E3S Web of Conferences **716**, 06001 (2026) https://doi.org/10.1051/e3sconf/202671606001
IAQVEC 2026

Human-AI Interaction in Large Language Model-Integrated Building Energy Management System: User Prompt Strategy

Wooyoung Jung[1*], *Prosper* Babon-Ayeng[1], and *Kahyun* Jeon[2]

[1]Civil and Architectural Engineering and Mechanics, The University of Arizona, 85719 Tucson, Arizona, USA
[2]Civil, Architectural, and Environmental Engineering, Illinois Institute of Technology, 60616 Chicago, Illinois, USA

Abstract. This study investigates how users formulate and utilize prompts when interacting with large language model (LLM)-integrated building energy management system (BEMS), with the objective of understanding their interaction strategies. Prior studies often relied on (1) rigid or pre-scripted user input interfaces in BEMS due to the limited capabilities of early language models or (2) posed single-turn user prompts to evaluate the performance of LLM-integrated BEMS. To address these limitations, this study conducted an experiment where participants were tasked with identifying five energy-saving behavioral changes while interacting with an LLM-integrated BEMS. This experimental setup enabled us to examine how users would naturally formulate prompts throughout their interactions with the system. In total, we collected 156 prompts (comprising 252 sentences) from 20 participants and classified them via six prompt strategies derived from decision support systems (DSS) theory. Our key takeaways were the following: (1) participants intuitively understood core principles of DSS, such as goal articulation, information seeking, and solution evaluation; (2) most participants began their interactions with GPT by directly articulating their goals, without offering any analytical guidance, showing a strong reliance on GPT's analytical capabilities; (3) participants were less engaged with solution evaluation, alternative seeking, and decision-making support, possibly owing to the domain knowledge required for such strategies. This study provides insights into the human side of human-AI collaboration by analyzing the content in prompts, demonstrating how users would communicate their needs and expectations using LLMs for building energy management tasks.

Keywords. Building Energy Management System, Large Language Model, User Prompt Strategy, Human-AI Collaboration, Decision Support Systems Theory

1 Introduction

Building energy – accounting for approximately 27.6% of total energy consumption in the U.S. [1] – is a substantial resource in modern life, essential for daily living, maintaining comfort, safeguarding health, and supporting productivity. Despite being the primary agents of energy usage within buildings, occupants are often unaware of how their everyday actions influence energy ecosystems. In most cases, the only feedback they receive is through periodic energy bills, which reflect aggregated monetary outcomes and lack both granularity and immediacy. This limited form of feedback often fails to provide timely and actionable insights and constraining occupants' ability to make informed decisions that balance energy use with comfort and financial considerations.

Efforts to digitize energy consumption through smart meters – which have reached over 80% penetration in the U.S. [2] – and smart plugs have made building energy data more accessible. Consequently, utilities offer mobile apps for customers to monitor their energy use during on- and off-peak hours daily, and monthly [3], [4]. However, this surface-level energy information is unlikely to serve as a meaningful tool for educating users or fostering deeper engagement with energy-saving interventions.

Research suggests that moving beyond surface-level information using personalized eco-feedback and energy-saving nudges can promote behavioral change [5], [6]. Yet, no effective systems have emerged that can communicate such information in ways tailored to users' differing energy literacy levels and diverse questions or concerns in practice. Several studies integrated graphical or voice user interfaces, designed to deliver adaptive eco-feedback or nudges; however, their rigid, pre-scripted user interfaces and limited capabilities in natural language processing and contextual adaptivity inhibited effective user engagement [7], [8], [9].

Integration of large language models (LLMs) into building energy management systems (BEMSs) can be a game changer, addressing this limited human-building interaction capabilities. LLMs can potentially facilitate the development of an effective, interactive tool that

* Corresponding author: wooyoung@arizona.edu

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

delivers personalized, real-time, data-driven feedback due to their capabilities in understanding and generating human-like language, as presented in Figure 1.

Recent studies revealed this potential by evaluating a broad range of potential user prompts, covering various themes such as energy analysis, device control, user preference reflection, and device scheduling (e.g., What is the energy use today? Can you adjust a setpoint?) [10], [11], [12]. However, their experimental setups were constrained to single-turn prompt-response trials, which may not fully capture real-world interaction patterns, where occupants typically engage in multi-turn interactions to iteratively accomplish their goals [13]. In other words, understanding multi-turn user prompt usage and strategies in human-AI collaboration, particularly within the context of BEMS, is essential for uncovering diverse needs and expectations users have when engaging with AI tools.

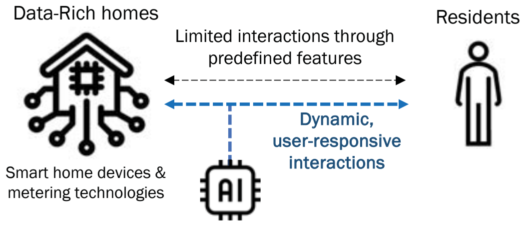

Figure 1. Envisioned human-AI interaction for building energy management.

Accordingly, this study aims to address the following research question:

What types of prompts do users employ when interacting with LLM-integrated BEMS to support their decision-making process?

By addressing this question, this study provides insights into the user side of human-AI collaboration in the context of BEMS, uncovering how users communicate their needs and expectations naturally when engaging with LLMs for energy management tasks. This knowledge can inform the development of LLMs that are better aligned with user intentions and energy-related goals, thereby enabling the refinement of general-purpose models into more specialized, user-responsive applications tailored for BEMS contexts.

2 Methodology

2.1 Data Collection: Experimental Study

We conducted an experimental study in which participants interacted with an LLM-integrated BEMS. Specifically, participants received instructions, outlining the goal of the experiment: identifying the top five behavioral changes to reduce an energy bill of a household in Austin, Texas. To encourage realistic engagement, participants were asked to assume the role of the household's occupant and make decisions as if the energy bill were their own. Then, the experimenter pre-loaded the GPT conversation with an initial prompt containing the appliance-level home energy data (Table 1) and time-of-use (TOU) pricing data (on-peak hours: 26.1¢/kWh from 2 to 8 pm; off-peak hours: 9.31¢/kWh). With this data as context, GPT (OpenAI GPT-4o) could reference the household's actual energy use patterns when responding to queries, functioning as an LLM-integrated BEMS effectively.

Participants then began their interactions freely with GPT through the basic prompting window without using advanced tools, like deep research or web search provided by OpenAI. They were given sufficient time to interact freely with GPT without imposed turn limits, allowing natural, multi-turn dialogue to unfold. When, participants concluded the experiment, all conversation history data were automatically recorded in the user interface and then manually exported to Word documents to keep their formatting and content structure.

This study was approved by the University of Arizona institutional review board in January 2025 as a minimal-risk study (STUDY00005726). The experimental studies were conducted from January 2025 to February, 2025.

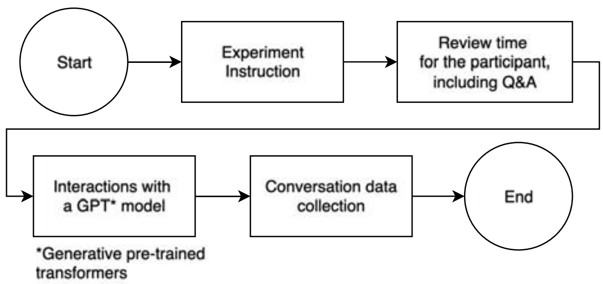

Figure 2. Experimental study's flowchart

Table 1. Details of appliance-level home energy data, obtained through [14].

Attribute	Details
Location	Austin, Texas
Appliances	Heating, ventilation, and air-conditioning (HVAC) unit; microwave; oven 1; oven 2; dishwasher; electric vehicle charger; washing machine; clothes dryer; pool pump; cooktop; refrigerator; electric water heater; bathroom; bedroom; living room; and kitchen
Data interval	15 minutes
Unit	kW
Timeframe	July and August, 2023

 https://doi.org/10.1051/e3sconf/202671606001

2.1.1 Participant Demographic

We had a total of 20 participants whose demographic data are illustrated in Figure 3. Most participants were between 25 and 34 years old, and a majority identified as male. The sample showed a relatively balanced distribution in terms of educational attainment. Also, participants represented a range of ethnic backgrounds. Overall, the participant pool reflected a diverse young adult population, appropriate for exploring interactions with emerging AI tools.

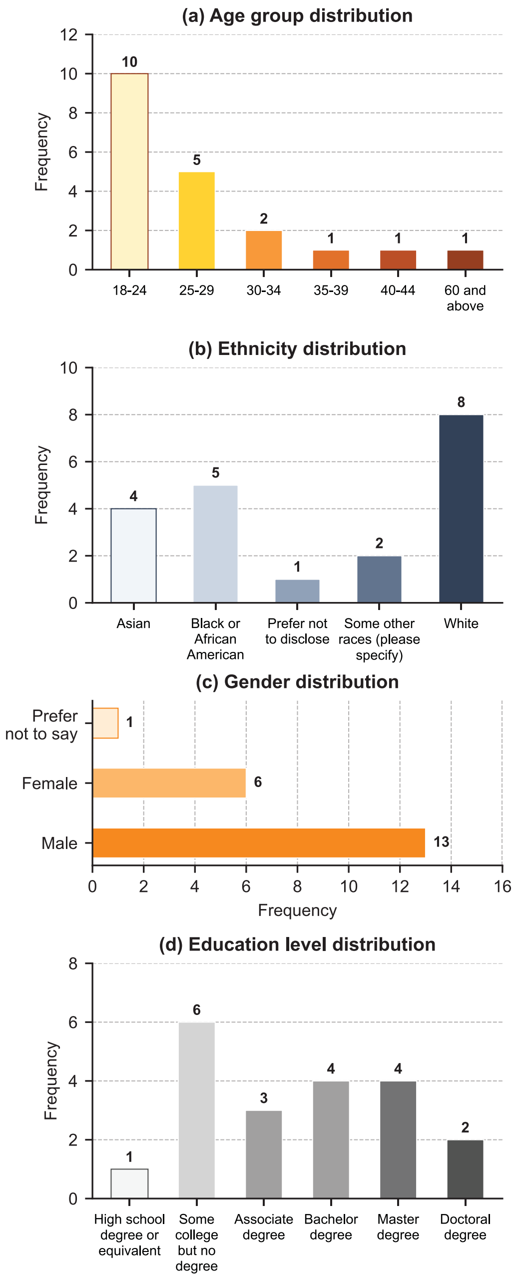

Figure 3. Participant demographic data

2.2 Prompt Categorization and Analysis

Figure 4 presents the analysis framework employed in this study following the data collection process. We extracted participant prompts and supplied GPT (OpenAI GPT4o-mini) with a set of prompt strategies identified through a literature review, allowing the model to classify each prompt accordingly. These initial classifications were manually reviewed and validated for accuracy. Then, we conducted an analysis of the categorized prompts to identify user strategies and conversational patterns.

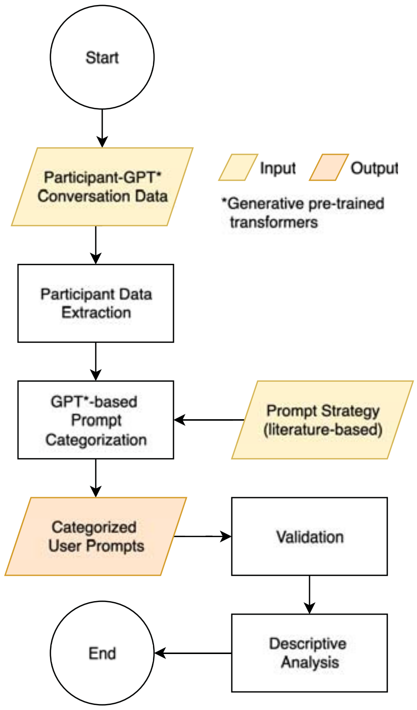

Figure 4. Human-AI interaction data analysis framework

2.2.1 Prompt Extraction and Categorization

We defined six prompt strategies, as detailed in Table 2, for systematically classifying user prompts based on the principles drawn from decision support systems (DSS) theory [13]. DSS theory focuses on the design of computer-based systems that support decision-making activities by providing relevant information, modeling tools, and user interfaces tailored to varying user needs and decision contexts. Grounding our classification in this theory allowed us to capture the range of ways users seek assistance, communicate constraints, and interact with GPT to support energy-saving decisions.

Figure 5 shows the prompt used to instruct GPT to classify each sentence using six prompt strategies. These strategies, prioritized using chain-of-thought instructions, were derived from Table 2. GPT was allowed to assign multiple strategies to each sentence, reflecting the multifaceted nature of user intents. In addition, we requested the GPT model to quantitatively score the confidence of its classification based on pre-defined scoring guidelines. It is noteworthy that the confidence scores were not presented in the Results section because they were primarily used as an internal reference to guide

GPT and to support manual validation and were not intended for direct interpretation or comparison.

Table 2. Six prompt strategies defined through DSS theory

Prompt Strategy	Definitions
Information seeking	Requesting specific energy-related information, appliance details, usage patterns, cost savings estimates, or technical explanations about energy consumption or energy consuming appliances
Constraint communication	Expressing limitations about acceptable lifestyle changes, comfort preferences, budget constraints, technology restrictions, or routine boundaries
Goal articulation	Stating specific energy-saving objectives, desired bill reduction amounts, or target outcomes (e.g., find five behavioral changes, reduce summer energy costs)
Alternative seeking	Requesting different energy-saving options, additional behavioral changes beyond initial suggestions, or alternative approaches to reduce energy consumption
Solution evaluation	Providing substantive evaluation of energy-saving suggestions, including detailed reactions about feasibility, effectiveness, or appropriateness of proposed behavioral changes, explaining why suggestions would or would not work
Decision-making support	Seeking help to choose between energy-saving strategies, requesting cost-benefit comparisons, or asking for prioritization of behavioral changes

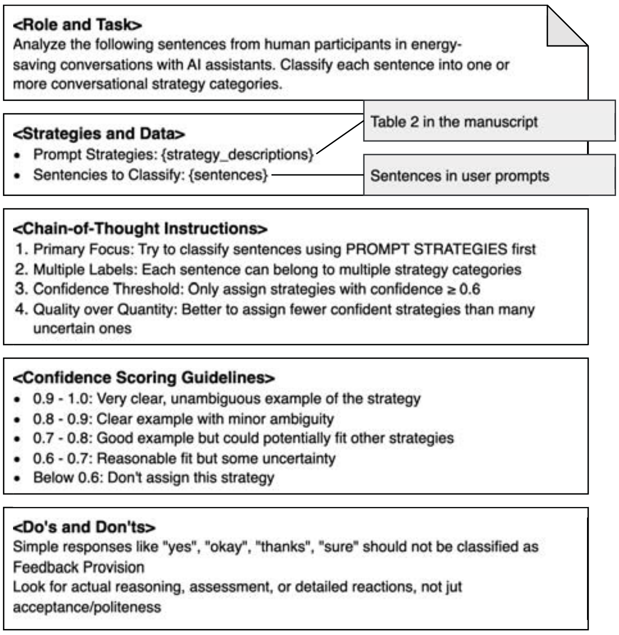

Figure 5. Prompt for categorizing prompts based on six pre-defined Strategies.

Following this GPT-based categorization process, we conducted a manual review to validate the accuracy of the classifications, ensuring that the assigned strategies appropriately reflected the intent and content of each user prompt. A total of 13 sentences (5.2%) were re-classified through this manual validation process.

2.2.2 Categorized prompt analysis

Lastly, we conducted an analysis of categorized prompts, attempting to understand the underlying patterns of participant intent, interaction style, and strategic behavior when engaging with LLMs for energy-related decision-making.

3 Results

We collected a total of 156 prompts, comprising 252 sentences, from 20 participants through our experiment and they were categorized as presented in Table 3.

Table 3. Prompt strategies identified from user prompts

Prompt Strategy	Number of sentences (%)	Number of participants (%)
Information seeking	93 (36.9%)	19 (95%)
Goal articulation	53 (21.0%)	18 (90%)
Constraint communication	44 (17.4%)	14 (70%)
Solution evaluation	44 (17.4%)	12 (60%)
Decision-making support	20 (7.9%)	8 (40%)
Alternative seeking	19 (7.5%)	11 (55%)

Percentages sum to more than 100% because sentences could be assigned to multiple strategies.

Information seeking is the most dominant strategy that participants used, highlighting their tendency to rely on GPT for acquiring specific details, explanations, or contextual data (e.g., climate) to guide their understanding and decision-making. When 'energy saving' information was requested, GPT leveraged the provided home energy and TOU rate data to calculate potential energy savings and suggested targeted behavioral changes tailored to participants' conditions. This widespread use reveals that participants viewed GPT as a valuable informational resource, capable of filling gaps in their knowledge in finalizing their decisions. Examples include (prompts are italicized when quoted in this manuscript. Only typos were corrected, while preserving participants' original phrasing and intent): "How is the morning temperature during summer in Austin, Texas?" "So, let's say I pre-cooled the house before 5am down to 70F, how long will this last if I don't open any windows throughout the day?" "Could you provide a detailed action plan and total energy saving analysis if all the discussed suggestions were implemented?"

The majority of participants began their interactions by asking GPT explicitly to recommend five behavioral changes that could lead to energy savings, demonstrating a clear articulation of their primary objective from the outset. A representative example was "Please find five behavioral changes that could reduce my energy bills," mirroring the experimental task instructions. This request was paraphrased by participants in various ways (tones, specificity, and phrasing), yet consistently conveying the same underlying objective. This pattern indicates participants' reliance on GPT as an assistant capable of synthesizing complex data and offering actionable guidance, even in the absence of clearly defined evaluation criteria such as energy consumption metrics or appliance usage frequency. This goal articulation strategy without specific analysis guidance may provide room for GPT to freely analyze the datasets and generate recommendations, but it also raises concerns about the consistency, reliability, and transparency of the outputs, when users lack the expertise to critically assess the model's responses.

Constraint communication was the third dominant strategy used by participants. In this strategy, participants conveyed specific limitations, preferences, or contextual constraints that they wanted GPT to consider when generating recommendations. These constraints included factors including comfort preferences, appliance usage habits, or time availability. Examples were "I cannot shift kitchen use to off peak hours since I work that time and need to use the kitchen during peak hours," and "I don't think I can reduce the frequency of cooking and dishing or laundry". Such constraint communication helped tailor the responses to align more closely with personal contexts and objectives, thereby enhancing the relevance and applicability of the recommendations provided by GPT.

Solution evaluation was a strategy that demonstrated participants' active engagement in refining behavioral change recommendations. This strategy revealed participants' willingness to critically evaluate suggestions and offer constructive feedback to arrive at the most contextually appropriate, feasible, and impactful actions. Notably, solution evaluation often required a degree of domain knowledge, especially when GPT's recommendations involved technical components or assumptions that needed to be validated or adjusted by the user. Examples were "I want you to combine those 1,2, and 3 options that you just provided and make it as one option," and "Keep this as my option 1 and I will go over others to select."

Decision-making support was a less frequent, but noteworthy strategy found in participant prompts, reflecting users' desire for assistance in assessing and choosing energy-saving suggestions. This strategy suggests that some participants approached the interaction with decision uncertainty and sought GPT's response to weigh trade-offs between comfort, lifestyle, and savings. Examples include: "If this is the way to save the money without change my lifestyle, I would go with this," and "Assuming in the summer when I sleep, I turned off the HVAC or set the HVAC to 3 degrees warmer than usual would that lead to extra savings." These prompts demonstrated how users looked to GPT for validation or optimization of their proposed strategies.

Alternative-seeking was another strategy that demonstrated users' reliance on GPT's analytical capabilities when initial recommendations were misaligned with their preferences, constraints, or feasibility. This behavior highlights participants' desire to explore multiple pathways and assess trade-offs between various energy-saving actions, treating GPT as a collaborative partner capable of generating refined alternatives in response to evolving criteria or dissatisfaction with prior outputs. Examples were "for number 4 suggest something different, these changes should not mean I can use new technology," "Is there any other recommendation in that case?"

4 Discussion

The diversity and sequencing of these prompt strategies, naturally employed by participants to support their decision-making, reinforce the potential of LLM-integrated BEMS. Users had an intuitive understanding of the envisioned system's potential, applying DSS principles like goal articulation, information seeking, and solution evaluation without explicit instruction or prior training by the experimenter. However, we also noticed that participants did not fully leverage the complete range of DSS principles; consequently, their interactions tended to underutilize the system's capacity to support structured decision-making. Participants intuitively engaged in goal articulation and information seeking; however, more advanced DSS-informed strategies, such as solution evaluation, decision-making support, were used less frequently or consistently. This gap may suggest that such strategies may require domain knowledge as a prerequisite; however, it also highlights an opportunity to improve user outcomes by incorporating interaction scaffolds that encourage deeper more structured engagement, thereby enabling user to leverage the full spectrum of DSS capabilities regardless of their expertise level.

In addition, the diversity in prompt strategies reflects variations in energy literacy, AI literacy, engagement styles, and decision-making needs among participants. This underscores the importance of delivering more personalized responses that consider each user's unique characteristics and interaction patterns. However, in the current structure of human-AI interactions, especially those facilitated through chatbots, AI systems tend to adopt a reactive stance, primarily responding to user prompts rather than guiding the conversation proactively. In our experiment, we saw that GPT often concluded its responses with optional suggestions or follow-up topics after addressing user queries. However, the participant retained full control over the conversational direction; as a result, GPT's role remained passive. Specifically, GPT lacked the initiative or authority to request additional user information that could improve the relevance of its recommendations. Consequently, the burden of sustaining effective interaction fell largely on users, ultimately limiting the AI's ability to deliver personalized, context-aware guidance in energy management scenarios.

5 Conclusion

This study shared insights into how users would formulate and use prompts in the context of LLM-integrated BEMS. In our experimental study, users initiated conversations in chatbot-based environments, and their prompts revealed a range of communication strategies aimed at obtaining energy-saving recommendations and supporting decision-making. We asked participants to identify top five energy saving behavioral changes in an LLM-integrated BEMS environment using GPT-4o and categorized their prompts and sentences using six pre-defined strategies, drawn from DSS theory.

Key takeaways were the following: (1) participants intuitively applied core principles of DSS, including goal articulation and information seeking; (2) most participants demonstrated a strong reliance on GPT's analytical capabilities by allowing it to independently generate behavioral recommendations without offering specific evaluative criteria; and (3) participants engaged less with strategies such as solution evaluation, alternative seeking, and decision-making support, possibly owing to the domain knowledge or AI literacy needed to effectively implement these more advanced decision-making processes.

This study provides valuable insight into how users of LLM-integrated BEMS would interact with AI systems to support their decision-making processes. Understanding these interaction patterns is essential for both guiding users on how to more effectively engage with AI and for engineering more responsive, adaptive AI systems that accommodate diverse levels of user knowledge, promote structured decision-making, and ultimately enhance the effectiveness of AI-assisted energy management.

This study has a few limitations. First, the sample size was limited to 20 participants, which may constrain the generalizability of the findings. A larger participant pool is required to enable more statistical analyses and to capture a broader range of user communicative and prompt strategies. Second, we primarily conducted frequency and distribution of prompt strategies without examining how these prompts evolved sequentially within individual conversations. Understanding the temporal and sequential patterns, such as whether users progressively shift from information seeking to solution evaluation, could reveal how decision-making strategies develop over the course of an interaction. Third, our use of a predefined DSS-based framework, while theoretically grounded, may overlook emergent interaction patterns that do not align with the established categories. Fourth, the controlled experiments may not fully replicate real-world conditions, where users would interact with their own household data and have a personal financial stake in the outcomes. This difference in motivation and familiarity with the data may influence prompt strategies.

As for our future research, we plan to address these limitations and will investigate a prompting template, designed to unlock the full potential of general-purpose GPT in the context of BEMS. This template can guide users to formulate more structured, context-aware prompts that align closely with core DSS principles. By scaffolding user input, the template may encourage deeper engagement with the system, facilitate more tailored recommendations, and support a wider range of decision-making strategies regardless of the user's prior domain knowledge or AI experience. Also, we will complement our top-down, theory-driven classification with an inductive approach, such as grounded theory, to identify novel interaction patterns or concepts that may not be captured by the DSS framework. This dual approach could help address potential overlap among categories and uncover richer insights into user-AI interaction behavior. In addition, we plan to investigate sequential patterns in user-LLM interactions, examining how prompt strategies transition and evolve across conversation turns. Such analysis could uncover common interaction trajectories and inform the design of LLM-integrated BEMS that proactively support users at different stages of their decision-making process. Lastly, we will aim to enhance ecological validity by deploying LLM-integrated BEMS in real household settings, where occupants interact with their own energy data over extended periods, enabling observation of authentic interaction patterns.

This research was partially supported by the Salt River Project under grant SRP-UA 08 24-25 and the National Science Foundation under grant #2519054. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Lastly, the authors would like to appreciate all participants in the experiment.

References

1. U.S. Energy Information Administration. How much energy is consumed in U.S. buildings? 2024 [cited 2024 12/11]; Available: https://www.eia.gov/tools/faqs/faq.php?id=86&t=1
2. J.S. Jones. Smart meter penetration surpasses 80% in North America. 2024 [cited 2025 07/22]; Available: https://www.renewableenergyworld.com/news/smart-meter-penetration-surpasses-80-in-north-america/

3. Tucson Electric Power (TEP). TEP mobile app. n.d. [cited 2024 12/12]; Available: https://www.tep.com/mobileapp/
4. Salt River Project (SRP). SRP mobile apps. n.d. [cited 2024 12/24]; Available: https://www.srpnet.com/account/mobile-apps
5. R.K. Jain, J.E. Taylor, and G. Peschiera. (2012). Assessing eco-feedback interface usage and design to drive energy efficiency in buildings. Energy and Buildings. **48**: p. 8–17. https://doi.org/10.1016/j.enbuild.2011.12.033
6. Y.A.A. Strengers, Designing eco-feedback systems for everyday life, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. (2011), Association for Computing Machinery: Vancouver, BC, Canada. p. 2135–2144.
7. H. Kim, S. Ham, M. Promann, H. Devarapalli, G. Bihani, T. Ringenberg, V. Kwarteng, I. Bilionis, J.E. Braun, J.T. Rayz, L. Raymond, T. Reimer, and P. Karava. (2022). MySmartE – An eco-feedback and gaming platform to promote energy conserving thermostat-adjustment behaviors in multi-unit residential buildings. Building and Environment. **221**: p. 109252. https://doi.org/10.1016/j.buildenv.2022.109252
8. H. Kim, I. Bilionis, P. Karava, and J.E. Braun. (2023). Human decision making during eco-feedback intervention in smart and connected energy-aware communities. Energy and Buildings. **278**: p. 112627. https://doi.org/10.1016/j.enbuild.2022.112627
9. M.L. Chalal, B. Medjdoub, N. Bezai, R. Bull, and M. Zune. (2022). Visualisation in energy eco-feedback systems: A systematic review of good practice. Renewable and Sustainable Energy Reviews. **162**: p. 112447. https://doi.org/10.1016/j.rser.2022.112447
10. T. He and F. Jazizadeh. LLM-based Building Energy Management Assistants. in i3ce 2024. (2024). Pittsburgh, PA, USA.
11. T. He and F. Jazizadeh. (2025). Context-aware LLM-based AI Agents for Human-centered Energy Management Systems in Smart Buildings. arXiv preprint arXiv:2512.25055.
12. J. Rey-Jouanchicot, A. Bottaro, E. Campo, J.-L. Bouraoui, N. Vigouroux, and F. Vella. (2024). Leveraging Large Language Models for enhanced personalised user experience in Smart Homes. arXiv preprint arXiv:2407.12024.
13. Y. Li, X. Shen, X. Yao, X. Ding, Y. Miao, R. Krishnan, and R. Padman. (2025). Beyond single-turn: A survey on multi-turn interactions with large language models. arXiv preprint arXiv:2504.04717.
14. Pecan Street Inc. Better tech, better decisions, a better world – we're making it happen. 2025 [cited 2025 07/07]; Available: https://www.pecanstreet.org/