

Primitives and world model for road transportation foundation models

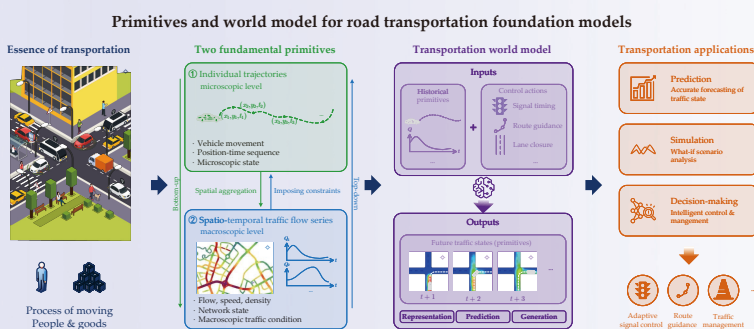
Haiyang Yu¹, Binghong Jiang², Xiaobo Qu³, Li Li²

Cite this article: Yu H, Jiang B, Qu X, et al. *Commun Transp Res* 2026, 6(3): 9640042. <https://doi.org/10.26599/COMMTR.2026.9640042>

ABSTRACT: Transportation foundation models are regarded as key infrastructure for achieving the leap from informatization to intellectualization in road transportation systems and thus have become one of the most focused research directions among transportation researchers. Starting from first principles and based on the fundamental transportation data currently available, this study abstracts two types of primitives as the core and fundamental data building blocks for transportation foundation models. Specifically, individual travel

data are represented by vehicle trajectories as the first type of primitive data, with travel attribute data serving as labels for trajectory data. For data at specific spatial scales, traffic flow spatiotemporal series data of people or vehicles within a given time window are adopted as the second type of primitive data. Furthermore, this study discusses how to construct a world model for transportation that generates primitive data representing future traffic states. This world model takes as input both sensor-acquired primitive data describing historical traffic states and time-series data of potential control and management actions (e.g., signal timing, route guidance, or lane closure plans), enabling it to simulate the outcomes of different intervention strategies. These primitive data will provide critical data foundations for reasoning, simulation, and decision-making in upper-level applications of transportation foundation models, serving as a bridge connecting data perception and intelligent decision-making.

KEYWORDS: transportation; foundation models; primitives; world model; large language models (LLMs); tokens



1 Introduction

With the advancement of artificial intelligence (AI), the concept of foundation models was introduced in approximately 2020 and has since rapidly become a research hotspot in AI and numerous application domains (Bommasani et al., 2021). Foundation models are pretrained machine learning models built on massive amounts of data. They are typically pretrained through unsupervised learning on large-scale general-purpose data (e.g., text, images), possessing cross-task generalizability, and are then adapted to specific scenarios via fine-tuning or prompt engineering.

Two concepts are frequently mentioned alongside foundation models. The first is artificial general intelligence (AGI), which refers to AI systems capable of learning, reasoning, and adapting to environments as broadly as humans, aligning with human values and ethical norms. AGI can not only handle single tasks but also transfer knowledge across domains and even set goals autonomously, making independent decisions across various tasks

and achieving or surpassing human-level performance. Foundation models are largely viewed as one promising path toward AGI, providing foundational knowledge and capabilities, although additional techniques and methods are still required to realize full AGI. The second is large language models (LLMs), which are pretrained on massive text corpora to capture semantic structures, thereby learning and mastering general language generation and comprehension capabilities. LLMs are generally considered a relatively mature type of foundation model and have found increasingly widespread and important applications in natural language processing and its derivative fields, such as scientific exploration, education, and healthcare (Zhou et al., 2025).

Intelligent transportation systems (ITSs), after decades of development, have gradually transitioned from the early digitization of road transportation systems toward transportation intellectualization, shifting research focus from data collection and preprocessing to a deep understanding and synthesis of information. However, traditional research has often built

¹ School of Transportation Science and Engineering, Beihang University, Beijing 100191, China. ² Department of Automation, Tsinghua University, Beijing 100084, China. ³ School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China.

✉ Corresponding authors. E-mail: X. Qu, xiaobo@tsinghua.edu.cn; L. Li, li-li@tsinghua.edu.cn

Received: May 20, 2026; Revised: July 26, 2026; Accepted: July 31, 2026

© The Author(s) 2026. This is an open access article under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0, <http://creativecommons.org/licenses/by/4.0/>).

dedicated data pipelines for data collection, processing, and application development tailored to different transportation applications. These pipelines frequently fail to fully leverage multimodal transportation data, the accuracy of intelligent models for individual transportation applications cannot be mutually validated with other transportation application models, and the resulting intelligent models often suffer from insufficient generalization performance.

Before delving into the design of primitives, it is important to clarify the scope of this study. While the term "transportation" broadly encompasses various modes of mobility, including air, maritime, rail, and road, the scope of this study is primarily focused on road traffic systems. In the rest of this study, we usually omit the term "road" for presentation simplicity.

Stimulated by the remarkable achievements of LLMs, transportation researchers have also begun to explore the application of foundation model technology in the transportation domain, hoping to address the challenges faced by ITSs (Li et al., 2024; Wandelt et al., 2024; Zhang et al., 2023; Zhao et al., 2022). The success of LLMs has led researchers to believe that developing appropriate transportation foundation models, by learning from multimodal and multipattern transportation data to extract universal features and patterns, can provide a powerful foundation for various applications in the transportation field (Nie et al., 2025; Zhou et al., 2024a).

LLMs inherently take the smallest textual unit—the token—as their fundamental building block, converting input text into sequences of tokens and completing tasks by analyzing contextual tokens to predict and generate new sequences of tokens as output. However, what exactly constitutes the primitives of transportation foundation models remains an open question (Zou et al., 2025). Without a definition of such primitives, subsequent research on transportation foundation models would be like building a tower on sand, making robust development difficult.

To this end, this study aims to return to the first principles of transportation systems, design primitive data for transportation foundation models, and thereby establish a solid data foundation and core architecture. We first argue that the essence of transportation systems—"the movement of people and goods in space and time"—can be abstracted into the two most fundamental types of data primitives: Trajectory sequences describing microscopic individual movements and spatiotemporal flow rate sequences describing macroscopic aggregate states. These two types of primitives possess completeness, composability, and interpretability and naturally align with both machine processing and human cognition, making them ideal atomic units for building higher-level intelligent applications.

Furthermore, this study proposes that a mature transportation foundation model must inherently include a "world model" capable of continuously inferring future system states based on historical primitive data. We find that existing advanced frameworks in the transportation domain that integrate data preprocessing with state prediction already constitute, in essence, efficient and reliable specialized world models. These models differ substantially from traditional reductionism-based microscopic traffic simulation, representing a paradigm shift from "rule-driven simulation" to "data-driven state inference".

The rest of this paper is organized as follows: Section 2 provides a literature review. Section 3 systematically elaborates on the design requirements for primitives of transportation foundation models and presents our two-primitive scheme. Section 4 demonstrates in detail why the integrated processing framework constitutes a transportation world model, contrasts it with

traditional simulation, and clarifies its necessity as a core component of foundation models. Finally, Section 5 concludes the paper.

2 Literature review

Current discussions on transportation foundation models are still in their early stages. Researchers either analogize the working mechanisms of LLMs to conceptualize future architectures of transportation foundation models or directly design problem-solving approaches for transportation tasks based on LLMs.

Zhang et al. (2024a) categorized the application of LLMs to mobility prediction tasks in transportation systems into three paradigms based on execution approaches: Fine-tuning existing LLMs to adapt to transportation tasks; zero-shot/few-shot inference that directly utilizes pretrained models for prediction; and integration into larger frameworks as feature extractors or intermediate processors. These methods collectively enhance the accuracy, interpretability, and cross-scenario adaptability of traffic prediction. Regarding input data, the authors discuss various strategies for converting data into LLM-comprehensible input formats, including tokenization (e.g., converting spatiotemporal points into tokens), prompt engineering (injecting domain knowledge and task instructions), and embedding representations (encoding data into semantic vectors), thereby enabling multimodal fusion of structured time-series data (e.g., traffic flow, speed) with unstructured contextual information (e.g., weather, event reports).

Gruver et al. (2023) directly converted numerical traffic flow values into corresponding characters to achieve zero-shot learning for traffic flow time series. However, this approach suffers from insufficient compression and extraction of input data, low processing efficiency, and inadequate preprocessing and mining of traffic features. Similar strategies have been adopted by Xu and Liu (2025) and Zhang et al. (2024b).

Li et al. (2024) proposed a three-step spatiotemporal-text alignment approach for tokenization in input data processing. First, spatiotemporal feature encoding: Raw spatiotemporal data (e.g., pedestrian/vehicle flows) are mapped into embedding vectors through linear layers, followed by multigranularity temporal feature extraction via gated dilated convolutional layers and multilevel correlation injection layers to form spatiotemporal dependency representations. Second, projection and token insertion: These representations are projected into the LLM latent space through a lightweight alignment module and converted into special token sequences, such as <ST_start>, <ST_HIS>, ..., <ST_end>, which are embedded into instructions as part of the input prompt. Third, textual instruction enhancement: Temporal information (day of week, time of day) and spatial information (city, region, points of interest (POIs)) are simultaneously incorporated into instructions in natural language form to provide semantic context. Similar approaches that employ convolutions to synthesize temporal and spatial dependencies in traffic flow time series include Liu et al. (2024) and (2025a). Lee and Gu (2025), on the other hand, considered using graphs to describe and learn the spatial dependencies of traffic flow time series.

On the other hand, with the advancement of information technology and wireless communication, trajectory data from traditional taxis and emerging massive shared mobility vehicles have provided increasingly abundant data resources for transportation research, creating new opportunities for analyzing link travel times and estimating changes in travel demand (Wang et al., 2019). Researchers have begun to explore the use of large models to analyze traffic trajectory data, aiming to overcome

challenges related to limited data scale and quality, as well as insufficient generalization capability. For instance, Zhou et al. (2024b) designed the PLM4Traj model, which leverages pretrained language models (PLMs) to recognize movement patterns and travel purposes from trajectories. The PLM4Traj model comprises two core components: A "trajectory semantic embedder" that converts spatiotemporal features of trajectories into embedding representations processable by PLMs and a "trajectory prompt" that integrates movement pattern and travel purpose information in natural language form, with a task prompt tuning mechanism to enhance model adaptability to different downstream tasks. The model also introduces a cross-reconstruction pretraining task to improve trajectory representation learning. Experimental results on two real-world trajectory datasets validate that PLM4Traj effectively extracts semantic information from trajectories while maintaining good interpretability in the tasks of travel time estimation and destination prediction.

Zhou et al. (2025) proposed an innovative framework called transport reasoning with intelligence progression, which constructs a dual-state-space theory that divides transportation system states into a physical state space (e.g., vehicle speed, position) and a semantic state space (e.g., traffic events, rules), achieving bidirectional alignment and fusion between the two through learnable mappings. This framework addresses the challenge of fragmentation between physical states and semantic information in ITSs. By integrating the semantic understanding capabilities of LLMs with the physical modeling strengths of world models and employing hierarchical reinforcement learning for progressive reasoning, the framework simulates the intelligent process of human experts from semantic understanding to physical prediction and then to decision execution. This enhances the decision-making capabilities of transportation systems in complex dynamic environments while ensuring interpretability, robustness, and real-time performance.

Chen et al. (2024a) and Xie et al. (2024) discussed the use of LLMs to analyze user requests expressed in natural language, combined with multimodal data, to generate trajectories for multimodal travel, demonstrating the superiority of LLMs in understanding and processing segmented trajectory data.

However, these studies have not yet carefully addressed the fundamental question of how primitives of transportation foundation models should be designed, how traffic flow time-series data and trajectory data should be fused and utilized, and how future transportation foundation models should be constructed upon these fundamental data.

3 Primitives of road transportation foundation models

3.1 Fundamental transportation data

The essence of transportation is the process of moving people or goods in space and time. Starting from first principles, it becomes evident that the essence of transportation data lies in the spatiotemporal location information of people or goods. Specifically, the relevant information currently collected and their relationships can be summarized in Fig. 1. Spatial information (or road network information when limited to ground transportation) serves as the foundation upon which transportation data are built; it is closely related to transportation data while forming an independent entity of its own. The data describing the process of moving people or goods in space and time mainly fall into four categories. The first is spatiotemporal trajectory information of people or goods, which refers to a series of data records indicating that a specific person or a specific object appears at a particular spatial location at a particular time. The second is the spatiotemporal speed information of people or goods, which refers to the time required for a specific person or a specific object to traverse a specific spatial extent within a given time window. The third is flow rate information of people or goods, which refers to the number of people or objects passing through a specific spatial extent within a given time window. The fourth is attribute information of people or goods within a specific time window and spatial extent, including identification, travel mode, travel purpose, and other related attributes.

In fact, almost all upper-level transportation applications and services can function using only these four types of processed information, without directly utilizing raw data collected from underlying transportation detection systems (Table 1) (Boukerche

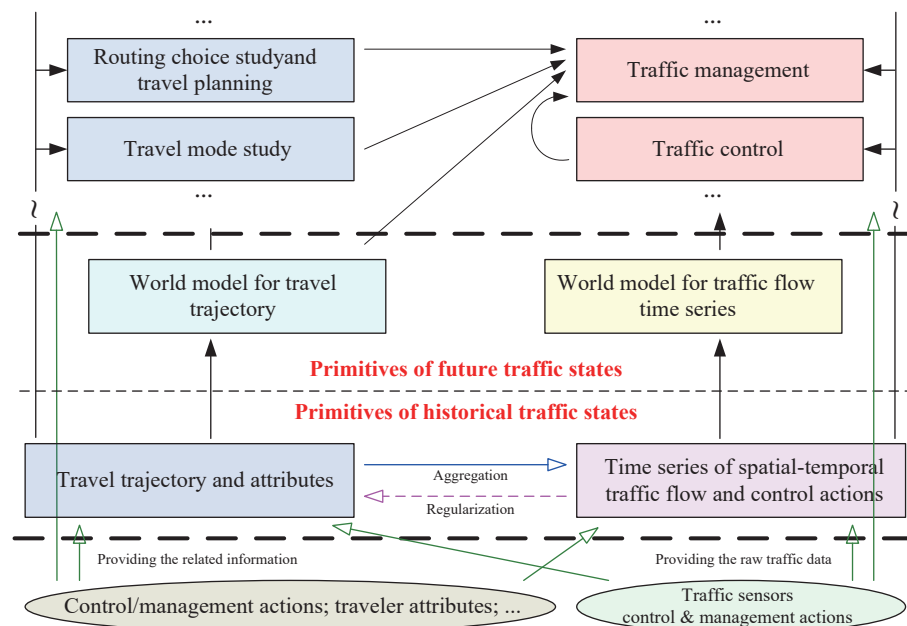


Fig. 1 Primitive design of transportation foundation models and their adaptation to upper-level transportation applications.

Table 1 Collection methods of various raw transportation data and their processed data

Raw data source	Processed data
Inductive loop detector	Vehicle speed at a specific spatial point; number of vehicles passing within a specific time window
Magnetic nail	Vehicle speed at a specific spatial point; number of vehicles passing within a specific time window; time difference of a vehicle passing upstream and downstream magnetic nails (vehicle speed through a specific spatial extent)
Camera/radar	Trajectories of people or vehicles through a specific spatial extent
Floating car	Vehicle trajectories within a specific time window
Mobile phone	Trajectories of people or vehicles within a specific time window
Questionnaire, etc.	Travel mode (vehicle type, transfers, etc.), travel purpose
Identification detection device	Identification of people or objects
Control/management actions	Signal timing, route guidance, or lane closure plans

and Coutinho, 2019; Guerrero-Ibáñez et al., 2018; Liu et al., 2018).

Synthesizing all relevant research reports published to date, the vast majority of researchers agree that the primitives of transportation foundation models should be derived from four types of fundamental information: spatiotemporal trajectory information, spatiotemporal speed information, flow rate information, and attribute information of people or goods.

3.2 Design requirements for primitives of transportation foundation models

Primitives, as the fundamental units for constructing all complex cognition and reasoning, serve as the "atoms" of a model. A well-designed primitive should satisfy the following requirements:

1) **Informational completeness and minimality:** A primitive should contain sufficient information to support reasoning without being redundant. It must strike a balance between "completely describing the state" and "computational efficiency".

2) **Composability and scalability:** Primitives should support multilevel abstraction, enabling smooth abstraction and transformation from individual traveler information in continuous spatiotemporal space (e.g., vehicle spatiotemporal trajectories) to aggregated traffic flow information (e.g., "average speed of traffic flow"). Higher-level primitives are aggregations and simplifications of lower-level primitives. Like Lego bricks, primitives should be combinable in different ways to describe and deduce the spatiotemporal transfer processes and phenomena of various traffic entities, from the microscopic level (individual trajectories) to the macroscopic level (network-wide traffic flow).

3) **Physical interpretability:** Primitives should directly correspond to observable and understandable physical quantities or behaviors in transportation systems. Spatiotemporal trajectories, speed, and flow rate precisely meet this requirement. Primitives based on these quantities make the model's decision-making process transparent, understandable, and unambiguous to traffic engineers and decision-makers.

4) **Alignment with machine and human cognition:** Ideal primitives should be efficiently processable by machine learning models (e.g., as vector/tensor representations) while also being easily mappable to human traffic concepts (e.g., "congestion"), facilitating human-machine collaboration and model evaluation.

5) **Support for prediction and generation:** Primitives should not only describe the current state of the transportation system; their definition and dynamic rules must also support the prediction of future system states under both natural evolution and external control actions and even the generation of entirely new, plausible traffic scenarios. This requires the primitive data framework to be compatible with action inputs, enabling the

world model to simulate the effects of management strategies. If future transportation foundation models are to fully leverage the paradigm established by LLMs, they must inherently support prediction and generation.

6) Distinction between data-level primitives and model-level tokens: It is essential to clarify that the primitives proposed in this study are defined at the transportation data level, not at the model token or embedding level. They represent the minimal complete physical units for describing the state of a transportation system—analogue to "atoms" in the physical world—rather than the discrete atomic units drawn from a compact vocabulary in LLMs (e.g., word pieces or BPE tokens). LLMs operate on token sequences where each token is drawn from a finite, predefined codebook (typically of size 104–105) and transportation primitive data (e.g., continuous trajectories or high-dimensional spatiotemporal flow tensors) that exhibit near-infinite variability in both spatial and temporal dimensions. Directly compressing these continuous, high-dimensional sequences into a similarly compact discrete space would inevitably incur severe information loss and is computationally impractical. Therefore, the conversion from these data-level primitives to model-level tokens or embeddings is not a trivial tokenization step but rather a challenging abstraction problem that likely requires hierarchical mechanisms (e.g., extracting semantic "phrases" or "situation descriptors" from primitive sequences) rather than direct sequence-to-token mapping. In this study, we focus on establishing the data-level foundation that defines what constitutes the minimal complete units for describing traffic states. The subsequent problem of how to translate these primitives into reasoning-friendly representations for LLMs or other foundation models is left as a critical open direction for future research, which we briefly discuss in Section 5.

3.3 Primitive design for transportation foundation models

Based on the above analysis, we argue that the primitives of transportation foundation models should comprise two categories, organized from the perspective of individual travelers and from the perspective of specific spatial extents. For individual travel data, vehicle trajectories are adopted as the first type of primitive data, with travel attribute data serving as labels for the trajectory data. For data at specific spatial extents, spatiotemporal traffic flow series of people or vehicles within a given time window are adopted as the second type of primitive data (Fig. 1).

From the perspective of informational completeness and minimality, individual travel trajectories can completely describe the motion state of a traffic participant in space and time. They constitute the minimal complete unit for describing microscopic dynamics in transportation systems. Any information with less

content (e.g., location without time) cannot uniquely determine the state. Meanwhile, speed/flow rate parameters within a specific spatial extent represent the classical and minimal complete set for describing macroscopic traffic flow states (core variables in the fundamental diagram of traffic flow theory, with density derivable from speed and flow rate). These parameters discard all individual identity information while retaining the most critical information for describing collective motion characteristics (e.g., free flow, congestion, queue dissipation). Individual trajectories serve as microscopically complete primitives, while spatiotemporal traffic flow series data serve as macroscopically complete primitives, each achieving a balance between informational completeness and minimality at their respective descriptive levels.

From the perspective of composability and scalability: Bottom-up, spatiotemporal series data for a specific spatial extent can be estimated by aggregating sufficient traffic trajectories of individual travelers passing through that spatial extent; estimation accuracy increases as the number of individual travelers passing through that spatial extent within a given time window increases. Top-down traffic data for a specific spatial extent (macroscopic speed and flow rate) cannot directly generate high-precision individual trajectories but can provide crucial spatiotemporal context and physical constraints for trajectory reconstruction, interpolation, and completion. In scenarios where individual trajectory data are sparse, traffic data for specific spatial extents provide valuable prior knowledge of collective behavior. When individual trajectory sampling is sparse, macroscopic data can serve as prior information, effectively guiding the model to probabilistically fill and smooth missing or low-frequency sampled individual trajectory points, thereby improving the rationality, smoothness, and consistency of trajectory estimation, avoiding trajectory inferences that violate macroscopic fluid dynamics, and increasing the confidence of trajectory reconstruction. Based on these two types of primitives, more complex traffic descriptions can be composed. For example, two continuously approaching individual trajectories can be composed of a "car-following" relationship; a sudden drop in flow rate accompanied by a sharp rise in density within a region can be composed of a "congestion event". Thus, these two types of primitives form a clear and actionable compositional pyramid: Individual trajectories $\rightarrow$ (spatial aggregation) $\rightarrow$ macroscopic flow rate; macroscopic flow rate $\rightarrow$ (imposing constraints) $\rightarrow$ individual trajectory generation. Together, they support seamless scaling from points to lines, from lines to areas, and from micro to macro levels.

From the perspective of physical interpretability: Both types of primitives originate from the most direct observations of the physical world. Their meanings are supported by over a century of research history and well-defined engineering definitions in the transportation field, making the decision-making process completely transparent, understandable, and unambiguous.

From the perspective of alignment with machine and human cognition, trajectory primitive data can be represented as temporal tensors, while spatial extent data can be represented as spatiotemporal grid tensors or graph structures. Both are compatible with current mainstream deep learning models such as recurrent neural networks (RNNs), convolutional neural networks (CNNs), graph neural networks (GNNs), and transformers. At the same time, the semantics of both types align well with human experts' mental models and everyday language, making them ideal media for human-machine collaboration.

From the perspective of supporting prediction and generation: There is already a substantial body of research on the prediction and generation of individual trajectories. Predicting future

trajectory points based on historical trajectory sequences is a core task in autonomous driving and traffic simulation (trajectory prediction). Generating continuous trajectories that conform to physical laws enables the construction of virtual traffic scenes. The prediction and generation of macroscopic flow rates likewise rest on a solid research foundation. Predicting future network-wide traffic flow and speed based on historical spatiotemporal flow rate data has become a core task in traffic management and travel services (traffic forecasting). By generating reasonable macroscopic flow fields, we can effectively evaluate planning schemes or conduct congestion simulations.

The preceding five perspectives establish how our two primitives satisfy the design requirements. We now clarify why only these two are elevated to the status of primitives. The answer lies in a fundamental distinction grounded in the very definition of transportation: Transportation is the process of moving people or goods in space and time. This definition describes the manifestation of movement itself as an outcome but does not describe the reasons behind that movement.

Following this first-principles reasoning, we classify all transportation-relevant information into two categories:

The effect (the movement itself): This is the direct manifestation of "people or goods moving in space and time". At the microscopic level, this effect is fully captured by individual trajectories (position + time for each traveler). At the macroscopic level, when aggregated over a specific spatial extent and time window, this effect is fully captured by spatiotemporal flow and speed series. These two forms of data are not arbitrary choices; they are the only direct, complete, and irreducible representations of the movement process across micro and macro scales. They are the "what" of the transportation system.

The causes (everything else): All other transportation-relevant information pertains to the drivers, contexts, or interpretations of the movement. These include the following: (1) Road network topology: The spatial carrier on which movement unfolds. (2) Control and management actions (e.g., signal timing, route guidance, lane closures): Interventions that perturb the system and influence future movement but are not constitutive of the movement state. (3) Weather conditions, incidents, and special events: Environmental contexts that affect the evolution of movement (e.g., rain reduces speed, incidents cause disruptions) but are not themselves movement states. (4) Traveler attributes (e.g., travel purpose, mode choice, driver characteristics): Semantic labels that explain why and how an individual travels but are not part of the spatiotemporal description of the movement itself. (5) Semantic information (e.g., POIs, land use): Contextual knowledge that enriches interpretation (e.g., why congestion occurs near a stadium after an event) but is not necessary for state representation.

The implication of this classification is profound: Primitives capture the universal, mode-agnostic outcome of transportation, while all causal, contextual, and semantic information is treated as auxiliary. This is precisely what makes primitives primitive: They are the fundamental, irreducible atoms of the transportation system's state, independent of the myriad reasons that produce that state. By separating effect from causes, our primitive scheme achieves both generality (the same primitives serve all applications) and extensibility (new causes can be added as auxiliary inputs or labels without altering the core primitive definition).

This principle also guides the architecture of our world model: The world model takes historical primitives (the effect observed in the past) and causal inputs (actions, context) to generate future

primitives (the effect to be predicted). Upper-level applications can then attach causal interpretations as labels or semantic annotations without compromising the integrity of the primitive representation. We illustrate this layered architecture in Fig. 1.

As the number of connected vehicles continues to increase, the collection of road traffic data will gradually shift toward primarily collecting individual trajectory data (Li et al., 2020). Therefore, transportation foundation models must incorporate built-in functionality to estimate spatiotemporal series data for specific spatial extents by aggregating individual travel trajectories passing through those spatial extents. Considerable research has been conducted in this area, and several recent studies have achieved sufficiently high accuracy in performing this conversion (Chen et al., 2024b; Jiang et al., 2026; Xu et al., 2025; Zhang et al., 2020).

4 Primitives of transportation foundation models and the world model

4.1 World model for transportation

We have synthesized and summarized the major research topics related to road transportation discussed in studies published in top-tier transportation journals over the past two decades (Cobo et al., 2012; Jiang et al., 2020; Modak et al., 2019; Ran et al., 2012; Sun and Yin, 2017; Tang et al., 2014) and explored whether these topics can be directly analyzed and processed using the primitives of transportation foundation models. Our analysis indicates that many upper-level transportation applications rely on only a single type of transportation primitive data. More notably, however, a significant number of upper-level transportation applications require both estimation of the current state of the transportation system and prediction of its future state. Therefore, we propose in this study that transportation foundation models must possess the capability to establish a world model.

In the field of AI, a world model refers to a computational model capable of internally modeling an external environment or a specific system, understanding its operational laws, and predicting its future evolution (Ha and Schmidhuber, 2018). In brief, a world model is an internal simulator in the mind of an agent (whether AI or human) about "How the world works." A complete world model should possess three core capabilities: (1) Representation: The ability to encode and store knowledge about the state of the world in a compact and structured form. (2) Prediction: The ability, given the current state and possible actions, to infer (simulate) future states. (3) Generation: The ability to generate logical, unseen yet plausible new states or sequences based on learned laws.

Transportation systems are typical complex adaptive systems characterized by highly dynamic interdependencies and uncertainties across both temporal and spatial dimensions. Traditional "sense-and-respond" management approaches are no longer sufficient to address these challenges. To establish a transportation foundation model for modeling, analyzing, and controlling transportation systems, we need an intelligent core capable of "reasoning, simulating, and deciding". This is precisely the value of a world model for transportation. Specifically, a world model for transportation offers the following benefits:

First, it enables us to proactively address the "butterfly effect" and achieve forward-looking management and control. The propagation of traffic congestion resembles ripples on water, where a minor disturbance at one intersection can potentially paralyze the entire network. Traditional models are limited to localized and short-term predictions. In contrast, an integrated

framework serving as a world model, through precise modeling of spatiotemporal correlations, can simulate state propagation chains over long distances and extended time horizons. This allows managers to foresee potential "storms" and implement interventions before congestion occurs (e.g., pretuning signals, issuing guidance information), shifting from "passive response" to proactive defense and resilience enhancement.

Second, it provides far-sighted support for "what-if" analysis of traffic management measures, serving transportation decision-making and planning. The formulation of urban transportation policies (e.g., establishing bus lanes, adjusting toll zones) or the management of large-scale events cannot be tested through repeated real-world experimentation (Guo et al., 2019; Li et al., 2014; Wang et al., 2024, 2025). We need a reliable "digital sandbox". A world model serves as the dynamic engine of this sandbox. It can accept various "what-if-then" queries: If the signal timing scheme at this intersection is changed, how will network-wide traffic flow be redistributed? If a new subway station is added, how much will the load on surrounding roads be reduced? By quantitatively simulating the consequences of different strategies, it provides computable and evaluable bases for scientific decision-making.

Third, it enables us to obtain "future" primitive transportation data based on "historical" primitive transportation data, jointly empowering upper-level transportation applications. Advanced transportation AI applications (e.g., adaptive signal control, connected and automated vehicle coordination, and dynamic route guidance) all depend on a deep understanding of both the current and future states of the transportation system. An integrated world model for transportation framework functions as a continuously updated digital twin of the transportation system, outputting complete, clean, and predicted future state sequences as transportation system primitive data. This provides a shared and stable cognitive environment for all upper-level intelligent applications, facilitating their planning and coordination, and collectively building a layered, coherent, and efficient ecosystem for transportation intelligence applications.

In fact, over the past decade, transportation researchers have designed and implemented integrated processing frameworks based on spatiotemporal traffic flow series primitive data. These frameworks integrate preprocessing (including missing data imputation, anomaly detection and alarming, and data compression) with the estimation/prediction of spatiotemporal traffic flow series data (Cui et al., 2020; Li et al., 2013, 2015). In recent years, researchers have also begun to investigate integrated processing frameworks specifically for trajectory data (Jiang et al., 2026; Teoh et al., 2025).

These integrated processing frameworks are, in essence, world models. In terms of "representation", these frameworks take spatiotemporal trajectories and traffic flow time series as primitives, which constitute the minimal complete set for describing traffic dynamic states and form the atoms of the internal representation of the transportation world. In terms of "prediction", these frameworks place data cleaning, imputation, and prediction within a unified optimization paradigm. This means they not only predict the future but also continuously revise and update the "best estimate" of current and past states—a process that inherently maintains an internally consistent representation of the world state. In terms of "generation", through embedded spatiotemporal dynamic models, these frameworks can start from any initial state and simulate continuous state sequences that conform to physical laws, thereby satisfying the capability to generate plausible scenarios.

4.2 Difference between the world model for transportation and traffic simulation

Traditional microscopic traffic simulation is a classic tool in the field of transportation engineering, with its core philosophy rooted in reductionism. It treats the transportation system as a complex system composed of a large number of individual travelers (vehicles, pedestrians) possessing independent decision-making capabilities. Through a "bottom-up" approach, it meticulously simulates the behavior of each individual and their interactions, thereby emerging macroscopic traffic flow phenomena.

The implementation of traditional microscopic traffic simulation typically follows the following standardized process:

1) Input Preparation: Construction of a high-precision digital road network containing details such as lanes, connectors, traffic signals, and gradients; definition of traffic demand, typically provided in the form of an origin–destination (OD) matrix with departure times.

2) Behavioral Modeling: Car-following models describe the longitudinal interaction between leading and following vehicles (e.g., Gipps model, intelligent driver model (IDM)), calculating the acceleration of the following vehicle based on variables such as distance headway and speed difference. Lane-changing models describe the lateral decisions made by vehicles to seek improved driving conditions (e.g., higher speed, correct lane), typically including submodels such as target lane selection and gap acceptance. Route choice models describe how drivers make selections among alternative paths based on real-time or historical traffic conditions.

3) Simulation Execution: The simulation core is a time-stepping loop (typically with a step size of 0.1–1 s). At each time step, the simulator sequentially updates the state of all individuals: Calculating the acceleration, speed, and new position of each individual based on the current surrounding environment (preceding vehicle position, traffic signal status) and its own behavioral models; detecting and resolving conflicts between individuals (e.g., intersection conflicts).

4) Output and Analysis: After the simulation concludes, individual trajectory data are aggregated to generate macroscopic indicators such as link flow, speed, density, travel time, and queue length for evaluating network performance.

The advantages of traditional microscopic traffic simulation lie in its clear mechanisms and strong interpretability, enabling the reproduction of complex traffic phenomena based on simplified traffic mechanism models. However, its modeling heavily relies on a large number of preset parameters (driving behavior parameters, OD matrices), and the simulation is computationally intensive. A single large-scale network simulation typically takes minutes to hours, essentially constituting a computationally expensive forward inference process that struggles to achieve second-level or minute-level real-time state updates and rapid predictions.

Nevertheless, this paradigm of "real-time simulation" or "online simulation" faces inherent bottlenecks. First, the modeling chain is long and fragile: From raw data cleaning, OD back-calculation, and parameter calibration to simulation execution, there are numerous stages, and errors accumulate and propagate through each stage. Second, computational latency is severe: Even with high-performance computing, the complete cycle of parameter updating and simulation inference cannot meet the demand for second-level response and prediction of instantaneous dynamics (e.g., rapid congestion propagation caused by incidents). Third, real-time data adaptation is difficult: Simulation models are strongly mechanistic, and their ability to simulate unexpected events depends on whether corresponding rules have been

preprogrammed. For anomalous patterns or emerging collective behaviors reflected in the data that are not defined in the model (e.g., local disturbances caused by concentrated pick-up and drop-off of ride-hailing vehicles), the model struggles to quickly capture and adapt to them. Therefore, traditional approaches often face practical challenges in terms of computational latency and calibration complexity when addressing the core requirements of real-time prediction and control.

The world model in our proposed transportation foundation model is built upon a completely different philosophy and technology. At its core, we abandon the indirect modeling path of "individual behavioral rules → collective emergence" and instead directly model the "state" of the transportation system in space and time. As described in the previous section, we distill the state of the transportation system into two types of primitive data: Spatiotemporal trajectories of individual travelers (representable as temporal tensors) and flow rate and speed within specific spatial extents (representable as spatiotemporal grid tensors or graph structures). These primitive data constitute direct, compact, and complete observational summaries of the system state.

Based on this, we can abstract the evolution of traffic states as a pure "sequence-to-sequence" mapping problem. Leveraging the power of deep learning (e.g., transformers, spatiotemporal graph neural networks), the model can directly learn, end-to-end from massive historical state sequence data, the complex mapping function from states over the past N time steps to states over the future M time steps. This process requires no explicit encoding of behavioral rules such as car-following or lane-changing, nor does it require back-calculation of OD matrices. Through learning, the model inherently and implicitly masters the full set of traffic dynamics and, upon receiving new real-time state observations, can infer future system states with extreme speed (typically millisecond to second-level) via forward inference. This state-centric, data-driven, deep learning-empowered modeling paradigm differs substantially from traditional microscopic behavioral simulation (Table 2).

In summary, compared to traditional microscopic simulation—which relies on numerous parameters and rules—this novel transportation world model based on deep learning sequence-to-sequence mapping offers distinct advantages in real-time performance, adaptability, and scalability, making it well suited for real-time prediction and control applications. It can operate in near real-time with the physical world, enabling rapid perception and simulation, and thus provides a powerful core engine for urban traffic management and dynamic optimization. Moreover, the world model can undergo rapid online iteration based on newly collected traffic data, allowing it to adapt more responsively to ever-changing transportation systems. This represents a significant shift in transportation system modeling—from a predominantly "mechanistic simulation" paradigm toward a more "data-driven intelligence" paradigm. We note, however, that traditional simulation retains its strengths in scenarios requiring explicit interpretability of behavioral mechanisms, offline scenario analysis, and planning-stage "what-if" evaluations where real-time constraints are less stringent. The two approaches are therefore complementary rather than mutually exclusive, and their integration may offer the best of both worlds in future transportation modeling frameworks.

4.3 World model for transportation as an essential part of transportation foundation models

Although some researchers consider that such integrated processing frameworks or world models do not belong to

Table 2 Comparison between traditional microscopic traffic simulation and the proposed world model

Comparison dimension	Traditional microscopic traffic simulation	Primitive-based and deep learning-powered transportation world model
Modeling philosophy	Reductionism: individual behavioral rules → collective emergence	Holism/state-space approach: directly learn spatiotemporal evolution laws of system states (trajectories, flow fields)
Core input	Network topology, OD matrix, detailed driving behavior parameters	Historical and current observations of system state (trajectory points, flow rate, speed tensors) and control/management actions (e.g., signal plans, guidance strategies, lane closures)
Computational essence	Rule-based iterative computation, with each individual computed at each time step	Neural network-based forward propagation, processing the entire spatiotemporal state as input through one-time parallel computation
Speed and real-time performance	Slow (minutes to hours), difficult to use for real-time closed-loop control	Extremely fast (milliseconds to seconds), well-suited for high-frequency real-time prediction and rapid decision-making
Data-drivenness and adaptability	Rigid, dependent on preset rules, poor adaptability to novel patterns; requires frequent and complex parameter calibration	Highly flexible, capable of automatically learning any pattern (including anomalies and emerging behaviors) from data, strong adaptability
Application mode	Offline analysis, scenario evaluation; primarily used for "what-if" analysis in planning stages	Online perception, real-time prediction, rapid simulation; can be directly embedded into management and control systems for second-level "perception-prediction-control" closed loops
Scalability and unification	Complex models, extending to new regions or integrating new data sources (e.g., video semantics) requires extensive remodeling	Unified architecture (e.g., transformer), easy to scale to larger networks and integrate multimodal data (e.g., incorporating video features as another input tensor)

foundation models or large models, we argue that future transportation foundation models will inevitably incorporate the world model for transportation as an essential component. The reasons are as follows:

First, some researchers believe that transportation foundation models must be "large" in terms of data scale and model scale. LLMs typically have data volumes reaching the terabyte (TB) or even petabyte (PB) level. They are usually based on the transformer architecture, which is unified in structure but massive in capacity, with parameter counts typically reaching hundreds of millions (10^8) or even hundreds of billions to trillions (10^{11} – 10^{12}). This is the intuitive manifestation of "large".

However, transportation primitive data do not have such large data volumes. Take the well-known California PeMS dataset in the United States as an example. Assume there are 5000 detectors in the Los Angeles area, recording data every 5 min (flow, occupancy, speed, etc., assuming a total of 10 float fields). The data volume per day is approximately $5000 \times (24 \times 60/5) \times 10 \times 4$ bytes $\approx 5000 \times 288 \times 40 \approx 57.6$ MB/day. The data volume over 20 years is approximately $57.6 \text{ MB} \times 365 \times 20 \approx 420$ GB. Such a sample data volume is indeed "small" compared to the trillion-token text corpora.

Meanwhile, spatiotemporal traffic flow series data exhibit clear daily periodicity, short-term trends, and certain randomness and are not as complex as natural language. Based on daily periodicity and short-term trends, sufficiently accurate traffic prediction can already be achieved. For example, on the PeMS dataset, a carefully designed lightweight time series model (with core parameter count typically below one hundred thousand, i.e., 10^5) can already capture these patterns well and achieve practical accuracy exceeding 90%. The existence of randomness also determines that no matter how much the model is scaled up, prediction accuracy cannot be significantly further improved; in contrast, overfitting and performance saturation are highly likely to occur.

Therefore, we argue that the "large" in transportation foundation models should not refer to parameter expansion for a single prediction task but rather to the expansion of model capacity and knowledge reserves for multitask general capabilities. In traffic prediction, there exists a trade-off among accuracy, interpretability, and computational cost. Increasing the model

scale beyond a certain point often leads to diminishing marginal gains in accuracy while potentially increasing computational overhead and reducing interpretability. Therefore, the "large" in transportation foundation models should not refer solely to parameter expansion for a single prediction task but rather to the expansion of model capacity and knowledge reserves for multitask general capabilities.

Second, some researchers believe that transportation foundation models must adopt a pretraining + prompting/fine-tuning paradigm. That is, they first "self-learn" from massive unlabeled traffic data to acquire general spatiotemporal understanding capabilities and then lightly adapt to various downstream applications. The same pretrained model, through different "prompts" or a small amount of fine-tuning, can directly handle nearly all traffic tasks, including data imputation, anomaly detection, short-term prediction, long-term simulation, scenario generation, and strategy evaluation.

However, for traffic state prediction applications where the physical laws are clear and the task definition is well specified, supervised learning is the most efficient and direct path. It directly leverages human domain knowledge (labeled data or defined loss functions) to guide the model. Blindly pursuing unsupervised pretraining may be a detour. The world model for transportation has already successfully demonstrated that, based on feature engineering and specialized model design, designing optimization algorithms and model structures for the specific task chain of "data preprocessing-prediction" can establish efficient traffic state prediction applications.

Future transportation foundation models are likely to adopt a "large model + small model" hybrid architecture, similar to the latest LLMs. The world model for transportation serves as a small model, outputting future traffic states as model primitives. In addition, multiple "small models" will be established as domain experts. For example, a specialized small model could perform signal timing optimization. The "large model" in future transportation foundation models is likely to be a dispatching hub that, through technologies such as a mixture of expert routing layers, intelligently invokes one or a combination of multiple expert "small models" based on the type of input problem (e.g., "predict future flow" vs. "explain the cause of an incident" vs.

"generate simulation scenarios"). During training, only the parameters of the invoked expert "small models" are updated, greatly improving efficiency.

Third, specifically regarding the world model for transportation, it internally encapsulates the generation mechanism of transportation primitive data and the state transition laws of the transportation system. In the broader architecture of transportation foundation models, this integrated framework can play the role of a highly reliable prediction expert that can be invoked. When upper-level applications (e.g., traffic management and control strategy evaluation) require future traffic information, they can invoke it. For example, upper-level applications may encounter the following needs: (1) Understanding traffic: "Analyze the propagation chain and root causes of this congestion event." (2) Travel planning: "Plan the fastest route for this ambulance and coordinate the traffic signals along the way." (3) Traffic management decision-making: "Recommend which traffic management scheme to adopt during tomorrow's large-scale event." (4) Traffic dynamics generation: "Simulate how the surrounding network traffic flow would change if a new subway station were built here." In these tasks, the foundation model needs to frequently query the integrated framework (the prediction expert) to obtain "what-if-then" simulation results to support its decision-making.

Furthermore, the ability to process action time series is indispensable for bridging the gap between prediction and decision-making. A foundation model that only forecasts future states under historical patterns is insufficient for active traffic control. By integrating action sequences as part of the input, the world model provides a unified interface for upper-level optimization modules (e.g., reinforcement learning agents or classical controllers) to query the outcomes of candidate strategies, thereby closing the loop from perception to action.

Therefore, explicitly defining the integrated processing framework as the world model for the transportation system represents a significant paradigm shift. It marks the transition of our goal from building a single accurate prediction tool to constructing a "computable, queryable, and simulatable digital core of the transportation system". We argue that the world model is a critical and mature infrastructure in the transportation foundation model ecosystem—a reliable data processing and state simulation engine. It will not be replaced but rather integrated and enhanced, playing an even more powerful role under a more intelligent dispatching system.

4.4 Positioning relative to other AI-driven transportation frameworks

Having introduced the proposed world model and contrasted it with traditional microscopic traffic simulation in Section 4.2, we now position our framework relative to three other major AI-driven paradigms in the transportation domain: Traditional traffic prediction models, digital twins, and LLM-based transportation applications. These paradigms share certain overlaps with our work but differ fundamentally in objectives, methodologies, and roles within the broader transportation intelligence ecosystem.

Compared with traditional traffic prediction models, a vast body of literature has focused on forecasting future traffic states from historical observations using models ranging from classical time-series methods to deep learning architectures. These models are typically designed for a single task (e.g., predicting flow, speed, or travel time) and operate on preprocessed, clean input data. In contrast, our proposed world model serves a broader purpose. It

integrates data preprocessing (including missing data imputation, anomaly detection, and compression), state estimation, and prediction under a unified optimization framework. More importantly, it is explicitly designed to accept external control actions (e.g., signal timing adjustments, route guidance commands, or lane closure plans) as conditional inputs, enabling it to simulate the traffic system's response to potential interventions. This transforms the world model from a passive forecasting tool into an active decision-support engine capable of answering "what-if" queries for traffic management and control. Prediction in our framework is not an end in itself but rather a means to comprehensive, actionable simulation for planning and real-time operations.

Compared with digital twins, digital twins have gained considerable attention in transportation as virtual replicas of physical infrastructure systems, enabling real-time monitoring, visualization, and asset management. A digital twin typically mirrors "what is" in the physical world, emphasizing high-fidelity representation and data synchronization. Our world model complements and extends this concept by focusing on predictive simulation and intervention analysis—inferring "what will be" and "what if" under different management strategies. In this sense, the world model can be viewed as the intelligent simulation core of a transportation digital twin: It is the component that turns a passive mirror into an active reasoning engine. A digital twin asks, "what is happening now?"; our world model is designed to answer "what would happen if we took this action?" and "what will the state be in 15 min?" This predictive and prescriptive capability distinguishes our framework from most existing digital twin implementations, which tend to prioritize real-time data fusion and visualization over forward simulation.

Compared with LLM-based transportation applications, a growing number of studies have explored the direct application of LLMs to transportation problems (Li et al., 2024; Wandelt et al., 2024; Xu and Liu, 2025; Zhang et al., 2023). These approaches typically convert traffic data into textual representations or token sequences and then leverage LLMs' semantic understanding and reasoning capabilities to generate predictions, recommendations, or explanations. While these efforts demonstrate the power of LLMs in handling unstructured information and human-computer interaction, they also face inherent limitations: Traffic time-series data are long, continuous, and high-dimensional, making direct tokenization into a compact codebook impractical without severe information loss; moreover, LLMs are not inherently designed for physics-grounded simulation and may generate plausible but physically inconsistent outputs. Our work takes a fundamentally different and complementary approach. Rather than relying on LLMs for state prediction or simulation, we focus on defining the data-level primitives and constructing a reliable world model that generates clean, complete, and predicted primitive sequences. This separation concerns the world model as a credible simulator and the LLM as a reasoner and communicator. Such a method leverages the respective strengths of both paradigms: the physical rigor and computational efficiency of the world model for simulation and the semantic richness and interactivity of LLMs for high-level reasoning.

In summary, our proposed framework occupies a distinct and foundational niche within the landscape of transportation AI. It is not a replacement for prediction models, digital twins, or LLM-based applications but rather a necessary data-level and system-level foundation upon which these higher-level capabilities can be built. By formalizing the primitives of traffic states and establishing a world model that generates future primitives under

interventions, we provide a shared, reliable, and actionable simulation backbone for the diverse ecosystem of transportation intelligence applications.

4.5 Illustrative examples: How primitives and the world model support real-world tasks

To make the preceding conceptual discussion more concrete, we give illustrative examples showing how the proposed primitives and world model can be applied to real-world transportation tasks. Each example follows the same workflow: The world model takes historical primitives (the observed "effect") and causal inputs (actions and context) to generate future primitives, which are then consumed by upper-level applications for decision-making. These examples are deliberately simplified to highlight the core logic rather than the technical implementation details.

Example 1: incident-responsive route guidance.

Suppose a major accident occurs on a highway segment, causing severe congestion. A traffic information service center needs to provide timely route guidance to affected drivers. The world model takes as input:

Historical primitives: Individual vehicle trajectories (microscopic primitive) and macroscopic flow/speed series from the affected corridor and surrounding network, covering the period before and immediately after the incident.

Causal inputs: The incident itself is encoded as an environmental context (e.g., location, severity, estimated clearance time), and potential route guidance strategies (e.g., "divert 30% of traffic to parallel routes") are encoded as action sequences.

The world model simulates future traffic states under different guidance strategies, generating predicted trajectories and flow patterns for the next hour. The upper-level routing module then evaluates these predictions by comparing estimated travel times, network spillover effects, and secondary congestion risks to select and disseminate the optimal guidance strategy. This demonstrates how the framework integrates both microscopic primitives (for individual routing) and macroscopic primitives (for network-wide impact assessment) within a unified simulation engine.

This uniform interface decouples the simulation engine from the application logic, allowing different upper-level modules (e.g., optimization algorithms, rule-based systems, or LLM-based reasoners) to leverage the same reliable, actionable predictions without duplicating the underlying modeling effort. We view this as a key advantage of our framework: It provides a shared cognitive layer for the entire transportation intelligence ecosystem.

5 Conclusions

Starting from the first principle of the essence of transportation systems—"movement in space and time"—this study systematically demonstrates the dual foundations necessary for constructing transportation foundation models: Primitives and the world model. We propose that individual travel trajectories and spatiotemporal traffic flow series at specific spatial extents are two types of minimal complete primitives that satisfy the requirements of completeness, composability, and human-machine alignment, jointly constituting the "atomic" description of transportation systems at both micro and macro levels. The aggregation and constraint relationships between these primitives provide a hierarchical cognitive skeleton for the model. Furthermore, the world model can help generate future primitive data from historical primitive data.

Looking forward, the evolution of transportation foundation models will proceed along two major directions:

First, "primitive expansion" and "information fusion". Although current primitives capture motion states, they have not yet fully integrated environmental semantics (e.g., weather, events, social media sentiment), individual attributes (e.g., travel purpose, driving style), and high-dimensional perceptions (e.g., scene understanding from video). Future world models need to map or associate massive multimodal, unstructured data into a unified primitive space, thereby uncovering deeper causal relationships and behavioral motivations (Hao et al., 2024; Liu et al., 2025b). For example, by fusing mobile signaling data with trajectory primitives, we can not only know "where traffic congestion occurs" but also understand "why it occurs" (e.g., dispersal after a large-scale event), thereby providing semantically enhanced decision support for management and control strategies.

Second, the expansion of "service objects". While this study primarily focuses on system-level traffic management and control optimization, another important mission of transportation foundation models is to empower individual travel services. A transportation foundation model equipped with an advanced world model can provide highly personalized prediction and planning for each traveler. For instance, by combining personal historical trajectories, real-time traffic simulation, and individual preferences, the model can dynamically generate optimal travel plans and provide forward-looking risk alerts during the journey (e.g., abnormal congestion ahead, suggested transfer), truly realizing the transition from "managing traffic flow" to "serving every individual".

Third, and perhaps most fundamentally, future research must address the abstraction gap between data-level primitives and model-level tokens. A mature transportation foundation model will likely require a two-stage architecture: First, the world model transforms raw sensor data into clean, complete, and predicted primitive sequences (the scope of this study); second, an upper-level reasoning engine further abstracts these high-dimensional primitive sequences into higher-level semantic units—such as "traffic situation phrases" or "operation descriptors"—that can be consumed by LLMs or other reasoning modules. To illustrate this, consider a user query: "I need to travel from point A to point B." Rather than asking the world model to directly compute an optimal route, the foundation model would (1) query the world model to simulate and compare the travel times along several candidate routes under current and predicted traffic conditions; (2) summarize the simulation results into concise semantic phrases (e.g., "Route 1: 25 min, congestion clears in 10 min"; "Route 2: 30 min, no delays"); and (3) feed these phrases into a reasoning module to generate an interpretable recommendation with supporting evidence. In this architecture, the world model serves as a credible simulator and evidence provider, while the upper-level LLM serves as a reasoner and communicator. The bridge between them is not direct tokenization but a deliberate abstraction layer that translates numerical simulation results into semantically meaningful "facts" that can be processed by language models. This represents a rich and largely uncharted research direction that goes beyond the scope of the current study, but we believe it is a necessary step toward truly ITSs.

Future transportation foundation models will move from scattered task-specific models to a new era of infrastructure centered on unified state perception and simulation. Transportation primitive data and the world model are not only the essential path to building powerful transportation foundation models but will also become the core engine of future transportation digital twin cities—a bridge connecting data perception and intelligent decision-making and a key infrastructure for realizing the leap of transportation systems from informatization to intellectualization. Building upon this

foundation, integrating multisource information and bridging macrolevel management and control with microlevel services will be the key pathway to achieving full-element, full-lifecycle intelligence in transportation systems. Upper-level general large models, through mechanisms such as mixture of experts (MoE), flexibly invoke multiple domain experts, including the world model, to collectively accomplish complex tasks ranging from understanding and planning to decision-making and generation.

Author contributions

Haiyang Yu: Writing – original draft, Conceptualization. **Binghong Jiang:** Writing – review & editing. **Xiaobo Qu:** Writing – review & editing, Conceptualization. **Li Li:** Writing – review & editing, Conceptualization, Supervision, Methodology, Funding acquisition.

Acknowledgements

This study is supported by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (No. JYB2025XDXM123).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article. The author Xiaobo Qu is the Editor-in-Chief of this journal.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used DeepSeek to help fix the grammar mistakes. After utilizing this tool, the authors thoroughly reviewed and edited the content as necessary and take full responsibility for the final content of the published article.

References

- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., et al., 2021. On the Opportunities and Risks of Foundation Models. <https://doi.org/10.48550/arXiv.2108.07258>
- Boukerche, A., Coutinho, R. W. L., 2019. Crowd management: The overlooked component of smart transportation systems. *IEEE Commun Mag*, **57**, 48–53.
- Chen, Y., Pesaranghader, A., Sadhu, T., Yi, D. H., 2024a. Can We Rely on LLM Agents to Draft Long-Horizon Plans? Let's Take TravelPlanner as An Example. <https://doi.org/10.48550/arXiv.2408.06318>
- Chen, X., Qin, G., Seo, T., Yin, J., Tian, Y., Sun, J., 2024b. A macro-micro approach to reconstructing vehicle trajectories on multi-lane free-ways with lane changing. *Transp Res Part C Emerg Technol*, **160**, 104534.
- Cobo, M. J., Lopez-Herrera, A. G., Herrera, F., Herrera-Viedma, E., 2012. A note on the ITS topic evolution in the period 2000–2009 at T-ITS. *IEEE Trans Intel Transport Syst*, **13**, 413–420.
- Cui, Z., Ke, R., Pu, Z., Wang, Y., 2020. Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transp Res Part C Emerg Technol*, **118**, 102674.
- Gruver, N., Finzi, M., Qiu, S., Wilson, A. G., 2023. Large language models are zero-shot time series forecasters. In: Proceedings of the 37th International Conference on Neural Information Processing Systems, 19622–19635.
- Guerrero-Ibáñez, J., Zeadally, S., Contreras-Castillo, J., 2018. Sensor technologies for intelligent transportation systems. *Sensors*, **18**, 1212.
- Guo, Q., Li, L., Jeff Ban, X., 2019. Urban traffic signal control with connected and automated vehicles: A survey. *Transp Res Part C Emerg Technol*, **101**, 313–334.
- Ha, D., Schmidhuber, J., 2018. World Models. <https://doi.org/10.48550/arXiv.1803.10122>

- Hao, H., Wang, Y., Chen, J., 2024. Empowering scenario planning with artificial intelligence: A perspective on building smart and resilient cities. *Engineering*, **43**, 272–283.
- Jiang, B., Guo, Z., Chang, C., Zhang, J., Li, L., 2026. Scene-to-scene microscopic traffic dynamics deduction: Integrating simulation, prediction, and real-time monitoring. *IEEE Intel Transp Syst Mag*, **18**, 41–52.
- Jiang, C., Bhat, C. R., Lam, W. H. K., 2020. A bibliometric overview of Transportation Research Part B: Methodological in the past forty years (1979–2019). *Transp Res Part B Methodol*, **138**, 268–291.
- Lee, X., Gu, Y., 2025. STEP-LLM: Spatial-temporal-enriched prompting with large language models for traffic prediction. In: The Proceedings of 2024 International Conference of Electrical, Electronic and Networked Energy Systems. Singapore: Springer Nature Singapore, 248–255.
- Li, L., Jiang, R., He, Z., Chen, X. M., Zhou, X., 2020. Trajectory data-based traffic flow studies: A revisit. *Transp Res Part C Emerg Technol*, **114**, 225–240.
- Li, L., Li, Y., Li, Z., 2013. Efficient missing data imputing for traffic flow by considering temporal and spatial dependence. *Transp Res Part C Emerg Technol*, **34**, 108–120.
- Li, L., Su, X., Zhang, Y., Lin, Y., Li, Z., 2015. Trend modeling for traffic time series analysis: An integrated study. *IEEE Trans Intel Transp Syst*, **16**, 3430–3439.
- Li, L., Wen, D., Yao, D., 2014. A survey of traffic control with vehicular communications. *IEEE Trans Intel Transp Syst*, **15**, 425–432.
- Li, Z., Xia, L., Tang, J., Xu, Y., Shi, L., Xia, L., et al., 2024. UrbanGPT: Spatio-Temporal Large Language Models. <https://doi.org/10.48550/arXiv.2403.00813>
- Liu, C., Hettige, K. H., Xu, Q., Long, C., Xiang, S., Cong, G., et al., 2025a. ST-LLM+: Graph enhanced spatio-temporal large language models for traffic prediction. *IEEE Trans Knowl Data Eng*, **37**, 4846–4859.
- Liu, C., Yang, S., Xu, Q., Li, Z., Long, C., Li, Z., et al., 2024. Spatial-Temporal Large Language Model for Traffic Prediction. <https://doi.org/10.48550/arXiv.2401.10134>
- Liu, Z., Zhou, Z., Gu, Z., Liu, S., Liu, P., Zhang, Y., et al., 2025b. TRIP: Transport reasoning with intelligence progression—a foundation framework. *Transp Res Part C Emerg Technol*, **179**, 105260.
- Liu, B., Zhou, W., Zhu, T., Gao, L., Xiang, Y., 2018. Location privacy and its applications: A systematic study. *IEEE Access*, **6**, 17606–17624.
- Modak, N. M., Merigó, J. M., Weber, R., Manzor, F., de Dios Ortúzar, J., 2019. Fifty years of Transportation Research journals: A bibliometric overview. *Transp Res Part A Policy Pract*, **120**, 188–223.
- Nie, T., Sun, J., Ma, W., 2025. Exploring the roles of large language models in reshaping transportation systems: A survey, framework, and roadmap. *Artif Intel Transp*, **1**, 100003.
- Ran, B., Jin, P. J., Boyce, D., Qiu, T. Z., Cheng, Y., 2012. Perspectives on future transportation research: Impact of intelligent transportation system technologies on next-generation transportation modeling. *J Intel Transp Syst*, **16**, 226–242.
- Sun, L., Yin, Y., 2017. Discovering themes and trends in transportation research using topic modeling. *Transp Res Part C Emerg Technol*, **77**, 49–66.
- Tang, S., Li, Z., Chen, D., Chen, Z., Liu, W., Liu, X., et al., 2014. Theme classification and analysis of core articles published in IEEE transactions on intelligent transportation systems from 2010 to 2013. *IEEE Trans Intel Transport Syst*, **15**, 2710–2719.
- Teoh, J., Tomar, M., Ahn, K., Hu, E. S., Pearce, T., Sharma, P., et al., 2025. Next-Latent Prediction Transformers Learn Compact World Models. <https://doi.org/10.48550/arXiv.2511.05963>
- Wandelt, S., Zheng, C., Wang, S., Liu, Y., Sun, X., 2024. Large language models for intelligent transportation: A review of the state of the art and challenges. *Appl Sci*, **14**, 7455.
- Wang, X., Jerome, Z., Wang, Z., Zhang, C., Shen, S., Kumar, V. V., et al., 2024. Traffic light optimization with low penetration rate vehicle trajectory data. *Nat Commun*, **15**, 1306.
- Wang, S., Li, L., Ma, W., Chen, X., 2019. Trajectory analysis for on-demand services: A survey focusing on spatial-temporal demand and supply patterns. *Transp Res Part C Emerg Technol*, **108**, 74–99.
- Wang, P., Lv, Y., Ge, J., Li, L., 2025. Boosting the training of deep rein-

forcement learning traffic control by using the world model. *IEEE Intel Transp Syst Mag*, 17, 58–66.

- Xie, J., Zhang, K., Chen, J., Zhu, T., Lou, R., Tian, Y., et al., 2024. TravelPlanner: A Benchmark for Real-World Planning with Language Agents. <https://doi.org/10.48550/arXiv.2402.01622>
- Xu, M., Fang, J., Bansal, P., Kim, E. J., Qiu, T. Z., 2025. Physics informed deep generative model for vehicle trajectory reconstruction at arterial intersections in connected vehicle environment. *Transp Res Part C Emerg Technol*, 171, 104985.
- Xu, Y., Liu, M., 2025. GPT4TFP: Spatio-temporal fusion large language model for traffic flow prediction. *Neurocomputing*, 625, 129562.
- Zhang, S., Fu, D., Zhang, Z., Yu, B., Cai, P., 2023. TrafficGPT: Viewing, Processing and Interacting with Traffic Foundation Models. <https://doi.org/10.48550/arXiv.2309.06719>
- Zhang, H., Liu, H. X., Chen, P., Yu, G., Wang, Y., 2020. Cycle-based end of queue estimation at signalized intersections using low-penetration-rate vehicle trajectories. *IEEE Trans Intel Transp Syst*, 21, 3257–3272.
- Zhang, Z., Sun, Y., Wang, Z., Nie, Y., Ma, X., Li, R., et al., 2024a. Large Language Models for Mobility Analysis in Transportation Systems: A Survey on Forecasting Tasks. <https://doi.org/10.48550/arXiv.2405.02357>
- Zhang, K., Zhou, F., Wu, L., Xie, N., He, Z., 2024b. Semantic understanding and prompt engineering for large-scale traffic data imputation. *Inf Fusion*, 102, 102038.
- Zhao, C., Dai, X., Lv, Y., Tian, Y., Ren, Y., Wang, F. Y., 2022. Foundation models for transportation intelligence: ITS convergence in TransVerse. *IEEE Intel Syst*, 37, 77–82.
- Zhou, Z., Gu, Z., Qu, X. B., Liu, P., Liu, Z. Y. 2024a. Urban multimodal transportation generative pretrained transformer foundation model: Hierarchical techniques and application scenarios of spot-corridor-network decomposition. *China J Highw Transp*, 37, 253–274. (in Chinese)
- Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., et al., 2025. A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT. *Int J Mach Learn Cybern*, 16, 9851–9915.
- Zhou, Z., Lin, Y., Wen, H., Xu, Q., Guo, S., Hu, J., et al., 2024b. TrajCogn: Leveraging LLMS for Cognizing Movement Patterns and Travel Purposes from Trajectories. <https://doi.org/10.48550/arXiv.2405.12459>
- Zou, X., Yan, Y., Hao, X., Hu, Y., Wen, H., Liu, E., et al., 2025. Deep learning for cross-domain data fusion in urban computing: Taxonomy, advances, and outlook. *Inf Fusion*, 113, 102606.



Haiyang Yu received the Ph.D. degree in traffic environment and safety technology from Jilin University, China, in 2012. He is currently a Full Professor with the School of Transportation Science and Engineering, Beihang University, China. His research interests include traffic big data, traffic control and simulation, and intelligent vehicle infrastructure cooperative systems.



Binghong Jiang received the B.S. degree in automation from Xi'an Jiaotong University, China, in 2023. He is currently working toward the Ph.D. degree with the Department of Automation, Tsinghua University, Beijing, China. His research interests include ITSs and intelligent vehicles.



Xiaobo Qu has been a Chair Professor with the School of Vehicle and Mobility, Tsinghua University since 2021. His research is focused on ITSs, ground-air cooperation and vertical transportation systems, and emerging transport mode informed mobility services. He has authored or coauthored over 200 journal articles published in top tier journals, including 21 highly cited ESI studies. He has been an elected Member of Academia Europaea—the Academy of Europe since Aug 2020. He was a Faculty Member at two Australian universities from 2012 to 2017, joined Chalmers University of Technology, Sweden, as a Professor in Feb 2018, and promoted to Chair Professor rank in 2020.



Li Li received the Ph.D. degree in systems and industrial engineering from the University of Arizona, USA, in 2025. He is currently a Professor at Tsinghua University. His research interests include AI, ITSs, and intelligent vehicles.