Supplementary Information for: Theory-Informed Generative Agents for Human Mobility Modeling

Haoyang Li[1], Runzhou Liu[1], Yao Li[2,3], Amy Wesolowski[4], Sen Pei[5], and Hongru Du[1,*]

[1]Department of Systems & Information Engineering, University of Virginia, Charlottesville, VA, USA
[2]Department of Earth, Environmental, and Geographical Sciences, University of North Carolina at Charlotte, Charlotte, NC, USA
[3]Center for Applied Geographic Information Science, University of North Carolina at Charlotte, Charlotte, NC, USA
[4]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[5]Department of Environmental Health Sciences, Columbia University, New York, NY, USA
[*]Email: hongrudu@virginia.edu

1 Supplementary Data

1.1 Classification and Aggregation of POI

We aggregate POIs from the North American Industry Classification System (NAICS)[1] into six functional categories to represent residents' core activity spaces. This aggregation was grouped based on economic sectors with similar mobility functions and travel motivations. For instance, "Wholesale Trade, Retail Trade, Transportation and Warehousing" are merged into "Retail & Trade" because both involve the acquisition of consumer goods and commercial centers. This semantic simplification allows generative agents to reason about core life function needs, rather than fragmented industrial sectors. The detailed mapping is provided in Supplementary Table 1.

Supplementary Table 1: Mapping of NAICS sector codes to functional activity categories. The table categorizes POIs into six core functional categories according to NAICS sectors. This classification is used to construct the agents' semantic activity space and environmental context based on the primary mobility function of each sector.

NAICS Sector Code	NAICS Sector Full Name	Aggregated Category	Logic of Aggregation
42, 44, 45, 48, 49	Wholesale Trade; Retail Trade; Transportation and Warehousing	Retail & Trade	Commercial logistics and consumer purchasing centers.
61	Educational Services	Education	Routine educational activities and campus-based trips.
62	Health Care and Social Assistance	Health Care	Necessary medical services and essential social care.
71	Arts, Entertainment, and Recreation	Arts & Recreation	Discretionary leisure, cultural events, and recreational activities.
72	Accommodation and Food Services	Food & Accommodation	Social dining, hospitality, and catering services.
All others	Mining, Utilities, Finance, Professional Services, etc.	Others	Diverse professional, industrial, and public services.

1.2 City-related Information

Socio-demographic characteristics of the study area are derived from the 2019 5-year American Community Survey (ACS)[2]. Spatial structure metrics, including administrative land areas and CBG definitions, are sourced from the 2019 U.S. Census Gazetteer Files[3] and Shape Files [4]. Environment data, including the location and categories of POIs, were acquired from the SafeGraph dataset[5]. We utilize the 2019 data as the primary temporal baseline to ensure consistency with empirical mobility records and to capture urban structure prior to the COVID-19 pandemic.

Supplementary Table 2: Summary of spatial, functional, and socio-demographic characteristics across the four study areas. Data are aggregated at city level from the 2019 5-year ACS[2] (demographics and occupation), 2019 U.S. Census Gazetteer Files[3] (land areas), Shape Files [4] (CBGs) and SafeGraph dataset[5] (POI distributions). Socio-economic classifications represent the percentage of CBGs in each tier based on city-level thresholds as demonstrated in Supplementary Information Section 2.1. POI category distributions are calculated using the SafeGraph Core POI dataset and the mapping defined in Supplementary Information Section 1.1. For brevity, Norfolk-Virginia Beach is abbreviated as Norfolk-VB in this table.

Metric	New York City	Chicago	Orlando	Norfolk-VB
Spatial Structure				
Total Population	8,059,011	2,527,136	386,914	677,026
Number of CBGs	6,493	2,183	144	491
Total POI Locations	57,883	43,957	19,788	33,346
Land Area (km^2)	778.0	588.9	286.3	771.8
Occupation Industry. (%)				
Agri., forestry, fishing, hunting, and mining	0.1	0.2	0.4	0.2
Construction	5.1	4.0	6.4	6.7
Manufacturing	3.2	8.2	4.2	6.0
Wholesale trade	2.1	2.3	2.4	1.9
Retail trade	9.1	8.5	11.4	11.4
Transp., warehousing, and utilities	6.5	6.9	6.4	4.4
Information	3.8	2.2	2.2	1.7
Finance, real estate, and rental	9.4	8.3	7.4	7.1
Prof., scientific, and admin. services	13.9	16.6	14.7	12.4
Edu. services, health, and social care	26.9	23.0	17.5	23.0
Arts, entertainment, and food services	10.7	10.8	19.5	11.6
Other services (excl. public admin.)	5.3	4.9	4.7	4.6
Public administration	3.8	4.1	2.7	9.1
POI Distribution (%)				
Retail & Trade	24.1	20.5	24.8	22.7
Education	3.1	3.8	3.1	3.3
Health Care	15.7	13.1	11.2	13.7
Arts & Recreation	6.0	5.1	4.2	6.8
Food & Accommodation	20.5	19.3	18.0	17.4
Others	30.5	38.2	38.7	36.2
Demographics (Census %)				
Sex (Male / Female)	47.7 / 52.3	48.6 / 51.4	48.4 / 51.6	50.2 / 49.8
Age Group (<18 / 18-60 / >60)	20.8 / 59.0 / 20.2	20.9 / 61.4 / 17.7	21.3 / 62.6 / 16.1	21.4 / 60.4 / 18.2
Race (White / Black / Other)	44.3 / 25.2 / 30.4	51.6 / 30.4 / 18.0	63.2 / 25.7 / 11.0	62.8 / 28.3 / 8.9

1.3 Computational Cost

Supplementary Table 3: Computational scale and cost. The table summarizes the simulation scale and the associated costs for LLM inference. To optimize computational efficiency, we group the synthetic population into unique demographic profiles ($\Theta_u = \{\mathbf{v}_u, \mathscr{H}_u\}$, as provided in Equation 5 and 6). LLM reasoning is performed once per unique demographic profile using Gemini-2.5-Pro, and the derived behavioral rules are applied to all agents sharing that profile ($n = 10$ agents per CBG). All API calls were directed to the Google Cloud without any fine-tuning.

Cities	Total Simulated Agents	Unique Agent Types	Approx. Cost (USD)
New York City	64,930	1,569	$\approx$ \$80.00
Chicago	21,830	1,324	$\approx$ \$67.52
Orlando	1,440	466	$\approx$ \$23.77
Norfolk-Virginia Beach	4,910	650	$\approx$ \$33.15
Total	**93,110**	**3,009**	$\approx$ **\$ 204.44**

1.4 POI Mobility Estimation

Distance calculation. To determine the spatial impedance of travel, we use the Haversine distance[6] to calculate the distance d between the origin centroid and the destination coordinates. This formula accounts for the Earth's curvature, providing a more accurate measure of urban mobility than Euclidean distance. For any two points with latitude ϕ_1, ϕ_2 and longitude λ_1, λ_2, the distance is defined as:

$$d = 2R \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right), \tag{1}$$

where R is the Earth's radius (approximately 6,371 km). In the analysis, we also exclude records with $d > 100$ km to filter out non-urban long-distance travel.

Empirical data processing and comparison. The empirical mobility analysis is constructed using the SafeGraph dataset[5], which provides aggregated visit records between residential home CBGs and visited POIs. Each destination POI is identified by its unique SafeGraph place ID, allowing us to link travel records with geographic coordinates, destination CBGs, and POI categories as defined in Supplementary Information Section 1.1. To ensure comparative consistency between empirical observations and simulation outcomes, we evaluate all mobility patterns using a home-based distance calculation. Although agents in the simulation actually perform continuous trajectories, their mobility behaviors are assessed based on the centroid of the visitor's residential home CBG to the destination POI coordinates, thus maintaining alignment with the record format of the SafeGraph data.

We apply a unified population-weighting method to keep statistical parity. For the SafeGraph empirical data, we apply a scaling factor k_c to each CBG c to adjust the device sample to the total census population (Pop_c):

$$k_c = \frac{\text{Pop}_c}{\text{Dev}_c}. \tag{2}$$

where Dev_c is the number of unique devices residing in that area. Correspondingly, in the TIMA and baseline simulation environment where $n = 10$ agents are instantiated in each CBG (see Supplementary Information Section 2.1 for details), each agent is assigned a statistical weight w to match the actual neighborhood CBG visit density:

$$w_u = \frac{\text{Pop}_c}{n}. \tag{3}$$

This normalization ensures that all visit densities, flow matrices, and socio-economic indicators are aggregated and compared at the CBG level, thus representing the empirical records and simulated results in the same physical magnitude.

1.5 Situational Context

To simulate mobility adaptation to external shocks, we construct a situational context based on the urban and policy environment of Manhattan, New York, during the early stages of the COVID-19 pandemic. Governmental mandates are derived from the Oxford COVID-19 Government Response Tracker (OxCGRT)[7], focusing on the "New York State on PAUSE" executive order implemented on March 22, 2020. The situational context included specific regulatory constraints, including the recommended closure of non-essential businesses like bars, gyms, theaters, and shopping malls; the restriction of the food service to takeout and delivery only; the prohibition of non-essential gatherings; and enforcing social distancing protocols requiring individuals to maintain a distance of at least 6 feet from each other. Detailed information on the specific inputs used in our simulation is provided in Supplementary Information Section 2.8.

Statistical information regarding the epidemiological environment is supplied using real-time surveillance data from the JHU CSSE dashboard[8]. The context specifies that New York City had over 12,000 laboratory-confirmed cases and around 100 fatalities at the start of the simulation, while noting an exponential growth trend of infections doubling every few days. To ensure the generative agents reason strictly based on the provided information rather than retrieving the training corpus, the virus name (i.e., COVID-19) is omitted from all prompts. This situational description is balanced by essential mobility demands and allowance aligning with real-world needs, which informs the agents that leaving their home was still permitted for necessity, such as purchasing food and pharmaceuticals, or for solitary physical activity in public parks. To assess TIMA's predictive accuracy, we compare the simulation results with SafeGraph empirical data from the week of March 23, 2020, which reflects the behavioral restructuring that occurred immediately after the implementation of these policy measures.

2 Supplementary Method

2.1 Agent initialization

Demographic Synthesis via IPF: To represent the socio-demographic heterogeneity of the urban population, we synthesized representative agents for each CBG using IPF. This algorithm reconstructs the joint probability distribution of four key demographic dimensions: sex $\in$ Male, Female, age $\in$ {<18. 18$-$60, >60}, race $\in$ {White, Black, Others}, and industry of occupation $\in$ {13 categories}, including:

- Agriculture forestry fishing and hunting and mining
- Construction
- Manufacturing
- Wholesale trade
- Retail trade
- Transportation and warehousing and utilities
- Information
- Finance and insurance and real estate and rental and leasing
- Professional scientific and management and administrative and waste management services
- Educational services and health care and social assistance
- Arts entertainment and recreation and accommodation and food services

- Other services except public administration
- Public administration

The synthesis procedure utilizes marginal aggregates extracted from ACS. Specifically, we prepare three sets of marginal distributions for each CBG: 1) joint sex-age marginal distribution, 2) race marginal distribution, and 3) joint sex-industry marginal distribution. These distributions provide the necessary constants to ensure that the synthetic population reflects the demographic composition of each CBG.

The IPF algorithm starts with an initial uniform distribution X^0, representing all possible combinations of demographic attributes. In each iteration k, the cell counts are sequentially adjusted by a scaling factor f to match each target marginal distribution M_d:

$$X_i^{k+1} = X_i^k \times \frac{M_{d,j}}{\sum_{i \in j} X_i^k}, \tag{4}$$

where $M_{d,j}$ is the target total for the j-th category of dimension d. This process continues until the relative change in cell counts falls below a tolerance of 1×10^{-4}. After the joint distribution $\mathbb{P}$(sex, age, race, occupation) is estimated for each CBG, we perform weighted sampling without replacement to extract the core demographic feature set for 10 discrete agents per CBG.

Discretizations of CBG level socio-economic indicators: In addition to individual-level demographic features derived from probability sampling, we also integrate CBG-level socio-economic indicators, including home neighborhood CBG's income and education levels, which are directly extracted from the metadata of the residential CBG. To facilitate semantic reasoning by the LLM, these continuous variables are discretized into a three-level scale: {Low, Medium, High}. The classification of the income level is determined by the median household income of the agent's home CBG, while the education level is defined by the proportion of residents holding a bachelor's degree or higher. We use the city-wide tertiles of these two distributions as objective thresholds to ensure the socio-economic classification is contextually consistent across the metropolitan area.

Assembly of the agent persona: The final agent profile vector $\mathbf{v}_u$ is constructed by concatenating the individual-level demographic features from the IPF process with discretized CBG-level socio-economic indicators (income_u and education_u for the agent's home CBG's income and education level, respectively). Formally, for each agent u, this vector is defined as:

$$\mathbf{v}_u = [\text{sex}_u, \text{age}_u, \text{race}_u, \text{occupation}_u, \text{income}_u, \text{education}_u] \tag{5}$$

Construction of the home environmental context $\mathscr{H}_u$: The residential context $\mathscr{H}_u$ encapsulates the functional characteristics of the agent's home neighborhood CBG. This vector informs the generative agent of local resource availability and typical distribution, thus facilitating the grounding of behavioral preference generation.

To maintain computational feasibility while preserving individual-level context, we implement a representative sampling during the LLM inference phase. We first categorize the synthetic population into unique demographic types based on the intersection of all demographic and socio-economic attributes in $\mathbf{v_u}$. For each distinct agent type, we randomly select a representative agent and identify its home CBG c^*. Subsequently, the functional distribution of CBG c^* is used as the urban semantic context $\mathscr{H}_u$ for the entire demographic group with the persona of $\mathbf{v}_u$.

This home environment vector $\mathscr{H}_u$ is constructed by aggregating all POIs within the representative CBG c^* into six functional categories as provided in Supplementary Information Section 1.1. Let N_{i,c^*} be the count of POIs belonging to category i in CBG c^*, then the vector $\mathscr{H}_u$ can be defined as the normalized distribution:

$$\mathscr{H}_u = \left[\frac{N_{1,c^*}}{\sum_{j=1}^{6} N_{j,c^*}}, \frac{N_{2,c^*}}{\sum_{j=1}^{6} N_{j,c^*}}, \ldots, \frac{N_{6,c^*}}{\sum_{j=1}^{6} N_{j,c^*}} \right] \tag{6}$$

In cases where no POI is recorded in the sampled CBG, we assign a uniform distribution to represent a functionally neutral environment. This design ensures that the "vibe" provided to the LLM is derived from real-world urban environments, ensuring a concrete reference for generative agents to output mobility rules.

2.2 Hyperparameter configuration of local opportunity radius

The local opportunity radius (D_{max}) serves as the spatial constraint on the availability of functional opportunities in the formation of visiting intentions (Stage 1, as illustrated in Eq. 14). We set differentiated city-specific D_{max} values to ensure that generative agents can reason at a consistent opportunity spatial scale of opportunities in different urban environments, thereby eliminating the bias caused by urban morphological heterogeneity and city scale.

This setting is grounded in the rank-based mobility law[9]. The theory posits that the probability of an individual located at origin u reaching a destination v at a distance r is governed by their rank $k(v)$, defined as the number of available opportunities closer to u than v: $P(u \to v) \propto k(v)^{-\alpha}$, where α is the universal scaling exponent. In the TIMA framework, we operationalize "opportunity" as the spatial counts of POIs. To maintain a physically consistent decision-making environment, we define D_{max} such that the expected number of opportunities K within the radius is constant across different cities. Under the assumption of a locally uniform spatial distribution of POI density ρ, the rank k of an opportunity at a distance r corresponds to the cumulative count within the circular region:

$$k(r) = \int_0^r \rho \cdot 2\pi x\, dx = \rho \cdot \pi r^2 \tag{7}$$

To keep the opportunity set size K constant (i.e., $k(D_{max}) = K$), the relationship between spatial constraint D_{max} and the opportunity density ρ must satisfy:

$$D_{max} = \sqrt{\frac{K}{\pi\rho}} \quad \Rightarrow \quad D_{max} \propto \rho^{-1/2} \tag{8}$$

This relationship defines the inverse square root law between the spatial constraint and opportunity density.

We utilize Chicago ($D_{max} = 1.000$ km) as the reference baseline for a high-density urban environment in the U.S. The functional density ρ is operationalized as the number of CBGs divided by the administrative land area, with both metrics provided in Supplementary Table 2. Using these data, the functional density of Chicago was calculated as $\rho_{CHI} \approx 3.71$ CBGs/km^2, and that of New York City was $\rho_{NYC} \approx 8.35$ CBGs/km^2, with a density ratio of $\rho_{NYC}/\rho_{CHI} \approx 2.25$. Following the inverse square root law in Eq. 8:

$$D_{max}^{NYC} = D_{max}^{CHI} \times \left(\frac{\rho_{NYC}}{\rho_{CHI}}\right)^{-1/2} \approx 1.000 \times 2.25^{-0.5} \approx 0.667 \text{ km} \tag{9}$$

In our simulations, a local opportunity radius of 0.675 km was used for New York City to maintain mathematical consistency and facilitate simulation implementation. For low-density environments such as Orlando and Norfolk-Virginia Beach, the functional density is significantly reduced ($\rho_{ORL} \approx 0.50$ CBGs/km^2), approximately 7.37 times lower than in Chicago. While the scaling law theoretically suggests a local opportunity radius of approximately 2.72 km to encompass a comparable opportunity set K, we implement a saturation cap of 1.500 km. This upper limit ensures that the parameters remain within the geographic scope of a "local neighborhood" context, as an opportunity radius of 2.72 km would cover a vast area (approximately 23.2 km^2), exceeding the routine urban accessibility assumed in the visiting intention formation (Stage 1). By combining mathematical scaling with geographical scale normalization, we ensure that agents in different cities reason over a functionally consistent local opportunity set. The detailed values of the local opportunity radius across cities are provided in Supplementary Table 4.

Supplementary Table 4: Local opportunity radius (D_{max}) across cities. The local opportunity radius (D_{max}) defines the maximum distance for the initial semantic search of POIs during the exploration phase. This parameter is calibrated based on the typical urban density and transit accessibility of each metropolitan area.

Study Area	Local Opportunity Radius (D_{max})
New York City	0.675 km
Chicago	1.000 km
Orlando	1.500 km
Norfolk-Virginia Beach	1.500 km

2.3 LLM inference for heterogeneous preference

The prompt is constructed by background information and three different task instructions listed below:

Background Information

You are a resident living in a city. Your task is to define your mobility behavior by writing a Python code snippet.

[Agent Profile Context]

Your Resident Profile

- **Sex:** *<Agent Sex>*
- **Age Group:** *<Agent Age Group>*
- **Race:** *<Agent Race>*
- **Job Sector:** *<Agent Industry>*
- **Home CBG's General Education Level:** *<Home CBG Education>*
- **Home CBG's General Income Level:** *<Home CBG Income>*
- **Home CBG's Vibe (Context Only):** *<Home CBG's POI Distribution List>*

  *(Note: This shows what is currently physically around you. It sets the context, but **DO NOT** simply give high scores to POI types just because they are abundant nearby. Output your **intrinsic** interests based on your age, job, etc.)*

The POI types correspond to the list indices:

0: Wholesale & Retail Trade, Transportation and Warehousing
1: Others
2: Educational Services
3: Health Care and Social Assistance
4: Arts, Entertainment, and Recreation
5: Accommodation and Food Services

Task Instruction #1

Your Task: Define Interest Scores

Write a Python code snippet to define a list named 'scores' containing 6 floats (0.0 to 1.0).

Requirements:

1. Define a list 'scores = [...]'.
2. Add comments explaining your logic based on your profile.

Output Format Constraint:
Return **ONLY** the Python code block wrapped in "'python ... "'.
Example:

```
# As a <demographic> person...
scores = [<float>, <float>, <float>, <float>, <float>, <float>]
```

Task Instruction #2

Your Task: Define CBG Preferences
Write a Python code snippet to define a dictionary named 'cbg_preferences'.
Requirements:

1. The dictionary MUST have keys 'income' and 'race'.
2. 'income' maps 'High', 'Medium', 'Low' to a score (<float> 0.5 to 1.5).
3. 'race' maps 'White', 'Black', 'Other' to a score (<float> 0.5 to 1.5).
4. **1.0 is neutral**. Higher means preference (homophily), lower means avoidance.
5. Add comments explaining your logic.

Output Format Constraint:
Return **ONLY** the Python code block wrapped in "'python ... "'.
Example:

```
# Based on my profile...
cbg_preferences = {
   'income': {'High': <float>, 'Medium': <float>, 'Low': <float>},
   'race': {'White': <float>, 'Black': <float>, 'Other': <float>}
}
```

Task Instruction #3

Context: Mobility Reproduce
Imagine you are living in a city and you need to decide: How likely you are to **Explore** new places vs. **Return** to places you have already visited.
Definitions:

- **Explore**: Choosing to visit a brand new place you haven't been to before.
- **Return**: Choosing to revisit a place you have been to before (Preferential Return). You are more likely to return to places you visit frequently.

Your Task: Define Mobility Dynamics
Write a Python code snippet to define 'exploration_probs'.
Requirements:

1. **`exploration_probs`** (list of 6 floats in [0,1]):
   - Probability of **exploring** a new place given the number of unique places (S) visited this week.
   - As S increases, exploration decreases.
   - Provide values for:

$S = 0$ Start of the week (usually high).
$S = 5$ You know 5 places.
$S = 10$ You know 10 places.
$S = 15$ You know 15 places.
$S = 20$ You know 20 places.
$S \geq 25$ You know 25+ places (routine likely established).

2. **Add comments**: include brief Python comments (#) explaining your choices.

Output Format Constraint:
Return **ONLY** the Python code block wrapped in "'python ... "' containing the definitions of the two variable.
Example:

```
# Exploration probabilities based on visited count (S)
# S=0, S=5, S=10, S=15, S=20, S>=25
exploration_probs = [<float>, <float>, <float>, <float>, <float>, <float>]
```

2.4 Exploration and preferential return

At the beginning of each discrete time step $t \geq 1$, agent u occupies a location (POI) $l_u^{t-1} \in L$ and chooses the next visited location $l_u^t \in L$. Each POI l has an associated category label $k(l) \in \{1, \ldots, K\}$ and is contained in a CBG $c(l) \in C$, where C is the set of all CBGs. The agent maintains a trajectory history and associated sufficient statistics, which determine whether the agent explores a new location or preferentially returns to a previously visited one.

History representation. We represent the ordered visit history of agent u up to time t as $\mathscr{T}_u^t := (l_u^0, l_u^1, \ldots, l_u^t)$. The corresponding set of distinct visited locations is $S_u^t := \text{Unique}(\mathscr{T}_u^t) = \{ l \in \mathscr{L} : \exists \tau \leq t \text{ such that } l_u^\tau = l \}$, and we denote the number of distinct visited locations by $|S_u^t|$. For each location $l \in \mathscr{L}$, the cumulative visit count up to time t is

$$f_{u,l}^t := \sum_{\tau=0}^{t} \mathbf{1}[l_u^\tau = l]. \tag{10}$$

These quantities $(S_u^t, s_u^t, f_{u,l}^t)$ are updated deterministically from the trajectory $\mathscr{T}_u^t$ and serve as input for the agents' explore or return decisions.

Exploration probability. At time t, agent u explores a new location (i.e., selects $l_u^t \notin S_u^{t-1}$) with probability

$$P_{u,e}^t = \mathscr{P}_u(|S_u^{t-1}| \mid \mathbf{v}_u, \mathscr{H}_u, \mathscr{S}), \tag{11}$$

where $\mathbf{v}_u$ is the agent persona, $\mathscr{H}_u$ is the home semantic context, and $\mathscr{S}$ is the (optional) situational context for the counterfactual analysis. The function $\mathscr{P}_u(\cdot)$ is an agent-specific exploration profile inferred by the LLM agent using Task Instruction #3 once per agent (and cached for reuse during simulation). The TIMA defines specific exploration probabilities based on discrete ranges of previously visited history, using thresholds at $S = 0$, 5, 10, 15, 20, and 25 or more locations, aiming to capture the agent's individual mobility capacity, daily needs, and intrinsic preference for exploration. The saturation threshold of $S = 25$ is selected according to empirical observations that people usually maintain a routine visiting set of approximately 25 active locations[10]. This design preserves the core EPR intuition while enabling heterogeneous exploration tendencies across agents with different profiles.

Preferential return distribution. If the agent does not explore at time t (with probability $1 - P_{u,e}^t$), the agent returns to a previously visited location by sampling from S_u^{t-1} with probability proportional to historical visit counts:

$$P_u^t(l) = \frac{f_{u,l}^{t-1}}{\sum_{l' \in S_u^{t-1}} f_{u,l'}^{t-1}}, \qquad l \in S_u^{t-1}. \tag{12}$$

We then sample $l_u^t \sim P_u^t(\cdot)$, based on equation above.

Edge cases and initialization. At $t = 1$, if the agent has visited only the initial location l_u^0, then $S_u^0 = \{l_u^0\}$ and $f_{u,l_u^0}^0 = 1$. In this case, the return distribution in Eq. (12) is degenerate (the agent returns to the initial location) unless exploration occurs. In practice, exploration is typically nonzero for small s_u^{t-1} via Eq. (11), allowing the trajectory to expand beyond the initial home location.

2.5 Semantic destination selection during exploration

When agent u explores at time t, the destination choice should reflect both (i) heterogeneous intent (what the agent wants to do) and (ii) external opportunity and feasibility constraints (what is accessible). We implement exploration as a two-stage choice: (1) select a POI category k under local opportunity constraints; (2) select a specific destination POI l (equivalently, its containing CBG $c(l)$) using a semantic-gravity mechanism that combines socioeconomic affinity with distance friction.

Stage 1: Opportunity-constrained POI category selection. Let $w_{u,k} \in [0,1]$ denote agent u's semantic interest score for category k. The scores $\{w_{u,k}\}_{k=1}^K$ are inferred by the LLM Φ_θ using Task Instruction #1, conditioned on the agent persona $\mathbf{v}_u$, home context $\mathscr{H}_u$, a textual description of category k, and situational context $\mathscr{S}$ (when applicable). To ensure feasibility, we restrict attention to categories for which at least one POI is locally accessible within a radius $D_{\max}$ from the agent's current position. Specifically, define the locally accessible candidate set for category k as:

$$\Omega_k(l_u^{t-1}, D_{\max}) = \left\{ l \in \mathscr{L} : \; k(l) = k, \;\; d_{l_u^{t-1},l} \le D_{\max} \right\}, \tag{13}$$

and the corresponding opportunity count $|\Omega_k(l_u^{t-1}, D_{\max})|$, which quantifies the number of locally accessible POIs of category k within the distance constraint. The probability that agent u selects category k during exploration is:

$$P_u^t\big(k \mid l_u^{t-1}\big) = \frac{w_{u,k} \cdot \mathbf{1}\big[|\Omega_k(l_u^{t-1}, D_{\max})| > 0\big]}{\sum_{k'=1}^K w_{u,k'} \cdot \mathbf{1}\big[|\Omega_{k'}(l_u^{t-1}, D_{\max})| > 0\big]}. \tag{14}$$

This formulation preserves semantic preference through $w_{u,k}$ while enforcing an opportunity constraint through the indicator $\mathbf{1}[|\Omega_{k'}(l_u^{t-1}, D_{\max})| > 0]$.

Stage 2: Semantic-gravitational location selection. Conditional on the selected category k, the agent samples a destination POI l from the global set of all POIs belonging to category k in the city $\mathscr{L}_k = \{l \in \mathscr{L} : k(l) = k\}$ using a gravity-style rule:

$$P_u^t!\left(l \mid k, l_u^{t-1}\right) \propto A_{u,c(l)}; d_{l_u^{t-1},l}^{-\alpha_u}, \qquad l \in \mathscr{L}_k. \tag{15}$$

Here, $\alpha_u > 0$ is the distance-decay exponent (capturing travel friction) and $A_{u,c(l)} \in [0.5, 1.5]$ is the socioeconomic affinity between agent u and the CBG $c(l)$ containing destination l (capturing social preference/avoidance). Consistent with prior research showing that social segregation is primarily shaped by socioeconomic and ethnic factors[11,12], this affinity weight is parameterized for each agent-CBG pair along two corresponding dimensions, household income level and racial composition, with parameter values informed by the LLM. Specifically, LLM generates preference scores for different income levels (i.e., High, Medium, Low) and racial categories (White, Black, Others) based on the agent's own demographic profile $\mathbf{v}_u$ and home environmental context $\mathscr{H}_u$. These scores are anchored at a neutral baseline of 1.0, where values greater than 1.0 indicate socioeconomic affinity, and values less than 1.0 reflect social avoidance or visiting penalty. This agent–CBG affinity parameterization is generated via prompt Task Instruction #2. The final value of $A_{u,c(l)}$ is calculated as the product of these two-dimensional scores. In the baseline (non-counterfactual) simulation, we set $\alpha_u = 2.0$ to align with standard gravity scaling and to ensure comparability with EPR-based baselines; in counterfactual scenarios, α_u may be inferred as a function of $(\mathbf{v}_u, \mathscr{H}_u, \mathscr{S})$ (see Section 2.8 for details).

Integration of local intent and global destination choice We decouple the triggering of activity intent from the spatial selection of destinations. While the intent to visit locations belongings to category k is informed by the availability of local opportunities defined by $D_{\max}$ (category-availability check in Eq. (14)), the subsequent choice of a specific location is made within the city-wide candidate set $\mathscr{L}_k$ and follows the gravity-style sampling in Eq. (15). This mechanism ensures that mobility decisions remain grounded in local resource context while preserving the physical plausibility of travel distributions (as governed by the distance-decay exponent α_u) across the entire urban space.

Summary. The two-stage exploration destination selection can be written compactly as a product of (i) an opportunity-constrained semantic category preference and (ii) a conditional semantic-gravity destination preference:

$$P_u^t\big(l_u^t = l \mid l_u^{t-1}, \text{explore}, \Theta_u\big) = \underbrace{P\big(k_u^t = k \mid l_u^{t-1}, \Theta_u\big)}_{\text{POI-category preference}} \cdot \overbrace{P\big(l \mid k(l) = k, l_u^{t-1}, \Theta_u\big)}^{\text{CBG socioeconomic preference}}. \tag{16}$$

The corresponding pseudocode is provided in Algorithm 1.

Algorithm 1 TIMA

Require: Agents U; POIs $\mathscr{L}$ with $k(l), c(l)$; distance $d(\cdot,\cdot)$; Time step T
Require: For each agent u: start l_u^0, cached $\mathscr{P}_u$, $\{w_{u,k}\}$, $\{A_{u,c}\}$, α_u
Require: Radius $D_{\max}$

for all $u \in U$ **do**
  $\mathscr{T}_u^0 \leftarrow l_u^0$; $S_u^0 \leftarrow \{l\}$; $f_{u,l}^0 \leftarrow 1$
  for $t = 1$ to T **do**
    if Bernoulli$(P_{u,e}^t) = 0$ **then** ▷ Preferential return
      Sample $l' \in S_u^t$ with Equation 12
    else ▷ Explore
      Build $\Omega_k = \{l' \in \mathscr{L} : k(l') = k,\ d_{l_u^{t-1},l'} \le D_{\max}\}$ for each k
      Sample k' with $\Pr(k') \propto P_u^t\big(k' \mid l_u^{t-1}\big)$ (Equation 14)
      Define $\mathscr{L}_{k'} = \{l' \in \mathscr{L} : k(l') = k'\}$ ▷ City-wide candidates
      Sample $l' \in \mathscr{L}_{k'}$ with $\Pr(l') \propto P_u^t\big(l' \mid k', l_u^{t-1}\big)$ (Equation 15)
    end if
    $\mathscr{T}_u^t \leftarrow l'$; $f_{u,l'}^t \leftarrow f_{u,l'}^{t-1} + 1$; $S_u^t \leftarrow S_u^{t-1} \cup \{l'\}$
  end for
  Store $(\mathscr{T}_u^t, S_u^t, f_{u,l'}^t)$ for agent u
end for

2.6 Model configuration and hyperparameters

To capture individual-level behavioral heterogeneity and ensure the following of physical law, we adopt a profile-conditional parameterization without global constants like uniform distance decay exponent in classical mechanistic models[13,14], and the core parameters in TIMA are dynamically derived by LLM based on each agent's demographic information. The detailed definitions for all key parameters are provided in Supplementary Table 5.

Supplementary Table 5: Key model parameters.

Parameter	Description	Configuration Rationale
LLM-Derived Behavioral Parameters or Functions		
α_u	Personalized distance decay exponent is defined as the product of the classical EPR constant (2.0) and an agent-specific heterogeneity multiplier. This multiplier is inferred by the LLM from the agent's profile Θ_u to reflect an individual's sensitivity to travel distance. In the default setting, α_u is fixed at 2.0 for all agents; it varies across agents only in counterfactual analyses.	Anchors the model's spatial and distance constraints to the universal scaling law[15] while also ensuring individual heterogeneity, thereby capturing the differences in travel intentions among different populations under the laws of physics.
$\mathscr{P}_u$	The personalized exploration probability function governs the likelihood of visiting a new location given counts of previously visited distinct places ($\mathscr{P}_u \in [0,1]$). This function is generated by an LLM conditioned on the agent profile Θ_u and the number of distinct locations previously visited $\lvert S_u^t \rvert$.	Replaces the fixed exploration decay form in classical EPR models ($\rho S^{-\gamma}$) with an LLM-parameterized function that captures empirically observed population heterogeneity in exploration propensity reported in mobile phone–based mobility[16].
$w_{u,k}$	The agent-specific semantic interest score with a range of $[0,1]$, which is a weight vector representing the agent's intrinsic preference for a certain POI category. It is derived by LLM based on the agent's profile Θ_u to reflect lifestyle-driven functional needs.	Overcomes the semantic limitations of traditional physical-based models by prioritizing locations according to individual lifestyle needs and functional necessities[17]. This allows the choice of visiting destination to be driven by semantic intent rather than treating all locations as functionally equivalent.
$A_{u,c(l)}$	The socioeconomic affinity measures the agent's preference for a target CBG in terms of income level and racial composition. The range of this weight is $[0.5,1,5]$, where 1 represents a neutral attitude. It is generated by the LLM from the agent's profile Θ_u, conditioned on the socioeconomic attributes of the target CBG.	Integrates the social segregation mechanisms[18,19] into the physical mobility models by considering the social barriers and homophily that may either encourage or discourage agents' visitation to a particular area.
Environmental Parameters		
D_{max}	The local opportunity radius is a spatial threshold that defines the set of accessible visiting opportunities for the agent's semantic decision-making. It acts as a hard constraint during the formation of visiting intention, ensuring that the selection of a specific POI category to visit is controlled by its local availability within that radius.	This mechanism aligns humans' subjective intentions with objective environmental constraints and intervening opportunities[20], reflecting adaptive human mobility.

2.7 Experienced segregation

To quantify the socio-economic segregation of social interactions, we adopted the experienced segregation index[21]. This index operationalizes segregation as a dynamic process, capturing the extent to which individuals are exposed to diverse income groups during their daily travels. The calculation follows three consecutive parts: constructing the income dissimilarity matrix, calculating experienced income dissimilarity at activity sites, and deriving the aggregate experienced segregation index.

Income dissimilarity matrix (D_{ij}): This metric is based on the concept of "income distance," measuring the relative social proximity between neighborhood CBGs based on their economic rankings. Let R_i denote the rank of a CBG i among total N CBGs ordered by median household income, the income disparity $D_{i\to j}$ from a focal neighborhood CBG i to the target neighborhood CBG j is defined as the proportion of all other neighborhood CBGs that are "closer" to i than j in terms of income ranking. Formally, defining the set of mediating communities $K = \{k \mid |R_k - R_i| < |R_i - R_j|\}$, the pairwise income dissimilarity is calculated as:

$$D_{i\to j} = \frac{|K| + 0.5}{N - 1}. \tag{17}$$

This metric ranges from 0 to 1, where a higher value indicates that neighborhood CBG j is socio-economically distant from neighborhood CBG i relative to the city-wide distribution. We pre-calculate these values to form a global income dissimilarity matrix $\mathbf{D}$ of size N^2.

Experienced income dissimilarity (E_{il}): When an individual from home CBG i visits location l, they are exposed to a mixture of visitors from different neighborhoods CBGs j. The experienced income dissimilarity is defined as the weighted average of $D_{i\to j}$ for all co-present visitors. In our implementation, we construct a visitor matrix $\mathbf{V}$ of size MN (where M is the number of POIs), where the element V_{lj} represents the total weight of visits from CBG j to location l. Then, we derive a row-normalized probability matrix $\mathbf{P}$, where $P_{lj} = V_{lj} / \sum_k V_{lk}$ represents the proportion of visitors to location l that come from CBG j to location l. Subsequently, the experienced dissimilarity matrix $\mathbf{E}$ can be calculated:

$$\mathbf{E} = \mathbf{P} \cdot \mathbf{D}^T, \tag{18}$$

where element E_{li} represents the dissimilarity experienced by a visitor from home CBG i at location l.

Experienced segregation index (S): The final experienced segregation index of a neighborhood CBG or population group is defined as the complement of the average experienced dissimilarity. A value of $S = 0$ means that an individual's visiting perfectly represents the broader socio-economic diversity of the city (maximum dissimilarity exposure), while $S = 1$ indicates complete segregation within one's own income group. The global experience segregation index S_{global} is calculated by aggregating the experienced dissimilarity of all visits and weighting them by the count of visits V_{li}:

$$S_{global} = 1 - \frac{\sum_{l,i} V_{li} E_{li}}{\sum_{l,i} V_{li}}. \tag{19}$$

This index serves as our primary metric for assessing whether the simulation reproduces the stratified social interaction patterns and segregated visiting patterns observed in the empirical data.

2.8 Counterfactual Analysis for COVID-19 Pandemic.

Situational context input. The specific text used to describe the pandemic context in the prompt is as follows, where the bold black text signifies the situational context and semantic hints injected specifically for the pandemic, and the gray text indicates content consistent with the baseline scenario (see Supplementary Information Section

2.3). This narrative ensures that the agent's decisions are based on the policy and health environment observed on March 23, 2020:

Full Prompt for Counterfactual (Pandemic) Situational Context

You are a resident living in a city. Your task is to define your mobility behavior by writing a Python code snippet. (same as the baseline scenario)

[Situational Context Module]
CURRENT SCENARIO: You are located at NYC Manhattan

- **The Situation: A new contagious virus is spreading rapidly in New York City.**
- **The Stats: News reports say NYC has over 12,000 confirmed cases and around 100 deaths. The numbers are doubling every few days.**
- **The Policy: The "New York State on PAUSE" executive order took effect last week.**
  1. **Non-essential businesses are ADVISED to close (bars, gyms, theaters, malls).**
  2. **Restaurants are Take-out/Delivery ONLY.**
  3. **All non-essential gatherings of any size are ADVISED to be banned.**
  4. **Social Distancing: People are told to stay 6 feet apart.**
- **However, it is NOT a mandatory lockdown. You are not welded to your home. You are allowed to go out for groceries, medicine, and solitary exercise (walking/running in Central Park or by the river).**

[Agent Profile Context] (same as the baseline scenario)
Your Resident Profile

- **Your Sex:** <Agent Sex>
- **Your Age Group:** <Agent Age Group>
- **Your Race:** <Agent Race>
- **Your Job Sector:** <Agent Industry>
- **Your Home Neighborhood's General Education Level:** <Home CBG Education>
- **Your Home Neighborhood's General Income Level:** <Home CBG Income>
- **Your Home Neighborhood's Vibe (Context Only):** <Home CBG's POI Distribution List>
  *(Note: This shows what is currently physically around you. It sets the context, but **DO NOT** simply give high scores to POI types just because they are abundant nearby. Output your **intrinsic** interests based on your age, job, etc.)*

The POI types correspond to the list indices:
0: "Wholesale & Retail Trade, Transportation and Warehousing" **(Includes Grocery)**
1: "Others"
2: "Educational Services" **(Schools - CLOSED, but maybe food pickup or admin?)**
3: "Health Care and Social Assistance" **(Health)**
4: "Arts, Entertainment, and Recreation" **(Arts/Parks - Theaters closed, but PARKS are OPEN for relief)**
5: "Accommodation and Food Services" **(Food - Takeout only)**

[Task Instruction #1: Define Interest Scores] (same as the baseline scenario)
Write a Python code snippet to define a list named "scores" containing 6 floats (0.0 to 1.0). Define a list `scores = [...]`. Add comments explaining your logic based on your profile.

[Task Instruction #2: Define CBG Preferences]
Write a Python code snippet to define a dictionary named "cbg_preferences". The dictionary MUST have keys 'income' and 'race'. 'income' maps 'High', 'Medium', 'Low' to a score (0.5 to 1.5). 'race' maps 'White', 'Black', 'Other' to a score (0.5 to 1.5). 1.0 is neutral. (same as the baseline scenario)
Logic: During a crisis, do you stick to neighborhoods like yours (safety/familiarity)? Or do you not care?

[Task Instruction #3: Define Mobility Dynamics]
Imagine you are living in a city and you need to decide: How likely you are to Explore NEW places vs. Return to places you have already visited. Write a Python code snippet to define "exploration_probs" **and "alpha_multiplier".**
1. "exploration_probs" (List of 6 floats): Probability based on visited unique places $S = 0, 5, 10, 15, 20, \geq 25$. (same as the baseline scenario)
2. "alpha_multiplier" (Float):

- **This factor controls how much Distance stops you from visiting a place.**
- **Base Logic: Attraction = Interest / (Distance ^ (2.0 * alpha_multiplier))**
- **How to choose: 1.0 is standard. > 1.0 (e.g., 1.5, 2.0): You stay VERY LOCAL. You hate traveling far. < 1.0 (e.g., 0.5, 0.8): You are Willing to Travel. Distance doesn't bother you much.**

Dynamic behavior adaptation. Conditioned on the inputs of both the agent side and the environment side, LLM will re-decide three key behavioral choices to reflect adaptive responses in the crisis situation:

- **Personalized distance decay exponent (α_u).** To capture the changing physical friction of travel under an emergency context while ensuring theoretical consistency with baseline models, we define the personalized distance decay exponent as :

$$\alpha_u(\mathscr{S}) = a \cdot m_u(\mathbf{v}_u, \mathscr{H}_u, \mathscr{S}), \tag{20}$$

  where $a = 2.0$ represents the universal constant which is utilized in the normal period's simulation of TIMA and EPR-based baselines (e.g., EPR and d-EPR models) and is a value observed in previous empirical studies[15]. m_u is a dynamic multiplier inferred by the LLM according to each agent's demographic information $\mathbf{v}_u$, home environment $\mathscr{H}_u$, and the situational context $\mathscr{S}$. This multiplier adjusts the agent's sensitivity to travel distance between two locations $d(l_u^{t-1}, l)$ based on their vulnerability (e.g., the elderly are less likely to travel far away; people who cannot afford a car also cannot travel far away) and the described situational constraints (e.g., the requirements of social distancing and stay-at-home orders).

- **Shift on semantic interest $\mathbf{w_u}$.** The semantic interest vector for each agent is re-mapped to $w_{u,k}(\mathscr{S})$ through a "visiting demand-realistic constraint" alignment process. Each agent evaluates the necessity and accessibility of each POI category according to the restriction detailed in the situational context $\mathscr{S}$. Therefore, the distribution of the visiting interest is structurally re-distributed: preference for high-risk or prohibited locations (e.g., arts, entertainment, and recreation under closure orders) is reduced, while weights for survival-critical POIs (e.g. health and social assistance) are maintained or amplified, ensuring that the destination selection reflects a necessity-driven prioritization as in the real world under the crisis condition.

- **Context-adaptive exploration decision $\mathscr{P}_u$.** The personalized exploration tendency was recalibrated as $\mathscr{P}_u(S \mid \mathbf{v}_u, \mathscr{S})$ to reflect the agent's heterogeneous behavioral adaptations to external crisis. The TIMA framework reasons about socio-economic constraints and necessities derived from the agent's persona vector $\mathbf{v}_u$. This design aims to capture stratified responses from different social classes, where exploration probabilities for remote-capable and affluent agents may be drastically suppressed, while essential workers or low-income groups should retain a necessity-driven travel requirement.

2.9 Comparative baselines

We benchmark the performance of TIMA by comparing it with four established physical-based models. To ensure comparability, all baseline models are instantiated using the exact same synthetic population and spatial configuration as the TIMA framework, including initializing $n = 10$ agents for each CBG. The choice of four baselines is aimed to cover both macroscopic physics and microscopic social dynamics:

- **Gravity Model[22]:** A classic macroscopic model, which determines mobility flows based on population mass and spatial distance, treating human travel as a memoryless process. In gravity model, at each time step, the probability of an agent located in CBG i selecting a destination POI in CBG k to visit is given by $P_{ij} \propto (\mathrm{Pop}_i \cdot \mathrm{Pop}_k)/d_{ij}^{\alpha}$, where, the distance decay exponent α is set to 1.0 following the formulation proposed by Zipf[22], allowing that visiting frequency is inversely proportional to distance.

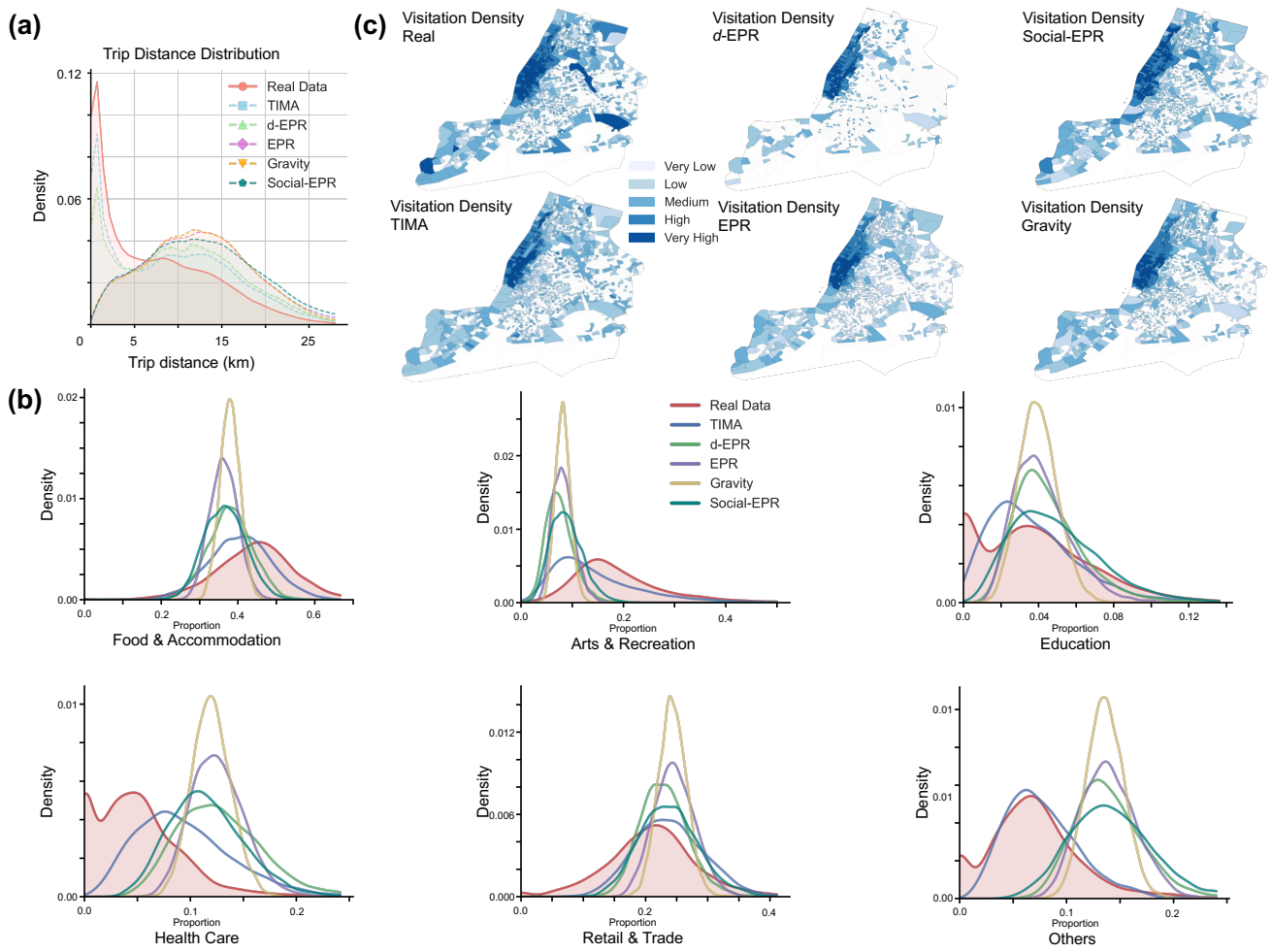
- **EPR Model[13]:** This model introduces "exploration" (the tendency to visit new locations) and "preferential return" (revisit familiar visited locations) mechanisms, where the probability of exploration is defined as $P_{new} = \rho S^{-\gamma}$, and S is the number of unique locations visited. We adopt the parameter values $\rho = 0.6$ and $\gamma = 0.21$ as established in the original study. In the exploration stage, the destination location is chosen uniformly at random from all unvisited POIs.

- ***d*-EPR Model[14]:** This model extends the EPR model by imposing spatial constraints on the exploration mechanism. While the decision to explore or return follows the same P_{new} function as the standard EPR model, the selection of a new destination j is governed by the gravity law: $P(j) \propto M_j/d_{ij}^{\alpha}$, where $\alpha = 2.0$. Due to the telecommunications data used to define the mass M_j in the original study is unavailable, we use the resident population of the destination's CBG as a reasonable proxy for location attractiveness.

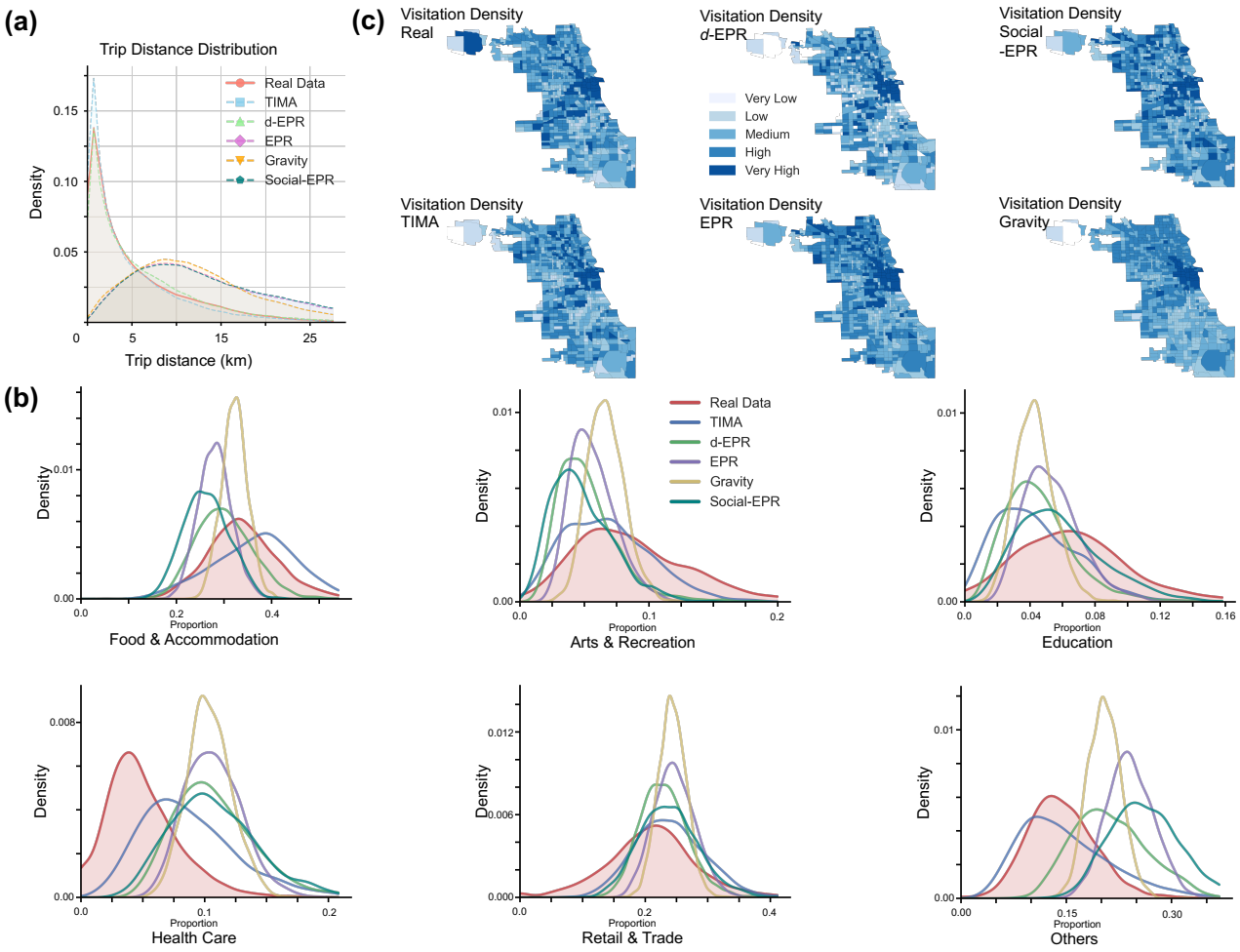
- **Social-EPR Model[19]:** Also based on the EPR model, this model refines the exploration stage with an idea of "social exploration," meaning that individuals differ in their willingness to visit locations dominated by other income groups. Agents are characterized by a social exploration rate σ_s, which dictates the probability of choosing a location where their income group is the minority. To replicate the empirical distribution observed in the original study (specifically, the σ_s distribution shown in Figure 3c of Moro et al. [19]), we sample this parameter for each agent from an income-dependent Beta distribution. Here, we use $\mathrm{Beta}(5,1)$ for high-income agents to reproduce their empirically higher social exploration tendency (distribution skewed towards 1), and $\mathrm{Beta}(2,1.5)$ for low-income agents to capture their more limited social range. In the preferential return stage, agents revisit locations proportional of their historical visit frequencies, consistent with the standard EPR model.

3 Supplementary Results

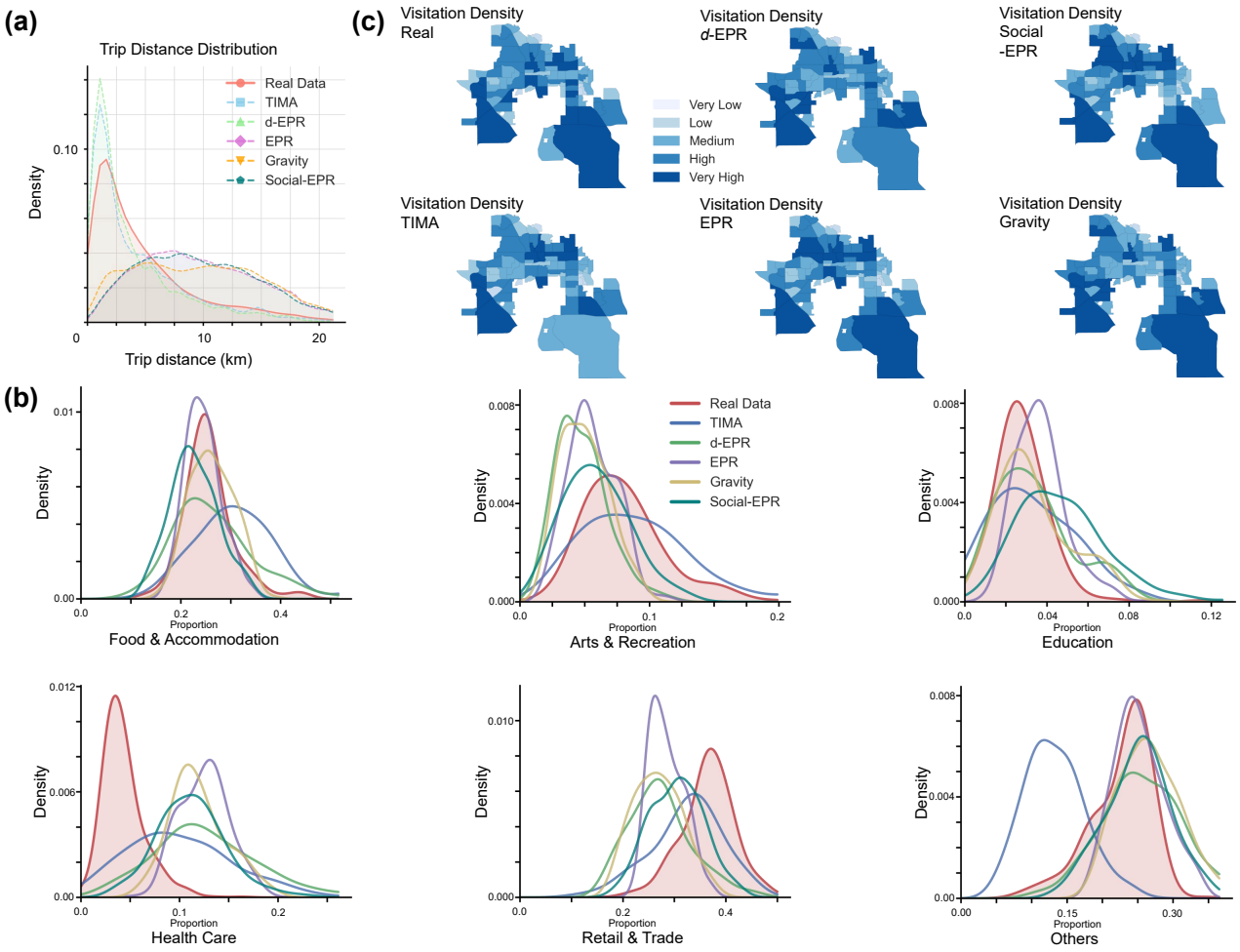
3.1 Multi-scale Mobility Patterns across Cities and Models



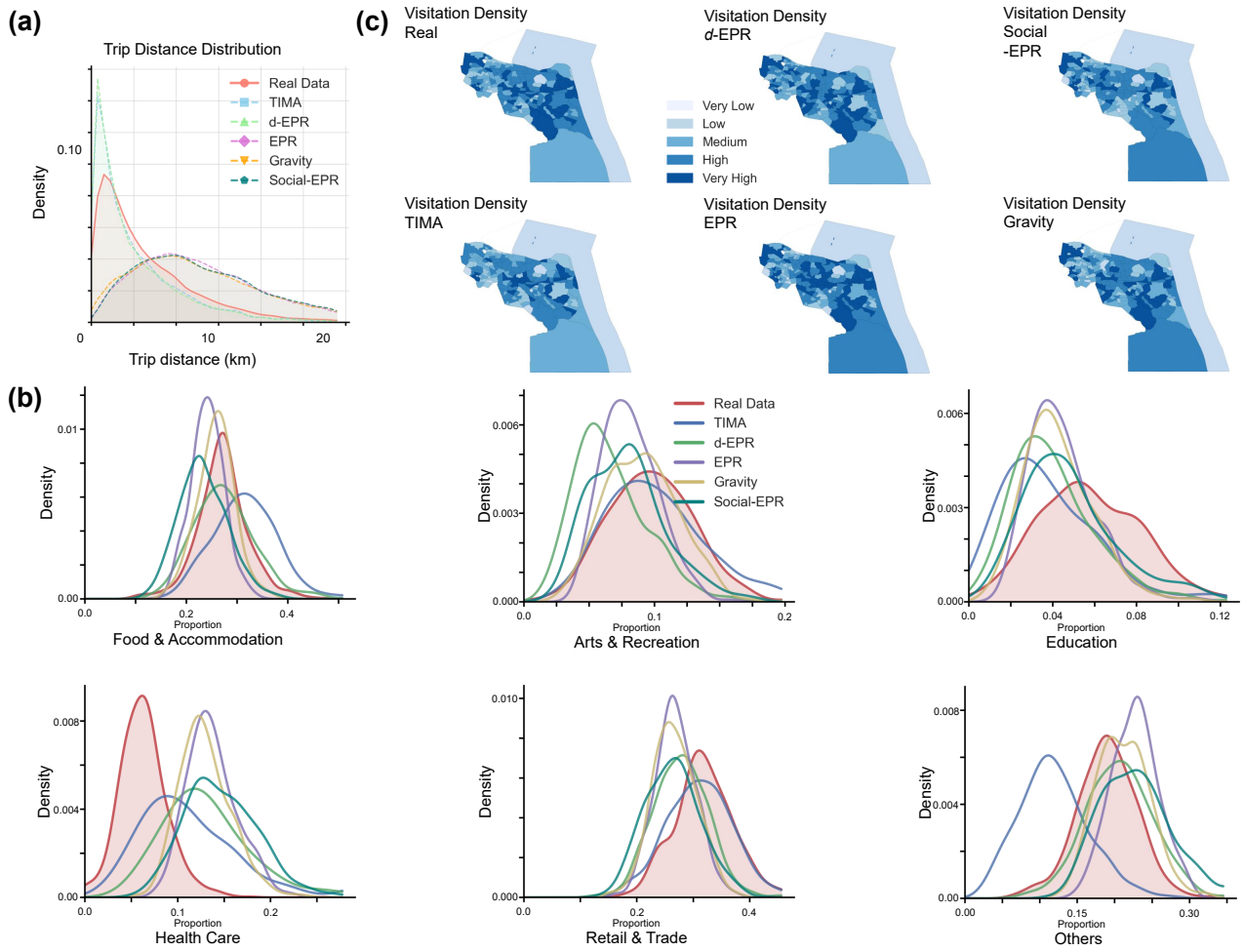
Supplementary Figure 1: Extended validation of multi-scale mobility patterns in New York City. a, Trip distance distribution (km) comparing empirical trajectories to simulations from TIMA and baseline mobility models (d-EPR, EPR, Social-EPR, and Gravity). **b,** CBG–level distributions of neighborhood visit composition (proportion of visits) across POI categories (Food & Accommodation, Health Care, Arts & Recreation, Retail & Trade, Education, and Others). **c,** Tract-level spatial visitation density maps (binned into very low–very high) for real data and each model. The color scale (very low, low, median, high, very high) is defined by the Fisher-Jenks natural breaks ($k = 5$) according to the visitation-density distribution across locations.



Supplementary Figure 2: Extended validation of multi-scale mobility patterns in Chicago. a, Trip distance distribution (km) comparing empirical trajectories to simulations from TIMA and baseline mobility models (d-EPR, EPR, Social-EPR, and Gravity). **b,** CBG–level distributions of neighborhood visit composition (proportion of visits) across POI categories (Food & Accommodation, Health Care, Arts & Recreation, Retail & Trade, Education, and Others). **c,** Tract-level spatial visitation density maps (binned into very low–very high) for real data and each model. The color scale (very low, low, median, high, very high) is defined by the Fisher-Jenks natural breaks[23] ($k = 5$) according to the visitation-density distribution across locations.

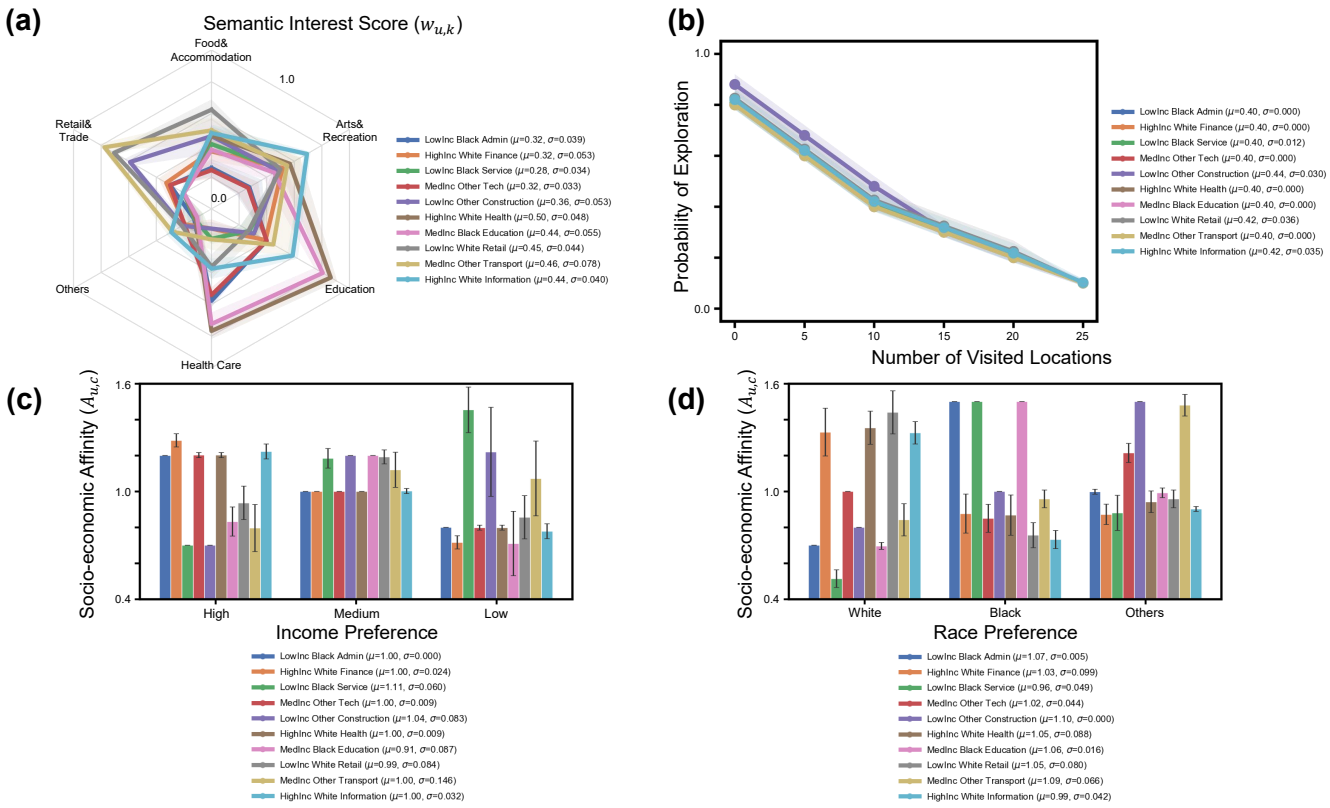


Supplementary Figure 3: Extended validation of multi-scale mobility patterns in Orlando. a, Trip distance distribution (km) comparing empirical trajectories to simulations from TIMA and baseline mobility models (d-EPR, EPR, Social-EPR, and Gravity). **b,** CBG–level distributions of neighborhood visit composition (proportion of visits) across POI categories (Food & Accommodation, Health Care, Arts & Recreation, Retail & Trade, Education, and Others). **c,** Tract-level spatial visitation density maps (binned into very low–very high) for real data and each model. The color scale (very low, low, median, high, very high) is defined by the Fisher-Jenks natural breaks[23] ($k = 5$) according to the visitation-density distribution across locations.



Supplementary Figure 4: Extended validation of multi-scale mobility patterns in Norfolk-Virginia Beach. **a,** Trip distance distribution (km) comparing empirical trajectories to simulations from TIMA and baseline mobility models (d-EPR, EPR, Social-EPR, and Gravity). **b,** CBG–level distributions of neighborhood visit composition (proportion of visits) across POI categories (Food & Accommodation, Health Care, Arts & Recreation, Retail & Trade, Education, and Others). **c,** Tract-level spatial visitation density maps (binned into very low–very high) for real data and each model. The color scale (very low, low, median, high, very high) is defined by the Fisher-Jenks natural breaks[23] ($k = 5$) according to the visitation-density distribution across locations.

3.2 Robustness analysis of agent behavioral parameters generation across different demographic profiles



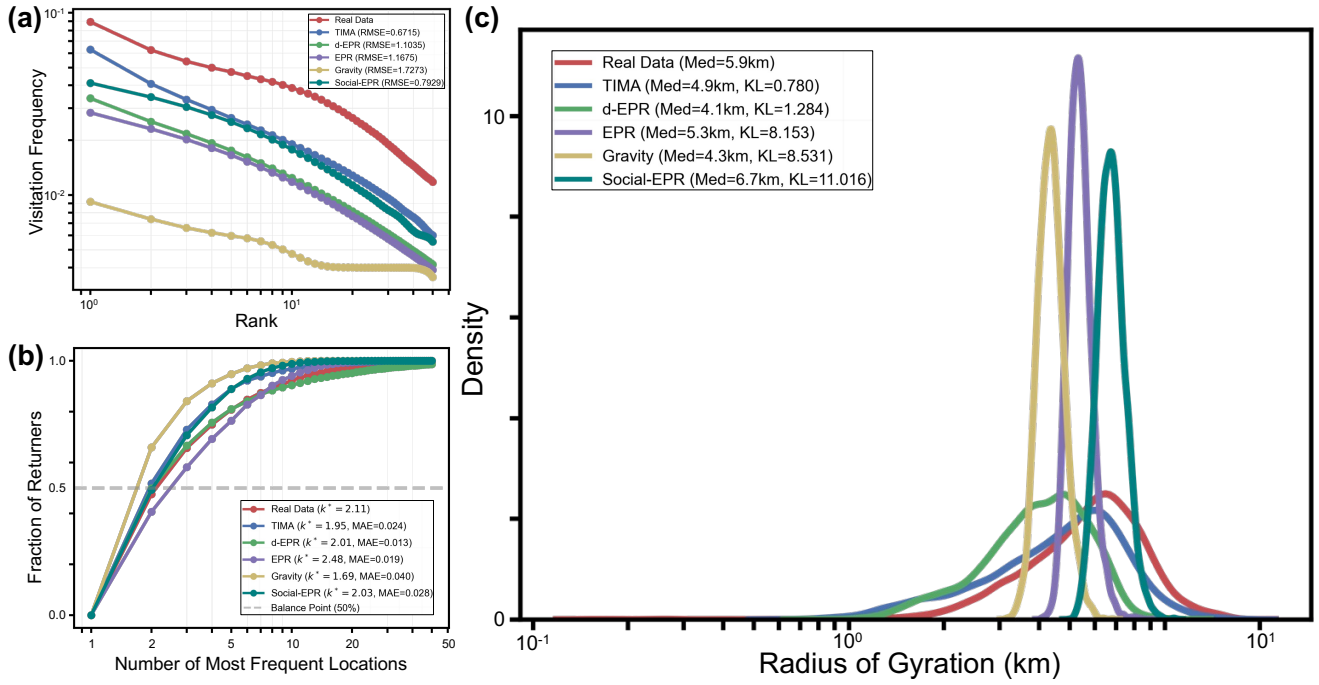
Supplementary Figure 5: Robustness and heterogeneity of LLM-inferred behavioral parameters across 50 independent runs. To verify stability, we perform 50 stochastic inference iterations for 10 representative agent profiles, covering different socioeconomic statuses. **a,** Semantic interest scores ($w_{u,k}$) for different demographic groups. The solid line represents the mean value, and the shaded area represents the standard deviation across 50 runs, demonstrating that agents maintain distinct yet stable lifestyle preferences (e.g., high-income agents consistently prefer Arts & Recreation venues across multiple runs). **b,** Personalized exploration probability ($\mathscr{P}_u$) as a function of the number of unique visited locations (S). The shaded bands (standard deviation) demonstrate the high robustness of the exploration decay rate inferred by different agents. **c, d,** Socio-economic affinity scores ($A_{u,c}$) for income and race dimensions. The clear separation of distributions among groups indicates that agents consistently exhibit unique social preferences, thereby driving heterogeneous mobility behaviors rather than acting as a single pattern.

3.3 Sensitivity analysis of device coverage

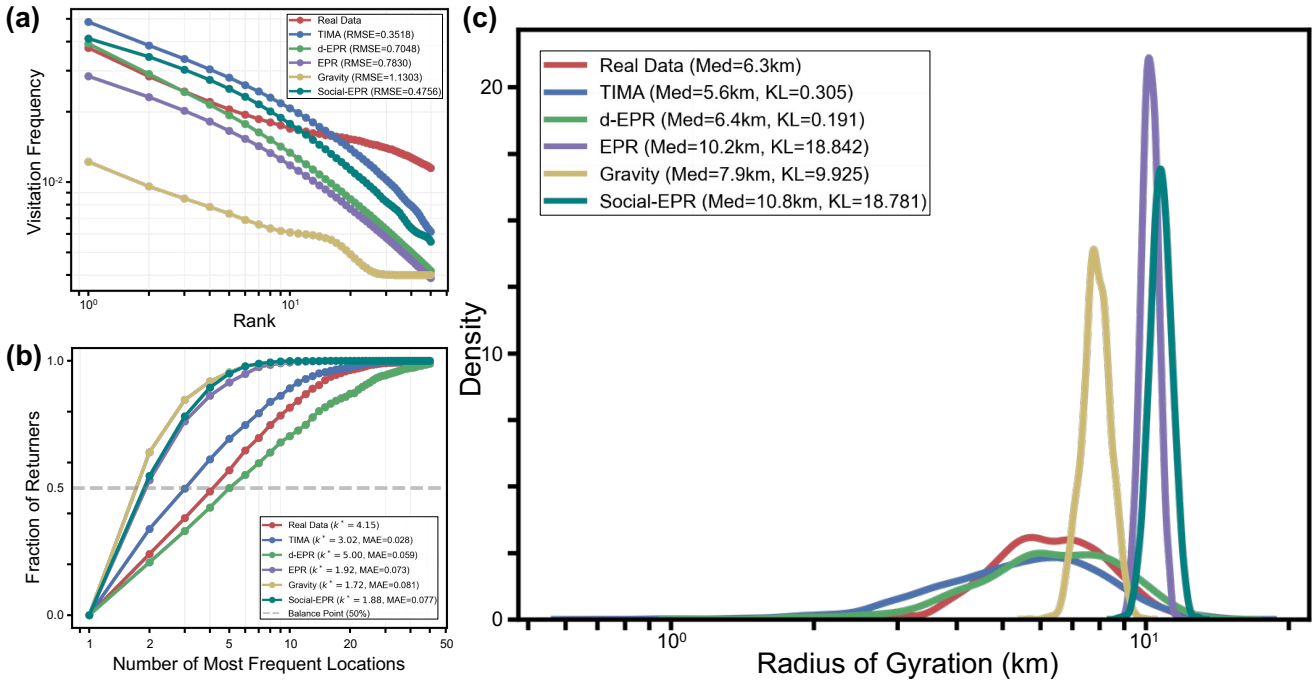
Supplementary Table 6: Sensitivity analysis of model performance under different device coverage thresholds. This evaluation filters CBGs based on the ratio of recorded devices to census population data (0%, 1%, and 5% thresholds). TIMA exhibits strong robustness, outperforming the gravity model at all spatial granularities regardless of data sparsity. While the Gravity model's performance significantly degrades in low-coverage scenarios (e.g., hierarchical OD fidelity), TIMA maintains high fidelity in travel distance and OD flow. Notably, as the sampling threshold increased (retaining CBGs with higher device coverage), TIMA improves performance on semantic metrics (POI Proportion KL), indicating that higher-quality input data helps further refine the generative agent's semantic decision-making capabilities.

Sampling Threshold	Model	Active CBGs	Trip Distance (KL) ↓	OD Flow (CPC) ↑	Vis. Density (MSE) ↓	POI Prop. (KL) ↓	Strat. OD (CPC) ↑
0%	TIMA	6,493	**0.054** (+87.5%)	**0.289** (+17.5%)	0.006 (0.0%)	**0.181** (+76.3%)	**0.231** (+14.3%)
	Gravity		0.432 (Ref.)	0.246 (Ref.)	0.006 (Ref.)	0.763 (Ref.)	0.202 (Ref.)
1%	TIMA	6,271	**0.055** (+86.0%)	**0.290** (+17.4%)	0.006 (0.0%)	**0.181** (+86.9%)	**0.232** (+14.9%)
	Gravity		0.394 (Ref.)	0.247 (Ref.)	0.006 (Ref.)	1.386 (Ref.)	0.202 (Ref.)
5%	TIMA	4,602	**0.057** (+85.2%)	**0.264** (+15.3%)	0.007 (0.0%)	**0.185** (+87.1%)	**0.216** (+13.1%)
	Gravity		0.385 (Ref.)	0.229 (Ref.)	0.007 (Ref.)	1.435 (Ref.)	0.191 (Ref.)

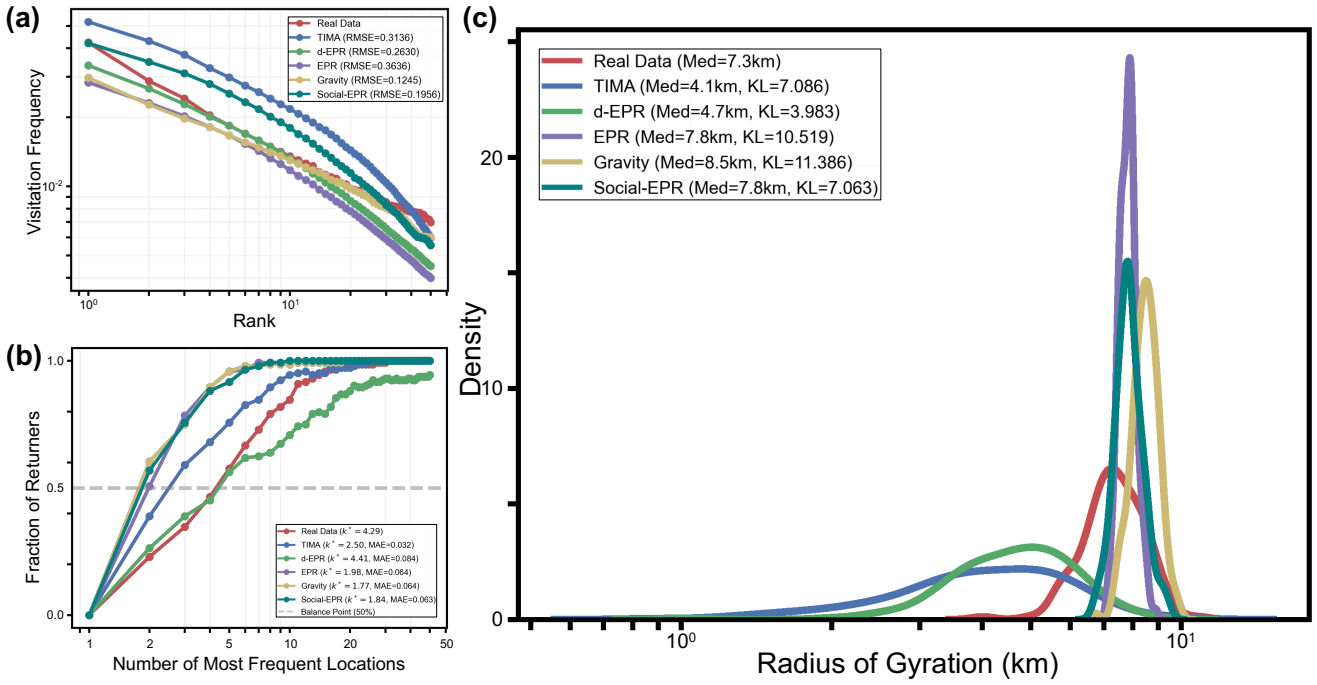
3.4 Comparison of fundamental mobility laws across cities and models



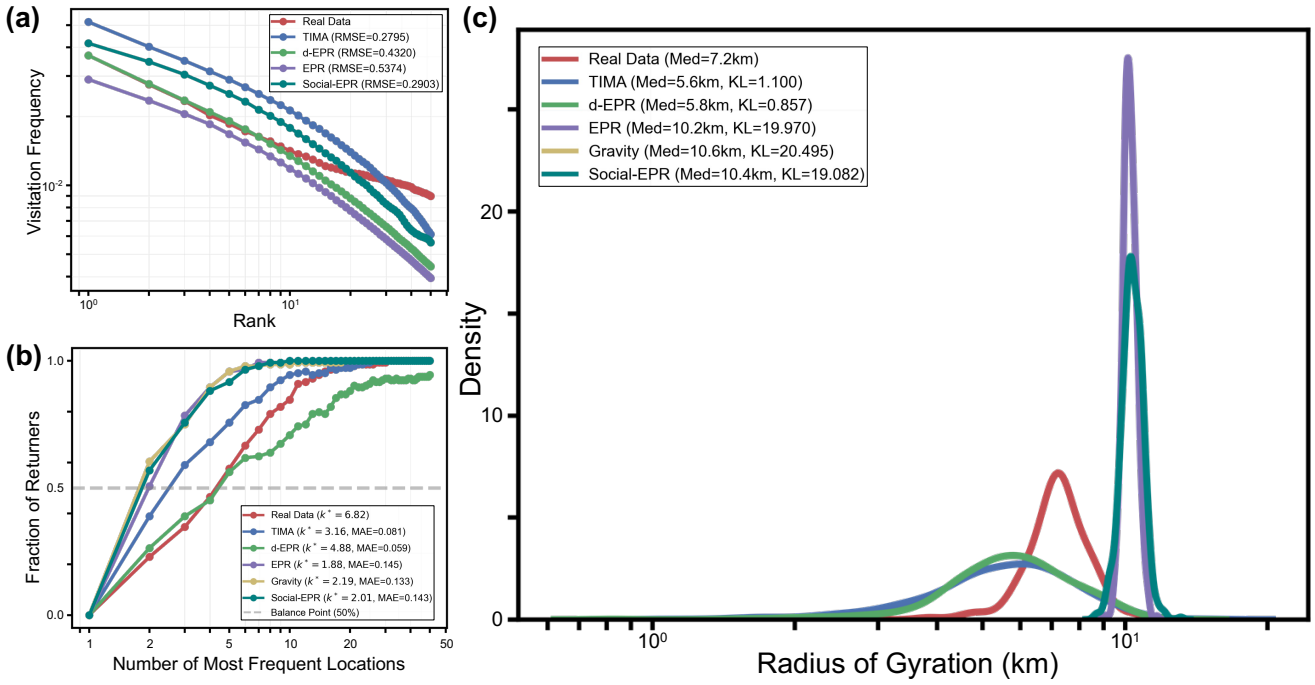
Supplementary Figure 6: Validation of fundamental mobility laws in New York City across models. a, Visitation frequency distribution follows Zipf's law, where the frequency of visits to the k-th most frequently visited location decays power-law. TIMA (blue) closely matches the slope (red) of empirical data. **b,** The proportion of "returners" as a function of the number of visited locations (k). The TIMA accurately reproduces the balance point where returners dominate explorers. **c,** The probability density function of the radius of turn (r_g). TIMA captures the heavy-tailed distribution characteristics of individual spatial extent, consistent with empirical observations.



Supplementary Figure 7: Validation of fundamental mobility laws in Chicago across models. a, Visitation frequency distribution follows Zipf's law, where the frequency of visits to the k-th most frequently visited location decays power-law. TIMA (blue) closely matches the slope (red) of empirical data. **b,** The proportion of "returners" as a function of the number of visited locations (k). The TIMA accurately reproduces the balance point where returners dominate explorers. **c,** The probability density function of the radius of turn (r_g). TIMA captures the heavy-tailed distribution characteristics of individual spatial extent, consistent with empirical observations.



Supplementary Figure 8: Validation of fundamental mobility laws in Orlando across models. a, Visitation frequency distribution follows Zipf's law, where the frequency of visits to the k-th most frequently visited location decays as a power law. TIMA (blue) closely matches the slope (red) of empirical data. **b,** The proportion of "returners" as a function of the number of visited locations (k). The TIMA accurately reproduces the balance point where returners dominate explorers. **c,** The probability density function of the radius of turn (r_g). TIMA captures the heavy-tailed distribution characteristics of individual spatial extent, consistent with empirical observations.



Supplementary Figure 9: Validation of fundamental mobility laws in Norfolk-Virginia Beach across models. a, Visitation frequency distribution follows Zipf's law, where the frequency of visits to the k-th most frequently visited location decays as a power law. TIMA (blue) closely matches the slope (red) of empirical data. **b,** The proportion of "returners" as a function of the number of visited locations (k). The TIMA accurately reproduces the balance point where returners dominate explorers. **c,** The probability density function of the radius of turn (r_g). TIMA captures the heavy-tailed distribution characteristics of individual spatial extent, consistent with empirical observations.

3.5 Comparison of experienced segregation

Supplementary Table 7: Comparison of experienced segregation indices across cities and models. This table reports the experienced segregation index (S), which quantifies the income homophily of social interaction and exposure. The detailed calculation method are provided in Supplementary Information Section 2.7. TIMA consistently maintains a closer consistency with empirical observations compared to baseline models (e.g., closest match in New York City and Chicago), accurately capturing the intensity of socio-economic stratification in urban mobility.

City	Empirical	Ours	Baselines			
	Real Data	TIMA	d-EPR	EPR	Social-EPR	Gravity
New York City	0.584	0.570	0.530	0.529	0.546	0.522
Chicago	0.639	0.648	0.557	0.544	0.581	0.529
Orlando	0.664	0.785	0.660	0.650	0.739	0.654
Norfolk-Virginia Beach	0.611	0.649	0.579	0.548	0.597	0.546

3.6 Comparison of results across different seasons

Supplementary Table 8: Temporal stability of model performance across seasons. To verify TIMA's generalizability beyond a single snapshot, we evaluate the simulation results with the weekly average mobility patterns in winter (January), spring (March), summer (June), and fall (October) of 2019. The "Baseline" row corresponds to the specific week of June 10, 2019, which is used in the main analysis. The consistency of performance metrics across different temporal windows (e.g., OD Flow CPC consistently remaining above 0.33 and trip distance KL remaining at a low level) confirms that the TIMA framework can reliably reconstruct routine travel structures regardless of seasonal variations.

Seasons	Trip Distance (KL) ↓	OD Flow (CPC) ↑	Visitation Density (MSE) ↓	POI Proportion (KL) ↓	Stratified OD Fidelity (CPC) ↑
Baseline (Week of Jun. 10, 2019)	0.054	0.289	0.006	0.181	0.231
Winter (Weekly Avg. of Jan. 2019)	0.061	0.334	0.007	0.198	0.276
Spring (Weekly Avg. of Mar. 2019)	0.057	0.338	0.006	0.247	0.279
Summer (Weekly Avg. of Jun. 2019)	0.054	0.344	0.006	0.338	0.283
Fall (Weekly Avg. of Oct. 2019)	0.072	0.333	0.006	0.255	0.275

3.7 Temperature stability verification

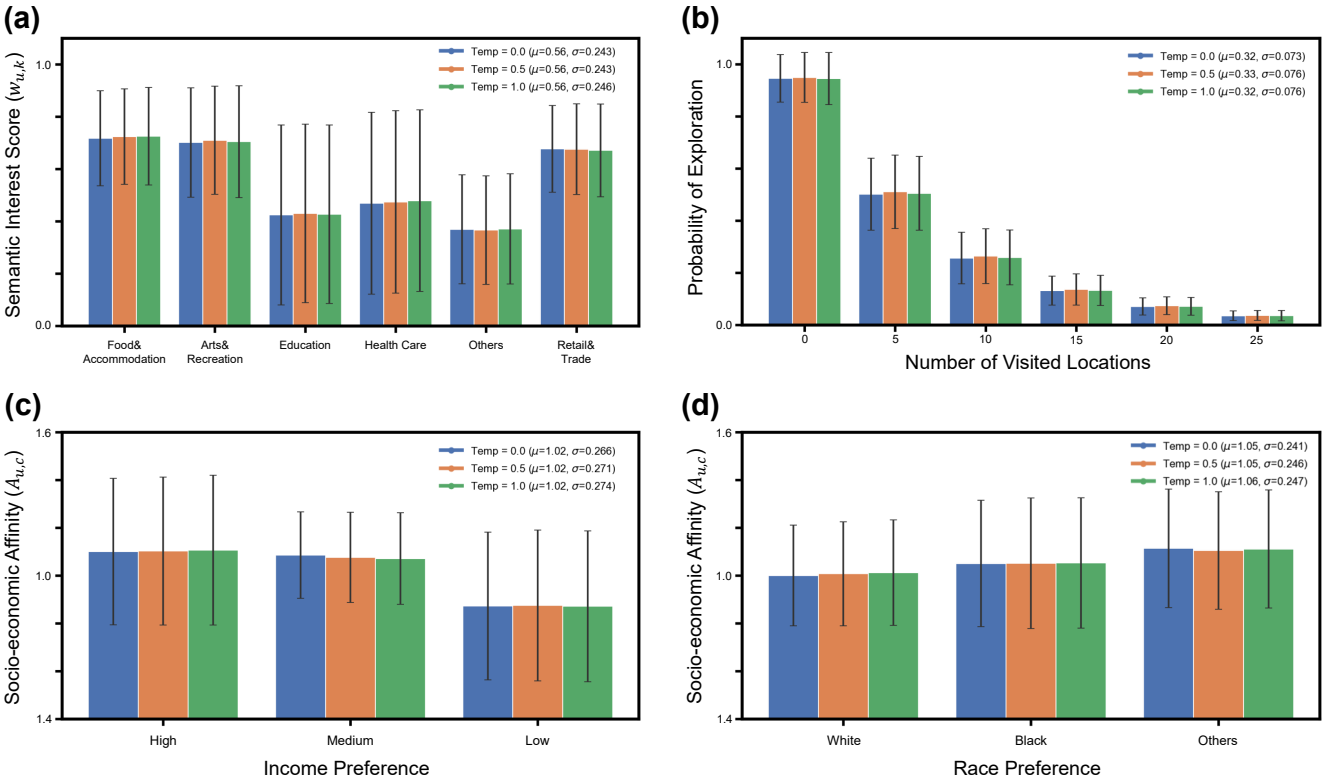
To assess the sensitivity of generative agents to stochasticity during inference, we conduct simulations at three temperature settings (0.0, 0.5, and 1.0). Supplementary Table 9 shows that overall mobility performance remains highly stable, with tiny fluctuations in key metrics (e.g., OD flow CPC remains ≈ 0.290, and Visitation Density remains constant). At the microscopic level, Supplementary Table 10 and Supplementary Figure 10 confirm that the distribution of inferred agent behavioral parameters, including personalized probability of exploration $\mathscr{P}_u$, semantic interest score $w_{u,k}$, and socio-economic affinity $A_{u,c}$, are structurally consistent across different temperatures. This indicates that the LLM's inference is robustly driven by the agent's semantic persona rather than random sampling noise.

Supplementary Table 9: Stability of overall mobility performance metrics under different temperature settings.

Temperature	Trip Distance (KL) ↓	OD Flow (CPC) ↑	Visitation Density (MSE) ↓	POI Proportion (KL) ↓	Stratified OD Fidelity (CPC) ↑
Temp.=0.0	0.054	0.289	0.006	0.181	0.231
Temp.=0.5	0.054	0.290	0.006	0.162	0.231
Temp.=1.0	0.054	0.289	0.006	0.162	0.231

Supplementary Table 10: Statistical summary of agent behavioral parameters across temperatures. Values represent the Mean and standard deviation (SD) of the inferred parameters ($\mathscr{P}_u$, $w_{u,k}$, $A_{u,c}$).

Temperature	Exploration Prob. ($\mathscr{P}_u$, Mean)	Semantic Interest Score ($w_{u,k}$, SD)	Socio-economic Affinity ($A_{u,c}$)	
			Income [Mean (SD)]	**Race** [Mean (SD)]
Temp. = 0.0	0.323	0.279	1.019 (0.291)	1.055 (0.246)
Temp. = 0.5	0.328	0.279	1.018 (0.296)	1.054 (0.250)
Temp. = 1.0	0.325	0.280	1.016 (0.298)	1.058 (0.251)

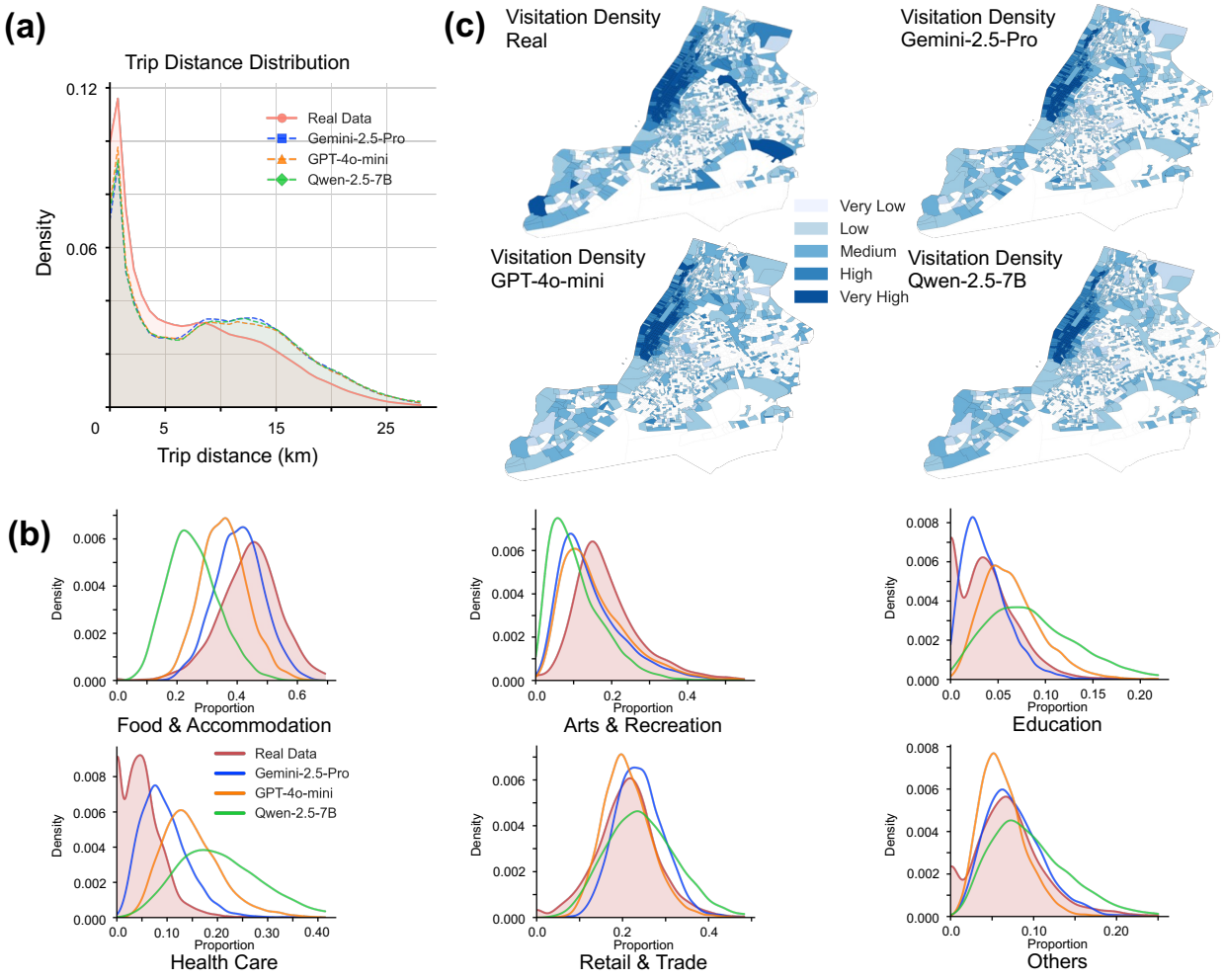


Supplementary Figure 10: Comparison of behavioral parameter distributions across temperatures. The consistent bar heights and error bars across settings (0.0, 0.5, 1.0) for **a,** semantic interest scores, **b,** personalized probability of exploration, and **c, d,** socioeconomic affinity scores confirm the insensitivity of agent reasoning to temperature changes.

3.8 LLM stability verification

Supplementary Table 11: Performance of TIMA driven by different LLMs. We compared the Gemini-2.5-Pro (baseline model in our study) against GPT-4o-mini and Qwen-2.5-7B-Instruct. The consistent performance across metrics indicates that the TIMA framework is model-agnostic. While semantic preference (POI Proportion) varies with model capabilities, the framework functions effectively with various LLMs

LLM	Trip Distance (KL) ↓	OD Flow (CPC) ↑	Visitation Density (MSE) ↓	POI Proportion (KL) ↓	Stratified OD Fidelity (CPC) ↑
Gemini-2.5-Pro	0.054	0.289	0.006	0.181	0.231
GPT-4o-mini	0.045	0.292	0.007	0.457	0.234
Qwen-2.5-7B-Instruct	0.054	0.286	0.007	1.047	0.227



Supplementary Figure 11: Generalizability of TIMA across different LLMs a, Trip distance distributions produced by TIMA driven by Gemini-2.5-Pro, GPT-4o-mini, and Qwen-2.5-7B-Instruct. All models successfully recover the Lévy flight characteristics. **b,** Comparative density maps of POI category visitation preferences. **c,** Spatial visitation density heatmaps. While minor variations exist in semantic reasoning capabilities (as reflected in POI Proportion KL provided in Supplementary Table 11), the TIMA framework ensures that mobility patterns remain consistent in both macro- and microscopic level across different underlying LLM backbones.

3.9 Prompt sensitivity analysis

Supplementary Table 12: Sensitivity analysis of agent behavioral parameters across different prompts. To evaluate the robustness of TIMA to changes in prompt engineering, we construct 10 different prompt permutations by altering the phrasing of task instructions and the structured presentation of agent profiles. The LLM inference for this analysis is performed using GPT-4o-mini to assess the consistency of the output. The table reports the Mean and standard deviation (SD) of key inferred behavioral parameters ($\mathscr{P}_u$, $w_{u,k}$, $A_{u,c}$). The low SD relative to the means indicates that the agent's reasoning is primarily driven by the semantic content of the persona, rather than by the specific prompt instructions.

Experimental Setting	**Exploration Prob.** ($\mathscr{P}_u$, Mean)	**Semantic Interest Score** ($w_{u,k}$, SD)	**Socio-economic Affinity** ($A_{u,c}$)	
			Income [Mean (SD)]	**Race** [Mean (SD)]
Test 1	0.435	0.189	0.971 (0.374)	0.967 (0.398)
Test 2	0.417	0.167	0.983 (0.322)	1.021 (0.329)
Test 3	0.444	0.175	0.987 (0.312)	1.002 (0.335)
Test 4	0.400	0.175	0.990 (0.335)	1.023 (0.352)
Test 5	0.400	0.176	0.996 (0.308)	1.015 (0.336)
Test 6	0.529	0.201	0.985 (0.273)	1.037 (0.317)
Test 7	0.526	0.152	1.004 (0.236)	1.056 (0.280)
Test 8	0.526	0.173	0.995 (0.255)	1.043 (0.308)
Test 9	0.515	0.142	0.996 (0.283)	1.039 (0.331)
Test 10	0.531	0.178	0.981 (0.288)	1.000 (0.355)

References

1. U.S. Census Bureau. 2022 NAICS manual (2022).
2. U.S. Census Bureau. American community survey 5-year data 2019 (2021).
3. U.S. Census Bureau. 2019 U.S. gazetteer files: Places (2019).
4. U.S. Census Bureau. Cartographic boundary files – shapefile (2024). Accessed 2026-02-03.
5. SafeGraph. Weekly patterns (2020).
6. Sinnott, R. W. Virtues of the haversine. *Sky telescope* **68**, 158 (1984).
7. Hale, T. *et al.* A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat. Hum. Behav.* **5**, 529–538, DOI: 10.1038/s41562-021-01079-8 (2021).
8. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases* **20**, 533–534 (2020).
9. Noulas, A., Scellato, S., Lambiotte, R., Pontil, M. & Mascolo, C. A tale of many cities: Universal Patterns in human Urban Mobility. *PLoS ONE* **7**, e37027, DOI: 10.1371/journal.pone.0037027 (2012).
10. Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S. & Baronchelli, A. Evidence for a conserved quantity in human mobility. *Nat. human behaviour* **2**, 485–491 (2018).
11. Schelling, T. C. Dynamic models of segregation. *J. mathematical sociology* **1**, 143–186 (1971).
12. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a Feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444, DOI: 10.1146/annurev.soc.27.1.415 (2001).
13. Song, C., Koren, T., Wang, P. & Barabási, A.-L. Modelling the scaling properties of human mobility. *Nat. physics* **6**, 818–823 (2010).
14. Pappalardo, L. *et al.* Returners and explorers dichotomy in human mobility. *Nat. communications* **6**, 8166 (2015).
15. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
16. Barbosa, H. *et al.* Human mobility: Models and applications. *Phys. Reports* **734**, 1–74 (2018).
17. Eagle, N., Pentland, A. S. & Lazer, D. Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci.* **106**, 15274–15278, DOI: 10.1073/pnas.0900282106 (2009).
18. Schelling, T. C. Dynamic models of segregation. *The J. Math. Sociol.* **1**, 143–186, DOI: 10.1080/0022250X.1971.9989794 (1971).
19. Moro, E., Calacci, D., Dong, X. & Pentland, A. Mobility patterns are associated with experienced income segregation in large us cities. *Nat. communications* **12**, 4633 (2021).
20. Stouffer, S. A. Intervening opportunities: A theory relating mobility and distance. *Am. Sociol. Rev.* **5**, 845–867 (1940).
21. Zhou, Y. & Lu, Y. Varying relationships between experienced income segregation and travel behaviour across neighbourhood social and urban contexts. *Nat. Commun.* **16**, 11236, DOI: 10.1038/s41467-025-66585-z (2025).
22. Zipf, G. K. The $\frac{P_1 P_2}{D}$ hypothesis: on the intercity movement of persons. *Am. sociological review* **11**, 677–686 (1946).
23. Jenks, G. The data model concept in statistical mapping. *Int. Yearb. Cartogr.* **7**, 186–190 (1967).