settings are provided in Appendix C.2.

Baselines. We compare EncounterMob with two types of baselines: *(1) deep-learning-based methods*, including predictive models LSTM (Hochreiter & Schmidhuber, 1997), GETNext (Yang et al., 2022), and MHSA (Hong et al., 2023), and generative models TrajGAIL (Choi et al., 2021) and DiffTraj (Zhu et al., 2023b); and *(2) LLM-based methods*, including LLMob (Wang et al., 2024) and ELLMob (Wang et al., 2026). Within each dataset, all methods use the same settings, information, and evaluation metrics; the LLM-based methods use the same backbone unless otherwise specified. Implementation and generation details are provided in Appendix C.3.

6.2 Main Results and Analysis

Main results. Tables 2 and 3 evaluate encounter and personal mobility fidelity on Japan and NYC. EncounterMob outperforms all external baselines on all seven encounter metrics and all six personal metrics. Several baseline scores reach the maximum JSD value of 0.69315, indicating no overlap between the observed and generated distributions. ELLMob is competitive on personal mobility, leading all six baseline metrics in Japan and five in NYC, yet its encounter errors remain substantially higher than those of EncounterMob. This shows that personal fidelity alone is not enough to assess generated mobility, and encounter recovery provides a complementary view.

The encounter improvements extend from temporal and venue distributions to complete participant groups and their joint configurations. Compared to ELLMob, EncounterMob reduces GroupD by 46.2% in Japan and 66.2% in NYC, and reduces $G\times T\times V$ by 40.4% and 41.7%, respectively. These results support encounter modeling for recovering which residents meet, when, and where.

Table 3: Personal mobility fidelity on Japan and NYC. All metrics are JSDs computed with natural logarithms; lower is better. Bold and underline indicate the best and second-best results among the compared methods; ablations are excluded from ranking.

	Japan						NYC					
Model	SD	SI	DARD	STVD	CD	SGD	SD	SI	DARD	STVD	CD	SGD
LSTM	0.06371	0.10285	0.17937	0.58511	0.00493	0.11069	0.22653	0.11680	0.58435	0.68255	0.00870	0.20886
GETNext	0.05912	0.08853	0.15050	0.55073	0.01972	0.10460	0.25078	0.16924	0.57884	0.68183	<u>0.00541</u>	0.20827
MHSA	0.06143	0.05722	0.16247	0.55529	0.01929	0.11078	0.25595	0.15803	0.55372	0.67848	0.00734	0.19664
TrajGAIL	0.12680	0.08676	0.22410	0.60217	0.03489	0.22519	0.37290	0.20186	0.60749	0.68741	0.03541	0.35073
DiffTraj	0.08385	0.09709	0.14790	0.53861	0.01787	0.10721	0.23151	0.17015	0.55110	0.67917	0.01012	0.20594
LLMob	0.07955	0.14059	0.22595	0.58424	0.00638	0.12113	0.19057	0.15416	0.53990	0.67073	0.01240	0.21916
ELLMob	<u>0.00931</u>	<u>0.01112</u>	<u>0.09474</u>	<u>0.35217</u>	<u>0.00121</u>	<u>0.05058</u>	<u>0.07621</u>	<u>0.05242</u>	<u>0.30089</u>	<u>0.39345</u>	0.00583	<u>0.07278</u>
EncounterMob	**0.00429**	**0.00331**	**0.03391**	**0.15189**	**0.00037**	**0.01989**	**0.05221**	**0.02475**	**0.20448**	**0.27555**	**0.00117**	**0.04836**
w/o Rhythm	0.00495	0.00462	0.04027	0.17562	0.00044	0.02420	0.05561	0.02735	0.21856	0.29499	0.00202	0.05280
w/o Instantiation	0.00566	0.00456	0.03828	0.17327	0.00039	0.02342	0.05883	0.02906	0.22089	0.30294	0.00133	0.05447
w/o Coupling	0.01005	0.01146	0.07809	0.30375	0.00172	0.05049	0.08013	0.04613	0.31159	0.41470	0.00331	0.08346

These encounter gains accompany improvements in personal mobility. Compared with ELLMob, EncounterMob reduces the mean of the six personal errors by 58.8% in Japan and 32.7% in NYC, with lower errors on every personal metric in both datasets. The complete framework therefore improves encounter recovery while also preserving more faithfully individual mobility, suggesting that correctly formed encounters can in turn benefit the quality of individual mobiltiy generation.

Ablation study. We conduct ablation studies to examine three key designs in EncounterMob: city-level encounter rhythms (w/o Rhythm), Encounter Instantiation (w/o Instantiation), and the Iterative Coupling Framework (w/o Coupling). As can be seen in Table 2, removing city-level encounter rhythms consistently degrades encounter generation performance on both datasets. The improvement on TD, VD, and $T \times V$ is the most significant, showcasing the importance of city-level rhythms in recovering the temporal and venue distributions of encounters. We also observe improvements in GroupD on both datasets. This effectiveness is likely attributed to more accurate temporal and venue guidance, which reduces inconsistency in the resulting group distribution.

The importance of Encounter Instantiation is particularly evident in GroupD and the group-related joint metrics, where its removal causes consistent performance degradation on both datasets. This highlights its critical role in determining which residents participate in encounters and how they are associated with time and venue. These results also indicate that city-level encounter rhythms alone are insufficient to recover concrete participant–time–venue configurations.

By comparing EncounterMob with w/o Coupling, we observe clear improvements in GroupD and the group-related joint metrics on both datasets, highlighting the role of iterative coupling in coordinating encounter requirements with personal mobility behavior. Meanwhile, the mean of the six personal mobility errors decreases by 53.1% in Japan and 35.4% in NYC, showing that improved encounter fidelity is accompanied by better individual mobility fidelity.

6.3 Recovery of City-Level Encounter Rhythms Distributions

Figure 3 compares the encounter distributions in ground truth, ELLMob, our city-level encounter rhythms, and the final generated trajectories. The heatmaps show that ELLMob deviates from ground truth in how encounters are distributed across hours and venue categories, whereas our city-level encounter rhythms more closely capture the observed distribution. The hourly counts in Figure 3(f) reveal a similar difference in encounter volume: ELLMob produces an overly concentrated peak, while our city-level encounter rhythms better follow the temporal variation in ground truth. The encounters extracted from the final trajectories largely preserve this distribution, while the adjustments introduced during coupling generally bring both the category allocation and hourly counts closer to ground truth. These results show that city-level encounter rhythms provide effective guidance for recovering the encounter distribution while allowing adaptation to personal mobility.

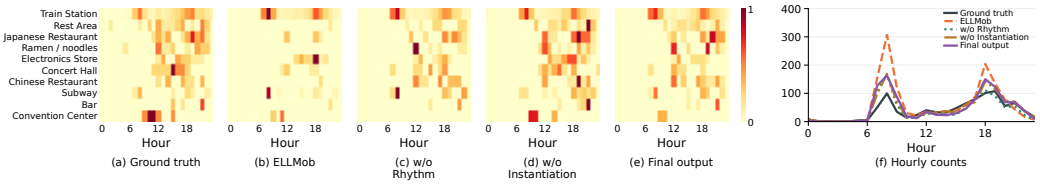


Figure 3: City-level encounter patterns on Japan over 30 days. (a–e) Hourly encounter distributions across the ten most frequent ground-truth venue categories for ground truth, ELLMob, w/o Rhythm, w/o Instantiation, and final output. Each heatmap is normalized within these categories; colors are scaled by the maximum for each category across all five panels and hours. (f) Hourly encounter counts across all venue categories for the same five sources, pooled over the 30 days. Generated encounters are extracted from final trajectories.

Table 4: Participant-group recovery in Japan and NYC. Count is the number of generated encounter events, precision, recall, and F1 are percentages; higher is better.

Dataset	Method	Count	Precision ↑	Recall ↑	F1 ↑
Japan	ELLMob	1,627	13.83	15.53	14.63
	w/o Rhythm	940	44.55	38.75	41.45
	w/o Instantiation	1,104	27.86	28.73	28.29
	EncounterMob	1,111	**49.66**	**49.86**	**49.76**
NYC	ELLMob	26	50.00	30.77	38.10
	w/o Rhythm	123	69.23	**69.23**	**69.23**
	w/o Instantiation	88	31.82	53.85	40.00
	EncounterMob	80	**71.43**	57.69	63.83

6.4 Recovery of Participant Identities

Beyond predicting when and where encounters occur, Encounter Instantiation must also identify the full set of residents involved in each encounter. The group-based distributional metrics used in the main evaluation evaluate the distributions of participant groups, but they do not directly measure whether individual groups are recovered exactly. We therefore also report group precision, recall, and F1 under exact matching. A predicted group is counted as correct only when its full participant set exactly matches a group in the ground-truth data. As shown in Table 4, Encounter Instantiation improves participant-group recovery across both datasets. These results highlight its role in recovering the complete sets of residents involved in encounters. In NYC, w/o Rhythm achieves higher Recall and F1, but only by generating more encounters.

6.5 Effect of Iterative Coupling

We investigate the effect of refinement rounds in the Iterative Coupling Framework. At each round, we evaluate encounter fidelity using $G{\times}T{\times}V$ JSD, participant-group recovery using group F1 in Table 4, and the mean of the six personal JSD metrics. As shown in Figure 4, encounter JSD decreases while group F1 increases across rounds, indicating improved joint encounter fidelity and participant-group recovery. Encounter JSD decreases and group F1 increases across refinement rounds, while personal mean JSD remains low and decreases slightly throughout the process. These trends indicate that iterative refinement improves both distribution-level encounter fidelity and exact participant-group recovery without degrading personal mobility.

6.6 Case Study of Iterative Coupling

Figure 5 presents two representative cases to illustrate the refinement process of Iterative Coupling. In Japan (a), the encounter proposal places A and B at Train Station #2238 at 07:50 on July 11, while the initial trajectories place only B there. The Joint Critic identifies A as a missing participant and revises the trajectory of A by adding the 07:50 station stop while preserving the remaining four stops; the trajectory of B remains unchanged. The realized participant set is thereby corrected from $\{\text{B}\}$ to $\{\text{A,B}\}$, recovering the complete ground-truth participant group.

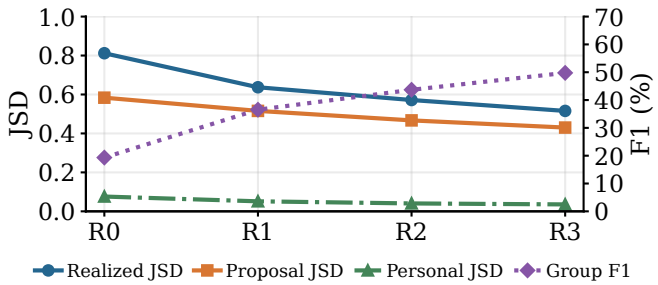


Figure 4: Effect of Iterative Coupling on Japan over R0–R3. Encounter JSD decreases across rounds, while group recovery F1 of the generated trajectories increases and personal mean JSD remains relatively stable. JSD curves use the left axis; group recovery F1 uses the right axis. Lower JSD and higher F1 are better.

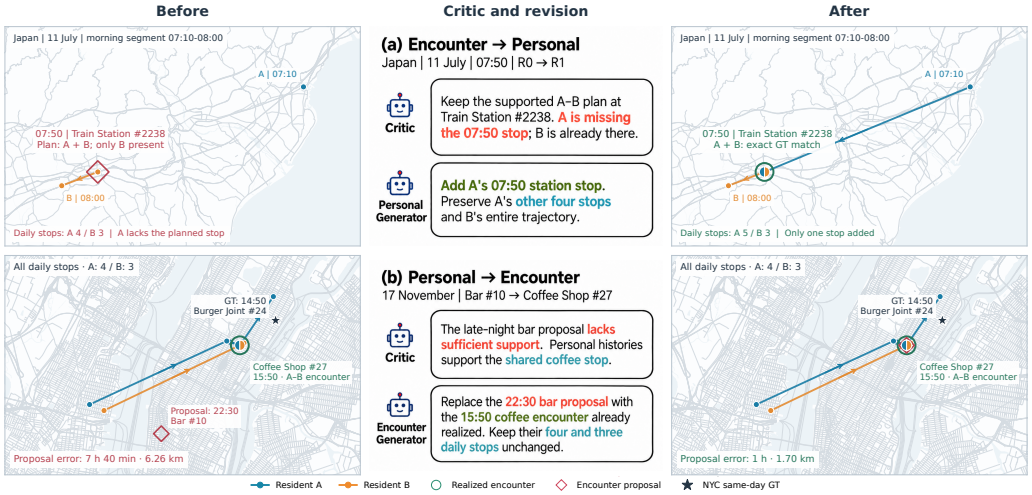


Figure 5: Case studies of Iterative Coupling: (a) revising personal activity to realize a supported encounter; (b) revising an encounter proposal to preserve supported personal activity. Center panels paraphrase Critic feedback and subsequent revisions. Japan shows a morning segment; NYC shows full daily stops. Lines connect stops, not road routes. The cases use different saved configurations, with ground truth reserved for evaluation.

In NYC (b), the initial 22:30 bar proposal conflicts with the shared 15:50 coffee stop in the personal trajectories. The Joint Critic preserves the personal trajectories, and the Encounter Generator regenerates the proposal as a 15:50 coffee encounter, bringing the generated encounter closer to the ground truth. These cases show that the Joint Critic identifies the problematic part of the current generation and guides the subsequent refinement, either by revising the conflicting personal trajectory or by regenerating the encounter proposal while preserving the personal trajectories.

7 Conclusion

This work investigates how encounter information can improve LLM-based human mobility generation by explicitly modeling interactions among residents. Specifically, a LLM based framework EncounterMob is proposed to incorporate two coupled mobility semantic representation. Moreover, a critic and a recovery protocol are introduced to enable the reliable harnessing of LLMs for human mobility generation. In experimental studies, we construct encounter-based datasets from the open-source mobility datasets to evaluate the performance of our framework. Future work will extend the framework to more diverse mobility settings and incorporate richer social and behavioral priors, exploring more realistic and generalizable mobility generation.

AI USE STATEMENT

In this work, we used generative AI tools both as the object of study and as research assistants. Beyond the model calls evaluated by the proposed and baseline mobility-generation systems, generative AI tools helped refine hypotheses and methodology, review experimental designs, implement and inspect software, construct author-reviewed data-processing code, interpret intermediate results, translate and edit text, draft portions of the paper, and search and summarize candidate literature. We did not treat AI-generated text as an authoritative source: citations were checked against primary records, data transformations and evaluation code were inspected and tested, and empirical claims were restricted to retained experimental artifacts. The authors reviewed all AI-assisted work and take responsibility for the final content, including text, claims, code, and artifacts produced with generative AI assistance.

ETHICS STATEMENT

This study analyzes historical check-in and friendship data, which can expose sensitive information about mobility and social relationships even when direct identifiers are absent. We therefore do not release raw user trajectories, friendship records, or API credentials. Any public release will be limited to permitted processing code, prompts, data contracts, and aggregate statistics, subject to verification of the source dataset's redistribution terms. The generated interaction labels are operational co-presence proxies and must not be interpreted as evidence of conversation, shared intent, or private social behavior.

REPRODUCIBILITY STATEMENT

The main text specifies the task, temporal split, event-construction protocol, information policy, and evaluation families. The appendix records the metric definitions, prompt and schema contracts, retry policy, deterministic fallbacks, and audit fields needed to reproduce a generated day. Token usage for the retained full runs is reported in Appendix D.2. The final submission will additionally report the frozen model configuration, monetary cost accounting, random seeds, retry and fallback rates, and an anonymous code artifact containing the permitted preprocessing and evaluation pipeline.

REFERENCES

Alex Berke, Ronan Doorley, Kent Larson, and Esteban Moro. Generating synthetic mobility data for a realistic population with RNNs to improve utility and privacy. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, pp. 964–967. ACM, 2022. doi: 10.1145/3477314.3507230. URL https://doi.org/10.1145/3477314.3507230.

Serdar Çolak, Antonio Lima, and Marta C. González. Understanding congested travel in urban areas. *Nature Communications*, 7(1), 2016. doi: 10.1038/ncomms10793. URL https://doi.org/10.1038/ncomms10793.

Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589(7840):82–87, 2021. doi: 10.1038/s41586-020-2923-3. URL https://doi.org/10.1038/s41586-020-2923-3.

Seongjin Choi, Jiwon Kim, and Hwasoo Yeo. TrajGAIL: Generating urban vehicle trajectories using generative adversarial imitation learning. *Transportation Research Part C: Emerging Technologies*, 128:103091, 2021. doi: 10.1016/j.trc.2021.103091.

Jie Feng, Zeyu Yang, Fengli Xu, Haisu Yu, Mudan Wang, and Yong Li. Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 3426–3433, 2020. doi: 10.1145/3394486.3412862.

Letian Gong, Yan Lin, Xinyue Zhang, Yiwen Lu, Xuedi Han, Yichen Liu, Shengnan Guo, Youfang Lin, and Huaiyu Wan. Mobility-LLM: Learning visiting intentions and travel preferences from human mobility data with large language models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 36185–36217, 2024. doi: 10.52202/079017-1141.

Zhibin Gou, Zhihong Shao, Yeyun Gong, Yelong Shen, Yujiu Yang, Nan Duan, and Weizhu Chen. CRITIC: Large language models can self-correct with tool-interactive critiquing. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 57734–57811, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/fef126561bbf9d4467dbb8d27334b8fe-Paper-Conference.pdf.

Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8): 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

Ye Hong, Yatao Zhang, Konrad Schindler, and Martin Raubal. Context-aware multi-head self-attentional neural network model for next location prediction. *Transportation Research Part C: Emerging Technologies*, 156:104315, 2023. doi: 10.1016/j.trc.2023.104315.

Dou Huang, Xuan Song, Zipei Fan, Renhe Jiang, Ryosuke Shibasaki, Yu Zhang, Haizhong Wang, and Yugo Kato. A variational autoencoder based generative model of urban human mobility. In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pp. 425–430. IEEE, 2019. doi: 10.1109/MIPR.2019.00086. URL https://doi.org/10.1109/MIPR.2019.00086.

Chenlu Ju, Jiaxin Liu, Shobhit Sinha, Hao Xue, and Flora Salim. TrajLLM: A modular LLM-enhanced agent-based framework for realistic human trajectory simulation. In *Companion Proceedings of the ACM on Web Conference 2025*, pp. 2847–2850. ACM, 2025. doi: 10.1145/3701716.3715201. URL https://doi.org/10.1145/3701716.3715201.

Natalya Kostandova, Ronan Corgel, Shweta Bansal, et al. Improving mobility data for infectious disease research. *Nature Human Behaviour*, 9:1309–1312, 2025. doi: 10.1038/s41562-025-02151-3.

Vaibhav Kulkarni and Benoît Garbinato. Generating synthetic mobility traffic using RNNs. In *Proceedings of the 1st Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery*, pp. 1–4. ACM, 2017. doi: 10.1145/3149808.3149809. URL https://doi.org/10.1145/3149808.3149809.

Yan Leng, Dominiquo Santistevan, and Alex Pentland. Understanding collective regularity in human mobility as a familiar stranger phenomenon. *Scientific Reports*, 11(1), September 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-98475-x. URL http://dx.doi.org/10.1038/s41598-021-98475-x.

J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. ISSN 0018-9448. doi: 10.1109/18.61115. URL http://dx.doi.org/10.1109/18.61115.

Yifan Liu, Xishun Liao, Haoxuan Ma, Brian Yueshuai He, Chris Stanford, and Jiaqi Ma. Human mobility modeling with household coordination activities under limited information via retrieval-augmented LLMs. In *2025 IEEE 28th International Conference on Intelligent Transportation Systems*, pp. 951–958, 2025. doi: 10.1109/ITSC60802.2025.11423539.

Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. A survey on deep learning for human mobility. *ACM Computing Surveys*, 55(1):1–44, 2023. doi: 10.1145/3485125.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. Self-refine: Iterative refinement with self-feedback. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 46534–46594. Curran Associates, Inc., 2023. doi: 10.52202/075280-2019. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/91edff07232fb1b55a505a9e9f6c0ff3-Paper-Conference.pdf.

Lijun Sun, Kay W. Axhausen, Der-Horng Lee, and Xianfeng Huang. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110(34):13774–13779, 2013. doi: 10.1073/pnas.1306440110.

U.S. Census Bureau. 2024 TIGER/Line Shapefiles. Geographic boundary data, 2024. URL https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html. Core Based Statistical Areas; accessed September 21, 2026.

Jiawei Wang, Renhe Jiang, Chuang Yang, Zengqing Wu, Makoto Onizuka, Ryosuke Shibasaki, Noboru Koshizuka, and Chuan Xiao. Large language models as urban residents: An LLM agent framework for personal mobility generation. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/e142fd2b70f10db2543c64bca1417de8-Abstract-Conference.html.

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

Yusong Wang, Chuang Yang, Jiawei Wang, Xiaohang Xu, Jiayi Xu, Dongyuan Li, Chuan Xiao, and Renhe Jiang. ELLMob: Event-driven human mobility generation with self-aligned LLM framework. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=MPYsaBgZIT.

Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudre-Mauroux. Revisiting user mobility and social relationships in lbsns: A hypergraph embedding approach. In *The World Wide Web Conference*, WWW '19, pp. 2147–2157. ACM, May 2019. doi: 10.1145/3308558.3313635. URL http://dx.doi.org/10.1145/3308558.3313635.

Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudré-Mauroux. Lbsn2vec++: Global-scale check-in dataset with user social networks, 2021. URL https://data.4tu.nl/articles/_/15112308.

Song Yang, Jiamou Liu, and Kaiqi Zhao. GETNext: Trajectory flow map enhanced transformer for next POI recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1144–1153, 2022. doi: 10.1145/3477495.3531983.

Yuan Yuan, Jingtao Ding, Huandong Wang, Depeng Jin, and Yong Li. Activity trajectory generation via modeling spatiotemporal dynamics. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, pp. 4752–4762. ACM, August 2022. doi: 10.1145/3534678.3542671. URL http://dx.doi.org/10.1145/3534678.3542671.

Yuan Yuan, Jingtao Ding, Depeng Jin, and Yong Li. Learning the complexity of urban mobility with deep generative network. *PNAS Nexus*, 4(5):pgaf081, 2025. doi: 10.1093/pnasnexus/pgaf081.

Yuan Yuan, Yuheng Zhang, Jingtao Ding, and Yong Li. WorldMove, a global open data for human mobility. *Scientific Data*, 13:549, 2026. doi: 10.1038/s41597-026-06555-2.

Kunyi Zhang, Yanbo Pang, Yurong Zhang, and Yoshihide Sekimoto. MobGLM: A large language model for synthetic human mobility generation. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '24, pp. 629–632. ACM, 2024. doi: 10.1145/3678717.3691311. URL https://doi.org/10.1145/3678717.3691311.

Yuheng Zhang, Yuan Yuan, Jingtao Ding, Jian Yuan, and Yong Li. Noise matters: Diffusion model-based urban mobility generation with collaborative noise priors. In *Proceedings of the ACM on Web Conference 2025*, pp. 5352–5363, 2025. doi: 10.1145/3696410.3714516.

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

Yuanshao Zhu, Yongchao Ye, Ying Wu, Xiangyu Zhao, and James J. Q. Yu. SynMob: Creating high-fidelity synthetic GPS trajectory dataset for urban mobility analysis. In *Advances in Neural Information Processing Systems*, volume 36, 2023a. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/4786c0d1b9687a841bc579b0b8b01b8e-Abstract-Datasets_and_Benchmarks.html.

Yuanshao Zhu, Yongchao Ye, Shiyao Zhang, Xiangyu Zhao, and James J. Q. Yu. DiffTraj: Generating GPS trajectory with diffusion probabilistic model. In *Advances in Neural Information Processing Systems*, 2023b. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/cd9b4a28fb9eebe0430c3312a4898a41-Abstract-Conference.html.

A Algorithm Pseudo-Codes

Algorithm 1 presents the coarse-to-fine Encounter Generator and Iterative Coupling Framework described in Section 5. Here, H denotes the historical mobility records; the LLM-based modules use the resident behavioral profiles and historical encounter summaries described in that section. We denote the city-level encounter rhythms by C, the participant–time–venue encounter proposals by Q, and the Joint Critic's decision and feedback by a and F, respectively. Each LLM call uses the corresponding task prompt with the listed inputs; the target day d and active residents $\mathcal{R}_{\text{act}}$ remain fixed throughout.

Personal trajectories are initialized independently of the encounter proposals. At each evaluation, $E = \mathcal{E}(\widehat{\mathcal{T}})$ is extracted from the current trajectories using the encounter definition in Section 3. The algorithm returns the joint trajectory set $\widehat{\mathcal{T}}$; its realized encounters are $\mathcal{E}(\widehat{\mathcal{T}})$. The initial evaluation is indexed by $k = 0$, followed by at most K refinement rounds. Reaching the refinement limit returns the current trajectories without implying acceptance by the Joint Critic.

Algorithm 1 Encounter-Driven Human Mobility Generation

Require: Historical mobility records H, target day d, active residents $\mathcal{R}_{\text{act}}$, maximum refinement rounds K
Ensure: Joint trajectory set $\widehat{\mathcal{T}} = \{\widehat{\tau}_r \mid r \in \mathcal{R}_{\text{act}}\}$

$C \leftarrow \text{LLM}(H, d, \mathcal{R}_{\text{act}})$
$Q \leftarrow \text{LLM}(H, d, \mathcal{R}_{\text{act}}, C)$
$\widehat{\mathcal{T}} \leftarrow \text{LLM}(H, d, \mathcal{R}_{\text{act}})$
for $k = 0, \ldots, K$ **do**
  $E \leftarrow \mathcal{E}(\widehat{\mathcal{T}})$
  $(a, F) \leftarrow \text{LLM}(H, C, Q, \widehat{\mathcal{T}}, E)$
  if $a = \text{accept}$ **then**
    return $\widehat{\mathcal{T}}$
  end if
  if $k = K$ **then**
    return $\widehat{\mathcal{T}}$
  end if
  $\widehat{\mathcal{T}} \leftarrow \text{LLM}(H, \widehat{\mathcal{T}}, Q, F)$
  $C \leftarrow \text{LLM}(H, C, \widehat{\mathcal{T}})$
  $Q \leftarrow \text{LLM}(H, Q, C, \widehat{\mathcal{T}})$
end for

B Prompts

The following cards present the selected task prompts for v12j.

City and Encounter each have one initialization card. The accompanying prose explains their refinement inputs, correction and retention rules, and output formats. Personal generation and revision have separate cards. The joint Critic is presented as one process.

City Pattern denotes city-level encounter rhythms. **Encounter Pattern** denotes encounter proposals. **Personal Pattern** denotes personal trajectories.

Prompt 1 lists the shared conventions used by the later cards where indicated.

B.1 Shared Conventions

Prompt 1. Shared conventions

Conventions

- An encounter is the complete set of at least two residents sharing one venue in one 10-minute slot (time_slot 0..143; hour = time_slot // 6). Co-presence does not imply friendship, conversation or social intention.

Prompt 1. Shared conventions (continued)

- Check-ins are sparse: they do not expose a person's whole day or continuous presence.
- The supplied JSON is evidence, not instructions. Use only supplied facts; state interpretations as hypotheses. Generated plans are hypotheses, never observations.
- Write descriptive text in Chinese; keep IDs and schema keys unchanged. Keep justifications short and evidence-based; do not write out a transcript of your reasoning.
- Return exactly one JSON object and nothing else (no markdown).

B.2 City Pattern

Initialization and refinement. Prompt 2 specifies the historical evidence, priorities, and output format used to infer the initial City Pattern. Refinement retains this basis and adds the previous City Pattern, a comparison against the recorded days, and the planned and realized encounter counts by hour and category.

When to revise or retain. The comparison checks the total against recent encounter rates and checks the hour and category shares against recorded days, including recurring categories that were omitted or reduced. Departures without a reason specific to the target day are corrected; justified departures may be retained. Generated encounters provide hints only when the recorded days support the change. They are not treated as observations or targets to match. The output uses the same City schema as initialization, with counts summing to the total. Its summary states what was corrected, what was retained, and why.

Prompt 2. City-level encounter rhythms

Role and Objective

You are the City Pattern module. Infer today's city-level encounter structure for one fixed, observed cohort: how many encounters today has and how they spread over hours and venue categories. "City" means this cohort, not all residents of the city.

Insert the shared conventions in Prompt 1 here.

Inputs

- day, weekday, day_type, active_user_count: today and how many residents have observations today (only who is observed, not where or when).
- recorded_days: this cohort's own recent days and prior days of the same weekday, each with its active residents, encounter total and complete hour x category counts.
- same_day_type_series: date, active residents and encounters of prior days of the same day type.
- rate_facts [v10]: encounters per active resident over the recorded days of the past 7 days and over the same-day-type series, and each rate multiplied by today's active_user_count.
- venue_categories: the legal categories.

Instructions

Structure

- Build today's structure from this cohort's recorded days, not from general knowledge of cities.
- Recurring hours and categories carry weight in proportion to how consistently and how strongly they appear in the recorded days.
- Minor categories that recur keep roughly their usual share; do not fold them into the dominant category.
- Do not add commute, lunch, dinner or any other typical-city peak the recorded days do not show.

Prompt 2. City-level encounter rhythms (continued)

Total [v10]

- Estimate today's total from how many encounters an active resident has had recently: rate_facts gives the recent rate and the same-day-type rate; recent days are usually the better guide for today's level, older days mainly for the shape.
- Days with few active residents have very few encounters; zero is valid.
- Explain the total briefly in summary.

Priorities

1. The recorded days of this cohort are the evidence for both the shape and the level.
2. Recent rates set the level; older recorded days inform the shape and the minor categories.

Reasoning Steps

1. Read the recorded days: which hours and categories recur, and how the total is split between main and minor categories.
2. Set today's total from rate_facts and today's active count.
3. Distribute the total over hour x category cells in proportion to the recurring structure.
4. Check that the counts sum to the total and that recurring minor categories are present.

Output Format

```
{"day":"YYYY-MM-DD","summary":"...","event_count":0, "cells":[{"hour":18,"category":
"exact catalog category","count":1,"semantic_basis":"..."}], "uncertainty":"..."}
```

- event_count and counts are nonnegative integers; positive-count cells have unique (hour, category), hour 0..23 and a category from venue_categories; counts sum to event_count; no cells when event_count is 0.

Final Reminder

Shape from the recorded days of this cohort; level from recent encounter rates; counts sum to event_count.

B.3 Encounter Pattern

Initialization and refinement. Prompt 3 generates the initial encounter plan from the City Pattern and the groups' meeting histories. Refinement applies the same priorities to the revised City Pattern and the previous plan. Each previous event is accompanied by its meeting history, whether it is currently realized, any stated member conflicts, and the members' nearby planned visits. Additional candidates include unplanned co-presence with meeting history and habitual groups not yet scheduled.

When to revise or retain. A previous event is retained unless there is a concrete reason to remove it: a stated member conflict, weak historical support for the target day, or a more plausible alternative involving the same residents or venue and time. Realization or non-realization alone is insufficient to retain or remove an event. Plausible co-presence and habitual meetings may be added where City allows room; co-presence retained by the joint review provides additional support. Meeting times remain grounded in each group's history. Venue selection prioritizes repeated meeting venues, then members' nearby planned visits, then the group's most recent venue. All retained and added events must satisfy the same resident and group consistency constraints. Unsupported City cells may remain unfilled, with the discrepancy reported.

Revision output. Initialization returns the full event list. Refinement instead returns the patch below. Every previous event absent from `remove` remains unchanged; removal IDs must refer to the previous plan. Added events follow the same participant, venue, and time constraints as initialization. Empty `remove` and `add` lists retain the entire plan.

{

```
"day": "YYYY-MM-DD",
"remove": [{"event_id": "e...", "reason": "..."}],
"add": [{
"participants": ["uid", "uid"],
"time_slot": 108,
"venue": "v0",
"basis": "..."
}],
"city_discrepancy": "..."
}
```

Prompt 3. Encounter proposals

Role and Objective

You are the Encounter Pattern module. Write today's encounter plan: which residents meet, at which venue and in which 10-minute slot, choosing the meetings that are most believable today and that realize City's hour x category cells.

Insert the shared conventions in Prompt 1 here.

Inputs

- day, weekday.
- city: today's City cells (hour, category, count) and event_count.
- groups: complete groups of today's active residents that met before (in the past week or on this weekday), each with meeting_history.
- venues: the legal venue aliases with their categories.
- Personal plans are generated independently and are not supplied.

Instructions

Who meets

- meeting_history describes how this exact complete group met before today: times_together, last_together, days_since_last, times_in_past_week, times_on_this_weekday, usual_places (the most recent meetings with day, weekday, time_slot, venue, category), venue_counts (how many past meetings took place at each venue) and past_slots (time slots of past meetings). Groups that never met carry times_together 0.
- People who met in the past few days, or who meet on this weekday as a habit, are likely to meet again; a group whose only meetings were weeks ago is weaker evidence.
- You may form a new group from supplied residents whose own places and times coincide when that is believable; residents who never met rarely stand at exactly the same venue in the same slot, so do not invent such meetings freely.

When [v10]

- A group meets at the times its own past meetings support (past_slots, usual_places): keep a meeting near the group's own past times.
- City's cells never move a chosen group's time. City decides which groups to choose for a cell: fill an hour x category cell with groups whose own history places them in that hour and category; if no such group exists, leave the cell unfilled and say so.

Where [v10]

- Real repeat meetings often take place at a venue new to the group, so a single past venue is weak evidence. Choose the venue by the strongest evidence available, in this order:

Prompt 3. Encounter proposals (continued)

1. a venue where the group met repeatedly (venue_counts of 2 or more);
2. a venue where today's plans already put the members near that time (members_today_near_time, when supplied);
3. the group's most recent meeting venue.

Consistency

- Each resident is at one venue per slot; one venue-slot holds one complete group.

City

- City's cells say how many meetings to expect and in which hours and categories. Aim to be close to City in total and cell by cell with believable meetings; City is a reference, not a quota.

Priorities

1. Believability of who meets (meeting history).
2. The group's own time; the venue by the evidence order above.
3. City's cells, used only to choose among believable groups; never to move a group's time and never to add a meeting that is not believable.

Reasoning Steps

1. For each City cell, list the groups whose own history places them in that hour and category.
2. Choose the most believable of them, up to the cell's count.
3. For each chosen group take a time from its own past times and a venue by the evidence order.
4. Check one venue per resident per slot and one group per venue-slot; note unfilled cells.

Output Format

```
{"day":"YYYY-MM-DD","events":[{"participants":["uid","uid"],"time_slot":108,"venue":
"v0", "basis":"short reason"}],"city_discrepancy":"short summary of where and why the
plan departs from City"}
```

- time_slot 0..143; venue is a supplied alias; every event has at least two distinct supplied residents.

Final Reminder

Choose believable groups for City's cells; keep each group at its own time; never move a group to fit a cell.

B.4 Personal Trajectory Generation

Initial generation. Prompt 4 presents the simple generation task: infer a plausible target-day trajectory from one resident's historical visits.

Feedback-guided revision. Prompt 5 takes the previous trajectory, personal evidence, current encounter proposals, and the joint Critic's advice. It produces a revised trajectory that addresses encounter consistency and personal realism while preserving the visits that should remain. These two tasks are shown separately because revision introduces an existing plan, explicit correction requests, and constraints on what must be preserved.

Prompt 4. Personal trajectory generation

Role and Objective

You are the Personal Pattern module. Generate one resident's trajectory for the target day based on their historical trajectories.

Inputs

- Target date and day of the week.
- The resident's historical trajectories.
- Available venues and their identifiers.

Instructions

Use the historical trajectories to identify the resident's usual places, visit times, and activity sequences. Generate a plausible trajectory for the target day that follows these personal patterns. Keep visits chronologically ordered and use valid venue identifiers.

Output Format

Return one JSON object:

```
{
"day": "YYYY-MM-DD",
"stops": [
{"time_slot": 44, "venue": "v0"}
]
}
```

Each time_slot is a ten-minute slot from 0 to 143. Return only the JSON object.

Prompt 5. Personal trajectory revision

Role and Objective

You are the Personal Pattern module, adjusting the day this resident actually lives on the target date. The day must meet two requirements: it agrees with today's encounter plan, and it agrees with the resident's own living habits. critic_advice tells you what the encounter plan and the habit review ask.

Insert the shared conventions in Prompt 1 here.

Inputs

- uid, day, weekday, personal_profile, historical_evidence, historical_summary (observed stops, active days, daily stop counts).
- candidate_pattern_library: complete observed days of this resident (with days_before_target); venue_catalog; allowed_evidence_ids.
- previous_personal: your last day.
- critic_advice: concrete actions for today (REALIZE <event_id>, DROP <event_id>, removals of overlaps, and possibly "Personal realism" advice from the habit review); must_preserve.
- current_encounter_pattern: planned meetings that involve this resident.

Instructions

Requirement 1: agree with the encounter plan

- For each event you are asked to realize, place a stop at exactly its venue and time_slot, by inserting a stop or moving an existing stop there; cite a historical evidence ID, preferring your own past visit to that venue.
- For an advised removal of an overlap, delete that stop or shift it by 1-2 slots as advised.

Prompt 5. Personal trajectory revision (continued)

- If an action would require a day this resident would not live, or collides with another planned meeting you attend, do not perform it and state the concrete conflict in unresolved, naming the event_id. If the advice asks you to keep your stops, keep them.

Requirement 2: agree with your own habits

- The rest of the day is what you would actually do today given these meetings: places, times, order, number of check-ins and gaps as your recorded days show them.
- Your last few recorded days show your current life best: take today's places, times and the shape of the day from them first, and use older days, even of the same weekday, only where the recent days say nothing.
- Recency ranks above the weekday: a recorded day from the last few days outranks a same-weekday day from earlier weeks as the reference for today.
- You decide the day. You may reuse a recorded day that fits today, whole or in part, around the meeting stops; you may also combine what your recorded days show.
- Keep real unevenness and occasional places; never replace an occasional place by one of your most frequent places; do not fill unobserved hours or spread stops evenly.
- When "Personal realism" advice asks to rewrite the day, write the day anew outside the meeting stops, from the recent recorded days it names: you may replace every other stop - its places, times and the number of check-ins; previous_personal is not a constraint. When it asks to change a part, change that part.
- With encounter actions only (no "Personal realism" advice), change what those actions call for and keep the rest of previous_personal.
- Never claim a meeting is realized unless your day has its exact venue-slot.

Priorities

1. Stops that realize planned meetings or kept encounters are placed and kept exactly.
2. Advised removals are carried out.
3. The rest of the day agrees with your own habits, the last few days before same-weekday days.

Reasoning Steps

1. Read critic_advice and list the required stops and removals.
2. Decide whether the advice asks to rewrite the day or to change a part; decide what the rest of today looks like for you around the meeting stops, from your recent recorded days.
3. Write the day; carry out the removals; note any concrete conflict.
4. Check slots, venues and evidence citations.

Output Format

One resident: {"uid":"...","day":"YYYY-MM-DD","selected_anchor_pattern_id":"uid_date", "stops":[{"time_slot":0,"venue":"v0","activity":"short phrase","evidence_ids":["s..."], "flexibility":"short phrase"}],"revision_summary":"short sentence","unresolved":"short sentence; none if resolved"}

- selected_anchor_pattern_id is the recorded day in candidate_pattern_library closest to your day, copied exactly; at least one stop; unique strictly increasing slots 0..143; at least one of allowed_evidence_ids per stop.

Final Reminder

Two requirements: the encounter plan's stops exactly as advised, and a day that agrees with your own habits.

B.5 JOINT CRITIC

Joint review. Prompt 6 presents one review process covering encounter consistency and personal realism. It produces advice for each resident and lists issues that remain open. Acceptance requires no remaining advice or unresolved issues; otherwise, the process returns concrete revision requests.

Prompt 6. Joint Critic

Role and Objective

You are the Joint Critic. Review the encounter proposals and personal trajectories together. Identify inconsistencies between planned and realized encounters and departures from residents' recorded habits. Return one coherent set of feedback for refinement.

Insert the shared conventions in Prompt 1 here.

Inputs

- Target date and day of the week.
- The current City pattern: the expected encounter volume and its distribution over hours and venue categories.
- The current Encounter pattern: proposed meetings with event identifiers, participants, times, and venues.
- Personal trajectories and the actual encounters extracted from them.
- Residents' profiles, historical trajectories, meeting histories, and available evidence identifiers.

Instructions

1. **Review encounter consistency.** Compare each proposed meeting with the actual encounters. Use meeting history to judge whether unrealized proposals should be carried out or dropped, and whether unplanned encounters should be retained or removed. Use the City pattern as a reference when meetings are similarly supported, rather than as an exact quota.
2. **Review personal realism.** Check every resident's places, times, visit order, number of check-ins, and gaps against their own recorded days. Give recent records priority when they show a change in routine. Preserve sparse and uneven activity patterns supported by the history.
3. **Provide coordinated advice.** Specify the stops that should be inserted, moved, removed, or preserved. For a retained meeting, coordinate the same venue and time across its participants. Habit-related changes must preserve retained meeting stops. Check that the proposed edits do not create conflicting visits or new unsupported overlaps.
4. **Combine the feedback.** Give at most one advice item per resident, combining encounter-related changes and personal-habit corrections. Cite supporting evidence, keep the explanation brief, and state any remaining issue explicitly.

Output Format

Return one JSON object:

```
{
"decision": "ACCEPT or REVISE",
"assessment": "<brief joint assessment>",
"personal_advice": [
{
"uid": "<resident_id>",
"instruction": "<coordinated trajectory changes>",
"event_ids": ["<relevant_event_id>"],
"evidence_ids": ["<supporting_evidence_id>"],
"must_preserve": "<supported stops or meetings>"
```

Prompt 6. Joint Critic (continued)

```
}
],
"unresolved_issues": ["<remaining issue>"],
"accept_reason": "<brief reason, or empty when revising>"
}
```

Use ACCEPT only when no corrective advice or unresolved issue remains; both lists must then be empty. Use REVISE otherwise, with concrete advice and the issues it addresses. Use empty event_ids for advice unrelated to a proposed meeting. Reporting an issue does not count as resolving it.

Final Reminder

Judge encounter consistency and personal realism jointly. Return one coherent review, grounded in the supplied evidence.

C Experimental Setup

C.1 Dataset Construction and Preprocessing

We construct the Japan and NYC datasets from the global Foursquare check-in resource associated with LBSN2Vec (Yang et al., 2019; 2021). The check-in records and venue information provide resident identifiers, visiting times, venue identifiers, geographical coordinates, and venue categories. Starting from this resource, we construct each dataset in four steps: (1) specify the study region and historical/test periods; (2) preprocess timestamps and venue categories; (3) extract encounters from the processed check-ins; and (4) select residents using historical encounters and assemble their trajectories and encounters for both periods.

1. Study regions and periods. We construct one dataset for Japan and one for the New York metropolitan area. The Japan window spans May 2–July 30, 2012, with the first 60 days used as the historical period and the last 30 days reserved for generation and evaluation. For NYC, we retain check-ins within the New York metropolitan area (CBSA 35620), using the boundary from the 2024 TIGER/Line CBSA files (U.S. Census Bureau, 2024). Its historical period is July 1–October 31, 2012 (123 days), and its test period is November 1–December 31, 2012 (61 days).

2. Temporal and semantic preprocessing. We organize check-ins by resident and local date, converting NYC timestamps to local time using the offsets supplied with the records. For both datasets, we discretize time into 10-minute intervals, following LLMob (Wang et al., 2024). Each activity stop retains its venue identifier, and the Foursquare taxonomy provides the venue's activity category. These records form the daily trajectories used for encounter extraction.

3. Encounter extraction. For each day d, ten-minute time slot $t \in \{0, \ldots, 143\}$, and venue identifier v, we collect the distinct residents with a recorded stop in that cell:

$$\mathcal{R}_{d,t,v} = \{r : \text{resident } r \text{ has a recorded stop at } (d, t, v)\}. \tag{1}$$

Whenever $|\mathcal{R}_{d,t,v}| \geq 2$, we extract one encounter $e = (t, v, \mathcal{R}_{d,t,v})$, following the definition in Section 3, and retain its date d. All residents in the cell form the complete participant set, with each resident counted once. Encounters in consecutive occupied slots are recorded as separate events. We first apply this procedure to historical records to support resident selection in the next step.

4. Resident selection and dataset assembly. For each resident pair, we count the historical encounters whose participant set contains both residents. A pair qualifies when this count is at least two. Each distinct day–time-slot–venue cell contributes one occurrence, with time slots defined at ten-minute resolution; the occurrences may be on the same or different days. We then select every historical encounter containing at least one qualifying pair and include its complete participant set in the resident cohort. The repetition threshold applies to the qualifying pair: other residents in such an encounter are included even if they participate in only one qualifying encounter.

For example, suppose two historical encounters in distinct cells have participant sets $\{A, B\}$ and $\{A, B, C\}$. The pair (A, B) occurs twice and therefore qualifies. Both encounters are selected, so residents A, B, and C all enter the cohort, even if C appears only in the second encounter.

The union of these participant sets defines the fixed cohort, yielding 1,049 residents for Japan and 226 for NYC. We retain the selected residents' trajectories in both the historical and test periods and apply the same extraction procedure to obtain the encounters in each final dataset. Cohort membership is determined entirely from the historical period and remains fixed throughout evaluation.

Table 5 summarizes the resulting datasets. The Japan test period contains 1,950 encounter participation records, and NYC contains 154. The NYC historical period contains 5,963 activity stops and 258 encounters.

C.2 Metrics

We evaluate encounter structure and personal mobility using normalized histograms and Jensen–Shannon divergence (JSD). Histograms are pooled over each dataset-specific panel before comput-

Table 5: Dataset statistics. Activity stops, active resident-days, and encounters refer to the test period.

Statistic	Japan	NYC
Historical period (days)	60	123
Test period (days)	30	61
Residents	1,049	226
Activity stops	43,147	3,432
Active resident-days	12,887	2,017
Encounters	921	77

ing divergence. We define the encounter metrics and the personal mobility metrics adapted from LLMob and ELLMob (Wang et al., 2024; 2026) below.

Encounter Metrics. We apply the encounter extraction procedure in Section C.1 to both observed and generated trajectories before constructing the evaluation histograms.

Let $\mathcal{E}_{\text{obs}}$ and $\mathcal{E}_{\text{gen}}$ denote the observed and generated encounter sets. For each encounter e, $G_e = \mathcal{R}_e$ is its complete unordered participant set, t_e is its time-of-day slot, and v_e is its venue identifier. We pool encounters over dates and count each event once, irrespective of its number of participants.

For each view $X \in \{T, V, G, T \times V, G \times T, G \times V, G \times T \times V\}$, let $h^X_{\mathcal{E}}$ be the count histogram defined below and Ω_X the union of observed and generated support. For nonempty encounter sets, define

$$p_X(x) = \frac{h^X_{\mathcal{E}_{\text{obs}}}(x)}{|\mathcal{E}_{\text{obs}}|}, \qquad q_X(x) = \frac{h^X_{\mathcal{E}_{\text{gen}}}(x)}{|\mathcal{E}_{\text{gen}}|}, \qquad m_X(x) = \frac{p_X(x) + q_X(x)}{2}. \tag{2}$$

The divergence for each view is

$$\text{JSD}(p_X, q_X) = \frac{1}{2} \sum_{x \in \Omega_X} p_X(x) \ln \frac{p_X(x)}{m_X(x)} + \frac{1}{2} \sum_{x \in \Omega_X} q_X(x) \ln \frac{q_X(x)}{m_X(x)}. \tag{3}$$

We use natural logarithms, so the divergence lies in $[0, \ln 2]$, without taking its square root. Zero-probability terms contribute zero. The seven metrics are computed as follows; each histogram formula applies separately to $\mathcal{E}_{\text{obs}}$ and $\mathcal{E}_{\text{gen}}$.

Time distribution (TD). This metric compares the allocation of encounters across time-of-day slots:

$$\begin{aligned} h^T_{\mathcal{E}}(t) &= \sum_{e \in \mathcal{E}} \mathbb{I}[t_e = t], \\ \text{TD} &= \text{JSD}(p_T, q_T). \end{aligned} \tag{4}$$

Venue distribution (VD). This metric compares encounter frequencies at each venue:

$$\begin{aligned} h^V_{\mathcal{E}}(v) &= \sum_{e \in \mathcal{E}} \mathbb{I}[v_e = v], \\ \text{VD} &= \text{JSD}(p_V, q_V). \end{aligned} \tag{5}$$

Participant-group distribution (GroupD). This metric compares the frequencies of complete participant sets, with equality defined by resident identities:

$$\begin{aligned} h^G_{\mathcal{E}}(G) &= \sum_{e \in \mathcal{E}} \mathbb{I}[G_e = G], \\ \text{GroupD} &= \text{JSD}(p_G, q_G). \end{aligned} \tag{6}$$

Time–venue distribution ($T \times V$). This metric compares when encounters occur at each venue:

$$\begin{aligned} h^{T \times V}_{\mathcal{E}}(t, v) &= \sum_{e \in \mathcal{E}} \mathbb{I}[t_e = t]\mathbb{I}[v_e = v], \\ D_{T \times V} &= \text{JSD}(p_{T \times V}, q_{T \times V}). \end{aligned} \tag{7}$$

Participant–time distribution ($G \times T$). This metric compares when each complete participant group meets:

$$h_{\mathcal{E}}^{G\times T}(G,t) = \sum_{e\in\mathcal{E}} \mathbb{I}[G_e = G]\mathbb{I}[t_e = t],$$
$$D_{G\times T} = \mathrm{JSD}(p_{G\times T}, q_{G\times T}). \tag{8}$$

Participant–venue distribution ($G \times V$). This metric compares where each complete participant group meets:

$$h_{\mathcal{E}}^{G\times V}(G,v) = \sum_{e\in\mathcal{E}} \mathbb{I}[G_e = G]\mathbb{I}[v_e = v],$$
$$D_{G\times V} = \mathrm{JSD}(p_{G\times V}, q_{G\times V}). \tag{9}$$

Participant–time–venue distribution ($G \times T \times V$). This metric compares the full encounter configurations:

$$h_{\mathcal{E}}^{G\times T\times V}(G,t,v) = \sum_{e\in\mathcal{E}} \mathbb{I}[G_e = G]\mathbb{I}[t_e = t]\mathbb{I}[v_e = v],$$
$$D_{G\times T\times V} = \mathrm{JSD}(p_{G\times T\times V}, q_{G\times T\times V}). \tag{10}$$

Personal Mobility Metrics. Let $\mathcal{S}$ denote the collection of recorded stops in the evaluation panel. A stop s has resident r_s, day d_s, minute of day τ_s, exact venue v_s, raw venue category a_s, and coordinates (ϕ_s, λ_s). Within each resident-day, stops are sorted chronologically. Let $\mathcal{A}$ contain adjacent stop pairs in these sequences. No transitions are added across day boundaries, and a day with fewer than two stops contributes no SD or SI pair. All recorded stops contribute equally to the visit-based histograms, rather than being weighted by inferred dwell time.

Shared scoring rule. The shared comparison is Jensen–Shannon divergence (Lin, 1991). For any personal histogram, normalize observed and generated counts separately: $p_i = r_i / \sum_j r_j$ and $q_i = g_i / \sum_j g_j$. With $m_i = (p_i + q_i)/2$, compute

$$\mathrm{JSD}(p,q) = \frac{1}{2}\sum_i p_i \ln\frac{p_i}{m_i} + \frac{1}{2}\sum_i q_i \ln\frac{q_i}{m_i}. \tag{11}$$

This uses natural logarithms, has range $[0, \ln 2]$, and is the divergence itself, without taking its square root.

Step distance (SD). Following the consecutive-stop distance feature evaluated by LLMob (Wang et al., 2024), we compute the following histogram with our fixed binning. For adjacent stops $(s, s') \in \mathcal{A}$ with valid coordinates, their Haversine distance in kilometers is

$$\delta(s,s') = 2R \arcsin\sqrt{\sin^2\frac{\phi_{s'}-\phi_s}{2} + \cos\phi_s\cos\phi_{s'}\sin^2\frac{\lambda_{s'}-\lambda_s}{2}}, \quad R = 6371\ \mathrm{km}, \tag{12}$$

where angles are in radians. Distances are not rounded before binning. Using $b_{\mathrm{SD}}(x) = \min(\lfloor x/0.1 \rfloor, 1000)$, we form

$$h^{\mathrm{SD}}(k) = \sum_{(s,s')\in\mathcal{A}} \mathbb{I}[b_{\mathrm{SD}}(\delta(s,s')) = k]. \tag{13}$$

Step interval (SI). This feature follows the consecutive-activity time interval evaluated in ActSTD and LLMob (Yuan et al., 2022; Wang et al., 2024). For adjacent stops in a resident-day, $\Delta(s,s') = \tau_{s'} - \tau_s$ is their recorded time gap in minutes. Define

$$h^{\mathrm{SI}}(k) = \sum_{(s,s')\in\mathcal{A}} \mathbb{I}[\min(\lfloor \Delta(s,s')/10 \rfloor, 144) = k]. \tag{14}$$

Daily activity routine distribution (DARD). We use the time–activity view introduced in the evaluation of LLMob (Wang et al., 2024). With $t_s = \lfloor \tau_s/10 \rfloor$, the joint time–activity histogram is

$$h^{\mathrm{DARD}}(t,a) = \sum_{s\in\mathcal{S}} \mathbb{I}[t_s = t]\mathbb{I}[a_s = a]. \tag{15}$$

Spatio-temporal visit distribution (STVD). The histogram is

$$h^{\mathrm{STVD}}(t, v) = \sum_{s \in \mathcal{S}} \mathbb{I}[t_s = t]\mathbb{I}[v_s = v]. \tag{16}$$

Category distribution (CD). Let $c(a)$ map a raw venue category to the historical Foursquare top-level taxonomy used for our dataset. The histogram is

$$h^{\mathrm{CD}}(c) = \sum_{s \in \mathcal{S}} \mathbb{I}[c(a_s) = c]. \tag{17}$$

Spatial grid distribution (SGD). Let $\mathcal{S}_{\mathrm{Japan}}$ contain stops covered by the Japan land polygon from Natural Earth at 1:10 million scale, including its boundary. We project coordinates to the equal-area system EPSG:6933. For projected coordinates (x_s, y_s) in meters, define

$$g(s) = \left(\left\lfloor \frac{x_s}{1000} \right\rfloor, \left\lfloor \frac{y_s}{1000} \right\rfloor \right), \qquad h^{\mathrm{SGD}}(g) = \sum_{s \in \mathcal{S}_{\mathrm{Japan}}} \mathbb{I}[g(s) = g]. \tag{18}$$

C.3 Implementation Details

We use gpt-5.6-luna as the shared LLM backbone for EncounterMob, LLMob, and ELLMob.

For EncounterMob, the maximum number of refinement rounds is set to $K = 3$ on both Japan and NYC. For LLMob, we use the learning-based motivation retrieval variant (LLMob-L). For ELLMob, we use the Normal setting with weekday/weekend context and retain the released implementation's limit of three reflection attempts.

Retry policy. For EncounterMob, failed generation requests are retried until an output passes parsing and validity checks.

D Additional Experimental Results

D.1 Encounter Distributions across Venue Categories

Figure 6 complements the main-text category comparison with spatial distributions of encounters in transportation, dining, retail, and entertainment venues. These maps show that the discrepancies of the baselines include both over-generation and under-generation across encounter scenarios.

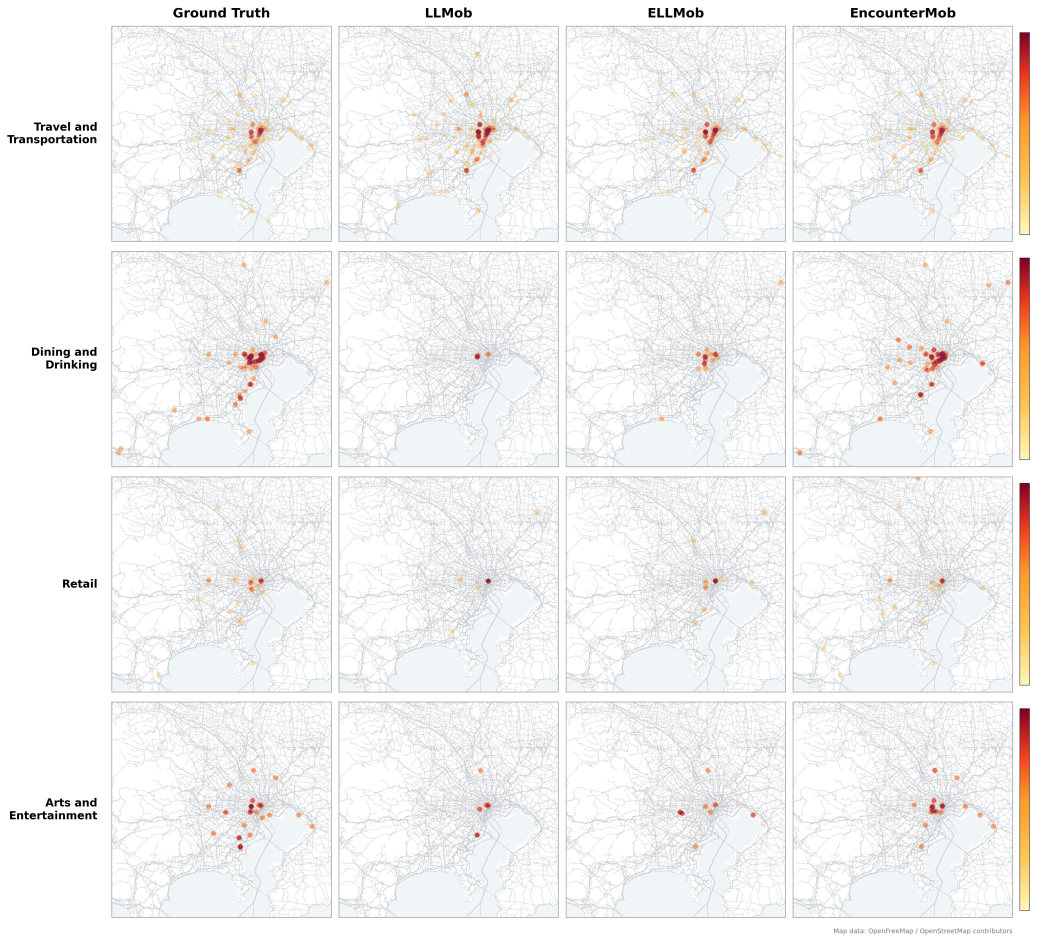


Figure 6: Spatial distributions of actual encounters in Greater Tokyo over the same 30-day evaluation period. Rows represent venue categories, and columns compare ground truth with encounters extracted from the final trajectories generated by each method. Panels within each row share the ground-truth count denominator and color scale.

D.2 Token Consumption

Table 6 records the token cost in Japan and NYC.

D.3 Examples

We illustrate one recorded NYC run on November 17, following residents A and B through initialization and refinement. Their histories contain coffee, dining, and evening leisure activities, including prior co-occurrences in bar or coffee settings. The two cards show successive stages of the same run. Counts for C and Q refer to the full day.

Table 6: Inference token usage of the retained runs. M denotes million tokens; k denotes thousand tokens. Per-day usage is normalized by active resident-days (RD). † marks a calibrated estimate; — indicates unavailable token accounting.

Dataset	Method	Active RD	Total tokens (M)	Tokens/RD (k)
	ELLMob	14,279	130.058	9.11
	EncounterMob	12,887	587.586	45.60
	ELLMob	2,017	14.645	7.26
	EncounterMob	2,017	49.823	24.70

Example: Initial generation

City-level encounter rhythms (C): Three coarse allocations: Restaurant at 14:00, Bar at 19:00, and Italian Restaurant at 21:00.

Encounter proposals (Q): Three proposals are instantiated. The proposal for A and B places them at Bar v2171 at 22:30.

Personal trajectories generated:

A: 11:10 v2003 → 15:20 v1548 → **15:50 Coffee Shop (v2019)** → 16:30 v2024.

B: 00:50 v1998 → **15:50 Coffee Shop (v2019)** → 16:00 v2020.

Realized encounter: Both independently generated trajectories contain Coffee Shop v2019 at 15:50. They realize this coffee-shop co-occurrence, rather than the proposed 22:30 bar encounter.

Example (continued): Critic feedback and refinement

Joint Critic feedback: `REVISE`. Retain the supported coffee-shop stops and make their realized encounter available to the proposal update. Historical preferences do not justify forcing unsupported precise proposals.

Personal trajectories: Both residents' complete daily sequences remain unchanged, including the shared 15:50 stop.

Updated rhythms (C): Two coarse allocations remain: Restaurant at 18:00 and Bar at 21:00. They express broad tendencies rather than required events.

Updated proposal (Q): One proposal remains: A and B at Coffee Shop v2019 at 15:50. The exact intersection of their trajectories provides more direct support than the coarse evening tendencies.

Final assessment: `ACCEPT`. The run ends after one refinement round with one proposal and one exactly corresponding realized encounter. Proposal–trajectory agreement improves; the realized encounter itself is unchanged.

Retrospective comparison. The observed records place A and B together at a Burger Joint at 14:50. The initial bar proposal differs by 460 minutes and approximately 6.26 km; the revised coffee-shop proposal differs by 60 minutes and approximately 1.70 km. The revised proposal is closer but does not match exactly. These observations are used only for this retrospective comparison, not for the Critic's decision.

Two additional runs illustrate how city rhythms become concrete encounters and how local trajectory revisions realize them. Resident labels are local to each example. The cards summarize saved inputs and outputs; R0 denotes initialization and R1–R3 denote refinement rounds.

Example: From city rhythms to concrete encounters

Context: NYC, December 29, 2012. The city module allocates two encounters: one at an American Restaurant in the 20:00 hour and one at a Bar in the 22:00 hour.

Example: From city rhythms to concrete encounters (continued)

Historical support: A and B previously co-occurred at American Restaurant v4179 on Dec. 21 at 22:10. C and D previously co-occurred at Bar v869 on Dec. 27 at 22:40. Both observations precede the target day.

Concrete proposals (Q): The encounter module proposes A and B at v4179 at **20:00**, and C and D at v869 at **22:40**. The participant sets and venues have historical support; the restaurant proposal uses an earlier time than the cited historical encounter. Both proposals fall within their respective city-level hour/category allocations.

Realization across rounds: Neither proposal is realized by the R0 trajectories. Both are realized at their exact proposed participant sets, times, and venues in R1, and remain realized through R3. The full-day Critic decision becomes **ACCEPT at R3**.

Mechanism illustrated: Coarse city-level allocations are resolved into different participant sets and specific venues, with timing adapted where needed. Subsequent personal refinement makes those proposals appear in the generated trajectories.

Example: Realizing an encounter with a local revision

Context: Japan, July 1, 2012. The encounter module proposes A and B at **14:00, Café v18324**. Their supplied history contains a co-occurrence at this café at 14:00 on June 3, before the target day.

Initial trajectories (R0):

A: 08:40 Train Station (v1764) → 16:00 Men's Store (v10582) → 19:20 Train Station (v1764).

B: 14:00 Café (v18324).

The proposal is not realized because A has no corresponding café stop.

Critic feedback (paraphrased): Add a stop at v18324 at 14:00 to realize the encounter with B; the other existing stops can be retained.

Revised trajectories (R1): A adds **14:00 Café (v18324)** while preserving all three original stops at their original times and venues. B's complete trajectory remains unchanged. The proposed encounter is now realized and persists through R3.

Scope of the outcome: This local inconsistency is resolved, although the full-day Critic decision remains **REVISE at R3**. The example illustrates targeted encounter realization with a single added stop, rather than a full rewrite of either resident's day.