Human Mobility Modeling with Household Coordination Activities under Limited Information via Retrieval-Augmented LLMs

Yifan Liu[1], Xishun Liao[1,*], Haoxuan Ma[1], Brian Yueshuai He[2], Chris Stanford[3], and Jiaqi Ma[1]

***Abstract*— Understanding human mobility patterns has long been a challenging task in transportation modeling. Due to the difficulties in obtaining high-quality training datasets across diverse locations, conventional activity-based models and learning-based human mobility modeling algorithms are particularly limited by the availability and quality of datasets. Current approaches primarily focus on spatial-temporal patterns while neglecting semantic relationships such as logical connections or dependencies between activities and household coordination activities like joint shopping trips or family meal times, both crucial for realistic mobility modeling. We propose a retrieval-augmented large language model (LLM) framework that generates activity chains with household coordination using only public accessible statistical and socio-demographic information, reducing the need for sophisticated mobility data. The retrieval-augmentation mechanism enables household coordination and maintains statistical consistency across generated patterns, addressing a key gap in existing methods. Our validation with NHTS and SCAG-ABM datasets demonstrates effective mobility synthesis and strong adaptability for regions with limited mobility data availability.**

I. INTRODUCTION

Understanding and accurately generating human mobility patterns remains a fundamental challenge in transportation research with implications for urban planning, public health, and even retail strategies [1], [2], [3], [4]. Accurate mobility modeling can enhance transportation efficiency and urban design, ultimately improving civilian quality of life.

Traditional activity-based models (ABMs) have advanced human mobility understanding by simulating daily activities based on socio-economic characteristics. Government agencies like SCAG have widely adopted these models for traffic analysis, urban planning, and commercial strategy development since 1999 [5], [6]. While ABMs effectively model behavioral dynamics, they require extensive local data and rely on many assumptions. Meanwhile, data-driven neural network approaches have emerged to capture mobility patterns using mobile phone and GPS datasets [7], [8], [9], [10].

However, these methods face several limitations, they require detailed individual travel diary data which raises privacy concerns, struggle to adapt to rapid urban changes, and rely on simplified behavioral assumptions that may not capture the flexibility of human decision-making in response to socio-economic changes [11].

[1]UCLA Mobility Lab, Department of Civil and Environmental Engineering, University of California, Los Angeles, Los Angeles, USA.

[2]Civil and Environmental Engineering Department, University of Louisville, Louisville, Kentucky, USA.

[3]Novateur Research Solutions, Ashburn, VA, USA.

*Corresponding author: xishunliao@ucla.edu

Recent advances in computational power have enabled Large Language Models (LLMs) to create new opportunities for human mobility modeling [12], [13]. Models like GPT-4 [14] excel at generating human-like text across domains and understanding complex sequences with strong interdependencies. Trained on diverse textual data, LLMs can incorporate a wide range of human experiences and behaviors, potentially leading to more nuanced and varied human mobility modeling compared with conventional methods, as daily routines often involve intricate chains of activities with subtle interrelations.

Building on this potential, we introduce a novel application of LLMs for mobility data synthesis. Based on the concept of activity chains [9], which reflects daily sequences of individual activities, our framework employs retrieval-augmented LLMs to generate realistic, demographically consistent mobility patterns considering the household coordination. Given by socio-demographic attributes and public available statistics, the model ensures statistical consistency and coordinated behavior without relying on extensive historical data or detailed behavioral assumptions. This approach enables scalable mobility data synthesis across diverse regions with limited data availability, supporting micro-simulation and integrated transportation modeling for urban planning. Our study makes several key contributions to the field of human pattern modeling compared with existing literature:

- We propose a novel LLM-based approach for activity chain generation using only basic socio-demographic and statistical data, reducing dependency on extensive mobility datasets while preserving privacy.
- We develop a retrieval-augmented LLM framework with feedback that ensures statistical consistency across generated patterns, producing reliable results for large-scale mobility modeling. To the best of our knowledge, this framework is the first application of retrieval-augmented LLMs with feedback for mobility data generation.
- We pioneer LLM-based household-coordinated activity generation, enabling realistic modeling of interdependent activities among household members.

II. LITERATURE REVIEW

A. *Human Mobility Modeling*

Human mobility modeling has evolved significantly since the 1940s when the "Law of Intervening Opportunities" first connected travel patterns to socio-economic factors [15]. Modern GPS and electronic tracking technologies have enabled sophisticated data collection and generative modeling

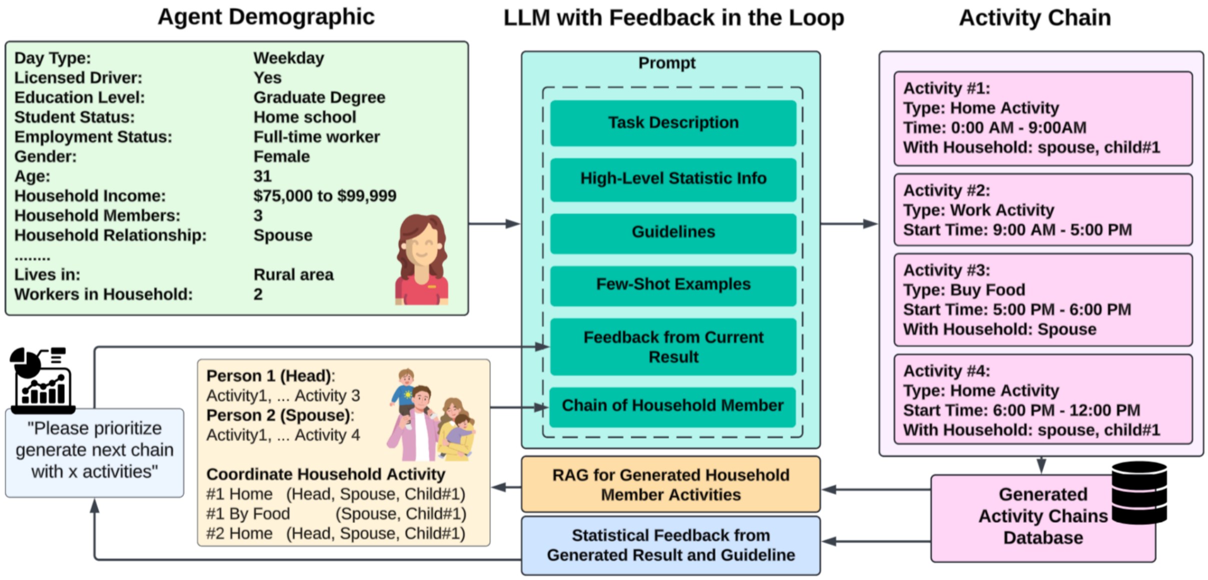

Fig. 1: Proposed retrieval-augmented LLM framework with feedback loop for activity chain generation.

approaches, with activity chain generation becoming a key focus area. ABMs represent a major advancement in transportation planning by simulating individuals' daily activities, using socio-demographic data, land use information, and transportation networks to construct detailed activity chains [16]. While SCAG implemented SimAGENT [6] to analyze regional travel behaviors, ABMs require extensive data collection and rely heavily on assumptions about travel patterns, limiting their transferability.

Learning-based methods using deep learning, Graph Convolutional Networks, and transformers offer alternatives when trained on mobility data from mobile devices [17], [9], but remain dependent on high-quality data that is often expensive and restricted. Both traditional and learning-based approaches face limitations from their reliance on comprehensive datasets or numerous behavioral assumptions, highlighting the need for training-free approaches like LLMs that can synthesize mobility patterns using more accessible data sources.

B. Household Activity Modeling

Early approaches treated household members as independent units, overlooking their natural interdependencies [18]. Modern household activity modeling now incorporates intra-household interactions, jointly modeling in-home and out-of-home activities to capture trade-offs and interdependencies between members [19]. Platforms like VirtualHome model complex household activities through atomic action sequences [20], while SMACH offers multi-agent simulations to study energy consumption patterns and behavioral impacts [21], collectively improving the realism of household activity models.

Despite advances, current models fail to adequately capture household schedule interdependencies, relying on rigid rules or extensive data that limits cross-cultural generalization. Computational demands also restrict scalability for large-scale implementations.

C. Large Language Models

LLMs trained on trillions of tokens have emerged as powerful tools with transformer architectures that excel across domains from personal assistance to vehicle navigation [14], [22]. Their flexibility enables rapid adaptation to new scenarios with minimal input [14].

Retrieval-augmented generation (RAG) enhances LLMs by allowing access to external databases, grounding responses in reliable information and addressing hallucination issues, though challenges in trustworthiness remain. Complementary approaches use automated feedback loops to iteratively refine responses, reducing hallucinations across various tasks [23].

Our framework leverages these technologies to generate realistic activity chains with minimal data requirements. By combining LLMs with specialized retrieval-augmentation, we ensure statistical consistency and enable household activity coordination without requiring extensive location-specific data, while capturing household activities' interdependencies.

III. Methodology

A. Overview

The problem addressed in this study is defined as generating the daily activity chain for individual agents based on their socio-demographic information throughout a day. For each agent i with his or her socio-demographic information collection $D_i = \{{d_i}^1, {d_i}^2, \ldots, {d_i}^n\}$, we aim to generate a daily activity chain C_i for that agent where each activity in the chain is defined by its type A, start time T_s, end time T_e, and household members participating H. The output activity chain C_i for agent i can be represented as

$$C_i = [{A_i}^1, {T_{s,i}}^1, {T_{e,i}}^1, {H_i}^1], \ldots, [{A_i}^n, {T_{s,i}}^n, {T_{e,i}}^n, {H_i}^n],$$

The architecture of our proposed framework is illustrated in Fig. 1. The model takes 9 socio-demographic attributes per agent as input, which feeds into a feedback loop LLM mechanism containing modules for task description, statistical information, generation guidelines, few-shot examples, and feedback. The framework outputs activity chains with temporal structure and household coordination. Once generated, activity chains are stored in a database. RAG retrieved household-related info and statistical feedback from these results will then be used in the feedback loop to continuously refine subsequent activity chain generation tasks, ensuring consistency across household members and alignment with empirical distributions.

TABLE I: Activity types aggregated in the NHTS 2017 dataset for the Los Angeles area.

1	Home	2	Work	3	School
4	Caregiving	5	Buy goods	6	Buy services
7	Buy meals	8	General errands	9	Recreational
10	Exercise	11	Visit friends	12	Health care
13	Religious	14	Something else	15	Drop off/Pick up

Table I presents the 15 activity types used in our framework, aggregated from the filtered NHTS 2017 dataset in Los Angeles area [24]. These categories encompass the full spectrum of daily human activities, from essential functions like home, work, and school to discretionary activities such as recreation, exercise, and social visits. This classification system provides a comprehensive foundation for generating realistic daily activity patterns.

B. Socio-Demographic Information

Agent socio-demographic information serves as model input to generate activity chains reflecting individual characteristics. Nine attributes (gender, age, education, student status, employment, household relationships, income, driver license status, and location) are converted to natural language descriptions, as shown in Fig. 1. This conversion from structured socio-demographic data into natural language expressions facilitates deeper contextual comprehension by the LLMs, enhancing the realism and relevance of the generated daily activity patterns.

C. Prompt Architecture

We provide the LLMs with a structured system prompt that guides the generation of activity chains. The components of the system prompt are designed to provide comprehensive context and clear instructions, ensuring the generated outputs are both logically reasonable and aligned with the socio-demographic data. The structured input consists of the following elements:

The Large Language Model is guided through a structured prompt system, as detailed in Fig. 2. This structured prompt integrates several critical components:

- **Task Description:** Defines the goal explicitly, instructing the LLM to generate realistic daily activity chains for agents based on socio-demographic inputs and empirical patterns from datasets like the NHTS survey.
- **High-Level Statistical Information:** Provides statistical context, such as activity type frequencies, typical activity durations, household coordination probabilities, and spatial-temporal patterns, that the LLM uses to anchor generated activities to realistic empirical distributions.

```
# Task Description
Generate a realistic one-day activity chain for a person based on their
demographic information, matching empirical patterns from NHTS survey
data. Also include information about household members participating in
each activity.
## Activity Type Codes:
1. **Home activities**: Sleep, household chores, remote work
2. **Work activities**: Professional or volunteer work
3. **School attendance**: Education activities
......
15. **Transport assistance**: Driving others
```

```
# Statistical Data
## Activity Type Frequencies:
| Code | Activity | Statistic % | Accompany by Household % | Notes |
|------|----------|----------|------------------------|-------|
| 1 | Home | 35-45% | 27.3% | Always start/end here |
| 2 | Work | 20-25% | 9.2% | For workers only |
...
| 15 | Transport | 4-8% | 58.8% | Common activity |
## Statistical Patterns to Match
Activity Chain Length: Each activity chain should vary in length from 3 to
14 activities, with a natural mix of shorter and longer chains....
Activity Duration Distribution: Use VARIABLE and realistic durations
based on activity type, not fixed slots. Shorter activities for shopping......
Activity Timing: Create a MORE CONTINUOUS distribution of activity
starts/ends throughout the day......
Household Coordination: For each activity, indicate how many household
members participate in the activity with the person. Consider which activities
are likely to be done together (dining, shopping, recreation) versus alone...
## Location and Time Context: United States, California, 2017
```

```
# Guidelines
## Create natural variation: Avoid fixed patterns or identical durations.
## Respect time constraints: Activities should flow logically......
## Demographic alignment:Employment status strongly affects daily
patterns. Age influences activity types and timing. Income level......
## Household coordination: Consider which household members would
logically participate in activities together......
## Output Format:
[[activity_code, start_quarter, end_quarter, household_members_count], ...]
## Example Activity Chains for Individuals and Household members
```

```
# Agent Demographic info and In Context Feedback
## Current Agent Demographic info: Liscensed Driver:Yes, Educational
Attainment: college, Gender: Male, Has Job: Yes....
## Current Agent Household info:
Current agent's Role in this household: Household head.
There are "x" household members in total and "y" out of "x" members are
generated.
Other household members' activity chains:
Person 1 (Spouse): Activity 1, Activity 2..., Activity 4......
Unfulfilled coordination activities that need to be addressed:
Activity type "a" starting at quarter "t1" ending at quarter "t2"
## Statistic Feedback based on generated Chains Dataset:
For this chain, please prioritize generating a chain with "n" activities...
```

Fig. 2: Example of input system prompt for LLMs.

- **Generation Guidelines:** Specifies constraints and standards for output realism, including the logical ordering of activities, duration distributions, and household coordination requirements. This ensures feasible and contextually coherent daily schedules.
- **Few-Shot Examples:** Demonstrates desired output formats through representative examples, helping the LLM internalize typical patterns of activity sequences, durations, and transitions. These examples enable the model to generate outputs adhering closely to observed human mobility patterns.
- **Retrieval-Augmented Generation Feedback:** Dynamically integrates statistical feedback and previously generated household member activities into the LLM's gen-

eration loop. This retrieval-augmented approach maintains consistency across household activity chains, supports realistic household coordination (such as shared home activities, joint travel, and synchronized activity timings), and aligns generated chains with observed statistical distributions.

Fig. 2 provides an example of the system prompt. These components collectively ensure that the LLM has a clear understanding of the task requirements and the contextual background needed to generate accurate and representative activity chains.

D. Retrieval-Augmentation Mechanism

Our framework incorporates a retrieval-augmentation mechanism to address limitations in sequential, agent-specific activity chain generation. LLMs typically struggle to maintain awareness of global statistical distributions when generating chains individually based on socio-demographic information. As shown in Fig. 2, we overcome this by continuously monitoring and storing generated results, tracking statistical attributes like activity chain lengths. This stored data provides real-time feedback to the LLM, guiding subsequent generations to maintain consistency with empirically observed distributions. The system adapts prompts by analyzing previously generated data, guiding the LLM to produce activity chains that align with desired statistical distributions. Our experimental analysis shows that providing statistical feedback exclusively on chain length distribution successfully enhances both activity type and temporal distributions by utilizing the LLM's natural inference capabilities. This streamlined approach that focuses only on constraining chain length statistics minimizes potential bias while enabling the model to create naturally coherent patterns throughout various aspects of human mobility behavior. For household coordination, our system implements a retrieval mechanism that incorporates previously generated activity patterns when generating chains for new household members. The system retrieves existing mobility patterns and coordination activities like family meals, shared transportation, and joint recreation. This household context enables the LLM to generate aligned joint activities while maintaining schedule coherence, reflecting realistic family dynamics and interdependencies.

The mechanism identifies and resolves scenarios where household coordination activities remain incomplete or unsynchronized, ensuring properly coordinated and temporally coherent activities.

By integrating retrieval-augmentation with iterative feedback, our framework produces robust activity chains that reduce unrealistic mobility patterns brought by LLM hallucinations while ensuring statistically consistent and contextually appropriate household coordination.

IV. EXPERIMENT

A. Dataset

1) National Household Travel Survey Dataset: The 2017 National Household Travel Survey (NHTS) dataset [25] serves as our primary reference, providing comprehensive U.S. travel behavior data. From approximately 129,000 households and 264,000 persons surveyed, we used 180,000 filtered person records. Activities were aggregated into 15 types shown in Table I. The dataset includes household coordinated activities, enabling analysis of travel behavior interdependencies among household members.

2) Activity-Based Model Dataset from Southern California Association of Governments: We also utilize synthetic results from the Southern California Association of Governments (SCAG) Activity-Based Model [6], which simulates travel patterns for 26 million people. The activity chains from this dataset are converted to match our activity type categories in Table I for consistent comparison.

B. Experiment and Result

We evaluated our approach using the NHTS dataset in the California area by randomly sampling 500 agents and generating their daily activity chains based on their socio-demographic data. These chains were then validated against the comprehensive daily activity records from the NHTS dataset and the SCAG dataset. In our experiments, we utilized three large language models: OpenAI's GPT-4o mini, Meta's Llama3.1-70b, and DeepSeek v3, with the temperature setting of 1.0 for each model, comparing their performance in accurately simulating daily human activities. The token size for input is around 700 and output around 4 on average for each activity chain generation. The generation speed for each instance was approximately 0.5 seconds for the GPT-4o mini and DeepSeek v3 API calls, and 1 second for the Llama3.1-70b model with 4-bit quantization running on an L40S GPU with 48GB memory.

The evaluation metrics will compare the distributions of activity type, start time, end time, duration, and the number of daily activities. We employed Jensen-Shannon Divergence (JSD) to quantify the differences between the generated activity chains and the reference activity chains from the NHTS and the SCAG dataset.

TABLE II: JSD values comparing different models with reference datasets. Lower values indicate better alignment with reference distributions. The SCAG-NHTS comparison (bottom row) serves as a baseline, showing inherent differences between the two reference datasets.

Comparison	Type	Time start	Time end	Duration	Length
NHTS-SCAG	0.031	0.007	0.006	0.006	0.003
NHTS-GPT	**0.023**	0.015	**0.011**	0.022	0.011
NHTS-Llama	0.024	0.090	0.081	0.023	0.011
NHTS-DeepSeek	0.040	**0.013**	0.013	0.026	**0.009**
SCAG-GPT	**0.028**	**0.009**	**0.009**	**0.020**	0.009
SCAG-Llama	0.045	0.061	0.056	0.027	**0.006**
SCAG-DeepSeek	0.065	0.013	0.016	0.020	0.007

We analyzed the JSD values between our approach and the reference datasets, as detailed in Table II. A JSD value closer to 0 indicates a more accurate approximation with the reference dataset's distribution. Overall, our LLM-based approach successfully captures the trends in human mobility patterns across all evaluated dimensions. Among the models we tested, GPT-4o mini demonstrates particularly strong performance, especially in activity type and end time modeling across both datasets. The JSD values between NHTS and SCAG datasets (ranging from 0.003 to 0.031) reveal inherent differences between survey-based data (NHTS) and synthetic data (SCAG-ABM), which provides important context for interpreting our results against each reference dataset.

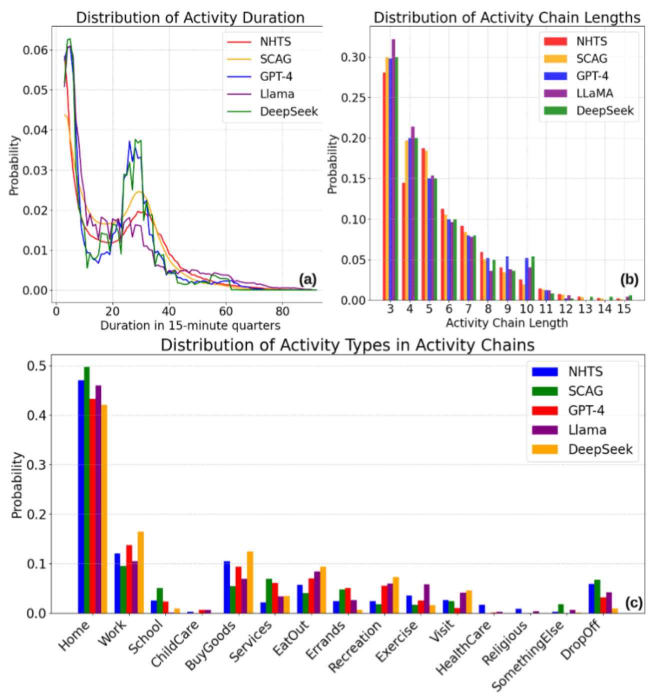

Fig. 3: Evaluation matrix on SCAG and NHTS dataset

C. Activity Type

Fig. 3c shows our approach accurately represents activity type distributions from reference datasets, especially for common activities like home and work. Our models capture both frequent and rare activity types with varying accuracy. GPT-4o mini aligns particularly well with reference distributions, accurately representing less common activities such as childcare, healthcare, and errands.

D. Activity Duration

Fig. 3a demonstrates our approach accurately models activity duration distributions from both reference datasets. The models capture the prevalence of shorter activities in NHTS and moderate-length activities in SCAG. All implementations perform well in this dimension, with GPT-4o mini showing the strongest alignment with both reference profiles, particularly for activities of various durations.

E. Activity Chain Length

Fig. 3b shows our approach captures the preference for shorter chains (3-6 activities) present in both reference datasets. All models generate realistic activity chain lengths, though they struggle with complex chains exceeding eight activities. This limitation presents an opportunity for future improvement in modeling extended daily routines, while confirming our approach successfully captures common activity chain patterns.

F. Activity Start Time

Fig. 4 reveals our approach captures daily temporal patterns from both NHTS and SCAG datasets. The models reproduce characteristic peaks matching typical daily schedules. While all models capture general temporal trends, GPT-4o mini shows the closest match with reference patterns, especially for activity end times. DeepSeek excels at modeling early-day start times, highlighting how different architectures may capture specific temporal aspects of mobility.

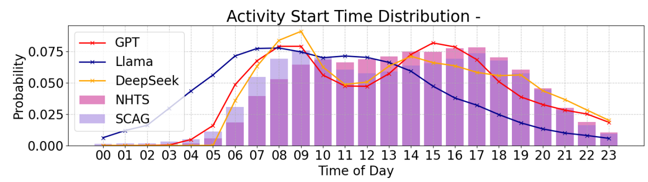

Fig. 4: Start time distribution comparation

G. Activity pattern in different social group

Our detailed analysis of activity start times for specific social groups, namely students and workers as illustrated in Fig. 5, demonstrates that our approach effectively captures distinct daily routines across different socio-demographic segments. The generated patterns successfully replicate characteristic peaks and time-of-day variations aligned with both NHTS and SCAG reference data. While GPT-4o-mini shows the best performance in modeling social group-specific mobility behaviors, our overall approach consistently represents realistic daily activity patterns for both students and workers. These results validate the effectiveness of our method in generating accurate human mobility patterns that reflect the temporal dynamics of different social groups.

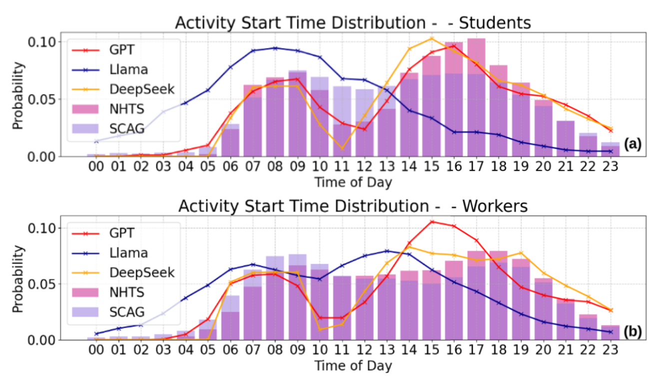

Fig. 5: Activity start/end time distribution for students and workers

H. Activity Pattern by Activity Type

The analysis of activity start times categorized by activity types, as shown in Fig. 6, highlights the robustness of our proposed approach in capturing nuanced temporal patterns. GPT-4o-mini demonstrates superior alignment with the reference datasets across all analyzed activity types, particularly in accurately reproducing the peak timing patterns for home, work, and dining activities. DeepSeek exhibits notable precision in capturing morning peaks for work-related activities, while Llama, though broadly accurate, tends to show more variability in capturing home and dining patterns. These findings emphasize the efficacy of our retrieval-augmented approach and underline GPT-4's particularly strong performance in modeling activity-specific timing trends.

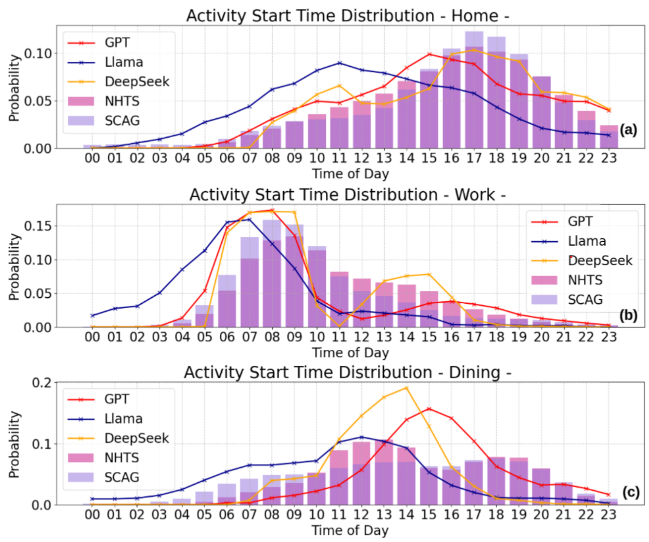

Fig. 6: Activity start/end time distribution for home, work, and dining activities

I. Household Coordination Activities

Based on GPT-4o mini's superior performance in previous evaluations, we focused our household coordination analysis on results from this model. Our analysis demonstrates excellent overall alignment with the NHTS reference dataset as shown in Fig. 7. The model successfully captures realistic patterns of joint household activities, particularly those where family members are most likely to participate together. We further analyzed specific relation pairs, namely head-spouse and head-child interactions, to evaluate the model's performance at a more granular level.

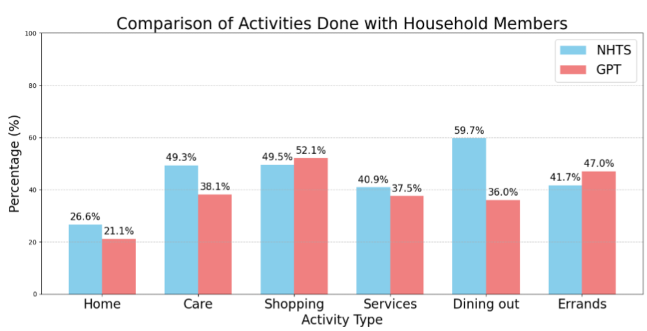

Fig. 7: Household coordination activities participate rate compared with NHTS dataset

As presented in Table III, household-interactive activities such as home, shopping, and errands show particularly strong alignment for both relation types, with minimal divergence measured by low JSD values. While care, dining out, and service-related activities—all common scenarios for joint household participation—exhibit slightly greater discrepancies, especially in head-child interactions, they remain within acceptable ranges. This detailed analysis by relation pair confirms that GPT-4o mini effectively reproduces household-coordinated behaviors across different family relationships, highlighting the strength of our retrieval-augmented approach in generating contextually consistent and realistic household interactions for the most common joint activities.

TABLE III: JSD values for household coordination activities participate rate compared with NHTS dataset

Relation	**Home**	**Care**	**Shop**	**Serv.**	**Dining**	**Errands**
Head-Spouse	0.0003	0.028	0.012	0.000	0.011	0.031
Head-Child	0.0003	0.049	0.005	0.005	0.057	0.011

V. Ablation Study and Hallucination Eliminating

To evaluate the effectiveness of our retrieval-augmented feedback approach in generating consistent activity chains with household coordination and reducing hallucinations, we conducted an ablation study comparing the model's performance with and without this mechanism.

TABLE IV: JSD values comparing activity chain statistics with and without retrieval-augmented feedback.

Comparison	Type	Time start	Time end	Duration	Length
NHTS-GPT	0.032	0.063	0.072	0.031	0.110
NHTS-GPT (Feedback)	**0.023**	**0.015**	**0.011**	**0.022**	**0.011**
SCAG-GPT	0.091	0.061	0.068	0.028	0.090
SCAG-GPT (Feedback)	**0.028**	**0.009**	**0.009**	**0.020**	**0.009**

As shown in Table IV, our retrieval-augmented feedback mechanism provides statistical distributions on the length of generated activity chains. Using GPT-4o mini, the best-performing model, we observed substantial improvements across all metrics when testing on both NHTS and SCAG datasets. These results clearly demonstrate that our retrieval-augmented feedback mechanism effectively constrains the model to generate activity chains that more closely match real-world statistical distributions in terms of activity type, timing, duration, and length.

TABLE V: Impact of Retrieval-Augmented Feedback on Household Activity Consistency

Approach	**Consistent**	**Inconsistent**
With RAG Feedback	1,011 (94.1%)	63 (5.9%)
Without RAG Feedback	509 (29.4%)	1,223 (70.6%)

Moving beyond general statistical distributions, we also examine how our approach impacts household coordination activities. Table V demonstrates the critical role of our retrieval-augmented feedback mechanism in eliminating hallucinations and maintaining household activity consistency. With the feedback loop, 94.1% of activities claimed to be performed with household members match corresponding activities by those members, compared to only 29.4% without this mechanism.

This difference highlights how our approach addresses a key challenge in activity chain generation: ensuring logical consistency across interdependent agents. The feedback mechanism reduces hallucinations while producing more realistic household dynamics by ensuring joint activities are actually shared among household members.

Our hallucination elimination strategy creates a cross-verification system where household members' activities are checked against others' reported schedules, reducing hallucinated joint activities from 70.6% to 5.9%. This demonstrates our approach not only generates more accurate activity patterns but also effectively mitigates a primary concern in using LLMs for simulation tasks requiring logical consistency across multiple agents.

VI. Conclusion and Future Work

This study presents a novel LLM-based approach for generating human mobility patterns using minimal socio-demographic data, leveraging GPT-4o mini and Llama2-70b with NHTS and SCAG datasets. Our framework demonstrates strong alignment with real-world patterns through low JSD values, with GPT-4o mini excelling in activity modeling while our retrieval-augmented feedback mechanism reduced hallucinations from 70.6% to 5.9%. While offering substantial benefits for urban planning through reduced data requirements, limitations remain. Future work will extend to multi-day forecasts with richer datasets, specialized fine-tuning, and hybrid models combining learning-based methods with LLMs to enable low-cost, large-scale simulations, establishing new benchmarks in mobility modeling while maintaining computational feasibility.

Acknowledgment

This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via the Department of Interior/Interior Business Center (DOI/IBC) contract number 140D0423C0033. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI/IBC, or the U.S. Government.

References

[1] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, "Human mobility: Models and applications," *Physics Reports*, 2018.

[2] V. M. G. Martinez, D. R. Campelo, and M. R. N. Ribeiro, "Sustainable intelligent transportation systems via digital twins: A contextualized survey," *IEEE Open Journal of Intelligent Transportation Systems*, 2025.

[3] L. Vanhaverbeke and C. Macharis, "An agent-based model of consumer mobility in a retail environment," *Procedia-Social and Behavioral Sciences*, vol. 20, pp. 186–196, 2011.

[4] M. Shen, C. Li, Y. Wu, X. Bi, and F. Xiao, "An extensible python open-source simulation platform for developing and benchmarking bus holding strategies," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 5, 2024.

[5] C. R. Bhat and F. S. Koppelman, "Activity-based modeling of travel demand," in *Handbook of transportation Science*. Springer, 1999.

[6] C. R. Bhat, K. G. Goulias, R. M. Pendyala, R. Paleti, R. Sidharthan, L. Schmitt, and H.-h. Hu, "A household-level activity pattern generation model for the simulator of activities, greenhouse emissions, networks, and travel (simagent) system in southern california," in *91st Annual Meeting of the Transportation Research Board, Washington, DC*, 2012.

[7] E. Ugurel and G. Wang, "Beat the morning rush: Survival analysis-informed dnns with collaborative filtering to predict departure times," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 6, 2025.

[8] G. Salierno, L. Leonardi, and G. Cabri, "A big data architecture for digital twin creation of railway signals based on synthetic data," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 5, 2024.

[9] X. Liao, Q. Jiang, B. Y. He, Y. Liu, C. Kuai, and J. Ma, "Deep activity model: A generative approach for human mobility pattern synthesis," *arXiv preprint arXiv:2405.17468*, 2024.

[10] F. Rocco Di Torrepadula, S. Di Martino, N. Mazzocca, and P. Sannino, "A reference architecture for data-driven intelligent public transportation systems," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 5, 2024.

[11] M. Luca, G. Barlacchi, B. Lepri, and L. Pappalardo, "A survey on deep learning for human mobility," *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–44, 2021.

[12] H. Ma, Y. Liu, Q. Jiang, B. Y. He, X. Liao, and J. Ma, "Mobility ai agents and networks," *IEEE Transactions on Intelligent Vehicles*, 2024.

[13] Y. Liu, C. Kuai, H. Ma, X. Liao, B. Y. He, and J. Ma, "Semantic trajectory data mining with llm-informed poi classification," *arXiv preprint arXiv:2405.11715*, 2024.

[14] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[15] S. A. Stouffer, "Intervening opportunities: a theory relating mobility and distance," *American sociological review*, vol. 5, no. 6, 1940.

[16] S. Rasouli and H. Timmermans, "Activity-based models of travel demand: promises, progress and prospects," *International Journal of Urban Sciences*, vol. 18, no. 1, pp. 31–60, 2014.

[17] X. Kong, K. Wang, M. Hou, F. Xia, G. Karmakar, and J. Li, "Exploring human mobility for multi-pattern passenger prediction: A graph learning framework," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16 148–16 160, 2022.

[18] S. Srinivasan and C. R. Bhat, "Modeling household interactions in daily in-home and out-of-home maintenance activity participation," *Transportation*, vol. 32, pp. 523–544, 2005.

[19] N. Rezvany, M. Bierlaire, and T. Hillel, "Simulating intra-household interactions for in-and out-of-home activity scheduling," *Transportation Research Part C: Emerging Technologies*, vol. 157, 2023.

[20] X. Puig, K. Ra, M. Boben, J. Li, T. Wang, S. Fidler, and A. Torralba, "Virtualhome: Simulating household activities via programs," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8494–8502.

[21] J. Albouys, N. Sabouret, Y. Haradji, M. Schumann, and C. Inard, "Smach: Multi-agent simulation of human activity in the household," in *Advances in Practical Applications of Survivable Agents and Multi-Agent Systems: The PAAMS Collection: 17th International Conference, PAAMS 2019, Ávila, Spain, June 26–28, 2019, Proceedings 17*. Springer, 2019, pp. 227–231.

[22] T. Cai, Y. Liu, Z. Zhou, H. Ma, S. Z. Zhao, Z. Wu, and J. Ma, "Driving with regulation: Interpretable decision-making for autonomous vehicles with retrieval-augmented reasoning via llm," *arXiv preprint arXiv:2410.04759*, 2024.

[23] B. Ni, Z. Liu, L. Wang, Y. Lei, Y. Zhao, X. Cheng, Q. Zeng, L. Dong, Y. Xia, K. Kenthapadi *et al.*, "Towards trustworthy retrieval augmented generation for large language models: A survey," *arXiv preprint arXiv:2502.06872*, 2025.

[24] J. Fish, J. Gonder, V. Garikapati, J. Cappellucci, B. Borlaug, J. Holden, and L. Boyce, "Transportation secure data center," DOE Open Energy Data Initiative (OEDI); National Renewable Energy Laboratory . . . , Tech. Rep., 2015.

[25] U.S. Department of Transportation, Federal Highway Administration, "2009 national household travel survey," https://nhts.ornl.gov, 2009.