Appendix

A. Simulation Setup

In this study, a real geographic environment is utilized as the simulation model to replicate the charging behavior of 100 taxi drivers over a seven-day period. Each EV starts with a battery energy of 42.755 kWh (total capacity of 50.3 kWh). The 100 taxi drivers are divided into two shifts, day and night, each consisting of 50 drivers. A simulation is conducted using GPT-4o, with a temperature setting of 0.1. The Amap API is used to obtain information such as calculating the distance between two locations and identifying nearby charging stations.

a) Property Generation: Based on whether the taxi driver is assigned to the day or night shift, the LLM uses prompts and Retrieval-Augmented Generation (RAG) to generate personal information, psychological characteristics, EV details, and charging habits.

b) Rule Generation: The rule generation process involves three phases: planning, memory storage, and reflection, which guide the agent's decision-making process. Initially, the LLM creates a preliminary plan based on the daily data of taxi drivers in a city and refines it according to trip-specific needs, resulting in a final plan. Then, decision information is stored in memory: short-term memory holds travel plans and decisions from the past three days, while long-term memory contains data from the last seven days. Finally, at 12:00 each day, the LLM conducts a daily reflection, reviewing whether the day's decisions met user expectations, assessing satisfaction with charging choices, and evaluating their cost-effectiveness.

c) Environmental Perception: The Amap API is used to obtain information such as calculating the distance between two locations and identifying nearby charging stations. Charging station information is matched with the database, assuming that all stations are available.

d) Behavior Simulation: At the end of each event, the vehicle's energy consumption is calculated, and a charging decision is made. Relevant information is provided to the LLM to make decisions that align with the user profile, and these decisions are stored in the database.

TABLE I: Experiment Setup for Simulating EV Charging Behavior

Parameter	**Details**
Environment	Shanghai, China
Duration	7 days
Role	Taxi Drivers
Number of people	100 (50 daytime drivers, 50 nighttime drivers)
Initial SoC	42.755 kWh (Total capacity: 50.3 kWh)
Vehicle type	Electric Vehicles
LLM	GPT-4o
Temperature	0.1
Property	Generates driver profiles.
Perception	Use the Amap API to get environmental information.
Rule	Generates work plans, memory and daily reflections.
Action	Simulates charging decisions.

B. Impact of Parameter in LLMs on Simulator

In order to verify the effectiveness of the proposed method, this research thoroughly examines the model's performance from two key angles: the influence of critical parameters on model accuracy and the model's ability to generalize across various conditions. The primary experimental findings and analyses are summarized as follows.

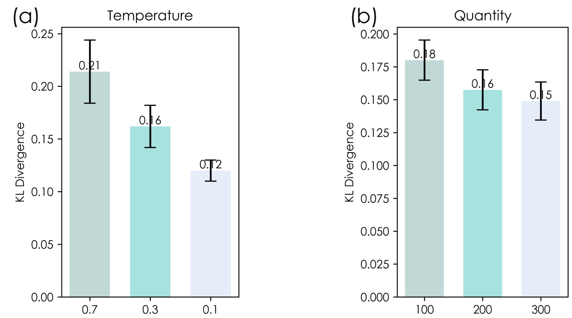

Fig. 8: The impact of different temperatures and varying quantities.

a) Impact of Temperature Parameter in LLMs on Simulator: Fig. 8(a) illustrates how the temperature parameter in a large language model affects the accuracy of the initial SOC in charging behavior simulation. The results indicate that at lower temperatures (e.g., 0.1), the model's generated behavior is more concentrated, showing greater determinism. Conversely, at higher temperatures (e.g., 0.7), the randomness in behavior increases, resulting in a higher KL divergence and a more dispersed behavior distribution. This suggests that the temperature parameter can effectively modulate the balance between randomness and determinism, allowing customization based on application needs. However, an excessively high temperature may lead to overly random behavior, potentially reducing the model's ability to accurately reflect actual user preferences. Consequently, it is essential to carefully calibrate the temperature parameter in practical applications to maintain model stability and ensure a realistic behavior distribution.

b) Influence of Electric Vehicle Quantity on Simulator: Fig. 8(b) shows how different numbers of EVs affect the accuracy of start-time simulation in charging behavior. The results reveal that as the EV count rises (for example, from 20 to 100), the behavior distribution generated by the model becomes more concentrated, and the alignment with the original distribution improves markedly. This relationship is confirmed by the changes in KL divergence: at 20 EVs, the KL divergence stands at 0.21, indicating a significant mismatch with the original distribution; as the EV count increases to 50, the KL divergence decreases to 0.13; and with 100 EVs, it further drops to 0.11. These findings demonstrate that a higher EV count enhances the model's ability to accurately replicate charging behavior distribution, bringing it closer to real-world data.

In more detail, with a smaller EV count, the model finds it challenging to create a stable behavior distribution, resulting in a larger deviation from the original data. However, as the EV count increases, the model's behavior distribution gradually becomes more focused, resulting in a higher level of alignment with the actual distribution. This trend indicates that increasing the EV count not only improves simulation accuracy but also enhances the model's applicability in areas such as demand forecasting and load management.

C. An example of the EV user decision making

In this section, an illustrative example of the decision-making process is presented undertaken by an EV user. As shown in Fig. 9, the scenario demonstrates a structured approach where the user evaluates several factors, including the current SoC, distance to charging stations, charging costs, and other key considerations, to determine the necessity of charging.

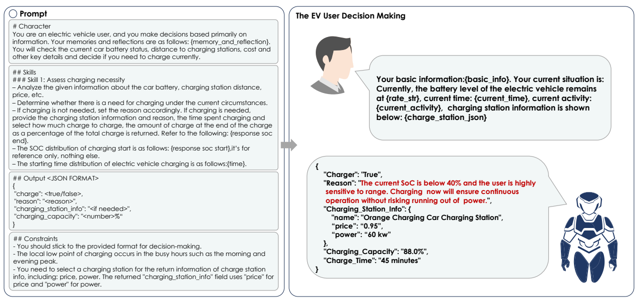

Fig. 9: Case studies on EV user charging decisions. The left side indicates prompt words, while the right side represents the actual dialogue process.

The user's initial situation is characterized by a SoC of rate_str, the current timestamp being current_time, and the user engaged in current_activity. Information regarding available charging stations is summarized in Table 1 as follows: charge_station_json. To assess the need for charging, the user systematically examines data on the remaining SoC, the proximity of charging stations, and the associated costs. If charging is deemed unnecessary, the rationale for this decision is documented. Conversely, if charging is required, the user selects an appropriate charging station based on the provided parameters (e.g., station name, price per kWh, and power capacity) and determines the amount of charge needed as well as the estimated charging time.

For example, in the illustrated scenario, the user's SoC is below 40%, indicating a high risk of running out of power. As a result, the user opts to charge immediately to ensure uninterrupted operation. The selected charging station, "Orange Charging Car Charging Station" offers a competitive price of CNY 0.95 per kWh and a power capacity of 60 kW. The estimated charging requirement is 88.0%, with a projected charging time of 45 minutes. This example highlights the importance of a systematic and data-driven decision-making process for EV users, ensuring optimized charging strategies that balance convenience, cost, and operational efficiency.

D. An example of the EV user charging decision analysis

In this section, an example analysis of the factors influencing EV users' charging decisions is presented. As shown in the Fig. 10 below, this scenario employs a structured approach, systematically analyzing each factor to elucidate their specific mechanisms of influence on user decision-making.

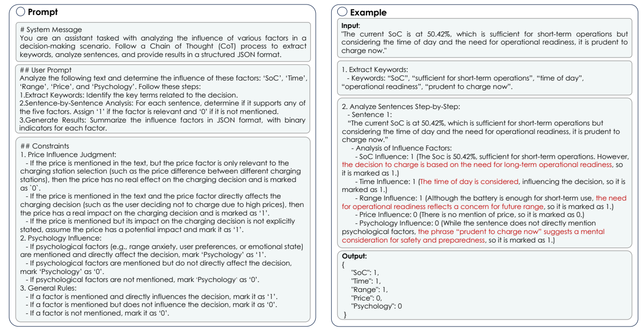

Fig. 10: Case studies on EV user charging decisions analysis. The left side indicates prompt words, while the right side represents the actual dialogue process.

In the illustrated scenario, the rationale behind the user decision to charge is described as follows: The current SoC is at 50.4%, which is sufficient for short-term operations but considering the time of day and the need for operational readiness, it is prudent to charge now.

To ensure the scientific rigor and systematic approach of the analysis, this study adopted a method that integrates keyword extraction with sentence-by-sentence analysis. The method commences with the identification of the principal elements affecting user decision-making using keyword extraction. The sentence-by-sentence analysis captures semantics and examines their interactions with the five components (SoC, Time, Range, Price, Psychology), thereby elucidating the specific influence of each element on user decision-making. Specifically, the first sentence indicated that although the SoC is sufficient for short-term operations, the user opted to charge in view of long-term operational requirements; this shows that the SoC is one of the factors considered in the decision-making process and is thus assigned a value of 1. The second sentence highlighted that time-windows and operational readiness directly prompted the charging decision, so they are also assigned a value of 1. Price and psychological factors are not mentioned in the text, and are therefore assigned a value of 0. This example demonstrates the interaction of many aspects in the charging decision-making process of EV users, affecting user behavior and offering crucial insights for optimizing charging techniques and improving user experience.