ISBN: 978-981-94-4198-3
2025 the 15th International Workshop on Computer Science and Engineering (WCSE 2025)
doi: 10.18178/wcse.2025.06.006

# Immersive Smart Meter Data Analytics: Leveraging eXtended Reality with LSTM and LLMs

Ru-Guan Wang[1], Chu Hsien Tsai[2], Mei Cheng Tseng[2], Rong-Cian Hong[2], HuaRong Syu[2], and Chien-Cheng Chou[3+]

[1] Postdoctoral Researcher at the Information Technology for Disaster Prevention Program, National Central University, Taoyuan, Taiwan

[2] Graduate Research Assistant at the Information Technology for Disaster Prevention Program, National Central University, Taoyuan, Taiwan

[3] Professor at the Information Technology for Disaster Prevention Program, National Central University, Taoyuan, Taiwan

**Abstract.** The rapid advancement of smart grid technologies has led to an exponential growth in smart meter data, creating new opportunities for more accurate energy consumption forecasting and immersive data visualization. This study proposes an integrated framework that combines eXtended Reality (XR), Long Short-Term Memory (LSTM) networks, and Large Language Models (LLMs) to enhance smart meter data analytics. The process begins with the application of LSTM to capture temporal dependencies in historical electricity usage data. Subsequently, the Large Language Models (LLMs) are employed to refine these textual forecasts, offering better predictions and explanations that are easily understandable by end-users. Finally, the enriched insights are presented through an XR environment, enabling users to interact with smart meter analytics in an immersive and intuitive way. By visualizing data trends, predictions, and explanatory narratives in a spatial computing interface, users can explore complex information more effectively. This multi-modal approach facilitates better decision-making for energy management, promotes user engagement, and supports smart city initiatives aiming for sustainable energy consumption. The integration of XR, LSTM, and LLMs technologies demonstrates a promising direction for future research and practical applications in smart energy systems.

**Keywords:** Long Short-Term Memory (LSTM), Large Language Models (LLMs), Extended Reality (XR), Smart Meters

## 1. Introduction

The ongoing digital transformation of the energy sector, driven by the rapid advancement of smart grid technologies, has resulted in an unprecedented increase in the volume and variety of data collected from smart meters. Smart meters, as critical components of modern smart grids, continuously monitor and record detailed information about electricity consumption patterns at both residential and commercial levels. This wealth of data provides new opportunities to enhance the accuracy of energy consumption forecasting and to develop more engaging methods of data visualization. However, effectively extracting actionable insights from this vast, complex, and time-dependent data remains a significant challenge. Traditional statistical models and simple visualization tools are often insufficient for capturing the intricate temporal patterns and contextual factors influencing energy usage, let alone presenting them in a way that is accessible and interactive for end-users.

To address these challenges, this study proposes an integrated analytical framework that synergizes three advanced technologies: eXtended Reality (XR), Long Short-Term Memory (LSTM) neural networks, and Large Language Models (LLMs). Each component of this framework plays a vital role in enhancing the depth and usability of smart meter data analytics. At the core of the system, LSTM networks are employed to model the sequential dependencies inherent in historical electricity consumption data. Unlike traditional time-series models, LSTMs are specifically designed to handle long-range dependencies and nonlinear patterns, making

+ Corresponding author. Tel.: +886-3-4227151 ext.34132; fax: +886-3-4252960.
*E-mail address*: ccchou@ncu.edu.tw.

them highly effective for predicting future consumption trends based on past meter readings. By learning from the historical behavior of energy usage, the LSTM component provides a robust foundation for forecasting demand fluctuations, seasonal changes, and anomalous consumption events.

Building upon the numerical outputs generated by the LSTM model, the next innovative step involves translating these quantitative predictions into descriptive textual formats. This transformation is critical, as raw numerical data can often be opaque or difficult to interpret for non-technical users. By converting numerical streams into textual descriptions, the system facilitates clearer communication of anticipated energy trends. To further refine these textual forecasts, the framework integrates state-of-the-art LLMs. LLMs, known for their advanced natural language understanding and generation capabilities, process the descriptive texts to enhance the clarity, accuracy, and contextual richness of the predictions. In doing so, LLMs not only improve the readability of the information but also provide explanatory narratives that help users understand the underlying factors driving future energy consumption patterns.

The inclusion of LLMs adds an important layer of interpretability and user-friendliness to the framework. Users are not simply presented with abstract predictions but are guided through understandable and context-aware explanations that support informed decision-making. Whether for residential consumers seeking to optimize their household energy use, or for utility companies aiming to balance supply and demand, this enhanced interpretability empowers stakeholders to take proactive measures based on reliable forecasts. Furthermore, LLMs can dynamically adjust the explanations according to the user's preferences or levels of expertise, making the system adaptable to a wide range of audiences.

The final and perhaps most transformative element of this framework is the integration of XR technology for immersive data visualization. XR encompasses a spectrum of immersive technologies, including augmented reality (AR), virtual reality (VR), and mixed reality (MR), all of which enable users to engage with data in three-dimensional, interactive environments. By presenting the enriched analytical insights through XR devices, the framework allows users to intuitively explore consumption trends, predictive scenarios, and explanatory narratives in a spatial computing interface. This immersive approach enhances user engagement by transforming abstract data into tangible experiences. Users can, for instance, virtually navigate through time-series visualizations, manipulate forecast parameters in real-time, or visualize comparative consumption scenarios across different time frames or user profiles.

Such an approach significantly elevates the decision-making process for energy management. The immersive environment fosters a deeper understanding of complex energy dynamics, making it easier to identify patterns, anticipate future needs, and evaluate the potential impact of energy-saving measures. Additionally, this multi-modal system aligns with the broader objectives of smart city initiatives, which prioritize sustainability, efficiency, and citizen engagement. By empowering users with accessible and actionable insights, the framework contributes to promoting responsible energy consumption behaviors and supporting policy goals related to energy conservation and carbon footprint reduction.

In conclusion, the integration of XR, LSTM, and LLM technologies within a unified framework represents a forward-thinking solution to the challenges of smart meter data analytics. This multi-layered approach not only advances the technical capabilities of energy forecasting systems but also transforms the way users interact with and understand their energy consumption. By combining the predictive power of LSTM, the interpretive strengths of LLM, and the immersive experience of XR, this research paves the way for more intelligent, user-centric energy management solutions. The proposed framework holds significant potential for future academic research and practical deployment in smart energy systems, offering a pathway toward more sustainable, efficient, and engaging energy ecosystems.

## 2. Proposed System

Figure 1 shows a CSV file of monthly household electricity usage provided by some buildings in Taiwan equipped with smart meters. The first column, User_ID, is typically presented in email format and is referred to as the household number in this study. The second column, w, is a floating-point number that records the total electricity consumption in watts as measured by the smart meter at that moment. The third column, channelid, represents the circuit number, where 0 stands for the main meter, 1 stands for the television, 3 stands for the refrigerator, 4 stands for the air conditioner, 5 stands for the water dispenser, and 6 stands for the washing machine. The last column, olid, represents the unique value of the household number and the circuit, which can be used to identify a specific smart meter.

Currently, the database contains monthly CSV files from 2021 to 2023, with approximately 5 to 10 million electricity usage records per file. The file sizes range from about 4GB to 9GB and cover data from over 700 households in Taiwan. In these data sets, the records for the main meter (channeled=0) are more complete. Data for other appliances with non-zero channel IDs often have missing entries. Therefore, the following analysis will focus on the main meter data.

```
User_ID,w,reporttime,channelid,olid
taifeng221@gmail.com,0.0,2022-12-01 00:00:00,4,II09000D6F0005B1FA7F
eedddwww21@gmail.com,0.0,2022-12-01 00:01:00,4,II09000D6F0005B1FA72
```

Fig. 1: Power consumption data CSV for one month.

In Taiwan, the standard electrical voltage is 110 V with a frequency of 60 Hz. Taking December 2022 as an example, electricity consumption data for that month were stored in a CSV file with an approximate size of 4,168 MB. These data were subsequently imported into a relational database, such as the open-source PostgreSQL. A total of 51,318,870 records were stored in the database table, comprising measurements not only from the main switch but also from several individual indoor circuits.

It is important to note that the smart meter's measurement intervals for the main switch were not consistent—ranging from approximately 14 to 130 seconds. This variability indicates that assuming a fixed time interval between records is inappropriate. As such, it is essential to retain the raw electricity consumption records in the database, and to generate derived records with suitable time granularity only during the analysis phase. Based on a review of existing literature and the research team's experience analyzing electricity usage data, several key challenges often arise when working with large-scale smart meter data:

- Variability in time intervals: While the power readings themselves are generally accurate (with error rates typically below 0.5%), the intervals between readings may vary from a few seconds to several minutes, potentially complicating analysis.
- Long-term behavior analysis: Accurate long-term usage analysis—such as identifying air conditioner usage patterns—requires data aggregated over longer intervals (e.g., one record per hour). In contrast, for analyzing short-term behaviors such as entertainment device usage, finer intervals (e.g., 15 minutes) are necessary. In this study, the term time interval refers to the actual time difference between two recorded measurements, while analysis time interval denotes the period spanned by the records aggregated for analysis (typically combining multiple actual records). Because actual measurement intervals (e.g., 1-minute readings) often differ from required analysis intervals (e.g., hourly), analysts must determine how many raw records are needed to represent each analysis interval. For instance, if 45 readings between 3:00 PM and 4:00 PM yield an average of 150 W, and the analyst requires at least 40 readings to ensure reliability, the data can be taken as representative of that hour. However, if only 10 readings are available, a conversion method—such as dividing the average by a factor (e.g., 6)—may be required to estimate the appropriate value for the analysis interval. In other words, when insufficient records are present within an analysis interval, the analyst must provide a conversion approach to estimate the average consumption.
- Handling missing data: Some analysis intervals may lack any readings. In such cases, the interval should be represented as having zero consumption during analysis.

The proposed data pre-processing approach involves the following steps: First, the analyst must define the desired analysis time interval (e.g., aggregated readings every 15 minutes). Then, based on the start and end dates of the analysis period, the analyst creates a list of target timestamps at fixed intervals—stored in a reference table, denoted as dt. For example, if the analysis period spans from 00:00:00 on October 1, 2021, to 00:00:00 on November 1, 2021, the second column of the dt table contains the time values at 15-minute intervals, while the first column contains corresponding serial numbers, starting from 1 and incrementing by 1 for each record.

## 3. XR-based Approach

Figure 2 showcases the utilization of HoloLens 2 in conjunction with the developed program, allowing the XR device to display sensed data or electrical power consumption at their respective locations. This research encompasses two distinct systems. The first is referred to as 'Commander,' serving as a demonstrative platform within the 3D virtual realm. The second is an XR system that utilizes the HoloLens 2 device. When users wear this XR equipment in real-world environments, they can perceive virtual world data superimposed onto their physical surroundings via the helmet's display. The primary advantage of employing an XR device lies in the fact that users experience genuine real-world imagery enriched with additional virtual world information. This sets it apart from purely virtual environments, which merely replicate similar scenes without the realism offered by XR.

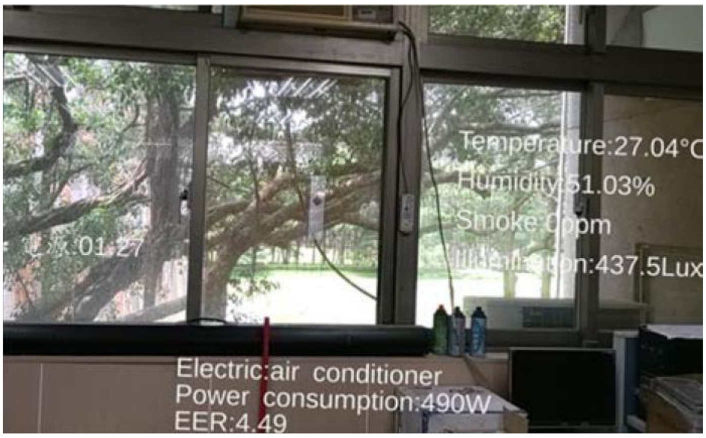


Fig. 2: Real-world IoT sensed data display for the lab environment using HoloLens 2.

Additionally, a home host is deployed, periodically scanning for all WiFi signals within its range. If a signal's name matches a predefined IoT device name, it establishes a connection using the specified password. Each IoT device hosts a web server, which the home host accesses upon connection. This allows the home host to retrieve the sensor output records from the IoT device. Consequently, the home host possesses the coordinate values of each IoT device, along with real-time measurement data. These sensory values can be amalgamated and presented within a 3D scene or XR device through data fusion.

The research team utilized three months of electricity consumption data from five households to develop individual deep learning models and construct corresponding knowledge graphs. These graphs were created using Neo4j's knowledge graph platform. Within the graph structure, the five households are represented as pink nodes, grouped under a shared community denoted by a green node. For the three households (Home02, Home03, and Home04) that exhibited instances of unsafe electricity usage, their respective power consumption data are illustrated as brown nodes. These households recorded maximum power usage exceeding the 4000 W threshold, which is marked by a blue node. In contrast, the remaining two households did not reach this threshold during the three-month observation period.

It is important to highlight that the number of unsafe electricity usage events varied across the households. To address this imbalance in the dataset, the team applied the trained deep learning models to synthetically generate additional unsafe usage records, ensuring that each household was associated with exactly 10 such events. Each event, along with its corresponding electricity consumption data from the preceding 165 minutes, collectively forms the attribute set for a single node within the knowledge graph.

## 4. Conclusions

This study presents a comprehensive framework for analyzing household electricity consumption behavior using smart meter data, combining deep learning models with knowledge graph technology. By focusing on real-world consumption records from households in Taiwan, we addressed key challenges such as inconsistent time intervals between measurements, the need for appropriate analysis time resolutions, and the management of missing data. The proposed preprocessing methodology allows for dynamic time aggregation that aligns with various behavior analysis needs, thereby enhancing the accuracy and usability of the data for downstream applications.

A critical part of this research was the transformation of large-scale CSV-based electricity consumption records into a structured relational database format, followed by the construction of a Neo4j-based knowledge graph. This graph effectively represents relationships between households, their electricity usage patterns, and associated safety risks. Through the application of the Fast Random Projection (FastRP) algorithm, we further demonstrated the capability to identify clusters of households with similar electricity consumption behaviors. This dimension reduction and node embedding approach enables intuitive visualization and facilitates the grouping of households for targeted interventions.

For example, households with recurring unsafe electricity usage patterns—such as exceeding a defined power threshold—can be grouped for the purpose of receiving customized alerts, energy safety education, or device-specific recommendations. The knowledge graph also serves as a foundation for continuously updating and expanding the dataset as more households are monitored or as behavioral models improve through retraining with new data.

In conclusion, this research highlights the value of combining structured data processing, AI-driven modeling, and graph analytics to support smarter, safer, and more responsive energy consumption environments. It lays a foundation for future advancements in personalized electricity management systems and opens opportunities for more intelligent demand-side energy solutions across broader urban and rural settings.

## 5. Acknowledgements

The research is supported by the National Science and Technology Council of Taiwan under Project Nos. MOST 111-2221-E-008-024-MY3, NSTC 113-2221-E-008-028-MY3, and NSTC 114-2811-E-008-003.

## 6. References

[1] M. Wu, S.-L. Dai,, S.-L. Dai, and C. Yang, "Mixed Reality Enhanced User Interactive Path Planning for Omnidirectional Mobile Robot," Applied Sciences, 10(3), 113, 2020.

[2] S. Wu, L. Hou, and G. Zhang, "Integrated Application of BIM and eXtended Reality Technology: A Review, Classification and Outlook," in Proceedings of the 18th International Conference on Computing in Civil and Building Engineering (ICCCBE 2020), São Paulo, Brazil, 2021.

[3] L. Adua, "Reviewing the complexity of energy behavior: Technologies, analytical traditions, and household energy consumption data in the United States," Energy Research & Social Science, 59, 101289, 2020.

[4] L. Cascone, S. Sadiq, S. Ullah, S. Mirjalili, H.U.R. Siddiqui, and M. Umer, "Predicting household electric power consumption using multi-step time series with convolutional LSTM," Big Data Research, 31, 100360, 2023.

[5] S. Yan, K. Li, F. Wang, X. Ge, X. Lu, Z. Mi, S. Chang, "Time frequency feature combination based household characteristic identification approach using smart meter data," IEEE Transactions on Industry Applications, 56(3), 2251-2262, 2020.