# User-Intent–Driven Dynamic Residential Energy Flexibility Management via Enhanced RAG

Xiaotian Zhou and Hao Liang
*Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada*
Emails: xzhou9@ualberta.ca, hao2@ualberta.ca

***Abstract*—Residential Energy Flexibility Management (REFM) has become a key approach to enhancing household energy efficiency and autonomy. In practice, users often issue real-time, ad-hoc requests for appliance usage rather than pre-scheduling their consumption, which necessitates dynamic re-optimization of appliance scheduling. However, residential users generally lack the expertise to formulate or solve such optimization problems and instead express their needs through natural language. To bridge this gap, the REFM problem is formulated as a user-intent–driven reward Markov Decision Process (MDP). Subsequently, a Retrieval-Augmented Generation (RAG)-based dynamic REFM framework is proposed as an interactive interface between users and the underlying energy management system. To improve the accuracy of user intent parsing, an enhanced RAG algorithm is developed for REFM based on the generative reinforcement policy optimization. A case study based on real-world residential demand data is presented to evaluate the proposed scheme, which employs DeepSeek-R1-Distill-Qwen-7B as the base model. The results demonstrate its effectiveness in accurately parsing user intent and performing dynamic REFM in compliance with user requirements.**

***Index Terms*—Residential energy flexibility management, retrieval-augmented generation, user-intent optimization.**

## I. Introduction

The rapid proliferation of distributed energy resources has transformed households from passive electricity consumers into active participants in the power systems. Residential Energy Flexibility Management (REFM) aims to harness this flexibility by optimally scheduling controllable appliances, storage, and generation to reduce costs and enhance user comfort. According to the U.S. Energy Information Administration, studies suggest that residential demand flexibility could provide up to 10–15% of peak-load reduction potential [1]. However, the effective utilization of household flexibility remains limited due to uncertainties in occupant behavior and the lack of adaptive, user-centric control mechanisms [2].

Traditional REFM approaches formulate energy scheduling as deterministic or stochastic optimization problems, such as rule-based strategy [3], model predictive control (MPC) [4], and reinforcement learning-based demand response [5]. These methods assume fixed user preferences and pre-defined objectives, thereby optimizing load shifting or cost minimization over a given time horizon. However, in practice, residential users frequently issue ad-hoc demand requests for appliance usage or adjustments in real time (e.g., “charge the EV earlier” or “reduce cost tonight”), which can invalidate prior schedules. This mismatch between static optimization frameworks and real-world user behavior highlights the need for REFM policies that can be adjusted dynamically in response to changing user intents.

Another major challenge of REFM lies in bridging human intent and machine optimization. Conventional schemes typically incorporate user comfort as a quantitative indicator to guide the optimization process, implicitly assuming that user preferences can be accurately expressed in numerical form. In practice, however, residential users often cannot quantify their demand with a single factor, as their needs are context-dependent and naturally conveyed through descriptive or real-time requests rather than pre-defined metrics [6]. Consequently, the system must accurately interpret such intents, translate them into actionable constraints or reward functions, and ensure consistency with ongoing optimization.

In addition, incorporating natural-language interaction introduces the risk of interpretation errors and uncertainty propagation within the REFM process. Large language models (LLMs) [7] offer a potential solution by enhancing natural-language understanding and enabling more flexible user interactions. A few recent studies have explored LLM-based user-interactive energy management frameworks that utilize zero- or few-shot learning to capture user schedules and preferences [8], [9]. However, these works primarily focus on intent formatting rather than optimization, and zero- or few-shot learning alone may be insufficient [10] to address the complex user-intent parsing objectives considered in this work. Retrieval-Augmented Generation (RAG) models [11] further improve language reliability by grounding LLM outputs in relevant external knowledge bases. Yet, existing RAG frameworks are primarily designed for static question answering and have not been tailored to dynamic energy management contexts, where incorrect retrieval or generation may lead to suboptimal control actions. Reinforcement learning provides a promising avenue for refining RAG behavior through reward-driven adaptation [12], [13], but integrating such mechanisms into REFM remains an open challenge.

To address these aforementioned challenges, this paper proposes a user-intent-driven dynamic REFM framework via enhanced RAG. By integrating natural language understanding, retrieval-augmented reasoning, and reinforcement learning, the proposed framework enables REFM policies to be dynamically adapted in response to changing user intents expressed through natural language. The main contributions of this paper are:

1) A residential energy flexibility management problem is formulated as an MDP with a user-intent–driven reward, where user requests can dynamically alter the reward function and state transition dynamics of the system;
2) A RAG-based dynamic REFM framework is proposed to incorporate user intents into home energy optimization through natural language interaction;
3) The accuracy of RAG in interpreting user energy-management intents is continually enhanced via a group relative policy optimization (GRPO) approach, which is

979-8-3315-8138-1

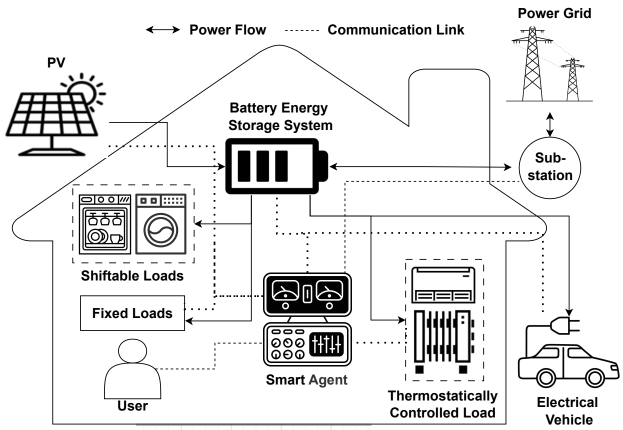

Fig. 1. An illustration of the REFM system.

specifically tailored based on the feedback from REFM applications.

The remainder of this paper is organized as follows. In Section II, the system model is described. Section III presents the dynamic REFM problem formulation. In Section IV, the RAG-based REFM problem solution is presented. The case study results are discussed in Section V, followed by the conclusion in Section VI.

## II. System Model

As illustrated in Fig. 1, the proposed residential energy flexibility management framework is built upon a typical smart home energy system that integrates conventional loads and flexible energy resources, including PV generation, a battery energy storage system (BESS), an electric vehicle (EV), thermostatically controlled loads (TCLs), and shiftable loads. All flexible devices are coordinated by a smart agent that manages their power flows and executes residential energy flexibility management strategies. When a user issues a natural-language request, the smart agent interprets the intent and dynamically re-plans the operational schedule to accommodate the request. System states are monitored through smart meters, which collect real-time operational status of appliances, while also integrating external signals, including weather conditions and electricity tariffs. The interaction among the user, the appliances, and the smart agent is supported by a local area network. In the following, the mathematical models of the flexible energy resources are presented.

### A. BESS Model

Within the REFM system, the BESS provides energy flexibility through controllable charging and discharging operations. Its charging and discharging behaviors can be scheduled in response to real-time electricity price signals, enabling economically optimal operation. The corresponding battery operation and state-of-charge (SoC) dynamics are given by

$$\begin{aligned}\mathrm{SoC}_{t+1} = \mathrm{SoC}_t + \eta^B \max\big(0, -P_t^B\big)\,\Delta t \\ - \frac{1}{\eta^B}\max\big(0, P_t^B\big)\,\Delta t, \quad \forall t\end{aligned} \tag{1}$$

$$\mathrm{SoC}^{\min} \leq \ \mathrm{SoC}_t \leq \mathrm{SoC}^{\max}, \quad \forall t \tag{2}$$

where $P_t^B$ represents the active power output of the BESS, which is positive during discharge and negative during charge, and is constrained by its rated power limits $P^{B,\min} \leq P_t^B \leq P^{B,\max}$. The parameter $\eta^B$ denotes the charging and discharging efficiency. The symbols $\mathrm{SoC}^{\min}$ and $\mathrm{SoC}^{\max}$ specify the allowable bounds of the state of charge. The battery degradation cost is $C_t^d = \tau(c_1 \mathrm{SoC}_t^d e^{c_2 \mathrm{SoC}_t^a} + c_3 e^{c_4 \mathrm{SoC}_t^d})\Delta Ah_t(P_t^B)$, where $\tau$, $c_1$, $c_2$, $c_3$, $c_4$, $SoC_t^a$, and $SoC_t^d$ are the replacement cost for the lost percentage of battery active capacity, the curve-fitting constants, and the SoC average and deviation, respectively, which are presented in [14].

### B. EV Model

The EV contributes to the REFM system through its controllable charging behavior, which can be scheduled within a user-specified plug-in window. The EV can strategically shift its charging load to low-tariff periods or periods with high renewable availability, thereby reducing peak demand. The corresponding state-of-energy (SoE) dynamics are given by

$$\mathrm{SoE}_{t+1} = \mathrm{SoE}_t + P_t^{EV}/\eta^{EV}\Delta t, \forall t \in \left[T_{\text{start}}^{EV}, T_{\text{end}}^{EV}\right] \tag{3}$$

$$\mathrm{SoE}^{\min} \leq \mathrm{SoE}_t \leq \mathrm{SoE}^{\max}, \forall t \in \left[T_{\text{start}}^{EV}, T_{\text{end}}^{EV}\right] \tag{4}$$

$$P_t^{EV} = 0, \quad t \notin \left[T_{\text{start}}^{EV}, T_{\text{end}}^{EV}\right] \tag{5}$$

$$\mathrm{SoE}_{T_{\text{end}}^{EV}} \geq \mathrm{SoE}_{tr}, \tag{6}$$

where $P_t^{\mathrm{EV}}$ denotes the active charging power of the EV when drawing power from the household, and is constrained by the EV's rated power limits $P^{\mathrm{EV},\min} \leq P_t^{\mathrm{EV}} \leq P^{\mathrm{EV},\max}$. The parameter $\eta^{\mathrm{EV}}$ represents the charging efficiency. The interval $\left[T_{\text{start}}^{EV}, T_{\text{end}}^{EV}\right]$ indicates the plug-in availability window during which charging or discharging can occur, with $P_t^{EV} = 0$ enforced outside this period. Moreover, $\mathrm{SoC}_{tr}^{EV}$ defines the minimum required state of charge at the departure time to satisfy the user's travel demand.

### C. Thermostatically Controlled Load Model

Thermostatically controlled loads (TCLs), such as air-conditioning (AC) units, are modeled through indoor temperature dynamics subject to user-defined comfort constraints. By adjusting power levels of TCLs, the REFM controller can shift or modulate thermal demand without violating comfort constraints $\left[\mathrm{Temp}^{\text{in,min}}, \mathrm{Temp}^{\text{in,max}}\right]$. The thermal dynamics are expressed as

$$\begin{aligned}\mathrm{Temp}_{t+1}^{\text{in}} &= \mathrm{Temp}_t^{\text{in}} \\ &- \frac{\left(\mathrm{Temp}_t^{\text{in}} - \mathrm{Temp}_t^{\text{out}} + \eta^{AC} R^{AC} P_t^{\mathrm{AC}}\right)\Delta t}{C^{AC} R^{AC}},\end{aligned} \tag{7}$$

$$\mathrm{Temp}^{\text{in,min}} \leq \mathrm{Temp}_t^{\text{in}} \leq \mathrm{Temp}^{\text{in,max}}, \forall t \tag{8}$$

where $\mathrm{Temp}_{\mathrm{t}}^{\text{in}}$ and $\mathrm{Temp}_{\mathrm{t}}^{\text{out}}$ denote indoor and external temperatures, respectively. The parameters $\eta^{AC}$, $R^{AC}$, and $C^{AC}$ represent the efficiency of the AC unit, the equivalent thermal resistance, and the thermal capacitance of the building envelope, respectively [15]. The active power of the AC unit is represented by $P_t^{\mathrm{AC}}$, which is constrained by its rated power limits $P^{\mathrm{AC},\min} \leq P_t^{\mathrm{AC}} \leq P^{\mathrm{AC},\max}$.

### D. Shiftable Load Model

Shiftable loads are appliances whose operation can be deferred within user-specified time windows while meeting completion requirements. Examples include laundry machine and

979-8-3315-8138-1

dishwasher. The operational status of each shiftable appliance $i$ is characterized by a binary variable $u_t^{\mathrm{S},i} \in \{0,1\}$, where the operation is non-stop once initiated. The corresponding operating logic is defined as follows

$$u_t^{\mathrm{S},i} = 0, \qquad \forall t \notin \left[T_{\text{start}}^{\mathrm{S},i},\, T_{\text{end}}^{\mathrm{S},i}\right],\ i \in \mathcal{I}, \tag{9}$$

$$P_t^{\mathrm{S},i} = u_t^{\mathrm{S},i} \mathbf{P}^{\mathrm{S},i}, \qquad \forall t, i \in \mathcal{I} \tag{10}$$

$$\sum_{t=T_{\text{start}}^{\mathrm{S},i}}^{T_{\text{end}}^{\mathrm{S},i}} u_t^{\mathrm{S},i} = T_{\text{req}}^{\mathrm{S},i}, \qquad \forall i \in \mathcal{I}, \tag{11}$$

$$T_{\text{end}}^{\mathrm{S},i} - T_{\text{start}}^{\mathrm{S},i} \geq T_{\text{req}}^{\mathrm{S},i}, \qquad \forall i \in \mathcal{I}, \tag{12}$$

where $P_t^{\mathrm{S},i}$ denotes the the instantaneous active power consumption of $i$th shiftable load, and $\mathbf{P}^{\mathrm{S},\, i}$ denotes the rated power of $i$th shiftable load. And parameters $T_{\text{start}}^{\mathrm{S},i}$ , $T_{\text{end}}^{\mathrm{S},i}$ , and $T_{\text{req}}^{\mathrm{S},i}$ denote allowed start time, allowed end time, and the total duration required to complete the task, respectively.

## III. Dynamic Residential Energy Flexibility Management Problem Formulation

In this section, the REFM problem is formulated. To address the stochastic nature within the REFM, such as variations in load, temperature, and PV generation, the problem is first formulated as an MDP. Furthermore, to accommodate ad-hoc user requests, the problem dynamics are extended through an intent-driven reward mechanism. First, the total residential active power consumption at time $t$ is defined as the sum of the inflexible load (base load $P_t^{\text{base}}$ and PV generation $P_t^G$) and all flexible loads, given by

$$P_t^{\text{total}} = P_t^{\text{base}} + P_t^{AC} + \sum_{i \in \mathcal{I}} P_t^{\mathrm{S},i} + P_t^{EV} - P_t^B - P_t^G, \tag{13}$$

while subject to a limit of $|P_t^{\text{total}}| <= P^{\text{total,max}}$. Then, we formulate the REFM problem as a finite-horizon MDP $\mathcal{M} = (\mathcal{S}, \mathcal{A}, R, \mathcal{P}, \gamma)$ with $t \in \mathcal{T} \triangleq \{0, 1, \ldots, T-1\}$ and step size $\Delta t$. The corresponding MDP components are:

*State space $\mathcal{S}$:* The system state aggregates controllable internal states and exogenous variables, which is defined as: $s_t = (\mathrm{SoC}_t, \mathrm{SoE}_t, \mathrm{Temp}_t^{in}, \mathrm{Temp}_t^{out}, \{\ell_t^i\}_{i\in\mathcal{I}}, P_t^{\text{base}}, P_t^G, \beta_t)$, where $\ell_t^i \in \{0, 1, \ldots, T_{\text{req}}^{\mathrm{S},i}\}$ is the remaining processing time (in slots) of shiftable appliance $i$ (with $\ell_t^i = 0$ meaning idle), and $\beta_t$ is the electricity tariff.

*Action space $\mathcal{A}(s_t)$:* At each $t$, the REFM smart agent determines the control actions for flexible energy devices, including the charging and discharging operations of the BESS and EV, the operating status of the AC, and the start decisions of shiftable loads: $a_t = (P_t^B, P_t^{EV}, P_t^{AC}, \{u_t^{S,i}\}_{i\in\mathcal{I}})$, with the state-dependent feasibility constraints.

*Reward $R$:* Given a user-intent request $q$, the REFM reward function may dynamically change due to the request. The request may contain a set of user-intent-oriented variables, which is $b_t(q) = (T_{\text{start}}^{EV}, T_{\text{end}}^{EV}, \mathrm{Temp}^{\text{in,min}}, \mathrm{Temp}^{\text{in,max}}, T_{\text{start}}^{\mathrm{S},i}, T_{\text{end}}^{\mathrm{S},i}, P_t^{\mathrm{S},i}, T_{\text{req}}^{\mathrm{S},i})$. The reward function consists of the instantaneous energy cost and the battery degradation cost. In addition, a penalty term $r_p$ is introduced when the scheduling policy fails to accommodate a user request. For instance, if the user specifies a desired completion time for the EV charging period and the charging task is not completed by that time, a penalty is imposed. The penalty serves as a surrogate measure of average user discomfort [16], representing a generic preference proxy without requiring users to explicitly parameterize individual comfort preferences. The intent-driven reward function is defined as:

$$R\left(s_t, a_t \mid b_t\right) = \begin{cases} -\beta_t\, P_t^{\text{total}}\, \Delta t - C_t^d, & \text{if satisfy } b_t, \\ -\beta_t\, P_t^{\text{total}}\, \Delta t - C_t^d - r_p, & \text{otherwise.} \end{cases} \tag{14}$$

*State transitions $\mathcal{P}$:* Given $s_t$ and $a_t$, the controllable components evolve deterministically, while exogenous components follow a stochastic process. The transitions of controllable states include $(\mathrm{SoC}_{t+1}|\mathrm{SoC}_t)$, $(\mathrm{SoE}_{t+1}|\mathrm{SoE}_t)$, $(\mathrm{Temp}_{t+1}^{in}|\mathrm{Temp}_t^{in})$, and $(\ell_{t+1}^i|\ell_t^i, b_t)$. Note that the transition $(\ell_{t+1}^i|\ell_t^i, b_t)$ can change dynamically in response to user requests that modify the shiftable-load workload. Exogenous processes (e.g., $P_t^{\text{base}}$, $\mathrm{Temp}_t^{out}$, and $P_t^G$) evolve according to a stochastic kernel $p(w_{t+1}|w_t)$, independent of the control given $w_t$. The full transition kernel factorizes as

$$\mathcal{P}(s_{t+1}|s_t, a_t) = \delta\big(s_{t+1}^{\mathrm{c}} - f(s_t^{\mathrm{c}}, a_t, w_t, b_t)\big) p(w_{t+1}|w_t) \tag{15}$$

where $s^{\mathrm{c}}$ collects the controllable states, $w_t$ collects exogenous components, $f(\cdot)$ is the deterministic update and $\delta(\cdot)$ is the Dirac measure.

Given a discount factor $\gamma \in (0, 1]$, the objective is to find a policy $\pi(a_t \mid s_t)$ that maximizes the intent-driven reward:

$$\max_{\pi}\ \mathbb{E}_{\mathcal{P},\pi}\left[\sum_{t=0}^{T-1} \gamma^t\, R\left(s_t, a_t \mid b_t\right)\right], \tag{16}$$

subject to the power balance and all state/action constraints of the energy flexibility devices within the REFM system. In the practical REFM system, delays may arise due to user-intent processing, optimization computation, and communication with flexible energy resources. In this work, such delays are considered to be bounded and sufficiently small relative to the decision time step size $\Delta t$, thereby preserving the validity of the discrete-time MDP formulation.

## IV. RAG-based User-Intent-Driven Dynamic Residential Energy Flexibility Management

To dynamically perform REFM based on user intent, user requests must be transformed into the optimization objectives defined in (16). However, users typically lack the expertise to directly modify these objectives and can only express their preferences through natural language. Therefore, the natural language request is parsed into a structured user-intent variable set $b_t(q)$. To accomplish this, a RAG framework is employed for the REFM smart agent. Since conventional RAG models may not accurately capture user intents, an enhanced RAG algorithm is developed based on the GRPO method. To further align the model behavior with REFM scenarios, an adaptive reward mechanism specifically tailored for REFM is introduced, which adjusts the reinforcement signal according to the user intent accuracy, feasible schedules, and structured input requirements of the optimization model.

The GRPO algorithm employs policy network learning to optimize a policy $\pi_\theta$ that enhances the accuracy of parsing

979-8-3315-8138-1

natural language requests into the structured variable set. For a user intent $q$, the policy $\pi_\theta$ generates $K$ candidate structured outputs $\{b_k\}_{k=1}^{K}$, where each $b_k$ corresponds to a possible interpretation of EV, TCL, or shiftable-load operating parameters. Each $b_k$ is evaluated using a weighted task-specific reward $r_k$. GRPO then compares the candidates within the same group by applying reward normalization, $\tilde{r}_k = \frac{r_k - \text{mean}(\{r_1,\ldots,r_K\})}{\text{std}(\{r_1,\ldots,r_K\})}$, and assigns $\tilde{r}_k$ as the advantage for $b_k$. This ensures that interpretations that better satisfy REFM requirements receive higher reinforcement relative to alternative candidates for the same intent. The overall GRPO objective $\mathcal{J}(\theta)$ is defined as

$$\mathcal{J}(\theta) = \mathbb{E}_{\{b_k\}\sim\pi_{\theta_{\text{old}}}}\left[\frac{1}{K}\sum_{k=1}^{K}\left(\tilde{\mathcal{A}}_k - \beta\,\mathbb{D}(\pi_\theta\|\pi_{\text{ref}})\right)\right], \quad (17)$$

where the per-candidate surrogate term $\tilde{\mathcal{A}}_k$ is computed as

$$\tilde{\mathcal{A}}_k = \frac{1}{|b_k|}\sum_{n=1}^{|b_k|}\min\Big(\mathcal{A}_{k,n}\,\tilde{r}_k,\ \text{clip}(\mathcal{A}_{k,n},\,1-\epsilon,\,1+\epsilon)\,\tilde{r}_k\Big), \quad (18)$$

where $\mathcal{A}_{k,n} = \frac{\pi_\theta(b_{k,n}|q,b_{k,<n})}{\pi_{\theta_{\text{old}}}(b_{k,n}|q,b_{k,<n})}$ and $n$ indexes the token position within the sequence $b_k$. The KL term $\mathbb{D}$ constrains the policy update, ensuring that the learned $q \rightharpoonup b$ mapping remains consistent with the reference model and avoids unstable divergence. It is computed using the unbiased estimator

$$\mathbb{D}(\pi_\theta\|\pi_{\text{ref}}) = \frac{\pi_{\text{ref}}(b_{k,n}|q,b_{k,<n})}{\pi_\theta(b_{k,n}|q,b_{k,<n})} - \log\frac{\pi_{\text{ref}}(b_{k,n}|q,b_{k,<n})}{\pi_\theta(b_{k,n}|q,b_{k,<n})} - 1. \quad (19)$$

To align the policy optimization with the objectives of residential energy flexibility, a set of task-specific, interpretable reward functions is designed to guide the GRPO toward accurate and user-aligned energy scheduling decisions:

- Intent-Parsing Reward $R_{\text{int.}}(b_k \mid q)$ – Encourages correct extraction of user-specified preferences, such as comfort range, appliance deadlines, and personal schedules, from natural language input. Outputs that follow the required representation receive a reward of 1; otherwise, 0.
- Scheduling Feasibility Reward $R_{\text{sch.}}(b_k \mid q)$ – Evaluates whether the generated energy re-planning schedule meets user objectives while satisfying physical and operational constraints (e.g., EV constraints, temperature limits). Feasible schedules receive a reward of 1; otherwise, 0.
- Structured Output Reward $R_{\text{str.}}(b_k \mid q)$ – Evaluates how well the generated output conforms to the predefined structured format required by the REFM optimization module. A reward of 1 is assigned when the generated intent representation fully satisfies syntactic and semantic format constraints, ensuring reliable downstream execution of the optimization process.

Then, the final composite reward for each generation step is given by

$$r_k = R_{\text{int.}}(b_k|q) + R_{\text{sch.}}(b_k|q) + R_{\text{str.}}(b_k|q) + \mathbb{I}_{\text{all}=1}\cdot R_{\text{bonus}} \quad (20)$$

where $\mathbb{I}_{\text{all}=1}$ denotes an indicator function that equals 1 when all three sub-rewards attain their maximum value (i.e., each equals 1), and 0 otherwise. Also, $R_{\text{bonus}}$ represents the additional consistency reward encouraging simultaneous success across intent parsing, scheduling, and structural correctness.

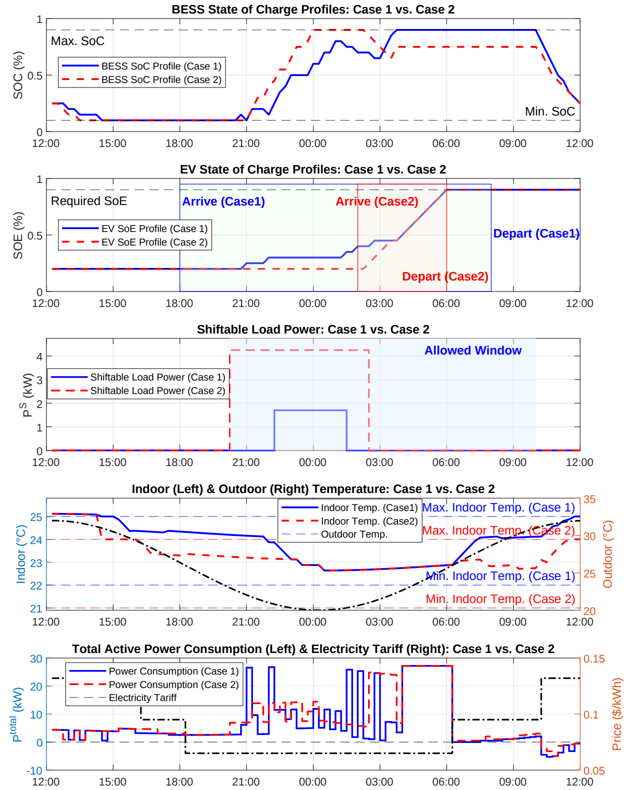


Fig. 2. REFM performance on energy-flexibility devices for Case I (normal day) and Case II (with user-intent request).

## V. Case Study

In this section, the performance of the proposed RAG-based dynamic REFM framework is evaluated. A typical residential electricity-demand dataset is considered [17]. The BESS and EV ratings follow the specifications of the Tesla Powerwall and a standard Tesla EV [18], respectively, with a round-trip efficiency of 92%. A time-of-use electricity tariff is considered [19]. The large language model DeepSeek-R1-Distill-Qwen-7B [20] is used to perform user-intent interpretation, and $K = 7$ candidate intent interpretations are generated for downstream evaluation. A set of user-intent requests is generated for each energy-flexibility device. Examples of user-intent requests include:

1) "The EV will arrive home at 2:00 a.m. Please charge it before 6:00 a.m."
2) "Activate the laundry machine and dishwasher tonight."
3) "Cool down the house by one degree Celsius today."

### A. Performance of the User-Intent–Driven Dynamic REFM

The performance of the proposed REFM on energy-flexibility devices is shown in Fig. 2. The figure illustrates the BESS SoC profiles, EV SoE profiles, shiftable-load power consumption, indoor and outdoor temperatures (with indoor temperature regulated by TCLs), as well as the total active-power consumption and the electricity tariff. To demonstrate the effectiveness of the framework in handling user-intent requests, we compare a normal-day setup (Case I) with a case in which a user issues several requests to the energy-flexibility devices (Case II), with both cases employing the proposed REFM framework. The simulation is conducted over a 24-hour period, from noon to noon of the following day.

TABLE I
COMPARISON OF INTENT PARSING PERFORMANCE

| Method | EV | Shiftable Load | TCL | Overall |
|---|---|---|---|---|
| Zero-shot [8] | 0.342 | 0.215 | 0.701 | 0.419 |
| Few-shot [9] | 0.503 | 0.242 | 0.825 | 0.523 |
| RAG [11] | 0.748 | 0.573 | 1.000 | 0.774 |
| Enhanced RAG (Ours) | **0.975** | **0.956** | **1.000** | **0.977** |

For Case I, the BESS charges during low-tariff periods and discharges during high-tariff periods, while EV charging and shiftable loads are scheduled in separate time windows to avoid concentrated power consumption. For Case II, the user issues three intent requests for the energy-flexibility devices considered in the case study. These requests are parsed into optimization objectives and incorporated into the re-planning process, typically specifying the EV plug-in availability window ($T_{\text{start}}^{EV}$, $T_{\text{end}}^{EV}$), the shiftable load allowed window and work load, ($T_{\text{start}}^{\text{S},i}$ ,$T_{\text{end}}^{\text{S},i}$ , $P_t^{\text{S},i}$, $T_{\text{req}}^{\text{S},i}$), and the required indoor temperature range ($\text{Temp}^{\text{in,min}}$, $\text{Temp}^{\text{in,max}}$). Compared to Case I, the system experiences a higher load level due to a limited EV plug-in window and increased shiftable-load demand, leading to a short congestion period (2:00–2:30) during which the EV charges at maximum power while shiftable loads are operating. To mitigate this overlap, the BESS is charged in advance and subsequently discharged to support the temporary overload. For the TCLs, indoor temperature is effectively regulated during periods of high outdoor temperature while remaining stable during off-peak hours. Overall, the proposed coordination scheme successfully satisfies user-intent requests while maintaining device operation within required limits and preserving overall operational efficiency.

To further assess the effectiveness of the proposed framework in terms of energy management cost reduction, we compare it with a rule-based strategy [3] and an MPC-based method [4] over a 7-day summer period. The proposed method achieves the lowest daily energy cost of $5.94, followed by the MPC-based method with a daily cost of $7.12, while the rule-based strategy results in the highest cost of $9.35. The improved performance of the proposed method is mainly attributed to its ability to dynamically adapt optimization decisions in response to user-intent requests, whereas the MPC method relies on fixed predictive models, and the rule-based strategy lacks adaptive optimization capability.

### *B. User-Intent Interpretation Evaluation*

To evaluate intent interpretation performance, accuracy is defined as the percentage of user requests for which all required intent components are correctly parsed. A prediction is considered correct only when the complete structured intent exactly matches the ground truth. The proposed method is compared with zero-shot [8], few-shot [9], and conventional RAG approaches [11]. As shown in Table I, zero-shot and few-shot prompting exhibit limited accuracy due to difficulties in interpreting domain-specific rules and complex parameters, particularly for EV and shiftable-load tasks. Incorporating external knowledge via RAG improves performance across all categories, with stronger results for TCL-related intents. However, EV and shiftable-load tasks still exhibit notable inaccuracies under standard RAG, which can degrade REFM performance. The proposed enhanced RAG method further improves consistency and correctness, achieving the highest overall accuracy of 0.977 and demonstrating strong effectiveness in handling both simple and complex user-intent requests.

## VI. CONCLUSION

This paper presented a user-intent–driven REFM framework that integrates user intent directly into the optimization process. By formulating the REFM problem as a user-intent–driven reward MDP and introducing a RAG-based interactive layer, the proposed approach enables dynamic re-planning that faithfully reflects user requirements. An enhanced RAG algorithm, built upon GRPO, further improves the accuracy of intent parsing. A case study validates that the framework can accurately interpret natural-language requests and adjust energy-flexibility device schedules accordingly in a manner that both satisfies user preferences and maintains operational efficiency. If desired, the proposed framework can be extended to broader distributed energy applications.

## REFERENCES


[1] U.S. Energy Information Administration, *Monthly Energy Review*. Accessed: Nov. 3, 2025. [Online]. Available: https://www.eia.gov/totalenergy/data/monthly/

[2] H. Li, Z. Wang, T. Hong, and M. A. Piette, "Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications," *Adv. Appl. Energy*, vol. 3, p. 100054, Aug. 2021.

[3] T. Kawakami, T. Yoshihisa, N. Fujita, and M. Tsukamoto, "A rule-based home energy management system using the Rete algorithm," in *Proc. IEEE 2nd Global Conf. Consumer Electronics (GCCE)*, Tokyo, Japan, Oct. 2013, pp. 162–163.

[4] R. Godina, E. M. Rodrigues, E. Pouresmaeil, J. C. Matias, and J. P. Catalão, "Model predictive control home energy management and optimization strategy with demand response," *Appl. Sci.*, vol. 8, no. 3, p. 408, Mar. 2018.

[5] L. Zhao, T. Yang, W. Li, and A. Y. Zomaya, "Deep reinforcement learning-based joint load scheduling for household multi-energy system," *Appl. Energy*, vol. 324, p. 119346, Oct. 2022.

[6] S.-J. Chen, W.-Y. Chiu, and W.-J. Liu, "User preference-based demand response for smart home energy management using multiobjective reinforcement learning," *IEEE Access*, vol. 9, pp. 161627–161637, Dec. 2021.

[7] S. Bubeck, V. Chandrasekaran, R. Eldan, *et al.*, "Sparks of artificial general intelligence: early experiments with GPT-4," *arXiv preprint*, arXiv:2303.12712, Apr. 2023.

[8] Y. Yang, Q. Yang, D. Li, L. Sun, L. Sen, and R. Liu, "LLM-MAC: LLM-driven multi-agent coordination for generalizable home energy management," in *Proc. Chin. Control Conf.*, Chongqing, China, Oct. 2025, pp. 7318–7323.

[9] F. Michelon, Y. Zhou, and T. Morstyn, "Large language model interface for home energy management systems," in *Proc. ACM e-Energy*, Jun. 2025, pp. 590–602.

[10] P. Han, P. Song, H. Yu, and J. You, "In-context learning may not elicit trustworthy reasoning: A-not-B errors in pretrained language models," in *Proc. EMNLP*, Miami, FL, USA, Sep. 2024, pp. 5624–5643.

[11] P. Lewis, E. Perez, A. Piktus, *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 9459–9474, Dec. 2020.

[12] J. Huang, S. Madala, R. Sidhu, C. Niu, H. Peng, J. Hockenmaier, and T. Zhang, "RAG-RL: advancing retrieval-augmented generation via RL and curriculum learning," *arXiv preprint*, arXiv:2503.12759, May 2025.

[13] J. Ren, Y. Xu, X. Wang, W. Li, W. Ma, and Y. Liu, "Effective and transparent RAG: adaptive-reward reinforcement learning for decision traceability," *arXiv preprint*, arXiv:2505.13258, Oct. 2025.

[14] L. Lam and P. Bauer, "Practical capacity fading model for Li-Ion battery cells in electric vehicles," *IEEE Trans. Power Electron.*, vol. 28, no. 12, pp. 5910-5918, Dec. 2013.

[15] J. Aldahmashi and X. Ma, "Real-time energy management in smart homes through deep reinforcement learning," *IEEE Access*, vol. 12, pp. 43155–43172, Mar. 2024.

[16] E. Togashi, "Reward function design in reinforcement learning for HVAC control: a review of thermal comfort and energy efficiency trade-offs," *Energy and Buildings*, vol. 348, p. 116439, Dec. 2025.

[17] M. A. Koksal, I. H. Rowlands, and P. Parker, "Energy, cost, and emission end-use profiles of homes: an Ontario (Canada) case study," *Applied Energy*, vol. 142, pp. 303–316, Mar. 2015.

[18] *Tesla*. Accessed: Nov. 3, 2025. [Online]. Available: https//www.tesla.com

[19] Ontario Energy Board Electricity Rates. Accessed: Nov. 3, 2025. [Online]. Available: https://www.oeb.ca/consumer-information-and-protection/electricity-rates

[20] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, S. Ma, P. Wang, X. Bi, and X. Zhang, "DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning," *arXiv preprint*, arXiv:2501.12948, Jan. 2025.


979-8-3315-8138-1