Adaptive Home Energy Management: Human-Centric RL Approach for Diverse Situations

1st Zachary Tchir
Electrical and Computer Engineering
University of Alberta
Edmonton, Canada
ztchir@ualberta.ca

2nd Petr Musilek
Electrical and Computer Engineering
University of Alberta
Edmonton, Canada
pmusilek@ualberta.ca

3rd Marek Z. Reformat
Electrical and Computer Engineering
University of Alberta
Edmonton, Canada
University of Social Sciences, Łódź, Poland
reformat@ualberta.ca

***Abstract*—The increased demand for Smart Home control technologies and the rapid growth of AI-based approaches provide an opportunity to develop systems that improve conveniences and reduce potential barriers to adopting renewable technologies, including Electric Vehicles, Photo Voltaic Solar Panels, and Household Batteries (or Energy Storage Systems). Reinforcement Learning is an AI tool for training systems to perform complex tasks and achieve challenging goals. This paper introduces a human-centric RL-based Home Energy Management System that manages a homeowner's requirements while maximizing comfort and convenience. The homeowner can alter the system's behaviour depending on existing situations and needs. The system can control the home's energy resources in the presence of the uncertainty and variability of solar power generation, the varying electricity demands, and the priorities of a homeowner.**

***Index Terms*—Reinforcement Learning, Energy Management, V2X, Human-Centric RL**

I. Introduction

Reinforcement learning (RL) algorithms represent competent mechanisms for training agents in various complex environments, such as control systems and gaming. In the realm of control systems, RL is effective in optimizing processes and managing dynamic settings. RL algorithms are particularly skilled at managing the variability often seen in these applications and unexpected situations. One distinct advantage of RL agents is their ability to learn without the need for extensive datasets. They can be trained in simulated environments, acquiring essential skills before being deployed in actual scenarios.

RL methods have been extensively explored in building energy management and microgrid energy management systems [1] [2]. The applications include model-based energy management [3], micro-grid energy management with flexible demand [4], battery control under cycle-based degradation [5], and energy storage arbitrage with degradation [6].

With the increasing importance of green solutions for homes and emerging Vehicle-to-Anything (V2X) technologies [7], there is a growing need for developing intelligent and human-friendly Home Energy Management Systems (HEMS). One of the most promising approaches is based on applying RL technologies. Numerous studies focusing on agents controlling various household appliances and the home energy storage system or charging of electrical vehicles (EVs) have been published [8] [9], [10] [11] [12]. In these home energy management systems, the agents are trained to control various home devices, including HVAC, lighting, and appliances such as washers and dryers. It results in a challenging deployment as agents require different decision processes. Their actions attempt to reach a balance between user comfort and the cost of electricity. However, it results in inconvenient situations – for example, the agents may prevent users from using certain appliances at times they would like to.

Further, to our knowledge, no studies consider factors that influence human-like decision-making processes, such as emotions and contextual issues. Such factors can significantly affect how the users view their goals; in one instance, they may prioritize saving money on their electricity bill; in another, the fear of grid instabilities may supersede the need to be cost-efficient or, yet in another case, the ability to go for a long trip replaces the cost efficiency.

In this paper, we expand on our previous work [13] to examine a human-centric home energy management system. The method proposed investigates training multiple agents to perform their tasks in diverse situations with three specific goals: ensuring cost efficiency, guaranteeing uninterrupted energy supply, and maximizing EV driving range. The operations of agents are interactively identified via a human setting up the goals of the RL agent.

II. Home Energy Management Agent

A. Problem Formulation

Let us envision a household that is equipped with an energy storage system (ESS) and photovoltaic (PV) panels. Additionally, the homeowners have just bought an electrical vehicle (EV). Such a house – not very futuristic yet energy-wise – would already benefit from an automated system/agent managing all sources and sinks of energy. It would eliminate

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00028

the owner's need to continuously monitor and control how energy is utilized, the state of devices requiring charging, and the cost of electricity.

The ultimate goal of our agent is to maximize the homeowners' comfort and security. We anticipate the agent has to work optimally in a number of different scenarios. To address such needs, we develop an agent capable of achieving the following, so-called, scenario-goals:

- **Cost-awareness**: to reduce the overall electricity cost via intelligent management of all sources of energy;
- **Range-awareness**: to secure readiness of the EV, i.e., to keep its state of charge to guarantee its full driving range as quickly as possible after returning to the house;
- **Outage-awareness**: to manage energy sources to maintain an uninterrupted energy supply even in case of outages.

The proposed system involves agents trained to prioritize one of the primary goals described above; the system will allow the homeowners to identify their primary goal(s) by selecting one of the scenarios. This paper focuses on comparing the performance of each operating mode. In addition to the primary goal, the agent will maintain some performance in the remaining two of the three scenario goals.

B. Training Environment

To develop an RL-trained energy managing agent that helps or even substitutes the homeowner's need to monitor, we need a data-driven simulator to train and test the agent. Such an RL simulator is modelled as a set of states, a set of activities or actions, and a reward function. The states (or observations) represent what the agent can observe/know about the environment. The actions are the methods that an agent performs to interact with an environment; these could be low-level control signals. A reward function provides feedback values representing how an agent's actions influence/change the environment. The RL mechanism 'teaches' our agent to choose the best action in a given state via maximizing the reward.

In the case of the household energy management system, the environment is composed of the household load, PV, ESS, and EV. The household load is the electrical load of a single home (this includes all potential loads connected to the household electrical panel) taken from [14] plus the energy consumption of the EV. The ESS specifications are based on the Tesla Power Wall 2 with 5 kW with a discrete set of charging and discharging actions, each a specific power setting (±10 % intervals) and a 13.5 kWh capacity. The PV system for the simulated household is a 20-panel system, each capable of generating 325 W of power with a total system output of 6.5kW. The panels are assumed to be 100 % efficient, and the solar generation is based on [15]. The EV's specs are based on a Tesla Model 3 with an 82 kWh battery with a maximum charging and discharging of 11.5 kWh. In this case, the charging is controllable via a discrete action space, ±10 % intervals up to the maximum charging and discharging levels, similar to ESS.

The agent manages the house's energy based on the previous 24 hours of electricity usage and PV (solar) generation, the state of charge (SoC) of the EV and the ESS, and whether or not the EV is connected to be charged. The agent controls the charging/discharge rate of the EV(s) and ESS.

The charging and discharging of the EV and ESS batteries are modelled using the formula (the same for EV and ESS)

$$C_{t+1} = \begin{cases} C_{max}, & \text{if } \; C_{max} - C_t \leq \eta \frac{p_{B,t}}{P_{max}} \\ C_{min}, & \text{if } \; C_{min} - C_t \geq \eta \frac{p_{B,t}}{P_{max}} \\ C_t + \eta \frac{p_{B,t}}{P_{max}}, & \text{otherwise.} \end{cases} \tag{1}$$

C_t is the SoC of the battery at time-step t, η is the charging/discharging efficiency (assumed 100 %), p_t^B is the charge or discharge power during t, while P_{max} is the capacity of the battery in kWh. C_{max} and C_{min} are the maximum and minimum charging values of the battery; for now, these are set to 0 and 100 %. The equation ensures the batteries cannot be charged or discharged on C_{max} and C_{min}.

Power is drawn from or returned (sold) to the grid every time step and obeys the following power balance equation

$$p_{util,t} = p_t + p_{EV,t} + p_{ESS,t} - p_{PV,t} \tag{2}$$

where p_t^{util} is the power to/from the grid, $p_{PV,t}$ is the power generated from the PV system at time-step t, which is always negative since the PV always generates energy. The powers p_t^{EV} and p_t^{ESS} can be positive (sink) or negative (source). Negative p_t^{util} means excess energy 'produced' by a household is sold to the grid.

Further, the cost or profit associated with the drawn or generated power is represented as

$$f_t = (\frac{\lambda_t - \lambda^*}{2} |p_{util,t}| + \frac{\lambda_t + \lambda^*}{2} p_{util,t}) \tag{3}$$

where f_t is the total electricity cost, λ^* is the fixed utility buyback rate \$0.082 /kWh and λ_t is the electricity price at a time-step t. The price λ_t follows a time-of-use model shown in Fig. 1. The peak rate is \$0.170/ kWh, mid-peak rate is \$ 0.113/ kWh and the off-peak rate is \$0.082/ kWh. If p_t^{util} is positive, then $f_t = \lambda_t \cdot p_t^{util}$, the agent (household) is buying power from the grid. If it is negative $f_t = \lambda^* \cdot p_t^{util}$, the agent is selling it to the grid. An important aspect of the training environment is the operation of the EV. The trips are randomized to ensure the agent can see a variety of situations. The electric vehicle takes a trip at a random time of day for

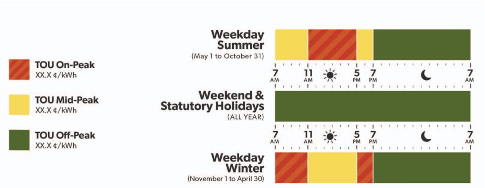

Fig. 1: Time-of-use electricity pricing used in the simulated smart home environment. [16]

a random duration to simulate realistic driving patterns, not a predictable 9-5 schedule. The EV's SoC is reduced according to the distance travelled during a trip. Fig 2 shows the distance travelled is modelled based on a gamma distribution of a frequent driver in the US [17].

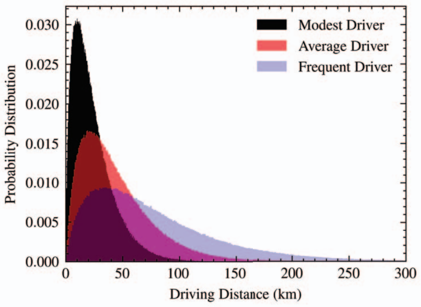

Fig. 2: Gamma Distribution for Driving Distances to Model EV Trips.

During a testing process and for comparison purposes, all scenarios executed during testing runs see consistent departures, arrivals, and driving distances of the EV throughout each run.

C. Reward Function

The reward function per time step r_t is defined below, and for each agent, the coefficients are changed depending on the human's goal.

$$
\begin{aligned}
r_t =& f_t + g_{EV,t} + g_{ESS,t} \begin{cases} P & \text{if } p_{util,t} > 0 \\ \alpha \cdot p_{util,t}^2 & \text{if } p_{util,t} \leq 0 \end{cases} \\
&+ \rho_{EV} \cdot |p_{EV,t}| + \rho_{ESS} \cdot |p_{ESS,t}| \\
&+ \begin{cases} \gamma_{s,0} \cdot S_{sEV} & \text{if } |sgn(p_{EV,t}) - sgn(p_{EV,t-1})| \neq 2 \\ \gamma_{s,1} \cdot S_{sEV} & \text{otherwise} \end{cases} \\
&+ \begin{cases} \gamma_{s,0} \cdot S_{sESS} & \text{if } |sgn(p_{ESS,t}) - sgn(p_{ESS,t-1})| \neq 2 \\ \gamma_{s,1} \cdot S_{sESS} & \text{otherwise} \end{cases}
\end{aligned} \tag{4}
$$

where f_t is the utility cost/profit, the outage anxiety penalty (representing uncertainty of accessing power during an outage) occurs only during an outage (at that time f_t is zero) – it is a fixed value P if the power requirements are not met, or $\alpha \cdot p_{util,t}^2$ if excess energy is used. ρ is the battery degradation coefficient and $p_{EV,ESS,t}$ is the energy use of EV and/or ESS, all at the time t.

To address the issue of insufficient charge of EV and ESS, we introduce the concept of range and backup anxieties. They represent the fear that SoC of EV and ESS, respectively, are too low to be useful. The mathematical equation modeling anxiety is depicted as

$$g_{EV,t} = \beta * max(C_{EV,Req} - C_{EV,t}, 0) \tag{5}$$

$$g_{ESS,t} = \beta * max(C_{ESS,Req} - C_{ESS,t}, 0) \tag{6}$$

g_t is the charge anxiety or backup anxiety at time-step t, β is the anxiety coefficient that represents the household anxiety about having EV and ESS sufficiently charged to meet the capacity requirement or backup power requirement, $C_{n,t}$ is the capacity of the battery in kWh and C_{Req} is the required battery capacity in kWh. The EV charge requirement is $80\,\%$, and the ESS requirement is $40\,\%$ for the simulated home.

The agent is biased towards action instead of holding a state-of-charge (SoC) at a constant value, typically at $0\,\%$ and $100\,\%$. It would mean that the agent performs charging and discharging actions that offset each other. These repeated charge and discharge actions are inefficient and, in the real world, would result in a shorter battery lifetime due to increased battery degradation. During experimentation, we found out that the degradation terms $\rho_{EV} \cdot |p_{EV,t}$ and $\rho_{ESS} \cdot |p_{ESS,t}$ in 4 alone has been ineffective when solving the agent's bias to action. To mitigate that problem, we have implemented a growing and decaying switching penalty term in the reward function.

In our RL reward function, the term switching is defined as consecutive charge/discharge actions executed on EV or ESS. The more such switching actions occur, the more the penalty $S_{s,...}$ increases for the corresponding device scaled by the switching coefficient $\gamma_{s...,0}$ increases. $S_{s,...}$ o The increase happens each time an alternating action occurs. If the agent's action is to wait or charge/discharge for more than three time steps, the penalty coefficient decays at a rate of $\gamma_{s...,1}$ each time the actions are not opposing. Such a mechanism helps reduce inefficient switching actions resulting in a net-zero change in SoC.

III. Training and Results

In the reported experiments, agents using the Stable-Baselines3 Masked PPO Algorithm [18]. Each training episode starts with a randomly selected date regarding the load [15], and solar-based generation [14]. The episode runs for one month, i.e., 2880 time steps of 15 minutes each. At the beginning of each month, the EV and the ESS SoC are randomly initialized. At the beginning of each day, a trip(s) is randomly scheduled – the EV will leave at a random time during the day and be away for a random amount of time. The distance traveled on the trip is determined based on a gamma distribution [17].

A. Training for Diverse Situations

Let us illustrate the agent's performance for each of the three scenario goals. The energy management carried out for **cost-awareness** goal is presented in Fig. 3 (top). It shows a single day of operations. The agent uses excess power available in the EV at the beginning of the day to sell power to the grid. Once the EV leaves (10:45), the agent begins to utilize the ESS to continue to sell power. When energy pricing switches to the off-peak rate, the agent begins to rely on the electricity grid to power the home and leverages the low rate to recharge the ESS and the EV when it is back (20:30).

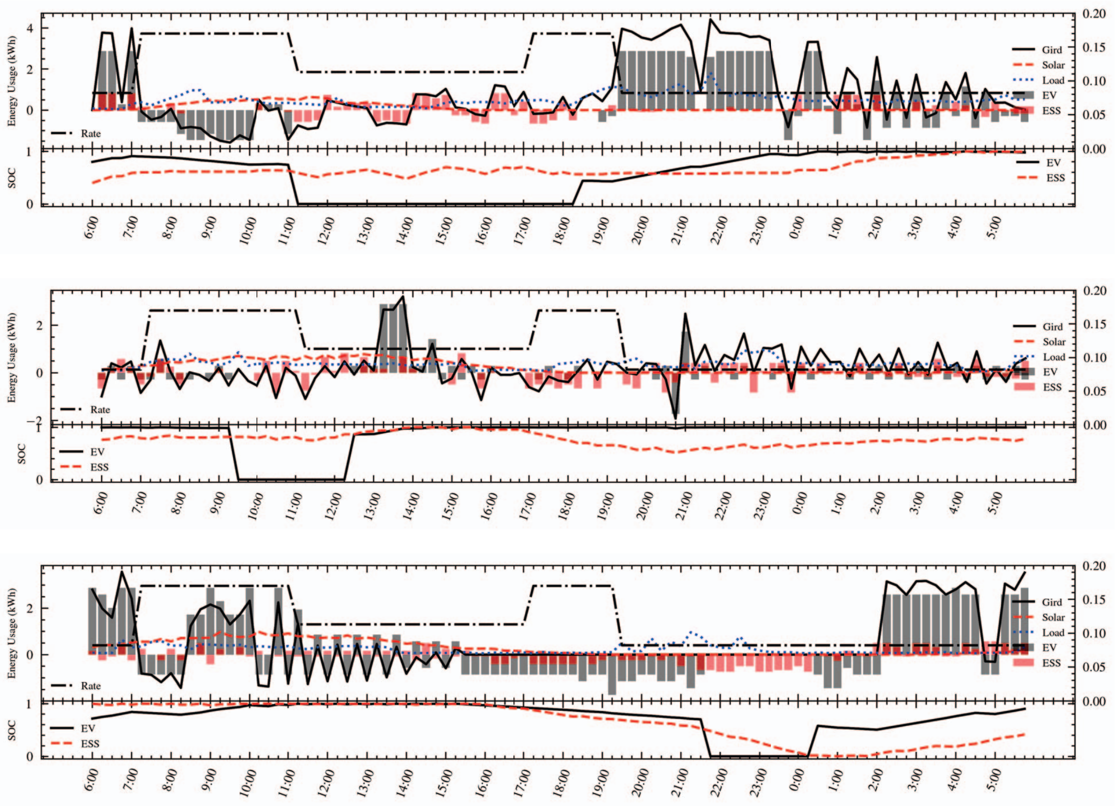

Fig. 3: Comparative analysis of RL-based agent performance in different scenarios: maximizing cost savings and profit (top); maximizing energy security during an outage (middle); and maximizing EV range (bottom).

The **range-aware** agent's operation can be seen in Fig. 3 (middle). In this case, it can be seen that the agent's primary goal is to charge the EV; it makes little use of the ESS and excess EV charge to sell back to the grid. During the day, the solar power generated by the PV is used directly to power the home, with the excess being sold. The agent uses the ESS to offset the load during peak rate between 17:00 and 19:00.

The agent's performance operating in the **outage-awareness** mode is shown in Fig. 3 (bottom). In this case, there is an outage from 15:30 to 2:00, and the agent supplies energy from the EV and ESS to meet the household power needs during the blackout. The agent maintains a cost-awareness goal as shown in Table III and maintains change beyond 75 %. Fig. 3 (bottom) indicates that the EV is charged when the electricity price is low. When excess charge is available in the electric vehicle at the beginning of the day, some of the electricity is used and sold back to the grid before leaving the home for the day. Finally, in the afternoon, the agent charges the ESS above the backup battery requirement set point to ensure that the ESS is sufficiently charged for any potential outages.

B. Comparison of Agent and User Performances

To illustrate the effectiveness of our RL-trained agent, we compare agent performances obtained for each goal, and to the performances achieved by a rule-based energy-conscious homeowner, called baseline cases. The homeowner's behaviour is modelled in our environment based on the following policy. We assume the homeowner in each baseline test case acts according to the following rules. Note. that the average winter and summer monthly costs for a household with only solar panels, no EV, no ESS and no outages are $\$61.74 \pm 28.35$ and $\$-6.25 \pm 18.19$, respectively. The high variance in the cost is due to the variance in the household load and solar data.

The **cost-focused** baseline; a homeowner rules are as follows: they manually charge their electric vehicle during off-peak hours to save money; this means a maximum charging rate when the electric vehicle is connected and the electricity rate is low. The ESS is charged while electricity is generated from the PV panels and discharged when the PV panels do not produce electricity. Thus, when PV panels generate electricity, the ESS is charged. The ESS is discharged when the sun goes down.

For the **range-focused** baseline; a homeowner executes the following rules: the EV is connected and charged when it arrives at home regardless of the electricity rate. Similar to the cost-focused homeowner, the ESS is charged when electricity is generated from the PV panels and discharged when no solar electricity is generated. Any excess solar electricity generated powers the home or is sold back to the grid depending on the load at a given time.

For a homeowner in the **outage-focused** baseline, the ESS is not discharged during regular operation, but only during outages. During outages, the EV is only charged when it is below 50 %. The ESS is charged when there is excess solar energy during an outage. In practice, the homeowner can also manually monitor their home solar system and the ESS SoC to determine when to use the EV as a power source.

The agent with **cost-awareness** goal has been trained to experience minimal power outages. The **range-awareness** agent has been trained to prioritize EV SOC over cost by increasing the charge anxiety coefficient. Finally, the agent with **outage-awareness** goal has been trained to experience an outage at a 50 % chance. Each agent and user has been 'exposed' to the same load and solar data for testing, and each test case experiences the same EV trip schedule (set with a random generator with the same seed).

Our experiments show that RL agents are much more flexible and can utilize the EV and ESS to help reduce electricity costs, maximize EV range, or secure power during periods of grid instability based on the homeowner's preference.

In the baseline cases, i.e., homes with limited or no RL-based control, there is a significant time without power during blackouts as shown in Tables II and III. With no Smart Control, the homeowner must maintain power during an outage. Thus, a homeowner must significantly adjust their habits to ensure that critical appliances (Modem, Heat Pump, Refrigerators, or Freezers) are powered throughout the outage. If not, they risk running out of energy stored in the ESS. For the fairness of the comparison, it is assumed that for both the simulated baseline case and the agent test case, the household load does not change from regular operations during an outage, i.e., the homeowners are not taking any additional actions to maintain power beyond the rules described above. Thus, the actions carried out to manage the ESS and EV are the only ones used to meet the household load demands.

The performance of the trained RL agents is evaluated over the three-year period. It focuses on the cost of electricity, the average EV charge when connected, the average ESS charge, and the number of time steps without power. Three scenarios are examined. The first one, with no outages experienced, simulates perfect grid operation. The second one introduces 2 % probability of outage. Finally, a scenario with 50 % probability of losing power represents an area of a highly unreliable electricity grid due to extreme weather events or unreliable grid infrastructure.

The obtained values for the electricity cost and EV SoC for the scenario with perfect gird operation are shown in Table I. We can see a significant decrease in monthly electricity cost with minimal effect on the EV SoC when leaving for a trip in the case of agents with cost- and range-awareness goals. The most significant decrease is observed with the cost-awareness agent – its actions save an average of \$10 per month with a very negligible effect on the SoC when the EV is leaving, in practice a 90% SoC when leaving would only be significant before a long trip. Note that the outage-awareness agent sees a higher average energy bill when compared to the baseline case; it is because it focuses on keeping the ESS and EV SoC at as high a level as possible. The range-awareness agent performs as well as the range baseline case but with an average savings of \$4 per month.

Table II shows the performance for the same operations in a grid with a 2% chance of an outage. We see significant savings of \$15 for the cost-awareness agent during winter months. It shows that the agent efficiently manages the EV and ESS to reduce costs when there is lower solar generation in winter. Solar generation primarily drives cost savings in the summer months. The range-awareness agent performs similarly to the baseline case. The most significant impact of using the RL agent for this test case is observed when there is no power. With the agent, its value decreases to only 12.77% as opposed to 48.29% for the baseline case and 54.09% for the cost-awareness agent.

The last baseline case is for a grid showing an extreme uncertainty – 50 % probability of outage, Table III; it would illustrate a severe weather situation. From a cost and range perspective, all agents perform as expected. The outage-awareness agent can provide continuous power to the entire home for 87.23 % of the time the grid is out. The agent efficiently uses the time when the grid is up and takes advantage of the solar panels, ESS, and EV during outages to ensure power can be provided in case of an outage.

IV. Conclusion

The paper investigates using Reinforcement Learning (RL) algorithms to develop a human-centric home energy management system (HEMS). The findings from our simulation results, with the Stable-Baselines3 Masked PPO Algorithm, reveal the capabilities of RL agents in balancing such aspects as minimizing energy costs, ensuring an uninterrupted power supply, and maximizing the electric vehicle (EV) range.

The RL agents are trained on different scenarios with varying primary goals, such as cost optimization, uninterrupted energy supply during outages, and maximizing the EV range. The results, as depicted in Figs. 3a and 3b, demonstrate the agents' effectiveness in managing energy usage and storage, showing their superiority over traditional rule-based systems.

In scenarios with no power outages, our cost-awareness goal agent optimizes energy costs while maintaining EV readiness, making this operating mode perfect for day-to-day usage. The outage-awareness agent, although resulting in higher energy costs, ensured the preparedness of the ESS and EV for potential outages, highlighting its focus on providing needed power without interruptions. Finally, the agent with the range-awareness goal is an ideal operating model for occasions when more extensive EV trips are required. The outage-awareness agent's performance is particularly notable. In more challenging scenarios with a 2 % and 50 % chance of power outages, it has significantly reduced the time without power

TABLE I: Comparison results for simulated environments with $p = 0\,\%$ outage probability.

Test Case	Winter Monthly Cost ($)		Summer Monthly Cost ($)		Avg. EV SoC@leaving (%)	
	Baseline	Agent	Baseline	Agent	Baseline	Agent
Cost-	133.71±31.79	118.83±27.48	52.44±21.99	51.57±20.12	100.0±1.0	88.0±9.0
Range-	141.04±32.15	136.12±29.18	59.64±21.76	62.18±20.49	100.0±0.0	99.0±2.0
Outage-	126.06±29.68	154.52±29.02	52.50±20.55	80.41±20.36	100.0±0.0	99.0±3.0

TABLE II: Comparison results for simulated environments with $p = 2\,\%$ outage probability.

Test Case	Winter Cost ($)		Summer Cost ($)		Avg. EV SoC@leaving (%)		Avg ESS SoC@OutageStart (%)		Time w/o Power (%)	
	Baseline	Agent	Baseline	Agent	Baseline	Agent	Baseline	Agent	Baseline	Agent
Cost-	125.08±33.22	110.83±29.18	55.64±17.82	59.62±15.73	97.0±12.0	78.0 ±21.0	57.0±47.0	75.0±24.0	63.29	50.25
Range-	141.04±32.15	122.49±34.7	59.64±21.76	69.87±16.15	100.0± 5.0	96.0±14.0	57.0±47.0	70.0±26.0	66.22	54.39
Outage-	143.94±30.86	191.31±31.54	87.99±21.82	121.26±31.95	100.0±5.0	91.0±15.0	98.0±9.0	94.0±13.0	39.73	10.13

TABLE III: Comparison results for simulated environments with $p = 50\,\%$ outage probability.

Test Case	Winter Cost ($)		Summer Cost ($)		Avg. EV SoC@leaving (%)		Avg ESS SoC@OutageStart (%)		Time w/o Power (%)	
	Baseline	Agent	Baseline	Agent	Baseline	Agent	Baseline	Agent	Baseline	Agent
Cost-	120.97±28.31	104.57±23.14	62.12±18.77	57.09±14.35	98.0±8.0	75.0±23.0	64.0±42.0	75.0±25.0	63.16	54.09
Range-	141.04±32.15	122.28±27.94	59.64±21.76	68.6±14.73	100.0±3.0	95.0±13.0	64.0±42.0	73.0±26.0	64.39	54.20
Outage-	141.83±29.89	192.01±28.89	82.45±18.07	134.55±24.36	100.0±3.0	88.0±20.0	97.0±12.0	92.0±15.0	48.29	12.77

by efficiently managing the EV and ESS. This capability is crucial in areas with unstable power grids or prone to extreme weather events, helping provide backup emergency power to homeowners.

The agents' flexibility in utilizing both the EV and ESS to meet various goals based on homeowner preferences further emphasizes the potential of RL in smart home energy management. This flexibility was evident in all tested scenarios, ranging from perfect grid operation to extreme grid instability. Yet, the research also highlights areas for future development. While RL agents offer significant improvements over rule-based systems, further refinement is needed to fully capture the complexity of real-world scenarios and integrate human habits, power usage and driving patterns, and homeowners' decision-making.

References

[1] Z. Wang and T. Hong, "Reinforcement learning for building controls: The opportunities and challenges," *Applied Energy*, vol. 269, p. 115036, Jul. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261920305481

[2] Q. Fu, Z. Han, J. Chen, Y. Lu, H. Wu, and Y. Wang, "Applications of reinforcement learning for building energy efficiency control: A review," *Journal of Building Engineering*, vol. 50, p. 104165, Jun. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352710222001784

[3] J. Arroyo, C. Manna, F. Spiessens, and L. Helsen, "Reinforced model predictive control (RL-MPC) for building energy management," *Applied Energy*, vol. 309, p. 118346, Mar. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0306261921015932

[4] T. A. Nakabi and P. Toivanen, "Deep reinforcement learning for energy management in a microgrid with flexible demand," *Sustainable Energy, Grids and Networks*, vol. 25, p. 100413, Mar. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352467720303441

[5] K.-b. Kwon and H. Zhu, "Reinforcement Learning Based Optimal Battery Control Under Cycle-based Degradation Cost," Jun. 2022, arXiv:2108.02374 [math]. [Online]. Available: http://arxiv.org/abs/2108.02374

[6] J. Cao, D. Harrold, Z. Fan, T. Morstyn, D. Healey, and K. Li, "Deep Reinforcement Learning-Based Energy Storage Arbitrage With Accurate Lithium-Ion Battery Degradation Model," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4513–4521, Sep. 2020, conference Name: IEEE Transactions on Smart Grid.

[7] N. S. Pearre and H. Ribberink, "Review of research on V2X technologies, strategies, and operations," *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 61–70, May 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032119300516

[8] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, and T. Jiang, "Deep Reinforcement Learning for Smart Home Energy Management," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2751–2762, Apr. 2020, arXiv:1909.10165 [cs, eess]. [Online]. Available: http://arxiv.org/abs/1909.10165

[9] A. Forootani, M. Rastegar, and M. Jooshaki, "An Advanced Satisfaction-Based Home Energy Management System Using Deep Reinforcement Learning," *IEEE Access*, vol. 10, pp. 47 896–47 905, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9766361/

[10] P. Lissa, C. Deane, M. Schukat, F. Seri, M. Keane, and E. Barrett, "Deep reinforcement learning for home energy management system control," *Energy and AI*, vol. 3, p. 100043, Mar. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2666546820300434

[11] H. Ding, Y. Xu, B. Chew Si Hao, Q. Li, and A. Lentzakis, "A safe reinforcement learning approach for multi-energy management of smart home," *Electric Power Systems Research*, vol. 210, p. 108120, Sep. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378779622003443

[12] Y. Ye, D. Qiu, X. Wu, G. Strbac, and J. Ward, "Model-Free Real-Time Autonomous Control for a Residential Multi-Energy System Using Deep Reinforcement Learning," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3068–3082, Jul. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9016168/

[13] Z. Tchir, M. Z. Reformat, and P. Musilek, "Home Energy Management with V2X Capability using Reinforcement Learning," in *2023 IEEE Conference on Artificial Intelligence (CAI)*. Santa Clara, CA, USA: IEEE, Jun. 2023, pp. 89–91. [Online]. Available: https://ieeexplore.ieee.org/document/10195059/

[14] S. DANE, "30 Years of European Solar Generation." [Online]. Available: https://www.kaggle.com/datasets/sohier/30-years-of-european-solar-generation

[15] U. M. LEARNING, "Household Electric Power Consumption." [Online]. Available: https://www.kaggle.com/datasets/uciml/electric-power-consumption-data-set

[16] "Electricity rates | Ontario Energy Board." [Online]. Available: https://www.oeb.ca/consumer-information-and-protection/electricity-rates

[17] H. Schwartz, "The computer simulation of automobile use patterns for defining battery requirements for electric cars," *IEEE Trans. Veh. Technol.*, vol. 26, no. 2, pp. 118–122, May 1977. [Online]. Available: http://ieeexplore.ieee.org/document/1622367/

[18] S. Huang and S. Ontañón, "A Closer Look at Invalid Action Masking in Policy Gradient Algorithms," *FLAIRS*, vol. 35,

May 2022, arXiv:2006.14171 [cs, stat]. [Online]. Available: http://arxiv.org/abs/2006.14171