

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Gamifying Sustainability: Harnessing AI and IoT for Human-Centric Energy Efficiency

Rafael Neves Teixeira



Mestrado em Engenharia Informática e Computação

Supervisor: Prof. Bruno Lima

Second Supervisor: Filipe Sousa

July 7, 2026

Gamifying Sustainability: Harnessing AI and IoT for Human-Centric Energy Efficiency

Rafael Neves Teixeira

Mestrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

President: Prof. João Carlos Viegas Martins Bispo

Examiner: Prof. Ângelo Manuel Rego e Silva Martins

Member: Prof. Bruno Miguel Carvalhido Lima

July 7, 2026

Resumo

O combate às alterações climáticas exige melhorias urgentes e contínuas na eficiência energética dos edifícios. Embora a proliferação das tecnologias da Internet das Coisas (IoT) tenha permitido a monitorização de energia em tempo real com grande granularidade, a apresentação de dados brutos através de painéis estáticos raramente promove mudanças de comportamento duradouras. Este desfasamento, conhecido como a lacuna entre intenção e ação, evidencia que, sem o envolvimento ativo do utilizador, o potencial de poupança de energia dos sistemas IoT permanece amplamente por concretizar. As soluções atuais dependem frequentemente de uma abordagem rígida e padronizada (*one-size-fits-all*), o que leva à fadiga de recompensas e à perda de interesse dos utilizadores ao longo do tempo.

Para colmatar esta lacuna, esta dissertação, desenvolvida em colaboração com a Fraunhofer Portugal AICOS, propõe o GreenShift: uma arquitetura de computação na borda (*edge-computing*) focada na privacidade, desenvolvida como uma integração personalizada para o Home Assistant. O GreenShift transforma a monitorização passiva de energia numa experiência envolvente e centrada no utilizador, integrando eco-feedback em tempo real, gamificação significativa e personalização impulsionada por Inteligência Artificial (IA). Para manter a motivação, o sistema adapta dinamicamente a dificuldade das tarefas de poupança de energia ao progresso do utilizador. Simultaneamente, um agente de Aprendizagem por Reforço com Feedback Humano (*Reinforcement Learning from Human Feedback*) ajusta as intervenções proativas aos contextos individuais. Ao modelar ativamente um Índice de Fadiga do utilizador, a IA regula de forma inteligente as notificações, prevenindo a exaustão digital enquanto maximiza a relevância de alertas acionáveis.

A eficácia da arquitetura foi validada empiricamente através de implementações em quatro ambientes reais distintos (dois laboratórios e duas casas inteligentes). Os resultados demonstram que a integração de gamificação dinâmica com IA adaptativa supera significativamente o eco-feedback estático tradicional, o qual se revelou altamente volátil. O envolvimento ativo com as mecânicas de gamificação gerou reduções de potência instantâneas substanciais, eliminando até quase 70% da carga base ativa em configurações localizadas, e estabilizou com sucesso o consumo de energia normalizado pelas condições meteorológicas. O estudo revelou ainda que as estratégias de gamificação devem estar estritamente alinhadas com a agência física do utilizador: os espaços localizados prosperaram com tarefas específicas para dispositivos, enquanto os espaços partilhados responderam melhor a comparações sociais normativas. Além disso, o agente de IA estabilizou com sucesso a sua política de tomada de decisão ao longo do tempo, bloqueando autonomamente cerca de 60% das notificações geradas para preservar a atenção do utilizador. Em suma, esta investigação valida que, quando a conservação de energia é enquadrada como uma experiência personalizada, ciente do contexto e intrinsecamente recompensadora, os ocupantes integram com sucesso hábitos sustentáveis nas suas rotinas diárias.

Palavras-chave: Aprendizagem por Reforço • Computação de Borda • Eco-feedback • Eficiência Energética • Gamificação • Internet das Coisas

Abstract

Addressing climate change requires urgent and sustained improvements in building energy efficiency. While the proliferation of Internet of Things (IoT) technologies has enabled granular real-time energy monitoring, presenting raw data through static dashboards often fails to foster lasting behavioural change. This disconnect, known as the intention-action gap, highlights that without active user engagement, the energy-saving potential of IoT systems remains largely unrealized. Current solutions frequently rely on a rigid "one-size-fits-all" approach, leading to reward fatigue and user disengagement over time.

To bridge this gap, this dissertation, developed in collaboration with Fraunhofer Portugal AICOS, proposes GreenShift: a privacy-preserving edge-computing architecture built as a custom Home Assistant component. GreenShift transforms passive energy monitoring into an engaging user-centric experience by integrating real-time eco-feedback, meaningful gamification and Artificial Intelligence (AI)-driven personalization. To maintain motivation, the system dynamically adapts the difficulty of energy-saving tasks to the user's progress. Concurrently, a Reinforcement Learning from Human Feedback agent tailors proactive interventions to individual contexts. By actively modelling a user's Fatigue Index, the AI intelligently throttles notifications, preventing digital burnout while maximizing the relevance of actionable alerts.

The architecture's efficacy was empirically validated through deployments in four distinct real-world settings (two laboratories and two smart homes). Results demonstrate that integrating dynamic gamification with adaptive AI significantly outperforms traditional static eco-feedback, which proved highly volatile. Active engagement with the gamified mechanics yielded substantial instantaneous power reductions, eliminating up to nearly 70% of the active baseload in localized setups, and successfully stabilized weather-normalized energy consumption. The study further revealed that gamification strategies must strictly align with the user's physical agency: localized settings thrived on specific device tasks, whereas shared spaces responded best to normative social comparisons. Furthermore, the AI agent successfully stabilized its decision-making policy over time, autonomously blocking approximately 60% of generated notifications to preserve user attention. Ultimately, this research validates that when energy conservation is framed as a personalized, context-aware and intrinsically rewarding experience, occupants successfully embed sustainable habits into their daily routines.

Keywords: Eco-feedback • Edge Computing • Energy Efficiency • Gamification • Internet of Things • Reinforcement Learning

The United Nations Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework for achieving a better and more sustainable future for all. It includes 17 goals (<https://sdgs.un.org/goals>) to address the world’s most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace and justice.

Because this dissertation focuses on the intersection of human behaviour, IoT and energy efficiency, the impact of the proposed architecture is specifically evaluated against three core goals:

SDG 7 Ensure access to affordable, reliable, sustainable and modern energy for all

SDG 12 Ensure sustainable consumption and production patterns

SDG 13 Take urgent action to combat climate change and its impacts

SDG	Target	Contribution	Performance Indicators and Metrics
7	7.3	The GreenShift integration actively reduces household and office power demand by bridging the intention-action gap through personalized AI-driven gamification.	Weather-normalized daily consumption reductions (E_{norm}) and mean instantaneous task power savings ($\Delta P_{task_savings}$).
12	12.2	Utilizing adaptive daily tasks and dynamic difficulty adjustments to build long-term sustainable consumption habits.	Daily task completion rates (R_{task}), self-reported habit formation and reduction in non-working hours power waste.
	12.8	Translating abstract energy data (kWh) into relatable disaggregated eco-feedback to increase energy literacy and reduce cognitive load.	Intrinsic Motivation Inventory (IMI) scores, sustained user engagement and proactive management of the user fatigue index ($I_{fatigue}$).
13	13.3	Providing real-time tangible feedback on avoided CO ₂ emissions (e.g., equivalent trees planted) to directly connect digital actions with physical climate impact.	Estimated cumulative CO ₂ avoided per user based on regional energy mix carbon intensity.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Bruno Lima, for his continuous guidance, patience and for always being available to help whenever I needed it. I am also deeply thankful to my co-supervisor, Filipe Sousa, and to Miguel Campos from Fraunhofer. Their invaluable tips, hands-on assistance and efforts in securing the necessary equipment were crucial to the development and success of this work. Additionally, I extend my appreciation to the reviewers for their time, constructive criticism and insightful feedback.

A special thanks goes to all the participants who volunteered their time to help me validate my solution. This research would not have been possible without your involvement.

Finally, on a personal note, I want to thank my family for their unwavering support and for providing me with the opportunities that allowed me to reach this point and be where I am today. To my girlfriend, Inês, thank you for your endless support, understanding and encouragement through every step of this journey.

Rafael Neves Teixeira

Declaration of honour

I declare that this dissertation is my own work and has not previously been submitted or assessed at this or any other institution. References to other authors (statements, ideas, thoughts), as well as the use of published works, strictly comply with the rules of attribution and are duly identified in the text and in the bibliographic references, in accordance with the applicable referencing standards. I am aware that the practice of plagiarism or self-plagiarism constitutes an academic offence, as established in the U.Porto Code of Ethics.

Rafael Neves Teixeira

Rafael Neves Teixeira

"Technology alone is not enough. It's technology married with the liberal arts, married with the humanities, that yields us the results that make our heart sing."

Steve Jobs

Contents

1	Introduction	1
1.1	Context & Motivation	1
1.2	Problem	2
1.3	Research Questions	3
1.4	Hypothesis	3
2	State of the Art	4
2.1	Review Methodology	4
2.1.1	Inclusion and Exclusion Criteria	4
2.1.2	Search Strategy and Paper Collection	5
2.1.3	Screening Process	7
2.1.4	Final Eligibility and Selection	7
2.1.5	Review Analytics	8
2.2	Occupant Behaviour as a Determinant of Building Performance	13
2.3	The Evolution of Eco-Feedback Strategies	14
2.4	From Traditional Bills to Real-Time IoT Feedback	15
2.4.1	The Evolution of Delivery Platforms	15
2.4.2	A Taxonomy of Feedback Methods	17
2.5	The Engagement Gap: Limitations of Feedback-Only Systems	19
2.5.1	Engagement Decay Over Time	19
2.5.2	Gaps Between Information and Action	19
2.5.3	Overriding Contextual Barriers	20
2.6	Gamification as an Engagement Strategy	20
2.6.1	Theoretical Foundations: Why It Works	21
2.6.2	Evidence from the Field	22
2.6.3	Limitations: Engagement Fatigue and the One-Size-Fits-All Problem . .	25
2.7	The New Frontier: AI-Driven Personalization and Adaptation	25
2.7.1	The Need for Adaptive Interventions	26
2.7.2	Algorithmic Enablers of Personalization	26
2.8	Advanced Architectural Frameworks for Sociotechnical Systems	28
2.8.1	Conceptual Architectures and Design Principles	29
2.8.2	From Concept to Actionable Recommendation	30
2.8.3	Digital Twins and Privacy-Preserving Edge Computing	30
2.8.4	Intelligent Scenario Orchestration	32
2.9	Empirical Validation: Real-World Case Studies	33
2.9.1	Closing the Intention-Action Gap: Verification via IoT	33
2.9.2	Optimizing Timing and Persuasion: The $(EM)^3$ Framework	34
2.9.3	Synthesizing AI and Gamification: Affective and Achievement Systems .	36

2.10	Research Gaps and Dissertation Positioning	39
2.10.1	Critical Limitations: Why Current Approaches Fail	39
2.10.2	Strategic Positioning: Harnessing High-Potential Innovations	41
3	GreenShift	44
3.1	Architectural Requirements and Design Principles	44
3.2	High-Level Architecture and Functional Flow	45
3.3	Physical Layer	46
3.4	Core Platform Layer (Home Assistant)	47
3.4.1	Integration Hub and State Machine	47
3.4.2	Database Strategy and Data Management	47
3.5	Intelligence Layer	48
3.5.1	The Integration Logic	48
3.5.2	Contextual Filtering: Environment and Working Hours Logic	49
3.5.3	The AI Engine	50
3.5.4	Adaptive Difficulty (Self-Efficacy)	55
3.6	User Interface Layer	55
3.6.1	System Onboarding and Device Configuration	55
3.6.2	Baseline Monitoring and Reporting	56
3.6.3	Consumption Visualization and Gamification	57
3.6.4	AI Notifications	60
3.6.5	Profile and System Configuration	62
3.7	Multilingual Support and Translation Mechanism	63
3.8	Data Reliability and Persistence Strategy	63
3.9	Quality Assurance and Software Testing	64
3.10	Design Decisions and Trade-offs	66
4	Validation	67
4.1	Validation Methodology	67
4.1.1	Experimental Topologies and Hardware Setup	67
4.1.2	Ethical Considerations	69
4.1.3	Experimental Procedure	70
4.1.4	Data Collection and Metrics	70
4.1.5	Data Analysis Strategy	73
4.2	Results and Discussion	75
4.2.1	Sample Characterization and Experimental Setup	76
4.2.2	Baseline Profiling and Shadow Learning (Phase 1)	77
4.2.3	Quantitative Impact on Energy Consumption	79
4.2.4	Gamification Efficacy and User Engagement	83
4.2.5	AI Agent Convergence and Nudge Interactions	88
4.2.6	Qualitative User Assessment	93
4.2.7	Summary of Validation Findings	96
5	Conclusion and Future Work	97
5.1	Synthesis of Findings and Hypothesis Validation	97
5.2	Main Contributions	98
5.3	Limitations and Future Work	99
	References	100

A	Qualitative User Assessment Questionnaire	112
A.1	Demographics and Context	112
A.2	System Usability Scale (SUS)	113
A.3	Gamification and Intrinsic Motivation	114
A.4	Intervention Impact and Cognitive Load	115
A.5	Behavioural Shifts and System Trust	116
A.6	Open-Ended Feedback	117
B	Scientific Article	119

List of Figures

2.1	Articles screening and filtering process (PRISMA flow diagram).	8
2.2	Distribution of selected articles by publication year.	9
2.3	Thematic evolution of article topics from 2019 to 2025.	10
2.4	Percentage of articles sourced from primary academic databases.	10
2.5	Articles related to the RQs presented.	11
2.6	Bibliometric analysis of keyword co-occurrence in the 114 selected articles. . . .	12
2.7	Relative contribution of occupant behaviour to total energy performance in buildings.	13
2.8	Normative feedback interface benchmarking daily room air conditioner consumption (top-right) against the group average.	16
2.9	Heat monitoring app.	16
2.10	Smart plugs used for energy awareness.	16
2.11	Disaggregated Feedback, isolating consumption by appliance category.	18
2.12	Normative Comparison interface showing the user's standing relative to other peers.	18
2.13	Framing strategy translating 100 kWh into relatable monetary and environmental equivalents.	19
2.14	The standard TPB model.	21
2.15	An extended TPB model for employee energy-saving behaviour, showing the integration of additional factors like habit, norms and comfort.	22
2.16	Dashboard interface displaying real-time and normative feedback, including social comparison with neighbouring peers and a goal-setting widget.	23
2.17	Interface for the gamified "pause hour" demand response challenge, showing (left) the sign-up prompt with a savings goal and (right) the results screen with social comparison.	24
2.18	App's main "Sustainable living" feedback screen, showing (left) social comparison feedback (normative) and (right) a general energy-saving tip.	24
2.19	CASER mobile app interfaces showing real-time consumption against the substitution threshold and predicted energy usage.	28
2.20	Conceptual Building Energy Management Systems (BEMS) architecture.	29
2.21	WebUI dashboard of the EMSS framework, showing analytics and recommendations.	30
2.22	3D model of the lab space created in "Sweet Home 3D" for the digital twin implementation.	31
2.23	Examples of (left) ecological and (right) economical persuasive recommendations sent to the user's mobile device.	31
2.24	System's recommendation generation process.	32
2.25	Floor plan illustrating four AI-driven scenarios for comfort and energy efficiency.	33
2.26	IoT sensors used for the BI-Tech platform.	34
2.27	BI-Tech platform interfaces.	34

2.28	(EM) ³ system workflow leveraging micro-moments for behavioural change. . . .	35
2.29	(EM) ³ mobile interface showing environmental data (left), consumption visualization with appliance control (middle) and generated recommendations (right). .	35
2.30	HomeMascot's system interface.	36
2.31	HomeMascot's finite state machine linking AI states to the mascot's mood. . . .	36
2.32	Gamification interfaces from the campus system, showing (a) the login page, (b) daily rewards page and (c) main screen.	37
2.33	Gamification interfaces from the campus system, showing (d, e) the shop page and (f) daily shared energy-saving tips page.	37
2.34	Gamification interfaces from the campus system, showing (g) the gameplay page, (h) experience points page and (i) level promotion page.	38
2.35	Greenely app interface.	39
3.1	High-level system architecture illustrating the data flow from physical sensors through the HA Core to the AI-driven Intelligence Layer and the user interface. .	45
3.2	System onboarding interface: (a) Temporal boundary configuration via work schedules for office environments; (b) User validation of auto-discovered sensor categorizations; (c) Spatial assignment of hardware to specific physical areas.	56
3.3	Baseline and reporting interface interfaces: (a) Status indicator explaining the initial 14-day monitoring phase; (b) The post-monitoring baseline report detailing historical consumption and targets.	57
3.4	Blocked interface tab displayed during the initial baseline phase, featuring a countdown to active intervention.	57
3.5	Detailed energy visualization interface, displaying aggregate metrics (total power load, cost and emission estimates) alongside granular area-by-area device consumption.	58
3.6	Main dashboard interface comparing current power usage against historical baselines, including area distributions and top consumers.	58
3.7	Daily Tasks interface featuring dynamic, AI-generated objectives, real-time status tracking and user feedback mechanisms for RL agent tuning.	59
3.8	Weekly Challenge interface displaying the collaborative community goal, real-time consumption status via a colour-coded gauge, financial savings estimates and a spatial breakdown of energy usage to guide user interventions.	60
3.9	Dedicated in-app Notifications interface for reviewing and interacting with AI-generated nudges.	61
3.10	Example of a received mobile push notification, demonstrating the absence of interactive buttons due to native platform limitations.	61
3.11	User profile dashboard displaying cumulative statistics, AI indices and tangible equivalents of environmental impact.	62
3.12	Settings interface for customizing reduction targets, financial preferences and managing the active sensor infrastructure.	63
4.1	Hardware topology for comprehensive power monitoring and environmental contextual sensing in the Fraunhofer laboratory.	68
4.2	Distributed hardware setup for appliance-level monitoring and contextual sensing in the FEUP laboratory.	69
4.3	Stacked bar chart showing the distribution of self-reported smart home familiarity (on a scale of 1 to 5) by deployment setting.	77

4.4	Evolution of max Q-value and detected anomalies during shadow learning across professional and residential environments.	78
4.5	Interrupted time-series analysis detailing the daily evolution of E_{norm} , illustrating the transition from passive visualization (Baseline) to active intervention (Active Phase).	79
4.6	Comparison of aggregate weather-normalized energy consumption (E_{norm}) between baseline and active phases across professional and residential environments.	81
4.7	Boxplots detailing the variance in working-hours power premium (ΔP_{WH}) between the active and baseline phases across professional settings.	82
4.8	Dual-axis charts correlating daily task completion rates (left axis) with the adaptive difficulty level (right axis) across professional and residential environments.	83
4.9	Intra-phase comparison of mean weather-normalized energy consumption (E_{norm}) clustered by active vs. passive engagement days.	85
4.10	RL convergence - Dual-axis charts illustrating the epsilon (ϵ) decay trajectory (left axis) against the episodic reward (right axis) across professional and residential environments.	89
4.11	Grouped scatter plot illustrating the distribution of suppressed notifications, crossing the calculated opportunity score with the user fatigue index ($I_{fatigue}$) across environments.	90
4.12	Notification block rate - Proportion of AI-generated nudges sent to the user versus those autonomously blocked by the system to prevent fatigue.	91
4.13	Notification outcomes by action type - Clustered bar charts detailing the count of ignored, rejected and accepted nudges.	92
4.14	User perceptions regarding system usability, gamified motivation and AI trustworthiness.	94
4.15	Assessment of the intervention's human impact: cognitive load (fatigue) versus long-term behavioural shifts.	95
A.1	Survey interface for demographic and baseline literacy data collection.	113
A.2	Standardized SUS forms presented to users.	114
A.3	Survey questions evaluating the gamification mechanics and intrinsic motivation.	115
A.4	Questionnaire items measuring notification intrusiveness and user fatigue.	116
A.5	Survey interfaces assessing long-term habit formation and perceived AI trustworthiness.	117
A.6	Form section for unstructured open-ended feedback.	118

List of Tables

2.1	Inclusion and exclusion criteria.	5
2.2	Number of search results per database for each query.	7
2.3	Comparison of energy reporting instruments.	15
2.4	Taxonomy of eco-feedback methods.	17
2.5	Design principles for sociotechnical BEMS.	29
2.6	Theoretical mapping of the gamification mechanics to their respective psychological drivers.	38
2.7	Challenges and mitigation strategies in gamified systems.	41
2.8	Overview of related studies and their specific influence on the proposed prototype.	43
3.1	Summary of quality assurance and test coverage.	65
3.2	Design decisions and trade-offs.	66
4.1	Definition of quantitative variables and metrics.	72
4.2	Definition of qualitative variables and metrics.	73
4.3	Summary of experimental environments and data collection.	76
4.4	Phase 1 baseline energy profiling.	77
4.5	Average instantaneous task savings relative to the active working-hours baseload.	86
4.6	Weekly challenge outcomes, target reductions and achieved savings per environment.	87

List of Acronyms

AC	Air Conditioner
ACM	Association for Computing Machinery
AI	Artificial Intelligence
API	Application Programming Interface
BEMS	Building Energy Management System
BLE	Bluetooth Low Energy
CDD	Cooling Degree Days
EMSS	Energy Management for Smart Spaces
FEUP	Faculty of Engineering University of Porto
GDPR	General Data Protection Regulation
HA	Home Assistant
HACS	Home Assistant Community Store
HCD	Human-Centric Design
HDD	Heating Degree Days
HVAC	Heating, Ventilation and Air Conditioning
IEA	International Energy Agency
IEEE	Institute of Electrical and Electronics Engineers
IHD	In-Home Display
IMI	Intrinsic Motivation Inventory
IoT	Internet of Things
LLM	Large Language Model
LSTM	Long Short-Term Memory
ML	Machine Learning
PDF	Portable Document Format
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QS	Query Sets
RL	Reinforcement Learning
RLHF	Reinforcement Learning from Human Feedback
RQ	Research Question
SDG	United Nations Sustainable Development Goal
SDT	Self-Determination Theory
SUS	System Usability Scale
TPB	Theory of Planned Behavior
URL	Uniform Resource Locator

Chapter 1

Introduction

Addressing climate change requires urgent and sustained improvements in energy efficiency. While the proliferation of Internet of Things (IoT) technologies has enabled sophisticated real-time energy monitoring, a critical challenge remains: providing users with raw data rarely leads to lasting behavioural change. The central problem addressed in this study is that without active engagement and motivation, the energy-saving potential of IoT systems is largely unrealized.

To bridge the gap between data presentation and sustained consumption reduction, this dissertation examines how eco-feedback, gamification and Artificial Intelligence (AI)-driven personalization can be integrated. The goal is to transform passive energy data into actionable engaging experiences that promote long-term efficiency. This research was developed in collaboration with Fraunhofer Portugal AICOS¹. This partnership not only provided the essential hardware infrastructure but also served as the primary environment to empirically evaluate the impact of the proposed interventions.

The remainder of this chapter outlines the specific context, motivation and objectives of the study. It closes by introducing the research questions (RQs) and hypothesis that shape the investigation, providing the foundation for the system design and evaluation presented in the chapters that follow.

1.1 Context & Motivation

The escalating urgency of climate change, coupled with the rising demand for sustainable energy practices, has intensified the pursuit of solutions that promote efficiency at both household and industrial domains [20, 53, 68, 86]. At the heart of these efforts lies energy efficiency, a central strategy for mitigating environmental impacts while simultaneously reducing operational and household costs [26, 53]. This approach directly aligns with the United Nations Sustainable Development Goals² (SDGs), leading governments and organizations worldwide to advance technologies that promote more responsible and sustainable consumption patterns.

¹<https://www.aicos.fraunhofer.pt/en/home.html>

²<https://sdgs.un.org/goals>

Among the most transformative of these technologies is the integration of IoT into energy systems. IoT has made it possible to monitor energy consumption at a highly granular level, offering an unmatched opportunity to optimize energy use and support sustainable lifestyles [22]. However, the effectiveness of these technologies does not depend on technical capacity alone. Evidence from behavioural science and interface design research shows that feedback and motivational strategies play a crucial role in shaping energy-related behaviour [51].

This dissertation is motivated by the need to bridge the gap between technical solutions for energy monitoring and behavioural strategies that ensure lasting user engagement and measurable efficiency gains. By combining eco-feedback with gamification techniques, abstract data can be transformed into actionable and rewarding experiences [89, 102]. In addition, AI holds the potential to personalize feedback by adapting recommendations to individual routines and lifestyles. However, this potential remains largely underexplored, as most existing solutions do not directly incorporate user behaviour into their eco-feedback mechanisms [16, 75].

1.2 Problem

Although IoT-based energy monitoring systems generate rich and valuable data, most existing solutions prioritize technical efficiency, such as accurate measurement and real-time tracking, over behavioural effectiveness, often overlooking the human factor. As a result, they often fail to support lasting user engagement or drive meaningful changes in everyday consumption habits [60].

Several factors contribute to this limitation. Dashboards frequently present complex or abstract data in ways that are not intuitive for non-technical users, leading to disengagement after initial use [22]. Traditional energy bills face an even greater challenge, as they provide only monthly usage totals, making it difficult for people to link specific daily behaviours to their overall energy consumption [22]. While IoT systems can provide more fine-grained data, this information often remains underutilized when presented in overly technical formats.

Attempts to address the issue through gamification have produced mixed results. While features such as points or badges can generate short-term interest, research indicates they rarely sustain engagement or lead to persistent energy savings, demonstrating a limited long-term impact on behaviour [100]. Many current systems rely on superficial reward structures without incorporating deeper motivational strategies, such as personalized goal-setting, nudges or social comparison, which are vital to foster lasting behavioural change [52].

Finally, personalization remains underexplored. Most current eco-feedback systems adopt a one-size-fits-all approach, even though users differ significantly in their routines, preferences and motivations. This lack of personalization reduces the effectiveness of such systems in promoting meaningful and sustained energy savings [68].

In summary, the central problem addressed in this dissertation is the absence of integrated IoT platforms that combine eco-feedback, meaningful gamification and AI-driven personalization to effectively engage users and support long-term energy-efficient behaviour.

1.3 Research Questions

To address this problem, the following RQs are proposed:

- **RQ1:** How effective are traditional eco-feedback strategies (e.g., historical comparison, normative feedback) in promoting energy-efficient behaviour in building contexts?
- **RQ2:** To what extent does the integration of meaningful gamification with real-time feedback lead to a greater reduction in energy consumption and sustained user engagement compared to traditional eco-feedback methods?
- **RQ3:** Which gamification strategies (e.g., challenges, rewards, social comparison) are most effective in sustaining user engagement and motivation within IoT energy-monitoring systems?
- **RQ4:** To what extent does AI-driven personalization provide an additive increase in energy reduction and sustained user engagement within a gamified eco-feedback system?

1.4 Hypothesis

Based on the problem previously identified and RQs formulated above, the following hypothesis is proposed as the central premise of this study:

An IoT platform integrating real-time eco-feedback, meaningful gamification and AI-driven personalization yields more significant and sustained energy savings and user engagement compared to traditional monitoring methods.

Chapter 2

State of the Art

The aim of this chapter is twofold. First, it describes the systematic process adopted to identify, collect and select relevant academic literature that forms the foundation of this dissertation. This process was designed to ensure transparency and reproducibility, allowing other researchers to follow the same steps and potentially reach a comparable set of studies.

Second, it provides a critical review of this selected literature, establishing the theoretical and empirical foundation for the proposed research. By situating the research within the existing academic landscape, this chapter moves from the broad societal challenge of energy consumption to the specific nuanced problems that this dissertation addresses. The review is structured to follow the story of technological and behavioural evolution: it begins by establishing human behaviour as a critical factor in energy consumption, traces the evolution of eco-feedback strategies, explores the engagement gap and gamification and finally converges on the current research frontier of AI-driven personalization.

2.1 Review Methodology

This literature review follows a systematic review methodology, adhering to the PRISMA¹ (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The process is structured in four stages: (1) inclusion and exclusion criteria definition, (2) paper collection, (3) screening and (4) final eligibility and selection.

2.1.1 Inclusion and Exclusion Criteria

The criteria in Table 2.1 were established to construct a focused, high-quality and modern review of the state of the art. The 2019 to 2025 publication window ensures the review is centred on the most current research, which is critical in rapidly advancing fields like AI personalization and IoT.

The thematic criteria were the most important for scoping this dissertation. To be included, a study had to focus on the intersection of technology and sustainable behaviour, such as IoT-enabled eco-feedback or gamification for energy efficiency. This was essential to deliberately filter

¹<https://www.prisma-statement.org/prisma-2020-checklist>

out the vast body of literature on generic hardware design, theoretical AI models or gamification mechanics that were not specifically applied to energy consumption or sustainable action in a building context.

Finally, to guarantee the academic rigour of the review, the selection was limited to complete and published works, such as journal articles and conference papers. This meant excluding works-in-progress, preprints, tutorials and workshop summaries, which may not have completed the validation process. The requirements for full-text access and English-language publication were practical constraints necessary for a feasible analysis.

Table 2.1: Inclusion and exclusion criteria.

#	Criteria
1	Published between 2019 and 2025 (inclusive)
2	Written in English
3	Full-text access available
4	Focus on at least one of the following topics: IoT-enabled eco-feedback, gamification for energy efficiency, behavioural change in energy consumption or AI-driven personalization for sustainability
5	Exclude studies unrelated to energy efficiency, consumption reduction or sustainable behaviours
6	Exclude works not applicable to building contexts
7	Exclude works limited to hardware design or generic gamification/AI not applied to sustainable behaviour
8	Not a duplicate
9	Not a retracted publication
10	Not a work-in-progress or preprint
11	Acceptable document types: journal article, conference paper, book chapter
12	Exclude seminars, workshop summaries, tutorials or lecture notes

2.1.2 Search Strategy and Paper Collection

The search was conducted across leading academic databases, including the Association for Computing Machinery (ACM) Digital Library², Institute of Electrical and Electronics Engineers (IEEE) Xplore³, Scopus⁴ and ScienceDirect⁵ (Elsevier). A consistent set of search operators was applied across all queries to balance precision and recall. Wildcards (e.g., gamif*, personaliz*) were employed to capture morphological variations of key terms. Proximity operators (e.g., energy W/2 saving) were used to enforce conceptual closeness between energy-related terms and their outcomes. Synonyms and related concepts were combined using the OR operator to ensure broad

²<https://dl.acm.org/>

³<https://ieeexplore.ieee.org>

⁴<https://www.scopus.com/>

⁵<https://www.sciencedirect.com/>

coverage across different disciplinary vocabularies. Spelling variants were accounted for with the ? operator (e.g., behavio?r, model?ing), thereby including both American and British English spelling conventions. For consistency, the queries are presented in Scopus syntax, however, adaptations were required when searching other databases, as not all supported the full range of operators available in Scopus.

To ensure a comprehensive retrieval of relevant literature, the search strategy was not limited to a single broad query. Instead, it was structured around four distinct Query Sets (QS), each designed to systematically address a specific facet of the research problem.

This multi-query approach allows for the capture of studies that directly test the dissertation's core hypothesis (the integration of all components), as well as more focused studies that independently examine the specific mechanisms of eco-feedback, gamification strategies and AI-driven personalization. The following subsections detail these four query sets (QS1-QS4).

2.1.2.1 The Core Hypothesis Intersection

Designed to retrieve literature that integrates all three key elements of the dissertation's proposed solution: IoT/smart home technologies, gamification and AI-driven personalization.

- **QS1:** ("IoT" OR "smart home") AND gamif* AND ("AI" OR "artificial intelligence" OR "machine learning" OR personaliz* OR "adaptive feedback") AND (energy W/2 saving OR consumption W/2 reduction)

2.1.2.2 Focused Eco-Feedback & Design

Focused on smart energy interfaces and eco-feedback mechanisms supporting sustained behavioural change and long-term engagement.

- **QS2:** ("IoT" OR "smart home" OR "smart buildings") AND ("eco-feedback" OR "energy dashboard" OR "real-time feedback" OR "user interface") AND ("sustained behavio?r" OR "long-term engagement" OR "behavio?ral change") AND (energy W/2 saving OR consumption W/2 reduction)

2.1.2.3 Focused Strategy Interaction Effectiveness

Isolated studies assessing gamification elements, historical feedback and normative feedback in energy-saving contexts.

- **QS3:** ("IoT" OR "smart home") AND (gamif* OR "historical comparison" OR "normative feedback") AND (energy W/2 saving OR consumption W/2 reduction OR "energy efficiency")

2.1.2.4 Focused AI & Personalization

Targeted personalization through AI, Machine Learning (ML) and adaptive recommendations to enhance eco-feedback and energy efficiency.

- **QS4:** ("IoT" OR "smart home") AND ("AI" OR "artificial intelligence" OR "machine learning" OR "user modeling" OR "behavioral modeling") AND ("personalized feedback" OR "personalized system" OR "adaptive recommendation" OR "recommender system*") AND (energy W/2 saving OR "energy efficiency" OR consumption W/2 reduction)

2.1.3 Screening Process

The database search, completed on 02 October 2025, regarding all four query sets yielded a total of **794** papers, depicted in Table 2.2. Following the removal of **249** duplicates identified using Mendeley's duplicate detection tool, **545** unique papers remained for preliminary screening. Titles, abstracts and, when necessary, conclusions were reviewed to assess their relevance to the research scope. Based on this stage, **155** papers were retained for full-text assessment. The full selection process is depicted in the PRISMA flow diagram (Fig. 2.1).

Table 2.2: Number of search results per database for each query.

Database	QS1	QS2	QS3	QS4	Total
ACM	8	10	36	17	71
IEEE	4	0	10	14	28
Scopus	35	13	63	41	152
ScienceDirect	109	157	149	128	543
Total	156	180	258	200	794

2.1.4 Final Eligibility and Selection

Several documents lacked associated URLs (Uniform Resource Locators) within Mendeley and were therefore linked manually. Likewise, PDF (Portable Document Format) files were manually added to the reference manager to facilitate organization and subsequent analysis. During this process, **13** records were identified as lacking full-text access and **9** additional papers were found to be duplicates, which Mendeley had failed to detect automatically. Furthermore, inconsistencies in publication year and other key article information were verified and corrected where applicable. The final eligibility assessment, based on comprehensive full-text analysis, resulted in **114** papers that met all inclusion criteria, as shown in Fig. 2.1. These articles, alongside the initial research from the dissertation manifest, are compiled in the References section. They collectively represent the current research frontier in IoT-based strategies for sustainable behaviour and energy efficiency.

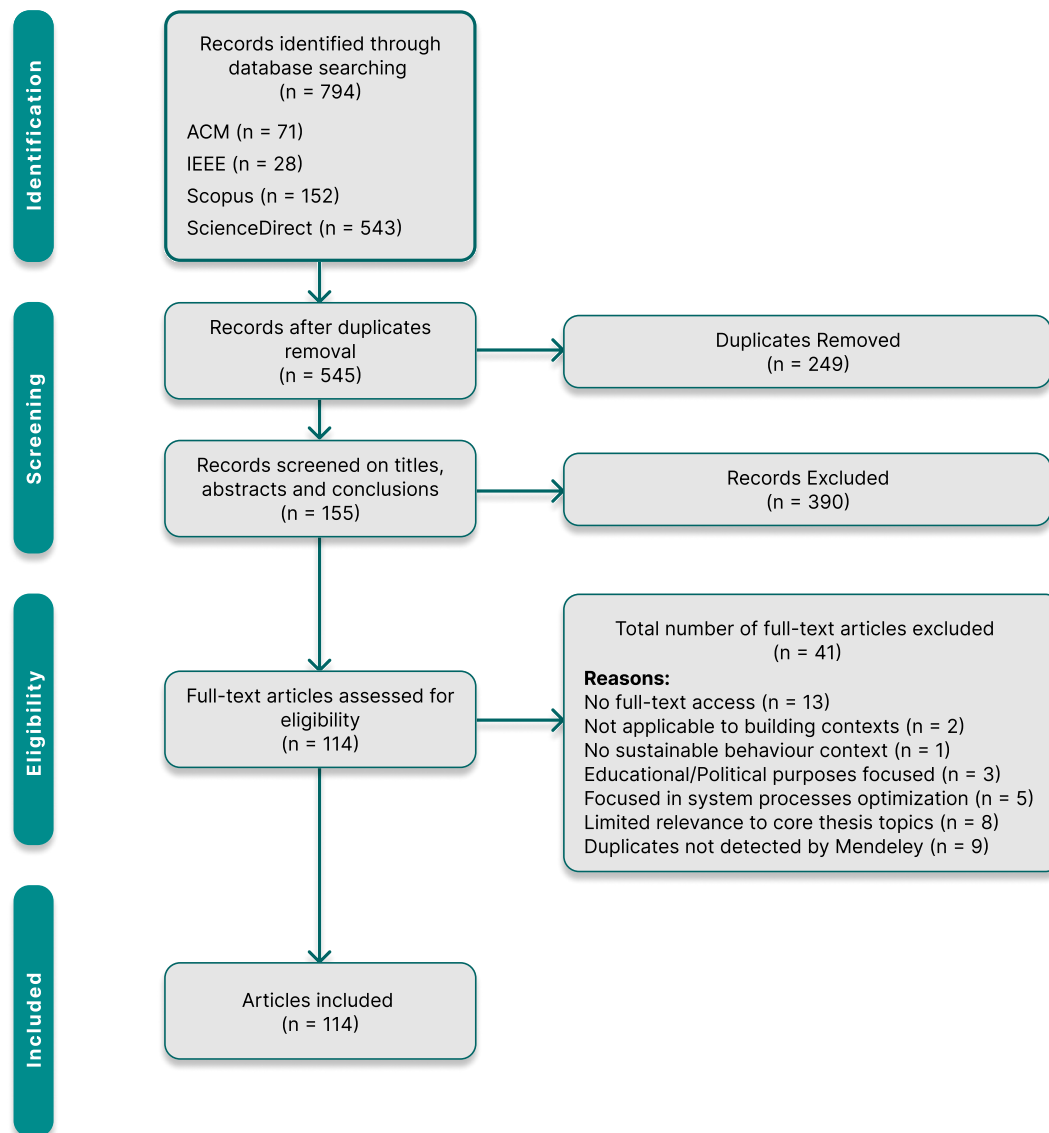


Figure 2.1: Articles screening and filtering process (PRISMA flow diagram).

2.1.5 Review Analytics

To validate the literature review's trajectory and thematic focus, the final set of **114** selected articles was analysed. The following charts visualize the results of this bibliometric analysis, revealing the field's publication trends, thematic evolution and disciplinary focus.

Fig. 2.2 illustrates the number of relevant articles published per year from 2019 to 2025. The trend shows a steady, foundational growth in the field from 2019 (8 articles) to 2023 (16 articles). After 2023, the field experienced a sharp acceleration, jumping to 25 articles in 2024 and 30 in 2025. This spike indicates that the research domain is not only growing but is at a peak of timeliness and relevance.

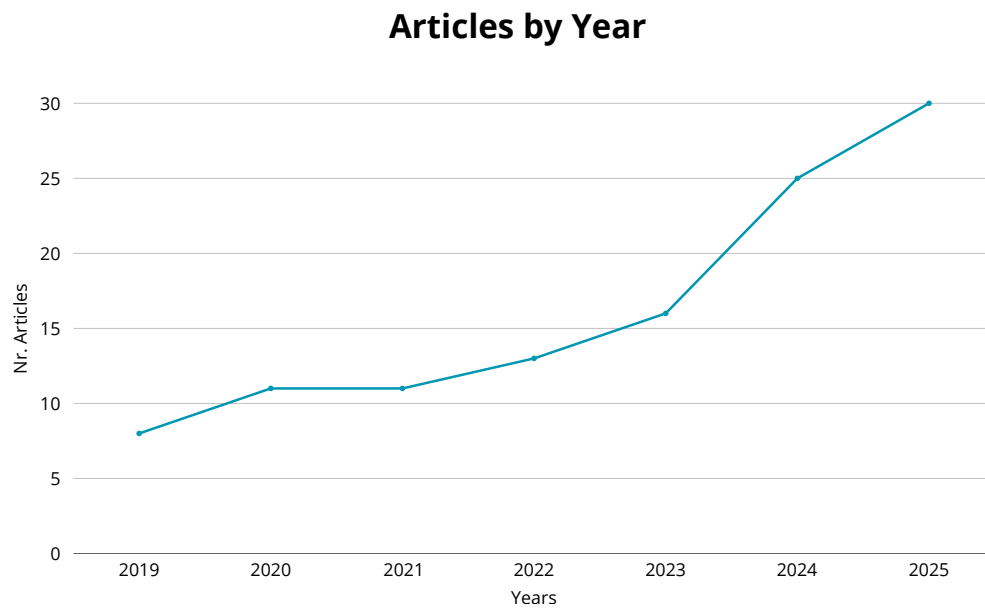


Figure 2.2: Distribution of selected articles by publication year.

Fig. 2.3 provides a thematic breakdown of the published articles, categorizing them into four key areas: Eco-Feedback, Gamification, Behaviour Analysis and AI Related. This chart clearly explains the acceleration shown in Fig. 2.2. From 2019 to 2023, the research was dominated by the foundational pillars of eco-feedback, gamification and behaviour analysis. However, the sharp spike in 2024 and 2025 is driven almost exclusively by a massive increase in the AI Related category. This demonstrates a clear pivot in the research community, moving from establishing the problem and initial solutions toward developing and integrating AI-driven systems.

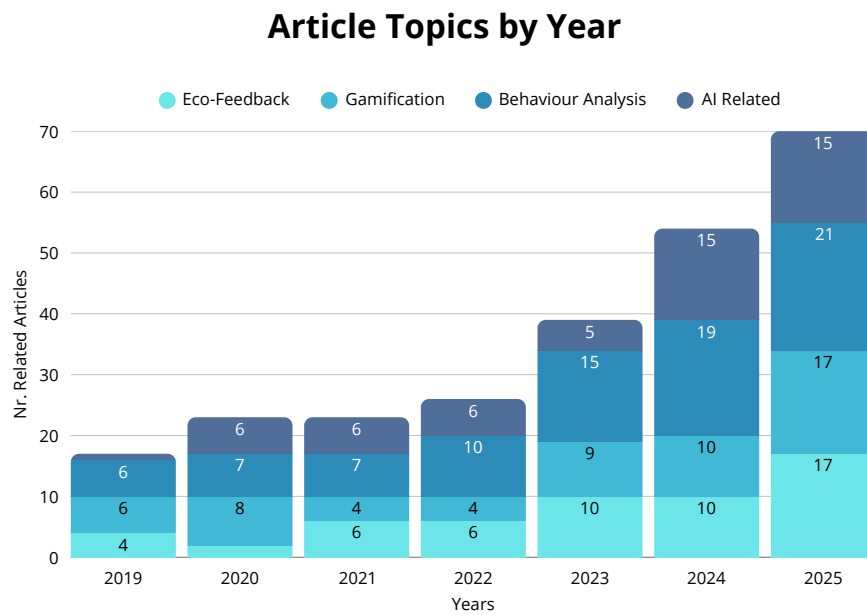


Figure 2.3: Thematic evolution of article topics from 2019 to 2025.

Furthermore, Fig. 2.4 shows the disciplinary focus of the literature based on the source databases. The field is dominated by publications in engineering and energy-focused databases, with ScienceDirect (59.6%) and Scopus (20.2%) accounting for approximately 80% of the articles. A significant and equal contribution comes from core computer science databases, ACM (9.6%) and IEEE (10.5%). This distribution confirms the interdisciplinary nature of the research, sitting at the intersection of energy science and human-computer interaction.

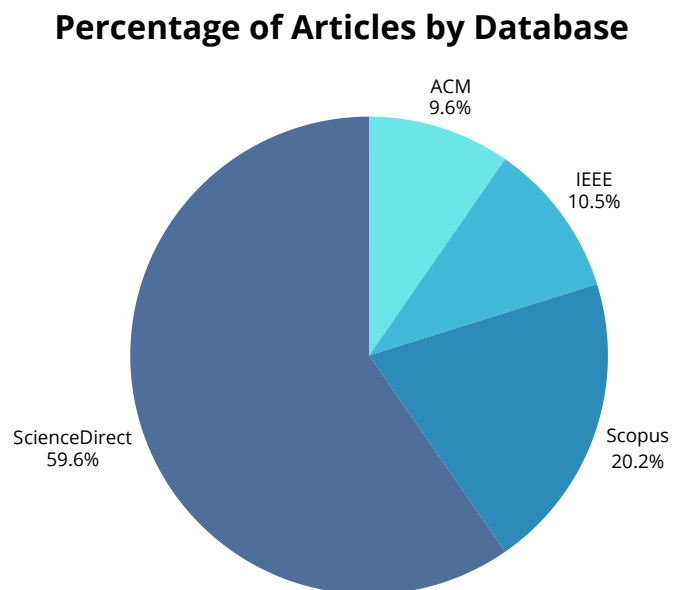


Figure 2.4: Percentage of articles sourced from primary academic databases.

To validate the alignment between the selected literature and the research objectives, the final set of articles was mapped against the four proposed RQs. Fig. 2.5 illustrates the number of papers contributing to each inquiry.

It is important to note that this classification is not mutually exclusive. Given the integrative nature of the field, a single study often addresses multiple dimensions simultaneously. For instance, a study demonstrating the superior efficacy of gamified systems over traditional feedback (relevant to RQ2) frequently achieves this by analysing the impact of specific mechanics such as points or social comparison (relevant to RQ3). This overlap confirms the interconnectedness of the topics and ensures a robust evidence base for all aspects of the dissertation.



Figure 2.5: Articles related to the RQs presented.

To further visualize the internal structure and thematic relationships of the selected literature, a bibliometric analysis was conducted on the title and abstract keywords of the 114 articles using VOSviewer, as shown in Fig. 2.6.

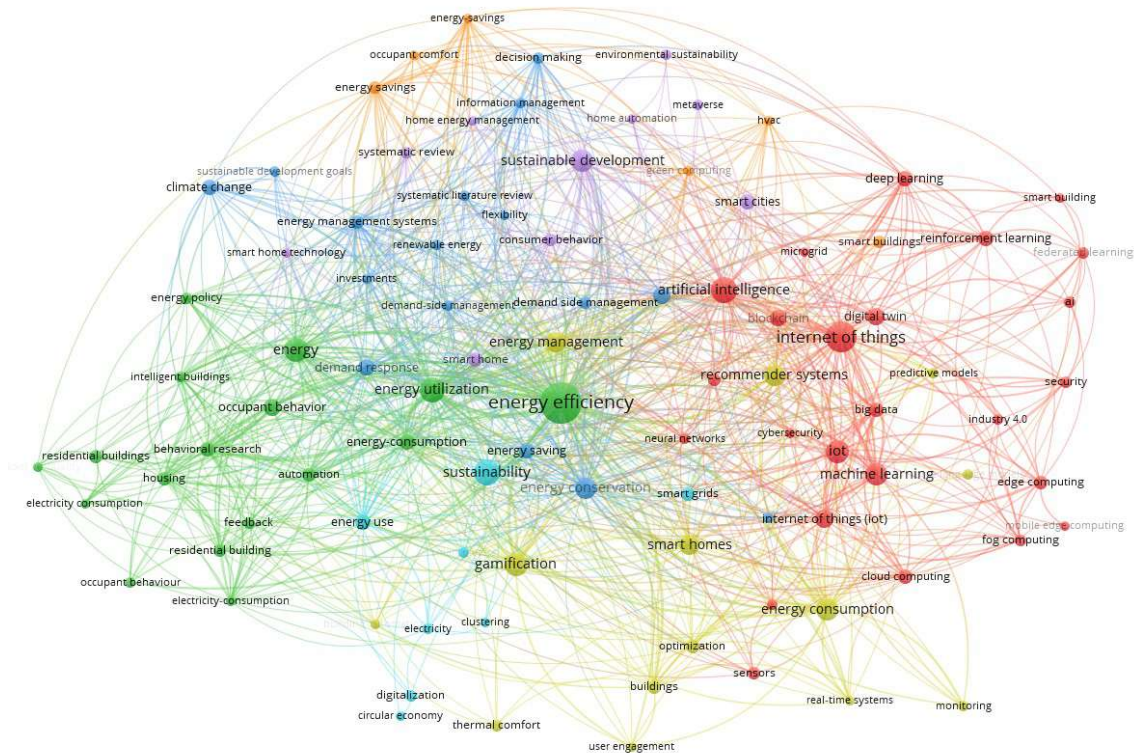


Figure 2.6: Bibliometric analysis of keyword co-occurrence in the 114 selected articles.

This analysis provides several key insights that validate the dissertation's focus. The map is fragmented into distinct thematic clusters, representing the "silos" of the research field. A large technical cluster (red) is visible, dominated by keywords like "artificial intelligence", "machine learning", "IoT" and "reinforcement learning". A separate behavioural cluster (green) investigates the core problem, focusing on "occupant behaviour", "feedback" and "residential buildings".

Critically, gamification (yellow) forms its own sociotechnical cluster, which is not strongly linked to the AI cluster. This visualization provides a clear empirical justification for the research gap this dissertation addresses: the lack of integration between advanced AI-driven personalization (the red cluster) and psychologically-grounded engagement strategies (the yellow cluster) to solve the behavioural problem (the green cluster).

In short, the analysis of the 114 selected articles strongly justifies the direction of this dissertation. The field is experiencing rapid accelerating growth and its primary thematic focus has pivoted decisively toward solving existing challenges, such as engagement decay and the one-size-fits-all approach, through the integration of AI-driven personalization.

To construct a robust theoretical foundation for the proposed architecture, the remainder of this chapter unpacks this exact trajectory. By critically examining the state of the art across these interconnected domains, the review aligns the empirical findings of the systematic search with established behavioural and technological theories.

2.2 Occupant Behaviour as a Determinant of Building Performance

While technological improvements, such as efficient Heating, Ventilation and Air Conditioning (HVAC) systems, are critical, recent literature reinforces a consensus: purely technical solutions are insufficient. End-user behaviour is frequently identified as the "invisible" variable, potentially influencing up to 50% of a building's energy consumption [29, 52, 61, 69, 97, 107]. Despite its impact, behaviour remains the least-integrated factor in energy forecasting models, introducing significant uncertainty that traditional technical models struggle to address [6, 12, 15, 27, 29, 52, 54, 77, 83, 91, 95, 97, 98, 104, 107].

The magnitude of this behavioural variable is illustrated in Fig. 2.7. Empirical studies have found that due to distinct habits and unconscious routines, identical housing units can exhibit electricity consumption variations of up to 600% [61, 103]. Unlike costly infrastructural retrofits, modifying these behaviours offers a path to immediate energy savings, estimated between 10% and 30% by the International Energy Agency⁶ (IEA), at virtually no capital cost [29, 97, 106].

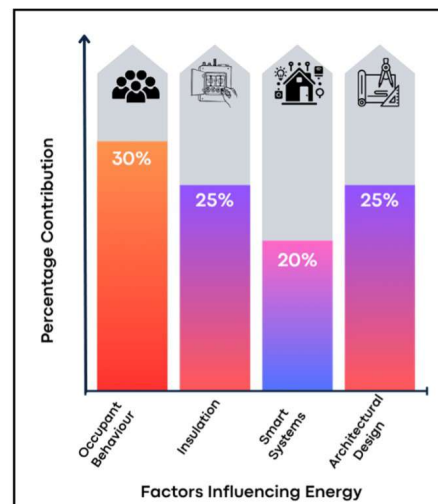


Figure 2.7: Relative contribution of occupant behaviour to total energy performance in buildings [97].

Capturing these savings is particularly challenging in non-residential contexts. Research has historically "focused on the residential sector", overlooking that the socio-psychological drivers at work differ significantly from those at home [47]. In office environments, employees face a "split-incentive" problem: they lack the direct monetary motivation to conserve because they are not responsible for the utility bills [90]. Consequently, interventions in these domains cannot rely on financial savings alone. Instead, they must activate alternative motivations, such as personal norms or the psychological needs for autonomy and relatedness [47, 90, 94]. Furthermore, the deployment of granular monitoring in workplaces faces barriers regarding user acceptance, specifically fears of surveillance and loss of control [90].

⁶<https://www.iea.org/>

These limitations have prompted a fundamental shift towards Human-Centric Design (HCD) [12, 35, 45]. This approach adopts a "human-in-the-loop" perspective, acknowledging that active user participation is crucial for successful energy transitions [5, 24, 45, 56, 69, 82]. Cottafava et al. [19] frame this explicitly as a shift toward viewing buildings as "socio-technical systems". In this view, energy consumption is the product of an interaction between a technical component (HVAC infrastructure, energy flows) and a social component (habits, comfort perception). Consequently, the primary goal of intervention is not merely reduction, but the decoupling of energy demand from indoor comfort, proving that efficiency does not require a sacrifice in well-being.

Within this HCD framework, providing feedback has emerged as the primary strategy to bridge the gap between "invisible" consumption and user awareness [21, 29, 106, 107]. However, behavioural science warns that information alone is often insufficient [17, 101]. People frequently fail to act in their own long-term interests because the rewards for sustainable behaviour are delayed, intangible or difficult to perceive [97]. This disconnect highlights the need for strategies that move beyond static data display to leverage deeper psychological drivers, transforming abstract energy data into actionable intrinsically rewarding experiences [82, 93].

Having established that behaviour is the critical "invisible variable", the immediate scientific challenge becomes making this variable visible. This necessitates an examination of eco-feedback strategies, which serve as the foundational interface between hidden energy flows and human awareness.

2.3 The Evolution of Eco-Feedback Strategies

As the primary strategy for human-centric energy intervention, eco-feedback rests on the principle that making energy consumption visible encourages efficiency investments and behaviour change [3, 47, 91, 106]. While empirical reviews confirm that feedback can contribute to savings of up to 33%, effectiveness is not guaranteed [3, 106, 109], as the impact can vary largely depending on *how*, *when* and to *whom* the feedback is delivered [9, 54, 106].

Regarding *how* information is presented, research highlights a comprehension gap where users struggle to interpret abstract technical metrics like kWh [107, 108]. Studies by Ferreira et al. [28] and Huseynli [43] suggest that feedback is significantly more persuasive when framed in relatable terms, such as monetary costs or environmental metaphors (e.g., "trees saved"), rather than raw engineering units.

The timing of delivery (*when*) is equally critical. Delayed feedback, typical of monthly reports, often fails to correct behaviour because users cannot link the cost to a specific action. In contrast, immediate feedback mechanisms, such as the ambient colour-changing smart plugs tested by Barreto et al. [9], have proven more effective by providing "just-in-time" visual cues directly at the point of consumption.

Finally, the recipient's profile (*to whom*) determines whether feedback motivates or discourages. Liu et al. [54] demonstrated this volatility in a field study where social comparison helped

high-energy users reduce consumption but accidentally triggered a "boomerang effect" among efficient users, who increased their usage to match the norm.

This complexity demonstrates that effective eco-feedback requires more than just data transmission. It demands a transition from static reporting to dynamic context-aware systems.

2.4 From Traditional Bills to Real-Time IoT Feedback

The effectiveness of eco-feedback is fundamentally tied to both the technological medium through which it is delivered and the psychological framing of its message. This section first explores the transition from delayed static reporting to interactive IoT platforms and subsequently outlines the specific communication methods used to translate raw data into actionable behavioural nudges.

2.4.1 The Evolution of Delivery Platforms

The trajectory of feedback technology begins with the traditional energy bill. While it remains the most ubiquitous interface, it is widely considered ineffective for behavioural change. Due to the significant temporal gap between action and reporting, its infrequency prevents users from linking specific actions to costs, often leading to a "fall-back effect" where initial savings dissipate [44, 65, 71, 106].

Enhanced billing represents the first step in addressing the comprehension gap by incorporating richer data and basic behavioural principles. By adding historical comparisons and web-based reports, enhanced billing constitutes a slight improvement but yields only modest savings averaging 3% [106].

However, the integration of IoT and smart metering has revolutionized this domain. By making energy "visible" through In-Home Displays (IHDs), mobile apps and web portals, these platforms allow users to immediately understand the impact of their consumption patterns [7, 32, 36, 61, 69, 106, 107]. A comparative evaluation of these reporting means is detailed in Table 2.3.

Table 2.3: Comparison of energy reporting instruments.

Instrument	Frequency	Info. Quality	Install. Cost	Display Capability
Standard Bill	Low (Monthly)	Low	None	Low (Static)
Enhanced Bill	Low (Monthly)	Medium	None	Low (Static)
Web Platform	Real-Time	High	Low	High (Interactive)
IHD	Real-Time	High	High	High (Interactive)
Mobile App	Real-Time	High	Low	High (Interactive)

Smartphones have emerged as the "ideal platform" for this intervention due to their interactivity and ubiquity, with over 90% of the population using them to browse the internet in many European countries [52, 69]. Field studies consistently demonstrate that real-time app feedback drives higher savings, typically between 5% and 8%, compared to static methods [46, 106]. This

effectiveness is driven by the ability to deliver granular timely insights, such as normative comparisons for air conditioning usage (Fig. 2.8) or specific heating breakdowns (Fig. 2.9). Hardware innovations like smart plugs (Fig. 2.10) further enhance this by providing "just-in-time" visual cues (e.g., LED colour changes) directly at the source of consumption [9].

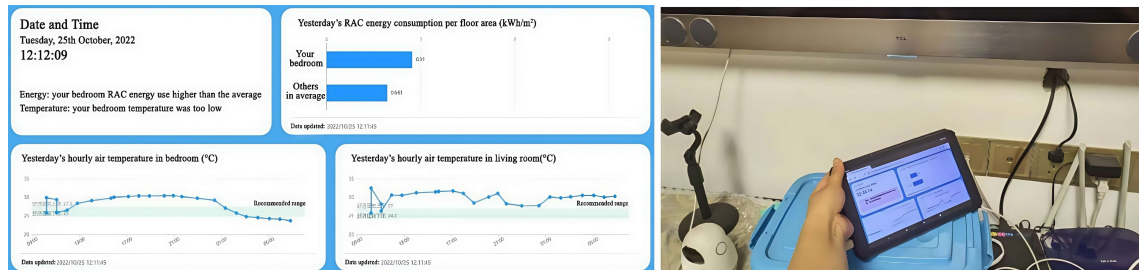


Figure 2.8: Normative feedback interface from Liu et al. [54], benchmarking daily room air conditioner consumption (top-right) against the group average.

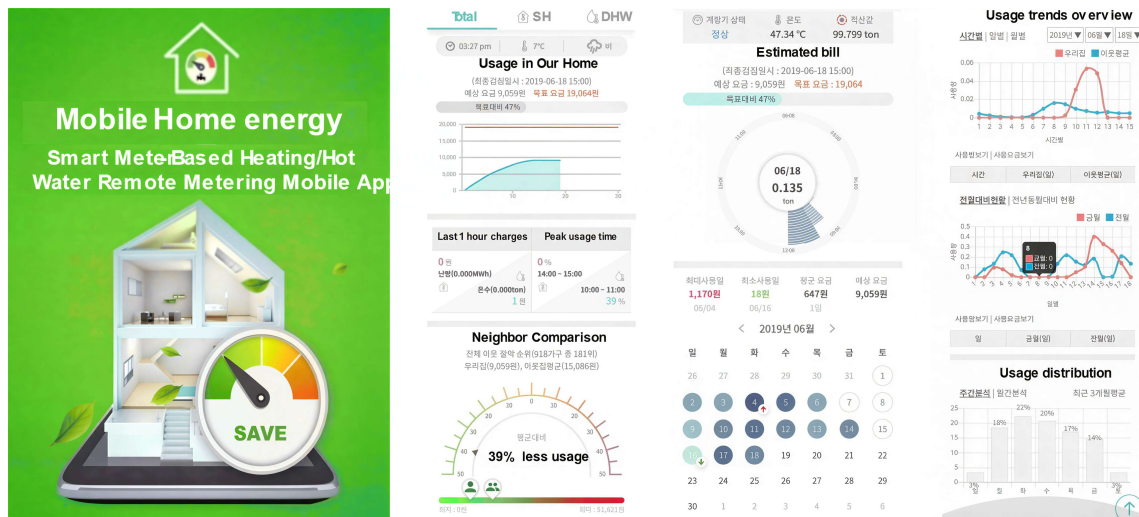


Figure 2.9: Heat monitoring app [46].

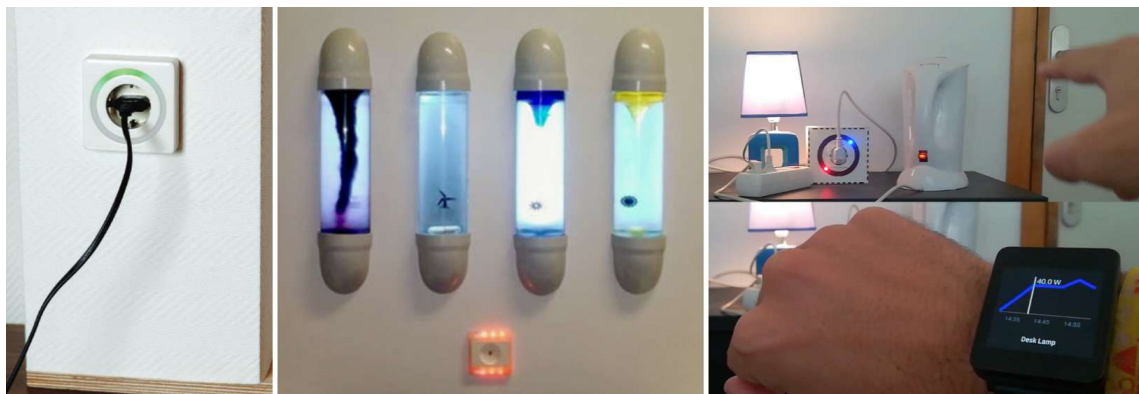


Figure 2.10: Smart plugs used for energy awareness [9].

2.4.2 A Taxonomy of Feedback Methods

This evolution in delivery has been paralleled by an evolution in communication style. These software-based approaches can be categorized by their method and delivery style [107]. As detailed in Table 2.4, these range from simple indicative data to complex social comparisons.

Table 2.4: Taxonomy of eco-feedback methods [107].

Feedback Method	Description
Indicative	Offers raw data (e.g., real-time graphs).
Disaggregated	Provides granularity (e.g., consumption by appliance).
Comparative	Places consumption in context through: •Historical comparison (current vs. past usage). •Normative comparison (vs. similar households). •Peer comparison (vs. specific friends).
Injunctive	Provides explicit norms or recommendations (e.g., a "red light" for high usage).
Ambient	Uses environmental cues (e.g., changing light colour) for "just-in-time" decision-making [9].
Artistic	Employs visual metaphors (e.g., a wilting flower/mascot) to foster emotional connection [75].

This taxonomy aligns with the behavioural science concept of "nudging", which seeks to influence choice without restricting it or increasing costs [11, 25, 65, 88, 93]. Specifically, research identifies several robust strategies for promoting sustainable behaviour, including informational nudges to provide context, social nudges that utilize comparative feedback reminders to prompt action and default nudges, which establish efficiency as the automatic choice unless actively changed [3, 6, 7, 47, 65, 69, 88, 92, 93, 98].

Informed by this taxonomy, and to anticipate the engagement challenges discussed in the following section, this dissertation synthesizes three specific methods to maximize impact.

2.4.2.1 Disaggregated Feedback

Standard aggregate data often hides inefficiencies. By providing appliance-level granularity, the system transforms general consumption into actionable insights. As illustrated in Fig. 2.11, this approach allows users to immediately identify which specific device contributes most to their bill, reducing the cognitive effort required to diagnose waste [36].

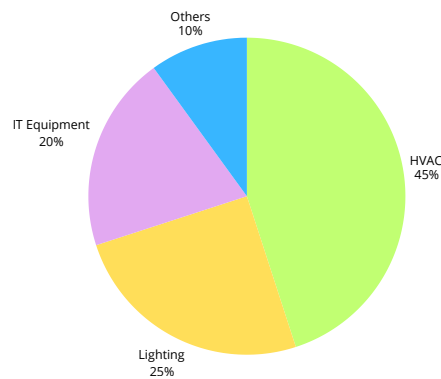


Figure 2.11: Disaggregated Feedback, isolating consumption by appliance category.

2.4.2.2 Normative Comparison

To leverage social motivation, this strategy uses "similar household" comparisons as a nudge. Fig. 2.12 depicts this mechanic, where a user's performance is benchmarked against an "efficient neighbour". However, acknowledging workplace findings that social competition can be perceived as toxic or unfair, this social metric is balanced with historical benchmarking (self-comparison) to ensure progress is measured against the user's own baseline [3, 69, 90].

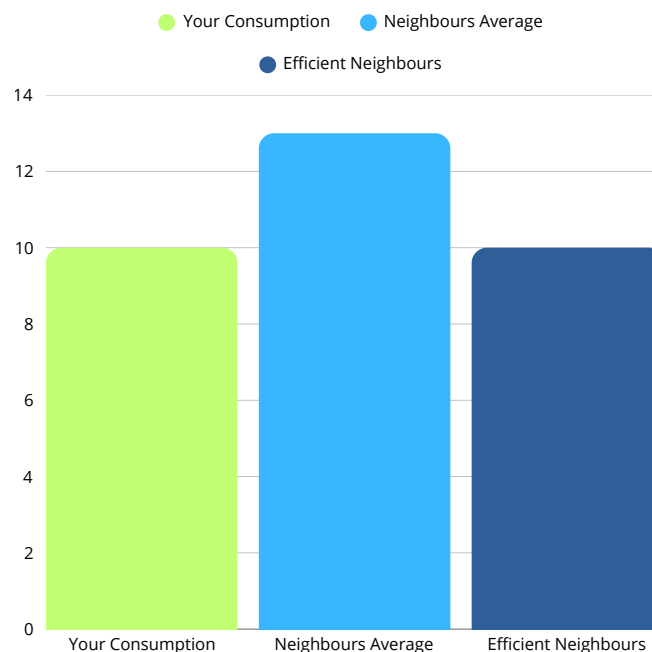


Figure 2.12: Normative Comparison interface showing the user's standing relative to other peers.

2.4.2.3 Framing and Metaphors

Finally, to bridge the comprehension gap, the system translates abstract metrics (kWh) into relatable units. Fig. 2.13 demonstrates this translation, converting technical energy units into monetary

savings or environmental impact (e.g., "trees saved"). This framing increases energy literacy and persuasive power by making the intangible tangible and aligning it with the user's personal values [3, 21, 28, 43, 92, 105, 107].

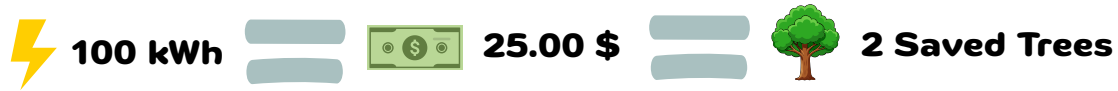


Figure 2.13: Framing strategy translating 100 kWh into relatable monetary and environmental equivalents.

2.5 The Engagement Gap: Limitations of Feedback-Only Systems

Despite the technical evolution from static bills to dynamic apps, the literature confirms that information alone is insufficient to guarantee durable behaviour change. Success is "heavily dependent on the engagement of end users", creating a critical engagement gap between data availability and user action [3, 43]. Consequently, the ultimate efficacy of feedback hinges not on the data itself, but on how it is "tailored to the household and presented in a clear, engaging way" [106].

2.5.1 Engagement Decay Over Time

The sustainability of behaviour change is plagued by the Frequency Paradox, where excessive updates lead to user fatigue, and the Duration Paradox, where attention naturally wanes over time, causing initial savings to dissipate [13, 74, 106]. This decline is often accelerated by "crisis fatigue", a psychological state where users become emotionally numb to constant urgent warnings or alarmist environmental narratives [28, 65]. Consequently, most short-term studies fail to demonstrate persistent behavioural shifts, with multiple authors concluding that feedback interventions remain "ineffective in the long term" without a robust engagement mechanism [4, 15, 61, 106].

2.5.2 Gaps Between Information and Action

Even when users remain engaged, several structural gaps prevent awareness from translating into savings.

The first hurdle is the comprehension gap, introduced previously in Section 2.3. Simply providing data does not automatically equip users with the knowledge to act [25, 36, 107, 108]. Abstract metrics like kWh are often "invisible" or confusing to non-technical users, failing to convey the true cost of consumption [47, 61]. To bridge this gap, data must be translated into relatable terms, such as monetary cost or environmental impact (e.g., "trees saved"), rather than raw technical units [28, 92, 109].

However, clarity alone does not guarantee execution. This disconnect is known as the intention-action gap, where high motivation fails to materialize into behaviour [15, 66]. Studies reveal a persistent 30-40% discrepancy between users' stated intentions and their actual actions [15]. Experiments confirm that while digital nudges can successfully boost motivation, they often fail to

statistically bridge the gap between attitude and behaviour, suggesting that the challenge is far more complex than simply reminding users of their values [29, 88].

Compounding these psychological challenges is a technological usability trade-off. As feedback strategies become more sophisticated, they risk becoming a burden for users rather than an asset [3, 24]. Al-Kababji et al. [44] quantified this tension, finding that while complex interactive plots (e.g., heatmaps) offer higher analytical accuracy, non-expert users rated them significantly lower in understandability compared to simple charts.

To mitigate this friction, Pombeiro et al. [71] argue that dashboards must minimize "cognitive overload" by speaking the user's language and ensuring immediate "visibility of system status". When these principles are ignored, field studies confirm that users actively reject complex apps, labelling them "useless" or "hard to use", expressing a clear preference for "straightforward numerical environment data" [3, 24, 108]. Poor design can even cause interventions to backfire. For example, a 2025 study found that the convenience of remote-control apps actually led users to operate their Air Conditioners (ACs) more frequently, prioritizing comfort over the intended energy savings [8].

To diminish this cognitive burden, modern architectures are shifting from passive data visualization to active guidance, using algorithms to interpret complex data on the user's behalf. However, this transition creates a dependency on the quality of the intervention. As noted by Fakhar, Yalcin, and Bilge [27], unless prompts are "precise and helpful", they are perceived as "impractical and vain", triggering user antagonism. This necessitates explainability as a core design requirement: to ensure compliance, the system must explicitly articulate the underlying rationale (why) and the quantifiable value (benefits) of every recommendation [39, 41, 64].

2.5.3 Overriding Contextual Barriers

Ultimately, feedback systems often fail because they do not account for the hierarchy of user needs [29, 111]. Qualitative analysis confirms that users consistently prioritize personal comfort and convenience over energy conservation [108, 111], while energy saving is often viewed merely as a "side effect" rather than a primary goal [88].

If an intervention conflicts with a user's primary task, such as a complex interactive challenge that distracts office workers, it will be rejected [95]. This creates a critical design paradox: interventions must be engaging enough to combat decay, yet seamless enough to avoid rejection. To address this, systems must look beyond static data to methods that actively support user motivation. This requirement drives the exploration of gamification, detailed in the following section, as a strategy to align energy efficiency with intrinsic user drivers.

2.6 Gamification as an Engagement Strategy

To address the engagement gap and bridge the intention-action gap defined in the previous section, research has turned to gamification as a "considerable strategy" to "achieve and maintain energy-efficient behaviour while ensuring occupant comfort" [24, 45, 47, 49, 52, 61, 64, 97]. Gamification

is defined not as playing games, but as "the use of game elements and design... to communicate with users, motivate engagement and educate them in adopting behaviours and habits that promote energy efficiency" [1].

While traditional feedback is knowledge-intensive, gamification aims to create a "psychological commitment" by transforming energy saving into a "rewarding experience" [13, 43, 49, 93]. Systematic reviews confirm this potential: while standard feedback yields modest results, gamified interventions have achieved average savings of 24%, with some studies maintaining impact for over 8 months [49]. This success stems from gamification's ability to align engagement, economic motivations (saving money) and sustainable behaviour with intrinsic drivers like achievement and social relatedness [1, 11, 25, 57, 75, 97]. Unlike one-off nudges, gamified systems can evolve with the user, making energy saving a dynamic and socially meaningful activity.

2.6.1 Theoretical Foundations: Why It Works

The success of gamification is grounded in human-focused design and relies on two primary psychological frameworks. The first, Self-Determination Theory (SDT), posits that sustained motivation depends upon satisfying three fundamental needs: Autonomy (control over goals), Competence (feeling effective via progress feedback) and Relatedness (a sense of belonging and social connectedness) [16, 47, 57, 58, 97, 111].

Complementing this is the Theory of Planned Behavior (TPB), which argues that behaviour is determined primarily by intention. It remains one of the most frequently used theories to describe energy-saving behaviour [10, 17, 37, 43, 47, 98, 111]. As illustrated in Fig. 2.14, TPB suggests that behaviour is shaped by Attitudes, Subjective Norms (social pressure) and Perceived Behavioural Control. Gamification explicitly targets these norms through mechanics such as leaderboards and social comparisons [1, 4, 42, 52, 57, 72, 97].

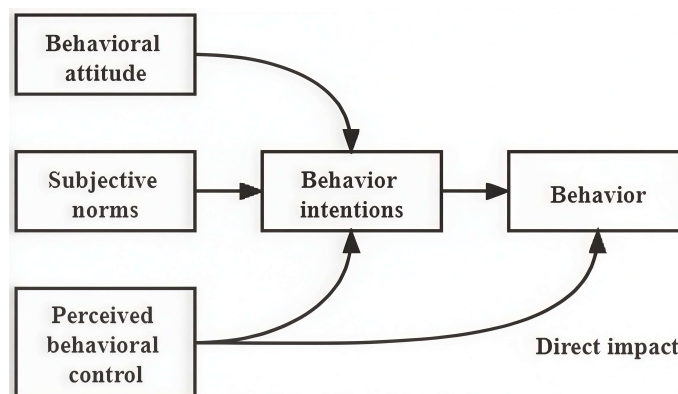


Figure 2.14: The standard TPB model [111].

This basic TPB model is often extended (Fig. 2.15) to include factors like Habit and Individual Comfort, which are critical in energy contexts [111]. A key component within Subjective Norms is the split between Descriptive Norms (what others actually do) and Injunctive Norms (what others approve of) [111]. Studies find that descriptive norms (e.g., seeing a colleague's visible

energy-saving actions) can have the most significant impact, which is why social feedback is so potent.

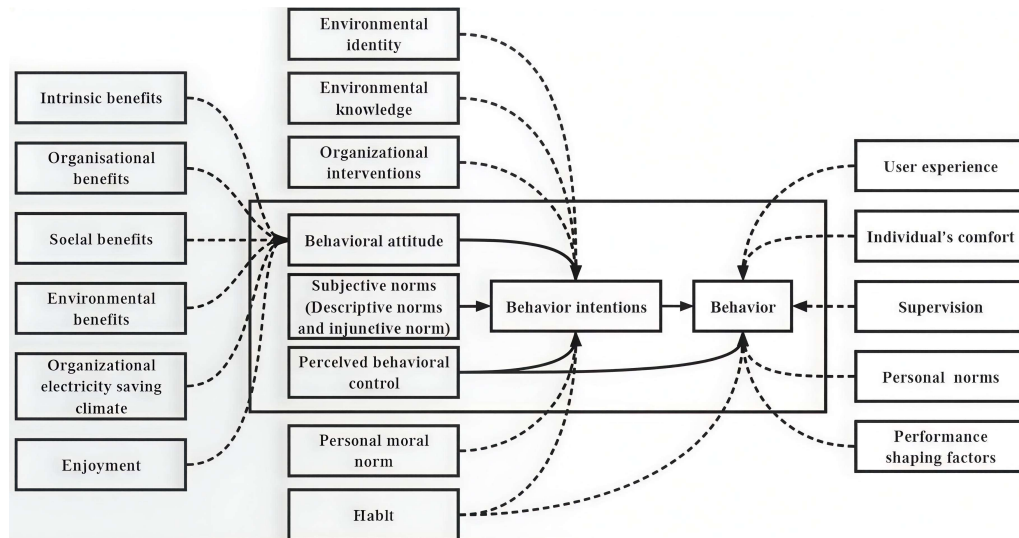


Figure 2.15: An extended TPB model for employee energy-saving behaviour, showing the integration of additional factors like habit, norms and comfort [111].

Other frameworks and theories like Goal-Setting Theory, the Norm Activation Model, Operant Conditioning and Habit Loop Theory also provide lenses for how gamification can strengthen behavioural intentions by enhancing self-efficacy, appealing to personal norms (a sense of moral obligation) or providing clear rewards [15, 31, 37, 47, 57, 79, 80].

Grounded in this psychological evidence, the approach of this dissertation will specifically leverage mechanics that target these proven drivers. It will incorporate (1) goal-setting and progress feedback to satisfy the need for Competence (SDT), (2) customizable challenges to support Autonomy (SDT) and (3) normative comparison against a historical baseline to activate Subjective Norms (TPB), alongside collaborative mechanics to foster Relatedness (SDT) and shared accountability.

2.6.2 Evidence from the Field

Empirical implementations reveal three distinct strategies for translating behavioural theory into successful practice: the use of tangible metaphors, the leverage of social norms and the integration of tokenized incentives in workplace settings.

2.6.2.1 Tangible and Visual Metaphors

Successful systems bridge the gap between abstract engineering data and user understanding by employing tangible metaphors that trigger emotional responses. A prime example is the Ant Forest app, which gamifies carbon reduction by allowing users to grow virtual trees that are eventually planted as real trees, successfully linking digital actions to physical environmental impact [42].

Similarly, Tiefenbeck et al. demonstrated that emotional connection can outweigh rational information, as their use of a "polar bear on a melting ice floe" display proved more effective at nudging users toward water conservation than numerical feedback alone [11]. This approach is further refined in the HomeMascot system (Fig. 2.30), which personifies energy conservation through a virtual character whose mood acts as an immediate, intuitive indicator of real-time consumption efficiency [75].

2.6.2.2 Social Comparison and Norms

Social nudges are among the most powerful tools for validation and motivation [17, 33, 63, 65, 66, 71, 87, 92]. Platforms like Opower have historically validated that comparing a user's usage to their neighbours consistently drives reductions [97]. Recent experimental work by Stieglitz et al. [88] reinforces this, finding that social norm nudges significantly promote energy-saving intentions, whereas simple self-commitment nudges often fail to produce a significant effect. This efficacy extends to gamified environments, as seen in the Social Power App, which achieved 8.5% electricity savings by integrating social media mechanics into the feedback loop [65].

Crucially, this strategy proves effective even in the absence of financial incentives. Trinh, Fung, and Straka [92] demonstrated that flat-rent tenants receiving normative feedback (neighbour comparisons) achieved 67% more dashboard views and higher savings than those viewing personal data alone (Fig. 2.16). This confirms that social currency drives behaviour even when financial currency is irrelevant. Similarly, Mi et al. [63] found that group-contrast feedback (comparing one group's performance against another) in student dorms reduced consumption by 24.23%.

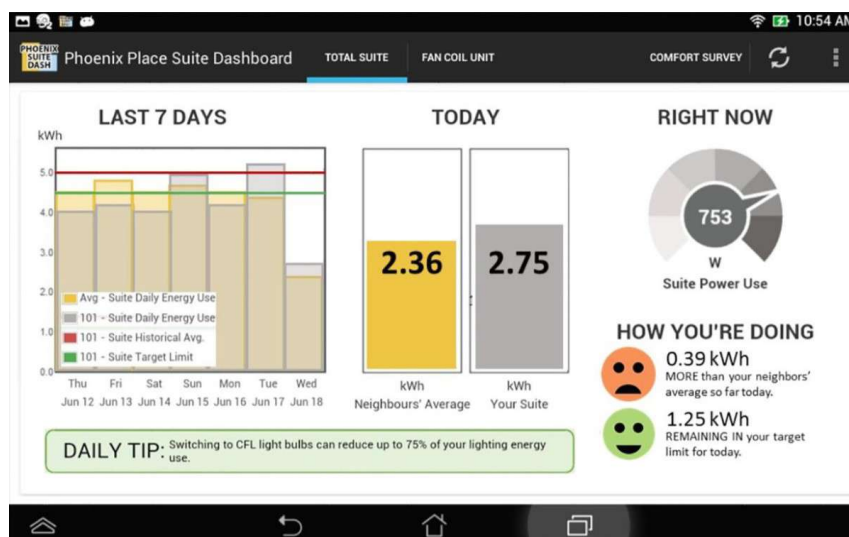


Figure 2.16: Dashboard interface displaying real-time and normative feedback, including social comparison with neighbouring peers and a goal-setting widget [92].

Furthermore, social mechanics can be targeted to address specific grid challenges. A 15-month study by Mäki-Verikko et al. [58] in student housing (Figs. 2.17, 2.18) combined a local social network with gamified "pause hours" (specific time windows where residents are challenged

to halt or defer consumption during grid peaks). While overall consumption did not drop, the gamified challenges led to a substantial 46% reduction in peak-load consumption, proving that social mechanics are highly effective for specific demand response events.

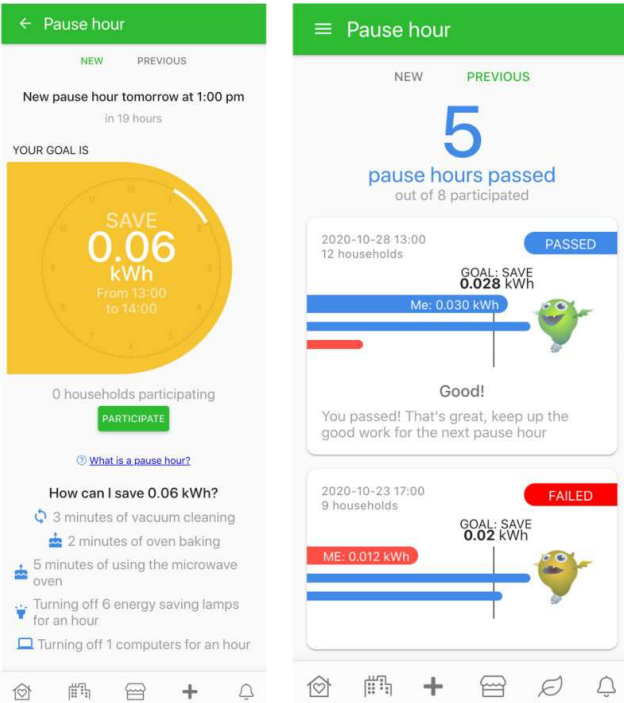


Figure 2.17: Interface for the gamified "pause hour" demand response challenge, showing (left) the sign-up prompt with a savings goal and (right) the results screen with social comparison [58].

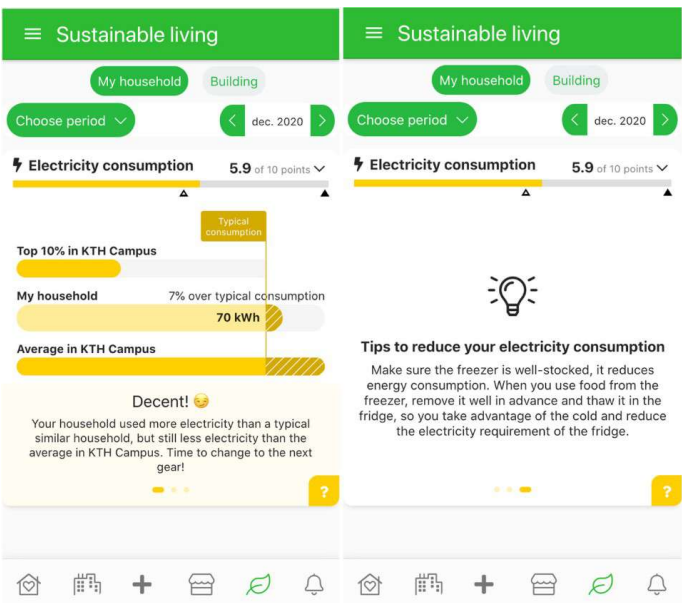


Figure 2.18: App's main "Sustainable living" feedback screen, showing (left) social comparison feedback (normative) and (right) a general energy-saving tip [58].

2.6.3 Limitations: Engagement Fatigue and the One-Size-Fits-All Problem

Despite these successes, the literature identifies a critical flaw in current implementations: the one-size-fits-all approach. Most systems apply the same mechanics (points, badges and generic leaderboards) to all users, regardless of their motivation profile [2]. This is a significant limitation, as the effectiveness of these mechanics is highly dependent on the environment and user type. For instance, a strategy designed for residential settings may fail to achieve the same results in an office or educational context [1, 61, 97].

This rigidity leads to two specific mechanisms of failure:

Reward Fatigue: Users disengage as tasks become repetitive and extrinsic rewards (points) lose their novelty, leading to a decline in usage after as little as 20 days [15, 97].

The Boomerang Effect: A 2024 study by Liu et al. [54] found that while normative feedback helped high-energy users reduce consumption (-11%), it caused low-energy users to increase their consumption. A single uniform social comparison message backfired because it failed to account for individual context.

To combat these limitations, recent systems propose an adaptive gamification layer that uses AI to personalize interventions [73, 88]. A 2025 study on water conservation found that AI-personalized nudges increased participation by 28% and led to a 12.5% reduction in consumption [73]. Large-scale studies often report lower savings precisely because a static approach fails to account for the "heterogeneity of participants" [3]. AI-driven personalization addresses this by clustering users based on behaviour (e.g., "unconcerned" vs. "active saver"), ensuring that social comparisons are made only between similar peers to prevent the boomerang effect [31].

This forms the core of this dissertation's technical contribution. The proposed approach diverges from using AI for purely technical building optimization. Instead, it positions AI as a human-centric tool. By dynamically tailoring challenges and difficulty levels, the system maintains user competence (per SDT) and prevents reward fatigue, utilizing semi-automated mechanisms that users prefer over full automation [90].

Ultimately, the failure of one-size-fits-all gamification highlights that while motivation is necessary, it must be relevant to be effective. This insight drives the transition to the final technological frontier: AI-driven personalization.

2.7 The New Frontier: AI-Driven Personalization and Adaptation

While gamification provides the engagement layer, the literature explicitly calls for AI-driven adaptation to overcome the rigidity of static systems, intelligently adapting to individual user profiles [1, 97].

2.7.1 The Need for Adaptive Interventions

The effectiveness of behavioural interventions is highly context-dependent [45, 65]. Standardized approaches often fail, as their pre-programmed fixed rules miss to account for changes in occupants' preferences or building usage [23, 28, 45, 61, 62, 70]. Moreover, a nudge that motivates one person may have negative impacts on another [65]. For instance, a "green nudge" emphasizing environmentalism may be ineffective for a user motivated solely by cost, while peer comparison can backfire for low-energy users, causing them to drift toward the higher group average [54, 65].

Recent segmentation studies underscore this heterogeneity. Webb, Olaru, and Rastin [99] identified distinct clusters of users ranging from "exemplary" (high intrinsic motivation requiring reinforcement), to "nudgeable" users (responsive to social norms) and "balanced" users (requiring carefully calibrated incentives). Similarly, Bae, Lee, and Chun [8] distinguished between "unconcerned" users, who prioritize comfort, and "active energy-savers". Static rules cannot accommodate this variance. AI enables the shift toward adaptive and personalized systems that interpret individual consumption needs and patterns, a capability shown to achieve energy savings of 7-52% in office cooling scenarios compared to static controls [50, 83].

2.7.2 Algorithmic Enablers of Personalization

Research emphasizes viewing AI and gamification as synergistic tools: gamification makes conservation approachable, while AI provides the personalization required to make it relevant [82, 93, 97]. To achieve this, the literature identifies three primary algorithmic strategies:

1. **Hybrid Prediction and Decision:** This approach decouples forecasting from action. Severiche-Maury and Arrubla-Hoyos [83] propose a model where a deep learning algorithm (e.g., LSTM - Long Short-Term Memory) forecasts usage, while a Reinforcement Learning (RL) agent acts as the decision engine. The RL agent learns from user feedback, continuously optimizing recommendations to balance comfort and efficiency [7, 30, 55, 83]. In a gamified context, RL agents can act as an adaptive optimization mechanism, fine-tuning the selection of energy-saving actions to individual user profiles to maximize relevance, whereas prediction models establish personalized reference baselines, enabling the generation of achievable data-driven goals rather than generic static targets.
2. **Pattern Mining and Anomaly Detection:** Unsupervised learning techniques, such as FP-Growth, are used to mine historical data for association rules (e.g., establishing that if event A occurs, event B follows). Concurrently, clustering algorithms detect abnormal outliers to flag waste in real-time or monitor appliance health by identifying irregular fluctuation patterns [23, 27, 40, 67]. Within gamified systems, these algorithms enable context-aware intervention triggers. Detected anomalies initiate real-time remedial challenges (corrective feedback loops) to normalize consumption, while association rules identify positive behavioural sequences, allowing the system to reinforce habit formation through multi-step reward mechanisms.

3. **Generative AI (LLMs):** The emergence of Large Language Models (LLMs) enables systems to generate natural language advice that is both explainable and persuasive [103]. Frameworks like GUIDE use domain-adapted LLMs to translate technical profiles into relatable advice, although privacy concerns are increasingly driving a shift toward local edge-based LLM implementations [18, 70]. In the context of gamification, LLMs can serve as a semantic contextualization layer. They convert quantitative energy metrics into persuasive narratives (thematic framing) to enhance user agency, while providing distinct empathetic feedback that simulates the social scaffolding necessary for sustained behavioural change.

2.7.2.1 Recommender Systems: The Central Engine

While the preceding algorithms supply the underlying intelligence, they remain isolated tools. Recommender systems serve as the overarching software architecture that unifies them. By integrating these disparate capabilities, a recommender system translates raw computational insights into a single synchronized intervention strategy [39].

Formalizing this process, the fundamental objective of a personalised recommendation system is to solve an optimization problem. Its objective is to maximize a utility function, $u(c, s)$ (Eq. 2.1), which identifies the optimal action s (e.g., an energy-saving action), from a set of candidates S , that maximizes utility for a specific user c , from a set of users C [55].

$$s_c = \arg \max_{s \in S} [u(c, s)] \quad (2.1)$$

To approximate this utility, systems typically employ four primary filtering strategies: content-based filtering (recommending actions similar to those previously accepted), collaborative filtering (recommending actions successful among similar peers), knowledge-based filtering (adhering to explicit user constraints, such as "no laundry after 10 PM") and utility-based recommendation (prioritizing actions based on calculated benefits, such as monetary savings) [27, 55, 66, 80].

2.7.2.2 Evolution of System Architectures

The implementation of these systems has evolved from rigid rule-based logic towards fully dynamic architectures. Bridging these paradigms is the hybrid approach, which integrates data-driven ML predictions with explicit rule-based constraints.

A prominent example of the hybrid approach is the Context-Aware Smart Energy Recommender (CASER) developed by Sitoula et al. [87]. CASER pairs support vector machine regression for forecasting with a rule repository module that accounts for specific tariff structures. Unlike generic nudges, CASER delivers highly personalized financial recommendations, such as suggesting load-shifting to cheaper time slots. As shown in Fig. 2.19, the interface explicitly visualizes the economic incentive alongside real-time data.

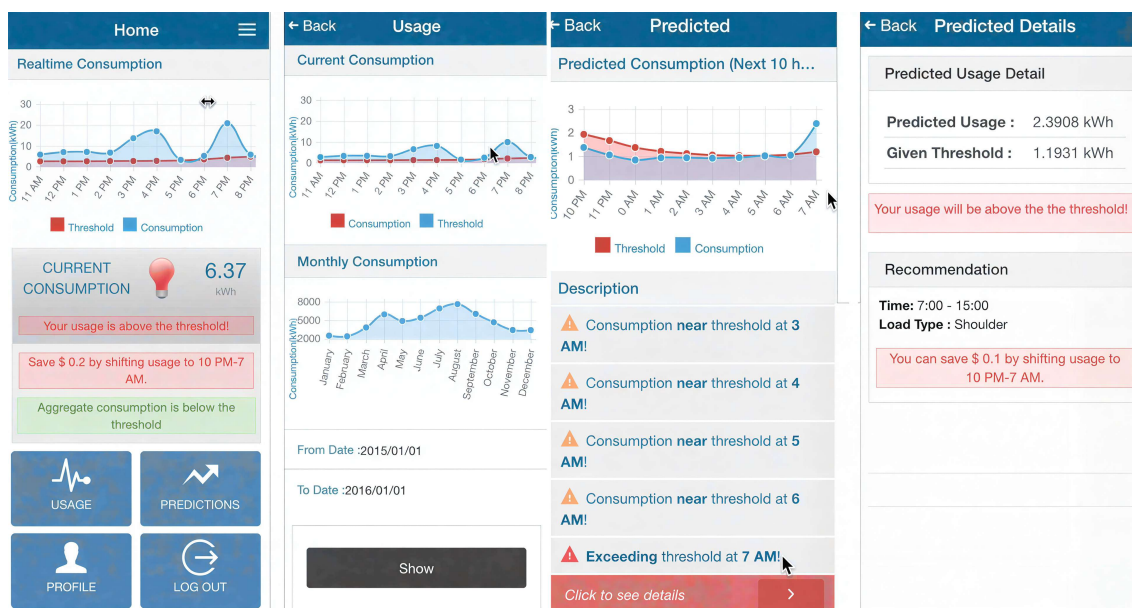


Figure 2.19: CASER mobile app interfaces showing real-time consumption against the substation threshold and predicted energy usage [87].

Despite its effectiveness, CASER's reliance on a static rule repository highlights a critical limitation: fixed rules can be brittle, often failing when user behaviour does not fit the pre-programmed economic logic. To address this, Alsalemi et al. [5] proposed a modular recommender system designed for architectural flexibility. Unlike monolithic systems that treat all energy usage similarly, this approach decomposes the problem into distinct contexts using specialized sub-components: an amplitude module (targeting high power consumption), a frequency module (profiling usage patterns) and an occupancy module (flagging active appliances in empty rooms). This modularity allows for "graceful adaptability". By generating advice based on specific behavioural contexts rather than generic price thresholds, the system provides more relevant interventions, with estimates suggesting potential annual savings of 10%.

The final leap in this evolution is the move toward RL from Human Feedback (RLHF). Even modular systems often rely on pre-defined logic that struggles to keep pace with evolving user habits. RLHF addresses this by continuously refining the utility function $u(c, s)$ based on real-time user reactions (e.g., accepting or dismissing a suggestion). This dynamic loop aligns the system's policy not just with technical efficiency, but with the complex, subjective and changing preferences of the human user [14].

2.8 Advanced Architectural Frameworks for Sociotechnical Systems

The algorithmic strategies discussed in the previous section supply the underlying computational intelligence, but integrating them with physical IoT networks and psychological gamification layers introduces severe architectural complexity. This section shifts the focus from individual algorithms to overarching system design. To ensure these disparate components function cohesively

without compromising user trust, the literature suggests moving beyond isolated interventions to view the building as a holistic sociotechnical system. The following subsections review the conceptual blueprints, edge-computing architectures and orchestration frameworks necessary to build these integrated environments.

2.8.1 Conceptual Architectures and Design Principles

Andolfi et al. [6] provide the blueprint for this integration, proposing a conceptual architecture (Fig. 2.20) governed by five core design principles (Table 2.5). This model explicitly links technical clusters, such as Actuation, Database Management and Scheduling, with a Rewarding/Reinforcing cluster designed to sustain user engagement, establishing a foundation for systems that are both technically efficient and socially engaging.

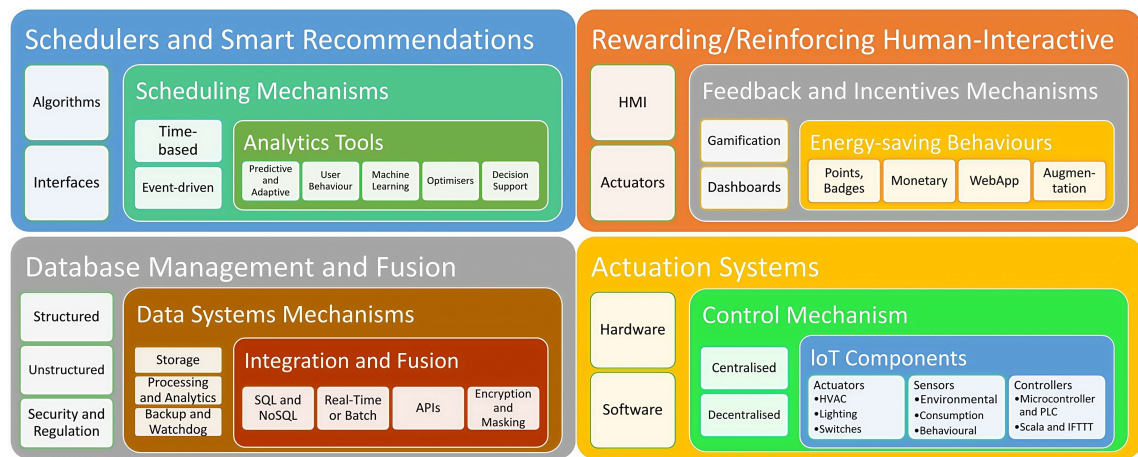


Figure 2.20: Conceptual Building Energy Management Systems (BEMS) architecture [6].

Table 2.5: Andolfi et al. [6] Design principles for sociotechnical BEMS.

Design Principle (DP)	Description
DP1: Build engagement	Reward interactions with personalised suggestions.
DP2: High information standards	Ensure data accuracy, reliability and timeliness, and regulatory compliance (e.g., GDPR ⁷ - General Data Protection Regulation).
DP3: Contextualisation	Integrate user behaviour with external data (e.g., weather, energy prices).
DP4: Flexibility and Reactivity	Dynamically adapt to external real-time environmental changes and user behaviours.
DP5: Efficiency by design	Use optimization algorithms, AI predictive controls and fault detection.

2.8.2 From Concept to Actionable Recommendation

Moving from concept to implementation, the Energy Management for Smart Spaces (EMSS) framework by Dasappa, G, and Somu [21] advances the field by fusing heterogeneous data sources (e.g., sensors and cameras) to generate high-quality recommendations. Crucially, EMSS addresses the explainability challenge, presented in the Subsection 2.5.2, by ensuring all advice meets four criteria: it must be actionable (e.g., "Turn OFF the AC"), explainable (providing the rationale, e.g., "Outdoor temperature is lower than indoor"), persuasive (quantifying benefits, e.g., "Save \$2.50") and personalized (learning from acceptance rates). As shown in Fig. 2.21, the dashboard visualizes these insights, while the system innovatively employs synthetic data generation to overcome the "cold start" problem (the inability to provide accurate recommendations for new users lacking historical data) inherent in training personalized AI models [21, 76, 110].

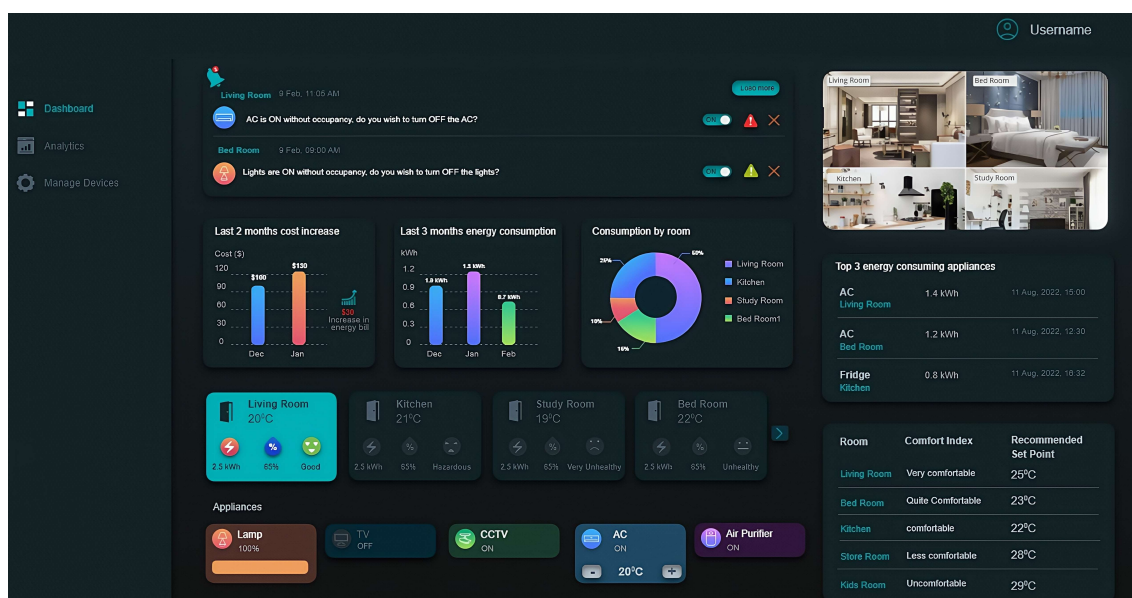


Figure 2.21: WebUI dashboard of the EMSS framework, showing analytics and recommendations [21].

2.8.3 Digital Twins and Privacy-Preserving Edge Computing

Advanced frameworks have begun to leverage digital twins (virtual representations of physical objects or systems that update in real-time using sensor data) to enhance context awareness while addressing privacy concerns. Sayed et al. [81] introduced a system that constructs a real-time 3D replica of a living space directly on an edge device (Fig. 2.22). By processing data locally rather than in the cloud, the system delivers context-aware recommendations (Fig. 2.23) that are both privacy-preserving and highly accurate. The system flow (Fig. 2.24) demonstrates how this local processing pipeline achieves an 80% acceptance rate, validating the efficacy of edge-based sociotechnical systems.

⁷<https://gdpr-info.eu/>

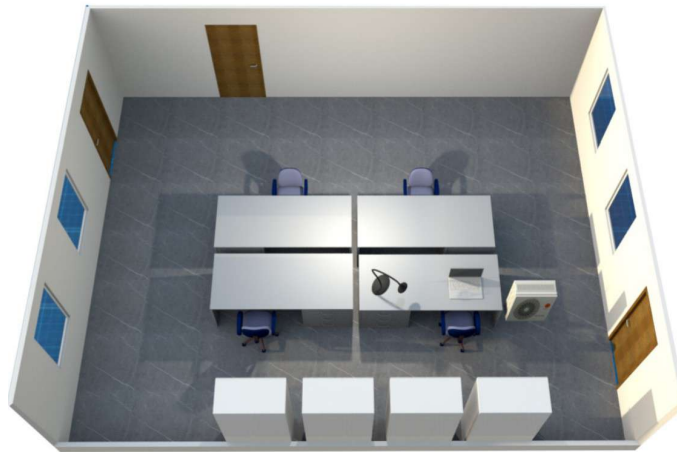


Figure 2.22: 3D model of the lab space created in "Sweet Home 3D" for the digital twin implementation [81].

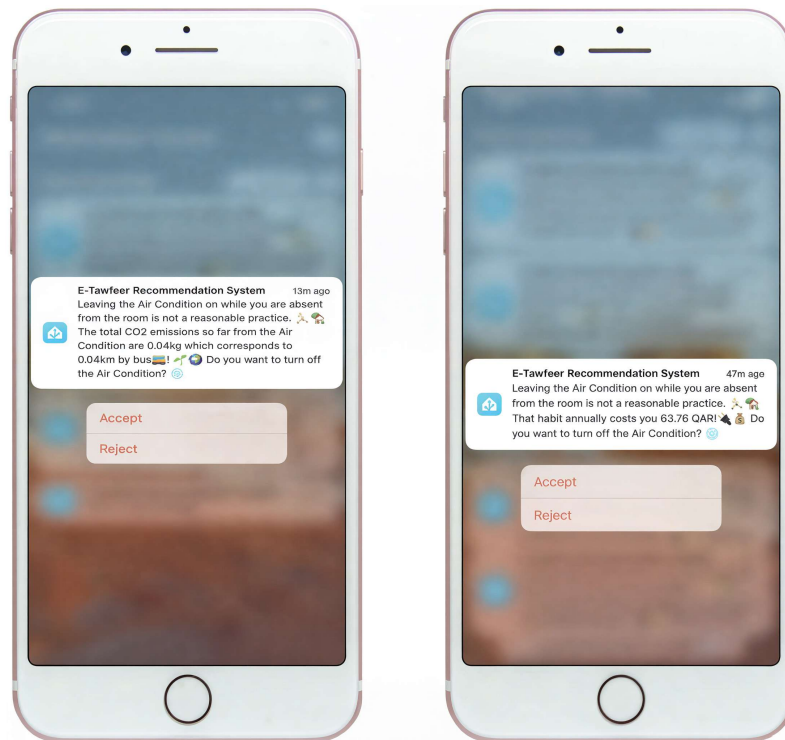


Figure 2.23: Examples of (left) ecological and (right) economical persuasive recommendations sent to the user's mobile device [81].

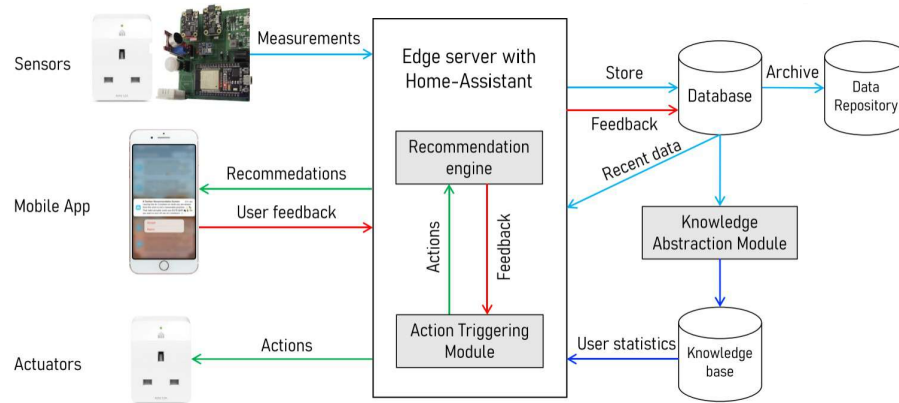


Figure 2.24: Sayed et al. [81] system recommendation generation process.

2.8.4 Intelligent Scenario Orchestration

Finally, the most advanced frameworks move beyond single recommendations to intelligent orchestration. Martins et al. [59] proposed a smart module combining LSTM and Multilayer Perceptron networks to learn complex user routines. Rather than suggesting isolated actions, the system suggests entire scenes (e.g., "Workday Routine", "Adaptive Acclimatization" or "Unoccupied Room") tailored to balance user comfort with energy efficiency. Critically, this framework addresses the issue of user autonomy by allowing users to set "auto-acceptance thresholds", providing a flexible balance between full automation and manual control (Fig. 2.25).

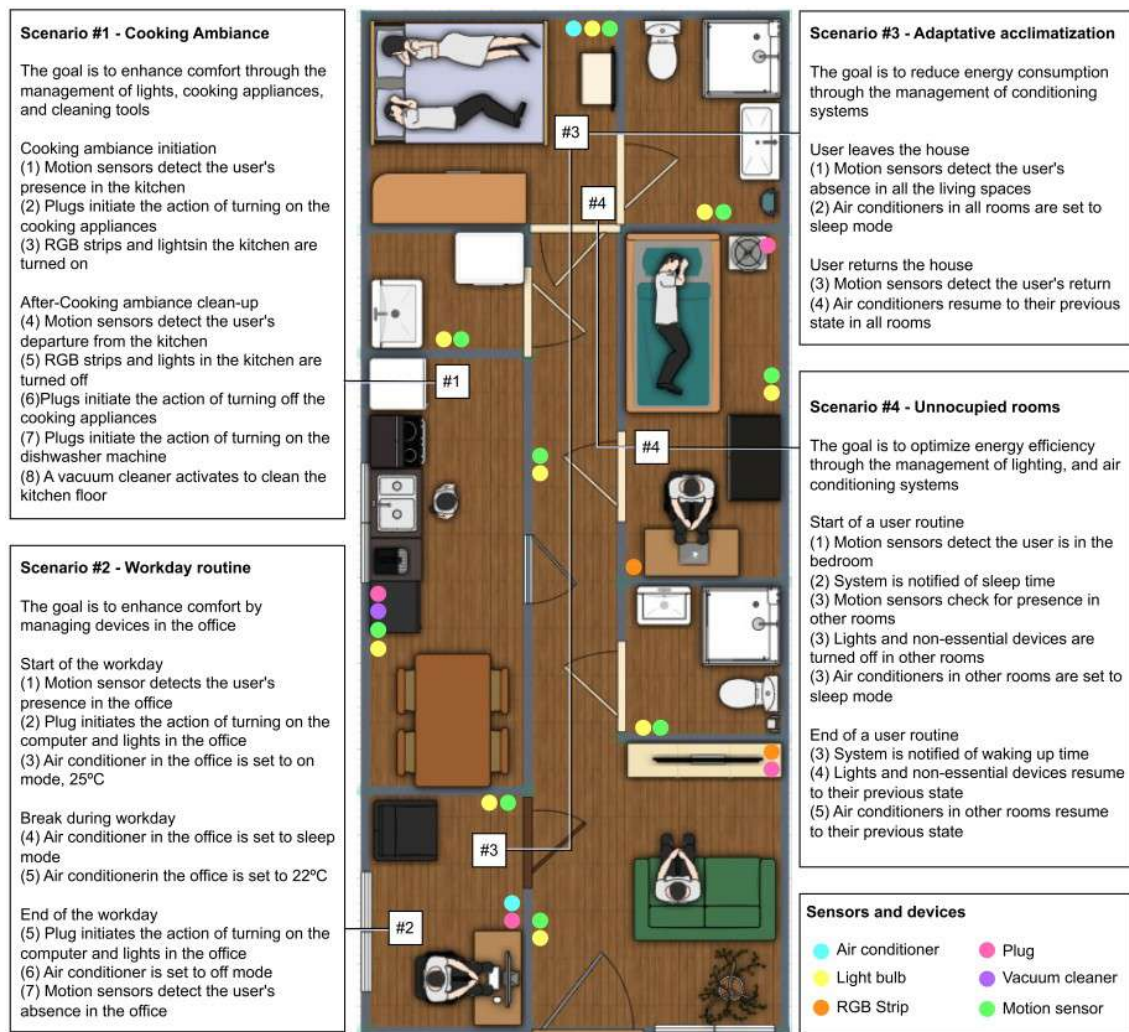


Figure 2.25: Floor plan illustrating four AI-driven scenarios for comfort and energy efficiency [59].

2.9 Empirical Validation: Real-World Case Studies

Transitioning from architectural blueprints to practical applications, this section examines how sociotechnical frameworks perform under real-world conditions. The following case studies provide empirical evidence of how the physical convergence of IoT, gamification and AI effectively solves practical challenges in verifying user actions, optimizing intervention timing and sustaining long-term engagement.

2.9.1 Closing the Intention-Action Gap: Verification via IoT

A primary challenge in behavioural interventions is reliance on self-reported data, which is often inaccurate. To address this, Chen et al. [15] developed the BI-Tech system (Figs. 2.26, 2.27). Unlike passive monitoring systems, BI-Tech actively uses IoT sensors to verify that users have

physically performed the energy-saving tasks they claimed. This synchronized validation creates a closed feedback loop as users are only rewarded for verified actions. By coupling this strict verification with app-based notifications and gamification, the system achieved a verified 23.3% decrease in non-HVAC consumption, proving that sensor-backed accountability significantly improves compliance.

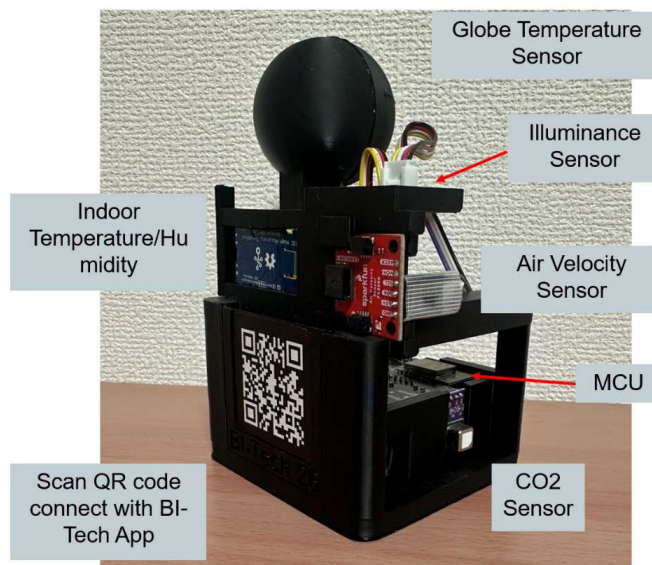


Figure 2.26: IoT sensors used for the BI-Tech platform [15].

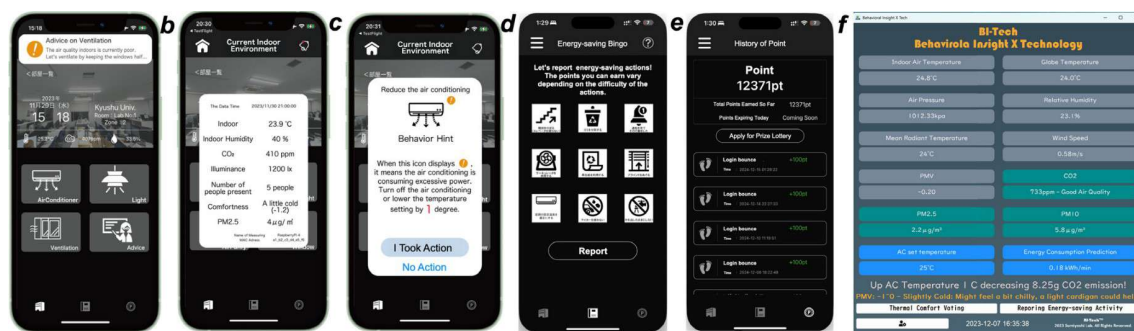


Figure 2.27: BI-Tech platform interfaces [15].

2.9.2 Optimizing Timing and Persuasion: The $(EM)^3$ Framework

While BI-Tech focuses on verifying action, the $(EM)^3$ framework focuses on the optimal timing and framing of the request itself (Fig. 2.28) [38, 41, 44, 78, 79, 96]. This framework applies the concept of micro-moments, an adaptation of Google's⁸ concept describing spontaneous smartphone interactions, to detect specific instances of energy waste (e.g., lighting left on as the last user exits a zone) [41, 79, 96]. By detecting specific windows of opportunity, the system injects

⁸<https://www.google.org/>

a timed recommendation that is far more likely to be accepted than a generic daily summary [78, 80].

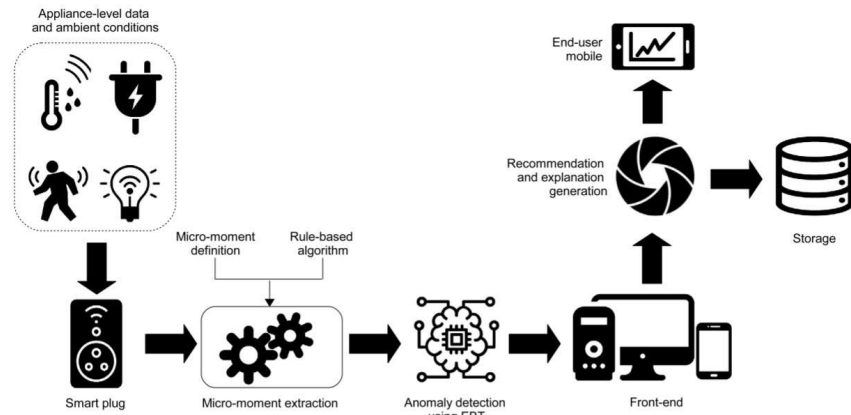


Figure 2.28: $(EM)^3$ system workflow leveraging micro-moments for behavioural change [41].

As shown in Fig. 2.29, the framework’s reference implementation visualizes this logic alongside real-time consumption graphs. This design proved highly effective: recommendations containing both explainability and persuasion achieved a 70% acceptance rate, compared to only 51% for plain recommendations. This provides robust empirical evidence that user compliance is directly linked to trust, further establishing explainability not merely as a technical feature, but as a fundamental psychological prerequisite for effective behavioural intervention.

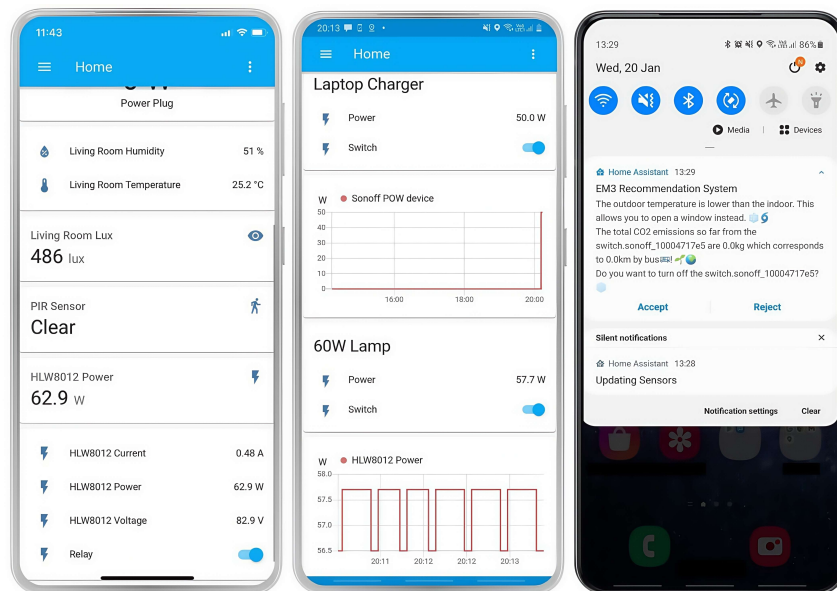


Figure 2.29: $(EM)^3$ mobile interface showing environmental data (left), consumption visualization with appliance control (middle) and generated recommendations (right) [41].

2.9.3 Synthesizing AI and Gamification: Affective and Achievement Systems

Finally, advanced case studies explicitly fuse AI with gamification to create systems that are not just smart, but emotionally resonant.

HomeMascot (Fig. 2.30) leverages "affective computing" by personifying energy conservation through a virtual polar bear. A rule-based AI finite state machine (Fig. 2.31) links real-time consumption data to the mascot's mood, transforming abstract data into an emotional cue. This approach proved effective at increasing energy awareness, with 90% of test households reporting a more positive attitude toward monitoring compared to standard graphs [75].

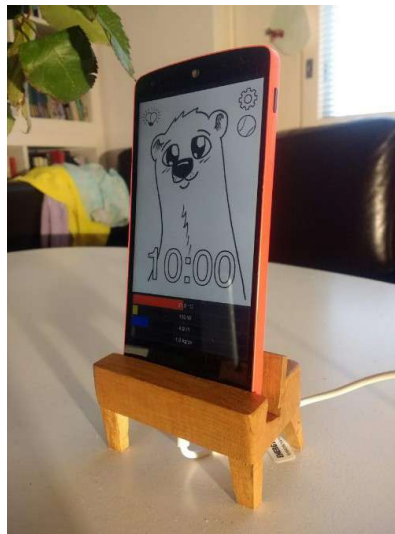


Figure 2.30: HomeMascot's system interface [75].

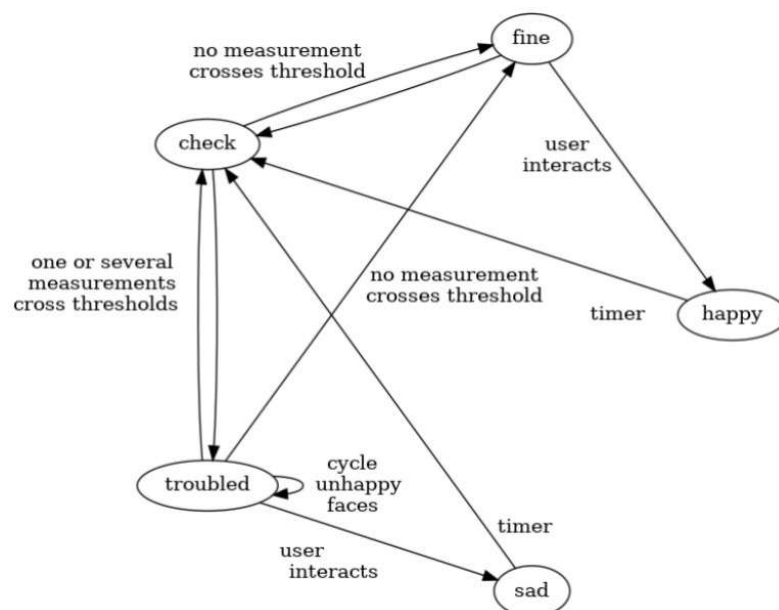


Figure 2.31: HomeMascot's finite state machine linking AI states to the mascot's mood [75].

Taking a more structured approach, Chen, Chang, and Wu [16] developed an adaptive smart campus system. This platform pairs a hybrid AI backend (Random Forest and Expert System) with a comprehensive gamified achievement system (Figs. 2.32 - 2.34). Unlike basic point systems, this design was explicitly mapped to SDT criteria (Table 2.6). By aligning game mechanics with psychological needs for autonomy and competence, the system significantly boosted user motivation and self-efficacy while maintaining high usability scores of 84 out of 100.

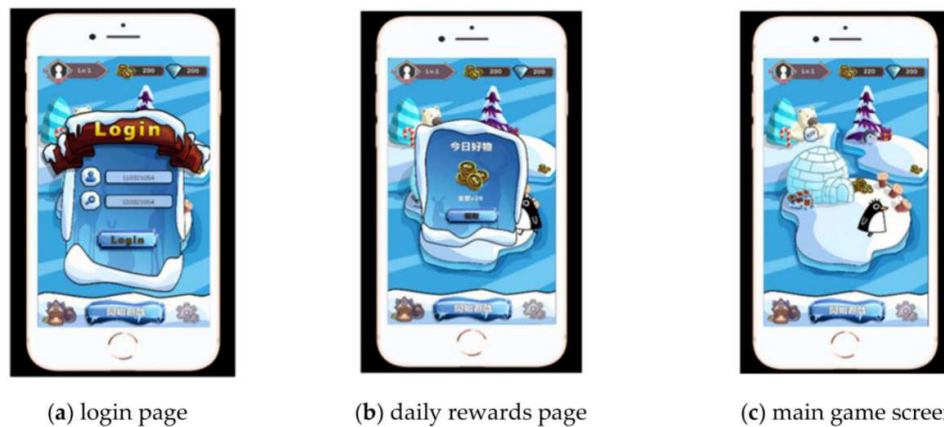


Figure 2.32: Gamification interfaces from the campus system, showing (a) the login page, (b) daily rewards page and (c) main screen [16].



Figure 2.33: Gamification interfaces from the campus system, showing (d, e) the shop page and (f) daily shared energy-saving tips page [16].

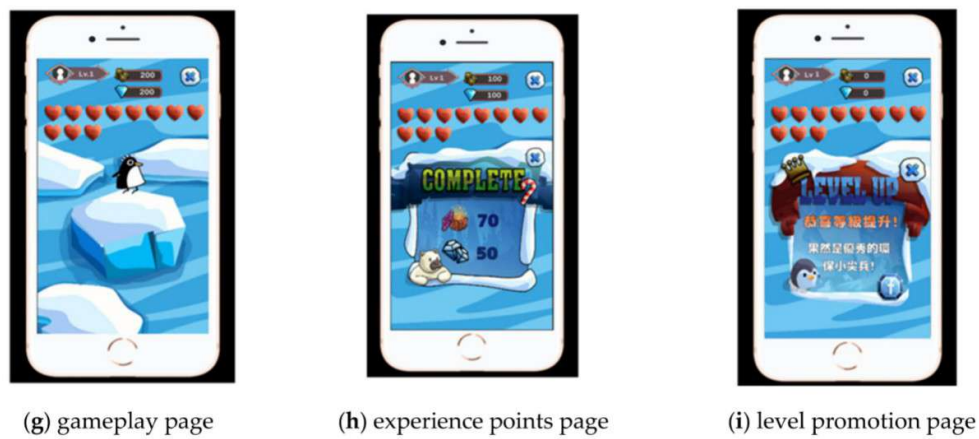


Figure 2.34: Gamification interfaces from the campus system, showing (g) the gameplay page, (h) experience points page and (i) level promotion page [16].

	Autonomy	Relatedness	Competence	Self-efficacy	Social Learning Theory	Operant Conditioning
(b)						●
(d) 、 (e)						●
(f)					●	
(g)	●					
(h)			●	●		●
(i)		●	●	●		

Table 2.6: Theoretical mapping of the gamification mechanics employed by Chen, Chang, and Wu [16] to their respective psychological drivers.

This synergy is further exemplified by commercial applications like Greenely (Fig. 2.35), which leverages IoT data to offer personalized feedback within a polished engaging framework, bridging the gap between academic theory and consumer adoption [82].

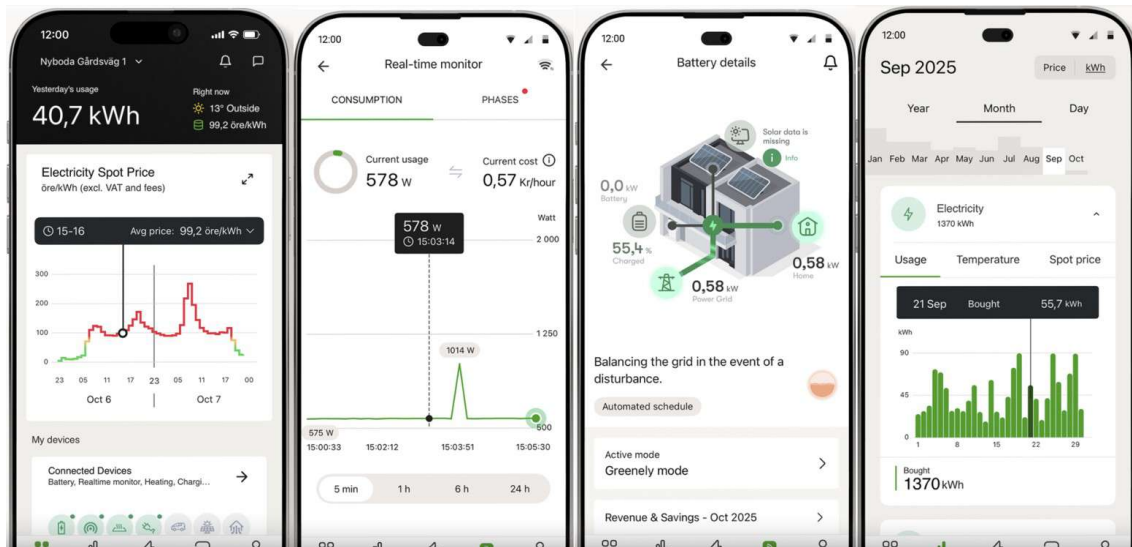


Figure 2.35: Greenely app interface from [Google Play Store](#).

2.10 Research Gaps and Dissertation Positioning

The review of the state of the art leads to a polarized conclusion. On one hand, the individual technologies required for energy efficiency (IoT monitoring, AI and gamification) have reached a high level of maturity. On the other hand, the integration of these technologies remains fragmented. Current research tends to treat these domains as separate silos, resulting in systems that are technically optimized but socially fragile.

This final section synthesizes the critical shortcomings of the state of the art, explaining why current approaches struggle to sustain long-term engagement. It specifically addresses five persistent failures: the lack of synergy between AI and gamification, the short-term validation bias, the reliance on synthetic data over real-world validation, the potential for toxicity in workplace competition and the persistent gap in data comprehension. Consequently, it delineates the design strategies adopted in this dissertation to overcome these barriers and build a more robust solution.

2.10.1 Critical Limitations: Why Current Approaches Fail

Despite the proliferation of smart energy systems, the literature reveals five persistent failures that prevent these solutions from achieving lasting impact:

1. **The Missing Synergy of AI and Gamification:** A clear dichotomy exists where AI is used for building control and gamification for user engagement, with little intersection. Current systems rarely use AI to personalize the gamified experience itself. As noted by Mabalay [57] and Abdurahmanovic and Cadenbach [1], this results in static interfaces that fail to account for user heterogeneity, leading to reward fatigue where users disengage once the novelty fades.

2. **The Validation Gap (Short-termism):** The field suffers from a significant novelty effect bias. Reviews indicate that most studies only last a few weeks, failing to capture the inevitable engagement decay [97]. Consequently, there is a lack of empirical evidence regarding whether personalization can sustain motivation in the long term or if savings dissipate as users revert to old habits [2, 7].
3. **The Methodological Gap (Synthetic vs. Real-World):** A widespread reliance on synthetic datasets for training AI models limits ecological validity. Advanced models (e.g., Graph Neural Networks) are often tested on static datasets rather than the chaotic variability of real human behaviour [48]. Sociotechnical frameworks remain largely conceptual, lacking practical implementation in operational settings, creating a critical need to move beyond simulation to capture the irrationality of actual occupants [6, 21].
4. **The Autonomy and Toxicity Gap:** Highly automated black-box systems (where decision logic is hidden from the user) are frequently rejected by users who feel a loss of control or fear surveillance from cloud-centric architectures [34, 85]. Furthermore, in workplace settings, competitive leaderboards can become toxic, demotivating low-performers and triggering a "boomerang effect" where efficient users may increase consumption to match the norm [54, 90].
5. **The Comprehension and Usability Gap:** Finally, while data availability has increased, data accessibility has not. Raw technical metrics (kWh) remain abstract to non-expert users [108]. Furthermore, complex dashboards often induce cognitive overload, leading users to reject systems that lack intuitive relatable framing [3, 24].

Beyond these theoretical gaps, practical implementation hurdles remain a barrier to widespread adoption. As summarized in Table 2.7, Wang et al. [97] identify these operational challenges and propose corresponding strategic mitigations that serve as foundational design constraints for this dissertation.

Table 2.7: Challenges and mitigation strategies in gamified systems [97].

Challenge	Description	Mitigation Strategy
Sustaining Engagement	Users lose interest over time as novelty wears off.	Introduce adaptive systems that evolve with user behaviour.
Privacy Concerns	Tracking energy usage may raise privacy issues.	Use data anonymization and secure user consent.
Digital Divide	Lack of access to smart technology in underserved communities.	Deploy inclusive low-tech support layers (e.g., SMS alerts, paper leaderboards or offline tracking charts) alongside smart apps.
Overemphasis on Competition	Competitive systems may discourage low-performing users.	Incorporate collaborative features alongside competition.
Implementation Costs	High costs of integrating gamification with smart systems.	Provide incentives like tax credits and modular system designs.

2.10.2 Strategic Positioning: Harnessing High-Potential Innovations

In response to these identified gaps, this dissertation does not seek to invent new psychological theories, but to synthesize the most efficient technological mechanisms identified in the review. The proposed solution integrates the following architectural pillars to ensure the system is both technically robust and socially sustainable.

2.10.2.1 Solving the Synergy Gap: Hybrid Intelligence (RLHF)

To counter reward fatigue, the proposal adopts the RLHF approach [14]. Unlike static systems, this innovation uses the user's reaction (acceptance or rejection of a nudge) as a reward signal. This allows the AI to dynamically tailor its interventions to the individual user, ensuring the system remains engaging even after the initial novelty period.

2.10.2.2 Solving the Privacy Gap: Edge Computing Architecture

To address the ethical barrier of surveillance and the black-box problem, the system adopts a decentralized edge computing approach [41, 81]. By processing behavioural data locally (e.g., on a Raspberry Pi) rather than in the cloud, the solution aligns with privacy-by-design principles. This ensures that granular personalization is achieved without exposing sensitive user habits to third-party servers.

2.10.2.3 Solving the Comprehension Gap: Cognitive Accessibility

To bridge the gap between abstract data and user understanding, the interface directly implements the disaggregated feedback and framing strategies selected and detailed previously in Subsection 2.4.2. By operationalizing these specific methods, the system translates technical units into relatable values (e.g., monetary cost or environmental metaphors) to maximize persuasive power and reduce cognitive load [28].

2.10.2.4 Solving the Validation Gap: Real-World Context Awareness

To move beyond synthetic data, the solution leverages a context-aware recommender system driven by real-time IoT sensors [31, 87]. By detecting specific micro-moments (e.g., an empty room with high consumption), the system validates its model against real-world unpredictability, ensuring that advice is actionable and physically verified rather than simulated.

2.10.2.5 Solving the Toxicity Gap: Self-Referential Design

Finally, acknowledging the risks of workplace tension, the gamification strategy prioritizes historical benchmarking, supplemented by cooperative normative baselines where data permits. This hybrid approach allows users to compare performance against their own history, shifting the focus from interpersonal rivalry to collective self-improvement. As suggested by Tekler, Low, and Blessing [90], this design eliminates the unfair comparison barrier and promotes intrinsic motivation without the risk of social friction.

Table 2.8 synthesizes the key studies analysed in this section, explicitly mapping their findings to the specific technical components adopted in the proposed prototype.

Table 2.8: Overview of related studies and their specific influence on the proposed prototype.

Study	Key Contribution to Prototype Design
[87]	Validates the shift from static dashboards to context-aware recommender systems. It demonstrates that combining user history with AI helps solve engagement decay by delivering personalized suggestions.
[90]	Suggests that for professional/workplace settings self-referential feedback (historical comparison) is safer and more effective than social comparison. It advocates for achievements (goals) over direct competition to avoid toxicity.
[73]	Quantifies novelty decay (stagnation after 7 days) and proves that dynamic gamification (e.g., themed challenges) can recover user retention to $\approx 90\%$.
[31]	Establishes the baseline effectiveness of real-time appliance-level feedback, showing it can drive 18-25% savings alone. It reinforces the need for actionable granular data rather than aggregate monthly bills.
[34]	Provides the technical basis for AI-based recommender systems. It demonstrates that modelling environmental context allows the system to intelligently balance energy efficiency with user comfort.
[14]	Introduces a hybrid RL approach that integrates Human Feedback to refine recommendations. It proves that allowing users to correct or guide the AI (human-in-the-loop) creates more robust models when reward functions are difficult to fully define mathematically.

Chapter 3

GreenShift

Following the sociotechnical gaps identified in Chapter 2, this chapter details the architectural design and implementation of the proposed research platform. As highlighted in the literature, existing solutions often suffer from a one-size-fits-all approach and privacy vulnerabilities due to cloud-centric processing [1, 81].

To address these limitations while ensuring broad accessibility and community adoption, the proposed system, named GreenShift¹, is designed as a custom component (plugin) for Home Assistant² (HA). This architectural choice shifts the paradigm from a siloed standalone application to an integrated ecosystem approach. By building directly atop the HA platform, GreenShift becomes inherently modular, open-source friendly and immediately accessible across both desktop and mobile devices via existing client interfaces. Crucially, to bridge the gap between academic research and practical deployment, GreenShift has been submitted to the Home Assistant Community Store³ (HACS), a premier distribution hub for third-party HA integrations, allowing the global user base to easily discover, install and autonomously update the system without requiring manual codebase manipulation.

3.1 Architectural Requirements and Design Principles

To ensure the proposed research platform can be seamlessly integrated into any environment utilizing HA, the architecture is governed by strict design principles aimed at scalability and hardware agnosticism. Because user infrastructures vary significantly, ranging from single-sensor setups to complex multi-protocol networks, the system is engineered to function robustly without imposing rigid hardware prerequisites.

Central to this adaptability is a highly modular decision-making architecture. By strictly decoupling the data acquisition layer from the intelligence logic, the AI model remains effective regardless of the specific sensors available in a given environment. This ability to dynamically

¹The source code for GreenShift is publicly available at: <https://github.com/RafaelNTeixeira/GreenShift>

²<https://www.home-assistant.io/>

³<https://hacs.xyz/>

adjust its state space to accommodate heterogeneous physical infrastructures is a critical feature for generalizability, which is detailed further in Subsection 3.5.3.

Beyond modularity, the system prioritizes privacy, interoperability and speed. It implements privacy-by-design through edge processing, where core data is aggregated locally via a home energy management system to mitigate surveillance risks and ensure data ownership. Interoperability is achieved by abstracting the heterogeneity of IoT devices across multiple protocols, such as ZigBee and MQTT. Finally, low latency is maintained through direct subscription to the internal event bus, enabling micro-moment interventions by removing network overhead between the state machine and the intelligence logic.

3.2 High-Level Architecture and Functional Flow

This system bridges the gap between passive energy monitoring and active user engagement. By integrating real-time physical data with intelligent decision-making, the architecture creates a dynamic ecosystem where the building not only senses its environment but actively nudges the user toward better habits. As shown in Fig. 3.1, this is achieved through four streamlined layers that maintain a continuous bidirectional flow of information.

1. **Physical Layer:** The sensor network responsible for raw data acquisition.
2. **Core Platform Layer:** A local instance of HA acting as the central hub for device management and state normalization.
3. **Intelligence Layer:** A plugin running natively within HA that hosts the AI model and gamification logic and manages application-specific data persisted via local databases.
4. **User Interface Layer:** The HA frontend (Lovelace), delivering custom dashboard cards for visualization and interaction to both mobile and desktop clients.

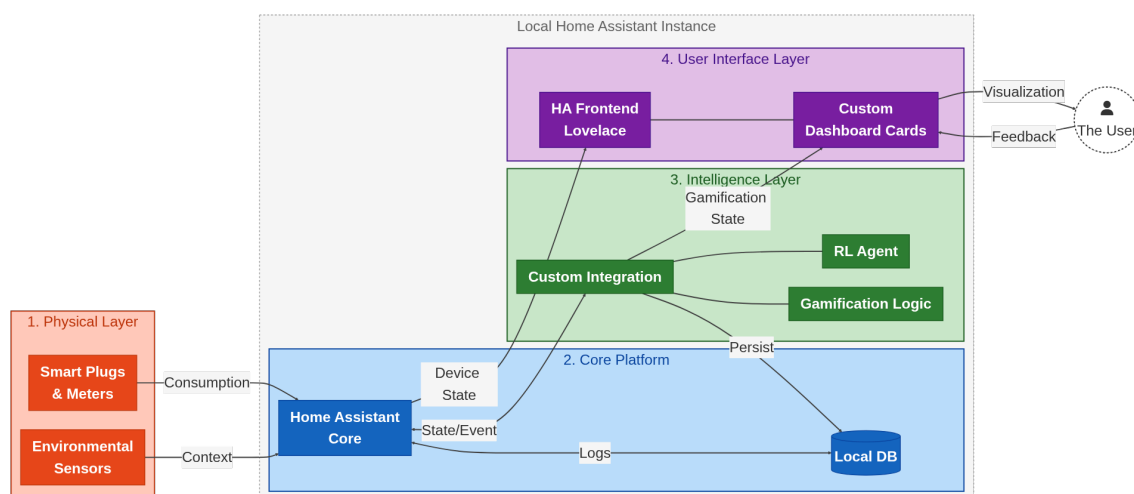


Figure 3.1: High-level system architecture illustrating the data flow from physical sensors through the HA Core to the AI-driven Intelligence Layer and the user interface.

This structure enables a continuous perceive-process-act cycle with minimal latency. It begins with ingestion, where sensors update entity states representing telemetry (e.g., Power P , Temperature T , Occupancy O) within the HA Core. GreenShift listens for these state changes via the internal event bus. During inference, the embedded Python-based RL agent evaluates the current state (S_t) (Eq. 3.1) and responds by updating virtual entities (software-defined variables used to store feedback or control signals) or issuing notifications/alerts. These updates are then automatically reflected in the user interface.

3.3 Physical Layer

To move beyond the limitations of traditional monthly billing, this layer provides granular appliance-level visibility through a distributed edge-computing topology. It is engineered to be highly agnostic, ensuring that the system is not bound to specific manufacturers or proprietary communication protocols. This flexibility relies on a cohesive integration of local computational nodes, multi-protocol communication interfaces and a tiered sensor network.

At the core of this physical infrastructure is the edge gateway. The architecture requires a local host, typically instantiated as a single-board computer or a local micro-server, capable of natively executing the HA Core. Deploying the processing gateway within the local area network ensures that the analytical logic and data persistence mechanisms operate continuously on the edge. This architectural decision strictly limits dependency on cloud-based computation, thereby preserving user privacy, ensuring data sovereignty and minimizing control latency.

To facilitate seamless interaction between the edge gateway and the physical environment, the architecture supports multi-protocol heterogeneous networks. By leveraging the integration layer of the core platform, the system can simultaneously ingest telemetry from devices communicating via diverse standards, such as Wi-Fi, Bluetooth Low Energy (BLE) or ZigBee. These heterogeneous signals are subsequently translated into standardized state events, establishing a uniform data pipeline for the Intelligence Layer regardless of the underlying physical medium.

Finally, to provide the high-resolution contextual data required by the RL agent, the topology incorporates a three-tiered sensor network. This network comprises whole-house smart energy meters for aggregate load profiling, appliance-level smart plugs for targeted device monitoring and environmental sensors capturing temperature, humidity, luminosity and occupancy metrics ($T_{in}, H_{in}, L_{in}, O_{status}$). This granular visibility allows the system to distinguish between waste and necessary consumption. Although GreenShift is strictly optimized for local edge execution to maintain offline resilience, the architecture retains the flexibility to integrate optional cloud-based data sources, such as weather forecasting Application Programming Interfaces (APIs), accommodating environments with severe infrastructural constraints.

3.4 Core Platform Layer (Home Assistant)

The central nervous system of the architecture is HA, an open-source home automation platform. HA was selected over proprietary cloud solutions not only to ensure interoperability and local data sovereignty but also because it was already deeply integrated into both Fraunhofer's infrastructure and the private smart home ecosystems of the participants.

3.4.1 Integration Hub and State Machine

Since both the institute and the individual participants had previously adopted HA for their respective building management tasks, utilizing it as the foundation for this platform minimized deployment friction and leveraged existing technical knowledge.

In this architecture, HA functions as the central integration hub and state machine. It seamlessly interfaces with the Physical Layer through native and manufacturer-specific integrations, allowing for the effortless incorporation of diverse sensor ecosystems. By easily ingesting and mapping raw telemetry into standardized software entities, HA ensures that the Intelligence Layer interacts with a consistent structured state registry, completely independent of the underlying hardware.

3.4.2 Database Strategy and Data Management

To prioritize data sovereignty, user privacy and system performance, the architecture manages all data persistence locally, bypassing cloud environments entirely. By maintaining records strictly on the edge server, sensitive occupancy and behavioural data never leave the premises. Rather than relying solely on HA's native Recorder, GreenShift partitions its storage across three specialized repositories (two SQLite databases and one JSON state file) each tailored to specific operational requirements:

- **Short-Term Telemetry Database (SQLite):** Operates on a rolling 14-day data retention policy to minimize the digital footprint and maintain fast read/write speeds. It stores high-frequency time-series sensor readings (Power, Temperature, Humidity, Occupancy, Luminosity) as well as area-based aggregated statistics. This short-term repository provides the immediate historical context required for predictive modelling and real-time anomaly detection.
- **Long-Term Research Database (SQLite):** Enforces a 4-month (120-day) retention policy to support longitudinal study analysis without compromising the performance of the main telemetry database. It archives all application events and behavioural data, including generated tasks, user interaction logs, nudges sent, notification acceptance rates and the RL transition logs. Specifically, it stores the sequential tuples $(S_t, A_t, M_t, R_t, S_{t+1})$ necessary for evaluating the RL agent's performance post-intervention.

- **Persistent Agent State (JSON):** A lightweight highly accessible configuration file used to preserve the active "brain" and configuration of the system across host reboots. It maintains critical operational keys, including the continuously updating RL states, the current system phase (baseline vs. active), established baseline consumption metrics and the dynamically calculated diagnostic variables, such as the behaviour and fatigue indices.

3.5 Intelligence Layer

This layer represents the primary technical contribution of the system. To ensure high responsiveness and maintain operational independence, the analytical logic is architected as a Custom Component (or "Custom Integration") residing natively within HA.

In the HA ecosystem, a Custom Component is a Python package that executes directly within the HA Core's asynchronous event loop. Unlike external scripts that require API communication, this architecture allows the component to operate as a native extension, sharing memory and resources with the platform for optimal performance.

3.5.1 The Integration Logic

The component functions as the bridge between the physical sensors and the abstract gamification engine. Its operation is defined by three key lifecycle stages.

3.5.1.1 Component Registration

Upon system startup, the component registers itself with the HA Component Registry. This initialization process loads the Python-based gamification engine and establishes connections to the database tables. By registering as a native component, the Intelligence Layer becomes discoverable to the rest of the system, allowing other services (like the user interface) to interact with it seamlessly.

3.5.1.2 Virtual Entity Creation

To expose the calculated gamification data and system states to the frontend, the component instantiates its own sensor entities. Rather than tracking physical hardware, these standard HA entities are specifically generated by the software to represent computed metrics, gamification progress and contextual flags. Two representative examples include:

- ***sensor.green_shift_task_streak*:** A numeric entity generated by GreenShift that tracks and displays the user's consecutive days of completed tasks, serving as a primary gamification metric to foster continuous engagement.

- ***binary_sensor.workday_sensor***: A boolean entity (provided by HA's native Workday integration) utilized by the system to evaluate whether the current day is a holiday. This dynamically dictates the AI's scheduling rules and working-hours filtering for both task generation and notification dispatch.

3.5.1.3 Event-Driven State Management

The core runtime loop operates by natively subscribing to HA's internal event bus. Rather than relying on inefficient polling mechanisms, the component listens specifically for *state_changed* events broadcast whenever any physical device in the network updates its status. This allows the system to passively monitor the environment with near-zero latency. For instance, if a physical presence sensor transitions to "unoccupied" while a connected smart plug continues to report high power consumption, the component instantly captures these state changes. Critically, because this listener operates internally on the event bus, the data ingestion occurs without the overhead of HTTP API calls or external network requests. This real-time telemetry immediately updates the virtual dashboard sensors and is seamlessly fed into the RL agent's next evaluation cycle, enabling highly responsive evaluations such as triggering an anomaly nudge for wasted energy in an empty area.

3.5.2 Contextual Filtering: Environment and Working Hours Logic

To ensure the AI interventions remain relevant and actionable, the component implements a temporal filtering mechanism based on the deployment setting. During the initial setup, the administrator defines the environment topology as either a "Home" or "Office".

This configuration dictates the temporal boundaries of the Intelligence Layer. In an office setting, users can only realistically intervene and alter consumption habits during their active presence. To prevent the AI from generating tasks that cannot be fulfilled or learning from periods where users have no agency (such as nighttime standby loads or weekends), the administrator inputs the specific working days and daily operational hours (start and end times) directly through the integration's configuration forms. Additionally, the system utilizes HA's native Workday integration exclusively to detect public holidays, ensuring interventions are dynamically paused when the office is legally closed.

When configured for an office environment, the data fed into the baseline calculations, daily gamified tasks and the RL agent's evaluation loop is strictly filtered to encompass only these custom-defined working hours on valid working days (excluding identified holidays). Any consumption patterns occurring outside of this schedule are explicitly ignored by the behavioural logic. Conversely, in a residential setup, this working-hours filter is disabled, allowing the system to monitor and evaluate energy behaviour continuously across a 24-hour cycle, reflecting the constant nature of domestic energy management.

3.5.3 The AI Engine

To address RQ4 (see Section 1.3) while ensuring the system remains deployable across diverse physical environments, the intelligence system is driven by an RL decision agent. It utilizes the Q-learning algorithm to optimize intervention policies through a combination of contextual pattern recognition and active decision-making. Specifically, the agent sends notifications with personalized nudges and tips to improve energy behaviours.

Crucially, this decision engine employs a context-aware architecture. Unlike monolithic systems that require a fixed sensor array, this architecture is capable of handling missing data or actuators without requiring the core model to be retrained. The personalization problem is modelled as a Markov Decision Process defined by the tuple $\langle S, A, M, P, R, \gamma \rangle$, where M represents the hardware availability mask that dynamically restricts the action space based on the current sensor environment.

3.5.3.1 State Space: Handling Missing Sensors

The State Space (S_t) (Eq. 3.1) utilizes a "Zero-Padding with Existence Flags" strategy to ensure the learning agent can operate seamlessly across both fully instrumented laboratories and sensor-sparse environments. The augmented state vector is constructed by pairing optional sensor readings with a binary validity flag (F):

$$S_t = [(P_{total}, F_{total}), (P_{app}, F_{app}), (T_{in}, F_T), (H_{in}, F_H), (L_{in}, F_L), (O_{status}, F_O), I_{anomaly}, I_{behaviour}, I_{fatigue}, C_{area_anomalies}, T_{day}, D_{week}] \quad (3.1)$$

In this representation, P_{total} denotes the total real-time facility power and P_{app} represents the device with the current highest power consumption. Basic contextual variables, such as T_{in} (Global Indoor Temperature), H_{in} (Humidity), L_{in} (Luminosity) and O_{status} (Occupancy), capture the overall environmental state. Crucially, each of these physical metrics is paired with a binary availability flag ($F_x = 1$ if the sensor is present, 0 if absent). During inference, the algorithm learns to weigh the input features by their corresponding flags, effectively neutralizing the impact of missing hardware data.

Furthermore, the vector integrates high-level diagnostic indices: $I_{anomaly}$ flags global irregular usage patterns, $I_{behaviour}$ tracks historical user engagement and compliance, and $I_{fatigue}$ quantifies the risk of user desensitization to frequent notifications.

Crucially, while the core environmental variables (such as T_{in}) rely on global averages to maintain a fixed-length state vector, the AI agent does not base its decisions solely on these overall metrics. Relying purely on global averages across a deployment can easily mask localized inefficiencies. For instance, if a sensor in the garage registers 10°C while the living room sensor reads an excessively hot 30°C , a global average would yield a deceptively normal 20°C , causing the agent to miss significant energy waste.

To prevent this loss of resolution, the system leverages HA's native area grouping to calculate independent anomaly scores for each distinct physical zone. This localized contextual sensing is injected back into the state vector via $C_{area_anomalies}$, which dynamically counts the number of specific areas currently registering significant environmental or power deviations. Finally, time-based normalizations (T_{day} and D_{week}) provide the temporal context necessary for the agent to distinguish between peak and off-peak behavioural patterns.

3.5.3.2 State Discretization and Q-Table Memory Management

While the continuous state vector (S_t) provides high-resolution context for immediate evaluation, directly mapping continuous variables to a Q-learning algorithm would result in an unmanageably large state space. This phenomenon, known as the curse of dimensionality, would prevent the AI from converging on optimal policies within the duration of the study and would rapidly exceed the memory constraints of an edge-processing device like a Raspberry Pi.

To resolve this, the system applies a discretization strategy to map the high-resolution state vector into a constrained, finite set of discrete states. This compression shrinks the learning environment to a maximum of 1,200 possible unique states, allowing the Q-table to remain lightweight in memory and converge efficiently based on limited real-world interactions. The discretized state key is composed of six distinct contextual dimensions:

- **Relative Power Level (5 bins):** Instead of raw wattage, power is classified relative to the environment's learned baseline consumption. This crucial design choice ensures the model generalizes perfectly across different building sizes, from small apartments to large laboratories. The bins are: Standby ($< 5\%$ of baseline), Low (5-50%), Medium/Normal (50-150%), High (150-300%) and Peak ($> 300\%$).
- **Anomaly Level (4 bins):** The continuous global anomaly index ($I_{anomaly}$) is bucketed into None (< 0.25), Low (0.25-0.50), Medium (0.50-0.75) and High (> 0.75).
- **Fatigue Level (3 bins):** The user's notification fatigue index ($I_{fatigue}$) is categorized into Low (< 0.33), Medium (0.33-0.66) and High (> 0.66) to heavily dictate intervention thresholds.
- **Spatial Anomaly Flag (2 bins):** A boolean flag (0 or 1) indicating whether any localized area (e.g., a specific room) is currently registering an anomaly, simplifying the multi-area evaluation without tracking every area individually.
- **Temporal Context (5 bins):** The 24-hour cycle is segmented into functional behavioural periods: Night/Inactive (22:00-07:00), Morning Peak (07:00-10:00), Midday (10:00-14:00), Afternoon (14:00-19:00) and Evening Peak (19:00-22:00).
- **Occupancy Status (2 bins):** A binary indicator denoting whether the building or specific monitored zones are currently occupied (1) or unoccupied (0).

3.5.3.3 Action Space: Handling Missing Actuators

The Action Space (A) (Eq. 3.2) defines the maximum set of potential interventions:

$$A = \{a_{noop}, a_{specific}, a_{anomaly}, a_{behavioural}, a_{normative}\} \quad (3.2)$$

Where a_{noop} represents no operation (preventing user annoyance), $a_{specific}$ triggers targeted device recommendations, $a_{anomaly}$ initiates notifications for irregular consumption (e.g., devices left on in unoccupied areas), $a_{behavioural}$ suggests long-term lifestyle adjustments and $a_{normative}$ leverages social proof against historical baselines.

To ensure robustness across different hardware tiers, the agent applies an Action Validity Mask (M) before selecting an intervention. M is implemented as a dynamic boolean vector (evaluating to True or False) that strictly defines which actions are physically and contextually possible in the current evaluation cycle. During the action selection process, this mask completely filters out impossible commands. For example, if individual smart plugs are absent, $a_{specific}$ is flagged as False. Similarly, $a_{anomaly}$ requires sufficient historical data and a high anomaly index to become True, ensuring the agent only explores relevant and executable interventions.

3.5.3.4 Q-Learning Optimization and Action Selection

To formalize the learning process, the decision agent optimizes its policy π using a Q-learning algorithm. This algorithm learns a state-action value function, represented as $Q(s, a)$, which estimates the expected cumulative reward for taking a specific intervention a in a given context s . These values are stored in a lookup structure known as a Q-table, where each entry is iteratively updated as the agent interacts with the environment according to the Bellman equation (Eq. 3.3):

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[R + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \quad (3.3)$$

In this formula, α represents the learning rate (which adapts depending on the deployment phase), R is the observed reward, γ is the discount factor for future rewards and $\max_{a'} Q(s', a')$ represents the maximum estimated value of the subsequent state.

Because the user's reaction to a specific nudge is stochastic, the agent balances the exploitation of known effective nudges with the exploration of new interventions through an ϵ -greedy strategy. Crucially, the exploration rate (ϵ) is not static. It decays multiplicatively as the agent accumulates real user feedback, transitioning naturally from an exploratory phase to an exploitative one while being floor-capped to a minimum value (ϵ_{min}) to ensure long-term adaptability.

3.5.3.5 Dual-Phase Reward Architecture

The learning architecture operates in two distinct modes, employing different reward functions and hyperparameter tunings to ensure the agent is pre-trained before interacting with the user.

Phase 1: Shadow Learning (Baseline) During an initial 14-day baseline phase, the AI operates in a silent mode. It actively monitors the environment and simulates decisions but halts the execution step, meaning no notifications are sent to the user. This "shadow learning" allows the Q-table to be warm-started based on energy patterns and temporal context.

Because there is no human interaction, the shadow reward (R_{shadow}) (Eq. 3.4) is strictly an estimation based on conservative weights:

$$R_{shadow} = 0.5 \cdot S_{potential} + 0.3 \cdot C_{alignment} + 0.2 \cdot T_{score} \quad (3.4)$$

Where $S_{potential}$ represents the current deviation from the mean consumption, $C_{alignment}$ measures how well the selected action matches the environmental context (e.g., suggesting a device-specific task when a single appliance dominates the load) and T_{score} evaluates the appropriateness of the time of day.

To prevent the Q-table from over-committing to these estimated rewards, the update mechanism (Eq. 3.3) utilizes a significantly lower learning rate ($\alpha = 0.05$) compared to the active phase ($\alpha = 0.1$). Conversely, because simulated actions carry no real-world cost (no risk of user annoyance), the shadow exploration rate is set higher. Furthermore, the discount factor (γ) is forced to 0 during this phase. Since no real action is taken, the subsequent state is identical to the current state, as using standard future-value estimation would artificially inflate Q-values in a continuous feedback loop without empirical validation.

Phase 2: Active Intervention and Delayed Feedback Following the baseline period, the agent transitions to the active phase, where decisions trigger real notifications. The reward function (R_t) (Eq. 3.5) evolves to incorporate actual empirical data:

$$R_t = \Delta P + 0.5 \cdot I_{feedback} - 0.3 \cdot I_{fatigue} \quad (3.5)$$

Where ΔP represents the actual measured power savings following the intervention, $I_{feedback}$ is derived directly from the user's explicit response (e.g., an acceptance yields a positive reward, while a rejection yields a penalty) and $I_{fatigue}$ acts as a penalty term reflecting the user's current desensitization risk.

A significant challenge during this active phase is the asynchronous nature of human feedback. To address this, the Q-table update employs a delayed learning mechanism: an episode is stored in memory and only evaluated when the user responds. During this delayed update, the system applies a dynamic discount factor (γ) based heavily on the user's response:

- **Acceptance:** γ is set to its standard value (e.g., 0.95), allowing the agent to perform normal Q-learning with future value estimation.
- **Rejection:** γ is forced to 0.0. A rejection is treated as a terminal state, meaning the agent ignores future value estimations and learns exclusively from the immediate negative feedback, preventing rejected actions from falsely inflating in value.

Furthermore, if the elapsed time between the notification execution and the user's response exceeds a 2-hour threshold, the power savings component (ΔP) is forced to zero to prevent non-causal daily load shifts from corrupting the Q-table. For example, if an agent sends a nudge during an afternoon peak (e.g., 18:00) but the user only dismisses the notification late at night (e.g., 23:30), comparing the initial high power state against the natural nighttime standby state would falsely credit the agent with massive power savings.

3.5.3.6 Adaptive Cooldown and Fatigue Prevention

To mitigate the risk of user desensitization, the decision agent implements an adaptive cooldown mechanism that dynamically throttles the system's output based on context and user receptiveness.

Before generating an intervention, the system calculates an Opportunity Score ($O_t \in [0, 1]$) (Eq. 3.6) to quantify the immediate value of breaking the silence:

$$O_t = 0.35 \cdot S_{potential} + 0.35 \cdot U_{urgency} + 0.20 \cdot R_{receptiveness} + 0.10 \cdot C_{context} \quad (3.6)$$

This composite evaluates the real-time deviation of power above baseline ($S_{potential}$), the severity of spatial anomalies ($U_{urgency}$), the inverse of the user's fatigue index ($R_{receptiveness}$) and the alignment of temporal and occupancy states ($C_{context}$).

Parallel to this evaluation, the agent maintains an adaptive cooldown timer. Instead of a static lockout period, the base cooldown duration (30 minutes) is continuously modulated. A Fatigue Multiplier scales the cooldown proportionally to $I_{fatigue}$ (increasing the lockout by up to $3\times$), while a Time Multiplier shrinks the cooldown during critical energy peak hours and extends it during nighttime standby hours.

To balance user comfort with critical energy events, the system evaluates the Opportunity Score (O_t) (Eq. 3.6) against the adaptive cooldown constraints:

- **Normal Opportunity** ($O_t < 0.6$): The system strictly enforces the full adaptive cooldown period.
- **High Opportunity** ($O_t \geq 0.6$): The required cooldown is halved ($0.5\times$), allowing the system to intervene sooner for significant energy waste.
- **Critical Opportunity** ($O_t \geq 0.8$): The adaptive cooldown is completely bypassed to address severe anomalies. However, to prevent technical spam from continuous sensory noise, a strict absolute minimum cooldown of 10 minutes is universally enforced.

Finally, as an absolute fail-safe against algorithmic runaway, the agent enforces a hard limit of 10 notifications per day, after which all further proactive interventions are suppressed until the next daily cycle.

3.5.4 Adaptive Difficulty (Self-Efficacy)

Complementing the RL agent, a rule-based mechanism addresses the engagement gap by dynamically adapting the difficulty of the daily generated tasks. Rather than presenting static goals, the system calculates appropriate reduction targets using predefined difficulty multipliers ranging from Level 1 (Easy, a $0.5\times$ multiplier) to Level 5 (Hard, a $1.5\times$ multiplier).

This progression relies heavily on explicit user feedback gathered through the dashboard interface. At any given time, users can voluntarily evaluate their assigned tasks by classifying the experience as "Too Easy", "Just Right" or "Too Hard". Depending on this feedback and the user's actual success rate, the system adjusts future targets according to two primary scenarios:

- **Scenario A (High Performance):** If the user successfully completes a task and marks it as "Too Easy", the system escalates the difficulty level for future generations of that specific task type. The required power reduction margin is increased (e.g., shifting the multiplier towards $1.25\times$ or $1.5\times$), pushing the user to identify new efficiency gains rather than stagnating.
- **Scenario B (Low Performance / Frustration):** If the user struggles and explicitly marks a task as "Too Hard" (or if the system automatically assumes this due to consecutive failures without feedback), the difficulty level is immediately lowered. By relaxing the reduction target (e.g., applying the $0.5\times$ multiplier), the system rebuilds momentum and prevents user attrition caused by insurmountable goals.

3.6 User Interface Layer

The frontend leverages Lovelace, the native card-based dashboarding framework of HA. This system organizes information into customizable views composed of modular widgets, allowing for a highly flexible and responsive user experience. By utilizing this established framework, the system is immediately accessible via any web browser on desktop and through the native Companion App on mobile, seamlessly supporting both Apple and Android ecosystems. This approach eliminates the need for users to install a separate unknown application, reducing friction.

3.6.1 System Onboarding and Device Configuration

During the initial setup, the component prompts the user to define the deployment environment (Home or Office). If an office environment is selected, the user must input a standard work schedule (Fig. 3.2a), which provides the system with the temporal boundaries necessary to determine appropriate windows for AI interventions. Subsequently, the system executes an auto-discovery protocol to identify available sensor entities, attempting to classify them into six core categories: power, energy, temperature, humidity, illuminance and occupancy.

This automated classification relies primarily on standard HA metadata, such as device classes or measurement units. However, if this metadata is absent or ambiguous, the system falls back

on heuristic keyword matching. Because the vast heterogeneity of IoT devices makes automated classification prone to errors, a mandatory validation interface is introduced (Fig. 3.2b). This allows the user to manually verify and correct the categorizations. Finally, the user maps each validated sensor to a specific physical area (Fig. 3.2c), establishing the spatial resolution necessary for the AI’s granular pattern recognition.

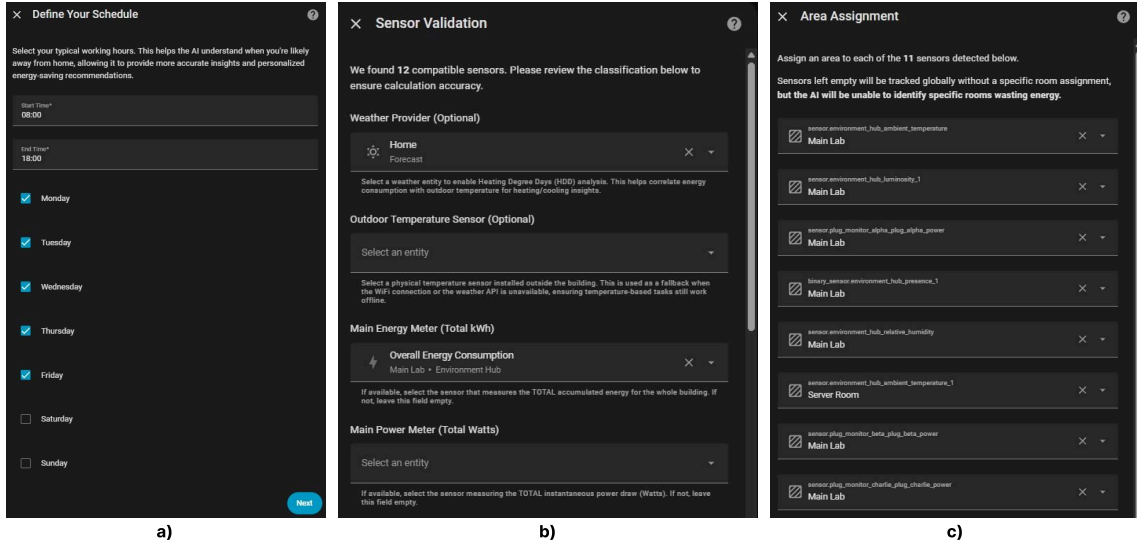


Figure 3.2: System onboarding interface: (a) Temporal boundary configuration via work schedules for office environments; (b) User validation of auto-discovered sensor categorizations; (c) Spatial assignment of hardware to specific physical areas.

3.6.2 Baseline Monitoring and Reporting

Once the hardware configuration is complete, the system transitions into its initial deployment phase, where the interface is strictly limited to passive observation. As shown in Fig. 3.3a, the user is presented with a status indicator explaining that, for an initial 14-day period, the system is exclusively monitoring energy consumption to establish an unbiased behavioural baseline.

During this 14-day phase, access to active intervention features is restricted. To keep users informed, the omitted tabs display a placeholder screen featuring a countdown (Fig. 3.4), indicating the number of days remaining until the active intervention phase begins.

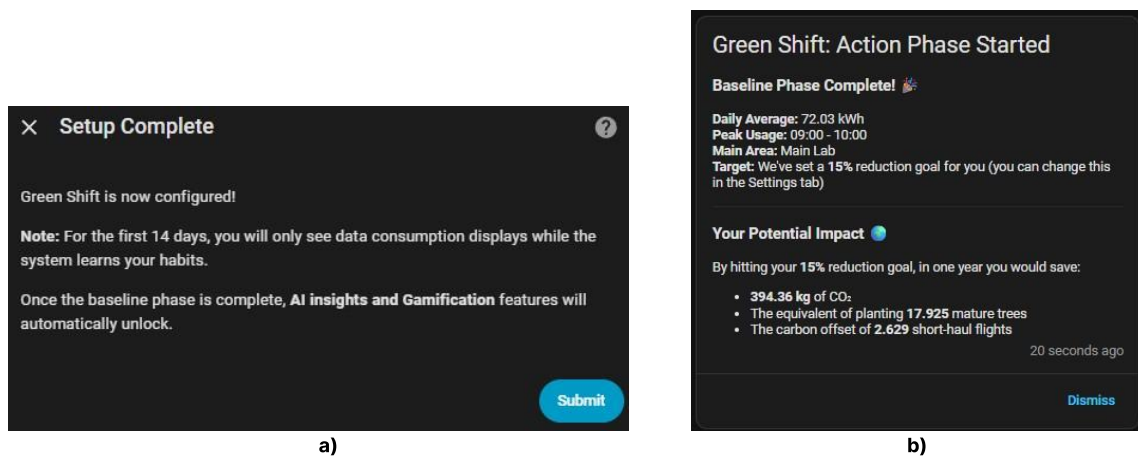


Figure 3.3: Baseline and reporting interface interfaces: (a) Status indicator explaining the initial 14-day monitoring phase; (b) The post-monitoring baseline report detailing historical consumption and targets.

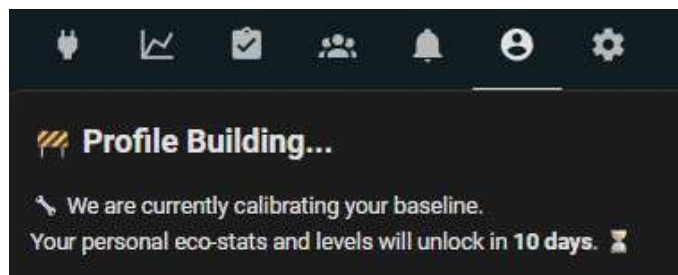


Figure 3.4: Blocked interface tab displayed during the initial baseline phase, featuring a countdown to active intervention.

Upon the conclusion of this period, the system generates a comprehensive baseline report (Fig. 3.3b). This report provides the user with an initial breakdown of their historical consumption and identifies the primary area of energy usage. Furthermore, it details the target reduction percentage and its projected yearly environmental impact, setting the stage for active AI intervention.

3.6.3 Consumption Visualization and Gamification

To promote energy awareness and encourage behavioural change, the component combines detailed consumption visualizations with interactive gamified elements. At a high level, the interface provides immediate context by displaying the total power consumption load alongside real-time estimates of hourly costs based on current usage. Additionally, it tracks cumulative daily metrics, including the total financial cost and daily CO₂ emissions. Beyond these aggregate metrics, the system ensures granular visibility through area-by-area device-level energy views (Fig. 3.5). This comprehensive breakdown allows users to pinpoint exactly how much energy individual appliances are consuming within specific physical spaces.

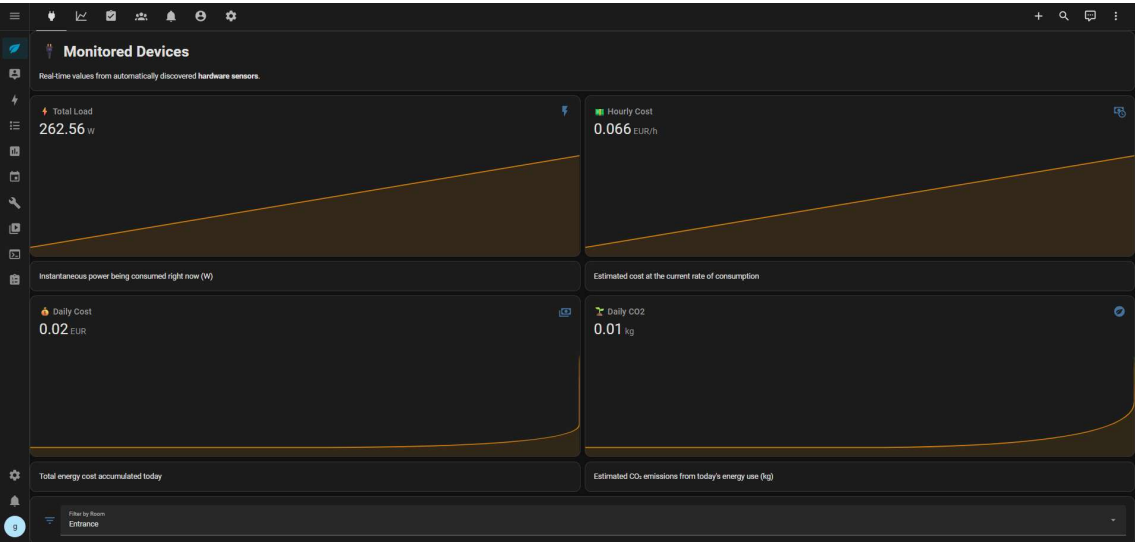


Figure 3.5: Detailed energy visualization interface, displaying aggregate metrics (total power load, cost and emission estimates) alongside granular area-by-area device consumption.

In the primary Dashboard tab (Fig. 3.6), users can track the evolution of their current power usage against their mean baseline consumption, fostering a better understanding of their daily habits. Additionally, this interface includes two crucial analytical sections: "Power Usage by Area", which illustrates the spatial distribution of energy consumption across the building and "Top 5 Power Consumers", which highlights the appliances with the most significant impact on overall usage.

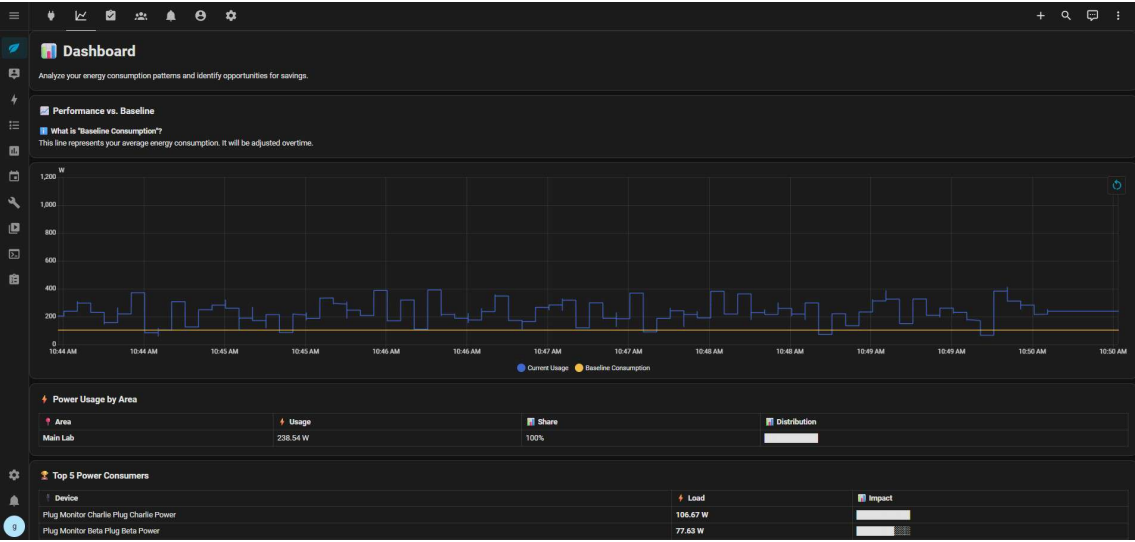


Figure 3.6: Main dashboard interface comparing current power usage against historical baselines, including area distributions and top consumers.

A fundamental component of the system’s behavioural framework is hosted within the Daily Tasks and Weekly Challenge tabs (Fig. 3.7, Fig. 3.8). The Daily Tasks interface (Fig. 3.7) serves as the primary touchpoint for the system’s adaptive difficulty mechanism. Rather than relying on a

static list, tasks are dynamically generated by a rule-based engine that evaluates sensor availability, current environmental context and historical consumption data. Furthermore, this screen captures direct user feedback for the gamification adaptation engine: users interact with buttons to rate tasks as "Too Easy", "Just Right" or "Too Hard".

To ensure accurate tracking, task progress is automatically validated every 15 minutes, however, users can override this interval by pressing a "Verify All Tasks" button for immediate evaluation. Following each validation cycle, the interface updates individual task statuses to provide continuous feedback. A task will clearly indicate whether it is successfully completed, waiting for a specific temporal window to commence (for schedule-based tasks) or failing/failed (e.g., if the current mean power consumption exceeds the established target).

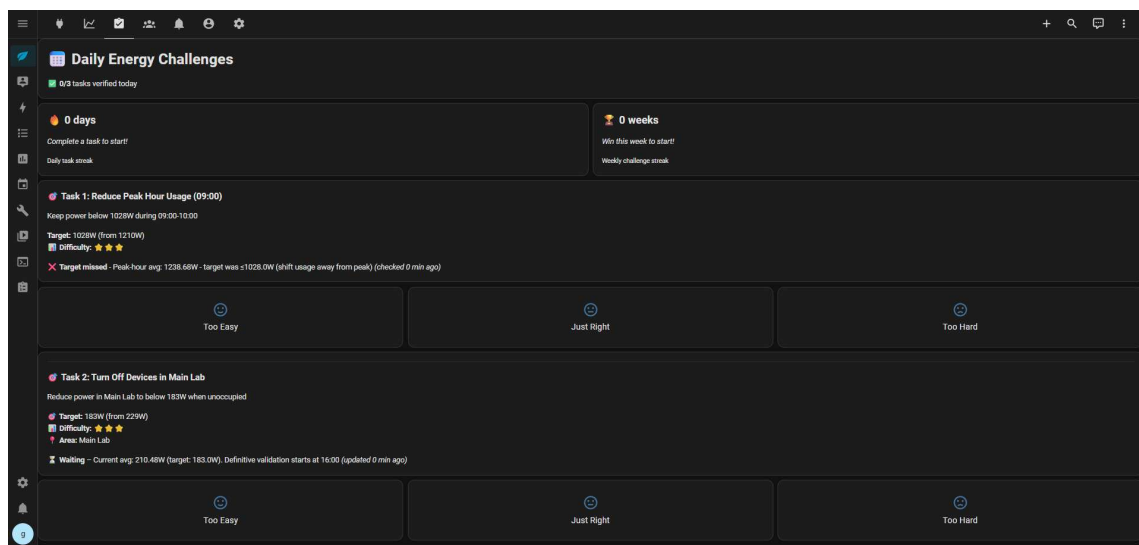


Figure 3.7: Daily Tasks interface featuring dynamic, AI-generated objectives, real-time status tracking and user feedback mechanisms for RL agent tuning.

The Weekly Challenge tab (Fig. 3.8) functions as a collaborative community goal. To successfully complete the challenge, all inhabitants of the building must work together to maintain their collective consumption below the established mean power threshold. A visual gauge provides immediate feedback: green indicates that consumption is comfortably below the target, orange signifies that it is approaching or slightly surpassing the limit and red warns that consumption has significantly exceeded the threshold.

To incentivize inhabitants to achieve this goal, an estimation of potential financial savings is prominently displayed alongside the gauge. Furthermore, to assist users in identifying actionable energy-saving opportunities, the interface features a spatial breakdown of current energy consumption per area. Particularly in larger buildings, this localized data enables users to quickly pinpoint specific regions where they can intervene to reduce overall usage and successfully meet the community challenge.

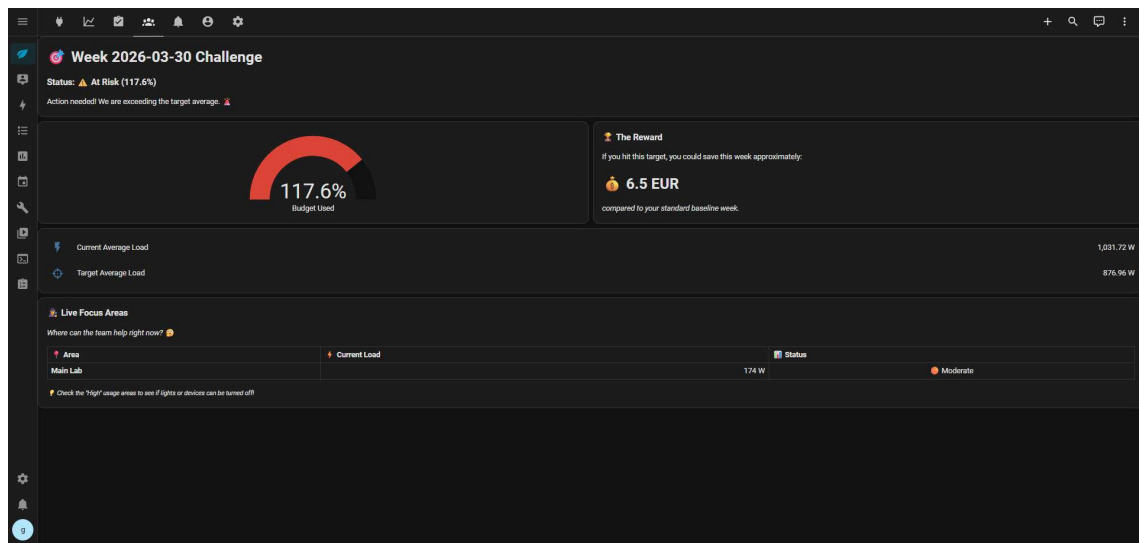


Figure 3.8: Weekly Challenge interface displaying the collaborative community goal, real-time consumption status via a colour-coded gauge, financial savings estimates and a spatial breakdown of energy usage to guide user interventions.

3.6.4 AI Notifications

To manage the targeted recommendations (nudges) sent by the AI agent, the system utilizes a dedicated notification tab (Fig. 3.9). A critical architectural decision was made regarding the collection of user feedback for these interventions. Due to technical constraints within HA's native push notification service, which lacks reliable support for custom stateful interactive buttons, the feedback loop was decoupled from the push notifications themselves. Instead, users review and evaluate the AI-generated nudges within this dedicated dashboard tab. This interaction provides the empirical response signal required to tune the RL agent's reward function ($I_{behaviour}$) and personalize future policies.

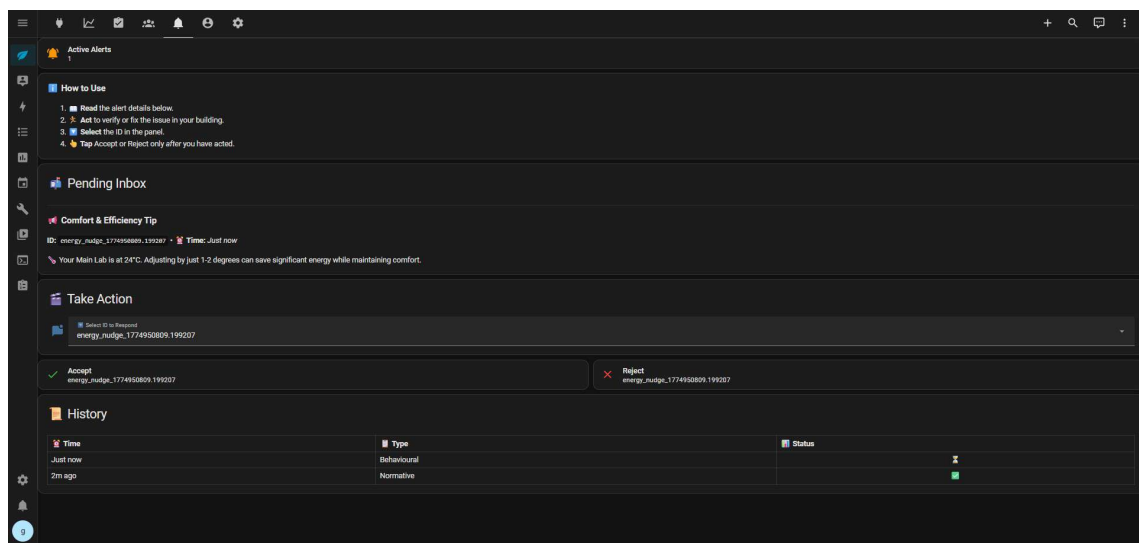


Figure 3.9: Dedicated in-app Notifications interface for reviewing and interacting with AI-generated nudges.

If the user has a mobile device connected to HA, the component will still dispatch a standard push notification to the device. Fig. 3.10 demonstrates an example of a notification received on a tablet.

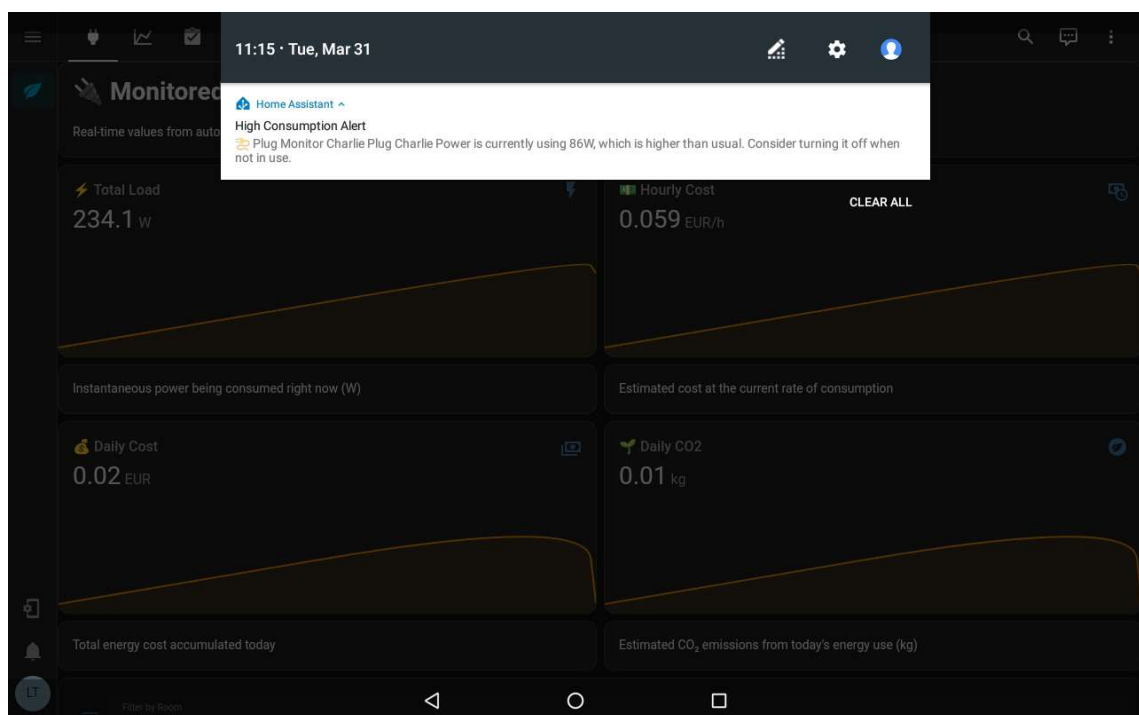


Figure 3.10: Example of a received mobile push notification, demonstrating the absence of interactive buttons due to native platform limitations.

3.6.5 Profile and System Configuration

To provide users with transparent insights into their progress and control over system parameters, the application includes dedicated modules for cumulative statistics and configuration.

The Profile View (Fig. 3.11) quantifies the user's overall impact. It displays their current progress level (calculated from tasks completed in the last 30 days), active streaks for daily tasks and weekly challenges, and gauges representing the AI's internal indices for user behaviour ($I_{behaviour}$) and fatigue ($I_{fatigue}$). Additionally, it highlights total financial savings and estimated CO₂ avoided. To make the environmental impact more tangible, the avoided emissions are translated into equivalent real-world metrics, such as the number of trees planted, short-haul flights offset and kilometres driven by car. To ensure the accuracy of these metrics, the system utilizes localized conversion factors.

While users can manually establish a specific energy price in the settings, the system defaults to 0.25 euros per kWh, reflecting the average household electricity price in Portugal as of February 2026, according to CountryEconomy⁴. Similarly, the carbon footprint is calculated using an estimated intensity of 0.1 kg of CO₂ per kWh, based on the Portuguese energy mix data from Nowtricity⁵ from the same period.

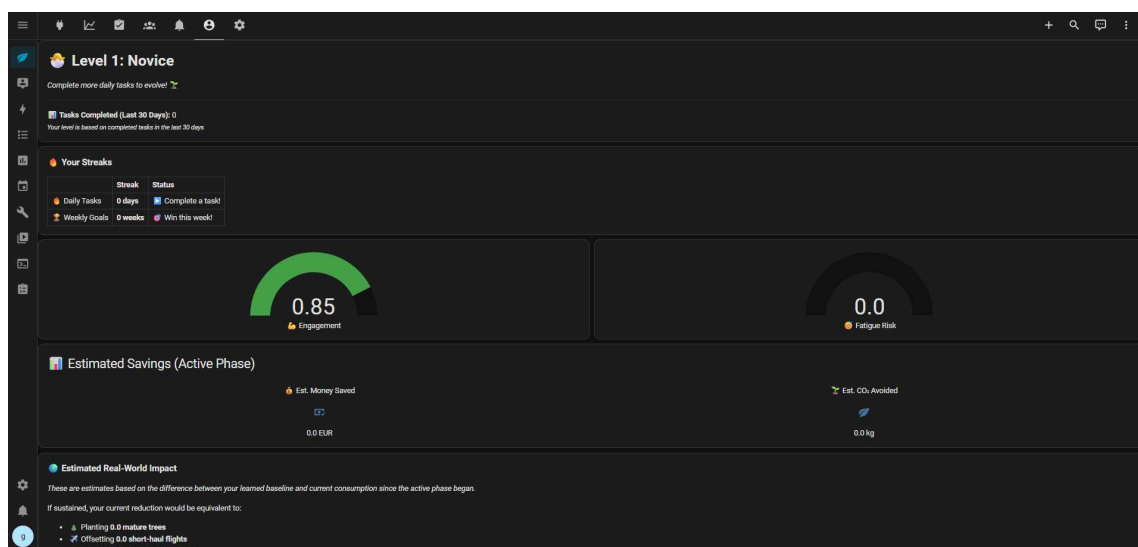


Figure 3.11: User profile dashboard displaying cumulative statistics, AI indices and tangible equivalents of environmental impact.

Finally, to align the application with individual preferences, the Settings Page (Fig. 3.12) provides several configuration options. Here, users can adjust their percentage reduction goal, toggle their preferred currency for financial visualizations and manually update the electricity price. Furthermore, a dedicated section displays the current sensor infrastructure established during the initial setup, allowing users to modify their device assignments as needed. Updating contextual

⁴<https://countryeconomy.com/energy-and-environment/electricity-price-household/portugal>

⁵<https://www.nowtricity.com/country/portugal/>

sensors, such as those for temperature and illuminance, results in a seamless transition without disrupting system operations. However, adding or modifying energy and power sensors will mandate a complete component reset. This strict measure forces the system to restart the initial 14-day monitoring phase from scratch, ensuring that the historical baseline is not biased by the introduction of new power measurements.

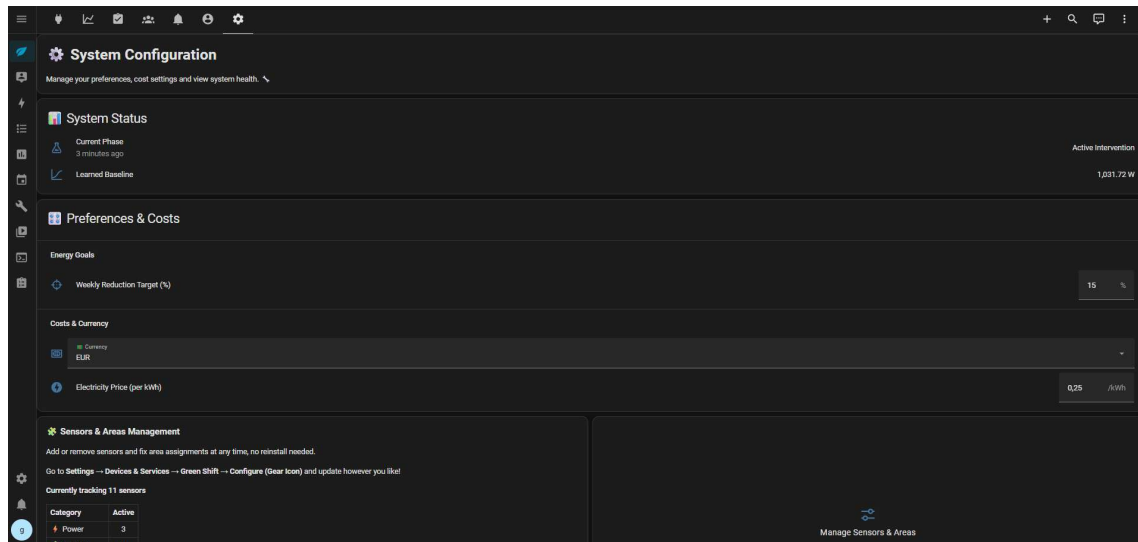


Figure 3.12: Settings interface for customizing reduction targets, financial preferences and managing the active sensor infrastructure.

3.7 Multilingual Support and Translation Mechanism

To ensure accessibility and user-centricity, the component incorporates a comprehensive translation framework. During the initial integration setup, users can explicitly define the system’s operational language, with full support for both English and Portuguese. This choice dictates the language of all frontend elements, including sensor names, dashboard titles and, most critically, the content of the proactive behavioural nudges and gamified tasks. By allowing this selection at the entry point, the system ensures that the behavioural intervention remains clear and effective regardless of the user’s primary language.

3.8 Data Reliability and Persistence Strategy

The architecture includes a multi-layered data protection suite designed to safeguard energy telemetry and the AI’s learned state against hardware failures or power interruptions. Reliability is maintained through three core technical pillars:

1. **Integrity Protections:** The system utilizes Write-Ahead Logging for SQLite databases to prevent data loss during crashes and employs atomic writes for the JSON state file to ensure that the AI’s Q-table and diagnostic indices are never partially written or corrupted.

2. **Lifecycle and Automated Backups:** Snapshots of the entire data environment (including the 14-day telemetry database, the 4-month research database and the AI state file) are managed across different triggers to ensure continuous protection. This includes:

- **Periodic Backups:** Generated every 6 hours, maintaining a rolling window of the last 10 snapshots (~2.5 days of protection).
- **Lifecycle Snapshots:** Mandatory backups are triggered at every integration startup and shutdown, retaining the last 2 versions of each to protect against state corruption during HA restarts.

3. **Manual Recovery Services:** For advanced maintenance, the component exposes dedicated HA services (e.g., *green_shift.create_backup* and *green_shift.restore_backup*), allowing users to manually archive or roll back the system to specific points in time.

All persistent data and backups are organized and separated into categorized subfolders (auto, startup, shutdown and manual) to prevent accidental data overwrites during recovery operations.

3.9 Quality Assurance and Software Testing

To validate the reliability of GreenShift, a rigorous testing suite was developed, comprising 1198 comprehensive tests. This approach achieved 100% total code coverage across all critical modules, ensuring that both the behavioural logic and the underlying infrastructure operate as intended. The testing distribution, shown in Table 3.1, focuses on the most sensitive architectural components:

Table 3.1: Summary of quality assurance and test coverage.

Component Module	Test Count	Functionality and Validation Focus
AI Decision Agent & Q-Learning	351 Tests (100%)	Validates the MDP framework, state discretization, epsilon-greedy action selection and the accuracy of Q-table updates based on reward signals.
Task Generation & Verification	143 Tests (100%)	Ensures that daily and weekly energy-saving tasks are correctly generated, tracked and verified against consumption data.
Database & Persistence	133 Tests (100%)	Covers all SQLite CRUD operations, ensuring data integrity for telemetry, research logs and persistent AI state files.
Sensor Entities	129 Tests (100%)	Verifies the correct behaviour, state updates and attribute mapping for all custom virtual sensors within the HA environment.
Data Collection & Energy Tracking	97 Tests (100%)	Validates real-time ingestion from physical sensors, power normalization and the calculation of consumption history.
Config Flow & Discovery	79 Tests (100%)	Tests the user setup interface, automated sensor discovery and the correct handling of environment-specific configurations (Home vs. Office).
Multilingual Support	68 Tests (100%)	Ensures that all UI elements, notification templates and task descriptions are correctly localized in both English and Portuguese.
Backup Management	46 Tests (100%)	Validates automated backup cycles (auto, startup, shutdown), manual restoration services and file system integrity.
Integration Logic & Helpers	152 Tests (100%)	Covers core setup/unload logic, service registrations and utility functions for time-of-day and holiday detection.
Total Coverage	1198 Tests	Overall System Coverage: 100%

This extensive validation ensures that the RL agent converges correctly, the backup system handles edge cases (such as power loss during writes) and the gamification engine accurately verifies user behaviour without false positives.

3.10 Design Decisions and Trade-offs

The architectural design prioritizes integration over isolation, aiming to create a system that is not only effective for the immediate research goals but also sustainable for long-term community adoption. This philosophy necessitated specific technical trade-offs, particularly regarding the choice to build within an existing ecosystem rather than alongside it. Table 3.2 summarizes these critical design decisions, contrasting the selected approaches against traditional alternatives.

Table 3.2: Design decisions and trade-offs.

System Architecture	Custom HA Integration (vs. Standalone Python App)	Rationale: Ensures native interoperability with thousands of devices and immediate access via existing mobile/desktop clients. Trade-off: Development is constrained by the host's specific Python environment and dependency management rules.
User Interface	Lovelace Framework (vs. Custom React/Web App)	Rationale: Eliminates installation friction for users; leverages a responsive and maintainable code base. Trade-off: Design flexibility is restricted to the specific card paradigm, preventing fully custom visual animations.
Data Ingestion & Normalization	Native HA Integrations (vs. MQTT-Exclusive Gateway or Direct Cloud API)	Rationale: Utilizes manufacturer-specific plugins to translate heterogeneous protocols (ZigBee/BLE/Wi-Fi) directly into standardized software entities, abstracting hardware differences while ensuring data remains on-premises. Trade-off: Increases system complexity by requiring the management of multiple integration dependencies within the local environment, rather than maintaining a single unified protocol broker (like MQTT).
Data Persistence	Local SQLite Databases (vs. TimescaleDB / InfluxDB)	Rationale: Utilizes the native HA storage format engine to ensure maximum system stability and zero-configuration deployment. Trade-off: Provides a simplified ACID-compliant local storage solution, but lacks the advanced time-series indexing and high-concurrency query performance of dedicated databases like InfluxDB for massive datasets.

Chapter 4

Validation

This chapter presents the empirical evaluation of GreenShift. Its primary objective is to validate whether this privacy-preserving, gamified AI system can effectively bridge the energy engagement gap. To ensure the transparency and replicability of the study, the chapter first outlines the experimental design, the deployment settings and the ethical considerations required for human-centric data collection. It then details the multi-phase experimental procedure, the specific evaluation metrics and the data analysis strategy. Finally, the chapter presents and discusses the empirical results from the real-world deployments, assessing both the system’s performance and the observed behavioural shifts.

4.1 Validation Methodology

The research employs a Quasi-Experimental Interrupted Time-Series Design [84] to examine whether the proposed system architecture contributes to improved energy efficiency. Due to the collaborative nature of the monitored spaces and privacy constraints that preclude individual-level monitoring, the unit of analysis is defined at the level of the entire intervention environment rather than the individual user.

The validation is conducted in two distinct settings to assess adaptability:

- **Professional:** Two laboratories (Fraunhofer Portugal AICOS and Faculty of Engineering University of Porto (FEUP)) characterized by shared zones and collective norms.
- **Residential:** Two smart homes where individual agency and personalization requirements are higher.

4.1.1 Experimental Topologies and Hardware Setup

To evaluate the architecture across varying infrastructural constraints and validate its hardware-agnostic design principles, the system was deployed using specific hardware configurations tailored to each environment.

4.1.1.1 Professional Settings

The core implementation utilized a distributed architecture to ensure low-latency edge processing. Fig. 4.1 illustrates the hardware topology deployed in the Fraunhofer laboratory, serving as the primary reference implementation. This specific environment was chosen to introduce the baseline architecture because it features a robust array of environmental sensors combined with facility-wide power metering. Additionally, detailing a laboratory environment at this stage, rather than a private home, protects the spatial privacy of the residential participants. In this setup, a Raspberry Pi 4 Model B (8GB RAM) acts as the local edge gateway. It communicates via BLE with ten custom Luxor sensors (integrating SHTC3 temperature/humidity and OPT 4001 luminosity modules). The system connects over Wi-Fi to an Everything Presence Lite sensor for accurate occupancy tracking and a Shelly Pro 3EM three-phase energy meter for facility-wide power measurement.

Furthermore, this deployment utilizes a Wi-Fi connection to fetch outdoor temperature data via HA's native *met.no* integration, acting as a practical substitute for a physical sensor to provide real-time context for gamified tasks and support future validation. The local HA instance for this node is hosted on an HP EliteOne 800 G5 All-In-One PC physically located within the laboratory.

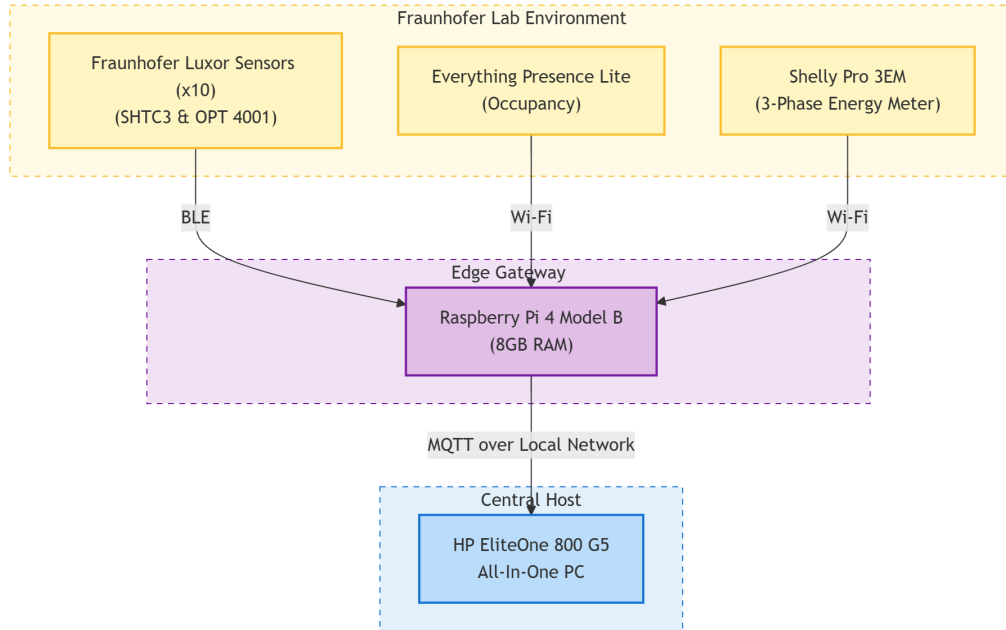


Figure 4.1: Hardware topology for comprehensive power monitoring and environmental contextual sensing in the Fraunhofer laboratory.

To validate the system's efficacy in highly restricted environments, the deployment was extended to the FEUP laboratory (Fig. 4.2). Here, the HA system is hosted locally on a Raspberry Pi 4 Model B (4GB RAM). Because access to the main energy grid is restricted, the setup relies on a localized hardware approach. Environmental context is captured using a Shelly BLU Motion Sensor alongside a Xiaomi Temperature and Humidity Sensor w/ Display 2. This is supplemented by

a physical Shelly BLU H&T for monitoring outdoor temperature, proving the system’s resilience by operating entirely independent of external cloud APIs. For energy tracking, three ZigBee smart plugs monitor specific loads (a TV, external monitors and computer charging sockets). Since the location lacked an accessible local area network, the Raspberry Pi was configured as a wireless hotspot to connect directly to the real-time feedback display (Lenovo Tab 4).

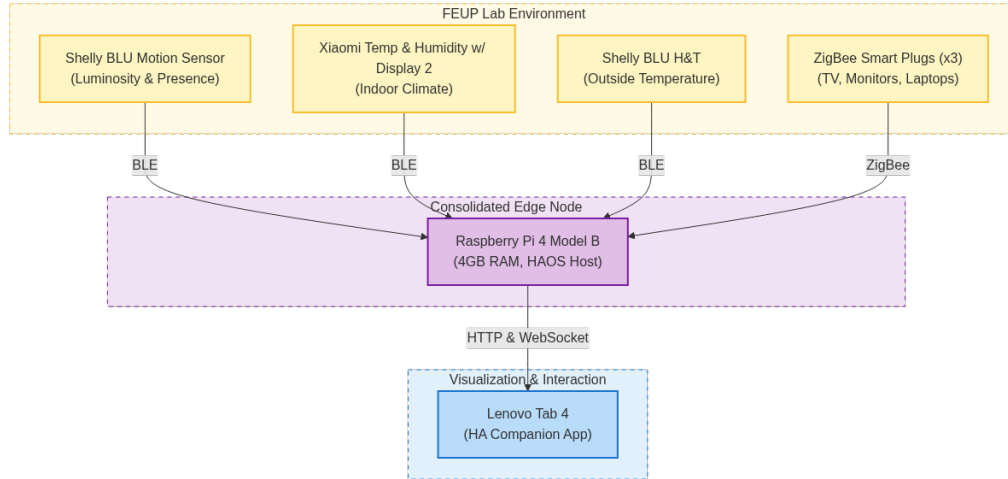


Figure 4.2: Distributed hardware setup for appliance-level monitoring and contextual sensing in the FEUP laboratory.

4.1.1.2 Residential Settings

Simultaneously, the system was deployed in two residential households, which functioned as the most comprehensive validation environments by fulfilling the complete architectural requirements. Rather than deploying novel standardized equipment, the study leveraged the participants’ pre-existing heterogeneous sensor networks.

To illustrate the infrastructural complexity, the households integrated varying scales of sensor networks, ranging from 10 to 19 devices per home. These operated concurrently across varied communication protocols like ZigBee, BLE and Wi-Fi. These ecosystems encompassed whole-house energy metering, distributed smart plugs for high-load appliances and multi-room environmental sensors capturing temperature, humidity and occupancy. Utilizing these established networks not only validates the behavioural intervention within private homes but also empirically demonstrates the architecture’s interoperability and resilience when confronting complex multi-vendor hardware setups.

4.1.2 Ethical Considerations

Given the collection of high-resolution behavioural data across both professional and residential settings, and the limited consideration of ethical implications in the current state of the art, this study is conducted in strict compliance with the GDPR.

To ensure ethical integrity, the research protocol adheres to several core principles. All data processing activities are conducted openly, with participants made fully aware of the system's operations to guarantee that no hidden tracking occurs. Processing is restricted solely to the validation of the energy management system, explicitly excluding incompatible uses such as employee performance monitoring. Furthermore, the system adheres to a strict minimization policy by collecting only the inputs essential for its core functioning, while avoiding extraneous personal details. Data accuracy is maintained through automated validation checks at the ingestion layer and raw telemetry is retained only for the duration of the active study before being purged or irreversibly anonymized. Finally, security and confidentiality are maintained via local edge processing and encryption to protect against unauthorized access or loss.

Additionally, the study guarantees specific rights to all participants. They are provided with a clear explanation regarding data collection methods and usage through direct engagement before the study begins. Participants retain the ability to request and view the data held about them at any time, as well as the option to correct any inaccuracies. Crucially, they may withdraw from the study and have their data permanently deleted from the system. Underlying these interactions is a specific, informed and unambiguous process to ensure that participation remains entirely voluntary.

4.1.3 Experimental Procedure

The experiment followed a three-phase timeline to ensure the AI model was grounded in empirical data and fully implemented prior to active intervention. As introduced earlier, Phase 1 and Phase 2 form the core of the deployment process, preceded by an initial offline foundational period:

- **Phase 0 (1 month) - Development & Data Acquisition:** Passive logging of power consumption and environmental metadata to create a synchronized dataset for the offline evaluation of the RL agent. Concurrently, the core integration logic of GreenShift was developed.
- **Phase 1 (14 days) - Silent Calibration:** The system was deployed in "Silent Mode". Users had access to a basic visualization dashboard (control period), while the RL agent refined consumption baselines without issuing interventions.
- **Phase 2 (Active Period) - Adaptation and RLHF:** Activation of the gamification engine and AI-driven nudges. The agent's policy was refined in real-time through RLHF, balancing energy savings against the user fatigue index ($I_{fatigue}$).

4.1.4 Data Collection and Metrics

Evaluating a behavioural energy intervention requires understanding the complex interactions between physical infrastructure, autonomous system logic and the psychological state of the occupants. Relying exclusively on aggregate power draw risks misinterpreting the system's true efficacy, creating evaluation blind spots that fail to capture the human cost of energy efficiency. To address this, the study employs a multi-modal data acquisition strategy.

4.1.4.1 Data Sources

To operationalize this strategy and prevent evaluation blind spots, data collection is deliberately triangulated across three synchronized sources:

1. **Physical Instrumentation:** The study employs a distributed IoT network consisting of smart plugs, smart energy meters and environmental sensors to capture raw physical telemetry from the environments.
2. **Software Telemetry:** The application leverages HA's internal event bus and two SQLite databases, a rolling (14-day operational window) and a long-term archive (retained for 120 days, spanning the full intervention duration), to automatically log user-interaction events, AI decision episodes, task outcomes, notification responses and daily aggregate statistics.
3. **Qualitative Instruments:** To capture explicit human factors, the study utilizes a standardized post-experiment questionnaire. This instrument measures long-term subjective variables such as habit formation, perceived notification fatigue and overall system trust.

By synthesizing these three distinct data streams, the methodology guarantees a holistic evaluation. The physical sensors establish the objective ground truth of energy reduction, the software logs provide a transparent audit trail of the AI's decision-making process and the qualitative surveys contextualize the human experience. Ultimately, this triangulation ensures that any observed energy savings are not achieved at the unacceptable cost of user fatigue or operational disruption, proving the intervention is both technically viable and psychologically sustainable.

4.1.4.2 Dependent Variables and Metrics

The impact of the intervention is measured through a combination of system telemetry and human factors evaluation. The quantitative metrics, derived from physical sensors and software logs, are detailed in Table 4.1.

Table 4.1: Definition of quantitative variables and metrics.

Construct	Variable	Measurement Metric
Energy Efficiency	Normalized Daily Consumption	Daily $\bar{P}$ (W), and daily E_{total} (kWh) weather-normalized via DD_{18} to yield E_{norm} (kWh/Degree Day).
Working-Hours Efficiency	Working-Hours Power Premium	$\Delta P_{WH} = \bar{P}_{working} - \bar{P}_{off}$ (W); quantifies active-use overhead and tracks reduction over time in professional settings.
Daily Task Efficacy	Daily Task Completion & Savings	$R_{task} = T_{completed}/T_{generated}$; tracked alongside mean instantaneous task savings ($\bar{\Delta P}_{task_savings}$) and relative impact (%) compared to $\bar{P}_{working}$.
Weekly Challenge	Challenge Success Rate	Binary outcome (achieved / failed) per week; $\Delta P_{savings}$ (W) and savings percentage vs. baseline.
Notification Effectiveness	Nudge Acceptance Rate	$R_{accept} = N_{accepted}/N_{rejected}$ disaggregated by action type.
Notification Fatigue	Fatigue Index ($I_{fatigue}$)	Continuous value $[0, 1]$ maintained by the agent; also measured via notification block rate $= N_{blocked}/(N_{sent} + N_{blocked})$.
RL Convergence	Q-value and Reward Trajectory	$\bar{Q}_{max}$ and $\bar{R}$ per day; ϵ decay curve; exploration-to-exploitation ratio.

To complement the system-level telemetry, qualitative human factors and psychometric variables were captured to assess the overall user experience and the psychological impact of the intervention. This data was collected through a comprehensive digital questionnaire, administered via Google Forms¹ at the conclusion of the active study period. The complete structure of the questionnaire, including the exact phrasing of the items, is provided in Appendix A. The primary constructs and their respective measurement metrics are defined in Table 4.2.

¹<https://docs.google.com/forms/>

Table 4.2: Definition of qualitative variables and metrics.

Construct	Variable	Measurement Metric
User Demographics & Context	Control Variables	Categorical (Age, Gender, Setting) and 5-point Likert scale smart home systems familiarity.
Usability and Acceptance	System Usability Scale (SUS)	Standardized 0-100 score derived from a 10-item 5-point Likert scale.
Intrinsic Motivation	IMI Subscale Scores	Standardized 7-point Likert-scale responses focusing on <i>Interest/Enjoyment</i> and <i>Value/Usefulness</i> dimensions.
AI Intrusiveness & Fatigue	Self-Reported Fatigue Index	5-point Likert scale evaluating notification timing/pressure, plus a categorical frequency metric for overall fatigue.
Behaviour Change & Habits	Self-Reported Habit Formation	5-point Likert scale evaluating off-screen energy consciousness and long-term habit continuation.
AI Explainability & Trust	Perceived System Trust	5-point Likert scale evaluating the perceived relevance, transparency and accuracy of AI recommendations.
Contextual Nuance	Open-Ended Feedback	Thematic analysis of optional free-text responses regarding desired changes and specific intervention examples.

4.1.5 Data Analysis Strategy

To provide a comprehensive evaluation of the proposed architecture, both system telemetry and survey-based psychometric data are processed and analysed using IBM SPSS Statistics² Version 30, supported by Python (Pandas³) for initial data formatting, cleaning and aggregation. The analysis is divided into quantitative system-level analytics and qualitative human factors evaluation.

4.1.5.1 System and Energy Analytics

This subsection details the quantitative methodologies applied to the raw physical telemetry and software logs. The primary objective is to isolate genuine energy savings, evaluate the convergence of the RL agent and quantify the direct physical impact of the gamified interventions.

²<https://www.ibm.com/products/spss-statistics>

³<https://pandas.pydata.org/>

Weather Normalization Because the Baseline (Phase 1) and Intervention (Phase 2) periods may occur under different climatic conditions, raw energy data is normalized to isolate behavioural impacts from building thermal loads. This is achieved using Degree Days (DD), specifically Heating Degree Days (HDD) or Cooling Degree Days (CDD) depending on the active thermal demand, calculated at a baseline temperature of 18°C. The normalized energy performance is calculated as shown in Eq. 4.1:

$$E_{norm} = \frac{E_{measured}}{DD} \quad (4.1)$$

Statistical Significance and Effect Size To evaluate the macro-impact of the proposed architecture, a Welch's t-test (chosen for its robustness against unequal variances) is applied to compare the mean E_{norm} between Phase 1 and Phase 2. For these statistical comparisons, the unit of analysis is defined as the daily aggregated consumption (i.e., individual days) rather than the individual user. For professional settings, the analysis is segmented between working and non-working hours to ensure structural unoccupancy is not misclassified as behavioural efficiency. Statistical significance is established at $p < 0.05$ and Cohen's d is calculated to quantify the practical magnitude (effect size) of the intervention.

Intra-Phase Gamification Efficacy To isolate the specific additive value of the gamification engine, an intra-phase analysis is conducted exclusively within Phase 2. Days are clustered into "Active" (where at least one task was completed) and "Passive" days (where tasks were ignored). By comparing the weather-normalized energy consumption (E_{norm}) between these two clusters, the study quantifies the overarching daily energy reduction attributable to the gamified mechanics. Furthermore, to evaluate the immediate physical impact of these interventions, the mean instantaneous task savings ($\overline{\Delta P}_{task_savings}$) are calculated exclusively from the tasks successfully completed during the active days.

RL Agent Convergence and Engagement Decay The algorithmic learning process is validated by analysing the RL agent's metrics over time. Time-series trend analysis is applied to the ϵ -decay rate, reward trajectories and Q-value stabilization to confirm the transition from exploration to exploitation. Concurrently, the efficacy of the system's self-throttling mechanism is validated by correlating the intervention's calculated opportunity score against the agent's internal fatigue index ($I_{fatigue}$), demonstrating how the system actively models user tolerance and suppresses notifications to prevent burnout.

4.1.5.2 Human Factors and Psychometrics

To complement the physical telemetry, this subsection outlines the statistical evaluation of the user experience. Standardized psychometric scales are utilized to measure the behavioural, motivational and cognitive responses to the gamified AI system, ensuring that energy savings are not achieved at the cost of user well-being.

Usability and Interface Acceptance The raw responses from the SUS are computed into a standardized 0-100 score. This composite metric provides a baseline for interface accessibility, ensuring that any observed lack of engagement is correctly attributed to behavioural resistance rather than poor UI/UX design.

Motivation and Gamification Perception Data from the Intrinsic Motivation Inventory (IMI) subscales (*Interest/Enjoyment* and *Value/Usefulness*) is analysed to differentiate between genuine habit formation and temporary compliance. High aggregate scores in these dimensions empirically validate whether the gamified mechanics fostered intrinsic motivation or were perceived as external coercion.

AI Explainability and Cognitive Load To evaluate user trust in the autonomous RL agent, the Likert-scale responses regarding perceived recommendation relevance and decision transparency are statistically aggregated. Crucially, self-reported metrics regarding annoyance and interruption frequency are cross-referenced with the system's calculated $I_{fatigue}$ to determine if the AI successfully and accurately modelled human tolerance thresholds.

Behaviour Change and Habit Formation The analysis of long-term behavioural shifts evaluates self-reported off-screen energy consciousness. By aggregating these scores, the study assesses the degree to which energy-saving behaviours translated into daily routines without the need for explicit system prompts.

Open-Ended Contextual Feedback To capture nuances that rigid psychometric scales cannot detect, thematic analysis is applied to the unstructured free-text responses. This qualitative process identifies recurring contexts of user interaction, such as specific frustrations with intervention timing or praise for particular nudge types, providing actionable design insights for future architectural iterations.

4.2 Results and Discussion

This section presents the empirical findings derived from the multi-phase validation procedure and critically examines them against the theoretical frameworks established in previous chapters. For readers seeking a concise overview of the core conclusions without navigating the detailed statistical and telemetry breakdowns, a comprehensive summary of the evaluation outcomes is provided at the end of this chapter (Subsection 4.2.7).

The discussion begins by characterizing the deployment environments and establishing baseline energy profiles. Following this contextualization, the section details the quantitative impact of the proposed architecture, evaluates the efficacy of the gamified mechanics and assesses the RL agent's convergence. By integrating system telemetry with qualitative human-centric outcomes,

this section directly addresses the RQs to evaluate whether a privacy-preserving gamified AI system can effectively bridge the energy engagement gap.

4.2.1 Sample Characterization and Experimental Setup

Data collection was successfully executed across the four previously defined environments, achieving a high integrity rate throughout the study period. Table 4.3 details the active participants, the number of monitored physical hardware entities (e.g., a smart plug’s raw energy consumption), and the days of telemetry captured in each location.

Table 4.3: Summary of experimental environments and data collection.

Environment	Type	Active Participants	HA Entities	Capture Days
Fraunhofer AICOS	Professional	2	33	62
FEUP	Professional	5	10	54
Home 1	Residential	4	32	52
Home 2	Residential	2	27	22

Across the four experimental environments, a total of 13 individuals actively participated in the study. While demographic data was collected for the entire sample, only 8 participants completed the final evaluation survey. The overall participant pool consisted of 8 males and 5 females. To ensure participant privacy, age was collected categorically via ranges rather than absolute numerical values. The reported age distribution was: 18-24 years (5 participants), 25-34 years (1 participant), 35-44 years (2 participants) and 45-54 years (2 participants), with the remaining 3 participants opting not to disclose their age. Given the contextual differences, the residential environments captured a broader demographic variance characteristic of family living arrangements, whereas the professional settings primarily involved researchers, university students and technical staff.

To contextualize the user interactions, technological literacy and familiarity with Smart Home environments (such as HA) were assessed. Fig. 4.3 illustrates the distribution of user familiarity across the specific settings for the participants who answered the survey. The overall sample indicates a predominantly high level of technological literacy, while still capturing valuable demographic variance across the deployment environments. The highest rating of 5 is the most frequent, representing over 60% of the responses and encompassing users from the FEUP laboratory, the Fraunhofer laboratory and the residential environments. Responses rating a 4 (25%) originated exclusively from the FEUP laboratory. Conversely, unlike the professional settings, the residential environments also captured responses from users with minimal prior exposure, specifically recording ratings of 1 (approximately 12.5%). Despite this variance in the residential cohort, the generally high baseline familiarity across the majority of the sample suggests that most usability challenges or engagement fluctuations observed during the study are unlikely to be caused by a lack of basic technological proficiency, ensuring that the evaluation successfully isolated the behavioural impact of the gamified AI interventions.

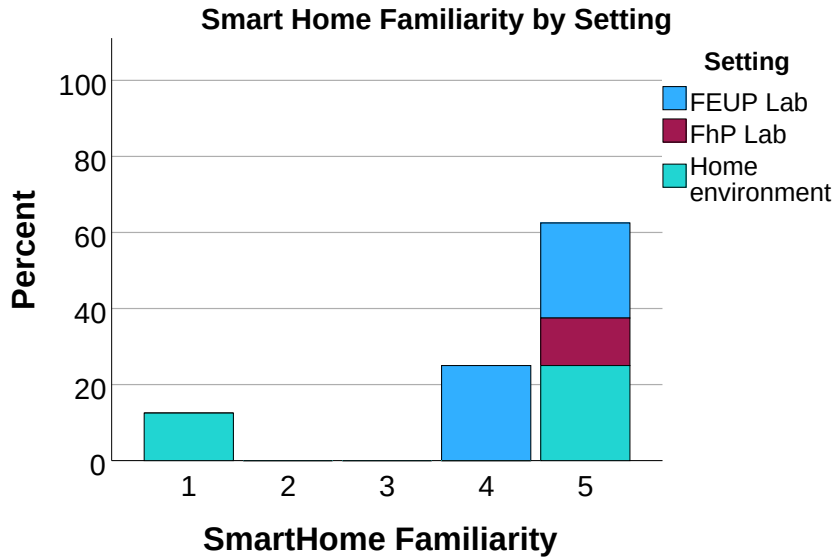


Figure 4.3: Stacked bar chart showing the distribution of self-reported smart home familiarity (on a scale of 1 to 5) by deployment setting.

4.2.2 Baseline Profiling and Shadow Learning (Phase 1)

During the 14-day silent calibration phase, the system monitored baseline energy profiles without issuing interventions. Table 4.4 presents the mean power consumption, highest peak power registered and the disparity between working and non-working hours for professional settings. As a reminder, in residential settings, average power is equivalent to working-hours power, since these environments are monitored continuously and no filtering is applied for non-working hours.

Table 4.4: Phase 1 baseline energy profiling.

Environment	Power (W)			
	Mean	Peak	Working Hrs ($P_{working}$)	Non-Working Hrs
Fraunhofer	5481.15	13301.11	6968.92	4851.32
FEUP	10.65	173.0	20.36	8.58
Home 1	1295.18	8722.67	1295.18	N/A
Home 2	321.83	6278.26	321.83	N/A

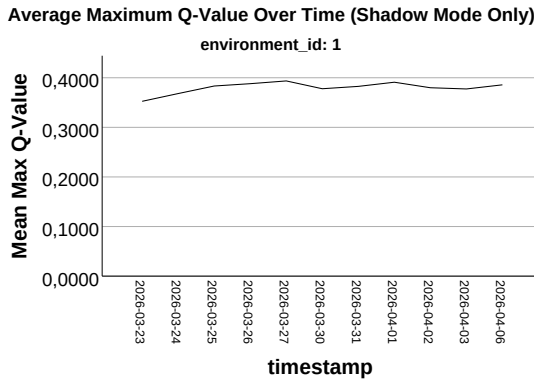
Simultaneously, the AI agent operated in Shadow Mode, effectively detecting anomalies and refining its policy through passive rewards. Fig. 4.4 demonstrates the evolution of the maximum Q-value over the baseline period across all environments.

The empirical telemetry reveals distinct learning trajectories heavily influenced by the sensor topology. In the Fraunhofer laboratory (Fig. 4.4a), the maximum Q-value exhibited a relatively smooth exploratory phase, peaking near 0.40 before stabilizing at approximately 0.38. Conversely, the FEUP laboratory (Fig. 4.4b) displayed significant and persistent volatility, oscillating between

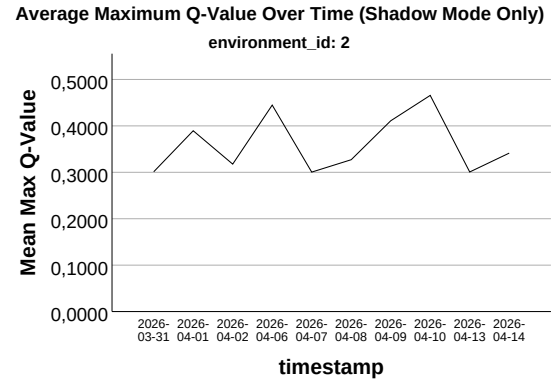
0.30 and 0.47. This instability is theoretically consistent with its localized hardware setup: monitoring individual volatile loads (e.g., smart plugs on monitors) creates a more chaotic state space than aggregate facility monitoring.

In the residential settings, the agent encountered more predictable state spaces. Home 1 (Fig. 4.4c) maintained stable Q-values around 0.35–0.40 for most of the period before experiencing a dip to 0.30 in the final days. Similarly, Home 2 (Fig. 4.4d) exhibited a highly stable learning trajectory, fluctuating minimally between 0.35 and 0.41.

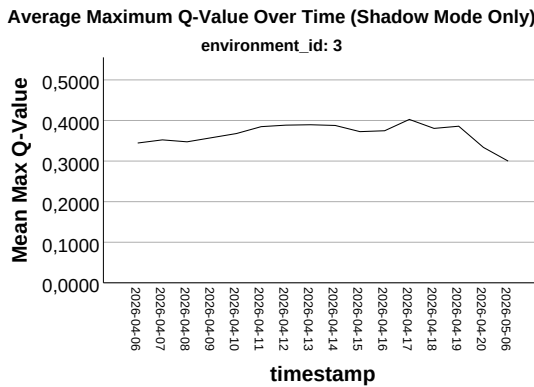
Ultimately, the smoother learning curves observed across the residential homes and the Fraunhofer laboratory share a common architectural denominator: the presence of whole-house aggregate metering. By measuring the total macro-level facility load, these setups naturally smooth out the sudden micro-fluctuations of individual devices. This stability allows the RL agent to calibrate with significantly less noise compared to the highly volatile isolated appliance-only monitoring deployed in the FEUP setup.



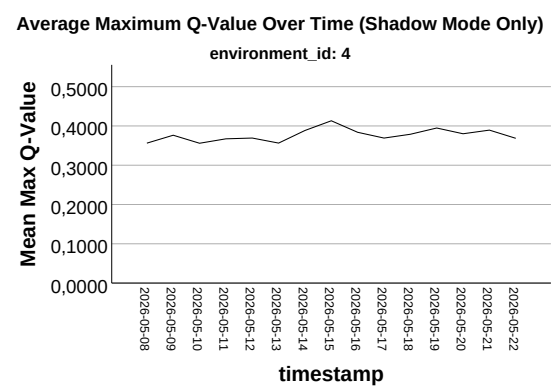
(a) Fraunhofer Laboratory



(b) FEUP Laboratory



(c) Residential Home 1



(d) Residential Home 2

Figure 4.4: Evolution of max Q-value and detected anomalies during shadow learning across professional and residential environments.

4.2.3 Quantitative Impact on Energy Consumption

The transition from Phase 1 (silent calibration and visualization) to Phase 2 (active gamification) yielded varying impacts on absolute energy usage depending on the environmental topology. To understand the trajectory of these behavioural shifts, and specifically to evaluate the standalone impact of the visualization dashboard, Fig. 4.5 presents the interrupted time-series evolution of weather-normalized consumption (E_{norm}) across the study duration.

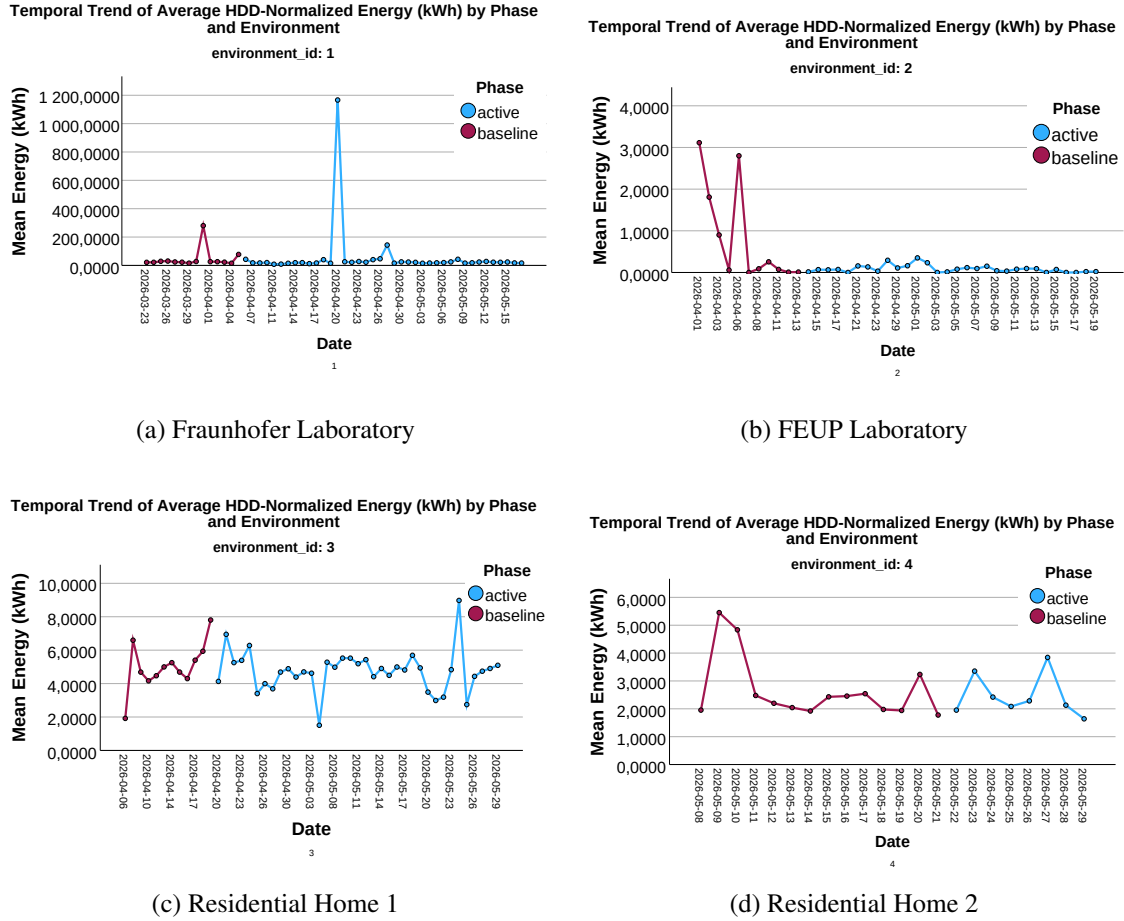


Figure 4.5: Interrupted time-series analysis detailing the daily evolution of E_{norm} , illustrating the transition from passive visualization (Baseline) to active intervention (Active Phase).

This temporal visualization provides critical nuance regarding the immediate human response to energy data exposure. In the Fraunhofer laboratory (Fig. 4.5a), the normalized energy consumption remained consistently low and structurally stable across both phases, with the exception of some occasional isolated operational peaks. Because the baseline values maintained this flat profile, drawing definitive conclusions regarding an overarching behavioural trend remains challenging in this specific facility.

In contrast, the baseline periods in user-controlled settings highlight that passive visualization does not produce constant or uniform results. The FEUP time-series (Fig. 4.5b) shows significant initial consumption peaks followed by a massive sharp decrease. This trajectory suggests

an immediate psychological reaction: once occupants were confronted with the real-time visualization dashboard, the sudden awareness of their high energy footprint prompted a deliberate and substantial reduction in usage.

A similar pattern was observed in Home 2 (Fig. 4.5d), which closely mirrored the FEUP environment's behaviour with a pronounced consumption spike in the early days of the baseline phase followed by a downward stabilization trend. Conversely, Home 1 (Fig. 4.5c) demonstrated the opposite baseline effect, where the normalized energy consumption ultimately experienced a progressive increase over the passive observation period.

Addressing RQ1 (formulated in Section 1.3), these mixed trajectories demonstrate that traditional eco-feedback alone yields highly volatile and context-dependent outcomes. While initial visual feedback can trigger a strong corrective response in spaces where occupants have direct control and high visibility of their actions (such as the FEUP laboratory and Home 2), it can also fail to prevent consumption increases over time (as seen in Home 1) or remain structurally constrained in low-agency communal environments (as seen in the Fraunhofer laboratory), underscoring the necessity of an active gamification layer to stabilize behaviours.

While the time-series illustrates the temporal trajectory, it is necessary to evaluate the macroscopic statistical efficiency between the two distinct phases, comparing the baseline visualization period against the active gamification and personalized AI period. To quantify this phase-to-phase impact, independent samples t-tests were conducted on the aggregated E_{norm} and average power metrics. Fig. 4.6 compares these shifts across the monitored environments.

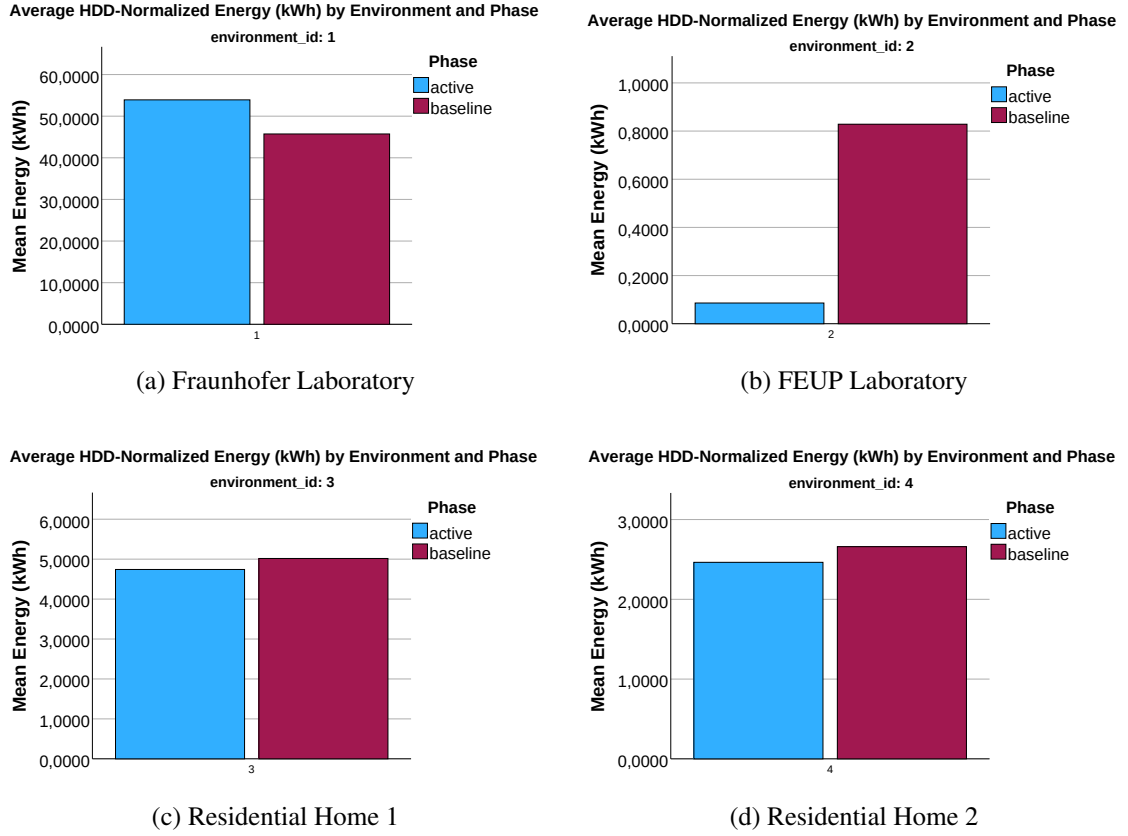


Figure 4.6: Comparison of aggregate weather-normalized energy consumption (E_{norm}) between baseline and active phases across professional and residential environments.

In the Fraunhofer laboratory, the mean E_{norm} shifted from 45.73 kWh/DD in the baseline phase to 53.92 kWh/DD (+17.9%) during the active phase (Fig. 4.6a). An independent t-test (equal variances assumed, Levene's $p = 0.639$) indicated this variance was not statistically significant ($t(52) = -0.164$, $p = 0.871$), with a negligible effect size (Cohen's $d = -0.051$). The average raw power consumption also remained remarkably stable, shifting only from 5481.16 W to 5464.54 W (-0.3%) ($p = 0.938$). This statistically validates that in a shared industrial laboratory, individual nudges were insufficient to systematically overcome the dominant inelastic structural loads of the facility.

Conversely, in the FEUP laboratory, the active gamification phase successfully locked in and deepened the initial efficiency gains. The overall baseline mean E_{norm} of 0.83 kWh/DD dropped drastically to an active phase mean of 0.09 kWh/DD (-89.2%) (Fig. 4.6b). Because Levene's test indicated unequal variances ($F = 63.845$, $p < 0.001$), a Welch's t-test was utilized. This reduction approached statistical significance ($t(10.04) = 2.068$, $p = 0.065$) and demonstrated a very large practical effect size (Cohen's $d = 1.223$), signifying a strong intervention impact. The average power draw correspondingly dropped from 10.65 W to 3.52 W (-66.9%). However, an independent t-test (equal variances not assumed, Levene's $p = 0.004$) indicated this raw power variance was not statistically significant ($t(13.25) = 1.117$, $p = 0.284$).

The residential environments provided a more nuanced perspective on the relationship between raw power and weather-normalized behaviour. In Home 1, the raw average power dropped significantly from 1299.69 W to 1103.81 W (-15.1%), yielding a statistically significant reduction ($t(50) = 2.140$, $p = 0.037$) with a moderate effect size (Cohen's $d = 0.669$). However, when normalized for weather (E_{norm}) (Fig. 4.6c), the shift from 5.02 kWh/DD to 4.74 kWh/DD (-5.5%) was not statistically significant ($t(47) = 0.645$, $p = 0.522$). This suggests that while occupants actively reduced their raw electrical load in response to the gamified AI, fluctuations in thermal demand partially masked the magnitude of these behavioural changes in the normalized data.

In Home 2, the metric utility was inverted. The raw power consumption actually increased slightly from 329.10 W to 367.83 W (+11.8%) during the active phase ($p = 0.621$). Nevertheless, the weather-normalized energy (E_{norm}) decreased from 2.66 kWh/DD to 2.46 kWh/DD (-7.4%) (Fig. 4.6d). Although this normalized reduction was not statistically significant ($t(20) = 0.439$, $p = 0.665$, Cohen's $d = 0.195$), it critically demonstrates the importance of weather normalization: without it, the raw data would have falsely implied a deterioration in energy habits, whereas the normalized metric proves that the household actually improved its operational efficiency relative to the active thermal constraints.

Finally, to determine if structural consumption patterns shifted during the intervention in the professional environments, the Working-Hours Efficiency metric (ΔP_{WH}) was analysed (Fig. 4.7).

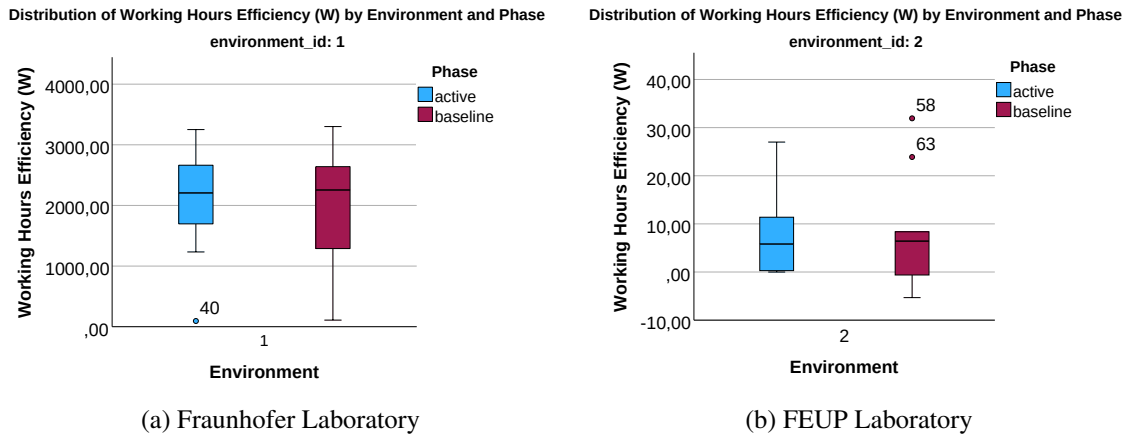


Figure 4.7: Boxplots detailing the variance in working-hours power premium (ΔP_{WH}) between the active and baseline phases across professional settings.

In the Fraunhofer setup (Fig. 4.7a), the power premium during working hours shifted slightly from 2127.04 W to 2025.39 W (-4.8%). An independent t-test (equal variances assumed, Levene's $p = 0.135$) confirmed this variance was not statistically significant ($t(38) = -0.371$, $p = 0.713$), presenting a minimal effect size (Cohen's $d = -0.135$). Similarly, the FEUP laboratory (Fig. 4.7b) experienced a marginal shift from 7.73 W to 7.13 W (-7.8%), which also failed to reach statistical significance ($t(35) = 0.177$, $p = 0.860$, Cohen's $d = 0.068$).

These results indicate that the proportional variance between working and non-working hours

did not structurally change over the observation period. This statistically validates that the underlying operational routines of the laboratories remained intact despite the active behavioural interventions, ensuring that any observed energy reductions were the result of targeted efficiency improvements rather than generalized operational shutdowns or decreases in overall laboratory activity.

4.2.4 Gamification Efficacy and User Engagement

User engagement was measured via the completion rates of daily tasks (R_{task}) and weekly challenges. To prevent user abandonment, the gamification engine dynamically adjusted the task difficulty based on ongoing user responsiveness. Fig. 4.8 illustrates this adaptive relationship across the monitored environments, correlating the daily completed tasks (bars) with the system's calculated difficulty level (line) over the intervention period.

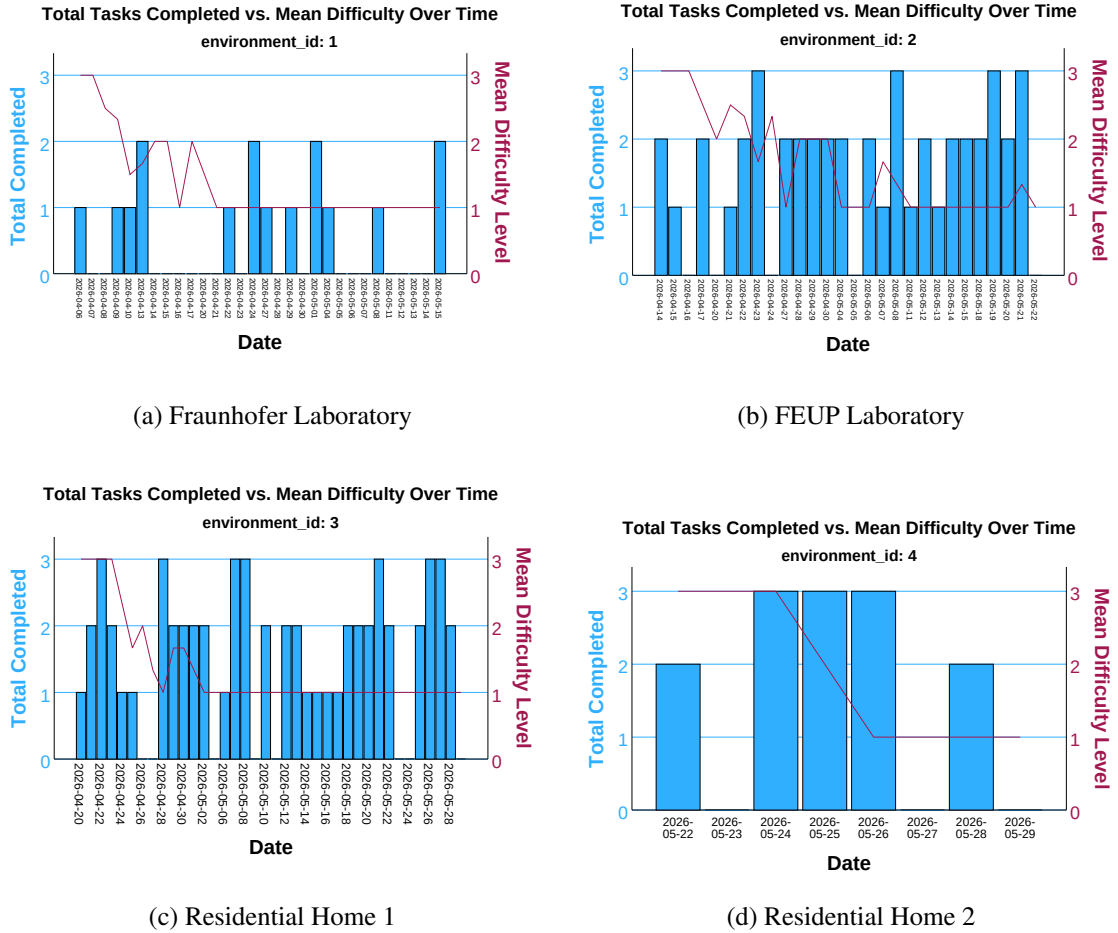


Figure 4.8: Dual-axis charts correlating daily task completion rates (left axis) with the adaptive difficulty level (right axis) across professional and residential environments.

In the Fraunhofer laboratory (Fig. 4.8a), the system initiated at a default difficulty level of 3 (on a scale up to 5). Following an initial period of non-engagement and explicit user feedback

indicating that the targets were too challenging, the adaptive engine rapidly scaled the difficulty down, plateauing at level 1 from mid-April onwards. While this dynamic calibration successfully prevented complete user abandonment, the overall task completion rate remained notably low at 23.5%, with the majority of tasks going uncompleted. There was not a single day where the maximum of three tasks was achieved. This limited engagement is theoretically consistent with the environment's professional demands, as occupants were likely highly focused on their primary work duties, treating secondary gamified energy interventions as a much lower priority in a shared space.

The FEUP laboratory (Fig. 4.8b) exhibited a similar trend regarding adaptive difficulty, but yielded a markedly different engagement outcome. Beginning at the default difficulty level of 3, the system systematically reduced the threshold in response to user feedback and interaction patterns. Unlike the Fraunhofer setting, as the difficulty stabilized around level 1 in the localized FEUP setup, it catalysed a much more consistent completion rate of 71.4%, allowing users to frequently achieve the maximum of 3 completed tasks per day.

In Home 1 (Fig. 4.8c), the difficulty level similarly stepped down from 3 to 1 over the first ten days. Once stabilized at this accessible level, the users maintained an extremely dense completion record, successfully finishing 1 to 3 tasks almost every single day of the active phase, achieving a completion rate of 51.8%. Home 2 (Fig. 4.8d) showed a brief compressed version of this curve, rapidly dropping the difficulty to 1, which was met with several days of 2 or 3 completed tasks and an overall completion rate of 59.3%. The robust performance in these residential settings indicates that gamified tasks are highly effective when integrated into private daily routines where users have total autonomy over their environment.

To isolate the genuine savings generated by the gamification engine, an intra-phase analysis was conducted exclusively within Phase 2. Days were clustered into "Active" (where at least one task was completed) and "Passive" days (where tasks were uncompleted or ignored). As depicted in Fig. 4.9, active days consistently registered lower mean weather-normalized energy consumption (E_{norm}) across all environments.

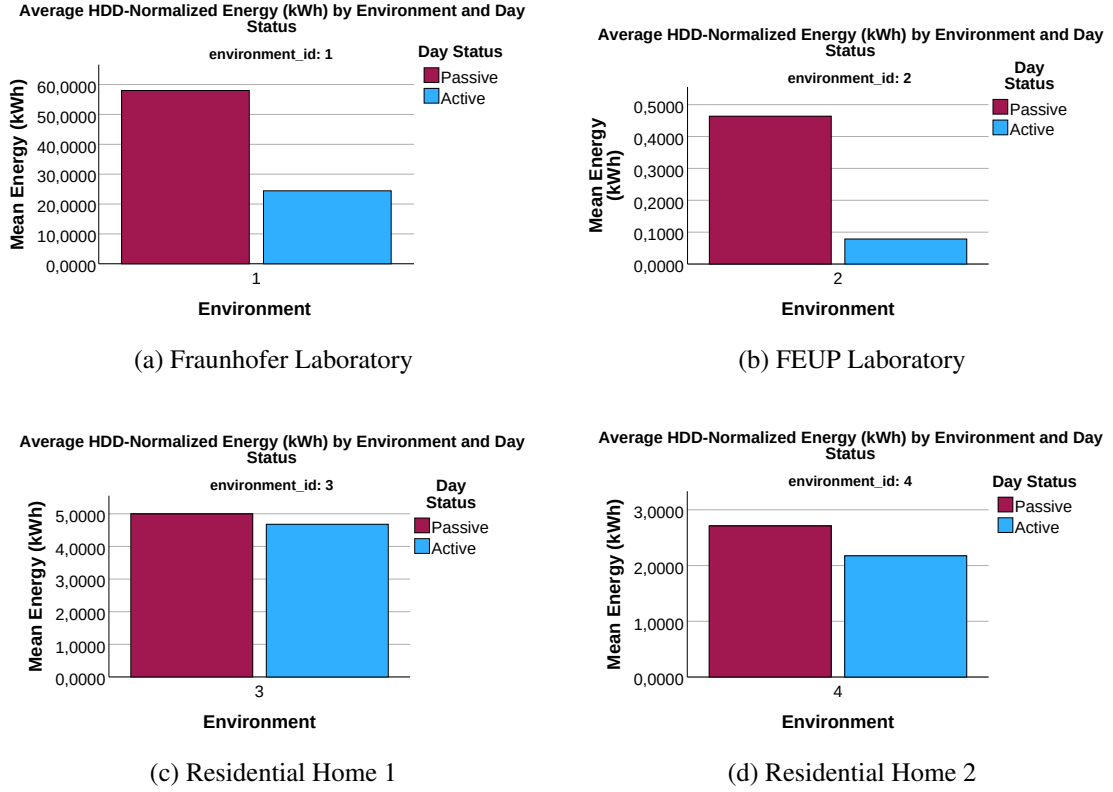


Figure 4.9: Intra-phase comparison of mean weather-normalized energy consumption (E_{norm}) clustered by active vs. passive engagement days.

In the Fraunhofer laboratory (Fig. 4.9a), active days averaged roughly 24.5 kWh/DD, representing a stark reduction (-57.8%) compared to passive days which exceeded 58 kWh/DD. Similarly, the FEUP laboratory (Fig. 4.9b) demonstrated active day averages of approximately 0.08 kWh/DD, a sharp contrast (-82.6%) to the near 0.46 kWh/DD observed on passive days. The residential environments also followed this pattern, albeit with varying magnitudes. Home 1 (Fig. 4.9c) showed a modest reduction, dropping from approximately 5.0 kWh/DD on passive days to 4.7 kWh/DD on active days (-6.0%). Home 2 (Fig. 4.9d) exhibited a more pronounced shift, dropping from nearly 2.7 kWh/DD to roughly 2.15 kWh/DD (-20.4%). Addressing RQ2 (defined in Section 1.3), this intra-phase analysis delivers empirical proof of gamification's additive value. By coupling real-time data with actionable game elements, the system created a verifiable wedge against energy waste, consistently outperforming days where users ignored or didn't complete the gamified interventions.

Furthermore, to quantify the immediate physical impact of these interventions, the mean instantaneous task savings ($\overline{\Delta P}_{task_savings}$) were calculated for all successfully completed tasks. As detailed in Table 4.5, the gamification engine drove substantial immediate power reductions, but their true impact is best understood when contextualized against the active working-hours (or baseline average) load of each environment ($\overline{P}_{working}$).

Table 4.5: Average instantaneous task savings relative to the active working-hours baseload.

Environment	Baseload ($\overline{P}_{working}$)	Mean Savings ($\overline{\Delta P}_{task_savings}$)	Impact
Fraunhofer	6968.92 W	590.00 W	8.5%
FEUP	20.36 W	10.80 W	53.0%
Home 1	1295.18 W	263.55 W	20.35%
Home 2	321.83 W	224.76 W	69.84%

While the Fraunhofer setup achieved the largest absolute drop in wattage per task (590.00 W), this represented only an 8.5% reduction relative to its massive infrastructural baseline. In stark contrast, the nominally smaller mean savings in the FEUP setup (10.80 W) effectively eliminated 53.0% of its localized working-hours load. This proportional disparity is mirrored in the residential setups: Home 2 achieved an exceptional 69.84% impact by reducing 224.76 W from a lighter 321.83 W baseline, whereas Home 1 achieved a 20.35% impact against a much heavier 1295.18 W baseline. These figures confirm that when users actively engaged with the system prompts, the gamified tasks effectively translated into quantifiable immediate drops in power demand, with the relative impact being drastically higher in individual-appliance and lighter-load setups where users exercise direct unmitigated control over the primary consumers.

Complementing these daily mechanics, the weekly challenges served as a collaborative normative gamification strategy designed to encourage broader sustained energy reduction milestones. The outcomes of these challenges are summarized in Table 4.6. By default, the system establishes a 15.0% target reduction relative to the baseline. While users have the configuration autonomy to manually adjust this percentage to create more achievable goals, the target remains structurally static, as it does not automatically adapt to consecutive successes or failures in the same manner as the daily tasks.

Table 4.6: Weekly challenge outcomes, target reductions and achieved savings per environment.

Week	Environment	Target (%)	Savings (W)	Savings (%)	Status
1	Fraunhofer	15.0%	392.89 W	5.63%	Failed
2	Fraunhofer	15.0%	-89.95 W	-1.29%	Failed
3	Fraunhofer	15.0%	-427.44 W	-6.12%	Failed
4	Fraunhofer	15.0%	265.48 W	3.80%	Failed
5	Fraunhofer	15.0%	6.16 W	0.09%	Failed
1	FEUP	15.0%	5.99 W	40.90%	Success
2	FEUP	15.0%	0.03 W	0.19%	Failed
3	FEUP	15.0%	6.63 W	45.24%	Success
4	FEUP	15.0%	5.17 W	35.28%	Success
5	FEUP	15.0%	7.76 W	53.01%	Success
1	Home 1	15.0%	149.54 W	11.55%	Failed
2	Home 1	15.0%	195.10 W	15.06%	Success
3	Home 1	15.0%	87.06 W	6.72%	Failed
4	Home 1	15.0%	143.29 W	11.06%	Failed
5	Home 1	15.0%	257.80 W	19.90%	Success
1	Home 2	15.0%	5.46 W	1.70%	Failed

In the Fraunhofer environment, the cohort maintained the ambitious 15.0% default. Although the system successfully stimulated absolute savings of 392.89 W during the first active week, the macroscopic target was never achieved, with consumption even exceeding the baseline in weeks two and three. Conversely, the localized monitoring in the FEUP environment yielded significantly greater compliance, achieving the target in four out of five weeks with maximum savings reaching 53.01%.

The residential environments highlight the challenge of applying static goals to whole-house loads. Home 1 successfully met the 15.0% target in weeks two and five (achieving 15.06% and 19.90% savings, respectively), proving that families can coordinate to reduce significant aggregate loads. However, they fell short in the remaining weeks. Home 2, which had a much shorter active deployment window and an inherently low baseline (averaging only 321.83 W), failed its single weekly challenge, saving only 1.70%.

Addressing RQ3 (see Section 1.3), these disparate outcomes yield critical insights into the design of energy-focused gamification.

First, agency emerges as the primary catalyst. The success of a gamification strategy depends heavily on the alignment between the scale of the goal and the user's direct physical authority over the environment. In FEUP, users had total authority over their specific monitored appliances, allowing them to easily translate a system goal into direct action. In Fraunhofer, individual conservation efforts were diluted by shared infrastructure (e.g., central HVAC), rendering the macroscopic goal demotivating.

Furthermore, the results indicate that static targets are inherently flawed. A fixed 15.0% reduction is an arbitrary metric that does not scale contextually. For a highly efficient low-consuming household like Home 2, carving out an additional 15.0% without compromising basic living standards is incredibly difficult. Conversely, for the localized FEUP setup, 15.0% was too conservative, as users easily cleared 40-50% reductions. To counter this, weekly targets should be dynamic, mirroring the difficulty adaptation mechanism of the daily tasks gamification engine. As demonstrated by the daily tasks, dynamically adapting the challenge to the user's actual capabilities and the historical elasticity of the load successfully maintained user engagement over a long period.

This adaptability explains why, across all environments, users exhibited much higher compliance with the daily tasks than with the rigid weekly challenges. Short-term, highly specific and dynamically scaling tasks foster a tighter feedback loop, continuously satisfying the psychological need for competence. In contrast, long-term static aggregate challenges risk inducing apathy if early failures make the target mathematically unattainable midway through the week.

4.2.5 AI Agent Convergence and Nudge Interactions

The RL agent progressively stabilized its decision-making policy by shifting its exploration-to-exploitation ratio in response to real-time human feedback. As the agent learned from user interactions, the probability of selecting a random exploratory action (ϵ) decreased, allowing it to exploit established optimal policies. Fig. 4.10 illustrates this learning trajectory across the monitored environments, correlating the ϵ -decay (left axis) with the episodic reward signal (right axis).

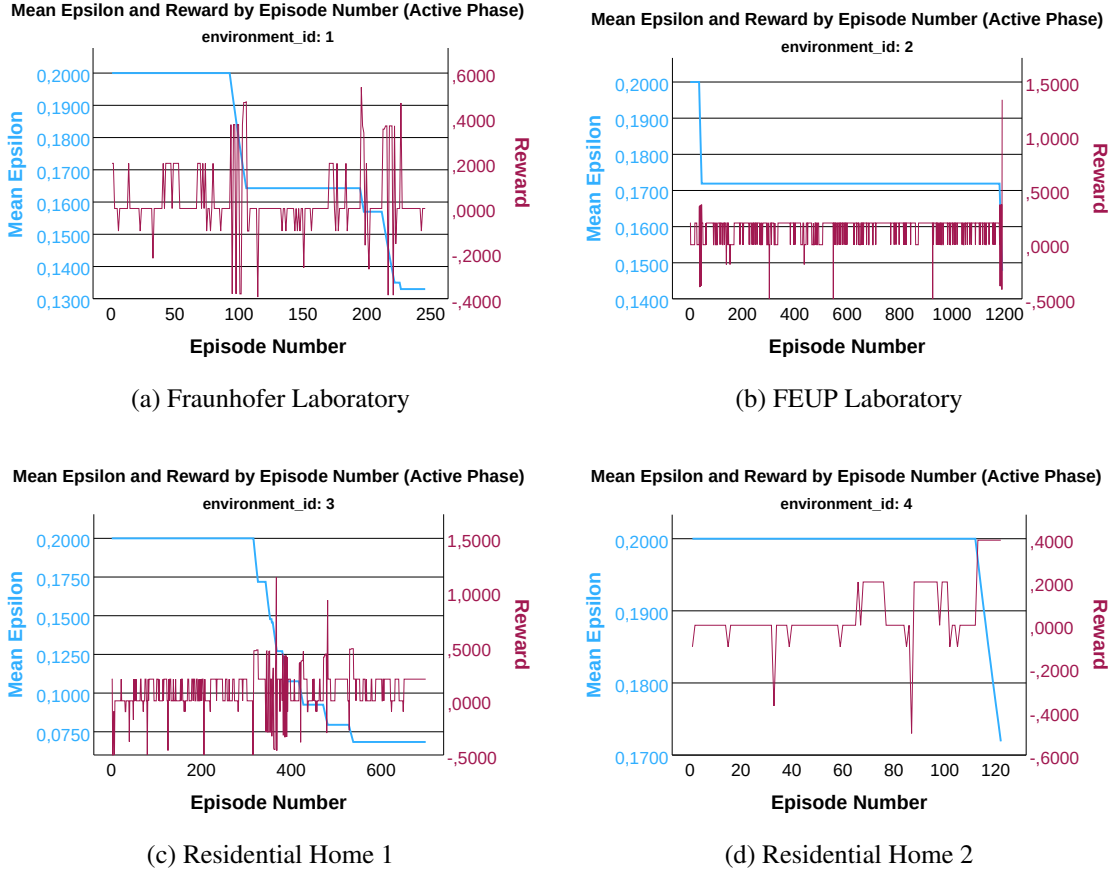


Figure 4.10: RL convergence - Dual-axis charts illustrating the epsilon (ϵ) decay trajectory (left axis) against the episodic reward (right axis) across professional and residential environments.

In the Fraunhofer laboratory (Fig. 4.10a), the epsilon decay followed a multi-step descent as the agent gathered sufficient empirical feedback to confidently narrow its policy. This gradual learning trajectory resulted in an overall behavioural distribution of 24.4% exploration and 75.6% exploitation. In contrast, the FEUP agent (Fig. 4.10b) experienced a sharp immediate drop in exploration early in the active phase, maintaining a long stable exploitation phase for the majority of the active period before a final stabilization adjustment. This aggressive shift is clearly reflected in its highly exploitative overall ratio, recording just 6.2% exploration against 93.8% exploitation.

The residential environments exhibited highly divergent learning paths. In Home 1 (Fig. 4.10c), the agent maintained a high exploration rate for a prolonged initial period (approximately 350 episodes) before executing a rapid staircase descent down to a highly exploitative state (below 0.1). This ultimately yielded a dense exploitation distribution, comprising 10.7% exploration and 89.3% exploitation. Conversely, Home 2 (Fig. 4.10d) maintained a flat exploration rate at 0.2 for the vast majority of its active phase. However, after approximately 115 episodes, the agent finally accumulated enough human feedback to trigger the decay mechanism, resulting in a sharp sudden drop to approximately 0.17 at the very end of the observation period. Because this convergence was significantly delayed due to the environment's shorter deployment window and fewer total episodes, the agent recorded a higher overall exploration rate of 23.8%, alongside 76.2%

exploitation. Nonetheless, this late decay confirms the agent’s capacity to eventually shift toward exploitation once the necessary empirical threshold is met.

Across all deployments combined, the RL architecture demonstrated a strong overall capacity to successfully identify and prioritize optimal user-specific interventions, achieving a global average of 8.9% exploration and 91.1% exploitation.

Answering RQ4 (introduced in Section 1.3) requires looking beyond sheer energy reduction to the preservation of digital mental health and sustained engagement. The core additive value of the AI agent was its ability to autonomously model and respect human boundaries by evaluating the necessity of an intervention against the user’s current cognitive load. Fig. 4.11 visualizes this dynamic evaluation, plotting the calculated opportunity score of a potential nudge against the user’s fatigue index ($I_{fatigue}$).

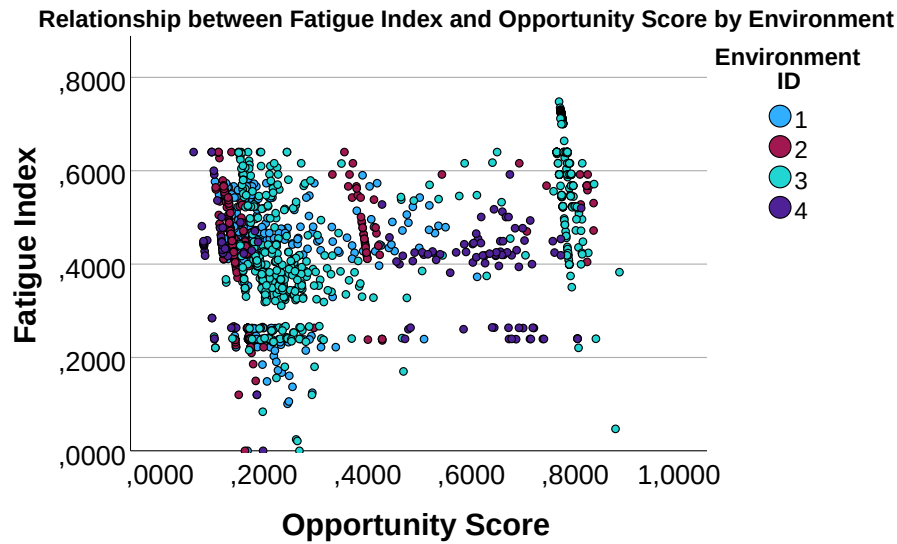


Figure 4.11: Grouped scatter plot illustrating the distribution of suppressed notifications, crossing the calculated opportunity score with the user fatigue index ($I_{fatigue}$) across environments.

Notably, Environment 2 (FEUP) and Environment 3 (Home 1) frequently generated substantially higher opportunity scores, with dense clusters of instances reaching or exceeding the 0.8 threshold. As established during the architectural overview, this disparity reflects the fundamental difference in monitoring resolution. Because the FEUP environment utilized localized appliance-level sensors and Home 1 featured a comprehensive blend of device and whole-house data, the AI could identify highly actionable and targeted intervention opportunities. In contrast, the facility-wide aggregate monitoring at Fraunhofer (Environment 1) was mostly clustered in lower opportunity scores (0.2-0.5). Lacking appliance-level resolution, the system could not pinpoint specific actionable waste with high confidence in this shared space.

Despite the identification of high-value opportunities, the scatter plot confirms that the agent actively prioritized the mitigation of user burnout across all settings. By mapping the opportunity

scores against the dynamically tracked fatigue index, the AI effectively prevented systemic annoyance, ensuring that even highly relevant nudges were suppressed if the user's cognitive load was determined to be too high.

To quantify this self-throttling mechanism, Fig. 4.12 details the notification block rate across the deployment environments.

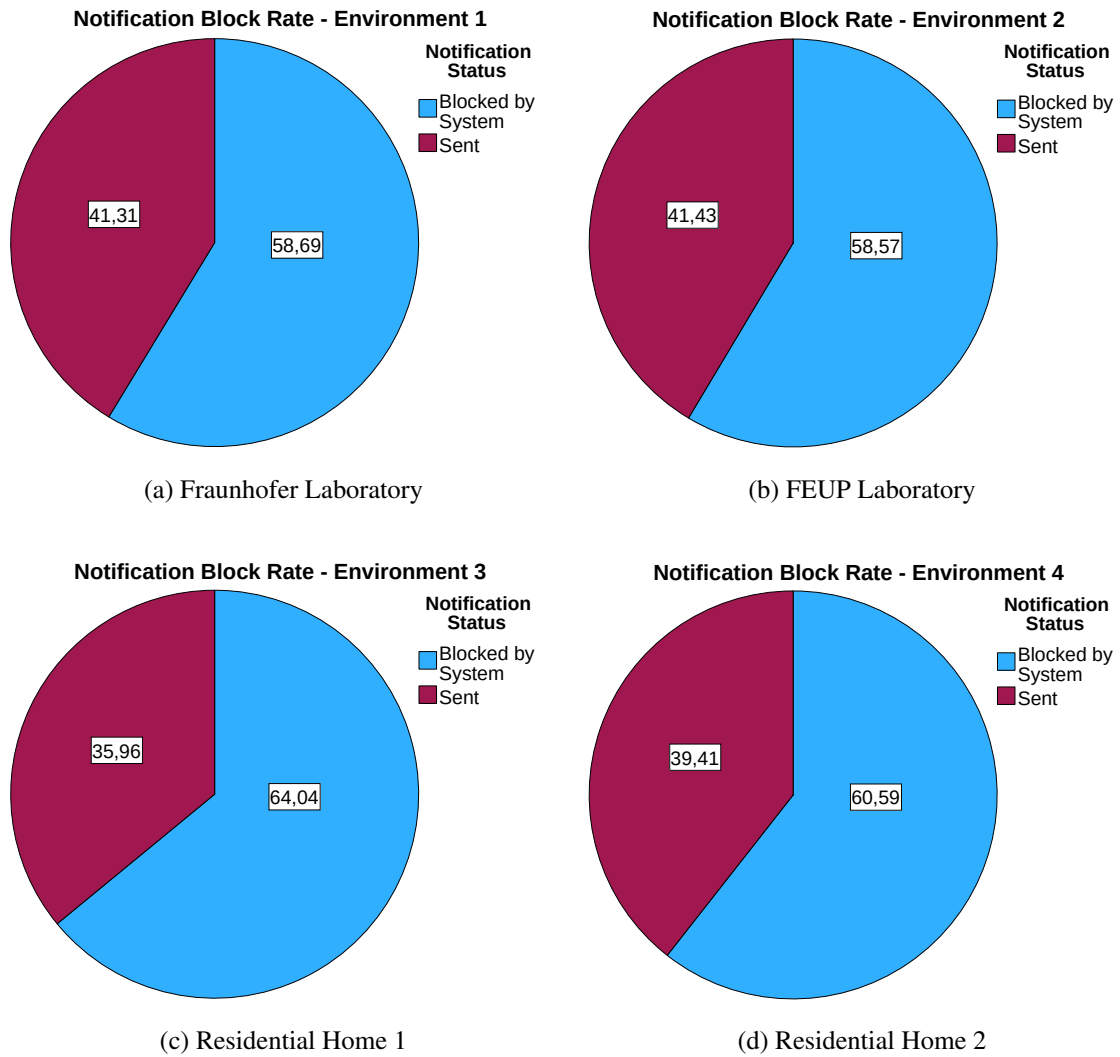


Figure 4.12: Notification block rate - Proportion of AI-generated nudges sent to the user versus those autonomously blocked by the system to prevent fatigue.

The AI agent autonomously suppressed the majority of all generated notification intents across every environment: 58.69% in the Fraunhofer laboratory, 58.57% in the FEUP laboratory, 64.04% in Home 1 and 60.59% in Home 2. By blocking roughly 60% of potential alerts universally due to high calculated fatigue scores, this mechanism successfully shielded users from technical spam and sensory overload, ensuring that only the most critical or highly likely-to-be-accepted nudges breached the interface.

To further address RQ3 and RQ4 (formulated in Section 1.3) regarding the most effective gamification strategies and the efficacy of AI-driven personalization respectively, the analysis of notification outcomes by action type (Fig. 4.13) uncovers critical context-dependent user preferences and engagement barriers.

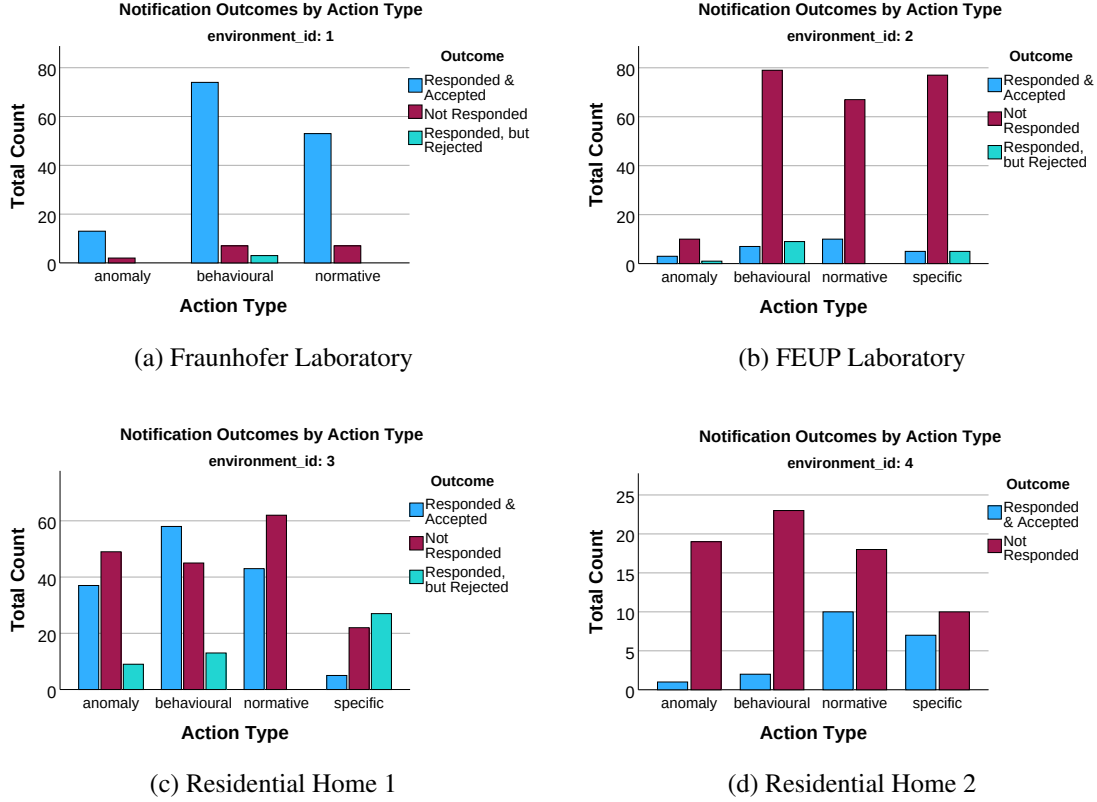


Figure 4.13: Notification outcomes by action type - Clustered bar charts detailing the count of ignored, rejected and accepted nudges.

In the Fraunhofer setting (Fig. 4.13a), occupants exhibited high acceptance rates (R_{accept}) for broad behavioural tips (96.1%) and normative nudges (100%). By focusing on historical self-comparison rather than toxic leaderboards, normative nudges fostered collective accountability without triggering workplace friction. Crucially, the "specific" targeted commands registered zero occurrences across all outcome categories. Because the Fraunhofer facility lacked appliance-level monitoring, the system's dynamic Action Validity Mask (M) successfully filtered out these impossible commands. This feature restricted the available action space to match the physical hardware, ensuring the system only generated interventions that could be accurately monitored and verified.

Conversely, the FEUP environment (Fig. 4.13b) and both residential settings (Figs. 4.13c, 4.13d) were characterized by a massive volume of "Not Responded" interactions across most nudge types. This phenomenon is fundamentally tied to the structural and routine realities of these spaces: the FEUP laboratory was frequently unoccupied and residents in the homes were regularly away during the day for work and other daily errands. Consequently, a large portion of the generated notifications naturally timed out before users could physically engage with the

system. When users did engage at FEUP, normative nudges were universally accepted (100.0%), alongside a solid acceptance of anomaly alerts (75.0%), whereas specific (50.0%) and behavioural (43.8%) prompts saw much lower adoption.

Despite this baseline absence, user engagement remained highly positive when participants were actually present. In Home 1, users still actively accepted a significant portion of anomaly (80.4%), behavioural (81.7%) and normative (100.0%) alerts when they interacted with the dashboard, though specific commands were largely ignored or rejected (15.6%), prompting the RL agent to successfully learn and subsequently suppress this intervention type over time. Home 2 presented an exceptionally unique profile: every nudge that was actively reviewed was accepted, yielding a flawless 100.0% acceptance rate across the anomaly, behavioural, normative and specific categories. While this indicates that the AI's contextual filtering was highly accurate for this environment when the users were home to receive it, this perfect compliance may also be partially attributable to a novelty effect, given the household's shorter active deployment window.

Amidst these environmental differences, a universal trend emerged regarding "anomaly" nudges. Anomaly alerts accounted for a lower total volume of notifications sent because significant load deviations or forgotten appliances are inherently rare and exceptional events rather than routine occurrences. Despite this lower frequency, they exhibited remarkably high acceptance and low explicit rejection rates across the board (100.0% at Fraunhofer, 75.0% at FEUP, 80.4% in Home 1 and 100.0% in Home 2). This indicates that when the system detects a true immediate anomaly, users readily recognize the high relevance of the alert and are highly unlikely to dismiss it.

4.2.6 Qualitative User Assessment

While the physical telemetry quantified structural energy reductions, post-intervention psychometric evaluations contextualized the user experience. A final survey was administered to capture the subjective dimensions of system usability, intrinsic motivation, perceived AI trustworthiness and intervention fatigue. The complete set of questionnaire items corresponding to the variables discussed in this section is detailed in Appendix A.

4.2.6.1 System Acceptability and AI Perception

To ensure that any observed lack of engagement was not erroneously attributed to behavioural resistance when it might stem from poor UI/UX design, the standardized SUS was applied. As shown in Fig. 4.14a, the raw item scores display the classic alternating pattern characteristic of the SUS framework, where users strongly agreed (scores ~ 4.5 out of 5) with positive statements (odd-numbered items) and strongly disagreed (scores ~ 1.2) with negative statements (even-numbered items). This consistency translates to a highly exceptional overall usability score of 90.0, confirming that the interface was intuitive and user friction was minimized.

Regarding gamified motivation (Fig. 4.14b), user responses on the 7-point IMI subscales were overwhelmingly positive, with average item scores ranging between 5.2 and 6.5. This indicates

a strong alignment with both the *Interest/Enjoyment* and *Value/Usefulness* dimensions, demonstrating that participants did not view the system merely as a monitoring tool, but recognized the practical and ecological value of their gamified tasks.

Furthermore, evaluations of the autonomous RL agent (Fig. 4.14c) revealed that participants maintained a high degree of trust in the system’s decision-making. With scores consistently clustered around 4.2 to 4.4 on a 5-point scale, users generally perceived the AI’s contextual nudges as highly transparent, accurate and relevant to their real-time environments, validating the agent’s underlying state-observation logic.

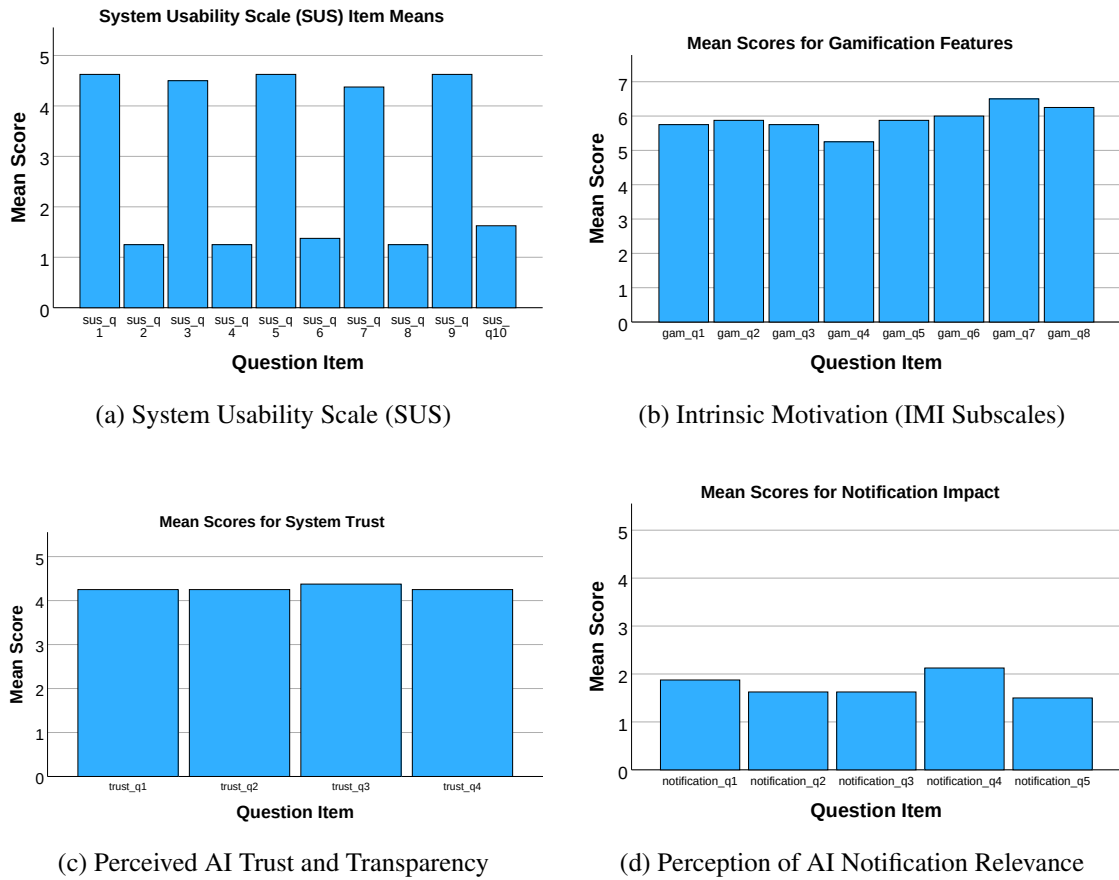


Figure 4.14: User perceptions regarding system usability, gamified motivation and AI trustworthiness.

4.2.6.2 Cognitive Load and Behavioural Shifts

A primary concern of proactive energy interventions is the accumulation of user annoyance and notification fatigue. Fig. 4.14d measures the self-reported impact and intrusiveness of the AI notifications. The low mean scores (ranging from 1.5 to 2.1 on a 5-point scale) are a highly positive indicator, revealing that users generally disagreed with statements characterizing the nudges as annoying, ill-timed or disruptive to their daily routines.

However, when evaluating overall accumulated fatigue across the specific deployment environments (Fig. 4.15a), a clear contextual divide emerges. Participants in the localized FEUP laboratory reported the lowest overall fatigue (mean of 1.5 out of 5), closely followed by the residential homes (mean of 2.0). In stark contrast, participants in the shared Fraunhofer laboratory reported a moderate fatigue level (mean of 3.0). This subjective assessment aligns perfectly with the quantitative telemetry discussed earlier: in environments where users lacked direct physical agency over the monitored loads (Fraunhofer), the nudges were perceived as more fatiguing and burdensome. When users possess control (FEUP and Homes), interventions are perceived as actionable guidance rather than spam.

Regarding long-term behavioural shifts (Fig. 4.15b), participants reported a high likelihood of maintaining energy-saving routines without explicit prompts, with scores ranging from 4.0 to 4.6 on a 5-point scale. While this suggests the nudges successfully encouraged lasting physical habits, it is crucial to note that these subjective self-reports were not corroborated by quantitative telemetry in the professional settings, where the observed power premium reductions lacked statistical significance.

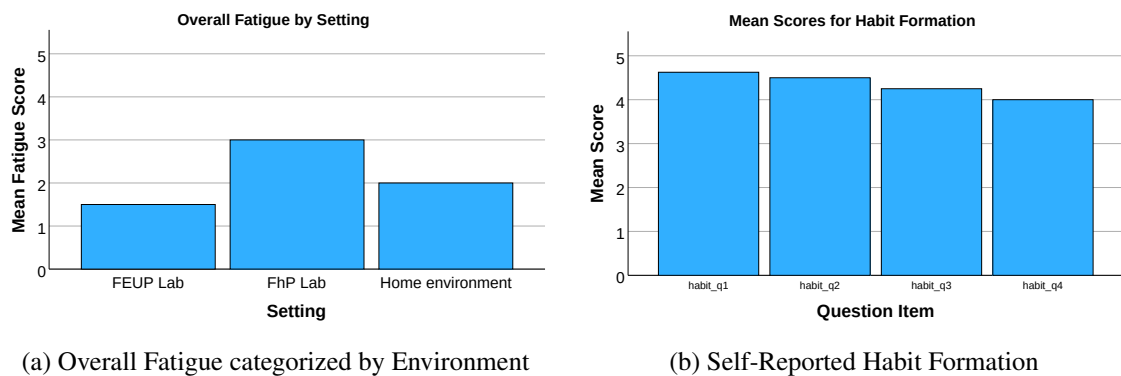


Figure 4.15: Assessment of the intervention's human impact: cognitive load (fatigue) versus long-term behavioural shifts.

4.2.6.3 Open-Ended Contextual Feedback

Thematic analysis of the unstructured free-text feedback provided critical context for the quantitative telemetry and highlighted specific areas for architectural refinement. Participants emphasized that targeted appliance-level suggestions were highly actionable, reinforcing that granular visibility catalyses immediate behavioural shifts.

However, users also identified a specific usability bottleneck: the lack of interactive "accept" and "reject" buttons directly within mobile push notifications. Participants explicitly noted that enabling lock-screen interactions would eliminate the friction of navigating to the dashboard, thereby streamlining the human-in-the-loop feedback cycle and improving response rates.

Additionally, users requested higher contextual resolution within the notifications. Consistently including exact details, such as the specific room, targeted device and precise timestamp, would enhance AI explainability and increase user compliance. Finally, participants recommended

expanding the variety of generated messages to prevent content repetitiveness and mitigate long-term notification fatigue.

4.2.7 Summary of Validation Findings

To consolidate the extensive statistical analysis and provide a clear accessible overview of the system's empirical impact, this subsection summarizes the core findings of the evaluation.

Overall, the deployment of GreenShift demonstrated that passive eco-feedback alone is insufficient to guarantee sustained energy reductions. While real-time visualization can trigger an initial decrease in energy consumption, particularly in environments where users have direct control over their appliances, these initial gains are often volatile and prone to degradation over time.

The introduction of active gamification proved essential in stabilizing these behaviours, yet the study revealed that its success is heavily dependent on user agency and physical occupancy. In shared macroscopic environments where individual efforts were diluted by communal infrastructure, gamification struggled to overcome inelastic baseloads. Conversely, in localized settings with high user agency, the gamified mechanics effectively translated intent into verifiable action. Furthermore, the evaluation established that granular dynamically scaling daily tasks consistently outperformed rigid static weekly challenges. By continuously adapting to the users' actual capabilities and historical consumption patterns, these short-term tasks fostered a much tighter feedback loop and drove significant instantaneous power reductions.

Crucially, the autonomous RL agent emerged as a vital component for sustaining long-term engagement. By continuously monitoring a dynamically calculated user fatigue index, the AI effectively throttled notifications, suppressing approximately 60% of generated alerts to prevent sensory overload. This self-regulating mechanism ensured that interventions remained relevant and actionable, shielding users from technical spam. Ultimately, this approach resulted in an exceptional SUS score of 90.0 and fostered deep intrinsic motivation among participants.

In conclusion, the study confirms that an edge-based AI system can successfully translate users' intentions into actual energy savings. However, to effectively bridge this gap, the system cannot rely on generic or intrusive approaches. Instead, its interventions must focus on specific controllable devices, adapt dynamically to the users' changing habits and actively limit notifications to prevent fatigue.

Chapter 5

Conclusion and Future Work

This final chapter consolidates the theoretical, technical and empirical contributions of this dissertation. The research journey was fundamentally driven by the need to bridge the intention-action gap in energy consumption, a challenge addressed through the development and deployment of GreenShift, a sociotechnical edge-computing architecture. To evaluate the success of this intervention, this chapter synthesizes the empirical findings to confirm the study’s central hypothesis and demonstrate how the initial research objectives were met. It then outlines the project’s primary theoretical and practical contributions, before acknowledging current limitations to establish a clear roadmap for future research.

5.1 Synthesis of Findings and Hypothesis Validation

The transition towards sustainable energy consumption is fundamentally a sociotechnical challenge. While advancements in IoT infrastructure have enabled the granular monitoring of energy flows, the provision of raw data is frequently insufficient to foster lasting behavioural change. To address this limitation, this dissertation engineered GreenShift, a custom HA integration, to investigate how eco-feedback, meaningful gamification and AI-driven personalization could be seamlessly integrated to bridge the critical intention-action gap.

By deploying GreenShift across professional laboratories and residential homes, the empirical validation provided clear answers to the core RQs posed at the outset. Regarding traditional eco-feedback (RQ1), the baseline profiling phase confirmed that passive data display yields highly volatile outcomes. While visualization can act as a powerful initial catalyst in environments where users have direct physical agency, its standalone effect often fails to prevent long-term consumption increases.

To stabilize these behaviours, the integration of real-time gamification (RQ2) proved essential. On a daily intra-phase basis, active user engagement with gamified mechanics consistently resulted in lower weather-normalized energy consumption (E_{norm}) and substantial instantaneous power reductions ($\Delta P_{task_savings}$) across all environments. However, the efficacy of specific gamification strategies (RQ3) was shown to be strictly dictated by user agency and environmental

topology. In localized setups, specific appliance-targeted nudges demonstrated moderate compliance and users achieved high completion rates. Conversely, in shared professional spaces lacking granular monitoring, cooperative normative social comparisons and behavioural tips proved highly effective. Furthermore, the results highlighted that dynamically scaling daily tasks maintained significantly higher long-term engagement than rigid static weekly targets, proving that gamification must adaptively scale to user capabilities to remain effective.

Crucially, the integration of AI-driven personalization (RQ4) successfully solved the widespread issue of reward fatigue. By autonomously calculating an opportunity score against a dynamically tracked Fatigue Index ($I_{fatigue}$), the RL agent effectively throttled notifications, blocking approximately 60% of generated intents to prevent user burnout. The AI's contextual filtering proved highly accurate, ensuring the system remained relevant while maintaining exceptional user trust.

Based on these combined physical and psychometric data, the central hypothesis of this dissertation is partially confirmed. Quantitatively, the architecture consistently drives immediate verifiable power reductions when users engage with the system. Qualitatively, it achieved an exceptional SUS score of 90.0 and fostered deep intrinsic motivation. Ultimately, the study proves that an integrated IoT platform yields significantly more sustained user engagement than traditional monitoring, although its macroscopic capacity to reduce facility-wide structural loads remains heavily dictated by the specific sociotechnical context and the user's degree of physical agency.

5.2 Main Contributions

A core philosophy of this research is that solutions addressing global challenges like energy efficiency should be accessible, reproducible and scalable. To ensure that the outcomes of this dissertation transcend theoretical validation and provide tangible immediate value to the broader smart home community, the proposed architecture was explicitly designed utilizing open-source frameworks.

Consequently, the core gamification engine, the RL-driven notification manager and the custom user interface developed for this study were packaged as a custom integration, named GreenShift¹. GreenShift has been submitted to HACS and is pending approval. By publishing this integration, developers, researchers and smart home enthusiasts worldwide can deploy the architecture within their own localized privacy-preserving edge networks. This contribution not only democratizes access to advanced behavioural energy tools but also establishes a foundation for future crowdsourced improvements and large-scale deployments across diverse cultural and infrastructural contexts.

Beyond the practical open-source software release, the theoretical frameworks and empirical findings derived from this study have been formalized into a scientific article. This manuscript (presented in Appendix B), submitted to "CIoT 2026: 9th Conference on Cloud and Internet

¹The source code for GreenShift is publicly available at: <https://github.com/RafaelNTeixeira/GreenShift>

of Things", details the novel integration of RLHF and meaningful gamification within edge-computing environments. By disseminating these insights, this research aims to advance the academic discourse on sociotechnical energy systems and provide a rigorously validated blueprint for future researchers.

5.3 Limitations and Future Work

While the architecture achieved significant engagement, the study encountered limitations that warrant future exploration. As observed in the weekly challenge outcomes, the reliance on a static default reduction target (e.g., 15.0%) proved rigid and occasionally demotivating when users faced structural baseline constraints. Future implementations must transition toward fully dynamic AI-generated milestones that actively scale based on the cohort's recent streaks and failures, ensuring that community goals remain challenging yet mathematically achievable.

Furthermore, the discrepancy in nudge acceptance highlights a critical avenue for future research. Because the shared laboratory setup in this study lacked appliance-level interventions, the system's ability to drive specific device-level actions in communal spaces remains untested. This architecture should be validated in a large-scale shared professional environment equipped with localized appliance-level monitoring, effectively isolating the behavioural response from the hardware limitations observed in this study.

Technologically, the reliance on single-modality motion sensors occasionally introduced noise in professional settings where occupants remained stationary at desks for prolonged periods. Future iterations should explore multi-modal presence detection (e.g., combining BLE triangulation with acoustic sensors). Additionally, the natural daily absences in both professional and residential environments resulted in a high volume of unanswered timed-out notifications. Future UI/UX designs should implement lock-screen interactive "accept" and "reject" buttons in mobile devices to eliminate application navigation friction and streamline the human-in-the-loop feedback cycle. Finally, longitudinal studies with extended deployment windows are necessary to fully map the long-term RL convergence and the endurance of habit formation.

References

- [1] Nermina Abdurahmanovic and Anna Cadenbach. “Enhancing energy efficiency through user engagement and behaviour change: A review on gamification approaches and serious games in energy systems”. In: *Energy* 334 (2025), p. 137496. ISSN: 0360-5442. DOI: <https://doi.org/10.1016/j.energy.2025.137496>. URL: <https://www.sciencedirect.com/science/article/pii/S036054422503138X>.
- [2] Ifeoma Adaji and Mikhail Adisa. “A Review of the Use of Persuasive Technologies to Influence Sustainable Behaviour”. In: *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, 2022, pp. 317–325. ISBN: 9781450392327. DOI: [10.1145/3511047.3537653](https://doi.org/10.1145/3511047.3537653). URL: <https://doi.org/10.1145/3511047.3537653>.
- [3] Rishika Agarwal et al. “A review of residential energy feedback studies”. In: *Energy and Buildings* 290 (2023), p. 113071. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2023.113071>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778823003018>.
- [4] Arghavan Akbarieh, Qi Han, and Alexander Klippel. “Playing for a purpose: A systematic review and framework for gamified sustainability interventions in urban built environments”. In: *Results in Engineering* 27 (2025), p. 105277. ISSN: 2590-1230. DOI: <https://doi.org/10.1016/j.rineng.2025.105277>. URL: <https://www.sciencedirect.com/science/article/pii/S2590123025013477>.
- [5] Abdullah Alsalemi et al. “A modular recommender system for domestic energy efficiency”. In: *Environmental Challenges* 13 (2023), p. 100741. ISSN: 2667-0100. DOI: <https://doi.org/10.1016/j.envc.2023.100741>. URL: <https://www.sciencedirect.com/science/article/pii/S2667010023000653>.
- [6] Laura Andolfi et al. “Sociotechnical design of building energy management systems in the public sector: Five design principles”. In: *Applied Energy* 377 (2025), p. 124628. ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2024.124628>. URL: <https://www.sciencedirect.com/science/article/pii/S0306261924020117>.
- [7] Luca Ardito. “Behavioural Modelling for Sustainability in Smart Homes”. In: *Proceedings of the 19th International Conference on Availability, Reliability and Security*. Association for Computing Machinery, 2024. ISBN: 9798400717185. DOI: [10.1145/3664476.3670946](https://doi.org/10.1145/3664476.3670946). URL: <https://doi.org/10.1145/3664476.3670946>.
- [8] Chihye Bae, Hyunji Lee, and Chungyoon Chun. “Beyond the thermostat: A behavioral study of residential cooling practices”. In: *Building and Environment* 282 (2025), p. 113250. ISSN: 0360-1323. DOI: <https://doi.org/10.1016/j.buildenv.2025.113250>. URL: <https://www.sciencedirect.com/science/article/pii/S0360132325007309>.

- [9] Mary Barreto et al. “Designing Smart Plugs for Interactivity and Energy Sustainability via a Survey and Thematic Analysis”. In: *Nordic Human-Computer Interaction Conference*. Association for Computing Machinery, 2022. ISBN: 9781450396998. DOI: [10.1145/3546155.3546681](https://doi.org/10.1145/3546155.3546681). URL: <https://doi.org/10.1145/3546155.3546681>.
- [10] Ariane L Beck, Sohum Chitalia, and Varun Rai. “Not so gameful: A critical review of gamification in mobile energy applications”. In: *Energy Research & Social Science* 51 (2019), pp. 32–39. ISSN: 2214-6296. DOI: <https://doi.org/10.1016/j.erss.2019.01.006>. URL: <https://www.sciencedirect.com/science/article/pii/S2214629618309034>.
- [11] Michelle Berger, Theresa Lange, and Bastian Stahl. “A digital push with real impact – Mapping effective digital nudging elements to contexts to promote environmentally sustainable behavior”. In: *Journal of Cleaner Production* 380 (2022), p. 134716. ISSN: 0959-6526. DOI: <https://doi.org/10.1016/j.jclepro.2022.134716>. URL: <https://www.sciencedirect.com/science/article/pii/S0959652622042883>.
- [12] Adel Binbusayyis and Mohemmed Sha. “Energy consumption prediction using modified deep CNN-Bi LSTM with attention mechanism”. In: *Heliyon* 11 (1 2025), e41507. ISSN: 2405-8440. DOI: <https://doi.org/10.1016/j.heliyon.2024.e41507>. URL: <https://www.sciencedirect.com/science/article/pii/S240584402417538X>.
- [13] S Chadoulos, I Koutsopoulos, and G C Polyzos. “Mobile Apps Meet the Smart Energy Grid: A Survey on Consumer Engagement and Machine Learning Applications”. In: *IEEE Access* 8 (2020), pp. 219632–219655. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2020.3042758](https://doi.org/10.1109/ACCESS.2020.3042758). URL: <https://doi.org/10.1109/ACCESS.2020.3042758>.
- [14] Spiros Chadoulos et al. “Reinforcement Learning with Partially Defined Rewards and Human Feedback for Energy Efficiency Recommendations”. In: *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*. Association for Computing Machinery, 2025, pp. 622–631. ISBN: 9798400711251. DOI: [10.1145/3679240.3734660](https://doi.org/10.1145/3679240.3734660). URL: <https://doi.org/10.1145/3679240.3734660>.
- [15] Y Chen et al. “BI-Tech: An IoT-Based Behavioral Intervention System for User-Driven Energy Optimization in Commercial Spaces”. In: *IEEE Access* 13 (2025), pp. 166853–166872. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2025.3607329](https://doi.org/10.1109/ACCESS.2025.3607329). URL: <https://doi.org/10.1109/ACCESS.2025.3607329>.
- [16] Z Chen, R.-I. Chang, and Q Wu. “Adaptive Smart System for Energy-Saving Campus †”. In: *Engineering Proceedings* 92 (1 2025). DOI: [10.3390/engproc2025092036](https://doi.org/10.3390/engproc2025092036). URL: <https://www.scopus.com/pages/publications/105009303310>.
- [17] Ming-Chuan Chiu, Tsai-Chi Kuo, and Hsin-Ting Liao. “Design for sustainable behavior strategies: Impact of persuasive technology on energy usage”. In: *Journal of Cleaner Production* 248 (2020), p. 119214. ISSN: 0959-6526. DOI: <https://doi.org/10.1016/j.jclepro.2019.119214>. URL: <https://www.sciencedirect.com/science/article/pii/S0959652619340843>.
- [18] C Chronis et al. “Privacy-Preserving Energy Recommendations Using Federated Learning and Local LLMs on the Edge”. In: *2024 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*. 2024, pp. 377–382. DOI: [10.1109/BDCAT63179.2024.00065](https://doi.org/10.1109/BDCAT63179.2024.00065). URL: <https://doi.org/10.1109/BDCAT63179.2024.00065>.

- [19] D Cottafava et al. “Crowdsensing for a sustainable comfort and for energy saving”. In: *Energy and Buildings* 186 (2019), pp. 208–220. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2019.01.007>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778818329955>.
- [20] Zurisaddai de la Cruz Severiche Maury, Ana Fernández Vilas, and Rebeca P. Díaz Redondo. “Low-Cost HEM with Arduino and Zigbee Technologies in the Energy Sector in Colombia”. In: *Energies* 15.10 (2022). ISSN: 1996-1073. DOI: [10.3390/en15103819](https://doi.org/10.3390/en15103819). URL: <https://www.mdpi.com/1996-1073/15/10/3819>.
- [21] Nirupam Sannagowdara Dasappa, Krishna Kumar G, and Nivethitha Somu. “Multi-sensor data fusion framework for energy optimization in smart homes”. In: *Renewable and Sustainable Energy Reviews* 193 (2024), p. 114235. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2023.114235>. URL: <https://www.sciencedirect.com/science/article/pii/S1364032123010936>.
- [22] Marco Dell’Isola et al. “An IoT Integrated Tool to Enhance User Awareness on Energy Consumption in Residential Buildings”. In: *Atmosphere* 10.12 (2019). ISSN: 2073-4433. DOI: [10.3390/atmos10120743](https://doi.org/10.3390/atmos10120743). URL: <https://www.mdpi.com/2073-4433/10/12/743>.
- [23] T Dilekh et al. “Dynamic Context-Aware Recommender System for Home Automation Through Synergistic Unsupervised and Supervised Learning Algorithms”. In: *Acta Informatica Pragensia* 13 (1 2024), pp. 38–61. DOI: [10.18267/j.aip.228](https://doi.org/10.18267/j.aip.228). URL: <https://www.scopus.com/pages/publications/85190846075>.
- [24] Jia Du and Wei Pan. “Behavioral energy efficiency with environment sensors: A case in Hong Kong”. In: *Energy and Buildings* 299 (2023), p. 113590. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2023.113590>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778823008204>.
- [25] Rolf Egert et al. “Exploring energy grid resilience: The impact of data, prosumer awareness, and action”. In: *Patterns* 2 (6 2021), p. 100258. ISSN: 2666-3899. DOI: <https://doi.org/10.1016/j.patter.2021.100258>. URL: <https://www.sciencedirect.com/science/article/pii/S2666389921000842>.
- [26] Angelica Errigo, Jun-Ki Choi, and Kelly Kisson. “Techno-economic-environmental impacts of industrial energy assessment: Sustainable industrial motor systems of small and medium-sized enterprises”. In: *Sustainable Energy Technologies and Assessments* 49 (2022), p. 101694. ISSN: 2213-1388. DOI: <https://doi.org/10.1016/j.seta.2021.101694>. URL: <https://www.sciencedirect.com/science/article/pii/S2213138821007086>.
- [27] Muhammad Zaman Fakhar, Emre Yalcin, and Alper Bilge. “A survey of smart home energy conservation techniques”. In: *Expert Systems with Applications* 213 (2023), p. 118974. ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2022.118974>. URL: <https://www.sciencedirect.com/science/article/pii/S0957417422019923>.
- [28] Marta Ferreira et al. “Connecting audiences with climate change: Towards humanised and action-focused data interactions”. In: *International Journal of Human-Computer Studies* 192 (2024), p. 103341. ISSN: 1071-5819. DOI: <https://doi.org/10.1016/j.ijhcs.2024.103341>. URL: <https://www.sciencedirect.com/science/article/pii/S1071581924001241>.

- [29] Marta Fonseca Gabriel et al. “Smartphone-enabled real-time data on home electricity use and indoor air quality to promote energy-saving behaviors”. In: *Energy and Buildings* 347 (2025), p. 116344. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2025.116344>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778825010746>.
- [30] B T Geetha et al. “HEMS-IoT: Smart Home Energy Saving with Big Data and Machine Learnings”. In: *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*. Vol. 7. 2024, pp. 400–404. DOI: [10.1109/IC3I61595.2024.10828998](https://doi.org/10.1109/IC3I61595.2024.10828998). URL: <https://doi.org/10.1109/IC3I61595.2024.10828998>.
- [31] Thirumalai Srinivasan Govindarajan, Bhavithrah Jegan, and Sangeetha Rengachary Gopalan. “Development and implementation of a user-centric real-time energy monitoring and feedback system for energy conservation”. In: *Energy and Buildings* 340 (2025), p. 115787. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2025.115787>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778825005171>.
- [32] Bisma Gulzar, Shabir Ahmad Sofi, and Sahil Sholla. “Exploring Personalized Internet of Things (PIoT), social connectivity, and Artificial Social Intelligence (ASI): A survey”. In: *High-Confidence Computing* 4 (3 2024), p. 100242. ISSN: 2667-2952. DOI: <https://doi.org/10.1016/j.hcc.2024.100242>. URL: <https://www.sciencedirect.com/science/article/pii/S266729522400045X>.
- [33] Laura Guzman, Stephen Makonin, and Alex Clapp. “CarbonKit: Designing A Personal Carbon Tracking Platform”. In: *Proceedings of the Fourth International Workshop on Social Sensing*. Association for Computing Machinery, 2019, pp. 24–29. ISBN: 9781450367066. DOI: [10.1145/3313294.3313385](https://doi.org/10.1145/3313294.3313385). URL: <https://doi.org/10.1145/3313294.3313385>.
- [34] Y He et al. “Enhancing Intelligent HVAC optimization with graph attention networks and stacking ensemble learning, a recommender system approach in Shenzhen Qianhai Smart Community”. In: *Scientific Reports* 15 (1 2025). DOI: [10.1038/s41598-025-89776-6](https://doi.org/10.1038/s41598-025-89776-6). URL: <https://www.scopus.com/pages/publications/85218475952>.
- [35] José L Hernández et al. “Challenges and opportunities in European smart buildings energy management: A critical review”. In: *Renewable and Sustainable Energy Reviews* 199 (2024), p. 114472. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2024.114472>. URL: <https://www.sciencedirect.com/science/article/pii/S1364032124001953>.
- [36] Melanie R Herrmann et al. “An empirical investigation of domestic energy data visualizations”. In: *International Journal of Human-Computer Studies* 152 (2021), p. 102660. ISSN: 1071-5819. DOI: <https://doi.org/10.1016/j.ijhcs.2021.102660>. URL: <https://www.sciencedirect.com/science/article/pii/S1071581921000781>.
- [37] Arsalan Heydarian et al. “What drives our behaviors in buildings? A review on occupant interactions with building systems from the lens of behavioral theories”. In: *Building and Environment* 179 (2020), p. 106928. ISSN: 0360-1323. DOI: <https://doi.org/10.1016/j.buildenv.2020.106928>. URL: <https://www.sciencedirect.com/science/article/pii/S0360132320302870>.
- [38] Y Himeur et al. “Detection of Appliance-Level Abnormal Energy Consumption in Buildings Using Autoencoders and Micro-moments”. In: *Lecture Notes in Networks and Systems*. Vol. 489 LNNS. 2022, pp. 179–193. DOI: [10.1007/978-3-031-07969-6_14](https://doi.org/10.1007/978-3-031-07969-6_14). URL: <https://www.scopus.com/pages/publications/85135039318>.

- [39] Yassine Himeur et al. “A survey of recommender systems for energy efficiency in buildings: Principles, challenges and prospects”. In: *Information Fusion* 72 (2021), pp. 1–21. ISSN: 1566-2535. DOI: <https://doi.org/10.1016/j.inffus.2021.02.002>. URL: <https://www.sciencedirect.com/science/article/pii/S1566253521000178>.
- [40] Yassine Himeur et al. “Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives”. In: *Applied Energy* 287 (2021), p. 116601. ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2021.116601>. URL: <https://www.sciencedirect.com/science/article/pii/S0306261921001409>.
- [41] Yassine Himeur et al. “Techno-economic assessment of building energy efficiency systems using behavioral change: A case study of an edge-based micro-moments solution”. In: *Journal of Cleaner Production* 331 (2022), p. 129786. ISSN: 0959-6526. DOI: <https://doi.org/10.1016/j.jclepro.2021.129786>. URL: <https://www.sciencedirect.com/science/article/pii/S0959652621039615>.
- [42] M Huang et al. “Unveiling the Mediating Role of Environmental Awareness: A Study on Gamified Rewards”. In: *Journal of Mobile Multimedia* 21 (1 2025), pp. 27–56. DOI: [10.13052/jmm1550-4646.2112](https://doi.org/10.13052/jmm1550-4646.2112). URL: <https://www.scopus.com/pages/publications/105004238697>.
- [43] B Huseynli. “Gamification in Energy Consumption: A Model for Consumers’ Energy Saving”. In: *International Journal of Energy Economics and Policy* 14 (1 2024), pp. 312–320. DOI: [10.32479/ijeep.14395](https://doi.org/10.32479/ijeep.14395). URL: <https://www.scopus.com/pages/publications/85184242313>.
- [44] Ayman Al-Kababji et al. “Interactive visual study for residential energy consumption data”. In: *Journal of Cleaner Production* 366 (2022), p. 132841. ISSN: 0959-6526. DOI: <https://doi.org/10.1016/j.jclepro.2022.132841>. URL: <https://www.sciencedirect.com/science/article/pii/S0959652622024350>.
- [45] Rana Asad Javid Khan and Jason D Lucas. “Building for human experiences: An in-depth exploration of human-centric design”. In: *Journal of Building Engineering* 106 (2025), p. 112493. ISSN: 2352-7102. DOI: <https://doi.org/10.1016/j.jobe.2025.112493>. URL: <https://www.sciencedirect.com/science/article/pii/S2352710225007302>.
- [46] Dongwoo Kim, Chihye Bae, and Jae Yong Lee. “Heat demand response of end users based on energy awareness level and energy-monitoring app usage”. In: *Journal of Building Engineering* 112 (2025), p. 113721. ISSN: 2352-7102. DOI: <https://doi.org/10.1016/j.jobe.2025.113721>. URL: <https://www.sciencedirect.com/science/article/pii/S2352710225019588>.
- [47] Dimosthenis Kotsopoulos, Cleopatra Bardaki, and Katerina Pramataris. “How to motivate employees towards organizational energy conservation: Insights based on employees perceptions and an IoT-enabled gamified IS intervention”. In: *Heliyon* 9 (5 2023), e16314. ISSN: 2405-8440. DOI: <https://doi.org/10.1016/j.heliyon.2023.e16314>. URL: <https://www.sciencedirect.com/science/article/pii/S2405844023035211>.

- [48] S Lai et al. “Advanced graph embedding for intelligent heating, ventilation, and air conditioning optimization: An ensemble learning-based recommender system”. In: *Case Studies in Thermal Engineering* 68 (2025). DOI: [10.1016/j.csite.2025.105888](https://doi.org/10.1016/j.csite.2025.105888). URL: <https://www.scopus.com/pages/publications/85218426486>.
- [49] M Laouadi, M Kissi, and O E El Beggar. “Gamification in energy consumption systems: A systematic literature review”. In: *2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA)*. 2023. DOI: [10.1109/SITA60746.2023.10373600](https://doi.org/10.1109/SITA60746.2023.10373600). URL: <https://www.scopus.com/pages/publications/85183323930>.
- [50] Clyde Zhengdao Li et al. “Advances in the research of building energy saving”. In: *Energy and Buildings* 254 (2022), p. 111556. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2021.111556>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778821008409>.
- [51] Lanlan Li, Xiaomeng Yuan, and Jingjing Li. “Understanding visual feedback mechanism from three-dimensional of information, time and display: A meta-analysis of feedback research for household electricity conservation”. In: *Energy and Buildings* 316 (2024), p. 114297. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2024.114297>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778824004134>.
- [52] W.-T. Li et al. “Gamification Approaches and Assessment Methodologies for Occupants’ Energy Behavior Change in Buildings: A Systematic Review”. In: *Buildings* 14 (6 2024). DOI: [10.3390/buildings14061497](https://doi.org/10.3390/buildings14061497). URL: <https://www.scopus.com/pages/publications/85197879585>.
- [53] Gunnar Lima et al. “Energy efficiency analysis: A household digital transformation”. In: *AIMS Energy* 12 (Jan. 2024), pp. 774–808. DOI: [10.3934/energy.2024037](https://doi.org/10.3934/energy.2024037).
- [54] Meng Liu et al. “A preliminary investigation of the effectiveness of feedback messaging on household room air conditioning usage in summer”. In: *Energy and Buildings* 318 (2024), p. 114443. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2024.114443>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778824005590>.
- [55] Fengji Luo, Gianluca Ranzi, and Zhao Yang Dong. “Chapter 14 - Occupant-to-grid integration”. In: *Building Energy Management Systems and Techniques*. Ed. by Fengji Luo, Gianluca Ranzi, and Zhao Yang Dong. Elsevier, 2024, pp. 295–311. ISBN: 978-0-323-96107-3. DOI: <https://doi.org/10.1016/B978-0-323-96107-3.00009-6>. URL: <https://www.sciencedirect.com/science/article/pii/B9780323961073000096>.
- [56] Nan Ma et al. “Blockchain + IoT sensor network to measure, evaluate and incentivize personal environmental accounting and efficient energy use in indoor spaces”. In: *Applied Energy* 332 (2023). ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2022.120443>. URL: <https://www.sciencedirect.com/science/article/pii/S0306261922017007>.
- [57] Adrian A Mabalay. “Gamification for sustainability: A systematic review of applications, trends, and opportunities”. In: *Computers in Human Behavior* 165 (2025), p. 108529. ISSN: 0747-5632. DOI: <https://doi.org/10.1016/j.chb.2024.108529>. URL: <https://www.sciencedirect.com/science/article/pii/S0747563224003972>.

- [58] A Mäkilievierikko et al. “Reducing electricity peak loads through ‘pause hours’ - a community-based behavioural demand response approach”. In: *Journal of Cleaner Production* 408 (2023). DOI: 10.1016/j.jclepro.2023.137064. URL: <https://www.scopus.com/pages/publications/85153860583>.
- [59] L Martins et al. “A Case Study of Smart Home Development”. In: *IEEE Software* (2024), pp. 1–7. ISSN: 1937-4194. DOI: 10.1109/MS.2024.3504732. URL: <https://doi.org/10.1109/MS.2024.3504732>.
- [60] Bruno Mataloto, João C. Ferreira, and Ricardo Pontes Resende. “Long Term Energy Savings Through User Behavior Modeling in Smart Homes”. In: *IEEE Access* 11 (2023), pp. 44544–44558. DOI: 10.1109/ACCESS.2023.3272888.
- [61] S McCarthy et al. “Digital Sustainability for Energy-Efficient Behaviours: A User Representation and Touchpoint Model”. In: *Information Systems Frontiers* 26 (6 2024), pp. 2077–2101. DOI: 10.1007/s10796-024-10509-7. URL: <https://www.scopus.com/pages/publications/85197692688>.
- [62] Esraa A Metwally, Mostafa Refat Ismail, and Ayman A Farid. “Development of an IoT occupant-centric metrics: A systematic literature review”. In: *Results in Engineering* 23 (2024), p. 102614. ISSN: 2590-1230. DOI: <https://doi.org/10.1016/j.rineng.2024.102614>. URL: <https://www.sciencedirect.com/science/article/pii/S2590123024008697>.
- [63] L Mi et al. “Assessing the effect of non-financial information intervention on promoting group-level energy savings”. In: *Science of the Total Environment* 720 (2020). DOI: 10.1016/j.scitotenv.2020.137533. URL: <https://www.scopus.com/pages/publications/85080084091>.
- [64] S Mischos, E Dalagdi, and D Vrakas. “Intelligent energy management systems: a review”. In: *Artificial Intelligence Review* 56 (10 2023), pp. 11635–11674. DOI: 10.1007/s10462-023-10441-3. URL: <https://www.scopus.com/pages/publications/85149937947>.
- [65] P Mochi et al. “Social Nudging for Sustainable Electricity Use: Behavioral Interventions in Energy Conservation Policy”. In: *Sustainability (Switzerland)* 17 (15 2025). DOI: 10.3390/su17156932. URL: <https://www.scopus.com/pages/publications/105013127526>.
- [66] A E Onile et al. “Uses of the digital twins concept for energy services, intelligent recommendation systems, and demand side management: A review”. In: *Energy Reports* 7 (2021), pp. 997–1015. DOI: 10.1016/j.egyr.2021.01.090. URL: <https://www.scopus.com/pages/publications/85100666778>.
- [67] Naeem Al-Oudat et al. “IoT-Based Home and Community Energy Management System in Jordan”. In: *Procedia Computer Science* 160 (2019), pp. 142–148. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2019.09.454>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050919316692>.
- [68] Akito Ozawa, Ryota Furusato, and Yoshikuni Yoshida. “Tailor-Made Feedback to Reduce Residential Electricity Consumption: The Effect of Information on Household Lifestyle in Japan”. In: *Sustainability* 9.4 (2017). ISSN: 2071-1050. DOI: 10.3390/su9040528. URL: <https://www.mdpi.com/2071-1050/9/4/528>.

- [69] Chandra Prakash Paneru and Ari K M Tarigan. “Reviewing the impacts of smart energy applications on energy behaviours in Norwegian households”. In: *Renewable and Sustainable Energy Reviews* 183 (2023), p. 113511. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2023.113511>. URL: <https://www.sciencedirect.com/science/article/pii/S1364032123003684>.
- [70] Alexios Papaioannou, Asimina Dimara, and Stelios Krinidis. “GUIDE: A prescriptive hybrid AI framework for energy-efficient appliances usage through behavioral modeling and LLM guidance”. In: *Energy and Buildings* 348 (2025), p. 116463. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2025.116463>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778825011934>.
- [71] H Pombeiro et al. “Displaying data is not enough: Incorporating User Behavior Transformation in domestic reporting systems”. In: *Sustainable Cities and Society* 48 (2019). DOI: [10.1016/j.scs.2019.101451](https://doi.org/10.1016/j.scs.2019.101451). URL: <https://www.scopus.com/pages/publications/85061693926>.
- [72] Pedro Ponce et al. “Tailored gamification and serious game framework based on fuzzy logic for saving energy in connected thermostats”. In: *Journal of Cleaner Production* 262 (2020), p. 121167. ISSN: 0959-6526. DOI: <https://doi.org/10.1016/j.jclepro.2020.121167>. URL: <https://www.sciencedirect.com/science/article/pii/S0959652620312142>.
- [73] Madhukrishna Priyadarsini, Rahul, and Reetanjali Paikra. “AI-driven IoT and gamification for smart water management: Real-time monitoring and predictive analytics”. In: *Environmental Modelling & Software* (2025), p. 106711. ISSN: 1364-8152. DOI: <https://doi.org/10.1016/j.envsoft.2025.106711>. URL: <https://www.sciencedirect.com/science/article/pii/S1364815225003950>.
- [74] Hamed Nabizadeh Rafsanjani, Ali Ghahramani, and Amir Hossein Nabizadeh. “iSEA: IoT-based smartphone energy assistant for prompting energy-aware behaviors in commercial buildings”. In: *Applied Energy* 266 (2020), p. 114892. ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2020.114892>. URL: <https://www.sciencedirect.com/science/article/pii/S0306261920304049>.
- [75] J Rasku, M Mäenpää, and S Alanko. “Your Very Own Polar Bear: Drawing Attention to Household Energy Efficiency Through Gamification”. In: *2024 International Workshop on Artificial Intelligence and Machine Learning for Energy Transformation (AIE)*. 2024, pp. 1–6. DOI: [10.1109/AIE61866.2024.10561363](https://doi.org/10.1109/AIE61866.2024.10561363). URL: <https://doi.org/10.1109/AIE61866.2024.10561363>.
- [76] Valentyna Riabchuk et al. “Utility-based context-aware multi-agent recommendation system for energy efficiency in residential buildings”. In: *Information Fusion* 112 (2024), p. 102559. ISSN: 1566-2535. DOI: <https://doi.org/10.1016/j.inffus.2024.102559>. URL: <https://www.sciencedirect.com/science/article/pii/S1566253524003373>.
- [77] Giuseppe Russo et al. “Strategies and technologies for the sustainable development of energy, water, and environmental systems”. In: *Energy Reports* 14 (2025), pp. 854–866. ISSN: 2352-4847. DOI: <https://doi.org/10.1016/j.egyr.2025.06.032>. URL: <https://www.sciencedirect.com/science/article/pii/S2352484725003993>.

- [78] C Sardinios et al. “Data Analytics, Automations, and Micro-Moment Based Recommendations for Energy Efficiency”. In: *2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService)*. 2020, pp. 96–103. DOI: [10.1109/BigDataService49289.2020.00022](https://doi.org/10.1109/BigDataService49289.2020.00022). URL: <https://doi.org/10.1109/BigDataService49289.2020.00022>.
- [79] C Sardinios et al. “Reshaping Consumption Habits by Exploiting Energy-Related Micro-moment Recommendations: A Case Study”. In: *Communications in Computer and Information Science*. Vol. 1217. 2021, pp. 65–84. DOI: [10.1007/978-3-030-68028-2_4](https://doi.org/10.1007/978-3-030-68028-2_4). URL: <https://www.scopus.com/pages/publications/85101522634>.
- [80] Christos Sardinios et al. “REHAB-C: Recommendations for Energy HABits Change”. In: *Future Generation Computer Systems* 112 (2020), pp. 394–407. ISSN: 0167-739X. DOI: <https://doi.org/10.1016/j.future.2020.05.041>. URL: <https://www.sciencedirect.com/science/article/pii/S0167739X19317327>.
- [81] Aya Nabil Sayed et al. “Enhancing building sustainability: A Digital Twin approach to energy efficiency and occupancy monitoring”. In: *Energy and Buildings* 328 (2025), p. 115151. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2024.115151>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778824012672>.
- [82] H N D Senyapar, S Ayik, and R Bayindir. “AI-Enhanced Gamification for Energy Efficiency: Strategies for Sustainable and Equitable Consumer Engagement”. In: *2025 IEEE 19th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*. 2025, pp. 1–6. ISBN: 2166-9546. DOI: [10.1109/CPE-POWERENG63314.2025.11027225](https://doi.org/10.1109/CPE-POWERENG63314.2025.11027225). URL: <https://doi.org/10.1109/CPE-POWERENG63314.2025.11027225>.
- [83] Z Severiche-Maury and W Arrubla-Hoyos. “Proposal of a Decision-Making Model for Home Energy Saving through Artificial Intelligence applied to a HEMS”. In: *2023 IEEE Colombian Caribbean Conference (C3)*. 2023. DOI: [10.1109/C358072.2023.10436203](https://doi.org/10.1109/C358072.2023.10436203). URL: <https://www.scopus.com/pages/publications/85186764559>.
- [84] William R Shadish, Thomas D Cook, and Donald T Campbell. “Quasi-experiments: interrupted time-series designs”. In: *Experimental and quasi-experimental designs for generalized causal inference* 1 (2002), pp. 171–205.
- [85] Leonie Nora Sieger et al. “User Involvement in Training Smart Home Agents: Increasing Perceived Control and Understanding”. In: *Proceedings of the 10th International Conference on Human-Agent Interaction*. Association for Computing Machinery, 2022, pp. 76–85. ISBN: 9781450393232. DOI: [10.1145/3527188.3561914](https://doi.org/10.1145/3527188.3561914). URL: <https://doi.org/10.1145/3527188.3561914>.
- [86] Jonas Sievers and Thomas Blank. “A Systematic Literature Review on Data-Driven Residential and Industrial Energy Management Systems”. In: *Energies* 16.4 (2023). ISSN: 1996-1073. DOI: [10.3390/en16041688](https://doi.org/10.3390/en16041688). URL: <https://www.mdpi.com/1996-1073/16/4/1688>.
- [87] Paras Sitoula et al. “Context-Aware Smart Energy Recommender (CASER)”. In: *Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia*. Association for Computing Machinery, 2020, pp. 13–19. ISBN: 9781450371780. DOI: [10.1145/3365921.3365935](https://doi.org/10.1145/3365921.3365935). URL: <https://doi.org/10.1145/3365921.3365935>.

- [88] Stefan Stieglitz et al. “The potential of digital nudging to bridge the gap between environmental attitude and behavior in the usage of smart home applications”. In: *International Journal of Information Management* 72 (2023), p. 102665. ISSN: 0268-4012. DOI: <https://doi.org/10.1016/j.ijinfomgt.2023.102665>. URL: <https://www.sciencedirect.com/science/article/pii/S0268401223000464>.
- [89] Yongbo Sun and Jiayuan Xing. “The Impact of Gamification Motivation on Green Consumption Behavior—An Empirical Study Based on Ant Forest”. In: *Sustainability* 15.1 (2023). ISSN: 2071-1050. DOI: [10.3390/su15010512](https://doi.org/10.3390/su15010512). URL: <https://www.mdpi.com/2071-1050/15/1/512>.
- [90] Zeynep Duygu Tekler, Raymond Low, and Lucienne Blessing. “User perceptions on the adoption of smart energy management systems in the workplace: Design and policy implications”. In: *Energy Research & Social Science* 88 (2022), p. 102505. ISSN: 2214-6296. DOI: <https://doi.org/10.1016/j.erss.2022.102505>. URL: <https://www.sciencedirect.com/science/article/pii/S2214629622000123>.
- [91] Le Na Tran, Gangwei Cai, and Weijun Gao. “Determinants and approaches of household energy consumption: A review”. In: *Energy Reports* 10 (2023), pp. 1833–1850. ISSN: 2352-4847. DOI: <https://doi.org/10.1016/j.egyr.2023.08.026>. URL: <https://www.sciencedirect.com/science/article/pii/S235248472301168X>.
- [92] Kevin Trinh, Alan S Fung, and Vera Straka. “Effects of Real-Time Energy Feedback and Normative Comparisons: Results from a Multi-Year Field Study in a Multi-Unit Residential Building”. In: *Energy and Buildings* 250 (2021), p. 111288. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2021.111288>. URL: <https://www.sciencedirect.com/science/article/pii/S0378778821005727>.
- [93] I Tsanov. “Design for Sustainable Behavior: A Focus on Sustainable Energy”. In: *2023 18th Conference on Electrical Machines, Drives and Power Systems (ELMA)*. 2023, pp. 1–4. DOI: [10.1109/ELMA58392.2023.10202440](https://doi.org/10.1109/ELMA58392.2023.10202440). URL: <https://doi.org/10.1109/ELMA58392.2023.10202440>.
- [94] Sanna Tuomela et al. “Impacts of home energy management systems on electricity consumption”. In: *Applied Energy* 299 (2021), p. 117310. ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2021.117310>. URL: <https://www.sciencedirect.com/science/article/pii/S0306261921007224>.
- [95] Ezianya Ubachukwu et al. “User engagement for thermal energy-efficient behavior in office buildings using dashboards and gamification”. In: *Applied Thermal Engineering* 266 (2025), p. 125598. ISSN: 1359-4311. DOI: <https://doi.org/10.1016/j.applthermaleng.2025.125598>. URL: <https://www.sciencedirect.com/science/article/pii/S1359431125001899>.
- [96] Iraklis Varlamis et al. “Smart fusion of sensor data and human feedback for personalized energy-saving recommendations”. In: *Applied Energy* 305 (2022), p. 117775. ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2021.117775>. URL: <https://www.sciencedirect.com/science/article/pii/S0306261921011120>.
- [97] T Wang et al. “Gamification as a catalyst for energy efficiency in buildings: Behavioural design, engagement dynamics, and implementation pathways”. In: *Results in Engineering* 27 (2025). DOI: [10.1016/j.rineng.2025.106807](https://doi.org/10.1016/j.rineng.2025.106807). URL: <https://www.scopus.com/pages/publications/105014093270>.

- [98] T Wang et al. “Research on energy consumption in household sector: a comprehensive review based on bibliometric analysis”. In: *Frontiers in Energy Research* 11 (2023). DOI: [10.3389/fenrg.2023.1209290](https://doi.org/10.3389/fenrg.2023.1209290). URL: <https://www.scopus.com/pages/publications/85182814594>.
- [99] Dave Webb, Doina Olaru, and Christopher Rastin. “Unlocking motivation for energy saving: A study of German electricity consumer segments”. In: *Energy Research & Social Science* 125 (2025), p. 104082. ISSN: 2214-6296. DOI: <https://doi.org/10.1016/j.erss.2025.104082>. URL: <https://www.sciencedirect.com/science/article/pii/S221462962500163X>.
- [100] Devon Wemyss et al. “Does it last? Long-term impacts of an app-based behavior change intervention on household electricity savings in Switzerland”. In: *Energy Research Social Science* 47 (2019), pp. 16–27. ISSN: 2214-6296. DOI: <https://doi.org/10.1016/j.erss.2018.08.018>. URL: <https://www.sciencedirect.com/science/article/pii/S2214629618302482>.
- [101] Georgina Wood et al. “Sensors, sense-making and sensitivities: UK household experiences with a feedback display on energy consumption and indoor environmental conditions”. In: *Energy Research & Social Science* 55 (2019), pp. 93–105. ISSN: 2214-6296. DOI: <https://doi.org/10.1016/j.erss.2019.04.013>. URL: <https://www.sciencedirect.com/science/article/pii/S2214629618303359>.
- [102] Ruowei Xiao, Zhanwei Wu, and Juho Hamari. “Internet-of-Gamification: A Review of Literature on IoT-enabled Gamification for User Engagement”. In: *International Journal of Human-Computer Interaction* 38.12 (2022), pp. 1113–1137. DOI: [10.1080/10447318.2021.1990517](https://doi.org/10.1080/10447318.2021.1990517). URL: <https://doi.org/10.1080/10447318.2021.1990517>.
- [103] Yifang Xu et al. “A large language model-based platform for real-time building monitoring and occupant interaction”. In: *Journal of Building Engineering* 100 (2025), p. 111488. ISSN: 2352-7102. DOI: <https://doi.org/10.1016/j.jobe.2024.111488>. URL: <https://www.sciencedirect.com/science/article/pii/S2352710224030560>.
- [104] Yizhou Yang, Qiuhua Duan, and Forooza Samadi. “A systematic review of building energy performance forecasting approaches”. In: *Renewable and Sustainable Energy Reviews* 223 (2025), p. 116061. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2025.116061>. URL: <https://www.sciencedirect.com/science/article/pii/S1364032125007348>.
- [105] Ming Hock Yew, Alemayehu Molla, and Vanessa Cooper. “Behavioural and environmental sustainability determinants of residential energy management information systems use”. In: *Journal of Cleaner Production* 356 (2022), p. 131778. ISSN: 0959-6526. DOI: <https://doi.org/10.1016/j.jclepro.2022.131778>. URL: <https://www.sciencedirect.com/science/article/pii/S0959652622013907>.
- [106] P Zangheri, T Serrenho, and P Bertoldi. “Energy savings from feedback systems: A meta-studies’ review”. In: *Energies* 12 (19 2019). DOI: [10.3390/en12193788](https://doi.org/10.3390/en12193788). URL: <https://www.scopus.com/pages/publications/85077497582>.
- [107] Thomas Zaragoza, Adel Nouredine, and Ernesto Exposito. “A systematic mapping study on software-based feedback for energy consumption”. In: *Renewable and Sustainable Energy Reviews* 222 (2025), p. 115889. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2025.115889>. URL: <https://www.sciencedirect.com/science/article/pii/S1364032125005623>.

- [108] Albin Zeqiri et al. “‘Eco Is Just Marketing’: Unraveling Everyday Barriers to the Adoption of Energy-Saving Features in Major Home Appliances”. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8 (1 Mar. 2024). DOI: [10.1145/3643558](https://doi.org/10.1145/3643558). URL: <https://doi.org/10.1145/3643558>.
- [109] Chaoqun Zhang et al. “The effect of customized information feedback on individual electricity saving behavior: Evidence from a field experiment in China”. In: *Technological Forecasting and Social Change* 193 (2023), p. 122602. ISSN: 0040-1625. DOI: <https://doi.org/10.1016/j.techfore.2023.122602>. URL: <https://www.sciencedirect.com/science/article/pii/S0040162523002871>.
- [110] Wenhao Zhang, Matias Quintana, and Clayton Miller. “Recommender systems and reinforcement learning for human-building interaction and context aware support: A text mining-driven review of scientific literature”. In: *Energy and Buildings* 329 (2025), p. 115247. ISSN: 0378-7788. DOI: <https://doi.org/10.1016/j.enbuild.2024.115247>. URL: <https://www.sciencedirect.com/science/article/pii/S037877882401363X>.
- [111] Yixiang Zhang et al. “Employee energy-saving behaviors: Review of theories, influencing factors, and interventions”. In: *Renewable and Sustainable Energy Reviews* 203 (2024), p. 114766. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2024.114766>. URL: <https://www.sciencedirect.com/science/article/pii/S1364032124004921>.

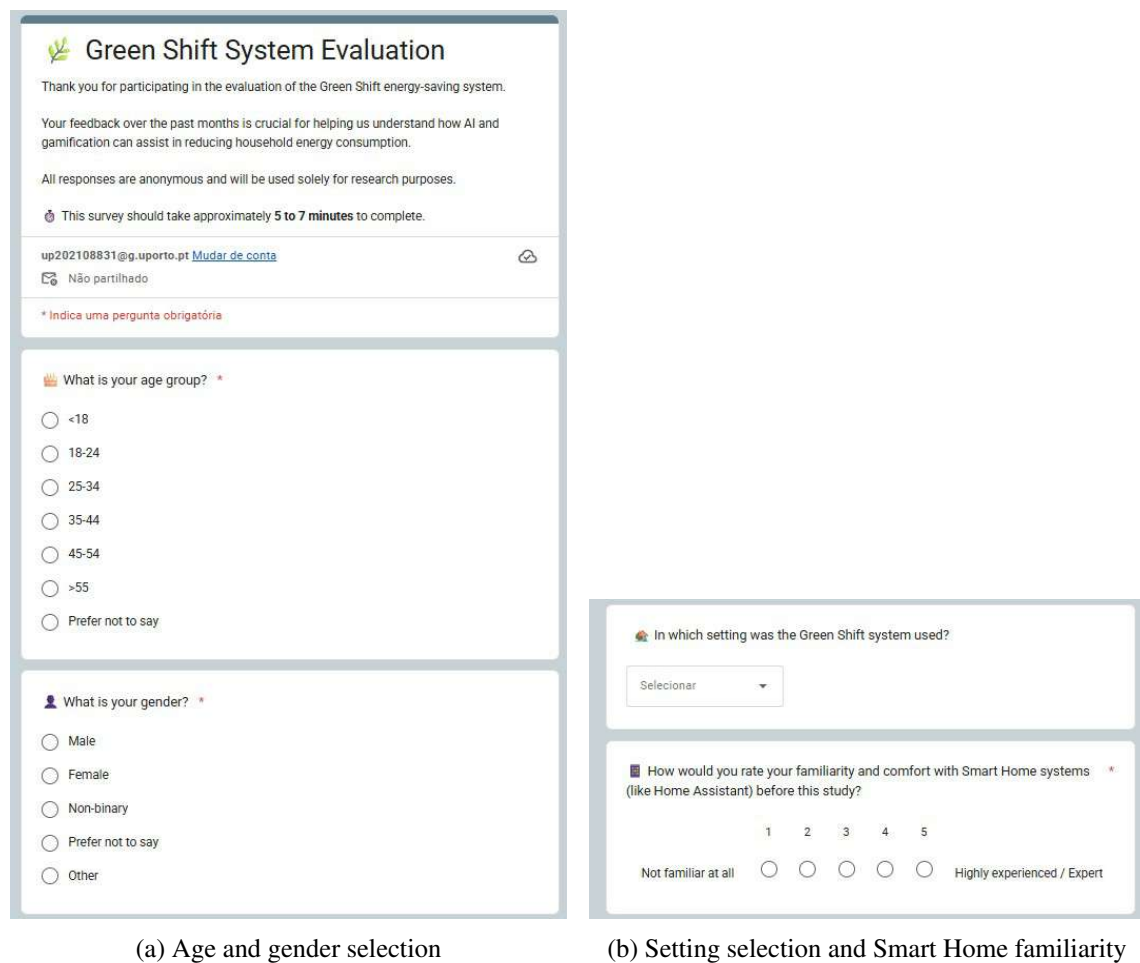
Appendix A

Qualitative User Assessment Questionnaire

The following pages present the complete set of forms and questions administered to the participants during the final qualitative evaluation. The survey was designed to assess baseline demographics, system usability, intrinsic motivation, cognitive load and long-term habit formation.

A.1 Demographics and Context

The initial section of the survey collected basic demographic data, the specific deployment environment, and the participant's baseline familiarity with Smart Home systems.



Green Shift System Evaluation

Thank you for participating in the evaluation of the Green Shift energy-saving system.

Your feedback over the past months is crucial for helping us understand how AI and gamification can assist in reducing household energy consumption.

All responses are anonymous and will be used solely for research purposes.

🕒 This survey should take approximately **5 to 7 minutes** to complete.

up202108831@g.uporto.pt [Mudar de conta](#)

🔒 Não partilhado

* Indica uma pergunta obrigatória

👤 What is your age group? *

☐ <18

☐ 18-24

☐ 25-34

☐ 35-44

☐ 45-54

☐ >55

☐ Prefer not to say

👤 What is your gender? *

☐ Male

☐ Female

☐ Non-binary

☐ Prefer not to say

☐ Other

🏠 In which setting was the Green Shift system used?

Selecionar

📊 How would you rate your familiarity and comfort with Smart Home systems (like Home Assistant) before this study? *

1 2 3 4 5

Not familiar at all ☐ ☐ ☐ ☐ ☐ Highly experienced / Expert

(a) Age and gender selection

(b) Setting selection and Smart Home familiarity

Figure A.1: Survey interface for demographic and baseline literacy data collection.

A.2 System Usability Scale (SUS)

To evaluate the UI/UX design and overall accessibility of the developed GreenShift custom component, the standard 10-item System Usability Scale was utilized.

⚙ Usability Rating *

1 (Strongly Disagree)

2

3

4

5 (Strongly Agree)

I think that I would like to use this system frequently.

☐

☐

☐

☐

☐

I found the system unnecessarily complex.

☐

☐

☐

☐

☐

I thought the system was easy to use.

☐

☐

☐

☐

☐

I think that I would need the support of a technical person to be able to use this system.

☐

☐

☐

☐

☐

I found the various functions in this system were well integrated.

☐

☐

☐

☐

☐

I thought there was too much inconsistency in this system.

☐

☐

☐

☐

☐

I would imagine that most people would learn to use this system very quickly.

☐

☐

☐

☐

☐

I found the system very cumbersome to use.

☐

☐

☐

☐

☐

I felt very confident using the system.

☐

☐

☐

☐

☐

I needed to learn a lot of things before I could get going with this system.

☐

☐

☐

☐

☐

(a) SUS Questionnaire (Part 1)

(b) SUS Questionnaire (Part 2)

Figure A.2: Standardized SUS forms presented to users.

A.3 Gamification and Intrinsic Motivation

This section assessed the effectiveness of the gamified mechanics using subscales derived from the IMI, focusing on interest, enjoyment and perceived value.

Gamification and Motivation

The following questions ask about your experience with the daily tasks and weekly challenges. Please read each statement and indicate how true it is for you.
(Scale for all questions: 1 = Not at all true to 7 = Very true)

Gamification and Tasks *

	1 (Not at all true)	2	3	4 (Somewhat true)	5	6	7 (Very true)
I enjoyed completing the daily energy-saving tasks very much.	☐	☐	☐	☐	☐	☐	☐
The weekly challenges were fun to participate in.	☐	☐	☐	☐	☐	☐	☐
I thought the gamification features (like streaks and tasks) were quite enjoyable.	☐	☐	☐	☐	☐	☐	☐
I would describe the energy-saving interventions as very interesting.	☐	☐	☐	☐	☐	☐	☐

I believe completing these daily tasks was of some value to me.	☐	☐	☐	☐	☐	☐	☐
I think that doing the weekly challenges is useful for improving my daily energy consumption habits.	☐	☐	☐	☐	☐	☐	☐
I believe this system could be of some value to help me save energy in the long term.	☐	☐	☐	☐	☐	☐	☐
I felt that doing these tasks was beneficial for my household.	☐	☐	☐	☐	☐	☐	☐

(a) Gamification Assessment (Part 1)
(b) Gamification Assessment (Part 2)

Figure A.3: Survey questions evaluating the gamification mechanics and intrinsic motivation.

A.4 Intervention Impact and Cognitive Load

To quantify the socio-technical boundaries of the AI agent, participants were asked to rate the intrusiveness of the notifications and their overall feeling of intervention fatigue.

Perceived Fatigue and AI Intrusiveness

The Green Shift system uses an AI agent to send timely notifications. We want to understand how these notifications impacted your daily routine.
(Scale for questions: 1 = Strongly Disagree to 5 = Strongly Agree)

Notification Impact *

	1 (Strongly Disagree)	2	3	4	5 (Strongly Agree)
The total number of energy-saving notifications I received per day was overwhelming.	☐	☐	☐	☐	☐
I felt annoyed by how often the AI suggested actions or tasks.	☐	☐	☐	☐	☐
The system frequently sent me notifications at inappropriate, highly disruptive, or inconvenient times.	☐	☐	☐	☐	☐
I felt that the AI learned to respect my schedule and tolerance for interruptions over time.	☐	☐	☐	☐	☐
I felt pressured or coerced by the AI's notifications rather than gently encouraged.	☐	☐	☐	☐	☐

Overall Fatigue: How often did you experience "notification fatigue" (the desire to ignore or disable the alerts) while using this system?

- ☐ Never
- ☐ Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

(a) Notification Impact

(b) Overall Fatigue Assessment

Figure A.4: Questionnaire items measuring notification intrusiveness and user fatigue.

A.5 Behavioural Shifts and System Trust

The final Likert-scale sections evaluated the agent's ability to foster lasting energy-saving routines and the degree to which users trusted the AI's autonomous decision-making.

Behavior Change & AI Trust

We are interested in how the AI influenced your habits and whether its decision-making made sense to you.

(Scale for questions: 1 = Strongly Disagree to 5 = Strongly Agree)

Habit Formation

1 (Strongly Disagree)2345 (Strongly Agree)

Because of this system, I am now more conscious of my energy usage even when I am not looking at the app.

I have continued to perform some energy-saving actions even when not explicitly prompted by a task.

The weekly challenges helped me build new, long-term habits regarding energy consumption.

I discuss energy saving more frequently with other members of my household now than I did 3 months ago.

System Trust

1 (Strongly Disagree)2345 (Strongly Agree)

I usually understood exactly why the AI was suggesting a specific energy-saving action at that moment.

The AI's recommendations felt directly relevant to my household's current context (e.g., time of day, occupancy).

I trusted the AI to make accurate suggestions regarding my energy consumption.

The system provided enough information for me to understand how it calculated my task difficulty and rewards.

(a) Habit Formation

(b) System Trust

Figure A.5: Survey interfaces assessing long-term habit formation and perceived AI trustworthiness.

A.6 Open-Ended Feedback

To capture qualitative nuances and usability bottlenecks not reflected in the scaled metrics, users were provided with an optional free-text field to detail their experience.

Open-Ended Feedback 💡

Please share any additional thoughts you have about your experience.
All answers in this section are entirely **optional**.

(You may answer both in English or Portuguese)

✍️ If you could change one thing about how the AI interacts with you, what would it be?

A sua resposta

🗨️ Do you have any specific examples of a time the AI gave a highly useful suggestion, or a highly inappropriate one?

A sua resposta

🗨️ Is there any other feedback regarding your experience over the past months that you would like to share with the research team?

A sua resposta

Figure A.6: Form section for unstructured open-ended feedback.

Appendix B

Scientific Article

As a formal dissemination of this research, the complete scientific manuscript derived from this study is provided below. The article was submitted to the CIoT 2026: 9th Conference on Cloud and Internet of Things.

Gamifying Sustainability: Harnessing AI and IoT for Human-Centric Energy Efficiency

Rafael Neves Teixeira
Faculty of Engineering,
University of Porto
Porto, Portugal
up202108831@fe.up.pt

Bruno Lima
LIACC and Faculty of Engineering,
University of Porto
Porto, Portugal
brunolima@fe.up.pt

Filipe Sousa
Fraunhofer Portugal AICOS
Porto, Portugal
filipe.sousa@acos.fraunhofer.pt

Abstract—Addressing climate change requires sustained improvements in building energy efficiency. While Internet of Things provides granular real-time energy monitoring, presenting raw data through static dashboards rarely fosters lasting behavioural change, a disconnect known as the intention-action gap. Current solutions frequently suffer from a “one-size-fits-all” approach, ultimately leading to user disengagement and notification fatigue. To bridge this gap, this paper introduces GreenShift, a privacy-preserving edge-computing architecture developed as a custom Home Assistant component. GreenShift transforms passive energy monitoring into a user-centric experience by integrating real-time eco-feedback, meaningful gamification and Artificial Intelligence (AI)-driven personalization. The system employs Reinforcement Learning from Human Feedback to dynamically tailor energy interventions to each user’s specific profile. By actively modelling a user’s Fatigue Index, the AI agent intelligently throttles notifications to prevent burnout while maximizing the relevance of actionable alerts. Empirically validated across four real-world environments, the architecture achieved up to a 69.8% reduction in monitored baseline loads within localized setups and successfully stabilized weather-normalized consumption. Furthermore, by autonomously blocking approximately 60% of generated notifications, the system effectively prevented user fatigue. The results demonstrate that personalized context-aware gamification outperforms traditional static eco-feedback, embedding sustainable habits into daily routines.

Index Terms—Eco-feedback, Edge Computing, Energy Efficiency, Gamification, Internet of Things, Reinforcement Learning

I. INTRODUCTION

The escalating urgency of climate change, coupled with the rising demand for sustainable energy practices, has intensified the pursuit of solutions that promote efficiency in both household and industrial domains [1]. At the heart of these efforts lies energy efficiency, a central strategy directly aligned with the United Nations Sustainable Development Goals¹. End-user behaviour is frequently identified as the “invisible” variable, potentially influencing up to 50% of a building’s energy consumption [2], [3]. Empirical studies indicate that distinct habits and unconscious routines can cause electricity consumption variations of up to 600% in otherwise identical housing units [4].

While the proliferation of Internet of Things (IoT) technologies has enabled sophisticated highly granular energy

monitoring [5], a critical challenge remains: providing users with raw data rarely leads to lasting behavioural change. Evidence from behavioural science demonstrates that without active engagement, the energy-saving potential of IoT systems is largely unrealized [6]. Traditional eco-feedback platforms face a persistent intention-action gap, where high motivation fails to materialize into behaviour, showing a persistent 30-40% discrepancy between users’ stated intentions and actual actions [7].

Dashboards frequently present complex data in abstract formats, such as raw kilowatt-hours (kWh), leading to cognitive overload and disengagement [8]. Furthermore, current gamification attempts often rely on rigid one-size-fits-all reward structures that fail to account for the heterogeneity of user routines [9]. This inevitably results in reward fatigue and, in some cases, a “boomerang effect” where initially efficient users increase their consumption to match a newly presented normative group average [10]. Furthermore, many AI-driven personalized systems rely heavily on synthetic datasets to pre-train models, failing to capture the chaotic variability of real human behaviour and suffering from the “cold start” problem when deployed in novel environments [11].

To systematically address these limitations, this study proposes *GreenShift*, a novel sociotechnical edge-computing architecture. GreenShift is developed as a native custom component for Home Assistant² (HA). HA was deliberately selected as the foundational platform due to its open-source nature, unparalleled interoperability across heterogeneous IoT protocols and its strict adherence to local data sovereignty, which completely eliminates the privacy vulnerabilities associated with cloud-centric alternatives. Furthermore, HA was chosen because it was already deeply integrated into the private and professional environments of the majority of the participants, thereby minimizing deployment friction and leveraging their existing technical familiarity. By building directly atop this platform, GreenShift integrates real-time IoT telemetry, adaptive gamification and an embedded Reinforcement Learning (RL) agent operating entirely on the edge.

This work contributes to the current literature in three significant ways. First, we introduce a sociotechnical edge

¹<https://sdgs.un.org/goals>

²<https://www.home-assistant.io>

architecture that synergizes real-time IoT telemetry, adaptive gamification and an embedded RL from Human Feedback (RLHF) agent directly at the edge to solve engagement decay. Second, we provide a rigorous real-world empirical validation across four distinct operational environments, spanning shared professional laboratories and private residential smart homes, thereby overcoming the inherent limitations of simulated datasets. By extending the validation period up to two months in specific environments, this study surpasses traditional short-term evaluations to explicitly address the “novelty effect” [3]. This prolonged observation proves that dynamic context-aware interventions can achieve up to a 69.8% reduction in targeted baseloads while autonomously suppressing 60% of notifications to successfully mitigate user fatigue. Finally, to bridge the gap between academic research and practical deployment, the GreenShift architecture has been publicly released as a fully functional open-source plugin on GitHub³ (<https://github.com/RafaelNTeixeira/GreenShift>), with formal integration into the HA Community Store⁴ (HACS) currently pending. This democratizes access to advanced behavioural energy tools and ensures broad reproducibility.

The remainder of this paper is organized as follows. Section II details the GreenShift methodology and system architecture, focusing on the RL agent and the adaptive gamification engine. Section III presents the empirical evaluation, discussing the experimental setup, quantitative energy reductions and qualitative user psychometrics. Section IV reviews the related work in eco-feedback, gamification and AI-driven personalization, contextualizing our contributions. Section V summarizes the main conclusions and outlines directions for future research.

II. GREENSHIFT: METHODOLOGY AND ARCHITECTURE

To ensure interoperability, privacy and low-latency decision-making while adhering to the core design principles of sociotechnical energy management systems [12], the proposed architecture is designed as a custom integration within HA, running entirely on local edge-computing nodes (e.g., Raspberry Pi). This ensures that sensitive occupancy and behavioural data never leave the premises [13], [14]. As illustrated in Fig. 1, the system consists of four primary layers: the Physical Layer (IoT sensors), the Core Platform (HA state machine), the Intelligence Layer (RL agent and gamification engine) and the User Interface Layer (user interaction).

A. Human-in-the-Loop Interface and Contextual Filtering

As illustrated in Fig. 2, the User Interface Layer leverages HA’s Lovelace⁵ framework to deliver a highly accessible and responsive dashboard across both mobile and desktop clients. Crucially, this interface serves as the primary conduit for the RLHF loop.

The strategic adoption of RLHF over conventional RL constitutes a fundamental architectural decision. Traditional

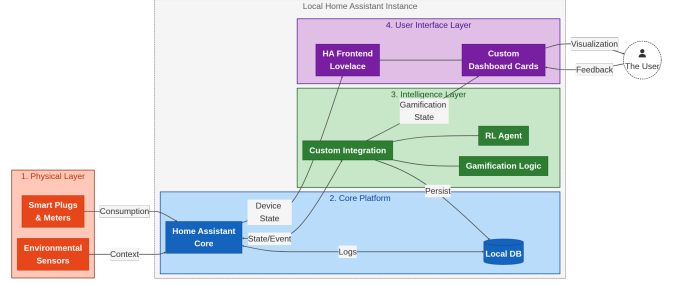


Fig. 1. High-level system architecture of GreenShift, illustrating the continuous data flow from physical sensors through the HA Core to the edge-based AI Intelligence Layer and culminating in user interaction via the interface.

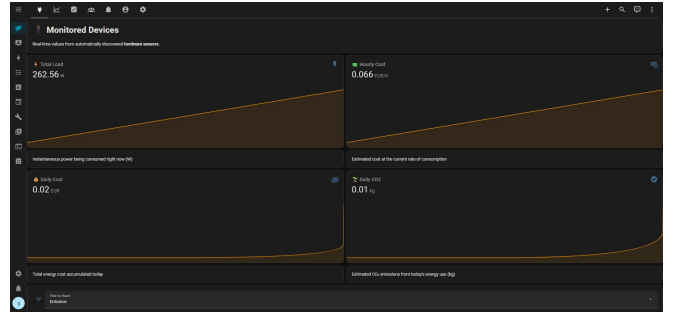


Fig. 2. The GreenShift dashboard interface, featuring interactive modules for gamified task evaluation and AI notification feedback, which continuously drive the RLHF loop.

RL agents optimize their policies based on rigid mathematically predefined reward functions (e.g., strictly maximizing kilowatt-hour reductions). However, such models frequently fail in sociotechnical environments because human comfort, behavioural tolerance and context are highly subjective and notoriously difficult to quantify mathematically. RLHF bridges this gap by treating the human occupant as an active evaluator [15]. Users interact directly with AI-generated notifications within the dashboard, explicitly classifying them as “Accepted” or “Rejected”. This categorical feedback supplies the critical empirical reward signal required by the agent. By learning directly from these interactions, the dynamic loop aligns the system’s policy not merely with technical efficiency, but with the complex, subjective and changing preferences of the human user [15], effectively preventing the rejection of the system.

Concurrently, users evaluate daily gamified tasks by rating them as “Too Easy”, “Just Right” or “Too Hard”. This direct input empowers the system to continuously calibrate subsequent task difficulty, ensuring sustained engagement by preventing both user frustration (goals that are too ambitious) and boredom (goals that lack challenge).

To ensure interventions remain highly actionable, the system applies contextual temporal filtering based on the deployment setting. In professional environments, the AI strictly filters data and schedules interventions according to custom-defined working hours and native holiday detection. This prevents

³<https://github.com/>

⁴<https://hacs.xyz/>

⁵<https://www.home-assistant.io/dashboards/>

the system from generating impossible tasks or learning from periods where users have no agency (such as nighttime standby loads or weekends). In residential setups, this filter is disabled to reflect the continuous nature of domestic energy management.

B. State and Action Space Formulation

The RL agent utilizes a Q-learning algorithm engineered to handle heterogeneous physical infrastructures. To accommodate environments with missing sensors, the continuous State Space (S_t) (Eq. 1) utilizes a "Zero-Padding with Existence Flags" strategy:

$$S_t = [(P_{total}, F_{total}), (P_{app}, F_{app}), (T_{in}, F_T), (H_{in}, F_H), (L_{in}, F_L), (O_{status}, F_O), I_{anomaly}, I_{behaviour}, I_{fatigue}, C_{area_anomalies}, T_{day}, D_{week}] \quad (1)$$

In this representation, P_{total} and P_{app} denote total facility and peak appliance power, respectively. Contextual variables including T_{in} (temperature), H_{in} (humidity), L_{in} (luminosity) and O_{status} (occupancy) capture the environment. Crucially, each metric is paired with a binary availability flag ($F = 1$ if present, 0 if absent) to seamlessly neutralize missing hardware data during inference. High-level indices model the human element: $I_{anomaly}$ flags irregular usage, $I_{behaviour}$ tracks compliance and $I_{fatigue}$ quantifies desensitization risk. Finally, to maintain memory efficiency, this continuous space is discretized into a finite set of states.

The Action Space (A) (Eq. 2) defines the maximum set of potential interventions:

$$A = \{a_{noop}, a_{specific}, a_{anomaly}, a_{behavioural}, a_{normative}\} \quad (2)$$

where a_{noop} maintains silence to prevent annoyance, $a_{specific}$ triggers targeted appliance commands, $a_{anomaly}$ alerts the user to irregular consumption, $a_{behavioural}$ suggests general lifestyle adjustments and $a_{normative}$ leverages historical comparisons.

To ensure robustness across different hardware tiers, the agent applies an Action Validity Mask (M). This dynamic boolean vector strictly filters out impossible commands. For example, if individual appliance-level smart plugs are absent, targeted device recommendations ($a_{specific}$) are masked out.

C. Q-Learning and Dual-Phase Reward Architecture

The decision agent optimizes its policy by learning a state-action value function, $Q(s, a)$, which estimates the expected cumulative reward for taking intervention a in context s [16]. The Q-table is iteratively updated according to the Bellman equation (Eq. 3):

$$Q(s, a) \leftarrow Q(s, a) + \alpha [R + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (3)$$

where α is the learning rate (which adapts depending on the deployment phase), R is the observed reward and γ is the discount factor for future rewards.

To directly combat the "cold start" problem and avoid relying on inaccurate synthetic datasets [11], the agent's deployment is split into two phases. Phase 1 (Shadow Learning) observes the environment silently for 14 days. During this period, the agent simulates decisions without sending notifications. Because there is no human interaction, the shadow reward (R_{shadow}) (Eq. 4) is strictly an estimation based on environmental alignment:

$$R_{shadow} = 0.5 \cdot S_{potential} + 0.3 \cdot C_{alignment} + 0.2 \cdot T_{score} \quad (4)$$

To prevent the Q-table from over-committing to these estimated rewards, the update mechanism utilizes a significantly lower learning rate ($\alpha = 0.05$) compared to the active phase ($\alpha = 0.1$). Furthermore, during this phase, the discount factor (γ) is forced to 0, preventing the artificial inflation of Q-values in a continuous feedback loop without empirical validation.

Phase 2 (Active Intervention) introduces gamified tasks and actual AI-generated nudges. The reward function (R_t) (Eq. 5) now integrates real-time RLHF:

$$R_t = \Delta P + 0.5 \cdot I_{feedback} - 0.3 \cdot I_{fatigue} \quad (5)$$

where ΔP represents actual measured power savings, $I_{feedback}$ is derived from the user's explicit response via the UI and $I_{fatigue}$ penalizes over-notification. The discount factor (γ) becomes dynamic: if a user accepts a nudge, γ uses its standard future-value estimation. However, if a user explicitly rejects a nudge, γ is forced to 0.0. This treats the rejection as a terminal state, heavily penalizing the action and ensuring the AI rapidly learns to stop annoying the user [15].

D. Adaptive Cooldown and Gamification Engine

To mitigate user desensitization, the system calculates an Opportunity Score ($O_t \in [0, 1]$) (Eq. 6) before generating an intervention:

$$O_t = 0.35 S_{potential} + 0.35 U_{urgency} + 0.20 R_{receptiveness} + 0.10 C_{context} \quad (6)$$

where $S_{potential}$ represents the real-time power deviation from the established baseline, $U_{urgency}$ quantifies the severity of detected spatial anomalies, $R_{receptiveness}$ reflects the user's current tolerance derived from the inverse of the fatigue index and $C_{context}$ evaluates the alignment of temporal and occupancy states. This composite score dynamically modulates a base cooldown timer. If an opportunity is deemed "Critical" ($O_t \geq 0.8$), the standard adaptive cooldown is bypassed.

Complementing the AI, a rule-based gamification engine dynamically adapts the difficulty of generated daily tasks (using multipliers from $0.5\times$ to $1.5\times$). Users can also explicitly rate tasks as "Too Easy" or "Too Hard", ensuring the difficulty scales to maintain psychological self-efficacy [17], [18].

E. Reliability, Multilingual Support and Testing

To ensure broad accessibility and continuous operation, GreenShift features a robust multi-layered data protection suite. This includes Write-Ahead Logging for SQLite databases to prevent data loss during power outages, alongside

automated lifecycle snapshots triggered during system startup and shutdown. Furthermore, the UI and behavioural nudges are fully localized in both English and Portuguese, ensuring interventions are clearly understood by the target demographic. To validate the reliability of this complex architecture, an exhaustive testing suite comprising 1,198 tests was developed, achieving 100% code coverage across all core modules, including the RL logic, state persistence and gamification mechanics.

III. EMPIRICAL EVALUATION

To rigorously evaluate the real-world efficacy and adaptability of the GreenShift architecture, a comprehensive empirical validation was conducted. This section details the experimental framework, transitioning from the theoretical design to practical deployment across diverse operational environments. By triangulating quantitative physical telemetry, such as weather-normalized energy consumption and algorithmic convergence metrics, with qualitative psychometric data, the subsequent analysis assesses the platform's capacity to bridge the intention-action gap, mitigate notification fatigue and sustain long-term energy-efficient behaviours.

A. Research Questions

To systematically evaluate the proposed GreenShift architecture, the experimental validation was designed to address four specific research questions: **(RQ1)** How effective are traditional eco-feedback strategies in promoting energy-efficient behaviour in building contexts? **(RQ2)** To what extent does the integration of meaningful gamification with real-time feedback lead to a greater reduction in energy consumption and sustained user engagement compared to traditional eco-feedback methods? **(RQ3)** Which gamification strategies (e.g., specific tasks, weekly challenges, normative comparison) are most effective in sustaining user engagement within IoT energy systems? **(RQ4)** To what extent does AI-driven personalization provide an additive increase in energy reduction and sustained user engagement by mitigating notification fatigue?

B. Experimental Setup and Phased Procedure

The architecture was deployed across four real-world settings to validate its hardware-agnostic design: two professional laboratories (Fraunhofer AICOS, featuring shared infrastructure; and FEUP, featuring localized appliance monitoring) and two residential smart homes (Home 1 and Home 2, integrating both whole-house metering and appliance-level sensors). The study involved a total of 13 active participants across these environments, characterized by diverse ages (18-54 years old) and generally high technological literacy, of which 8 completed the final psychometric evaluation. Furthermore, while the maximum continuous validation period spanned up to two months, deployment durations varied by environment. This extended timeline explicitly addresses the "novelty effect" bias by pushing beyond the short-term evaluations typical of most existing studies [3].

As an example of the deployed edge topology, Fig. 3 illustrates the hardware setup utilized in the Fraunhofer laboratory.

It relies on a local Raspberry Pi edge gateway communicating via Bluetooth Low Energy (BLE) with custom environmental sensors, while utilizing Wi-Fi to integrate a Shelly Pro 3EM whole-facility energy meter and a presence sensor, completely bypassing cloud-based analytics.

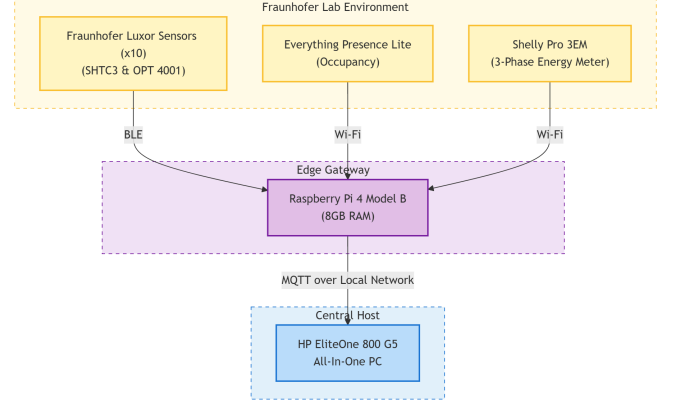


Fig. 3. Hardware topology deployed in the Fraunhofer laboratory, illustrating the distributed edge network for power monitoring and contextual sensing.

The experimental procedure followed a strict multi-phase timeline to ensure the AI model was grounded in empirical data before executing active interventions:

- **Phase 0 (1 month) - Development & Data Acquisition:** Passive logging of power consumption and environmental metadata to create a synchronized dataset for the offline evaluation of the RL agent. Concurrently, the core integration logic of GreenShift was developed.
- **Phase 1 (14 days) - Silent Calibration:** The system was deployed in "Silent Mode". Users had access to a basic visualization dashboard (control period), while the RL agent refined consumption baselines in the background without issuing interventions.
- **Phase 2 (Active Period) - Adaptation and RLHF:** Activation of the gamification engine and proactive AI nudges. The agent's policy was refined in real-time through RLHF, balancing energy savings against the user fatigue index.

C. Energy Reduction and Gamification Efficacy

To understand the temporal trajectory of these behavioural shifts and evaluate the standalone impact of the visualization dashboard, Fig. 4 presents the interrupted time-series evolution of weather-normalized consumption by Degree Days (E_{norm}) specifically for the FEUP laboratory.

During Phase 1 (Silent Calibration), the provision of passive real-time data yielded highly volatile results. In localized setups (e.g., the FEUP laboratory), occupants experienced an immediate psychological reaction: the sudden awareness of their high energy footprint prompted a sharp initial decrease in usage. However, without active reinforcement, passive visualization often struggles to maintain these drops long-term. In fact, passive monitoring failed to prevent progressive consumption increases over time in some residential settings (e.g.,

Home 1), and was largely ignored in shared infrastructural spaces (e.g., Fraunhofer) where individual agency is diluted.

The transition to Phase 2 (Active AI Gamification) successfully locked in and deepened these initial efficiency gains. In the localized laboratory setup, the overall E_{norm} dropped by 89.2% from the baseline, approaching statistical significance ($p = 0.065$) with a massive practical effect size (Cohen's $d = 1.223$). In residential setups, active gamification effectively mitigated thermal demand fluctuations, reducing raw power consumption by up to 15.1% in heavily loaded homes. Conversely, in the large Fraunhofer facility, where users lacked direct authority over the central HVAC, the average power shifted negligibly (-0.3%), proving that both visualization and gamification require physical user agency to succeed.

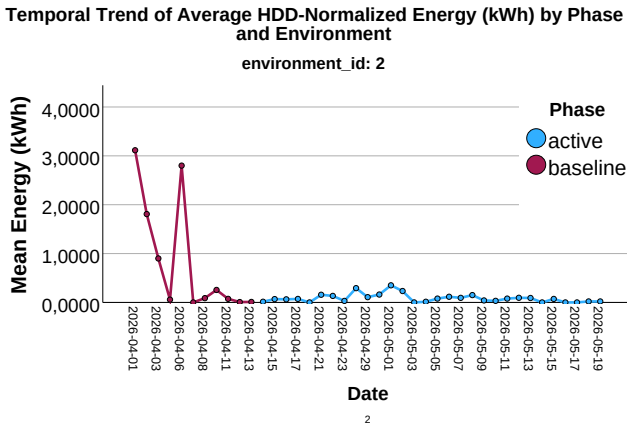


Fig. 4. Interrupted time-series analysis detailing the daily evolution of weather-normalized energy consumption (E_{norm}) at the FEUP laboratory, illustrating the high volatility of passive visualization (Phase 1) versus the stabilized reductions achieved during active gamification (Phase 2).

To isolate the specific impact of the gamified mechanics from general system awareness, an intra-phase analysis of the active period was conducted. Days clustered as "Active" (where at least one gamified task was completed) registered consistently lower E_{norm} than "Passive" days (where tasks were uncompleted/ignored). Furthermore, Table I illustrates the average instantaneous power savings registered during successfully completed daily tasks. Because these daily tasks dynamically adapted their difficulty based on real-time feedback, they fostered high completion rates (up to 71.4%) and sustained engagement, particularly in localized setups. The completion of these dynamically generated tasks eliminated up to 53.0% of the active baseline in the localized lab and 69.8% in Home 2. This provides empirical proof that interactive goal-oriented feedback vastly outperforms static reporting by creating a verifiable wedge against energy waste.

In addition to daily tasks, the system deployed cooperative weekly challenges featuring a default 15% reduction target, which users retained the freedom to manually adjust in the configuration settings. As detailed in Table II, the results revealed that static macroscopic targets are inherently flawed

TABLE I
AVERAGE INSTANTANEOUS TASK SAVINGS VS. BASELOAD

Environment	Baseload (W)	Savings (W)	Impact
Fraunhofer	6968.9	590.0	8.5%
FEUP	20.4	10.8	53.0%
Home 1	1295.2	263.6	20.4%
Home 2	321.8	224.8	69.8%

for heterogeneous environments. A fixed 15% reduction was virtually unattainable in the shared Fraunhofer facility, where massive infrastructural loads diluted individual efforts, resulting in zero successful weeks. Conversely, users in the localized FEUP setup routinely cleared 35-50% reductions, making the static goal trivial. The residential environments further highlighted this flaw: while Home 1 successfully coordinated to hit the target in two of its five weeks, Home 2, which possessed an inherently low baseline (averaging only 321.8 W), struggled to carve out an additional 15% without compromising basic living standards, ultimately saving only 1.7% during its single active week. This sharp contrast demonstrates the critical importance of dynamic auto-scaling targets, such as the daily tasks, for the success of behavioural interventions, as rigid goals risk becoming either mathematically unattainable or trivial, ultimately leading to demotivation.

TABLE II
WEEKLY CHALLENGE OUTCOMES

Environment	Week	Savings (%)	Status
Fraunhofer	1	5.6%	Failed
	2	-1.3%	Failed
	3	-6.1%	Failed
	4	3.8%	Failed
	5	0.1%	Failed
FEUP	1	40.9%	Success
	2	0.2%	Failed
	3	45.2%	Success
	4	35.3%	Success
	5	53.0%	Success
Home 1	1	11.6%	Failed
	2	15.1%	Success
	3	6.7%	Failed
	4	11.1%	Failed
	5	19.9%	Success
Home 2	1	1.70%	Failed

D. Nudge Acceptance and Strategy Efficacy

While the rule-based gamification engine managed the adaptive daily tasks, the RL agent drove proactive user engagement through targeted notifications. Operating within its defined action space, the AI engine inherently generates and classifies these nudges into four distinct types: Anomaly, Behavioural, Normative and Specific. As illustrated in Fig. 5, analysing the acceptance rates of these notification strategies reveals critical context-dependent user preferences.

In the shared Fraunhofer facility, occupants exhibited exceptionally high acceptance rates for broad behavioural tips (96.1%) and normative nudges (100.0%). By focusing on cooperative historical comparisons rather than toxic competitive leaderboards, normative nudges successfully fostered

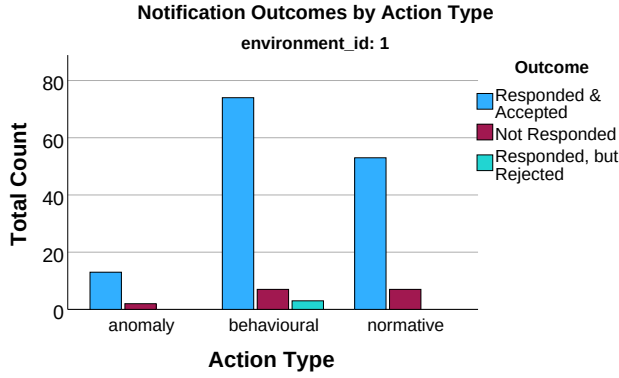


Fig. 5. Notification Outcomes by Action Type at Fraunhofer: Distribution of accepted, ignored and rejected nudges across different intervention strategies.

collective accountability without triggering workplace friction. Crucially, the "specific" targeted commands registered zero occurrences across all outcome categories. Because the Fraunhofer facility lacked appliance-level monitoring, the system's dynamic Action Validity Mask (M) successfully filtered out these impossible commands. This architectural feature prevented the AI from targeting shared infrastructure over which users had no direct physical authority, effectively shielding occupants from irrelevant prompts and potential frustration.

Conversely, in localized setups (FEUP and residential homes) where users possessed direct agency over the monitored appliances, specific appliance-level commands generally saw significantly higher engagement. However, the RLHF loop proved essential for adapting to unique household preferences. In Home 1, for instance, users largely ignored or explicitly rejected these specific nudges. As illustrated in Fig. 6, the "specific" action type experienced the highest rejection rate and was consequently generated far less frequently than other nudges. The AI agent successfully learned from this negative empirical feedback and autonomously suppressed the generation of "specific" interventions for that environment over time.

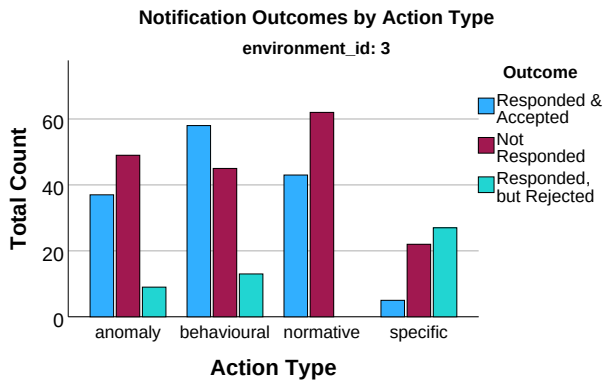


Fig. 6. Notification Outcomes by Action Type at Home 1: Distribution of accepted, ignored and rejected nudges, highlighting the high rejection rate and subsequent suppression of "specific" nudges.

Furthermore, a universal trend emerged across all environments regarding "anomaly" nudges: although these alerts were generated less frequently, as significant load deviations are exceptional events, they exhibited remarkably high acceptance and low rejection rates (e.g., 100.0% at Fraunhofer, 75.0% at FEUP, 80.4% in Home 1 and 100% in Home 2). This confirms that when the system detects a true immediate deviation, users immediately recognize the high relevance of the alert and are highly unlikely to dismiss it.

These findings prove that there is no universal one-size-fits-all gamification strategy. To successfully sustain long-term engagement, IoT systems must dynamically map intervention types to the user's level of physical control. Localized settings thrive on specific actionable device tasks, whereas shared infrastructural spaces must rely on cooperative normative social comparisons and anomaly detection [19].

E. AI Convergence and Mitigation of Fatigue

The RL agent successfully stabilized its decision-making policy. Utilizing an ϵ -greedy strategy, the agent shifted from exploration to exploitation as it gathered human feedback, averaging a global distribution of 8.9% exploration and 91.1% exploitation by the end of the deployments.

Crucially, the architecture successfully modelled and respected human boundaries. As shown in Fig. 7, by correlating the Opportunity Score against the Fatigue Index ($I_{fatigue}$), the AI autonomously suppressed (blocked) approximately 60% of all generated notification intents across every environment, preventing technical spam and sensory overload.

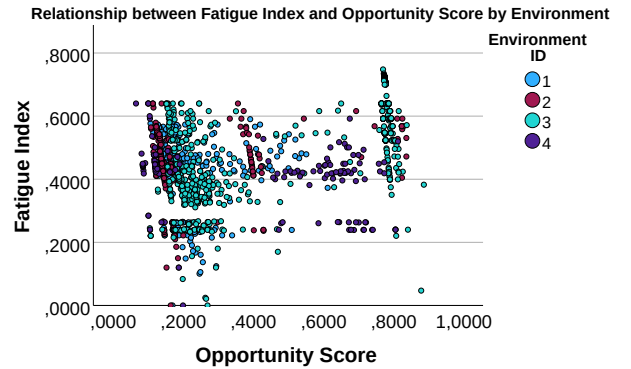


Fig. 7. Grouped scatter plot illustrating the distribution of suppressed (blocked) notifications, crossing the calculated AI opportunity score with the user fatigue index ($I_{fatigue}$) across environments.

F. Qualitative User Assessment and Psychometrics

While physical telemetry quantified structural energy reductions, post-intervention surveys contextualized the subjective user experience. To ensure that any lack of engagement was not due to poor interface design, the System Usability Scale (SUS) was administered. Users awarded GreenShift an exceptional overall SUS score of 90.0 out of 100, confirming the interface was highly intuitive.

Regarding gamified motivation, responses on the Intrinsic Motivation Inventory (IMI) subscales were overwhelmingly positive (averaging between 5.2 and 6.5 out of 7). This indicates strong alignment with the *Interest/Enjoyment* and *Value/Usefulness* dimensions, as the participants viewed the gamified tasks not as external coercion, but as practical and ecologically valuable tools.

Furthermore, evaluations of the autonomous RL agent revealed a high degree of trust. With scores consistently clustered around 4.3 on a 5-point scale, users perceived the AI's contextual nudges as highly transparent, accurate and relevant. Crucially, the reported fatigue aligned perfectly with physical agency: participants in localized setups (FEUP and residential homes) reported very low fatigue (1.5 to 2.0 out of 5), whereas those in shared infrastructure (Fraunhofer) reported moderate fatigue (3.0). This highlights that interventions are perceived as actionable guidance when users possess control, but lean toward burdensome when agency is lacking. Finally, participants reported a high likelihood of maintaining energy-saving routines without explicit prompts (scoring 4.0 to 4.6), indicating successful long-term habit formation.

G. Addressing the Research Questions

Based on the synthesized telemetry and psychometric data, the empirical findings yield the following conclusions:

RQ1: Traditional Eco-Feedback

Passive visualization is highly volatile and context-dependent. While it can act as a powerful initial catalyst in localized setups where users have direct control, it fails to prevent consumption increases over time in some residential settings and is largely ignored in shared infrastructural spaces where individual agency is diluted.

RQ2: Impact of Gamification

Active gamification adds significant quantifiable value on an intra-phase daily operational scale by locking in behaviour shifts and stabilizing consumption against erratic volatility. Days characterized by active task engagement consistently registered lower normalized consumption across all cohorts. Furthermore, at a micro-level, successfully completed daily tasks generated immediate verifiable power reductions, eliminating 53.0% and 69.8% of the baseload in the localized lab and Home 2, respectively. However, the macro phase-to-phase statistical analysis reveals that translating these localized daily gains into statistically significant long-term facility-wide reductions ($p > 0.05$ in most settings) remains heavily constrained by inelastic structural and thermal building demands.

RQ3: Efficacy of Specific Strategies

The efficacy of gamification is strictly dictated by user agency and environmental topology. In localized setups, specific appliance-targeted nudges demonstrate moderate compliance and users achieve high completion rates. Conversely, in shared professional spaces lacking granular monitoring, cooperative normative social comparisons and behavioural tips prove highly effective. Furthermore, anomaly alerts prove universally effective across all topologies due to their high immediate

relevance. Finally, dynamically scaling daily tasks maintain significantly higher long-term engagement than rigid static weekly targets, which often prove mathematically unattainable or demotivating.

RQ4: Role of AI Personalization

AI-driven personalization is the critical mechanism for preserving digital mental health and sustaining engagement. By autonomously evaluating real-time opportunity scores against a dynamic Fatigue Index, the RL agent successfully throttled interventions, blocking approximately 60% of generated notification intents. This self-throttling prevented user burnout, shielding occupants from technical spam while retaining exceptional system trust and high usability.

IV. RELATED WORK

While real-time IoT platforms and visual eco-feedback have largely replaced static utility bills, they frequently suffer from a persistent intention-action gap where short-term motivation fails to translate into sustained habits [7], [20], [21]. To combat this engagement decay, modern interventions heavily leverage gamification, utilizing frameworks like Self-Determination Theory or normative social comparisons to foster intrinsic motivation and accountability [18], [22], [23]. However, rigid "one-size-fits-all" mechanics often induce "reward fatigue" once the novelty wanes [3], or even trigger a "boomerang effect" where efficient users increase consumption to match a higher normative group average [10].

Recent literature explicitly calls for AI-driven adaptation to overcome this rigidity [24]. However, most intelligent solutions attempt to optimize energy efficiency autonomously, bypassing the occupant entirely and neglecting the vast potential of human behavioural shifts. While some recommender frameworks target waste "micro-moments" [13], [25], they predominantly rely on cloud processing, raising privacy concerns [26], and lack real-time human feedback. GreenShift extends the state of the art by explicitly keeping the human in the loop. By synergizing an interactive visualization tool and adaptive gamification with edge-based RLHF [14], [15], the architecture goes beyond mere automated control. It actively coaches the user to form sustainable habits, dynamically mapping interventions to physical building states and real-time user fatigue without compromising data sovereignty.

V. CONCLUSIONS AND FUTURE WORK

The transition to intelligent sustainable buildings cannot be achieved through algorithmic optimization and automated hardware alone: it requires the active willing collaboration of the occupants. This paper introduced GreenShift, a privacy-preserving edge-computing architecture that fuses the psychological drivers of meaningful gamification with the adaptive capabilities of RLHF.

Empirical validation across diverse real-world environments confirmed that while traditional eco-feedback generates volatile results, personalized and context-aware gamification effectively bridges the intention-action gap. The study demonstrated that intervention efficacy is strictly dictated

by physical user agency: localized environments thrive on active task engagement, yielding up to a 69.8% reduction in active baseloads, alongside specific appliance-level notifications, whereas shared infrastructural spaces respond best to cooperative normative comparisons. Furthermore, the RL agent proved highly capable of mitigating digital fatigue. By dynamically evaluating opportunity scores against user capacity, the system autonomously throttled approximately 60% of its interventions, preserving user well-being and maintaining an exceptional SUS score of 90.0.

The main contributions of this research include the development of a fully localized hardware-agnostic RL architecture that ensures data privacy at the edge, the empirical demonstration of dynamic gamification's superiority over static targets and the open-source release of the GreenShift integration on GitHub (<https://github.com/RafaelNTeixeira/GreenShift>) and submission to HACS to democratize access to human-centric energy tools.

To build upon the significant engagement achieved by the current architecture, future research can explore several promising avenues for enhancement. Transitioning static weekly challenges into fully dynamic AI-generated community milestones that scale based on cohort performance will further optimize collective motivation. To streamline the human-in-the-loop cycle and reduce timed-out notifications, future UI/UX designs should implement lock-screen interactive buttons in mobile devices. Furthermore, expanding the deployment into a shared professional environment fully equipped with comprehensive appliance-level monitoring will enable the evaluation of how occupants respond to highly specific granular nudges within communal spaces. Technologically, edge nodes can incorporate multi-modal presence detection (e.g., acoustic and BLE triangulation) to eliminate sensor noise from stationary occupants. Finally, extended longitudinal studies will help fully map long-term RL convergence and true habit endurance.

REFERENCES

- [1] G. Lima, A. Nascimento, M. Oliveira, and F. Dias, "Energy efficiency analysis: A household digital transformation," *AIMS Energy*, vol. 12, pp. 774–808, 01 2024.
- [2] W.-T. Li, O. Iuorio, H. Fang, and M. W. T. Mak, "Gamification approaches and assessment methodologies for occupants' energy behavior change in buildings: A systematic review," *Buildings*, vol. 14, 2024.
- [3] T. Wang, Z. Xin, X. Ma, S. Wang, D. S. Ametefe, D. John, and T. Darboe, "Gamification as a catalyst for energy efficiency in buildings: Behavioural design, engagement dynamics, and implementation pathways," *Results in Engineering*, vol. 27, 2025.
- [4] Y. Xu, S. Zhu, J. Cai, J. Chen, and S. Li, "A large language model-based platform for real-time building monitoring and occupant interaction," *Journal of Building Engineering*, vol. 100, p. 111488, 2025.
- [5] M. Dell'Isola, G. Ficco, L. Canale, B. I. Palella, and G. Puglisi, "An iot integrated tool to enhance user awareness on energy consumption in residential buildings," *Atmosphere*, vol. 10, no. 12, p. 743, 2019.
- [6] L. Li, X. Yuan, and J. Li, "Understanding visual feedback mechanism from three-dimensional of information, time and display: A meta-analysis of feedback research for household electricity conservation," *Energy and Buildings*, vol. 316, p. 114297, 2024.
- [7] Y. Chen, D. Sumiyoshi, Y. Morita, S. Ishibashi, X. Wang, T. Yamamoto, T. Ueno, and J. Oh, "Bi-tech: An iot-based behavioral intervention system for user-driven energy optimization in commercial spaces," *IEEE Access*, vol. 13, pp. 166 853–166 872, 2025.
- [8] A. Zeqiri, P. Jansen, J. O. Rixen, M. Rietzler, and E. Rukzio, "'eco is just marketing': Unraveling everyday barriers to the adoption of energy-saving features in major home appliances," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 8, 3 2024.
- [9] I. Adaji and M. Adisa, "A review of the use of persuasive technologies to influence sustainable behaviour," in *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, 2022, pp. 317–325.
- [10] M. Liu, M. Liang, B. Huang, Z. Liu, C. Fu, and Z. Li, "A preliminary investigation of the effectiveness of feedback messaging on household room air conditioning usage in summer," *Energy and Buildings*, vol. 318, p. 114443, 2024.
- [11] N. S. Dasappa, K. K. G., and N. Somu, "Multi-sensor data fusion framework for energy optimization in smart homes," *Renewable and Sustainable Energy Reviews*, vol. 193, p. 114235, 2024.
- [12] L. Andolfi, R. L. Baima, L. M. Burcheri, I. Pavić, and G. Fridgen, "Sociotechnical design of building energy management systems in the public sector: Five design principles," *Applied Energy*, vol. 377, p. 124628, 2025.
- [13] Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, I. Varlamis, G. Bravos, C. Sardianos, and G. Dimitrakopoulos, "Techno-economic assessment of building energy efficiency systems using behavioral change: A case study of an edge-based micro-moments solution," *Journal of Cleaner Production*, vol. 331, p. 129786, 2022.
- [14] A. N. Sayed, F. Bensaali, Y. Himeur, G. Dimitrakopoulos, and I. Varlamis, "Enhancing building sustainability: A digital twin approach to energy efficiency and occupancy monitoring," *Energy and Buildings*, vol. 328, p. 115151, 2025.
- [15] S. Chadoulos, I. Koutsopoulos, G. Polyzos, and N. Ipiotis, "Reinforcement learning with partially defined rewards and human feedback for energy efficiency recommendations," in *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, 2025, pp. 622–631.
- [16] F. Luo, G. Ranzi, and Z. Y. Dong, *Chapter 14 - Occupant-to-grid integration*. Elsevier, 2024, pp. 295–311.
- [17] Z. Chen, R.-I. Chang, and Q. Wu, "Adaptive smart system for energy-saving campus," *Engineering Proceedings*, vol. 92, 2025.
- [18] A. A. Mabalay, "Gamification for sustainability: A systematic review of applications, trends, and opportunities," *Computers in Human Behavior*, vol. 165, p. 108529, 2025.
- [19] Z. D. Tekler, R. Low, and L. Blessing, "User perceptions on the adoption of smart energy management systems in the workplace: Design and policy implications," *Energy Research & Social Science*, vol. 88, p. 102505, 2022.
- [20] P. Zangheri, T. Serrenho, and P. Bertoldi, "Energy savings from feedback systems: A meta-studies' review," *Energies*, vol. 12, 2019.
- [21] A. E. Onile, R. Machlev, E. Petlenkov, Y. Levron, and J. Belikov, "Uses of the digital twins concept for energy services, intelligent recommendation systems, and demand side management: A review," *Energy Reports*, vol. 7, pp. 997–1015, 2021.
- [22] P. Mochi, K. Pandya, K. B. Lindberg, and M. Korpås, "Social nudging for sustainable electricity use: Behavioral interventions in energy conservation policy," *Sustainability (Switzerland)*, vol. 17, 2025.
- [23] K. Trinh, A. S. Fung, and V. Straka, "Effects of real-time energy feedback and normative comparisons: Results from a multi-year field study in a multi-unit residential building," *Energy and Buildings*, vol. 250, p. 111288, 2021.
- [24] N. Abdurahmanovic and A. Cadenbach, "Enhancing energy efficiency through user engagement and behaviour change: A review on gamification approaches and serious games in energy systems," *Energy*, vol. 334, p. 137496, 2025.
- [25] P. Sitoula, D. Rahayu, P. D. Haghighi, S. Goodwin, and C. Ling, "Context-aware smart energy recommender (caser)," in *Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia*. Association for Computing Machinery, 2020, pp. 13–19.
- [26] L. N. Sieger, J. Hermann, A. Schomäcker, S. Heindorf, C. Meske, C.-C. Hey, and A. Doğançün, "User involvement in training smart home agents: Increasing perceived control and understanding," in *Proceedings of the 10th International Conference on Human-Agent Interaction*. Association for Computing Machinery, 2022, pp. 76–85.