Transportation Research Part D 136 (2024) 104470

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

ChatEV: Predicting electric vehicle charging demand as natural language processing

Haohao Qu [a,b], Han Li [a], Linlin You [a,*], Rui Zhu [c], Jinyue Yan [d], Paolo Santi [e], Carlo Ratti [e], Chau Yuen [f]

[a] School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China
[b] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
[c] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore
[d] Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hong Kong, China
[e] Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cam bridge, MA, USA
[f] School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Keywords:
Electric Vehicle
Charging demand
Time-series forecasting
Large Language Models

ABSTRACT

The increasing popularity of electric vehicles (EVs) in recent times has introduced considerable load conditions for urban power grids and transportation systems, which highlights the importance of accurately predicting charging demand to enhance charging efficiency. However, current forecasting methods still face challenges in effectively aligning diverse data and generating accurate predictions that can be applied to unseen scenarios. To overcome the challenges, this work introduces a novel perspective: employing large language models (LLMs) as EV charging demand predictors. First, we reformulate the prediction task into a text-to-text format, enabling seamless and effective alignment of various features within a unified language semantic space. Subsequently, we fine-tune a LLM using a meta-learning framework to adapt it specifically for EV charging prediction. Through comprehensive evaluations, it has been demonstrated that the proposed model, ChatEV, achieves outstanding performance in EV charging demand forecasting, particularly in scenarios with limited data.

1. Introduction

In recent years, there has been a notable increase in the adoption of electric vehicles (EVs) among vehicle consumers worldwide, leading to a significant purchasing trend. According to the Global EV Outlook (IEA, 2023), sales of EVs surpassed 10 million in 2022 and approached 14 million in 2023. This surge in EV adoption is primarily driven by the goal of aligning with climate ambitions, as EVs can make a substantial contribution to reducing harmful emissions and fostering a more sustainable transportation sector (You et al., 2024; Sacco et al., 2022). While such transition brings significant environmental benefits, the proliferation of EVs also poses challenges to urban power and transportation systems, largely due to the limited availability of public charging infrastructure (Unterluggauer et al., 2022; Kim et al., 2023; Zhang et al., 2021a). This highlights the important role of EV charging optimization in promoting the advancement of urban vehicle electrification (Pasha et al., 2024; Song and Hu, 2023; Wald et al., 2023), e.g., strategically scheduling the supply of public charging infrastructure across urban areas can effectively flatten the loads on power grids and transportation systems.

* Corresponding author.
E-mail address: youllin@mail.sysu.edu.cn (L. You).

https://doi.org/10.1016/j.trd.2024.104470
Received 19 May 2024; Received in revised form 10 October 2024; Accepted 13 October 2024
Available online 23 October 2024

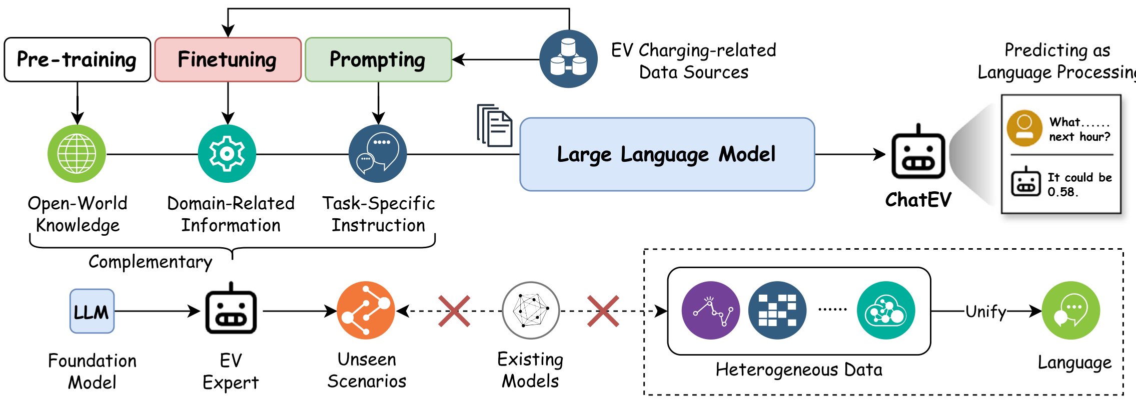


Fig. 1. The key idea of developing a LLM-empowered predictor for EV charging demand.

To enable such optimization, a crucial component is an effective EV charging demand prediction method. In general, the method is tasked with estimating future charging demand for EVs within a defined time frame and geographical location, which involves analyzing historical data on EV charging patterns, along with relevant contextual information. With accurate demand predictions, EV drivers can plan their trips more effectively and reduce cruising time for parking (Zhang et al., 2021b); regulators can adjust the power generation and distribution accordingly to ensure a stable and reliable energy supply (Ye et al., 2023). In the pursuit of accurate prediction, many studies have made significant contributions by developing advanced data-driven forecasting methods (Huang et al., 2024; Meng et al., 2024; Liu et al., 2023), which encompass three main aspects: (1) capturing time-series patterns in EV charging history through the utilization of statistical analysis (Yi et al., 2022); (2) modeling non-linear relationships between charging-related factors (e.g., time of day and weather conditions) by employing deep learning methods (Zhang et al., 2023; Wang et al., 2023b); and (3) incorporating geographical information around EV charging stations through the use of spatio-temporal models (Wang et al., 2023a; Kuang et al., 2024). However, the adoption of these data-dependent methods in EV charging demand forecasting still faces several challenges. First, the scarcity of high-quality EV charging data prevents existing models from being adequately trained. Second is their difficulty in understanding the relationships between heterogeneous data (Qu et al., 2023), such as time series, geographical information, and socio-economical factors. This hinders a comprehensive charging demand prediction as well as the charging-related managements. Finally, a generalizable foundation model capable of robust demand prediction in diverse areas remains to be explored, which is constructive and supportive for charging optimization policy-making.

Fortunately, the rapid development of Large Language Models (LLMs), like ChatGPT, has recently showcased notable milestones for significantly advancing various areas (Brown et al., 2020; Zhao et al., 2023), highlighting the immense potential of LLMs in revolutionizing the development of next-generation EV charging predictors. Equipped with million-scale parameters, these language models have exhibited unprecedented natural language understanding and reasoning abilities (Zhang et al., 2024), along with remarkable open-world knowledge that facilitate them to better generalize to time-series forecasting (Jin et al., 2024b). For example, Xue and Salim (2023) propose a novel LLM-empowered prompt-based predictor for time-series forecasting, named PromptCast, which exhibits impressive adaptation ability in the predictions of weather temperature, energy consumption, and customer flow. Another notable illustration is LLM-MPE (Liang et al., 2024a), which introduces a novel paradigm utilizing the Chain-of-Thought technique (Wei et al., 2022) to guide LLMs in predicting human mobility. This approach takes into account historical mobility patterns but also event features.

Despite the aforementioned success, exploring LLMs for forecasting EV charging demand poses significant challenges that have yet to be fully addressed. First, while there have been groundbreaking endeavors to utilize LLMs for general time series forecasting, the development of a tailored model specifically for EV charging demand prediction remains imperative to fulfill the diverse demands of real-world scenarios. For instance, companies in the EV charging industry may prioritize a specialized LLM-based predictor distinguished by personalized insights and cost-effectiveness over a generic solution. Furthermore, although demand predictions for EV charging and other forms of human activities in urban transportation systems, such as biking (Liang et al., 2023, 2024b) and driving (Wang et al., 2024), share similarities, they are influenced by distinct factors, such as charging prices, to varying degrees. These nuanced distinctions set EV charging demand prediction apart from general time-series forecasting, underscoring the need for a specialized LLM-based EV demand predictor. Finally, well-crafted instructions are crucial to fully harness the remarkable understanding and reasoning capabilities of LLMs for optimizing EV charging demand prediction, but this issue remains under-explored and warrants further investigation. To bridge this gap, this paper introduces a novel LLM-empowered predictor called **ChatEV**. As illustrated in Fig. 1, the proposed approach involves the finetuning and prompting of a widely-used pre-trained language model, Sentence-T5 (Ni et al., 2022), to turn it into a generalizable EV charging demand predictor. To achieve this, there are two essential steps: prompt-based task reformulation and multi-area alignment tuning. First, we deeply immerses the proposed predictor into a full language environment, wherein the task of EV charging demand prediction is reformulated into a text-to-text format by utilizing personalized prompts. By doing so, ChatEV can leverage the abundant open-world semantic knowledge embedded in the pre-training corpora of LLMs and align heterogeneous data within the same language semantic space, thus facilitating accurate predictions across diverse areas. Subsequently, we employ an efficient model-agnostic meta-learning method called Reptile (Nichol and Schulman, 2018) to fine-tune the LLM backbone, thus incorporating domain-specific knowledge related to EV charging demand prediction. In this process, the language modeling objective is used, i.e., generating the next token based on previous tokens, instead of using task-specific loss functions. This enforces the backbone model to deeply integrate language and forecasting semantics, so as to achieve an effective and efficient adaptation of the language model to EV charging demand prediction.

To evaluate our approach, this paper conducts extensive experiments on a real-world dataset (Qu et al., 2023) with 18,061 public EV charging piles, and study how ChatEV performs compared with both representative and recently developed prediction models on three forecasting scenarios, namely full-, few-, and zero-shot settings of training data. In addition, ablation studies are performed to demonstrate the effectiveness of each component in ChatEV. Overall, our main contributions are:

- This paper introduces ChatEV as a promising alternative for LLM-empowered EV charging demand prediction, by reformulating the prediction task into a text-to-text format that is compatible for LLMs. This enables a straightforward and effective fusion of heterogeneous charging-related factors, such as weather conditions, charging price, and road density, within a unified language semantic space. While numerous pioneering studies have investigated the application of LLMs for time-series prediction, to the best of our knowledge, this paper marks the very first endeavor in utilizing LLMs for EV demand forecasting.
- The proposed meta-learning-based alignment tuning method enhances the generalization capability of our LLM-based predictor, enabling accurate predictions across diverse and previously unseen scenarios.
- The evaluation results show that ChatEV attains outstanding performances compared to competitive methods, especially in few-shot and zero-shot forecasting tasks, demonstrating its exceptional generalization ability in EV charging demand prediction.
- We discuss the policy implications of promoting LLM-empowered EV demand predictors, exploring the current challenges and future directions in EV planning and management within the LLM era to provide in-depth insights for future research endeavors.

The remaining sections of this paper are organized as follows. Section 2 provides a literature review that summarizes existing solutions for EV charging demand prediction and LLM-based time-series forecasting. Next, Section 3 presents the proposed LLM-empowered predictor, which is evaluated in Section 4. In Section 5, we present a discussion of policy implications. Finally, Section 6 concludes the paper.

2. Related work

This section provides a brief review of the recently developed methods that focus on predicting EV charging demand. Furthermore, it highlights the emergence of time-series predictors that utilize LLMs as a promising approach.

2.1. EV charging demand prediction

At its core, predicting EV charging demand refers to the process of estimating the availability or demand status of charging stations/areas at a given time or in the near future, by analyzing historical data on EV charging patterns, such as charging sessions, charging duration, and charging locations, along with relevant contextual information. As the prediction models continue to advance, several challenges are emerging in the pursuit of efficient and effective forecasting, namely

- Data Scarcity: The number of EV charging stations is still relatively small compared to traditional fuel stations, resulting in a scarcity of data points and an inadequate training of prediction models.
- Factor Heterogeneity: Various factors influence the behavior of EV charging, including infrastructure distribution, pricing schemes, and weather conditions (Pasha et al., 2024). However, incorporating and analyzing these diverse factors poses a challenge for data-driven prediction models due to their heterogeneity.
- Model Generalizability: The utilization pattern of EV charging stations exhibits variations over time and across different locations (Meng et al., 2024). Consequently, there is a need for a unified and scalable predictor that can effectively handle both common patterns and specific contexts for the demand forecasting, even in unseen areas.

In recent years, many efforts have been made to tackle the aforementioned challenges. At the beginning, related solutions primarily focus on statistical analysis to enhance the interpretation of temporal patterns in charging records (Jeon et al., 2022; Yi et al., 2022). For example, Zhang et al. (2023) first leverage clustering methods with K-means kernels for analyzing the profiles of EV users, followed by introducing a simulation process for short-term charge demand prediction. More recently, with the progress of big data techniques, deep learning methods sparked in the field of time series forecasting, including EV charging demand modeling and prediction (Abdelaty et al., 2021), since these methods can effectively capture non-linear patterns. As illustrated by Wang et al. (2023b), a recurrent model with Long Short-Term Memory blocks achieves a significant performance improvement over statistical models in short-term EV charging demand prediction. Later on, there is an increased emphasis on incorporating spatial knowledge to enable forecasting methods to have a comprehensive understanding of charging situations (Kim and Kim, 2024). For example, HSTGCN (Wang et al., 2023a) utilizes graph convolutional layers to extract underlying relationships between neighboring areas. Moreover, PIAST (Kuang et al., 2024) combines graph and temporal attentions for capturing spatial and temporal correlations in charging occupancy as well as price, respectively. Even though these data-driven models can handle intrinsic characteristics of the charging data, they still struggle to align heterogeneous factors and support generalizable forecasting in unseen areas.

Table 1
Example for charging area/station characterization, which can be easily retrieved using area/station IDs.

Area/Station ID = 12
INFORMATION Coordinates = [22°32′29″N, 114°03′35″E]; Address = “No. 66 Gongchang Rd, Guangming Dt, Shenzhen, China”; Time: [5:29 pm, Tuesday, April 30, 2024] Charging Pile Number = 46; Road Length = 83.23 km; Weather = [22 °C, ...]; ......

2.2. Time-series forecasting empowered by LLMs

Recent years have experienced a booming success of large language models, such as BERT (Kenton and Toutanova, 2019), Sentence-T5 (Ni et al., 2022), and ChatGPT, which scales up their parameters to the million level over large-scale mixture-of-source corpora and have demonstrated groundbreaking capabilities in natural language understanding and reasoning. Hence, pioneering studies (Xue and Salim, 2023; Gruver et al., 2024; Lai et al., 2024; Li et al., 2024b) have attempted to explore the potential of LLMs for time-series forecasting using a new paradigm based on prompting, which encodes time series into natural language sentences and generates semantic tokens as model prediction. For instance, PromptCast (Xue and Salim, 2023) achieves LLM-empowered forecasting on weather temperature, energy consumption, and customer flow, without any modifications on the LLM architectures. Recent advancements in LLM-based foundation models for time-series, such as Time-LLM (Jin et al., 2024a) and UrbanGPT (Li et al., 2024a), have showcased their astonishing performance by reprogramming into a language task. Unlike conventional forecasting models that often involve complex parameter searching and training from scratch, the new paradigm (i.e., LLM-empowered time-series forecasting) offers a simpler and more accessible alternative and underscore their impressive generalizability in the realm of time series forecasting.

However, exploring LLMs for EV charging optimization is not a trivial task and remains largely unexplored. Firstly, it is challenging to design personalized prompts for reformulating this intricate task when considering various heterogeneous factors in neighboring areas. Moreover, while LLMs possess a vast amount of open-world knowledge, their training corpus may lack information regarding EV charging demand forecasting. Finally, even though several generalized models have been devised for time-series, a dedicated model tailored for EV charging demand remains essential to explore alternatives with reduced model size and the capability to offer personalized functions in specific scenarios. To tackle these shortcomings, this paper introduces a finetuned language model as predictor with prompts tailored for EV charging demand prediction. To the best of our knowledge, ChatEV represents the pioneering effort in approaching generalizable EV charging demand forecasting from a language-based perspective.

3. The proposed approach

In this work, we focus on the point prediction task but consider the average charging demand in neighboring areas. To better align natural language with EV charging demand prediction scenarios, we propose a novel LLM-empowered predictor, named ChatEV. As shown in Fig. 2, this is achieved by conducting two key steps, namely prompt-based task reformulation and multiarea alignment tuning. In particular, we first create a personalized prompt template to support the reformulation from numerical forecasting to text-to-text generation. Then, we perform a stable finetuning on a widely used pre-trained language models with million-scale parameters, i.e., Sentence-T5 (Ni et al., 2022), integrating within a meta learning framework known as Reptile (Nichol and Schulman, 2018). The details will be described in the following sections.

3.1. Prompt-based task reformulation

To reformulate the prediction task as natural language processing, we create a personalized prompt template, which covers fundamental information and additional instructions. In this section, our prompt design and task formulation will be introduced.

3.1.1. Prompt design

Generally speaking, a prompt includes two templates for model inputs and targets, along with a set of related metadata. Sanh et al. (2021). Accordingly, we design our prompts, as described below.

- **Area Characterization**. To provide the LLM backbone with personalized knowledge specific to a particular charging area or station, our prompts will include a detailed description of charging-related features using natural language. These features may include geographic information (e.g., coordinates and address), socio-economic features (e.g., road length density and POI density), weather conditions (e.g., temperature and humidity), and more. Moreover, we compute the average charging-related variables (such as demand and price) from neighboring areas and integrate them into the area description to incorporate spatial awareness. This strategy is inspired by the concept of simple graph message passing (Wu et al., 2019), which uniformly

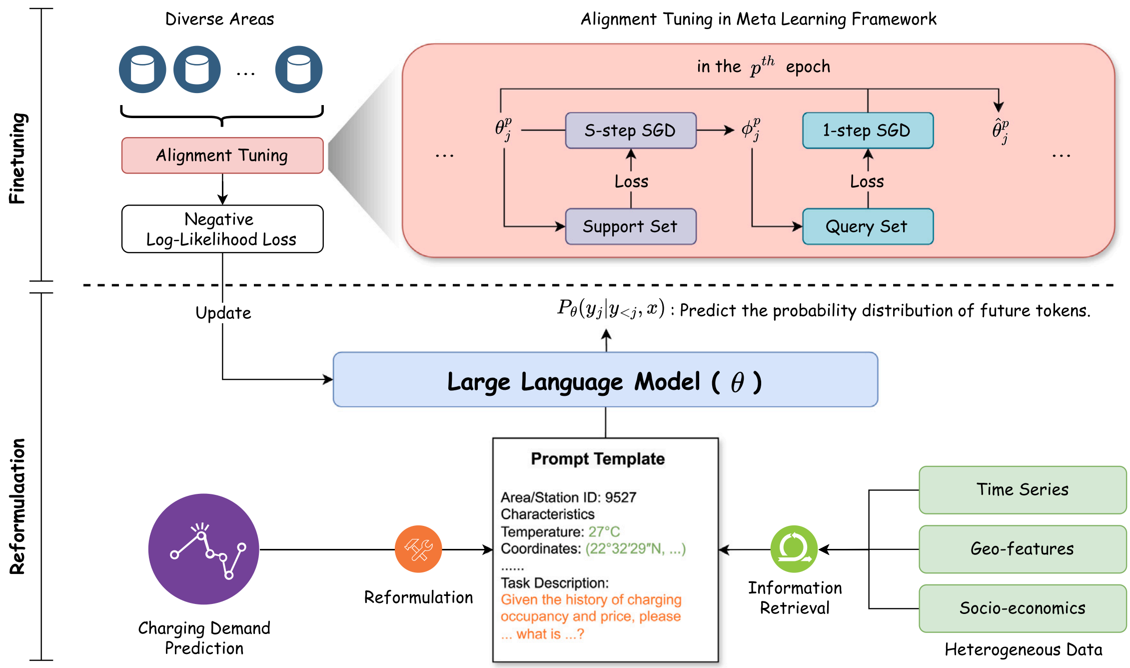


Fig. 2. The overall framework of ChatEV. Specifically, we reformulate the electric vehicle charging demand prediction task into a unified text-to-text format, and finetune the language model (i.e., Sentence-T5) within a meta-learning framework for knowledge alignment and model adaptation.

consolidates features from neighboring nodes. By doing so, ChatEV converts spatial data into textual elements, eliminating the need for additional training on time-series patches in current LLM-driven time-series forecasting models (Jin et al., 2024a; Li et al., 2024a). To sum up, such a characterization helps the LLM backbone to identify the studied charging area/station efficiently, thus facilitating personalized adaptation.

<u>**INPUT**</u>
You are an expert in electric vehicle charging management,
who is good at <charging demand prediction>.
We are now in {area characterization}.
Given the following time series of historical charging data,
Local Charging Occupancy = [..., 0.24, 0.23, 0.25, 0.26, 0.31, 0.35, ...];
Average Neighboring Charging Occupancy = [..., 0.33, 0.33, 0.32, 0.31, 0.28, 0.27, ...];
......
Now, pay attention! Your task is to <predict the charging demand in the area for the next hour> by analyzing the given information and leveraging your common sense.
In your answer, you should provide the value of your prediction only.

<u>**TARGET**</u>
ChatEV: 0.58.

- **Information Retrieval**. To attain up-to-date metadata from different sources for area characterization, we adopt an emerging technique, known as retrieval-augmented generation (RAG) (Lewis et al., 2020), which enriches LLMs' knowledge by pulling in information from a corpus of useful data, just like fetching a book from a library. In our case, RAG means to create a knowledge base with EV charging-related metadata, where corresponding information can be easily retrieved based on charging area/station IDs and inserted into our prompt template. Notably, all retrieved metadata will be transformed into textual tokens that is compatible for LLMs, as shown in Table 1. Integrated with RAG, our ChatEV can generate not only contextually accurate but also information-rich predictions.
- **Zero-shot Augmentation Instruction**. With the pursuit of enhancing the reasoning ability of our ChatEV, we apply a simple but effective zero-shot augmentation instruction in our prompt designing, i.e., role-playing. Role-playing entails the LLM to adopt a specific role, which the AI utilizes to perform the assigned task more proficiently (Shanahan et al., 2023). For example, the instruction can be presented as "*an expert in <specific field>, who is good at <specific task>*". Building upon this insight, we

define a detailed role for our LLM-empowered predictor. Drawing from the psychological principle, role-playing can help us program LLMs more effectively, which harnesses the power of the 'mindset shift', thus being expected to refine the outputs of our LLM-based predictor.

To be intuitive, we present our prompt template in the above text box, where texts in blue are the zero-shot augmentation instruction, texts in orange denotes the charging-related metadata retrieved by area IDs, and texts in red are the task descriptions. For traditional statistical or deep learning models, it is challenging to align these heterogeneous characteristics and model their relationships. Instead, ChatEV unifies them through natural language and leverage the open-world knowledge of LLMs to achieve an effective feature understanding. To sum up, our prompts involves instructing the LLM backbone, listing charging related factors, and presenting certain tasks within a sequence-to-sequence format.

3.1.2. Large language model-empowered forecasting formulation

In this section, we will give a unified mathematical definition of ChatEV's pipeline, including its input, output, and objective function. Firstly, for an area or station i, ChatEV's input $\mathcal{X}_i$ can be presented by the following combination in natural language according to the abovementioned prompt design:

$$\mathcal{X}_i = (\mathcal{P}, \mathcal{T}_i, \mathcal{O}_i), \tag{1}$$

where $\mathcal{P}$ denotes the prompt template; $\mathcal{T}_i$ represents the characterization of the area/station i, including time-series features (e.g., pile occupancy, charging price, and temperature) and static factors (e.g., charging type, road length, land area, and coordinates). $\mathcal{O}_i = \{\mathbf{o}_i, \mathbf{o}_j | j \in \mathcal{N}_i\}$ are the local and neighboring charging demand histories from time $(t-w)$ to current time t, where w is the lookback window size in recurrent forecasting tasks, and $\mathcal{N}_i$ denotes the set of neighbors of area i. Notably, the input $\mathcal{X}_i$ will be tokenized by a LLM tokenizer (i.e., SentencePiece) into in-vocabulary discrete tokens. Given these input tokens, we employ a widely-used LLM with millions of parameters, T5 (Ni et al., 2022), as our backbone to generate the future demand in area/station i based on its understanding of the task-specific instruction. Formally, the prediction $\mathcal{Y}_i$ can be calculated by:

$$\mathcal{Y}_i = \mathrm{LLM}(\mathcal{X}_i). \tag{2}$$

Specifically, $\mathcal{Y}_i$ is the token of target demand at time $(t+\lambda)$ in area/station i, where λ is the forecasting horizon. In a word, the objective of our LLM-empowered model lies on accurately predicting the future demand in specific areas or stations based on a range of textual information.

Finally, we reformulate the EV charging occupancy prediction objective as a conditional language generation task in an auto-regressive manner, i.e., generating the next token based on previously generated tokens. In this context, the model parameters θ are optimized by minimizing the negative log-likelihood (NLL) of target tokens $\mathcal{Y}_i$ given the input text $\mathcal{X}_i$, expressed as

$$\mathcal{L}_{\mathrm{NLL}} = -\sum_{i=1}^{I}\sum_{k=1}^{K} \log P_\theta(y_i^k | \mathcal{X}_i, y_i^{<k}), \tag{3}$$

where I and K represent the numbers of EV charging areas and target tokens, respectively; y_i^k is the k_{th} token of the output $\mathcal{Y}_i$ and $y_i^{<k}$ denotes the tokens before y_i^k; $P_\theta(\cdot)$ is the probability distribution of tokens based on the model parameters θ. During the inference phase, our model, ChatEV, has the capability to directly perform various tasks in both familiar and unfamiliar scenarios, and the proposed model simply uses greedy decoding to generate answers. As a result, the EV charging demand prediction for diverse scenarios are consolidate with a unified data format, a shared model, a single loss.

3.2. Alignment tuning for domain adaptation

Finetuning has recently been demonstrated as a promising technique to specialize pre-trained language models to perform downstream tasks. However, in real-world scenarios, the EV charging patterns usually varies across time and space. Direct fine-tuning using a variety of samples from different charging areas could potentially lead to the LLM backbone learning in the wrong direction, leading to poor performance on other datasets. To enhance the learning process, we apply a simple but effective Model-Agnostic Meta-Learning method, called Reptile (Nichol and Schulman, 2018) for knowledge adaptation. It simply works by repeatedly sampling each task and performs stochastic gradient descent (SGD) on each task in a standard way to achieve an unbiased and efficient tuning. Next, we will detail the Reptile tuning in our cases.

Assume that there are I EV charging stations/areas as source domains, our approach involves initially dividing their data into two distinct sets: the Support Set for exploration and the Query Set for harvesting, as shown in the finetuning part of Fig. 2. To elaborate further, we conduct S steps of SGD on the Support Set. This process enables us to obtain a set of learned parameters. Subsequently, we determine a well-established optimization direction by performing a one-step SGD on the Query Set based on the parameters learned from the Support Set. This step allows us to refine the optimization process. Formally, the whole tuning process of Reptile for ChatEV can be described in Algorithm 1. After the proposed tuning process, ChatEV is expected to be generalizable for EV charging demand prediction, even in unseen charging areas.

To sum up, the novelty of the proposed approach can be expressed as: (1) **Knowledgeable**. ChatEV can incorporate the open-world knowledge presented in LLMs to support EV charging demand prediction. (2) **Compatible**. In a unified language semantic space, ChatEV can easily fuse heterogeneous data and make full use of various information from different sources. and (3) **Generalizable**. ChatEV leverages the impressive reasoning ability in LLMs to enable accurate prediction on diverse unseen areas by prompt engineering and alignment tuning.

Algorithm 1 Procedures of the proposed alignment tuning.

Require: A pre-trained language model θ and I source domains from diverse charging areas for finetuning;
Divide the data of each area into Support and Query sets;
for epoch 1,2,3,...,p,... **do**
  Randomly sample domain i;
  Perform $S > 1$ steps of SGD using Adam optimizer on the Support set of area i, starting with the pre-trained parameters θ_i^p, resulting in temporary parameters ϕ_i^p;
  Perform one-step of SGD using Adam optimizer on the Query set of area i, starting with the temporary parameters ϕ_i^p, resulting in tuned parameters $\hat{\theta}_i^p$;
  Update: $\theta_{i+1}^p \leftarrow \theta_i^p + \epsilon(\hat{\theta}_i^p - \theta_i^p)$
end for
Return θ

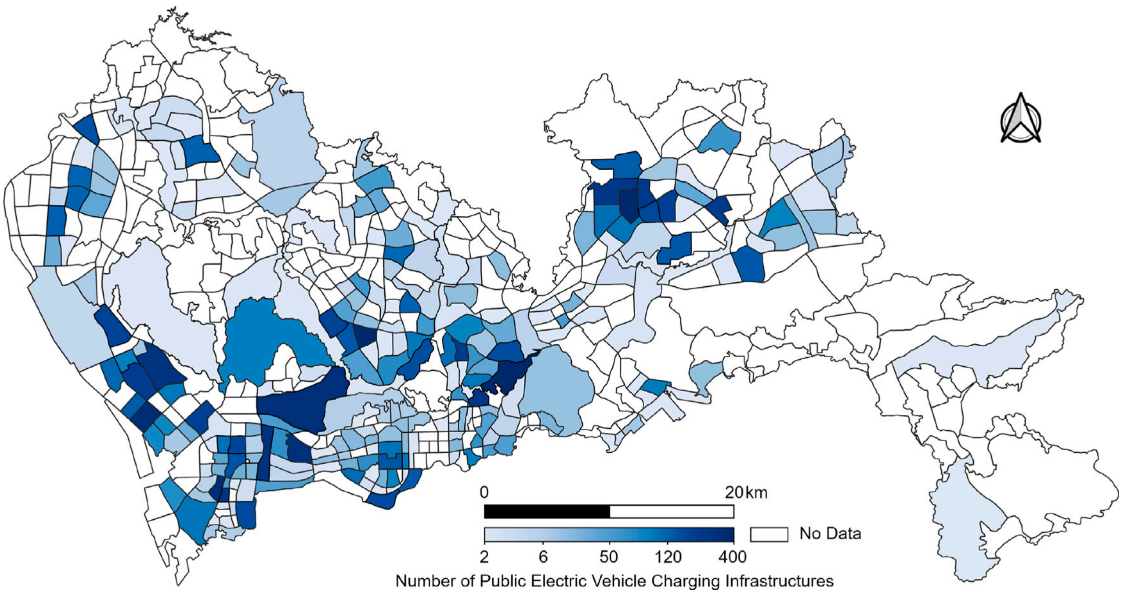


Fig. 3. Map of the 247 studied traffic zones. The image is created based on the dataset introduced by Qu et al. (2023).

4. Performance evaluation

In this section, we will evaluate ChatEV in comparison to representative forecasting methods, using a benchmarking dataset and identical evaluation metrics, to demonstrate the improvements accomplished.

4.1. Experimental settings

4.1.1. Dataset

To demonstrate the effectiveness of the proposed method, we conduct comprehensive experiments over an up-to-date real-world dataset: ST-EVCDP.[1] This dataset offers real-time charging occupancy information of 18,061 public EV charging piles for 247 traffic zones within Shenzhen, China. More specifically, it covers a period of 30 days, from 19 June to 18 July 2022, with a time interval of 5 min. Notably, this evaluation employs charging occupancy as the indicator of charging demand within a particular area. Besides, various important characteristics for the studied areas are also collected, including the charging pricing schemes, coordinates, area adjacency, road density, points of interest, and charging types (i.e., fast or slow charging). Notably, according to the zero-shot setting, we adopt a strategy where a portion of the traffic zones are designated as unseen areas, meaning they are not included in the fine-tuning corpus. The ratio of selection for these unseen areas is explored within the range of [0.2, 0.4, 0.6, 0.8]. For the traffic zones that are considered as seen areas, the available data is partitioned into training set (60%, Day 1–18), validation set (20%, Day 19–24), and test set (20%, Day 25–30) in a chronological order (see Fig. 3).

[1] https://github.com/IntelligentSystemsLab/ST-EVCDP.

4.1.2. Baseline methods

Two representative statistical methods, four competitive neural network methods, and a recently developed LLM-empowered method are used as baselines. To be specific, the statistical methods are **ARIMA**, which is a statistical analysis model that makes use of lagged moving averages to predict future values; and **Lasso** (Tibshirani, 1996), a regression analysis method that combines variable selection and regularization techniques to improve prediction accuracy while maintaining interpretability of the model. The compared neural networks are: **FCNN**, a vanilla full connected neural network; **LSTM**, a typical recurrent neural network with long short-term memory, which is employed by Wang et al. (2023b) for EV charging demand estimates; **GCN-LSTM** (Chen et al., 2022), a representative method for spatio-temporal traffic flow prediction; **STGCN** (Yu et al., 2018), a typical convolution-based model for spatio-temporal forecasting; **HSTGCN** (Wang et al., 2023a), a recently developed method that integrated graph convolutional layers and gated recurrent units to learn spatial and temporal patterns in the electric vehicle charging demand, respectively; **PIAST** (Kuang et al., 2024), a state-of-the-art model for electric vehicle charging demand prediction that considers the underlying influence of charging prices. The LLM-based baseline is **PromptCast** (Xue and Salim, 2023), which employs a pre-trained language model for prompt-based time series forecasting in the zero-shot setting; **LLMTIME** (Gruver et al., 2024), which tokenizes time-series data and transforms discrete distributions across tokens into highly adaptable digits. Notably, in order to testify the significance of adapting the LLM-based methods to downstream datasets, LLMTIME is allowed to be fine-tuned on the training corpus in the following experiments, while PromptCast is tested without finetuning.

4.1.3. Hyper-parameters

ChatEV is configured as following. First, a widely-used language model, i.e., Sentence-T5 (Raffel et al., 2020), is employed as the LLM backbone for our ChatEV. Second, it is trained with an advanced optimizer, AdamW (Loshchilov and Hutter, 2018), in a mini-batch manner, where the batch size and learning rate are searched in the ranges of {24, 32, 48, 64} and {0.0001, 0.001, 0.01, 0.1}, with optimal values of 48 and 0.001, respectively. Third, the fine-tuning process is set to run for 200 epochs, with an early stopping mechanism, which stops the process if the validation loss remains unchanged for 10 consecutive epochs. Fourth, the window sizes of historical and predicted data are 12 and 6 intervals, respectively. This means that the models being compared are required to forecast the charging volume for the next 30 min, taking into account the charging history of the previous hour. Finally, negative log-likelihood (NLL) is used as the loss function. Notably, in our model, we conduct a grid search on critical hyperparameters and implement them in the compared models. Meanwhile, the default hyperparameters for baseline methods are configured as recommended in the respective papers.

4.1.4. Running environment

Two common metrics for time series prediction are used, i.e., Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). All the experiments are conducted on a Linux workstation with two GeForce RTX 3090 GPUs.

4.2. Performance comparison

The compared methods are discussed in four aspects: (1) full-shot predicting to illustrate how good the model is in forecasting EV charging occupancy; (2) few-shot forecasting to demonstrate how fast the model is to adapt to specific areas; (3) zero-shot forecasting to show the ability of the model to handle the prediction tasks in unseen areas; and (4) the ablation experiments to testify each key component of ChatEV.

4.2.1. Full-shot forecasting

We first compare the full-shot forecasting performance between ChatEV and all baseline methods over the benchmark dataset. In this case, the compared prediction models and ChatEV are allowed to access the training data (i.e., Day 1–24) in all the studied traffic zones. Table 2 presents the performance comparison in the test set (i.e., Day 25–30) on the four forecasting intervals, i.e., 15 min, 30 min, 45 min, and 60 min, where the best and second best results are marked by **Bold** and underlined. We can make the following observations. First, our proposed ChatEV achieves the state-of-the-art performance and consistently outperforms all the baseline methods across all testing intervals in terms of all metrics. In average, ChatEV exceeds the strongest baseline by 3.7% and 4.6% in RMSE and MAE, respectively. Such improvement demonstrates the effectiveness of our proposed methods and the great potential of exploring finetuned large language models for EV charging demand prediction. In the full-shot scenario, it can be attributed to the exceptional learning capabilities obtained from the Transformer-based language model with million parameters, enabling our model to effectively capture the underlying patterns in the EV data. ChatEV also leverages the benefits of the proposed alignment tuning and the designed prompts, gathering comprehensive information spanning from temporal to spatial aspects. Second, as an earlier RNN method designed to capture hidden patterns in time series, LSTM outperforms ARIMA, Lasso, and FCNN. This observation suggests that the potential of integrating non-linear temporal knowledge for time-series forecasting. However, LSTM is inferior to the spatio-temporal predictors (i.e., GCN-LSTM, STGCN, HSTGCN, and PIAST), implying its insufficient ability to capture spatial information. Third, the spatio-temporal prediction methods show relatively stronger performance than the traditional methods (ARIMA and Lasso) and the representative recurrent method (i.e., LSTM). This highlights the significance of incorporating the spatial features via learning adjacency relationships. However, these advanced methods still struggle to incorporate heterogeneous data, such as socio-economic features, for EV charging demand prediction. For LLM-based methods, PromptCast is a competitive baseline as it achieves impressive prediction accuracy in all metrics, without the need for additional adaptation or fine-tuning. It accomplishes this by utilizing the natural language understanding capabilities of LLM (Language Model) to incorporate diverse

Table 2
Performance comparison of full-shot forecasting in different prediction intervals (5 min per interval).

Metric (10^{-2})	RMSE					MAE				
Model	3	6	9	12	Average	3	6	9	12	Average
ARIMA	4.57	6.48	7.85	9.06	6.99	2.49	3.71	4.69	5.60	4.12
Lasso	4.34	6.27	7.92	9.22	6.94	2.50	3.69	4.65	5.57	4.10
FCNN	4.52	6.41	7.86	8.89	6.92	2.45	3.60	4.50	5.32	3.97
LSTM	3.53	5.62	8.33	8.91	6.60	<u>1.87</u>	3.19	4.68	6.28	4.01
GCN-LSTM	3.29	5.69	7.26	8.66	6.23	1.95	3.30	4.27	5.21	3.68
STGCN	3.45	5.33	6.72	7.54	5.73	2.01	3.39	4.03	4.66	3.53
HSTGCN	3.34	5.27	6.52	7.44	5.64	2.00	3.30	<u>3.94</u>	4.55	3.45
PIAST	3.36	5.22	**6.36**	7.45	5.60	1.91	<u>3.11</u>	4.00	4.81	3.46
PromptCast	4.08	6.24	7.45	8.14	6.48	2.56	3.84	4.29	5.23	3.98
LLMTIME	<u>3.04</u>	<u>5.20</u>	6.58	<u>7.28</u>	<u>5.53</u>	1.98	3.26	4.02	<u>4.47</u>	<u>3.43</u>
ChatEV	**2.97**	**5.06**	<u>6.46</u>	**7.13**	**5.40**	**1.85**	**3.09**	**3.90**	**4.34**	**3.29**

Table 3
Performance on few-shot forecasting with limited training data, i.e.,the first {5%, 10%, 15%, 20%} of training time steps.

Metric (10^{-2})	RMSE					MAE				
Model	5%	10%	15%	20%	Average	5%	10%	15%	20%	Average
ARIMA	15.01	16.76	14.80	12.80	14.84	10.11	9.71	9.19	8.45	9.37
Lasso	16.09	15.63	12.10	14.38	14.55	10.62	9.38	8.56	8.45	9.25
FCNN	15.58	14.69	14.40	13.32	14.50	9.56	8.90	8.49	8.19	8.78
LSTM	14.72	13.75	13.63	12.42	13.63	9.16	8.18	8.40	8.10	8.46
GCN-LSTM	17.73	15.71	15.48	14.31	15.81	9.36	9.17	9.16	9.06	9.19
STGCN	17.26	16.22	15.74	14.43	15.91	9.46	9.32	9.24	9.06	9.27
HSTGCN	17.74	16.50	15.26	13.22	15.68	10.71	10.50	9.48	9.17	9.97
PIAST	10.17	9.88	9.41	8.96	9.61	6.95	6.20	6.01	5.66	6.21
PromptCast	6.48	6.48	6.48	6.48	6.48	3.98	3.98	3.98	3.98	3.98
LLMTIME	<u>6.36</u>	<u>6.20</u>	<u>5.72</u>	<u>5.65</u>	<u>5.98</u>	<u>3.84</u>	<u>3.60</u>	<u>3.55</u>	<u>3.48</u>	<u>3.62</u>
ChatEV	**5.84**	**5.77**	**5.60**	**5.49**	**5.67**	**3.55**	**3.52**	**3.49**	**3.38**	**3.48**

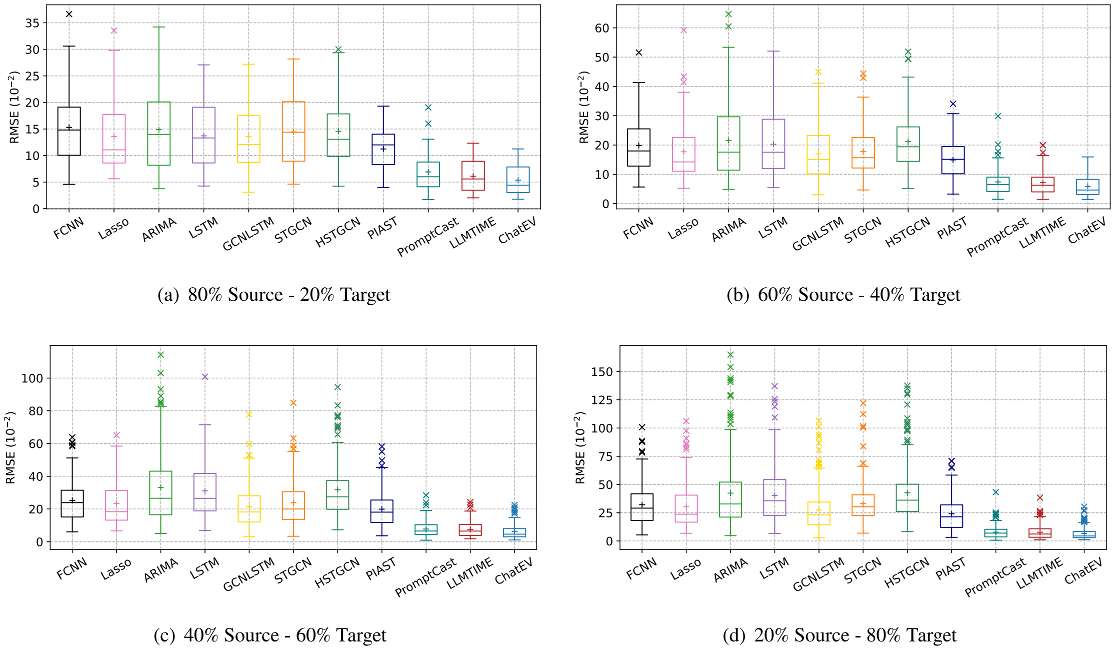
factors as textual contexts for forecasting. Furthermore, LLMTIME surpasses PromptCast, owing to its tokenization approach that inserts spaces between numbers and a typical fine-tuning process, which extracts specific knowledge from the EV data. These two LLM-empowered methods highlight the potential of natural language processing in serving as a promising alternative approach for next-generation EV charging demand forecasting. However, our proposed method outperforms PromptCast and LLMTIME because of the introduction of alignment tuning, which enhances the model's ability to align and capture relevant information effectively.

4.2.2. Few-shot forecasting

LLMs have recently demonstrated impressive capabilities in few-shot learning. Thus, in this section, we examine whether our finetuned LLM retains this ability when applied to EV charging demand forecasting tasks. Specifically, we evaluate scenarios with limited finetuning data (i.e., $\leq$ the first 20% of the training time steps). Our brief few-shot learning results are summarized in Table 3. The results reveal that existing data-driven forecasting methods, including the statistical and deep learning methods, perform poorly when faced with few-shot scenarios containing only {5%, 10%, 15%, and 20%} of the training data. Even though PIAST introduces additional prior knowledge (i.e. constraints) to help it outperform other learning models on the few-shot scenarios, its performance is still unsatisfactory. In contrast, the novel prompt-based forecasting method, PromptCast, outshines these data-driven methods significantly, which we attribute to the open-world knowledge present in its pre-training corpus. Note that PromptCast's prediction results remain consistent across all few-shot settings since it is a prompt-based model that does not undergo any fine-tuning process. Furthermore, LLMTIME incorporates the local knowledge in the limited training data and the prior knowledge derived from the LLM, leading to a better performance to PromptCast. More impressively, our approach surpasses all baselines, underscoring the potential prowess of language models as proficient EV charging demand predictors. Our findings also demonstrate that ChatEV can rapidly adapt to specific downstream prediction tasks in particular areas with only a small amount of finetuning data, showcasing the fast learning ability of our model.

4.2.3. Zero-shot forecasting

Beyond few-shot learning, LLMs hold potential as effective zero-shot reasoners (Kojima et al., 2022). In this subsection, the zero-shot learning capabilities of the proposed predictor are evaluated within the framework of cross-domain adaptation (Qu et al., 2022). To be specific, we examine how well a prediction model performs on unseen areas when it is optimized on other areas, where the model has not encountered any data samples from the target (i.e., unseen) areas. This means to randomly sample {20%, 40%, 60%, 80%} of the 247 studied traffic zones as source domains, when the remaining zones are used as target domains for testing. Notably, we use the 30-min forecasting protocol for this zero-shot evaluation, and the results on RMSE are presented by boxplots in Fig. 4, where the mean and median values of the prediction metrics are marked by "+" and the horizontal lines, respectively, and "×" denotes outliers. We can see that ChatEV shows remarkable generalizability in zero-shot cross-area adaptation, i.e., less outliers



(a) 80% Source - 20% Target

(b) 60% Source - 40% Target

(c) 40% Source - 60% Target

(d) 20% Source - 80% Target

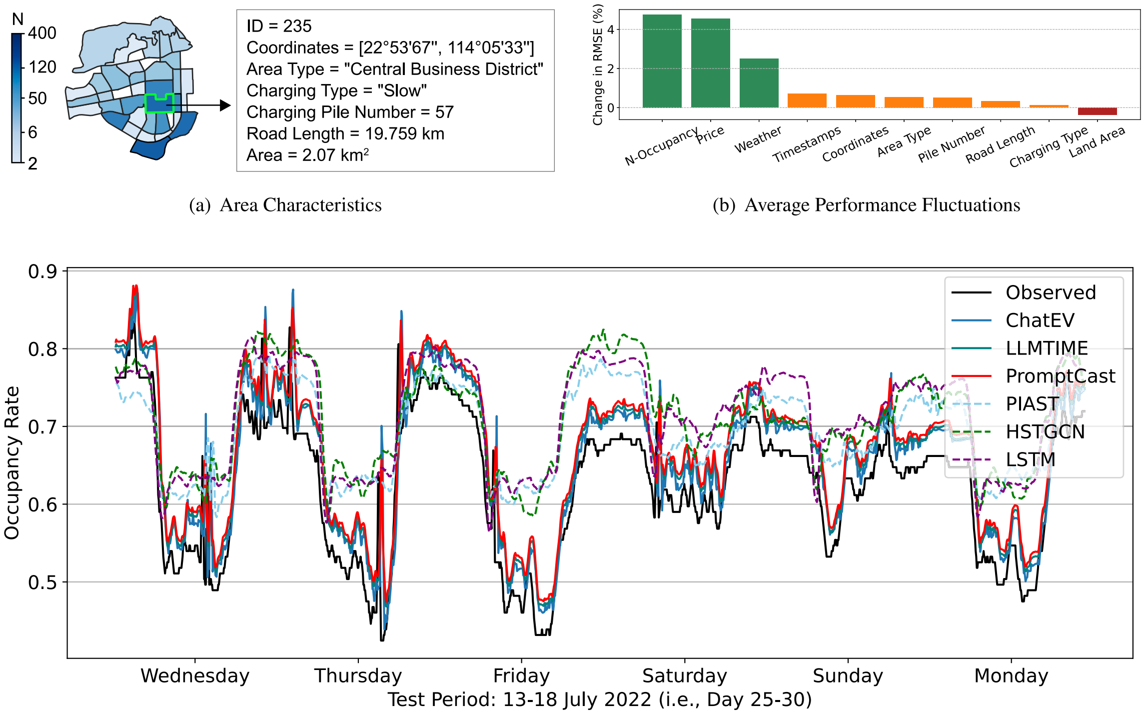
Fig. 4. Performance comparison on zero-shot forecasting within the framework of cross-area adaptation.

and lower forecasting errors than all baselines. This suggests that the proposed model can make reasonable predictions based on its open-world knowledge, even if it has not seen the relevant data in particular areas. Such a property generalizability underscores the great potential of ChatEV as a reliable online EV charging demand predictor in the real-world scenarios, where high-quality data are usually scarce. Similarly, PromptCast and LLMTIME utilize pre-trained language models for prompt-based forecasting, achieving fast adaptation second only to our model. This is because ChatEV capitalizes on the prompts tailored for electric vehicle charging demand prediction, enabling it to outperform other LLM-based methods such as PromptCast and LLMTIME, which adhere to prompts designed for general time-series forecasting. In contrast, the statistical and learning methods are stretched thin on this zero-shot forecasting task: as the source areas dwindle, their performances on target areas become worse and worse. It is difficult for these data-driven models to extract charging patterns common to all areas from such a small amount of data, which highlights again the significance of introducing open-world prior knowledge to support a generalizable forecasting task.

To be specific, an illustration of the zero-shot performance is presented in Fig. 5 to show directly on how efficient and effective the LLMs contribute to predict the charging demand (i.e., future occupancy). Specifically, sub-graph (a) showcases an example of area characteristics with diverse features; sub-graph (b) illustrates the significance of these features in prediction, represented by the performance changes resulting from removing each feature individually (where positive values indicate a decrease in model performance upon feature removal, and vice versa); sub-graph (c) displays the forecasting curves of baselines juxtaposed with the ground truth occupancy rate. Based on the figure, we can make the following observations. First, sub-graph (a) shows that diverse features from various data resources, such as coordinates, charging price, and weather conditions, can be fused efficiently in the natural language space that compatible for LLMs. In sub-graph (b), we can see that the static features like Coordinates, Area Type, Road Length, Pile Number, and Charging Type results in a performance decline. This highlights the importance of integrating static features as textual descriptions in LLM-based EV charging demand prediction to pinpoint target areas effectively. On the other hand, in our scenarios, the inclusion of "Land Area" information can introduce noise to the forecasting process. This is because the feature may vary irregularly across difference traffic areas; for example, many large areas might have a scarcity of charging piles instead. Compared to the static features, the dynamic variables, namely Neighboring Occupancy (N-Occupancy), Charging Price, and Weather (Temperature), exhibit a more significant impact on the prediction accuracy of ChatEV. Excluding these dynamic features from our model results in a notable decrease in prediction performance, indicated by an increase in RMSE. Finally, according to sub-graph (c), the two LLM-empowered predictors, namely PromptCast, LLMTIME, and ChatEV, is able to keep in sync with the ground-truth occupancy in the zero-shot settings. This impressive performance is attributed to the open-world knowledge provided by LLMs, which enables them to learn rapidly and generalize effectively. In contrast, the three deep learning baselines struggle with underfitting. Finally, ChatEV exhibits superior performance and achieves minimum prediction error thank to the personalized knowledge learned from the proposed prompts and aligning process.

4.2.4. Ablation study

In order to assess the key components in ChatEV (i.e., the designed prompt template, the alignment tuning, and the pre-trained knowledge), we conducted ablation experiments, where the impact of each component is eliminated separately: (1) w/o Prompting:



(a) Area Characteristics

(b) Average Performance Fluctuations

(c) Prediction Results

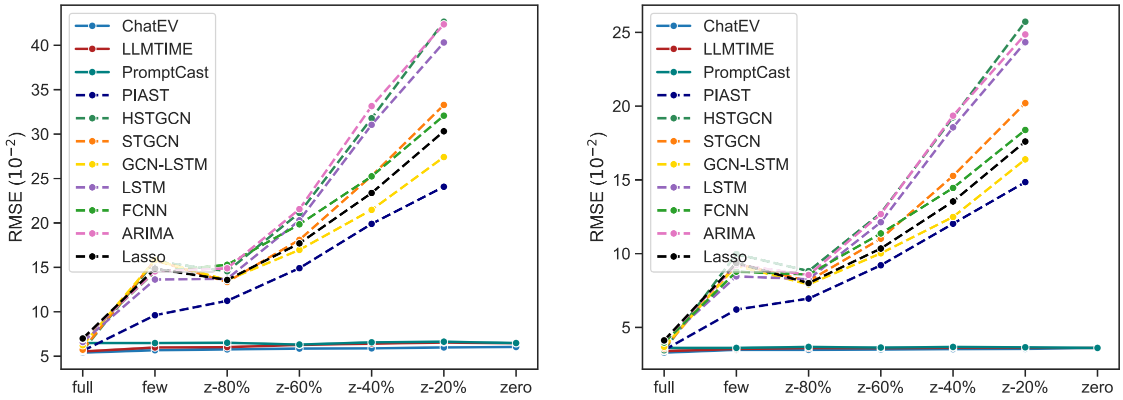
Fig. 5. Illustration of the forecasting results in the setting of (40% Source–60% Target). (a) Area characterization in area 235; (b) The average performance fluctuations of ChatEV in all areas resulting from removing each feature; (c) The prediction curves of compared models in area 235.

Table 4
Ablation results in full-, few-, and zero-shot forecasting scenarios. "w/o" denotes "without".

Metric (10^{-2})	RMSE				MAE			
Model	Full	Few	Zero	Average	Full	Few	Zero	Average
w/o pre-training	7.08	11.02	11.31	9.80	4.38	5.79	7.04	5.74
w/o finetuning	6.02	6.02	6.02	6.02	3.68	3.68	3.68	3.68
w/o aligning	5.57	5.83	5.95	5.77	3.51	3.62	3.81	3.65
w/o prompting	5.47	5.75	5.94	5.72	3.48	3.56	3.78	3.61
ChatEV	5.40	5.71	5.91	5.67	3.30	3.48	3.61	3.46

Remove the designed prompts, including area characterization and task instruction, while remain time series only as model input; (2) w/o Aligning: Deactivate the proposed alignment tuning process and perform a vanilla finetuning on our LLM backbone; (3) w/o Pre-training: A blank T5 model is used instead of the pre-trained ones to eliminate the open-world knowledge in the LLM; (4) w/o Finetuning: perform forecasting via prompting the language model only. Notably, these experiments are performed on the 30-min forecasting protocol. The ablation results in full-, few-, and zero-shot forecasting scenarios are shown in Table 4. In which, all few-shot metrics are averaged across four training data size: {5%, 10%, 15%, 20%}, while all zero-shot results are averaged from four sampling ratio of unseen areas: {20%, 40%, 60%, 80%}. From the ablation results, we can make the following observations. First, each component in our approach contributes to the overall performance since eliminating any one of them results in the performance decline. Second, a significant performance drop occurred when removing the open-world knowledge presented in large language models (i.e., without pre-training), especially in the few-shot and zero-shot scenarios. This emphasizes again the potential of employing LLMs as EV charging demand predictors. Additionally, without any finetuning, ChatEV experiences a performance reduction, but it can still achieve a comparable performance to PromptCast and LLMTIME, which demonstrates the effectiveness of the proposed prompts and the pre-trained knowledge.

To recap, Fig. 6 provides an overview of model evaluation across different scenarios, including full-shot, few-shot, and zero-shot scenarios. In which, the proportions of source areas available for training and finetuning {z-80%, z-60%, z-40%, z-20%} indicate the amount of training data from the source areas, while the "zero" scenario represents the absence of any training data. We can see that the utilization of Language Models (LLMs) for EV charging occupancy prediction is effective. Moreover, ChatEV exhibits superior performance in the full-shot settings, positioning it as a state-of-the-art solution. This improvement can be attributed to the integration of diverse information within the language space. Finally, the proposed model outperforms other baselines by a



(a) Performance in RMSE (b) Performance in MAE

Fig. 6. The overview of model evaluation demonstrates the performance of each method in various scenarios.

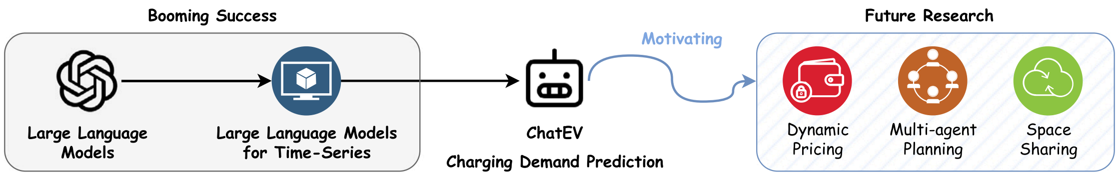


Fig. 7. ChatEV bridges the remarkable achievements of Large Language Models (LLMs) in the realm of time-series with the future investigations into LLM-based agents for the next-generation electric vehicle-related intelligent services.

Table 5
Comparison of the computational efficiency between LLM-based models and deep learning methods.

Model	GCN-LSTM	STGCN	HSTGCN	PromptCast	LLMTIME	ChatEV
Milliseconds per sample	8.57	5.81	6.83	36.52	47.73	39.58

significant margin in the few-shot and zero-shot settings, showcasing its exceptional generalization ability to adapt to new scenarios with limited data.

5. Policy implications

As an emerging branch in the transportation field, vehicle electrification has been witnessed a surge development motivated by the pursuits of gas energy conservation and emission reduction. Meanwhile, Large Language Models (LLMs), including advanced models like ChatGPT, have gained significant popularity and demonstrated tremendous capabilities across various areas and tasks in the field of Artificial Intelligence. As powerful and increasingly pervasive tools, LLMs have the potential to revolutionize the future of EV planning and management. In this context, this paper proposes to study new frontiers of EV charging demand prediction in the era of large language models, bearing both academic and practical impacts.

From the academic perspective, although numerous groundbreaking studies have explored the use of LLMs for time-series analysis, ChatEV represents the initial effort in employing LLMs for charging demand forecasting, a fundamental aspect of EV management. As illustrated in Fig. 7, the success achieved in LLM-empowered EV charging demand prediction is poised to inspire a resurgence of interest in addressing the research question: *How can large language models be harnessed for the development of next-generation AI agents for EV planning and management?* From the practical perspective, this paper emphasizes the importance of collecting and aggregating high-quality data in cloud-based platforms, facilitating a holistic understanding of charging dynamics and enabling strategic resource allocation to meet evolving needs. Furthermore, given that the computational efficiency of LLM-based predictors is roughly six times that of conventional deep learning methods, as depicted in Table 5, LLM-based approaches encounter difficulties in scenarios requiring high concurrency. Nevertheless, with response times under 0.05 s, they can still demonstrate significant efficacy in macro policy adjustments, where model performance and generalizability outweigh the importance of computational efficiency. Finally, urban regulators can benefit from simplified and user-friendly forecasting solutions with interactive interfaces provided by LLM-based agents, reducing reliance on complex coding and feature engineering. To sum up, as an emerging technique, large-scale pre-trained language models show promise in assisting electric vehicle management, thus facilitating the transition of vehicle power from gasoline to electricity.

Additionally, we present a series of policies in this section aimed at regulating LLM-based agents for electric vehicle charging prediction and management. First and foremost, urban administrations may be required to set forth regulations and norms concerning the utilization of large language models in overseeing electric vehicles, ensuring equitable and ethical practices. Such directives could encompass stipulations on data utilization, model transparency, and accountability. Second, authorities might contemplate providing incentives or subsidies to encourage the integration of electric vehicles managed by large language models, thereby expediting the shift towards more sustainable transportation alternatives. These strategies have the potential to hasten the transition towards cleaner transportation options. Moreover, to enable seamless communication and coordination between different systems and devices involved in electric vehicle management, policymakers may need to establish interoperability standards. This would ensure that data exchange and integration are smooth and to facilitate seamless communication and synchronization among various systems and devices involved in electric vehicle management, policymakers may need to establish interoperability standards, which would guarantee the smooth and efficient exchange and integration of data. Finally, policymakers may also need to allocate resources for research and development to bolster the capabilities of large language models for electric vehicle management. Financial support could foster innovation in key areas such as predictive maintenance, autonomous charging, and vehicle-to-grid technologies.

6. Conclusion

In this work, we highlighted a novel paradigm of exploring LLMs as EV charging occupancy predictors. On the one hand, by reformulating the prediction task into a text-to-text format, ChatEV provided an effective data fusion method to make full use of various information from different sources, such as weather conditions, coordinates, and time of day. On the other hand, ChatEV adopted a meta-learning-based alignment tuning method for knowledge adaptation, thus facilitating our LLM-empowered EV charging occupancy prediction. Through comprehensive experiments on 247 urban areas, we demonstrated that ChatEV can achieve state-of-the-art performance on charging demand prediction, while also exhibiting the capacity of being generalizable to unseen areas with limited data. Finally, we also engaged in a discussion regarding the potential implications of this research on EV charging planning and management in the era of LLMs. Moving forward, to further enhance the predictive capabilities of our approach and extend its applicability, future research could focus on the following aspects.

(1) To utilize larger models: Expanding the scale of language models, which involves utilizing the latest advancements in model architecture and training methodologies, to leverage their full potentials will enable more accurate predictions. The state-of-the-art models, e.g., the Llama (Touvron et al., 2023), GPT-4 (Achiam et al., 2023), etc., can be adopted.
(2) To integrate diverse tasks: In addition to predicting EV charging demands, ChatEV are expected to encompass other fundamental tasks associated with EV charging, such as data imputation and dynamic charging price estimation. This enhances its capacity to serve as an assistant for EV charging, offering comprehensive advice and services from multiple perspectives.
(3) To optimize model structure: Advanced external modules can be introduced, such as a Mix-of-Expert (Shazeer et al., 2017) layers for scalable content awareness and Adapter (Song et al., 2024) blocks for parameter-efficient tuning, so as to enhance the performance of our LLM-empowered predictor.
(4) To improve the interpretability of our model. Providing explanations for ChatEV's predictions is crucial for the LLM-based predictor to establish trustworthiness among policymakers. To achieve this, we will integrate advanced frameworks and metrics (Tang et al., 2023; Yuan et al., 2024) into our fine-tuning process, necessitate ChatEV to generate relevant explanations. Besides, it is a promising strategy to incorporate the Chain-of-Thought technique to instruct LLMs in sequential thinking (Liang et al., 2024a), bolstering their capacity for reasoning and handling complex EV-related tasks.

CRediT authorship contribution statement

Haohao Qu: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Han Li:** Writing – original draft. **Linlin You:** Writing – review & editing, Supervision, Conceptualization. **Rui Zhu:** Writing – review & editing. **Jinyue Yan:** Writing – review & editing. **Paolo Santi:** Writing – review & editing. **Carlo Ratti:** Writing – review & editing. **Chau Yuen:** Writing – review & editing.

Acknowledgements

This paper was partially supported by the National Key Research and Development Program of China (2023YFB4301900), the Guangdong Basic and Applied Basic Research Foundation (2023A1515012895), and the Department of Science and Technology of Guangdong Province (2021QN02S161).

Data availability

Data will be made available on request.

References

Abdelaty, H., Al-Obaidi, A., Mohamed, M., Farag, H.E., 2021. Machine learning prediction models for battery-electric bus energy consumption in transit. Transp. Res. D 96, 102868.

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

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.

Chen, Z., Lu, Z., Chen, Q., Zhong, H., Zhang, Y., Xue, J., Wu, C., 2022. Spatial–temporal short-term traffic flow prediction model based on dynamical-learning graph convolution mechanism. Inform. Sci. 611, 522–539.

Gruver, N., Finzi, M., Qiu, S., Wilson, A.G., 2024. Large language models are zero-shot time series forecasters. Adv. Neural Inf. Process. Syst. 36.

Huang, A., Mao, Y., Chen, X., Xu, Y., Wu, S., 2024. A multi-timescale energy scheduling model for microgrid embedded with differentiated electric vehicle charging management strategies. Sustainable Cities Soc. 101, 105123.

IEA, 2023. Global ev outlook 2023: Catching up with climate ambitions. URL: https://www.iea.org/reports/global-ev-outlook-2023.

Jeon, Y.E., Kang, S.B., Seo, J.I., 2022. Hybrid predictive modeling for charging demand prediction of electric vehicles. Sustainability 14, 5426.

Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J.Y., Shi, X., Chen, P.Y., Liang, Y., Li, Y.F., Pan, S., Wen, Q., 2024a. Time-LLM: Time series forecasting by reprogramming large language models. In: International Conference on Learning Representations. ICLR.

Jin, M., Zhang, Y., Chen, W., Zhang, K., Liang, Y., Yang, B., Wang, J., Pan, S., Wen, Q., 2024b. Position paper: What can large language models tell us about time series analysis. In: International Conference on Machine Learning. ICML.

Kenton, J.D.M.W.C., Toutanova, L.K., 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT. pp. 4171–4186.

Kim, H.J., Kim, M.K., 2024. Spatial–temporal graph convolutional-based recurrent network for electric vehicle charging stations demand forecasting in energy market. IEEE Trans. Smart Grid.

Kim, K., Moon, S., Lee, J., Kim, J., 2023. Environmental impacts of transportation policy and vehicle market share predictions. Transp. Res. D 114, 103550.

Kojima, T., Gu, S.S., Reid, M., Matsuo, Y., Iwasawa, Y., 2022. Large language models are zero-shot reasoners. Adv. Neural Inf. Process. Syst. 35, 22199–22213.

Kuang, H., Qu, H., Deng, K., Li, J., 2024. A physics-informed graph learning approach for citywide electric vehicle charging demand prediction and pricing. Appl. Energy 363, 123059.

Lai, Z., Wu, T., Fei, X., Ling, Q., 2024. Bert4st:: Fine-tuning pre-trained large language model for wind power forecasting. Energy Convers. Manage. 307, 118331.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.t., Rocktäschel, T., et al., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. Adv. Neural Inf. Process. Syst. 33, 9459–9474.

Li, Z., Xia, L., Tang, J., Xu, Y., Shi, L., Xia, L., Yin, D., Huang, C., 2024a. Urbangpt: Spatio-temporal large language models. In: ACM SIGKDD Conference on Knowledge Discovery and Data Mining. KDD.

Li, Z., Xia, L., Xu, Y., Huang, C., 2024b. Flashst: A simple and universal prompt-tuning framework for traffic prediction. In: The Forty-First International Conference on Machine Learning. ICML.

Liang, Y., Huang, G., Zhao, Z., 2023. Cross-mode knowledge adaptation for bike sharing demand prediction using domain-adversarial graph neural networks. IEEE Trans. Intell. Transp. Syst..

Liang, Y., Liu, Y., Wang, X., Zhao, Z., 2024a. Exploring large language models for human mobility prediction under public events. Comput. Environ. Urban Syst. 112, 102153.

Liang, Y., Zhao, Z., Ding, F., Tang, Y., He, Z., 2024b. Time-dependent trip generation for bike sharing planning: A multi-task memory-augmented graph neural network. Inf. Fusion 106, 102294.

Liu, X., Xia, Y., Liang, Y., Hu, J., Wang, Y., Bai, L., Huang, C., Liu, Z., Hooi, B., Zimmermann, R., 2023. Largest: A benchmark dataset for large-scale traffic forecasting. In: Advances in Neural Information Processing Systems.

Loshchilov, I., Hutter, F., 2018. Decoupled weight decay regularization. In: International Conference on Learning Representations.

Meng, W., Song, D., Huang, L., Chen, X., Yang, J., Dong, M., Talaat, M., Elkholy, M., 2024. Distributed energy management of electric vehicle charging stations based on hierarchical pricing mechanism and aggregate feasible regions. Energy 291, 130332.

Ni, J., Abrego, G.H., Constant, N., Ma, J., Hall, K., Cer, D., Yang, Y., 2022. Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. In: Findings of the Association for Computational Linguistics: ACL 2022. pp. 1864–1874.

Nichol, A., Schulman, J., 2018. Reptile: a scalable metalearning algorithm. p. 4, arXiv preprint arXiv:1803.02999 2.

Pasha, J., Li, B., Elmi, Z., Fathollahi-Fard, A.M., Lau, Y.y., Roshani, A., Kawasaki, T., Dulebenets, M.A., 2024. Electric vehicle scheduling: State of the art, critical challenges, and future research opportunities. J. Ind. Inf. Integr. 100561.

Qu, H., Kuang, H., Li, J., You, L., 2023. A physics-informed and attention-based graph learning approach for regional electric vehicle charging demand prediction. arXiv preprint arXiv:2309.05259.

Qu, H., Liu, S., Li, J., Zhou, Y., Liu, R., 2022. Adaptation and learning to learn (all): An integrated approach for small-sample parking occupancy prediction. Mathematics 10 (2039).

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J., 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res. 21, 1–67.

Sacco, D., Motta, G.P.A., You, L., Bertolazzo, N., Chen, C., et al., 2022. Smart cities, urban sensing and big data: mining geo-location in social networks. In: AICA 2013. AICA.

Sanh, V., Webson, A., Raffel, C., Bach, S., Sutawika, L., Alyafeai, Z., Chaffin, A., Stiegler, A., Raja, A., Dey, M., et al., 2021. Multitask prompted training enables zero-shot task generalization. In: International Conference on Learning Representations.

Shanahan, M., McDonell, K., Reynolds, L., 2023. Role play with large language models. Nature 623, 493–498.

Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J., 2017. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In: International Conference on Learning Representations.

Song, Y., Hu, X., 2023. Learning-based demand-supply-coupled charging station location problem for electric vehicle demand management. Transp. Res. D 125, 103975.

Song, L., Xue, R., Wang, H., Sun, H., Ge, Y., Shan, Y., et al., 2024. Meta-adapter: An online few-shot learner for vision-language model. Adv. Neural Inf. Process. Syst. 36.

Tang, J., Xia, L., Huang, C., 2023. Explainable spatio-temporal graph neural networks. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. pp. 2432–2441.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat. Methodol. 58, 267–288.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al., 2023. Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

Unterluggauer, T., Rich, J., Andersen, P.B., Hashemi, S., 2022. Electric vehicle charging infrastructure planning for integrated transportation and power distribution networks: A review. eTransportation 12, 100163.

Wald, D., Johnson, K., King, J., Comden, J., Bay, C.J., Chintala, R., Vijayshankar, S., Vaidhynathan, D., 2023. Shifting demand: Reduction in necessary storage capacity through tracking of renewable energy generation. Adv. Appl. Energy 10, 100131.

Wang, S., Chen, A., Wang, P., Zhuge, C., 2023a. Predicting electric vehicle charging demand using a heterogeneous spatio-temporal graph convolutional network. Transp. Res. C 153, 104205.

Wang, S., Chen, A., Wang, P., Zhuge, C., 2023b. Short-term electric vehicle battery swapping demand prediction: Deep learning methods. Transp. Res. D 119, 103746.

Wang, Q., Wang, S., Zhuang, D., Koutsopoulos, H., Zhao, J., 2024. Uncertainty quantification of spatiotemporal travel demand with probabilistic graph neural networks. IEEE Trans. Intell. Transp. Syst..

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V., Zhou, D., et al., 2022. Chain-of-thought prompting elicits reasoning in large language models. Adv. Neural Inf. Process. Syst. 35, 24824–24837.

Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., Weinberger, K., 2019. Simplifying graph convolutional networks. In: International Conference on Machine Learning. PMLR, pp. 6861–6871.

Xue, H., Salim, F.D., 2023. Promptcast: A new prompt-based learning paradigm for time series forecasting. IEEE Trans. Knowl. Data Eng..

Ye, Y., Wang, H., Cui, T., Yang, X., Yang, S., Zhang, M.L., 2023. Identifying generalizable equilibrium pricing strategies for charging service providers in coupled power and transportation networks. Adv. Appl. Energy 12, 100151.

Yi, Z., Liu, X.C., Wei, R., Chen, X., Dai, J., 2022. Electric vehicle charging demand forecasting using deep learning model. J. Intell. Transp. Syst. 26, 690–703.

You, L., Zhu, R., Kwan, M.P., Chen, M., Zhang, F., Yang, B., Wong, M.S., Qin, Z., 2024. Unraveling adaptive changes in electric vehicle charging behavior toward the postpandemic era by federated meta-learning. Innovation 5.

Yu, B., Yin, H., Zhu, Z., 2018. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. IJCAI.

Yuan, C., Xie, Q., Huang, J., Ananiadou, S., 2024. Back to the future: Towards explainable temporal reasoning with large language models. In: Proceedings of the ACM on Web Conference 2024. pp. 1963–1974.

Zhang, H., Chen, J., Yan, J., Song, X., Shibasaki, R., Yan, J., 2021a. Urban power load profiles under ageing transition integrated with future evs charging. Adv. Appl. Energy 1, 100007.

Zhang, W., Liu, H., Wang, F., Xu, T., Xin, H., Dou, D., Xiong, H., 2021b. Intelligent electric vehicle charging recommendation based on multi-agent reinforcement learning. In: Proceedings of the Web Conference 2021. pp. 1856–1867.

Zhang, J., Wang, Z., Miller, E.J., Cui, D., Liu, P., Zhang, Z., 2023. Charging demand prediction in beijing based on real-world electric vehicle data. J. Energy Storage 57, 106294.

Zhang, K., Zhou, F., Wu, L., Xie, N., He, Z., 2024. Semantic understanding and prompt engineering for large-scale traffic data imputation. Inf. Fusion 102, 102038.

Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., et al., 2023. A survey of large language models. arXiv preprint arXiv:2303.18223.