ILKOM Jurnal Ilmiah Vol. 18, No. 2, August 2026, pp. 333-350
Accredited 2nd by KEMDIKTISAINTEK No. 10/C/C3/DT.05.00/2025; E-ISSN 2548-7779 | P-ISSN 2087-1716

Research Article Open Access (CC–BY-SA)

IoT-Based Household Energy Monitoring with Preliminary Large Language Model-Assisted Recommendations

Atikah Tri Budi Utami [a,1*]; Hamdan Gani [a,2]; Annisa Dwi Damayanti [b,3]
[a] *Politeknik ATI Makassar, Jl. Sunu No.220, Makassar and 90211, Indonesia*
[b] *Universitas Hasanuddin, Jl. Perintis Kemerdekaan No.KM.10, Makassar and 90245, Indonesia*
[1] *atikah.tribudi@atim.ac.id;* [2] *hamdangani@atim.ac.id;*[3] *annisa.dd@unhas.ac.id*
* *Corresponding author*

Article history: Received December 16, 2025; Revised January 31, 2026; Accepted May 22, 2026; Available online August 10, 2026

Abstract

IoT-based household energy monitoring systems are effective for real-time data acquisition but often provide limited interpretive support for non-technical users. Most existing implementations focus on measurement and visualization, with less attention to converting raw consumption data into actionable recommendations. This study presents a prototype household energy monitoring system that integrates an ESP32-based IoT platform, PZEM-004T sensors, Firebase Realtime Database, and a cloud-hosted Large Language Model (LLM) in a hybrid edge-cloud architecture. Sensor validation was performed by comparing PZEM-004T readings with clamp meter measurements. The results showed low overall error, with voltage measurement achieving an MAE of 0.505 V, RMSE of 0.678 V, and MAPE of 0.231%, while power measurement achieved an MAE of 0.206 W, RMSE of 0.267 W, and MAPE of 0.889%. The LLM module was assessed through scenario-based functional evaluation, where it generated contextual explanations, identified potential energy-use anomalies, and produced quantified energy-saving suggestions. A real-user evaluation involving 50 respondents further indicated generally positive perceptions of the generated recommendations in terms of clarity, relevance, usefulness, trust, and behavioral intention, although the findings remain limited to perceived user responses rather than long-term behavioral outcomes. Overall, the results demonstrate the feasibility of employing an LLM as an interpretive layer in IoT-based household energy monitoring and indicate its potential to improve the accessibility of energy information for non-technical users.

Keywords: Large Language Model (LLM); Internet of Things (IoT); Household Energy Monitoring; ESP32; Smart Home

Introduction

The rapid growth of global electricity demand, together with the need to improve energy efficiency and reduce environmental impact, has made household energy management an important area of technological development [1]–[3]. In residential settings, conventional electricity meters are still commonly used to report aggregate energy consumption. However, such devices generally provide only total usage information and do not reveal appliance-level behavior, inefficient operating patterns, or opportunities for targeted energy-saving actions [4]–[6]. As a result, users often lack the detailed and understandable information needed to make informed decisions about their daily electricity consumption.

To address these limitations, Home Energy Management Systems (HEMS) supported by Internet of Things (IoT) technologies have been widely explored [7]–[9]. IoT-based monitoring systems can collect real-time electrical parameters, support remote supervision, and improve data accessibility through mobile or cloud interfaces [10], [11]. In many implementations, microcontrollers such as the ESP32 are used because they provide integrated wireless communication, low cost, and sufficient processing capability for household monitoring applications [10]–[12]. Likewise, sensing modules such as the PZEM-004T enable real-time measurement of voltage, current, power, and energy, while relay modules allow basic device control [13]. Although these systems are effective for data acquisition and visualization, most still emphasize monitoring functionality rather than interpretive assistance. In practice, users are often required to interpret dashboards and infer their own energy-saving actions, which can limit the usefulness of the system for non-technical households.

http://dx.doi.org/10.33096/ilkom.v18i2.3233.333-350

Recent studies have also explored artificial intelligence methods for energy applications, including machine learning for forecasting, anomaly detection, and consumption pattern analysis [14]–[16]. These approaches are valuable for pattern recognition, but they are often designed for prediction or classification tasks rather than for communicating findings to end users in natural language. At the same time, emerging studies on Large Language Models (LLMs) in IoT and smart-home environments suggest that LLMs can support natural-language interaction, routine generation, and contextual assistance [29]–[32]. However, existing LLM-related studies in smart-home settings generally focus on conversational control, task automation, or conceptual integration. They provide limited evidence on how LLMs can be connected to real household energy-monitoring data and used as an interpretive layer that translates sensor measurements into user-oriented explanations and recommendations.

This limitation defines the main research gap addressed in this study. Existing IoT-based household energy systems are generally strong in sensing and connectivity but weak in interpretation, while LLM-based smart-home studies highlight conversational intelligence but are rarely validated in direct connection with appliance-level energy monitoring data. Therefore, there remains a need for a practical prototype that combines real-time IoT sensing with an LLM-based interpretive module capable of producing contextual and understandable energy-related feedback for household users.

In response to this gap, this study develops a prototype household energy monitoring system that integrates an ESP32 microcontroller, PZEM-004T sensors, Firebase Realtime Database, and a cloud-hosted LLM within a hybrid edge-cloud architecture. The system is designed not only to monitor and control selected household appliances, but also to generate preliminary natural-language explanations and energy-saving suggestions based on sensor data. Unlike conventional IoT dashboards, the proposed approach introduces an additional interpretive layer intended to make energy information more accessible to non-technical users.

The main contributions of this study are as follows. First, it presents the design and implementation of an IoT–LLM prototype for household energy monitoring and recommendation support. Second, it provides quantitative validation of the sensing subsystem using formal error metrics, including MAE, RMSE, and MAPE, by comparing PZEM-004T measurements against clamp meter readings. Third, it reports a preliminary scenario-based functional evaluation of the LLM module in interpreting appliance usage patterns, identifying possible anomalies, and generating contextual recommendations. Through these contributions, the study aims to provide an initial step toward more human-centered household energy monitoring systems that combine data acquisition with interpretable feedback.

Literature Review

The escalating global demand for electrical energy, driven by technological advancements and population growth, presents a critical challenge for urban areas, with the residential sector being a significant contributor to overall energy consumption and greenhouse gas emissions [16], [19]–[21]. This increase in energy use, often compounded by a lack of real-time monitoring and user awareness regarding consumption patterns, frequently results in inefficiencies and unnecessary waste [22]. To address these issues, efficient and sustainable power management systems are imperative, requiring innovative solutions that surpass traditional energy meters.

Existing IoT-based energy monitoring systems offer a foundational step by enabling real-time data collection and remote access to electrical parameters [23]–[26]. Microcontrollers such as the ESP32, highly popular for their built-in wireless connectivity (Wi-Fi and Bluetooth), are widely utilized in these applications to gather data from various sensors like the PZEM-004T (for AC voltage, current, power, and energy), PZEM-017 (for DC parameters), and DHT11 (for temperature and humidity) [27], [28]. This data is then transmitted to cloud platforms, including Google Firebase, Google Spreadsheet, Thing Speak, or Blynk, allowing users to monitor performance remotely via Android applications and receive alerts for abnormal usage. While these systems successfully track consumption and offer basic control capabilities, the process of mere monitoring is often insufficient for achieving complete energy optimization. A critical gap remains in providing sophisticated, personalized, and easily understandable recommendations that can genuinely drive behavioral change and maximize efficiency [29].

This is where the integration of Machine Learning (ML) and LLMs becomes vital. Machine Learning algorithms can process vast amounts of collected IoT data to detect complex consumption patterns, predict future energy usage, and even identify specific appliances contributing to waste [14]–[16]. For instance, the K-Nearest Neighbors (K-

NN) algorithm has demonstrated high accuracy in predicting daily energy consumption for household appliances. Building upon this, LLM offers a transformative solution by enhancing the human-computer interaction aspect. LLMs, such as GPT-3.5 Turbo and GPT-4, possess advanced natural language understanding and generation capabilities, enabling them to interpret complex sensor data and translate these insights into intuitive, natural language-driven recommendations and warnings for users. This enables personalized advice for optimizing energy consumption, streamlining the creation of home automation routines, and educating users on sustainable practices in an engaging and human-like manner [30]. By integrating LLMs, the system can autonomously provide actionable intelligence, potentially even controlling household appliances through actuators to implement optimizations directly [31], [32].

Therefore, this research aims to bridge the identified gap by developing an integrated system that leverages the robust data collection capabilities of ESP32-IoT for energy monitoring, combined with the advanced analytical and conversational capabilities of LLMs for providing intelligent recommendations. This approach offers a powerful synergy: accurate real-time data from the physical world is processed by smart algorithms, and the resulting insights are delivered to users in an easily digestible, actionable format, fostering energy conservation, reducing costs, and ultimately contributing to a more sustainable energy future for urban households.

Method

This study employed a combined approach consisting of prototype development, technical validation, scenario-based functional evaluation, and real-user assessment. The prototype development phase focused on integrating the IoT monitoring subsystem with a cloud-hosted Large Language Model (LLM). The technical validation phase was conducted to assess the accuracy of the sensing subsystem. The LLM module was then evaluated through scenario-based functional testing to examine its ability to interpret appliance-level energy data and generate recommendations. Finally, a real-user evaluation was conducted to assess user perceptions of the generated recommendations.

A. Research Design

The research was designed as a prototype-based applied study. The system was developed to monitor household electricity usage in real time and to provide natural-language recommendations based on monitored sensor data. The study consisted of four stages:

1. Design and implementation of the IoT–LLM prototype,
2. Validation of sensor measurements against a reference instrument,
3. Scenario-based functional evaluation of the LLM module, and
4. Real-user evaluation of recommendation quality using questionnaire-based assessment.

This structure was selected to separately evaluate the technical reliability of the sensing subsystem and the interpretive usefulness of the LLM-generated output.

B. System Architecture

The proposed system adopts a hybrid edge-cloud architecture consisting of four interconnected layers:

1. Hardware Layer: This layer includes an ESP32 microcontroller, two PZEM-004T sensors, and a two-channel relay module. The two monitored appliances in this study were a lamp and a rice cooker. The ESP32 acts as the main controller for data acquisition and relay control.
2. Cloud Layer: Firebase Realtime Database was used as the central data repository. Sensor readings and device status information were transmitted from the ESP32 to Firebase, while control commands and recommendation outputs were also exchanged through the same platform.
3. LLM Analysis Layer: A cloud-hosted LLM module was connected to Firebase to retrieve historical monitoring data, process the data, and generate textual outputs in the form of explanations and energy-saving recommendations.

4. User Interface Layer: A mobile application developed with MIT App Inventor was used to display real-time monitoring data, receive control commands, and present the recommendation outputs generated by the LLM.

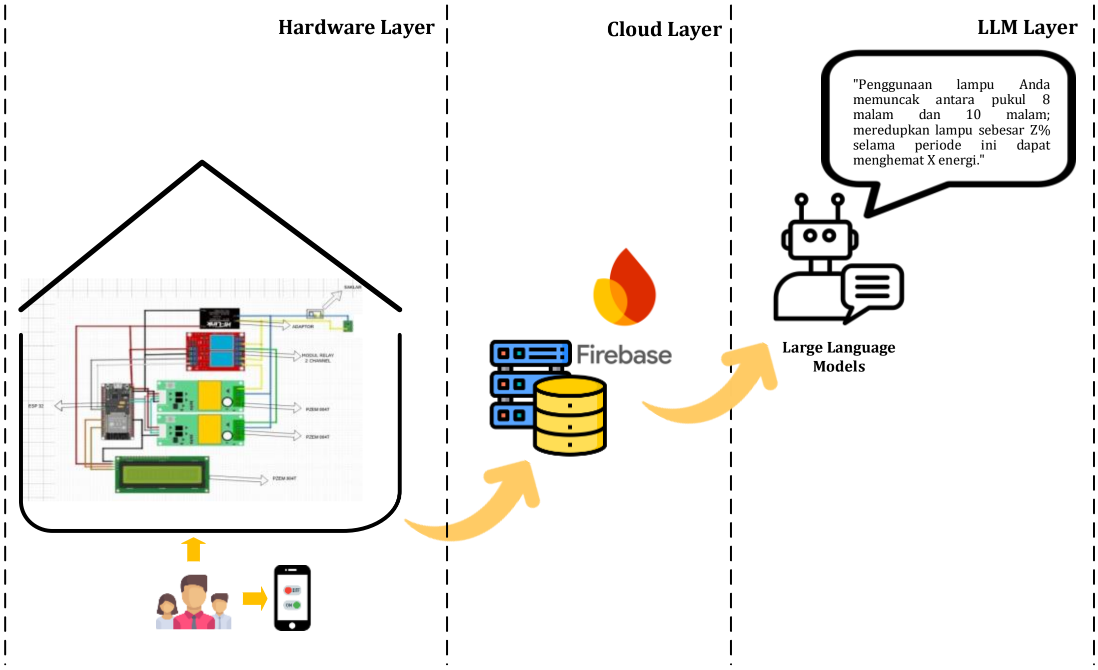


Figure 1. Architecture System

The overall architecture is shown in [Figure 1](#). The detailed interaction between the sensing, cloud, LLM, and user interface components is illustrated in [Figure 2](#).

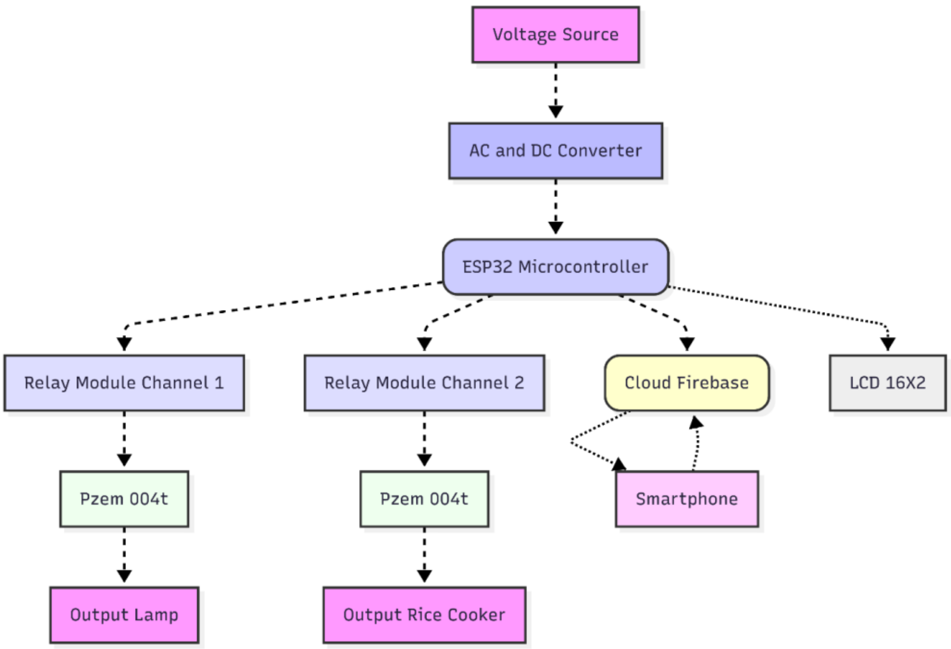


Figure 2. Diagram System IoT

C. Hardware Configuration

The hardware subsystem consisted of the following components:

1. ESP32 microcontroller as the main processing and communication unit
2. Two PZEM-004T sensors for voltage, current, power, and energy measurement
3. A two-channel relay module for appliance switching
4. A 16×2 I2C LCD for local display
5. Adapter and enclosure for power supply and protection

The lamp and rice cooker were selected as representative household appliances because they provide contrasting energy-use characteristics, namely low-power continuous use and higher-power thermal load operation.

Table 1. Hardware and Software Components

Component	Type	Function
ESP32	Hardware	Main controller
PZEM-004T	Sensor	Voltage/current/power measurement
Relay	Actuator	Device switching
LCD	Display	Local monitoring
Firebase	Cloud DB	Data storage
MIT App Inventor	Mobile App	User interface
Gemini 2.5 Pro	LLM	Recommendation engine

The prototype interface and hardware implementation are shown in Figure 3.

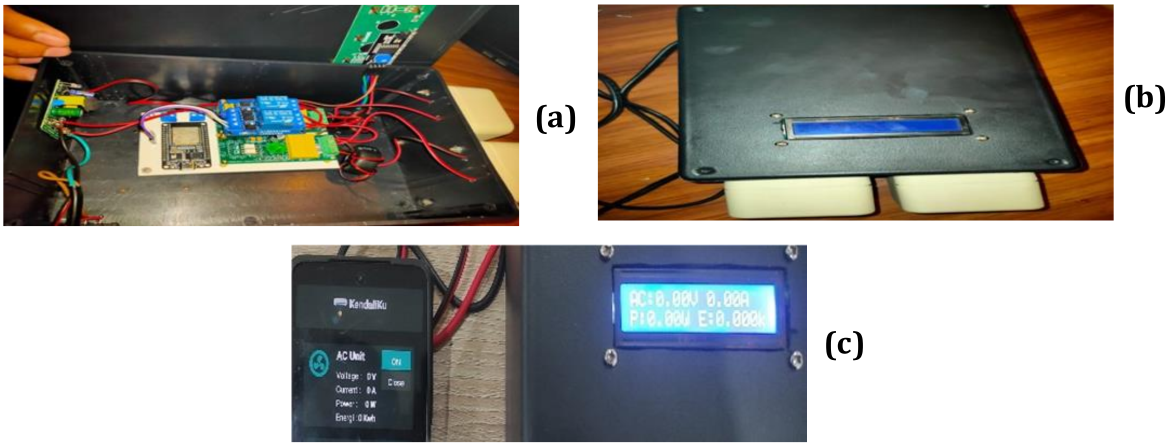


Figure 3. (a) External Display, (b) Internal Display, (c) Mobile Application Display

D. Software and LLM Configuration

The software subsystem included Firebase Realtime Database, MIT App Inventor, and an API-based LLM module. The LLM was not deployed on the ESP32 because of the hardware limitations of microcontroller-based systems. Instead, the system used a cloud-based inference approach.

The LLM configuration used in this study included:

1. Model: Gemini 2.5 Pro (API-based)
2. Deployment mode: cloud-hosted external inference
3. Input source: historical appliance monitoring data retrieved from Firebase Realtime Database

4. Input format: structured prompt containing monitoring context, appliance data summary, and task instruction
5. Output format: natural-language explanation and recommendation text
6. Processing schedule: triggered at predetermined intervals or scenario-based test conditions
7. Prompting style: structured instruction-based prompting with contextual numerical data
8. Language objective: recommendation text understandable for non-technical users

The LLM was designed to function as an interpretive layer rather than as a prediction engine. Its role was to convert appliance-level monitoring data into contextual textual explanations that could be more easily understood by end users.

E. Prompt Design

Structured prompting was used to guide the LLM in interpreting sensor data and generating outputs relevant to household electricity use. Each prompt consisted of three components:

1. Instruction – defines the role of the LLM and the communication objective
2. Contextual Data – provides appliance-level monitoring data or summarized usage patterns
3. Specific Task – requests anomaly detection, pattern interpretation, or recommendation generation

Three prompt scenarios were used in this study:

1. Anomaly detection and causal explanation,
2. Personalized energy-saving recommendation, and
3. Routine consumption pattern recognition with behavioral insight.

The prompts were designed to reduce ambiguity and improve consistency of generated responses by explicitly specifying the analytical objective and user-oriented output style. An example of the structured prompting format used in this study is shown in Figure 4.

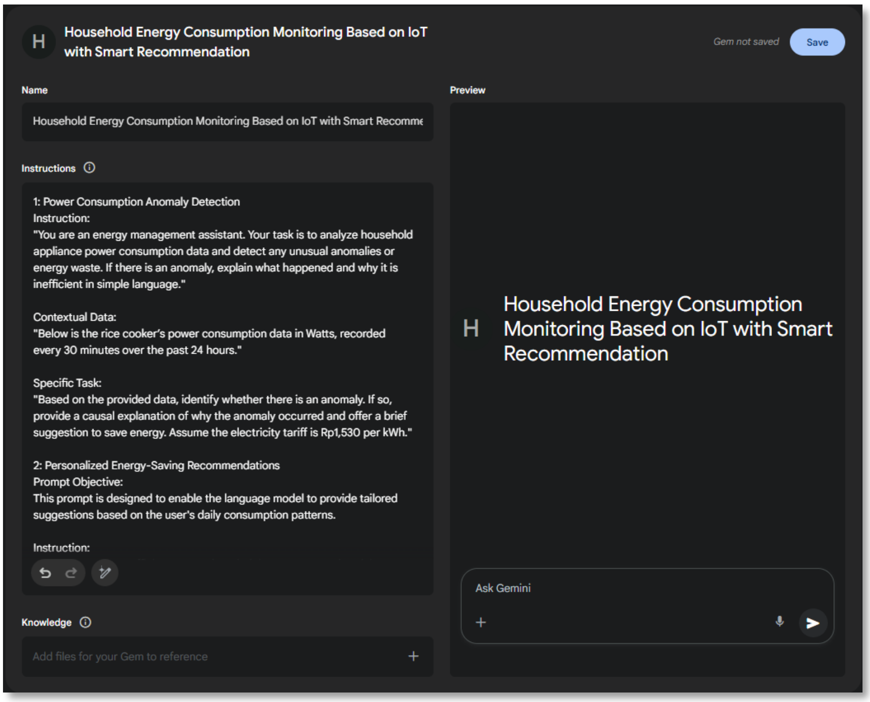


Figure 4. LLM Prompt

F. Sensor Validation Procedure

The technical accuracy of the sensing subsystem was evaluated by comparing PZEM-004T readings against measurements obtained from a clamp meter used as the reference instrument. Measurements were taken repeatedly for the lamp and rice cooker loads across several time intervals. The compared parameters were:

1. Voltage (V)
2. Current (A)
3. Power (W)

To quantify measurement performance, the following formal error metrics were calculated:

1. Mean Absolute Error (MAE)
2. Root Mean Square Error (RMSE)
3. Mean Absolute Percentage Error (MAPE)
4. Maximum Absolute Error

This evaluation was intended to verify whether the prototype produced measurement quality sufficient for downstream interpretation by the LLM module.

G. Scenario-Based Functional Evaluation of the LLM Module

The LLM module was evaluated through scenario-based functional testing. This evaluation was intended to assess whether the model could generate plausible and contextually relevant interpretations from real monitoring data. It was not designed as a benchmark comparison against other LLMs or rule-based systems. Three functional scenarios were used:

1. Anomaly Detection

   The LLM was provided with appliance usage data that reflected potential inefficient operation, particularly prolonged rice cooker power consumption. The purpose was to assess whether the model could identify unusual energy use and explain the inefficiency in natural language.

2. Consumption Pattern Analysis

   The LLM was tested on routine appliance data to determine whether it could identify stable or recurring usage patterns and describe them in an understandable way.

3. Behavioral Insight Generation

   The LLM was asked to interpret the usage pattern in relation to possible household behavior and to produce practical recommendations that aligned with the inferred routine. The outputs were then reviewed in terms of their ability to provide contextual explanations, anomaly indications, and actionable suggestions.

H. Real-User Evaluation

To assess the perceived quality of the LLM-generated recommendations, a real-user evaluation was conducted involving 50 respondents. Unlike the earlier synthetic approach, this phase used actual questionnaire responses collected through an online form.

1. Participants

   The respondents were recruited as non-technical household users or individuals familiar with common household electricity use. The evaluation aimed to capture perceptions from users with varying levels of age, education, and familiarity with household electrical appliances.

2. Instrument

   A structured questionnaire using a 5-point Likert scale was employed, where:

   1 = strongly disagree / very poor

2 = disagree / poor

3 = neutral

4 = agree / good

5 = strongly agree / very good

The questionnaire contained 10 Likert-scale items, grouped into five evaluation dimensions:

1. Clarity
2. Relevance
3. Usefulness
4. Trust
5. Behavioral Intention

In addition to the Likert items, demographic questions were included to describe the respondent profile.

3. Evaluation Procedure

Respondents were presented with LLM-generated recommendation outputs derived from representative monitoring scenarios. After reviewing the outputs, they were asked to rate the recommendations using the questionnaire. The goal was to measure how users perceived the generated outputs in terms of understandability, practical relevance, usefulness for energy-saving action, trustworthiness, and willingness to follow the recommendations.

I. Data Analysis

The data analysis in this study consisted of three parts:

1. Technical Performance Analysis

   Sensor readings were compared with clamp meter measurements using descriptive statistics and formal error metrics, including MAE, RMSE, MAPE, and maximum absolute error.

2. Functional Evaluation Analysis

   The LLM outputs from the three test scenarios were analyzed qualitatively to determine whether the model could produce relevant anomaly explanations, pattern summaries, and contextual recommendations.

3. Real-User Evaluation Analysis

   Questionnaire data from the 50 respondents were analyzed using:

   - Respondent frequency distributions for demographic variables
   - Mean and standard deviation for each evaluation dimension
   - 95% confidence intervals
   - Percentage of positive responses (score ≥ 4)
   - Cronbach's alpha to assess internal consistency of the questionnaire
   - A one-sample test against the neutral midpoint (3) to determine whether the overall perception score was significantly more positive than neutral

This combination of analyses was used to evaluate both the technical feasibility and the perceived usefulness of the proposed system.

Results and Discussion

This section presents the results of the prototype evaluation, consisting of three main parts: (1) technical validation of the IoT sensing subsystem, (2) scenario-based functional evaluation of the LLM module, and (3) real-user evaluation of the LLM-generated recommendations.

A. IoT Performance Testing

This section presents the technical testing results of the IoT subsystem. The purpose of this evaluation is to validate the accuracy and stability of the sensing layer before interpreting the data through the LLM module. In this test, measurements obtained from the PZEM-004T sensor and displayed on the LCD and mobile application were compared with values from a clamp meter as the reference instrument. This comparison is important because the quality of the LLM-generated recommendation depends on the reliability of the sensor data used as input.

Table 2 presents the raw measurement comparison for the lamp and rice cooker loads over repeated observation intervals. The results show that the measured values from the prototype were generally close to those obtained from the reference instrument, indicating stable system operation during the test.

Table 2. IoT Performance Test

TIME/MINUTE	LOAD		Voltage (V)			Amper (I)			Watt (W)		
			V1	V2	V3	I1	I2	I3	P1	P2	P3
10	LAMP	LCD	219	220.1	220	0.04	0.04	0.04	4.8	4.7	4.8
		PK	219	220.1	220	0.04	0.04	0.04	4.8	4.7	4.8
		CM	218	220	220	0.04	0.04	0.04	4.7	4.6	4.6
	RICE COOKER	LCD	218.5	218.9	218	1.34	1.34	1.34	292.8	292.7	292.5
		PK	218.5	218	218.9	1.34	1.34	1.34	292.8	292.7	292.5
		CM	218	218.8	218.4	1.34	1.34	1.34	293	293	293
20	LAMP	LCD	220	219.8	218.9	0.04	0.04	0.04	4.8	4.8	4.7
		PK	220	219.8	218.9	0.04	0.04	0.04	4.8	4	4.7
		CM	220	219	219	0.04	0.04	0.04	4.6	4.8	4.7
	RICE COOKER	LCD	218	217.9	217.5	1.34	1.34	1.34	292.5	291.6	291.8
		PK	218	217.9	217.5	1.34	1.34	1.34	292.5	291.6	291.8
		CM	218.8	218	218	1.34	1.34	1.34	293	291.8	292
30	LAMP	LCD	220	218.9	220.7	0.04	0.04	0.04	4.8	4.7	4.9
		PK	220	218.9	220.7	0.04	0.04	0.04	4.8	4.7	4.9
		CM	220	220	219	0.04	0.04	0.04	4.6	4.7	4.8
	RICE COOKER	LCD	218	217.5	218.2	1.34	1.34	1.34	292.5	291.8	291.4
		PK	218	217.5	218.2	1.34	1.34	1.34	292.5	291.8	291.4
		CM	218.8	218.9	218	1.34	1.34	1.34	293	292	291.8
40	LAMP	LCD	218.9	220.7	219.2	0.04	0.04	0.04	4.9	4.9	4.8
		PK	218.9	219.2	220.7	0.04	0.04	0.04	4.9	4.9	4.8
		CM	219	219	220	0.04	0.04	0.04	4.8	4.8	4.8
	RICE COOKER	LCD	217.5	218.2	218.11	1.34	1.34	1.34	291.8	291.4	291.7
		PK	217.5	218.2	218.11	1.34	1.34	1.34	291.8	291.4	291.7
		CM	218	218.9	218.4	1.34	1.34	1.34	292	291.8	292
50	LAMP	LCD	220.7	219.2	224.2	0.04	0.04	0.04	4.9	4.8	4.8
		PK	220.7	219.2	224.2	0.04	0.04	0.04	4.9	4.8	4.8
		CM	220	219	224	0.04	0.04	0.04	4.8	4.8	4.7

TIME/MINUTE	LOAD		Voltage (V)			Amper (I)			Watt (W)		
	RICE COOKER	LCD	218	218	218	1.34	1.34	1.34	291.4	291.7	291.8
		PK	218	218	218	1.34	1.34	1.34	291.4	291.7	291.8
		CM	218.9	218.4	218.6	1.34	1.34	1.34	291.8	292	292
60	LAMP	LCD	219.2	224.2	224.2	0.04	0.04	0.04	4.8	4.8	4.8
		PK	219.2	224.2	224.2	0.04	0.04	0.04	4.8	4.8	4.8
		CM	219	224	224	0.04	0.04	0.04	4.8	4.7	4.8
	RICE COOKER	LCD	218.11	218	218	1.34	1.34	1.34	291.7	291.8	291.3
		PK	218.11	218	218	1.34	1.34	1.34	291.7	291.8	291.3
		CM	218.4	218.6	218	1.34	1.34	1.34	292	292	292

*LCD: Liquid Crystal Display

*PK: Application Display

*CM: Clamp Meter

The raw observations indicate that the monitoring subsystem produced consistent voltage, current, and power measurements for both tested appliances. For the lamp load, the measured power remained in the range of 4.7–4.9 W, while the rice cooker measurements were consistently within the range of approximately 291–293 W. These values are close to the reference readings from the clamp meter, suggesting that the sensing subsystem is sufficiently stable for household monitoring purposes.

To strengthen the quantitative validation of the sensing subsystem, formal error metrics were calculated by comparing the PZEM-004T readings with clamp meter measurements, as shown in Table 3.

Table 3. Error Metrics of PZEM-004T Measurements Compared with Clamp Meter

Load	Variable	MAE	RMSE	MAPE (%)	Max Absolute Error
Lamp	Voltage (V)	0.505	0.745	0.230	1.700
Lamp	Current (A)	0.000	0.000	0.000	0.000
Lamp	Power (W)	0.078	0.105	1.663	0.200
Rice Cooker	Voltage (V)	0.504	0.604	0.231	1.400
Rice Cooker	Current (A)	0.000	0.000	0.000	0.000
Rice Cooker	Power (W)	0.334	0.362	0.114	0.700
Overall	Voltage (V)	0.505	0.678	0.231	1.700
Overall	Current (A)	0.000	0.000	0.000	0.000
Overall	Power (W)	0.206	0.267	0.889	0.700

For voltage measurement, the overall MAE was 0.505 V, the RMSE was 0.678 V, and the MAPE was 0.231%. For power measurement, the overall MAE was 0.206 W, the RMSE was 0.267 W, and the MAPE was 0.889%. The current values recorded under the tested scenarios were identical to the reference instrument. These results indicate that the proposed IoT subsystem provides sufficiently accurate and stable measurements for household energy monitoring applications. The slightly higher percentage error observed for the lamp power measurement is expected, since small absolute deviations become proportionally larger under low-power loads.

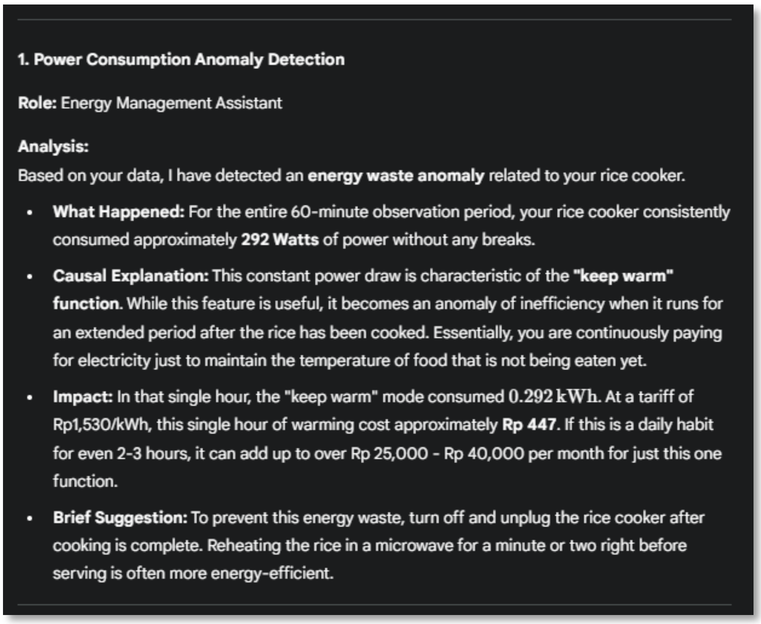
B. Scenario-Based Functional Evaluation of the LLM Module

Following the validation of the sensing subsystem, the LLM module was evaluated through scenario-based functional testing. This evaluation was intended to examine whether the LLM could interpret appliance-level energy data and generate meaningful natural-language feedback, rather than to benchmark the model against other LLMs or analytical baselines. Three functional scenarios were used: anomaly detection, routine consumption pattern analysis, and behavioral interpretation.

1. Detection Anomaly Testing

The first scenario assessed whether the LLM could identify unusual or inefficient energy consumption patterns from monitored appliance data. In this case, the rice cooker data were used to determine whether prolonged high-power operation could be recognized as a potential source of energy waste.

The results showed that the LLM was able to identify a potential anomaly in rice cooker operation. Based on the input pattern, the LLM associated the sustained power draw with an extended keep-warm mode and generated an explanation describing why such usage could be inefficient. It also produced a quantified energy-cost implication and a practical recommendation for reducing unnecessary consumption. This result suggests that the LLM can transform numerical sensor readings into a user-oriented explanation that is easier to understand than raw dashboard values alone.



1. Power Consumption Anomaly Detection

Role: Energy Management Assistant

Analysis:
Based on your data, I have detected an **energy waste anomaly** related to your rice cooker.

- **What Happened:** For the entire 60-minute observation period, your rice cooker consistently consumed approximately **292 Watts** of power without any breaks.
- **Causal Explanation:** This constant power draw is characteristic of the **"keep warm" function**. While this feature is useful, it becomes an anomaly of inefficiency when it runs for an extended period after the rice has been cooked. Essentially, you are continuously paying for electricity just to maintain the temperature of food that is not being eaten yet.
- **Impact:** In that single hour, the "keep warm" mode consumed $0.292\,\text{kWh}$. At a tariff of Rp1,530/kWh, this single hour of warming cost approximately **Rp 447**. If this is a daily habit for even 2-3 hours, it can add up to over Rp 25,000 - Rp 40,000 per month for just this one function.
- **Brief Suggestion:** To prevent this energy waste, turn off and unplug the rice cooker after cooking is complete. Reheating the rice in a microwave for a minute or two right before serving is often more energy-efficient.

Figure 5. LLM Answer for the Detection Anomaly Test

2. Consumption Pattern Analysis Test

The second scenario evaluated whether the LLM could identify stable and routine energy usage patterns from the monitored appliances. This capability is important because anomaly detection depends on the prior recognition of what constitutes normal or regular usage behavior.

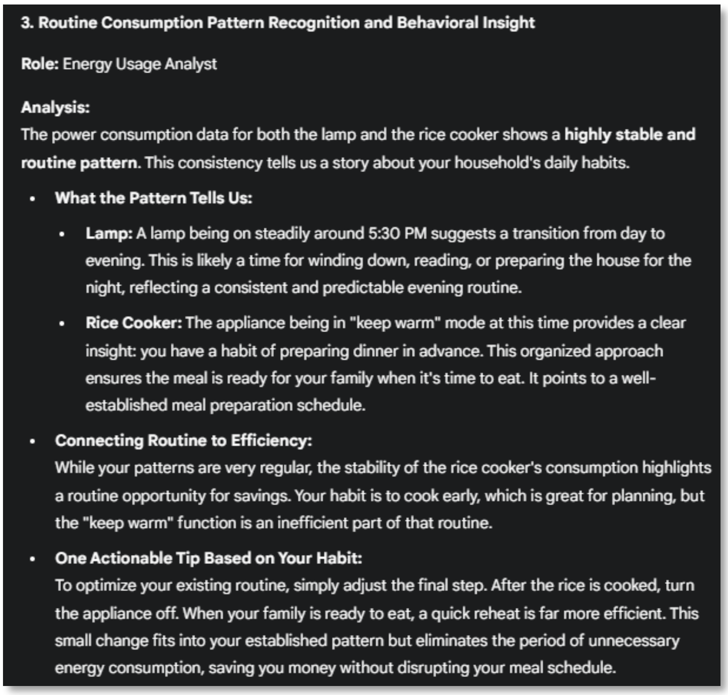
The results indicate that the LLM was able to identify a stable consumption pattern for both the lamp and the rice cooker. Based on the data in **Table 2**, the LLM correctly interpreted the lamp usage as consistently low-power and the rice cooker usage as persistently high-power during the observation period. This shows that the model can summarize repetitive numerical patterns into an understandable description of appliance behavior.

The ability to establish a routine baseline is relevant because it provides contextual support for later identifying deviations or inefficient behavior. In this sense, the LLM functioned as an interpretive layer that translated sensor logs into simplified behavioral summaries.

3. Routine Consumption Pattern Recognition and Behavioural Insight

The third scenario evaluated whether the LLM could go beyond pattern recognition and produce a behavioural interpretation of household electricity use. The goal was to determine whether the output could provide not only description, but also context-sensitive insight.

The LLM successfully translated the monitored lamp and rice cooker data into behavioural narratives. For the lamp, the stable evening usage was interpreted as part of a regular household routine. For the rice cooker, the model associated the operational profile with a recurring meal-preparation habit. Based on this interpretation, the LLM also generated a practical suggestion aimed at improving efficiency while remaining compatible with the existing user routine.



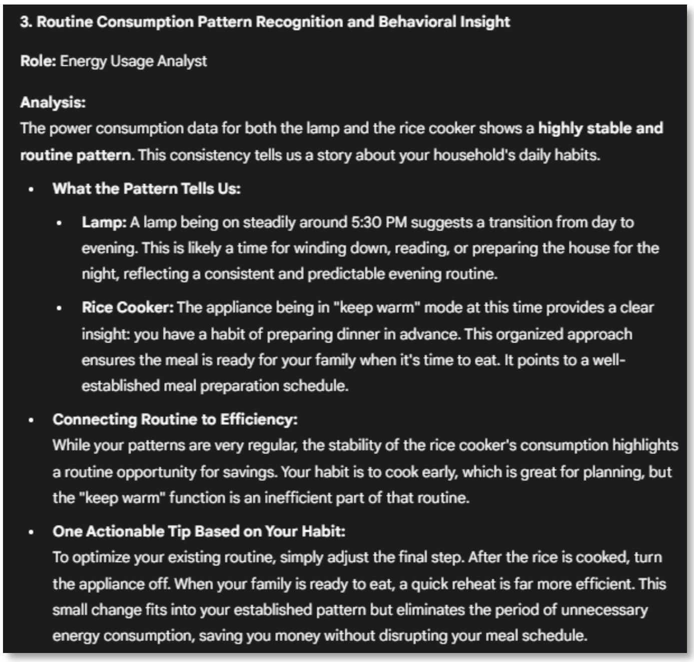
3. Routine Consumption Pattern Recognition and Behavioral Insight

Role: Energy Usage Analyst

Analysis:
The power consumption data for both the lamp and the rice cooker shows a **highly stable and routine pattern**. This consistency tells us a story about your household's daily habits.

- **What the Pattern Tells Us:**
  - **Lamp:** A lamp being on steadily around 5:30 PM suggests a transition from day to evening. This is likely a time for winding down, reading, or preparing the house for the night, reflecting a consistent and predictable evening routine.
  - **Rice Cooker:** The appliance being in "keep warm" mode at this time provides a clear insight: you have a habit of preparing dinner in advance. This organized approach ensures the meal is ready for your family when it's time to eat. It points to a well-established meal preparation schedule.
- **Connecting Routine to Efficiency:**
  While your patterns are very regular, the stability of the rice cooker's consumption highlights a routine opportunity for savings. Your habit is to cook early, which is great for planning, but the "keep warm" function is an inefficient part of that routine.
- **One Actionable Tip Based on Your Habit:**
  To optimize your existing routine, simply adjust the final step. After the rice is cooked, turn the appliance off. When your family is ready to eat, a quick reheat is far more efficient. This small change fits into your established pattern but eliminates the period of unnecessary energy consumption, saving you money without disrupting your meal schedule.

Figure 6. LLM Answer for Consumption Pattern Analysis Test



3. Routine Consumption Pattern Recognition and Behavioral Insight

Role: Energy Usage Analyst

Analysis:
The power consumption data for both the lamp and the rice cooker shows a **highly stable and routine pattern**. This consistency tells us a story about your household's daily habits.

- **What the Pattern Tells Us:**
  - **Lamp:** A lamp being on steadily around 5:30 PM suggests a transition from day to evening. This is likely a time for winding down, reading, or preparing the house for the night, reflecting a consistent and predictable evening routine.
  - **Rice Cooker:** The appliance being in "keep warm" mode at this time provides a clear insight: you have a habit of preparing dinner in advance. This organized approach ensures the meal is ready for your family when it's time to eat. It points to a well-established meal preparation schedule.
- **Connecting Routine to Efficiency:**
  While your patterns are very regular, the stability of the rice cooker's consumption highlights a routine opportunity for savings. Your habit is to cook early, which is great for planning, but the "keep warm" function is an inefficient part of that routine.
- **One Actionable Tip Based on Your Habit:**
  To optimize your existing routine, simply adjust the final step. After the rice is cooked, turn the appliance off. When your family is ready to eat, a quick reheat is far more efficient. This small change fits into your established pattern but eliminates the period of unnecessary energy consumption, saving you money without disrupting your meal schedule.

Figure 7. LLM Answer for the Behavioural Insight Test

These three functional scenarios indicate that the LLM is capable of generating contextualized textual explanations from appliance-level monitoring data. However, the present evaluation should be interpreted as functional feasibility testing rather than as definitive performance benchmarking.

4. Real-User Evaluation of LLM Recommendations

To address reviewer feedback regarding synthetic usability evidence, a real-user evaluation involving 50 respondents was conducted. The objective was to assess how the LLM-generated recommendations were perceived in terms of clarity, relevance, usefulness, trust, and behavioural intention using a 5-point Likert scale questionnaire.

4.1. Respondent Demographics

The demographic characteristics of the respondents are presented in **Table 4**.

Table 4. Respondent Demographics

Variable	Category	Frequency (n=50)	Percentage (%)
Gender	Male	30	60
	Female	20	40
Age Group	<20 years	5	10
	20–29 years	18	36
	30–39 years	17	34
	40–49 years	7	14
	≥50 years	3	6
Education	High School /Vocational	8	16
	Diploma	7	14
	Bachelor's Degree	16	32
	Postgraduate	19	38
Familiar with Household Electrical Appliances	Yes	46	92
	No	4	8
Knowledge of Electricity Bills / Consumption	Very Low	2	4
	Low	5	10
	Moderate	29	58
	High	12	24
	Very High	2	4

Table 4 shows that the respondents were diverse in age, education level, and familiarity with household electrical appliances, indicating suitable representation for a preliminary household-user evaluation. Most respondents reported familiarity with household electrical devices, and the majority had at least moderate awareness of electricity bills and consumption.

4.2. Reliability and Overall Descriptive Statistics

Before interpreting the user perception scores, the internal consistency of the questionnaire was assessed. The results are presented in **Table 5**.

Table 5. Reliability and Descriptive Statistics of Questionnaire

Statistic	Value
Number of Respondents	50
Number of Likert Items	10

Statistic	Value
Scale Range	1–5
Cronbach's Alpha	0.744
Overall Mean Score	4.193
Overall Standard Deviation	0.344
95% Confidence Interval (Lower)	4.095
95% Confidence Interval (Upper)	4.291
One-Sample Test vs Neutral Midpoint (3)	p < 0.001

The questionnaire demonstrated acceptable internal consistency, with Cronbach's alpha = 0.744, indicating that the measurement instrument was sufficiently reliable for the present evaluation. The overall mean score was 4.193 with a standard deviation of 0.344, and the overall 95% confidence interval ranged from 4.095 to 4.291. In addition, the one-sample test against the neutral midpoint of 3 showed $p < 0.001$, indicating that the overall perception of the LLM-generated recommendations was significantly more positive than neutral.

4.3. Mean Scores by Evaluation Dimension

The detailed results by evaluation dimension are presented in **Table 6**.

Table 6. Mean Scores by Evaluation Dimension

Dimension	Mean	SD	95% CI Lower	95% CI Upper	Positive Responses (≥4) %
Clarity	3.94	0.60	3.78	4.10	66
Relevance	4.25	0.50	4.11	4.39	82
Usefulness	4.34	0.60	4.18	4.50	82
Trust	4.14	0.50	4.00	4.28	84
Behavioral Intention	4.25	0.60	4.07	4.43	86

The highest score was observed for usefulness (M = 4.34), followed by relevance and behavioural intention (M = 4.25). Trust also showed a positive mean score (M = 4.14), while clarity received the lowest mean score (M = 3.94), although it remained above the neutral midpoint. These results indicate that respondents generally perceived the recommendations as practically helpful, relevant to the presented situations, and potentially influential in shaping energy-saving behaviour. At the same time, the clarity dimension suggests that some outputs could still be improved through simpler or more direct wording.

A visual summary of the mean scores is shown in Figure 8. The percentage of positive responses (score ≥ 4) further supports this trend. Positive responses reached 82% for relevance, 82% for usefulness, 84% for trust, and 86% for behavioural intention, while clarity reached 66%. This pattern suggests that users were generally willing to accept and potentially follow the recommendations, even though linguistic simplification may further improve comprehensibility.

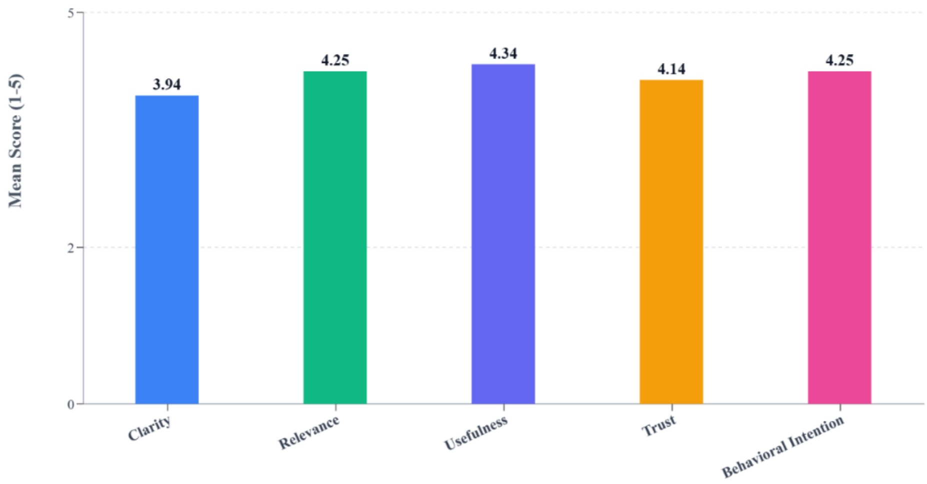


Figure 8. Mean Scores by Evaluation Dimension

Figure 8 visually confirms that usefulness obtained the highest mean score, while clarity received the lowest score among the five evaluated dimensions.

Discussion

The results of this study indicate that the proposed prototype successfully combines real-time IoT sensing with an LLM-based interpretive layer for household energy monitoring. From the hardware perspective, the low error values obtained from the PZEM-004T validation confirm that the sensing subsystem provides data of sufficient quality for downstream analysis. This is important because the usefulness of the recommendation layer depends directly on the reliability of the monitoring layer.

From the LLM perspective, the scenario-based functional tests show that the model can generate contextual explanations, identify potentially inefficient appliance usage, and produce understandable recommendations from monitored sensor data. This distinguishes the proposed system from conventional IoT dashboards that primarily focus on display and monitoring. Instead of presenting only numerical values, the prototype adds an interpretive layer that can help translate raw data into user-oriented feedback.

The real-user evaluation further strengthens this contribution. The results show that respondents evaluated the recommendations positively across all five dimensions, with the highest scores observed for usefulness, relevance, and behavioural intention. This suggests that the generated outputs were not only understandable but also perceived as practically meaningful and potentially actionable. The acceptable Cronbach's alpha also indicates that the questionnaire responses were internally consistent.

At the same time, the clarity dimension received the lowest mean score among the five dimensions. Although still positive, this result implies that the wording of some recommendations may need to be simplified or made more concise for broader household audiences. This observation is valuable because it suggests that improving the language style of the recommendation layer may further enhance the accessibility of the system, particularly for users with limited technical or energy-literacy backgrounds.

Overall, the findings support the feasibility of employing an LLM as an interpretive layer in household energy monitoring systems. However, several limitations remain. First, the LLM evaluation in this study was scenario-based and did not include benchmark comparison against rule-based or non-LLM recommendation systems. Second, the prototype was tested on a limited number of household appliances, which restricts the generalizability of the findings. Third, although the real-user evaluation provides stronger evidence than the earlier synthetic approach, it still measures perceived usefulness and acceptance rather than actual long-term behavioural change or energy savings. Future work should therefore include broader appliance coverage, comparative baselines, longitudinal household deployment, and measurement of actual behavioural impact.

Conclusion

This study presented a prototype household energy monitoring system that integrates an ESP32-based IoT platform, PZEM-004T sensors, Firebase Realtime Database, and a cloud-hosted Large Language Model (LLM) within a hybrid edge-cloud architecture. The results show that the proposed system is feasible for combining real-time household energy monitoring with natural-language interpretation of appliance-level data.

The technical validation demonstrated that the sensing subsystem provided sufficiently accurate and stable measurements for household monitoring applications. Comparison against clamp meter readings showed low overall error, with voltage measurement achieving an MAE of 0.505 V, RMSE of 0.678 V, and MAPE of 0.231%, while power measurement achieved an MAE of 0.206 W, RMSE of 0.267 W, and MAPE of 0.889%. These findings indicate that the IoT layer can provide reliable input data for higher-level analysis.

The scenario-based functional evaluation further showed that the LLM module was able to generate contextual explanations, identify potentially inefficient appliance usage, and produce quantified energy-saving suggestions. In addition, the real-user evaluation involving 50 respondents indicated generally positive perceptions of the generated recommendations across clarity, relevance, usefulness, trust, and behavioral intention. The highest mean score was observed for usefulness, while clarity received the lowest, suggesting that the recommendations were considered practically helpful but could still benefit from simpler wording.

Overall, the findings support the feasibility of employing an LLM as an interpretive layer in IoT-based household energy monitoring systems, particularly for improving the accessibility of energy information for non-technical users. However, this study has several limitations. The LLM evaluation was scenario-based and did not include comparison with rule-based or non-LLM baselines. The prototype was tested on a limited number of household appliances, and the user study evaluated perceived usefulness rather than long-term behavioral change or actual energy savings. Future work should therefore include comparative baseline testing, broader appliance coverage, longitudinal household deployment, latency and cost analysis, and direct measurement of behavioral and energy-saving outcomes.

Acknowledgement

The author gratefully acknowledges the research participants and all individuals who supported the completion of this study.

References

[1] L. Gorina, E. Korneeva, O. Kovaleva, and W. Strielkowski, “Energy-saving technologies and energy efficiency in the post- COVID era,” *Sustainable Development*, vol. 32, no. 5, pp. 5294–5310, Oct. 2024, doi: 10.1002/sd.2978.

[2] B. Gajdzik, R. Wolniak, R. Nagaj, B. Žuromskaitė-Nagaj, and W. W. Grebski, “The Influence of the Global Energy Crisis on Energy Efficiency: A Comprehensive Analysis,” *Energies*, vol. 17, no. 4, p. 947, Feb. 2024, doi: 10.3390/en17040947.

[3] A. Elkhatat and S. Al-Muhtaseb, “Climate Change and Energy Security: A Comparative Analysis of the Role of Energy Policies in Advancing Environmental Sustainability,” *Energies*, vol. 17, no. 13, p. 3179, Jun. 2024, doi: 10.3390/en17133179.

[4] M. S. Bakare, A. Abdulkarim, M. Zeeshan, and A. N. Shuaibu, “A comprehensive overview on demand side energy management towards smart grids: challenges, solutions, and future direction,” *Energy Informatics*, vol. 6, no. 1, p. 4, Mar. 2023, doi: 10.1186/s42162-023-00262-7.

[5] A. Zaboli, S. R. Kasimalla, K. Park, Y. Hong, and J. Hong, “A Comprehensive Review of Behind-the-Meter Distributed Energy Resources Load Forecasting: Models, Challenges, and Emerging Technologies,” *Energies*, vol. 17, no. 11, p. 2534, May 2024, doi: 10.3390/en17112534.

[6] A. Othman and N. H. Zakaria, “Energy Meter based Wireless Monitoring System using Blynk Application via smartphone,” in *IEEE International Conference on Artificial Intelligence in Engineering and Technology, IICAIET 2020*, 2020, pp. 1–5, doi: 10.1109/IICAIET49801.2020.9257827.

[7] A. Asit Kumar Majhi and S. Mohanty, "A Comprehensive Review on Internet of Things Applications in Power Systems," *IEEE Internet of Things Journal*, vol. 11, no. 21, pp. 34896–34923, Nov. 2024, doi: 10.1109/JIOT.2024.3447241.

[8] P. Franco, J. M. Martínez, Y.-C. Kim, and M. A. Ahmed, "A Cyber-Physical Approach for Residential Energy Management: Current State and Future Directions," *Sustainability*, vol. 14, no. 8, p. 4639, Apr. 2022, doi: 10.3390/su14084639.

[9] M. S. Aliero, K. N. Qureshi, M. F. Pasha, and G. Jeon, "Smart Home Energy Management Systems in Internet of Things networks for green cities demands and services," *Environmental Technology & Innovation*, vol. 22, p. 101443, May 2021, doi: 10.1016/j.eti.2021.101443.

[10] H. J. El-Khozondar *et al.*, "A smart energy monitoring system using ESP32 microcontroller," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 9, p. 100666, Sep. 2024, doi: 10.1016/j.prime.2024.100666.

[11] A. L. H. P S, M. Shafiulla, S. Mohammed Naveed, S. Ahmed, S. Mohammed Nawaz, and U. Kumar, "Home Automation Using Wi-Fi: ESP32-Based System for Remote Control and Environmental Monitoring," in *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, Apr. 2024, pp. 1–7, doi: 10.1109/ICDCECE60827.2024.10549726.

[12] F. Condon, J. M. Martínez, A. M. Eltamaly, Y.-C. Kim, and M. A. Ahmed, "Design and Implementation of a Cloud-IoT-Based Home Energy Management System," *Sensors*, vol. 23, no. 1, p. 176, Dec. 2022, doi: 10.3390/s23010176.

[13] M. S. Hossain and M. M. Hassan, "Smart Energy Meter with Real Time Monitoring System," in *2024 13th International Conference on Electrical and Computer Engineering (ICECE)*, Dec. 2024, pp. 761–766, doi: 10.1109/ICECE64886.2024.11024970.

[14] A. Diro, N. Chilamkurti, V.-D. Nguyen, and W. Heyne, "A Comprehensive Study of Anomaly Detection Schemes in IoT Networks Using Machine Learning Algorithms," *Sensors*, vol. 21, no. 24, p. 8320, Dec. 2021, doi: 10.3390/s21248320.

[15] D. Arsene, A. Predescu, B. Pahonțu, C. G. Chiru, E.-S. Apostol, and C.-O. Truică, "Advanced Strategies for Monitoring Water Consumption Patterns in Households Based on IoT and Machine Learning," *Water*, vol. 14, no. 14, p. 2187, Jul. 2022, doi: 10.3390/w14142187.

[16] J. O. Ojadi, C. S. Odionu, E. C. Onukwulu, and O. A. Owulade, "AI-Enabled Smart Grid Systems for Energy Efficiency and Carbon Footprint Reduction in Urban Energy Networks," *International Journal of Multidisciplinary Research and Growth Evaluation.*, vol. 5, no. 1, pp. 1549–1566, 2024, doi: 10.54660/.IJMRGE.2024.5.1.1549-1566.

[17] O. Arshi, A. Rai, G. Gupta, J. K. Pandey, and S. Mondal, "IoT in energy: a comprehensive review of technologies, applications, and future directions," *Peer-to-Peer Networking and Applications*, vol. 17, no. 5, pp. 2830–2869, Sep. 2024, doi: 10.1007/s12083-024-01725-8.

[18] G. Tyagi, "End-User and Human-Centric IoT, Including IoT Multimedia, Societal Impacts and Sustainable Development," in *The Next Generation Innovation in IoT and Cloud Computing with Applications*, Boca Raton: CRC Press, 2024, pp. 90–109.

[19] D. Lerede and L. Savoldi, "Might future electricity generation suffice to meet the global demand?," *Energy Strategy Reviews*, vol. 47, p. 101080, May 2023, doi: 10.1016/j.esr.2023.101080.

[20] J. Wang, S. Ghosh, O. A. Olayinka, B. Doğan, M. I. Shah, and K. Zhong, "Achieving energy security amidst the world uncertainty in newly industrialized economies: The role of technological advancement," *Energy*, vol. 261, p. 125265, Dec. 2022, doi: 10.1016/j.energy.2022.125265.

[21] M. Farghali *et al.*, "Strategies to save energy in the context of the energy crisis: a review," *Environmental Chemistry Letters*, vol. 21, no. 4, pp. 2003–2039, Aug. 2023, doi: 10.1007/s10311-023-01591-5.

[22] J. Dahihande, A. Jaiswal, A. A. Pagar, A. Thakare, M. Eirinaki, and I. Varlamis, "Reducing energy waste in

households through real-time recommendations," in *Fourteenth ACM Conference on Recommender Systems*, Sep. 2020, pp. 545–550, doi: 10.1145/3383313.3412212.

[23] P. Pal *et al.*, "IoT-Based Real Time Energy Management of Virtual Power Plant Using PLC for Transactive Energy Framework," *IEEE Access*, vol. 9, pp. 97643–97660, 2021, doi: 10.1109/ACCESS.2021.3093111.

[24] C. K. Rao, S. K. Sahoo, and F. F. Yanine, "A comprehensive review of smart energy management systems for photovoltaic power generation utilizing the internet of things," *Unconventional Resources*, vol. 7, p. 100197, Jul. 2025, doi: 10.1016/j.uncres.2025.100197.

[25] W. Yaïci, K. Krishnamurthy, E. Entchev, and M. Longo, "Recent Advances in Internet of Things (IoT) Infrastructures for Building Energy Systems: A Review," *Sensors*, vol. 21, no. 6, p. 2152, Mar. 2021, doi: 10.3390/s21062152.

[26] S. M. A. A. Abir, A. Anwar, J. Choi, and A. S. M. Kayes, "Iot-enabled smart energy grid: Applications and challenges," *IEEE Access*, vol. 9, pp. 50961–50981, 2021, doi: 10.1109/ACCESS.2021.3067331.

[27] D. S. Naga Malleswara Rao, A. Prathyusha, V. L. Priyanka, and M. Mrudhula, "IoT-Enabled Smart Energy Meter for Improved Billing and Usage Monitoring," in *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*, Jun. 2023, pp. 1019–1023, doi: 10.1109/ICPCSN58827.2023.00173.

[28] I. Plauska, A. Liutkevičius, and A. Janavičiūtė, "Performance Evaluation of C/C++, MicroPython, Rust and TinyGo Programming Languages on ESP32 Microcontroller," *Electronics*, vol. 12, no. 1, p. 143, Dec. 2022, doi: 10.3390/electronics12010143.

[29] M. Zong, A. Hekmati, M. Guastalla, Y. Li, and B. Krishnamachari, "Integrating large language models with internet of things: applications," *Discover Internet of Things*, vol. 5, no. 1, p. 2, Jan. 2025, doi: 10.1007/s43926-024-00083-4.

[30] M. Giudici, L. Padalino, G. Paolino, I. Paratici, A. I. Pascu, and F. Garzotto, "Designing Home Automation Routines Using an LLM-Based Chatbot," *Designs*, vol. 8, no. 3, p. 43, May 2024, doi: 10.3390/designs8030043.

[31] X. Luo, D. Liu, F. Dang, and H. Luo, "Integration of LLMs and the Physical World: Research and Application," in *ACM International Conference Proceeding Series*, 2024, pp. 1–5, doi: 10.1145/3674399.3674402.

[32] D. Rivkin *et al.*, "AIoT Smart Home via Autonomous LLM Agents," *IEEE Internet of Things Journal*, vol. 12, no. 3, pp. 2458–2472, Feb. 2025, doi: 10.1109/JIOT.2024.3471904.