Scilight

Journal of Social Physics
https://www.sciltp.com/journals/jsp

Article

A Language Modeling Framework for Generating and Adapting Human Mobility Trajectories

Takayuki Mizuno [1,*,†], Taizo Horikomi [1,†], Shouji Fujimoto [2,†] and Atushi Ishikawa [2]

[1] National Institute of Informatics, Tokyo 101-8430, Japan
[2] Department of Information Engineering, Kanazawa Gakuin University, Kanazawa 920-1392, Japan
* Correspondence: mizuno@nii.ac.jp
† These authors contributed equally to this work.

How To Cite: Mizuno, T.; Horikomi, T.; Fujimoto, S.; et al. A Language Modeling Framework for Generating and Adapting Human Mobility Trajectories. *Journal of Social Physics* **2026**, *1*(1), 2.

Received: 3 October 2025
Revised: 28 December 2025
Accepted: 13 January 2026
Published: 21 January 2026

Abstract: The ability to generate realistic and adaptive synthetic human mobility data is vital for applications ranging from urban planning and epidemiology to disaster response. However, most existing generative models are static, limiting their usefulness in dynamic, real-world scenarios. We introduce a paradigm shift by treating human mobility as a language. We represent sequences of locations and inter-event times as discrete tokens and train a Transformer (GPT-2) model from scratch to learn the underlying grammar of movement. Our approach offers two key capabilities. (1) Conditional generation: by prepending special tokens that encode personal attributes (e.g., gender, age) and environmental context (e.g., weekday/weekend, weather), the model produces trajectories consistent with subgroup-specific mobility patterns. (2) Rapid adaptation: a pre-trained mobility model can be fine-tuned to new, anomalous conditions (e.g., post-disaster mobility) using a small amount of data, achieving faster convergence and higher final accuracy than training from scratch. Across large-scale datasets, our Transformer outperforms Markov-chain and autoregressive baselines in long-horizon location prediction and inter-event time modeling, while closely matching real-world distributional statistics. These findings establish mobility-as-language as a powerful, flexible paradigm for controllable and adaptive trajectory simulation in social physics.

Keywords: human mobility; trajectory generation; transformer model

1. Introduction

The increasing availability of large-scale datasets on individual daily trajectories has become instrumental in decoding the complex patterns of human mobility. This understanding is critical for addressing a gamut of societal challenges, including the mitigation of traffic congestion, the modeling of infectious disease propagation, and the coordination of effective disaster response. For instance, detailed analysis of mobility patterns can identify the root causes of urban bottlenecks [1], leading to more effective traffic management strategies that balance economic activity with public health measures [2–4]. Similarly, during natural disasters or civil unrest, telecommunication data provides a vital channel for monitoring population evacuation and displacement in real time [5, 6]. The development of a generative model capable of replicating the realistic attributes of these trajectories allows for the simulation of urban population dynamics under hypothetical scenarios, such as the introduction of new infrastructure, the outbreak of an epidemic, or a major global event [7–10]. Furthermore, the ability to generate high-fidelity synthetic trajectory data serves as a crucial tool for protecting individual geo-privacy, a growing concern in an era of ubiquitous location tracking [11–13].

The pursuit of realistic trajectory generation has led to the development of a wide array of models, which can be broadly categorized into physics-based and machine learning methodologies. Traditional approaches include gravity models, preferential selection models, and Markov chains, which offer interpretable but often oversimplified representations of human movement [14, 15]. In recent years, deep generative models have emerged as the state-of-the-art, offering the capacity to learn complex, high-dimensional distributions directly from data [16, 17].

Architectures such as Generative Adversarial Networks (GANs) [18] and, more recently, Diffusion Models [19] have demonstrated remarkable success in producing realistic and continuous trajectory paths. Yet these models are typically static: once trained on historical data, they sample from a fixed distribution and are not designed to rapidly adapt to sudden, exogenous shocks (e.g., natural disasters or public-health emergencies). Moreover, much of the literature in this area has focused on short-term, kinematically constrained trajectories such as vehicles or pedestrians, where the main task is next-step prediction under physical constraints. By contrast, our work targets day-long, unstructured human mobility trajectories, which involve fundamentally different challenges. This distinction highlights the novelty of our approach and the need for models capable of capturing long-range dependencies in social contexts.

We introduce a paradigm shift by treating mobility as a language. We discretize continuous spatiotemporal trajectories into sequences of tokens, enabling the use of the Transformer architecture to model human movement. This reframing moves beyond replicating historical patterns and unlocks new capabilities for controllability (via prompt-like attribute tokens) and adaptability (via supervised fine-tuning).

Our contributions are threefold:

1. A Foundational Framework for Spatiotemporal Trajectory Generation: A robust framework is presented for generating human trajectories using a Transformer (GPT-2) architecture. This is achieved through a novel tokenization scheme that converts both geographic locations and the time intervals between them into a unified, sequential representation. The model is trained from scratch, learning the fundamental ”grammar” of human mobility directly from data.
2. High-Fidelity Conditional Generation: A novel method for conditioning trajectory generation on discrete variables is introduced. By prepending special tokens representing personal attributes (e.g., gender, age) or environmental context to the input sequence, the model can generate trajectories that are statistically consistent with the behavior of specific demographic subgroups. This acts as a form of ”prompting” for mobility generation.
3. Rapid adaptation via fine-tuning: We demonstrate that a pre-trained mobility model can be rapidly adapted to anomalous conditions—such as post-earthquake mobility—using only a small dataset, converging faster and to a better final state than training from scratch. We discuss the fine-tuning cost as a potential quantitative indicator of behavioral shock magnitude.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 details the foundational model, including tokenization and architecture. Section 4 presents conditional synthesis and adaptation via fine-tuning. Section 5 describes datasets, metrics, and results. Section 6 discusses implications, limitations, and future directions. Section 7 concludes.

2. Related Works

This section provides a comprehensive, expanded review of the field, situating the present work within the broader landscape of trajectory generation and modeling. The literature can be organized into three convergent streams: the development of powerful generative models for trajectory synthesis, the application of advanced sequence models like the Transformer, and the emerging demand for controllable and adaptive generation methods.

2.1. Generative Models for Trajectory Synthesis

The goal of synthetic trajectory generation is to learn a model of a data distribution $P(T)$ from a set of real trajectories T_{real} such that generated samples $T_{synth} \sim P(T)$ are indistinguishable from real ones. Several families of deep generative models have been applied to this problem.

Generative Adversarial Networks (GANs): GANs employ a two-player min-max game between a generator, which creates synthetic data, and a discriminator, which tries to distinguish synthetic data from real data. This adversarial process has proven effective for generating continuous and realistic paths. Early applications in mobility, such as Social GAN, focused on pedestrian motion forecasting by integrating adversarial training. More recent work has extended this to general human mobility. For instance, LSTM-TrajGAN combines an LSTM architecture with a GAN to generate privacy-preserving synthetic trajectories, demonstrating a key application of this technology [20]. Other variants, like the map-based Two-Stage GAN (TSG) [18], incorporate geographical context by using a Deep Convolutional GAN to learn general spatial patterns and an encoder-decoder network to generate GPS sequences constrained by road networks. While powerful, GANs are difficult to train, often suffering from issues like mode collapse and training instability.

Diffusion Models: As a more recent and highly successful class of generative models, diffusion probabilistic models have emerged as a state-of-the-art approach [21]. These models work by systematically adding Gaussian noise to the data in a "forward process" and then training a neural network to reverse this process, learning to denoise the data step-by-step. This iterative refinement process allows for the generation of high-fidelity and highly diverse samples. In the context of mobility, benchmarks including models like TrajGDM [19], MobilityGen [22], and TrajDD-GAN [23] have been proposed to capture universal mobility patterns or enhance generative efficiency. While these models excel at path-level refinement, our discrete Transformer-based approach enables intuitive semantic conditioning via attribute tokens and superior data efficiency for rapid adaptation through supervised fine-tuning.

Autoencoders (AEs) and Variational Autoencoders (VAEs): These models are primarily used to learn a compressed, low-dimensional latent representation of the data. A VAE, for example, was used in one of the earliest deep learning approaches to generate urban human mobility trajectories. Often, they are combined with other architectures. For instance, a framework by Demetriou et al. uses a Recurrent Autoencoder to learn latent representations of trajectories, which are then generated by a GAN, effectively handling the issue of variable trajectory lengths [24].

2.2. Sequence Modeling and the Rise of Transformers

Human mobility is inherently sequential. Consequently, models designed for sequential data have been central to trajectory analysis.

From Markov Chains to Recurrent Networks: Traditional approaches often relied on first or second-order Markov chain models, which predict the next location based only on the last one or two locations. While simple and interpretable, they fail to capture the long-range dependencies characteristic of human travel (e.g., the morning commute influencing the evening return). Recurrent Neural Networks (RNNs) and their more advanced variants, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, became the standard for sequence modeling by incorporating a hidden state that acts as a memory, allowing them to capture longer-term patterns [25].

The Transformer Architecture: The introduction of the Transformer architecture by Vaswani et al. represented a paradigm shift in sequence modeling [26]. Its core innovation, the self-attention mechanism, allows the model to weigh the importance of all elements in the input sequence regardless of their spatial or temporal distance. This provides a fundamentally more effective mechanism for capturing long-range dependencies compared to the sequential memory of RNNs. While initially developed for NLP, Transformers have been adapted for trajectory-related tasks, such as TrajGen for autonomous driving [27] and various pedestrian forecasting models. The foundational work for this paper by Mizuno et al. was among the first to apply a pure NLP approach, using a GPT-2 architecture trained from scratch on tokenized locations to generate unstructured, day-long trajectories [28]. This approach distinguishes itself by focusing on general human mobility patterns rather than kinematically constrained movements in controlled environments. Recent research has further extended this framework to complex indoor retail environments [29]. The adoption of the Transformer is critical for day-long mobility synthesis. As demonstrated in prior comparative studies [29], conventional recurrent models like LSTMs or GRUs often produce unrealistic "sliding" movements that fail to respect physical constraints (e.g., crossing through walls) or long-term goals, primarily due to the vanishing gradient problem and limited context retention over long sequences. In contrast, the self-attention mechanism ensures direct access to the entire historical context, significantly improving the global coherence of trajectories regardless of sequence length and providing the necessary theoretical basis for this work.

2.3. Conditional Generation and Model Adaptation

As the field matures, the focus is shifting from unconditional generation to creating models that are controllable and adaptable to specific contexts.

Conditional Trajectory Generation: A significant body of research is dedicated to generating trajectories conditioned on certain inputs, such as goals, intentions, or environmental factors. Conditional GANs (CGANs) have been used to generate trajectories conditioned on planning problems or user-controlled speed [30]. Other approaches build conditional probabilistic models to forecast goal-oriented trajectories by first estimating the probability distribution of intentional goals (e.g., destinations) and then generating paths conditioned on those goals [27]. Transformer-based models are also being developed for ”controlled” generation, where the model can fill in missing segments of a partially specified trajectory, analogous to text infilling in NLP [31]. This body of work provides the academic context for the attribute-conditioning contribution presented in this paper.

Model Adaptation and Fine-Tuning: A critical frontier in generative modeling is the ability of models to adapt to new data or changing conditions without being completely retrained. In the broader machine learning landscape, fine-tuning large, pre-trained models has become the dominant paradigm, especially for Large Language Models (LLMs). This approach is now emerging in the mobility domain. For example, recent cutting-edge research explores the use of Reinforcement Learning from Human Feedback (RLHF) to fine-tune generative trajectory models for autonomous driving [32]. In this paradigm, a pre-trained model is further refined using human preference data to align its output with desired driving styles (e.g., more or less aggressive). This paper contributes to this emerging direction by demonstrating a different, yet highly effective, form of adaptation: supervised fine-tuning for rapid adaptation to an exogenous shock (a natural disaster), a novel application that showcases the practical utility of the ”mobility as a language” paradigm for dynamic social systems.

3. A Foundational Sequence Model for Human Mobility

This section details the core methodology of the proposed framework, which involves transforming continuous spatiotemporal data into a discrete token sequence and applying a Transformer-based sequence model. This serves as the foundation upon which the advanced capabilities described in Section 4 are built.

3.1. Formal Problem Statement

Let a raw human mobility trajectory T be defined as a time-ordered sequence of spatiotemporal points, $T = (loc_1, ts_1), (loc_2, ts_2), ..., (loc_m, ts_m)$, where loc_i is a geographical coordinate (latitude, longitude) and ts_i is a timestamp. The primary objective is to learn a generative model $P(T)$ capable of producing synthetic trajectories T_{synth} that are statistically indistinguishable from real-world trajectories.

The approach taken here is to first define a deterministic transformation function $f : T \to S$, which maps a continuous trajectory T into a discrete sequence of tokens $S = (tok_1, tok_2, ..., tok_n)$. The problem is then reformulated as learning the autoregressive probability distribution of this token sequence:

$$P(S) = \prod_{i=1}^{n} P(tok_i | tok_1, tok_2, ..., tok_{i-1}), \tag{1}$$

A trained model can then generate new token sequences autoregressively, which can be converted back into spatiotemporal trajectories.

3.2. Spatiotemporal Tokenization: Creating the Vocabulary of Movement

A critical component of this framework is the tokenization scheme, which must effectively discretize both space and time. To convert continuous geographical coordinates into discrete tokens, a hierarchical grid system based on the Japanese regional grid code JIS X 0410 is utilized. This system recursively subdivides space. For example, a location at a 250-m resolution is represented by a five-character token $X = \zeta_1\zeta_2\zeta_3\zeta_4\zeta_5$.

- ζ_1 represents a primary grid, a unique area enclosed by a square with a 40-min difference in latitude and a 1-degree difference in longitude.
- ζ_2 represents an area formed by dividing the primary grid into an 8×8 matrix.
- ζ_3 is formed by dividing the secondary grid into a 10×10 matrix.
- Subsequent divisions are recursively split into two equal regions in each direction.

Each subdivision at each level is assigned a unique character. A complete location token is thus a “word” composed of “characters” that specify its position with increasing precision. This hierarchical structure is advantageous because it allows the model to learn spatial relationships; nearby locations will share common prefixes in their tokens, analogous to how related words in a language might share a common root.

While this study employs the Japanese grid code for its implementation, it is important to emphasize that the “Mobility as a Language” paradigm is agnostic to the specific spatial indexing system. The underlying logic of mapping hierarchical subdivisions to discrete tokens is directly compatible with global standards such as “Microsoft’s Quadkeys [33]” or “Uber’s H3 [34]”, which utilize similar recursive partitioning logic. Furthermore, this hierarchical nature ensures global scalability; by introducing a higher, “zeroth-level” tier of tokens to represent larger continental or geopolitical regions, the entire world could be mapped with a manageable vocabulary size, thus avoiding a combinatorial explosion of tokens. In the present analysis, a 250-m resolution (348 unique characters for Japan) is sufficient, as it aligns with the typical error margin of general smartphone GPS data.

Trajectories are often characterized not by observations at fixed intervals, but as a series of origin-destination movements with varying time intervals Δt between them. To incorporate this crucial temporal information into the token sequence, the continuous time interval (in minutes) is discretized using a logarithmic scale. A unique character token $r(\tau)$ is assigned to each discrete time interval τ, where τ is calculated as:

$$\tau = \mathrm{int}(\log_{1.5} \Delta t) + 1 \tag{2}$$

This logarithmic scaling is motivated by the observation that the perceptual and behavioral significance of a time difference is often relative. For example, the difference between a 5-min and a 10-min trip is far more significant than the difference between a 60-min and a 65-min trip. This discretization scheme effectively captures these non-linearities.

The location and time interval tokens are interleaved to form a complete representation of a daily trajectory. A trajectory starting at time t_0 from an initial location $X(t_0)$, moving to a second location $X(t_0 + \Delta t_1)$ after a time interval Δt_1, and so on, is represented as the sequence:

$$X(t_0)\_r(\tau_1)X(t_0 + \Delta t_1)\_r(\tau_2)X(t_0 + \sum_{k=1}^{2} \Delta t_k)\_\cdots, \tag{3}$$

where τ_k is the discretized token for the time interval Δt_k. For instance, a simplified day-long trajectory starting from home location X_A, moving to location X_B after a time interval $r(\tau_1)$, and finally returning home after $r(\tau_2)$ would be represented as the discrete token string: $X_A\_r_1X_B\_r_2X_A$. (where the final period marks the completion of the daily routine). Special delimiter tokens are appended to signify important events in the daily routine. A comma character "," is used to denote a temporary return to the home location during the day, and a period character "." signifies the final return home, marking the end of the daily trajectory sequence. The preceding and subsequent locations are connected with a "_" to signify the trajectory.

3.3. Model Architecture and Training

The model architecture used is the GPT-2 SMALL variant proposed by OpenAI. This architecture consists of 12 stacked Transformer decoder blocks. Each block contains a multi-head self-attention mechanism (with 12 attention heads) and a position-wise feed-forward network. The embedding and hidden states have a dimensionality of 768 [35].

It is important to clarify that while we utilize the GPT-2 model architecture (specifically the 12-layer, 12-head configuration), we do not utilize any pre-trained weights derived from natural language corpora (such as OpenAI's released models). Our model is initialized with random weights and trained entirely from scratch using only the tokenized mobility dataset. This ensures that the learned embeddings reflect purely spatiotemporal relationships rather than linguistic semantics. By training from scratch, the model learns the statistical patterns, dependencies, and "grammar" of human mobility directly from the spatiotemporal data, remaining entirely independent of commercial or pre-trained Large Language Models (LLMs) such as ChatGPT.

The training objective is to minimize the cross-entropy loss of predicting the next token in the sequence, given all previous tokens. This autoregressive framework naturally handles trajectories of varying lengths. The model learns the probability distribution of the end-of-sequence token (".") conditioned on the preceding sequence. Generation proceeds token by token until this end token is sampled, implicitly learning the distribution of trajectory lengths observed in the training data without requiring any explicit handling or padding to a fixed length during inference.

The model was trained for 10 epochs using the full dataset. We observed that the cross-entropy loss on a validation set sufficiently converged around the 4th epoch, and no signs of overfitting were detected upon the completion of the 10 epochs. The training was conducted on a single NVIDIA RTX A6000 GPU and took approximately three days to complete. Unless otherwise specified, all other hyperparameters, such as the learning rate, batch size, and optimizer settings, were set to the default values of the standard GPT-2 SMALL implementation [35].

4. Advanced Capabilities: High-Fidelity Conditioning and Rapid Adaptation

Building upon the foundational model, this section introduces the two primary novel contributions of this work: the ability to generate trajectories conditioned on personal attributes and the capacity for rapid adaptation to new scenarios via fine-tuning.

4.1. Conditional Trajectory Synthesis via Attribute Prompting

A key objective in synthetic population generation is to create a heterogeneous set of agents whose behaviors reflect real-world demographic diversity. The proposed framework achieves this through a method analogous to "prompting" in LLMs. To enable conditional generation, the model's vocabulary is expanded to include a set of unique “special tokens” that represent various personal attributes and environmental factors. For this study, we designated a total of 20 special tokens across eight categories:

- Environmental Factors
  - Day of the week: [$Weekday$], [$Weekend$]
  - Temperature: [$h < 25$ °C], [25 °C $\leq h < 30$ °C], [$h \geq 30$ °C]
  - Weather: [$Sunny$], [$Cloudy$], [$Rainy$]
  - Daily COVID-19 cases in Tokyo: [$n < 20,000$], [$20,000 \leq n < 30,000$], [$n \geq 30,000$]
- Personal Attributes
  - Gender: [$Male$], [$Female$]
  - Age: [$Under$ 29], [30 to 59], [$Over$ 60]
  - Home location: [$Urayasu\ city\ (H)$], [$Outside\ of\ the\ city\ (H)$]
  - Work location: [$Urayasu\ city\ (W)$], [$Outside\ of\ the\ city\ (W)$]

The selection of these specific attributes,“Gender, Age, and Weather” is informed by their established statistical significance in human mobility research. Demographic factors such as age and gender are well-documented to correlate with variations in trip purpose, frequency, and travel distance, while environmental conditions like weather significantly influence the choice of transportation mode and the duration of outdoor activities.

To formalize the generation process, we define the Core Prompt Template. Let $A = \{a_1, a_2, \ldots, a_k\}$ be a set of discrete attribute tokens selected from the categories above, and let $T = \{tok_1, tok_2, \ldots, tok_n\}$ be the sequence of spatiotemporal trajectory tokens. The complete input sequence S provided to the model is defined as:

$$S = A \oplus [\triangleright] \oplus T \tag{4}$$

where $\oplus$ denotes the concatenation operator and [$\triangleright$] is a special delimiter token that separates the attribute metadata from the trajectory body. The structure of this combined sequence is illustrated in Table 1.

Table 1. Structure of the Attribute-Conditioned Input Sequence. The sequence consists of a header of metadata tokens followed by the discretized mobility trajectory.

Sequence Segment	Token Representation Example
Attribute Tokens (A)	$[Male], [Under29], [Weekday], [Sunny]$
Delimiter Token	$\triangleright$
Trajectory Tokens (T)	$X(t_0), \_, r(\tau_1), X(t_0 + \Delta t_1), \ldots$

During training, for each trajectory, the corresponding sequence of attribute tokens is prepended as shown in Equation (4). For example, a trajectory for a young male on a weekday would start with:

$$[Male][Under29] \cdots \triangleright X(t_0)\_r(\tau_1)X(t_0 + \Delta t_1)\_ \cdots , \tag{5}$$

The model is then trained on this complete sequence. During inference, a user can specify a set of desired attributes to steer the generative process. If attributes are missing from the dataset, the corresponding special tokens are omitted.

To validate this capability, we conducted an experiment to determine if the model could reproduce gender-based mobility differences observed in real-world data, specifically the tendency for males to have longer average daily travel distances. The model was prompted to generate synthetic trajectories for both [$Male$] and [$Female$] tokens. As shown in Figure 1, the synthetic distributions (b) closely mirror the ground truth (a), with the male population exhibiting a right-shifted travel distance distribution. This demonstrates that the model learns the statistical relationships between attribute tokens and spatiotemporal patterns rather than merely memorizing sequences.

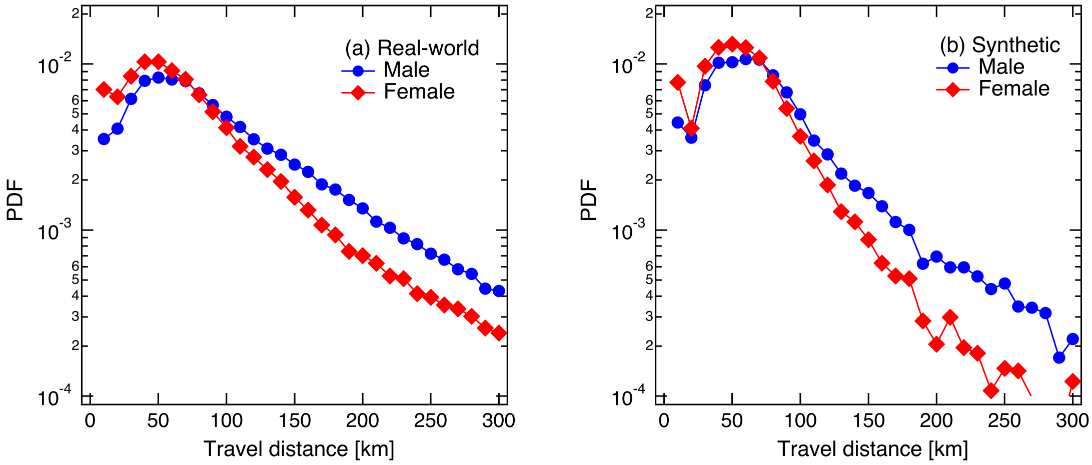

Figure 1. Comparison of Daily Travel Distance Distributions by Gender. Probability distributions of daily travel distance for male and female individuals. (**a**) Distribution observed in the real-world dataset. (**b**) Distribution observed in the synthetic dataset generated by the attribute-conditioned model. The model successfully captures the real-world tendency for males to have longer average travel distances.

4.2. Rapid Adaptation to Exogenous Shocks via Fine-Tuning

The most significant contribution of this work is demonstrating the framework's capacity for rapid adaptation. While traditional simulation models are static, real-world systems are dynamic and subject to sudden, disruptive events. The ability to quickly update a mobility model to reflect post-event behavior is critical for applications like disaster response and epidemiological modeling, forming the basis of a responsive "digital twin".

On 1 January 2024, a major M7.6 earthquake struck the Noto Peninsula in Ishikawa Prefecture, Japan, causing widespread damage and fundamentally altering the mobility patterns of the region's population. This event provides a powerful real-world case study to test the model's adaptability. The experiment was designed to compare the efficiency of fine-tuning versus training from scratch:

- Base Model: A foundational mobility model was first pre-trained on a large dataset of trajectories from Ishikawa Prefecture during August and September 2023. This model represents a comprehensive understanding of "normal" mobility patterns in the region before the disaster.
- Anomalous Data: A small dataset of trajectories was collected from 1 January 2024—the day of the earthquake. This dataset is small because such data is inherently scarce immediately following a major event.
- Two Scenarios:
  - Training from Scratch: A new model using the GPT-2 architecture was trained using only the small post-earthquake dataset.
  - The pre-trained base model was further trained (fine-tuned) using the same small post-earthquake dataset, with the same learning rate (5×10^{-5}) as the base training.

The results, shown in Figure 2, demonstrate the profound advantage of the fine-tuning approach.

The model trained from scratch (blue curve) begins with a high cross-entropy loss of approximately 7.4, indicative of a random initialization. It requires around 24,500 trajectories to converge to a final loss of about 2.5. In contrast, the fine-tuned model (red curve) starts its training on the post-earthquake data with a loss of only 2.7. This low initial loss signifies that the "grammatical rules" of mobility learned during pre-training (e.g., people move along roads, travel occurs between meaningful locations, trips have characteristic durations) are still largely applicable, even in a post-disaster context. The fine-tuning process then specializes this general knowledge to the new, specific patterns of post-earthquake movement, converging rapidly to a final loss of approximately 1.5 after 24,500 trajectories.

The implications are substantial. Fine-tuning is not only far more data-efficient but also results in a more accurate final model. This methodology provides a practical pathway for public authorities and researchers to rapidly update mobility simulations in the immediate aftermath of a crisis, when data is scarce but timely models are most needed. The "cost" of fine-tuning—in terms of data required or convergence time—could itself be interpreted as a novel quantitative metric for the magnitude of a societal behavioral shock. A larger deviation from pre-event patterns would necessitate a more substantial model update, providing a principled way to measure the impact of an event on a social system.

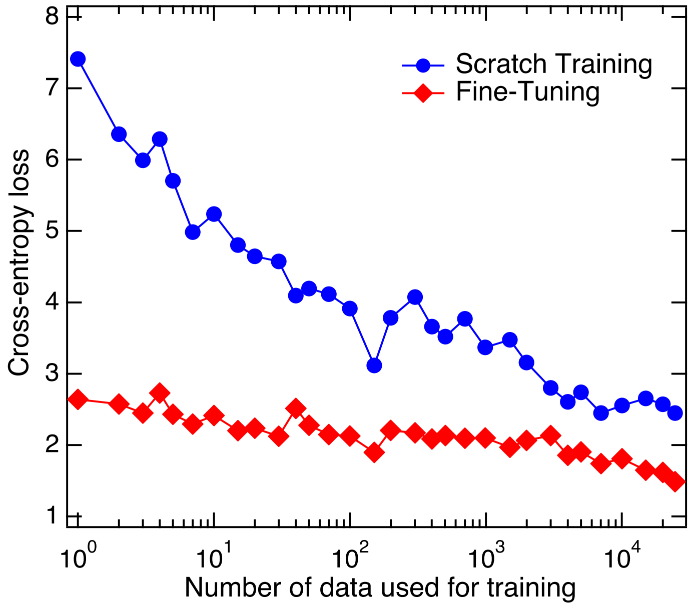

Figure 2. Comparison of Learning Curves for Scratch Training vs. Fine-Tuning. Cross-entropy loss as a function of the number of training trajectories for the post-earthquake scenario. The blue curve shows a model trained from scratch, which starts with a high loss and requires approximately 24,500 trajectories to converge. The red curve shows the pre-trained model being fine-tuned; it starts with a much lower loss and converges to an even better final state more quickly.

5. Experimental Evaluation

This section details the dataset, defines the evaluation metrics used to assess model performance, and presents a quantitative and qualitative analysis of the foundational and conditional models.

5.1. Dataset and Preprocessing

This study utilizes two distinct datasets provided by Agoop Corp. [36] for different experimental purposes.

Urayasu City Dataset (for Conditional Generation and Performance Evaluation): The primary dataset for the conditional generation (Section 4.1) and general performance evaluation (Section 5.3) consists of anonymized daily trajectory data from approximately 590,000 smartphones located in Urayasu City, Chiba, Japan, during August 2022. This region is notable for containing Tokyo Disney Resort, a major tourist destination, as well as significant residential and commercial areas. The raw data includes latitude, longitude, timestamps, and unique smartphone identifiers. The temporal resolution of the data is variable, with a mean interval between location points of 24.27 minutes. To protect geo-privacy, all location data within a 100-radius of each user's inferred home location were omitted from the dataset. The remaining data, comprising approximately 21 million coordinates, was used for the study. Crucially for the conditional synthesis experiments, this dataset also contains linked demographic information for a subset of users, including gender (71% of users) and age (64% of users). The full dataset was split into training and testing sets at a 4:1 ratio for these experiments.

Ishikawa Prefecture Dataset (for Fine-Tuning Experiment): For the rapid adaptation experiment (Section 4.2), trajectory data from Ishikawa Prefecture was used. The pre-training dataset, representing ”normal” behavior, was collected in August and September 2023 from approximately 2,175,000 smartphones. The post-earthquake dataset, representing “anomalous” behavior, was collected on 1 January 2024, from approximately 245,000 smartphones that were present in the prefecture on the day of the event.

5.2. Evaluation Metrics

To rigorously evaluate the models' performance, a combination of metrics is used to assess location accuracy, temporal accuracy, and distributional similarity.

Location Prediction Accuracy: This metric evaluates the model's ability to predict a user's future location. For a given trajectory from the test set, the first four locations and three time intervals are provided as an input prompt. The model then autoregressively generates the subsequent sequence. The “Probability that the prediction is within d km” is defined as the fraction of test trajectories where the Euclidean distance between the model's predicted location at a future time horizon h (e.g., 1 h, 2 h) and the ground-truth location at that same time is less than a distance threshold d (e.g., 3 km or 10 km).

Time Interval Prediction Accuracy: To assess the accuracy of the generated time intervals between locations, the Mean Absolute Logarithmic Error ($MALE$) is used. This metric is defined as $MALE = \frac{1}{N}\sum_{i=1}^{N} |log(\Delta t_{pred,i}) - log(\Delta t_{actual,i})|$. It is particularly suitable for data spanning several orders of magnitude, as is the case with human mobility time intervals.

Distributional Similarity: To evaluate how well the generated trajectories capture the macroscopic statistical properties of the real data, the cumulative distribution functions (CDFs) of key mobility indicators are compared. This includes the hourly moving distance (in a straight line) and the duration of time intervals between movements. A close match between the CDFs of the synthetic and real data indicates that the model has successfully learned the underlying distributions.

5.3. Performance of the Foundational and Conditional Models

The Transformer-based models (using the GPT-2 architecture) were benchmarked against simpler, standard time-series models: first and second-order Markov chain models for location generation and an Autoregressive model of order 3 (AR(3)) for time interval generation. While much of the existing literature on trajectory prediction focuses on short-term, kinematically constrained movements (e.g., vehicles or pedestrians) and thus employs different sets of baselines, our work addresses the distinct challenge of generating unstructured, day-long human mobility trajectories. For this less-common, long-horizon generation task, these fundamental time-series models serve as robust and appropriate baselines to clearly demonstrate the value of capturing long-range dependencies, a core contribution of our approach in the context of social physics.

While recurrent models such as LSTMs and GRUs are standard for short-term trajectory tasks, their inherent limitations in long-horizon generation—specifically the unrealistic "sliding" movements and failure to respect physical constraints described in Section 2.2—render them less suitable for day-long mobility synthesis. Our preliminary experiments confirmed that even when extending recurrent architectures with kinematic features (e.g., velocity and acceleration), they struggle to maintain global coherence in dense urban environments. Consequently, to clearly demonstrate the Transformer's superior ability to capture long-range dependencies, we prioritize baselines that provide a direct contrast in sequence modeling performance, such as Markov-chain and autoregressive models. Future work will continue to explore direct benchmarks against optimized recurrent frameworks.

Table 2 shows the location prediction accuracy. The Transformer-based models significantly outperform both Markov chain models across all time horizons. The accuracy of the Markov models decays rapidly with time, as they are incapable of capturing long-range dependencies. In contrast, the Transformer model maintains a hit rate of nearly 20% even for predictions many hours into the future. Furthermore, the inclusion of environmental and individual attributes in the conditional Transformer model consistently improves prediction accuracy over the foundational model. This suggests that these attributes provide valuable context that the model effectively integrates into its generative process. A detailed analysis of the internal mechanisms behind this improvement is provided in Section 6.

Table 2. Probability that the predicted location is within 3 km (10 km) of the actual location coordinates for various future time horizons. The Transformer-based models demonstrate superior long-term prediction accuracy compared to Markov models, and the inclusion of attributes further enhances performance.

Model	1 h	2 h	4 h	8 h	Final Time of Day
1st-order Markov chain	0.20 (0.36)	0.13 (0.28)	0.07 (0.17)	0.02 (0.06)	0.00 (0.00)
2nd-order Markov chain	0.25 (0.40)	0.17 (0.30)	0.09 (0.20)	0.03 (0.08)	0.00 (0.00)
Transformer (Foundational)	0.33 (0.61)	0.28 (0.51)	0.22 (0.42)	0.15 (0.29)	0.15 (0.22)
Transformer (with Attributes)	0.36 (0.64)	0.31 (0.54)	0.26 (0.47)	0.17 (0.32)	0.15 (0.27)

Table 3 compares the accuracy of time interval prediction. The Transformer model achieves a lower MALE than the AR(3) model for the next four predicted time intervals, indicating that the joint modeling of space and time provides benefits for predicting temporal patterns as well.

The distributional similarity plots (Figures 3 and 4) further confirm the model's fidelity. The CDF of hourly moving distance generated by the Transformer model closely matches the ground truth, capturing the heavy-tailed nature of human travel distances. Similarly, the CDF for time intervals shows that the Transformer model reproduces the distribution of stay/travel times with high precision, unlike the AR(3) model which deviates significantly.

Table 3. Mean Absolute Logarithmic Errors (MALE) for the next four predicted time intervals. The integrated spatiotemporal Transformer model outperforms the time-only AR(3) model.

Model	Next	Second	Third	Fourth
AR(3) with a lower bound	0.651	0.648	0.653	0.650
Transformer Model	0.581	0.574	0.577	0.578

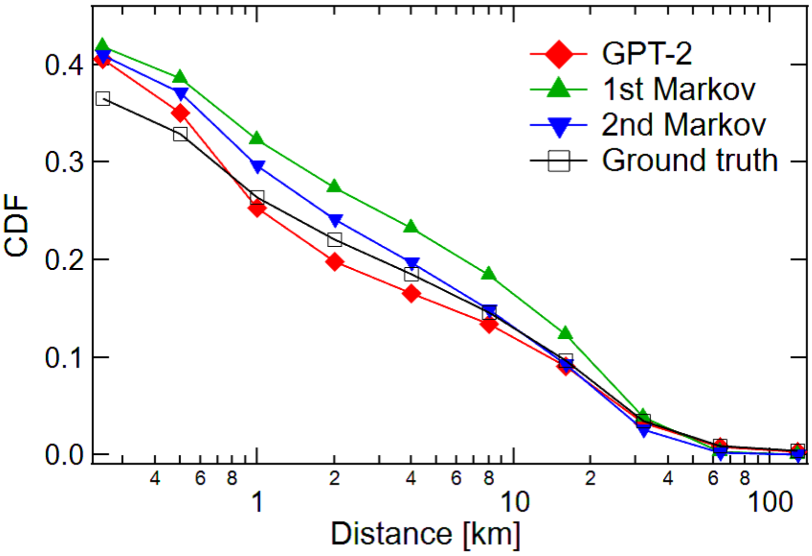

Figure 3. Cumulative distribution of hourly moving distance in a straight line. The distribution from (♦) the Transformer model closely tracks (□) the ground truth, outperforming (▲) the first-order and (▼) the second-order Markov models.

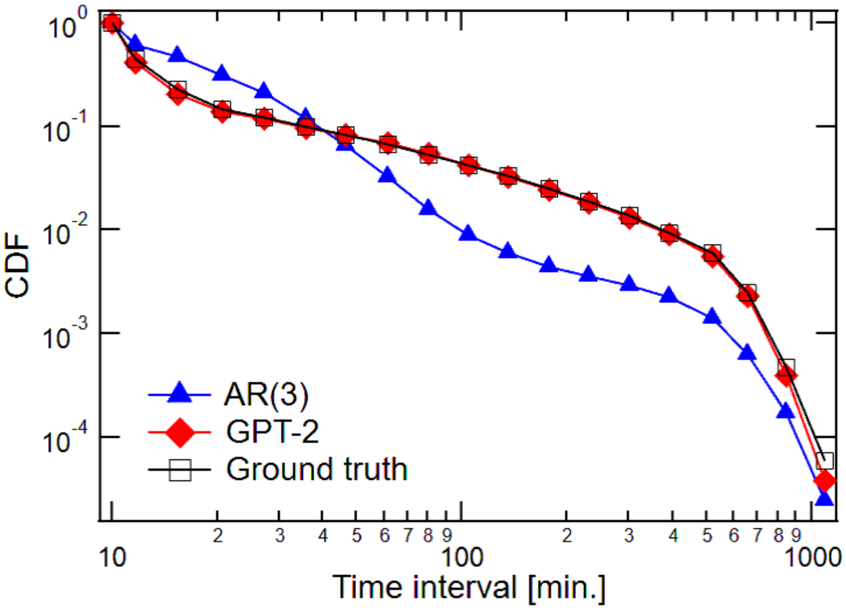

Figure 4. Cumulative distribution of time intervals between movements. (♦) The Transformer model's output aligns almost perfectly with (□) the ground truth, while (▲) the AR(3) model shows significant deviation.

5.4. Qualitative Analysis

To provide an intuitive sense of the model's generative capabilities, Figure 5 visualizes several example daily trajectories generated by the foundational Transformer model. Each trajectory starts from a different initial location. The model successfully generates plausible round trips, where the final location is near the initial location, typifying the common daily pattern of departing from and returning home. The orange and purple trajectories, initiated far from the main study area, demonstrate the model's ability to generate coherent long-distance travel as well.

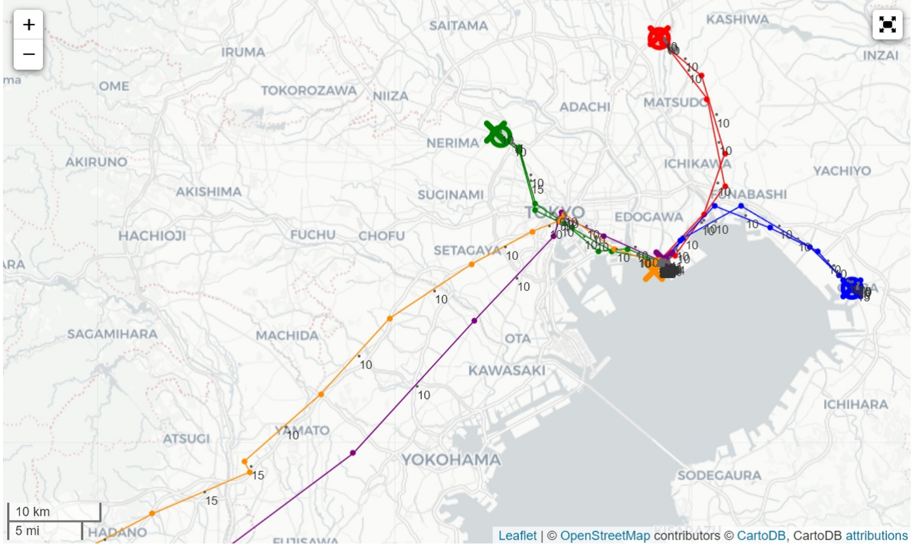

Figure 5. Five examples of individual daily trajectories generated by the Transformer model, each shown in a different color. Start points are marked with circles and end points with squares. The model generates plausible return trips (final location near initial location) for local trajectories and coherent paths for long-distance travel. The initial locations of the orange and purple trajectories were set to the locations of Nagoya City Hall and Kyoto Station, which are each more than 250 km away from Urayasu City in linear distance.

6. Discussion

The results presented in this paper strongly support the proposition that human mobility can be effectively modeled as a language. The success of attribute-conditioning, which functions as a form of "prompting", and the remarkable efficiency of fine-tuning for scenario adaptation, demonstrate the power of this paradigm. This approach elevates the task of trajectory generation from simple pattern mimicry to a more flexible, controllable, and adaptable form of synthesis. By treating mobility sequences as sentences, we unlock a rich theoretical and practical toolkit from NLP that can be applied to problems in social physics.

The fine-tuning experiment on the Noto earthquake data has implications that extend beyond mere technical efficiency. It offers a novel, data-driven method for analyzing and quantifying the resilience and adaptation of social systems. The fact that the pre-trained model provided a strong baseline (low initial loss) for the post-disaster data suggests that a core "grammar" of mobility persists even during crises. The fine-tuning process then learns the "delta"—the specific, context-dependent changes in behavior. The computational and data cost required to learn this delta can be interpreted as a quantitative measure of the magnitude of the behavioral shock. This "fine-tuning cost" could potentially be developed into a standard metric for assessing societal disruption, providing a new analytical tool for the field of social physics.

While the Transformer model operates as a black box, its internal mechanisms can offer preliminary insights into the learned "grammar" of movement. An analysis of the model's attention weights provides quantitative evidence for a hierarchical processing of information, as shown in Figure 6.

In the lower layers of the network (layers 1–6), attention is predominantly focused on recent locations and time intervals, learning local movement dynamics. This is evident from the high attention weights assigned to the 'X' (past locations) and 'r' (past intervals) tokens. As we move to the upper layers (7–12), the model begins to integrate higher-level context. The attention weights for the attribute tokens (such as age, gender, and day of the week) increase significantly. Notably, this effect is more pronounced when generating time intervals (right panel) compared to locations (left panel). In the upper layers of the time-interval generation process, attributes like the daily coronavirus case count ($a(cov)$), age ($a(age)$), weather ($a(wth)$), and gender ($a(gen)$) each receive an attention weight of approximately 0.1.

This hierarchical structure suggests that the model learns to first establish a baseline trajectory based on recent history and then refines its predictions—specifically stay durations and travel timing—using the broader context provided by the attributes. This mechanistic behavior explains the superior accuracy of the conditional Transformer observed in Section 5.3 and demonstrates that the model is capturing nuanced behavioral patterns consistent with social physics principles.

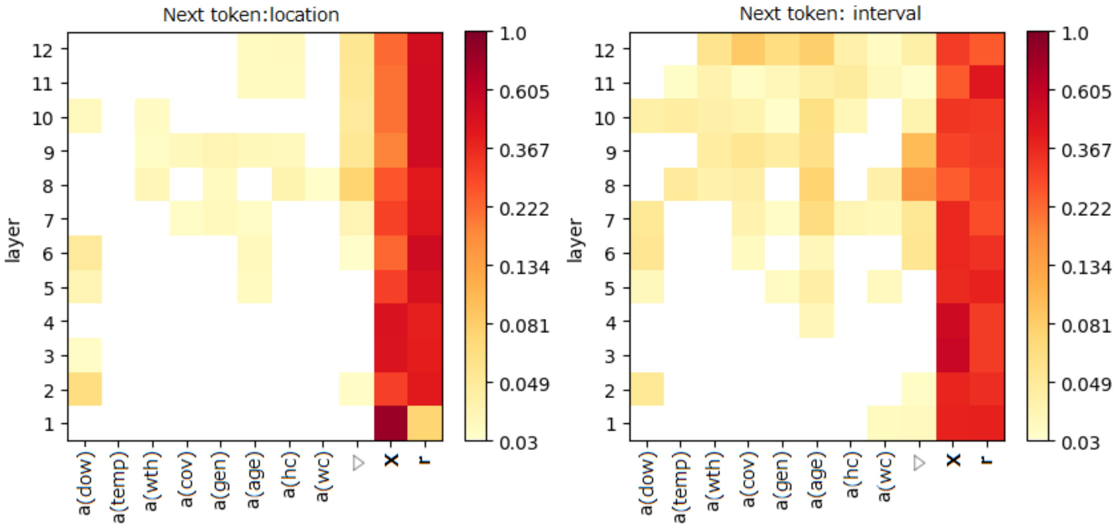

Figure 6. Attention weights for eight special attribute tokens (A), as well as cumulative weights for past location (X) and past time interval (r) tokens, across the 12 layers of the Transformer. The left panel shows weights when generating the next location, and the right panel shows weights for generating the next time interval.

This work positions the “Mobility as a Language” paradigm as a powerful and flexible framework for human mobility modeling. Its primary strengths lie in the ability to capture complex long-range dependencies and the efficiency of supervised fine-tuning for disaster adaptation. However, several limitations warrant further discussion. First, while the discrete tokenization of space results in a loss of spatial precision compared to continuous paths, this discretization is precisely what enables the model to leverage the Transformer’s attention mechanism. As demonstrated by our hierarchical analysis, this approach transforms trajectory generation from a black-box path-smoothing task into an interpretable process of semantic synthesis. The trade-off in spatial precision is thus counterbalanced by the gain in intuitive control via attribute prompting, a capability that remains challenging in purely continuous frameworks. Second, although training large Transformers from scratch is computationally expensive, our results show that once a foundational model is established, it can be adapted to localized or anomalous scenarios with minimal data and time.

The “mobility as a language” paradigm opens several exciting avenues for future research. The most immediate and promising direction is the development of a multi-modal framework for map-conditioned trajectory generation. This would involve extending the current model to accept a map image as an additional input. A Convolutional Neural Network (CNN) could be used to encode the road network and other geographical features into a feature embedding. This embedding would then condition the Transformer’s generation process, forcing the model to produce trajectories that are not only statistically realistic but also physically constrained to the underlying road network. This would combine the sequential modeling strength of the Transformer with the spatial awareness of CNNs.

Another critical area for future work is the generation of collective trajectories. Modeling the interactions between multiple agents is essential for simulating congestion and crowd dynamics. This could be explored by adapting multi-agent architectures or by incorporating interaction-aware attention mechanisms into the current framework.

7. Conclusions

This paper introduced a comprehensive and flexible framework for human mobility modeling based on the paradigm of “mobility as a language”. By tokenizing spatiotemporal data and applying a model based on the Transformer architecture, it is possible to not only generate realistic synthetic trajectories but also to control and adapt the generative process in powerful new ways.

The key contributions were the demonstration of two advanced capabilities. First, a method for high-fidelity conditional generation was presented, allowing for the synthesis of trajectories consistent with specific personal and environmental attributes by “prompting” the model with special tokens. Second, and most significantly, the paper provided the first demonstration of using supervised fine-tuning to rapidly adapt a pre-trained mobility model to a new, anomalous scenario. The case study of the 2024 Noto earthquake showed that this approach is vastly more data-efficient and effective than training a model from scratch, offering a practical methodology for creating the dynamic digital twins needed for real-time crisis response.

The success of this approach validates the ”mobility as a language” concept and opens up new avenues for dynamic modeling and quantitative analysis in social physics and computational social science. By bridging the gap between the static nature of traditional models and the dynamic reality of human societies, this work provides a foundational step toward more responsive, adaptive, and realistic simulations of our complex world.

Author Contributions

T.M.: conceptualization, resources, data curation, funding acquisition, supervision, visualization, methodology, writing—original draft, project administration, writing—review and editing; T.H.: formal analysis, validation, investigation; S.F.: formal analysis, funding acquisition, validation, investigation, visualization; A.I.: funding acquisition, methodology, writing – review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by JSPS KAKENHI (Grant Numbers JP25K01458, JP25K01453, JP25K08191, and JP22K04609), RIOS NII Open Collaborative Research (2025-251S9-22495), The Obayashi Foundation, and JST CREST (Grant Number JPMJCR20D3). The authors declare that financial support was received for the research and publication of this article.

Institutional Review Board Statement

Ethical review and approval were waived for this study. This is because the study involved a secondary analysis of a large-scale human mobility dataset provided by a third-party corporation. The data was fully anonymized by the provider before the authors gained access, making it impossible to identify any specific individuals.

Informed Consent Statement

Informed consent was waived for this study. This research is based on a secondary analysis of a fully anonymized dataset provided by the data vendor, Agoop Corp. The responsibility for obtaining consent for the collection and use of anonymized data for research purposes rested with the original data provider, in accordance to their user agreements and privacy policies.

Data Availability Statement

The human mobility data that support the findings of this study were provided by Agoop Corp. under a license agreement. Due to commercial and privacy restrictions, these data are not publicly available. This data is commercially available for purchase, and inquiries can be directed to Agoop Corp. (https://www.agoop.co.jp/).

Acknowledgments

The authors acknowledge Agoop Corp. for making the smartphone location data available for this study. An earlier version of this work was presented at the Conference on Complex Systems 2023 (CCS 2023), and we are grateful to the conference participants for their valuable feedback.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used Gemini 2.5 Pro solely for English-language proofreading. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Zhu, L.; Yu, F.R.; Wang, Y.; et al. Big data analytics in intelligent transportation systems: A survey. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 383–398.
2. Chang, S.; Pierson, E.; Koh, P.W.; et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **2021**, *589*, 82–87.

3. Deb, P.; Furceri, D.; Ostry, J.D.; et al. The economic effects of COVID-19 containment measures. *Open Econ. Rev.* **2022**, *33*, 1–32.
4. Mizuno, T.; Ohnishi, T.; Watanabe, T. Visualizing social and behavior change due to the outbreak of Covid-19 using mobile phone location data. *Open Econ. Rev*. **2021**, *39*, 453–468.
5. Sudo, A.; Kashiyama, T.; Yabe, T.; et al. Particle filter for real-time human mobility prediction following unprecedented disaster. In Proceedings of the SIGSPACIAL'16: 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Burlingame, CA, USA, 31 October–3 November 2016.
6. Rotman, A.; Shalev, M. Using location data from mobile phones to study participation in mass protests. *Sociol. Methods Res.* **2020**, *51*, 1357–1412.
7. Cutter, S.L.; Ahearn, J.A.; Amadei, B.; et al. Disaster resilience: A national imperative. *Environ. Sci. Policy Sustain. Dev.* **2013**, *55*, 25–29.
8. WMO-UNISDR. *Disaster Risk and Resilience*; Thematic Think Piece; UN System Task Force on the Post-2015 UN Development Agenda: Geneva, Switzerland, 2012.
9. Yabe, T.; Tsubouchi, K.; Sekimoto, Y. Cityflowfragility: Measuring the fragility of people flow in cities to disasters using GPS data collected from smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 1–17.
10. Yabe, T.; Tsubouchi, K.; Sudo, A.; et al. A framework for evacuation hotspot detection after large scale disasters using location data from smartphones: Case study of Kumamoto earthquake. In Proceedings of the SIGSPACIAL'16: 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Burlingame, CA, USA, 31 October–3 November 2016; Volume 44.
11. Fiore, M.; Katsikouli, P.; Zavou, E.; et al. Privacy in trajectory micro-data publishing: A survey. *Trans. Data Priv.* **2020**, *13*, 91–149.
12. Mir, D.J.; Isaacman, S.; Caceres, R.; et al. Dp-where: Differentially private modeling of human mobility. In Proceedings of the 2013 IEEE International Conference on Big Data, Santa Clara, CA, USA, 6–9 January 2013; pp. 580–588.
13. Pellungrini, R.; Pappalardo, L.; Simini, F.; et al. Modeling adversarial behavior against mobility data privacy. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 1145–1158.
14. Schlapfer, M.; Dong, L.; O'Keeffe, K.; et al. The universal visitation law of human mobility. *Nature* **2021**, *593*, 522–527.
15. Song, C.; Koren, T.; Wang, P.; et al. Modelling the scaling properties of human mobility. *Nat. Phys*. **2010**, *6*, 818–823. https://doi.org/10.1038/nphys1760.
16. Luca, M.; Barlacchi, G.; Lepri, B.; et al. A survey on deep learning for human mobility. *ACM Comput. Surv*. **2023**, *55*, 1–44. https://doi.org/10.1145/3485125.
17. Toch, E.; Lerner, B.; Ben-Zion, E.; et al. Analyzing large-scale human mobility data: A survey of machine learning methods and applications. *Knowl. Inf. Syst.* **2019**, *58*, 501–523.
18. Jiang, W.; Zhao, W.X.; Wang, J.; et al. Continuous trajectory generation based on two-stage GAN. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; pp. 4374–4382.
19. Chu, C.; Zhang, H.; Wang, P.; et al. Simulating human mobility with a trajectory generation framework based on diffusion model. *Int. J. Geogr. Inf. Sci.* **2024**, *38*, 847–878.
20. Rao, J.; Gao, S.; Kang, Y.; et al. LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection. In the Proceedings of the 11th International Conference on Geographic Information Science, Poznań, Poland, 27–30 September 2021; Volume 12.
21. Kong, X.; Chen, Q.; Hou, M.; et al. Mobility trajectory generation: A survey. *Artif. Intell. Rev.* **2023**, *56*, 3057–3098.
22. Hong, Y.; Zhang, Y.; Schindler, K.; et al. Deep Generative Model for Human Mobility Behavior. *arXiv* **2025**, arXiv:2510.06473. 2025.
23. Abbar, H.; Kassan, S.; Bidet, F.; et al. TrajDD-GAN: A Synthetic Mobility Trajectory Generation Solution Based on Diffusion Models. *IEEE Access* **2025**, *13*, 158018–158032.
24. Demetriou, A.; Alfsvåg, H.; Rahrovani, S.; et al. A Deep Learning Framework for Generation and Analysis of Driving Scenario Trajectories. *SN Comput. Sci.* **2023**, *4*, 251.
25. Graser, A.; Jalali, A.; Lampert, J.; et al. MobilityDL: A review of deep learning from trajectory data. *GeoInformatica* **2024**, *29*, 115–147.
26. Vaswani, A.; Shazeer, N.; Parmar, N.; et al. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762
27. Zhang, Q.; Gao, Y.; Zhang, Y.; ; et al. TrajGen: Generating Realistic and Diverse Trajectories With Reactive and Feasible Agent Behaviors for Autonomous Driving. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 24474–24487.
28. Mizuno, T.; Fujimoto, S.; Ishikawa, A. Generation of individual daily trajectories by GPT-2. *Front. Phys.* **2022**, *10*, 1021176.
29. Horikomi, T.; Mizuno, T. Generating in-store customer journeys from scratch with GPT architectures. *Eur. Phys. J. B* **2024**, *97*, 144.
30. Barbi, T.; Nishida, T. Trajectory Prediction using Conditional Generative Adversarial Network. Proceedings of the 2017 International Seminar on Artificial Intelligence, Networking and Information Technology, Bangkok, Thailand, 2–3 December 2017; pp. 193–197.
31. Fujimoto, S.; Ishikawa, A.; Mizuno, T. RoBERTa Trained from Scratch on GPS Trajectory Data. In Proceedings of

IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology, Venice, Italy, 26–29 October 2023; pp. 636–639.

32. Li, D.; Li, C.; Wang, Y.; et al. Learning Personalized Driving Styles via Reinforcement Learning from Human Feedback. *arXiv* **2025**, arXiv:2503.10434
33. Schwartz, J. Bing Maps Tile System (2024). Microsoft Corporation. Available online: https://learn.microsoft.com/en-us/bingmaps/articles/bing-maps-tile-system (accessed on 28 December 2025).
34. Brodsky, I. H3: Uber's Hexagonal Hierarchical Spatial Index. Uber Engineering Blog. 2018. Available online: https://www.uber.com/blog/h3/ (accessed on 28 December 2025).
35. Radford, A.; Wu, J.; Child, R.; et al. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.
36. Agoop-Corp. Dynamic Population Data (2025). Available online: https://www.agoop.co.jp/ (accessed on 1 October 2025).