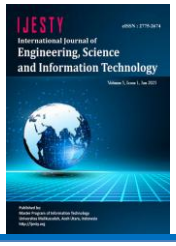
International Journal of Engineering, Science and Information Technology
Volume 5 No. 1 (2025) pp. 480-489
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i1.1288
Research Paper, Short Communication, Review, Technical Paper



Design of Ethical AI Frameworks for Sustainable and Adaptive Energy Management Systems

Mustafa Humadi[1], Haider Hadi Abbas[2], Hassan Waryoush Hilou[3], Nahlah. M. A. D. Najm[4], Ammar Abdulkhaleq Ali[5*], M. Batumalay[6]

[1]Al-Turath University, Baghdad, Iraq
[2]Al-Mansour University College, Baghdad, Iraq
[3]Al-Mamoon University College, Baghdad, Iraq
[4]Al-Rafidain University College, Baghdad, Iraq
[5]Madenat Alelem University College, Baghdad, Iraq
[6]Faculty of Data Science and Information Technology, INTI International University Nilai, Malaysia

**Corresponding author Email: ammar.ali@mauc.edu.iq*

The manuscript was received on 17 June 2024, revised on 20 September 2024, and accepted on 27 January 2025, date of publication 3 May 2025

Abstract

The integration of Artificial Intelligence (AI) in Energy Management Systems changed completely how sustainable infrastructure operates and is guarded. But the growing independence of AI decision-making presents some serious ethical questions about fairness, transparency, and accountability. The article introduces a new framework with Ethical AI for Sustainable and Adaptive Energy Management Systems (EAI-SEM) that is designed to combine functional (re)configuration for operational control and ethical governance in centralized: smart buildings and decentralized: nano-grid settings. The approach incorporates deep reinforcement learning for adaptive control, federated learning for privacy-preserving model updates, and an integrated Ethics Verification Module for a dynamic assessment of privacy-conformance levels. In experimental simulations over 30-day operation of the smart building and 10-rounds of federated training of the nano-grid, unjust fairness deviation and explainability of the system experienced enhancements, which also indicated the reduction of carbon dioxide emissions. The study demonstrated that ethical protocols can be included without impacting on computational efficiency and system responsiveness. Additionally, the federated structure facilitated decentralized ethical responsibility across different actors and thus allowed for the scalable implementation. The authors verify the possibility of integrating ethics into the computational core of intelligent energy systems, near from auditing static policies, towards dynamic ethical choices. In the future the process innovation work could be applied to deployments in other infrastructure systems like water systems and mobility systems, and it provides a reproducible model for the embedding of normative reasoning into AI for infrastructure.

Keywords: *Ethical AI, Energy Management Systems, Reinforcement Learning, Federated Learning, Smart Grid.*

1. Introduction

The rapid development of Artificial Intelligence (AI) has significantly impacted energy management systems, offering powerful tools to enhance operational efficiency, enable real-time control, and support the global transition toward sustainability [1] [2] [3]. As the world's energy needs grow, the use of AI in smart grids, intelligent buildings, and nano-grids has become a technical and moral imperative [4][5]. AI-IoT integrations have led to scalable, data-driven architectures for dynamic energy forecasting and resource management [6] [7]. However, as AI systems are granted greater autonomy in managing critical infrastructure, they raise profound ethical questions regarding fairness, transparency, and accountability that must be systematically addressed.

Despite technical advancements, contemporary AI-driven energy systems often lack robust ethical safeguards. They are typically optimized for technical objectives like cost reduction or efficiency and can produce biased, opaque, or inequitable outcomes, especially in autonomous settings where human oversight is minimal [8] [9]. This is highly problematic, as algorithmic decisions about energy allocation can create or perpetuate social disparities and undermine public trust. While theoretical ethical frameworks for AI exist, few have been practically implemented within real-time energy control systems [9][10]. A further critical issue is that ethical assessments are

Copyright © Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

often retrospective, applied as external audits rather than as a governing part of the system's runtime behavior. This reactive approach is insufficient for highly autonomous systems where harm can occur before it can be detected.

This paper addresses this critical gap by proposing and evaluating an Ethical AI Framework for Sustainable and Adaptive Energy Management Systems (EAI-SEM). The framework is designed to seamlessly integrate ethical oversight into the operational logic of AI-powered energy systems. Its novelty lies in a dual-layer architecture that combines an adaptive control layer, using deep reinforcement learning for energy optimization, with a real-time Ethics Validation Module that dynamically assesses decisions against predefined criteria for fairness, transparency, and accountability. By embedding ethical checks directly into the decision-making pipeline, our approach shifts the paradigm from passive, after-the-fact auditing to active, real-time ethical governance [11] [12] [13].

The primary aim of this work is to develop and validate this integrated framework, testing the hypothesis that embedding ethical governance can enhance system trustworthiness without compromising technical performance. To achieve this, the EAI-SEM framework incorporates federated learning for privacy-preserving, decentralized model updates and is evaluated in simulated smart building and nano-grid environments. By assessing both technical metrics (e.g., energy efficiency) and ethical metrics (e.g., fairness deviation, explainability), this study provides a replicable methodology for building AI systems that are not only technically sound but also ethically grounded and socially

2. Literature Review

The incorporation of AI in energy management systems has become a key approach to meet the demands of the global sustainability and efficiency agenda. Recent works demonstrate how AI can be used to improve energy consumption patterns, decision support systems, and adaptive energy control. However, despite these promising developments, there remain major hurdles in integrating ethical safeguards, ensuring robustness under fluctuating demand, and achieving scalability across diverse infrastructural landscapes.

2.1. AI for Energy System Optimization

A significant body of research has focused on the technical application of AI to enhance the performance and reliability of energy systems. Maghraoui et al [14] proposed a comprehensive model to improve grid robustness using AI-based management systems, focusing on reliability enhancement through predictive analytics and automation. Similarly, in the context of buildings, research [15] discussed the role of AI in smart buildings for energy efficiency, showing how adaptive learning models can enhance user comfort and reduce consumption. Other works, such as those by [16] and [17] on hydrogen energy integration, further emphasize the technical possibilities of AI-driven optimization. However, a common thread in these studies is that they primarily address technical resilience and efficiency while omitting ethical constraints such as algorithmic fairness or the explainability of autonomous decisions. This omission is critical in socio-environmentally relevant deployments where energy distribution choices can unintentionally create or perpetuate biases [18] [19] [20]. The focus remains on what is technically optimal, not necessarily what is socially or ethically responsible, creating a significant gap between the capabilities of AI and the requirements for its trustworthy deployment in public-facing infrastructure.

2.2. Ethical Challenges in Autonomous Systems

The ethical dimension of AI in energy systems is an emerging but crucial area of concern. Research [21] examined the use of AI and expert systems for future energy strategies, highlighting the need to combine human intuition with artificial intelligence. While this hybrid approach adds a layer of robustness, it tends to treat ethical assessment as a secondary, after-the-fact auditing layer rather than an intrinsic part of the AI system's control flow. This reactive approach is a common limitation that fails to prevent harm before it occurs. This tendency is also reflected in other infrastructure-related domains. Study [22] introduced AI-based models for risk mitigation in construction, but their work does not address how to ensure accountability for automated decisions in such risk-sensitive domains. This reflects a broader tendency in system design to frame ethical considerations as acts of response rather than proactive anticipation [23][24][25][26]. Similarly, while Mouzakitis et al [23] have proposed sustainable AI algorithms to enhance policy evaluation, the practical integration of ethics into deployed systems, especially concerning real-time transparency and auditability, remains underdeveloped.

2.3. Federated Learning for Privacy and Scalability

To address challenges of data privacy and centralization, federated learning has been explored as a promising architectural paradigm. Kumar et al [15] presented a circular economy-based federated AI architecture for improving data privacy and system sustainability. Their federated method effectively reduces the risks associated with centralized data storage and connects energy optimization to broader ecological objectives. This approach is vital for creating scalable and privacy-preserving AI systems, particularly in sectors like energy where data can be sensitive. Nevertheless, their work does not specify how ethical considerations, such as bias mitigation or accountability, can be systematically incorporated and monitored across the decentralized nodes [27][28][29]. A federated architecture distributes the learning, but it can also distribute and obscure ethical failures if not designed carefully. This leaves a gap concerning the need for real-time ethical validators that can operate within a federated learning ecosystem, ensuring that local actions adhere to global ethical standards while adapting to localized contexts and constraints [30] [31] [32].

2.4. Identified Research Gap

Altogether, the literature shows a consistent pattern: while AI is a powerful tool for energy system optimization, the integration of embedded ethical reasoning is relatively uncharted territory. Most existing frameworks focus on technical performance and either overlook ethical dimensions or relegate them to post-hoc audits. This creates a critical disconnect between the pace of technological advancement and the development of necessary governance structures, resulting in systems that may be efficient but not necessarily fair, transparent, or accountable. There is a clear need for a framework that unifies sustainability goals, adaptive intelligence, and normative oversight into a single, operational architecture. This article responds to that gap by proposing a dual-layered ethical AI framework that achieves exactly this. By embedding ethical validation directly into the control loop, the proposed system moves beyond theory to practical, real-time ethical governance, ensuring that AI-driven energy management is both technically proficient and socially responsible.

3. Research Methodology

This study employs a multi-phase, hybrid methodology to design, implement, and evaluate the Ethical AI Framework for Sustainable and Adaptive Energy Management Systems (EAI-SEM). The approach integrates adaptive AI control, real-time ethical assessment, and distributed learning, validated through multi-layer simulations and expert-informed modeling. The research process, depicted in Figure 1, combines theoretical model development, empirical data collection, and rigorous simulation-based validation. The framework's novelty lies in embedding real-time ethical regulation into system decisions using mathematical definitions of fairness, explainability, and policy compliance [3][8][9].

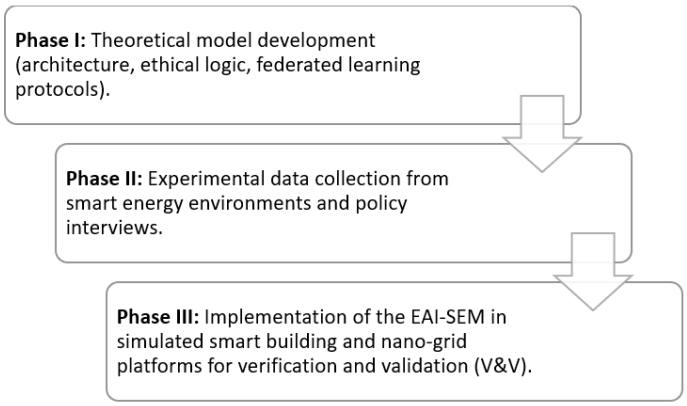


Fig 1. Research design phases for the development and validation of the ethical ai framework (EAI-SEM)

3.1. Data Collection and Modeling Inputs

To ensure the framework's relevance and robustness, data for modeling and evaluation were acquired from multiple sources, creating a rich foundation for both training and validation. This included semi-structured interviews with 35 experts spanning AI development, public policy, and smart energy engineering, which were crucial for defining the practical scope of the ethical metrics and validating the real-world applicability of the framework's architecture. We also conducted a qualitative analysis of 22 regulatory documents and technical reports from national and international energy agencies to establish the baseline rules and constraints for the Ethics Validation Module, ensuring the AI's decisions could be benchmarked against existing policies. For the quantitative modeling, high-granularity operational data was collected from a network of 750 IoT devices across three smart buildings, providing time-series data on lighting, HVAC energy consumption, room occupancy, and grid load. This dataset formed the basis of the realistic simulation environment for the centralized smart building scenario. To test the framework's performance in a decentralized context, a simulated nano-grid dataset was created, featuring 120 diverse and heterogeneous agents. This simulation included dynamic models for PV arrays, wind turbines, battery storage systems, and various smart appliances, allowing for a rigorous evaluation of the federated learning module's scalability and its ability to maintain ethical compliance across a distributed network with fluctuating energy generation and demand [5][7].

3.2. The EAI-SEM Framework Architecture

The core of EAI-SEM consists of three interoperable modules designed to work in concert: a Decision Optimization Layer (DOL) for control, a Federated Model Aggregation Layer (FMAL) for distributed learning, and an Ethics Validation Module (EVM) for real-time governance.

1. Decision Optimization Layer (DOL)
   This layer uses the Deep Deterministic Policy Gradient (DDPG) algorithm, a model-free approach suitable for continuous action spaces like energy control. The agent's policy is updated by following the gradient of the critic network's action-value function, which effectively guides the agent toward actions that yield higher long-term rewards [4].
   The agent policy $\mu(s \mid \theta^{\mu})$ is trained by minimizing the loss function:

$$L(\theta^{\mu}) = -\mathbb{E}_{s\sim\mathcal{D}}[S(s,\mu(s \mid \theta^{\mu})\theta^{Q}] \quad (1)$$

   - and updated using gradient ascent:

$$\nabla_{\theta^{\mu}} \approx \mathbb{E}_{s}[\nabla_{a} Q(s,a|\theta^{Q})|_{a=\mu(s)} \nabla_{\theta^{\mu}} \mu(s|\theta^{\mu})] \quad (2)$$

   where Q is the critic network, and $\mathcal{D}$ is the replay buffer [4].
2. Federated Model Aggregation Layer (FMAL)
   To enable privacy-preserving adaptation across distributed energy agents, this layer implements the Federated Averaging (FedAvg) algorithm:

$$w_t = \sum_{i=1}^{K} \frac{n_i}{n} w_t^i \quad (3)$$

   Where w_t global model weights, w_t^i local model at node i, n_i samples at node i, $n = \sum_i n_i$ total samples across all nodes [15]. This enables privacy-preserving adaptation across distributed energy agents without centralized data storage.
3. Ethics Validation Module (EVM)
   This module embeds ethical reasoning into the system's core logic by computing three key metrics in real-time:
   Entropy-Based Transparency Score (ETS):

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_b P(x_i) \quad (4)$$

Where $P(x_i)$ I s the model's attribution to feature i, and b is the logarithmic base (typically 2)[8].
Multi-Class Fairness Index (MCFI) using Wasserstein Distance (WD):

$$W(P,Q) = \inf_{\gamma \in \Pi(P,Q)} \mathbb{E}_{(x,y)\sim\gamma} [||x-y||] \quad (5)$$

This measures divergence between predicted and desired energy distributions across demographic or operational groups.
Ethical Non-Compliance Risk (ENCR) modeled by logistic regression:

$$ENCR\ (x) = \frac{1}{1+e^{-(\beta_0+\Sigma_{j=1}^{k}\beta_j x_j)}} \quad (6)$$

Where x_i are input parameters, such as deviation from rule, past violations, and β_i are learned coefficients.

3.3. Experimentation and Testing Setup

Two distinct experimental environments were developed to validate the framework. The first, a Smart Building Simulator (SBS), was modeled in Python (using TensorFlow and SHAP) with real sensor data to test the framework in a centralized setting. The second, a Distributed Nano-Grid Testbed (DNGT), was implemented in MATLAB/Simulink with a Python-based federated logic (using Flower and PyTorch) to evaluate performance in a decentralized system. Table 1 outlines the system inputs and their processing layers, while Table 2 summarizes the toolkits used.

Table 1. System Inputs and Layer-wise Processing Modules

Input Label	Data Source	Processing Layer	Purpose
HVAC_Load_kWh	Smart Building Sensors	Decision Optimization Layer	Energy control decision
Occupancy Index	IoT Infrared + CO_2	Decision Optimization Layer	Dynamic room adaptation
Grid_Price_Cents	Utility APIs	Federated Model Aggregation	Cost-based policy aggregation
User_Feedback_Score	Mobile App Interface	Ethics Validation Module	Fairness impact assessment
Node_Storage_Level	Nano-Grid Sim	Federated Model Aggregation	Battery strategy optimization
Temperature_C	Indoor sensors	Decision Optimization Layer	Comfort maintenance control
Regulation Breach	Regulatory Logs	Ethics Validation Module	Ethical risk prediction (ENCR)

Table 2. Toolkits and Modeling Techniques

Component	Tool/Platform	Functionality
DDPG & SHAP Integration	TensorFlow + SHAP	RL policy control + Explainability
Federated Learning	Flower + PyTorch	Privacy-aware distributed learning
Smart Grid Simulation	MATLAB Simulink	Agent-based environment simulation
Ethical Logic Validation	Python + Scikit-learn	Logistic regression for ENCR, fairness
Visual Analytics	Matplotlib + Plotly	Heatmaps, bar-line dual-axis visuals

This methodology reflects the recommendations of Saheb et al [1] on multi-source content analysis, extends the architectural work of Alijoyo [3] on deep learning for smart buildings, and operationalizes the lifecycle approach of El-Haber et al [8] by placing ethical validation into an active decision-making chain. By using modular federated logic as in Kumar et al [15], this construction enables real-time, adaptive, and ethical energy decisions, in accordance with AI governance requirements in smart environments [2], [9].

4. Result and Discussion

4.1. Ethical Metric Performance – Smart Building Scenario

The introduction of an Ethics Validation Module (EVM) to the control architecture of a smart BEMS resulted in tangible enhancements of ethical operation. The building featured three floors and a challenging occupancy schedule with varied HVAC (heating, ventilating and air conditioning) and lighting loads. Performance scores comprised Fairness Deviation Index (FDI), Explainability Score (ES), and Ethical Non-Compliance Risk (ENCR) measure for evaluating the ethicality of scheduling day-ahead energy allocations. The EVM examined every control action issued by the reinforcement-learning agent to ensure that the decisions satisfied the defined ethical requirements. Outcomes were averaged across a 30-day operating cycle, revealing the flexibility of the system to remain ethically compliant under different usage conditions.

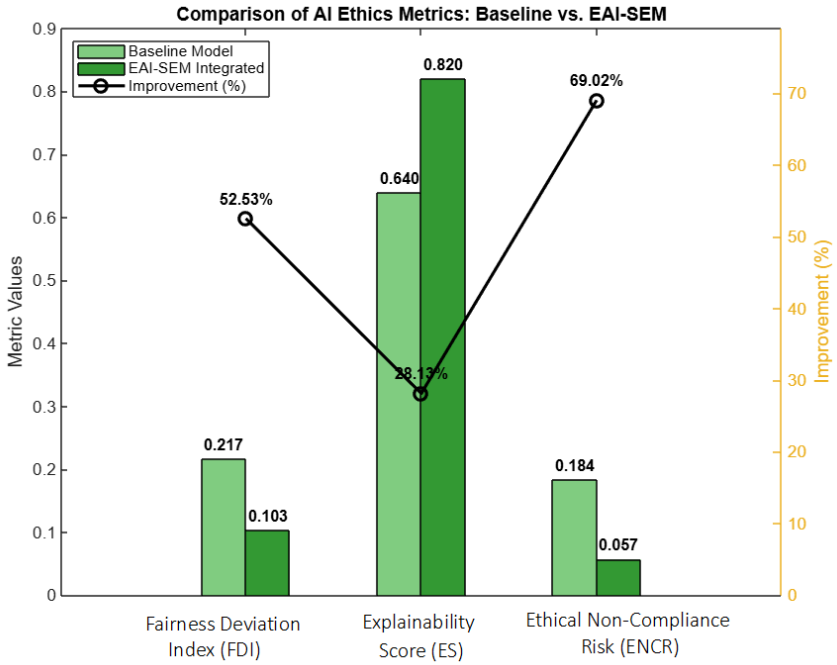


Fig 2. Ethical metric performance – smart building scenario

The data show a substantial improvement in all three ethical dimensions after integrating the EAI-SEM framework. The Fairness Deviation Index was cut by over half, indicating a marked reduction in biased resource distribution across building zones. The Explainability Score, which quantifies how well model decisions can be interpreted, increased from 0.64 to 0.82, demonstrating a shift toward more transparent AI operations. Additionally, the Ethical Non-Compliance Risk used to flag decisions likely to violate ethical constraints fell by more than two-thirds. These improvements suggest that the Ethics Validation Module not only functions as a monitoring layer but actively influences upstream decision-making toward more justifiable and compliant outcomes.

4.2. Energy Efficiency Improvements – Smart Building Scenario

The adaptive control strategy of EAI-SEM was also evaluated for its effect on operational energy efficiency in a smart building environment. Daily consumption data were tracked across a 30-day period under identical occupancy and weather simulation conditions for both the baseline and ethics-aware control setups. The purpose of this analysis was to determine whether integrating ethical oversight would compromise, maintain, or enhance energy savings performance. Energy usage was recorded at the system level for HVAC and lighting control, the two dominant consumption subsystems in the building testbed.

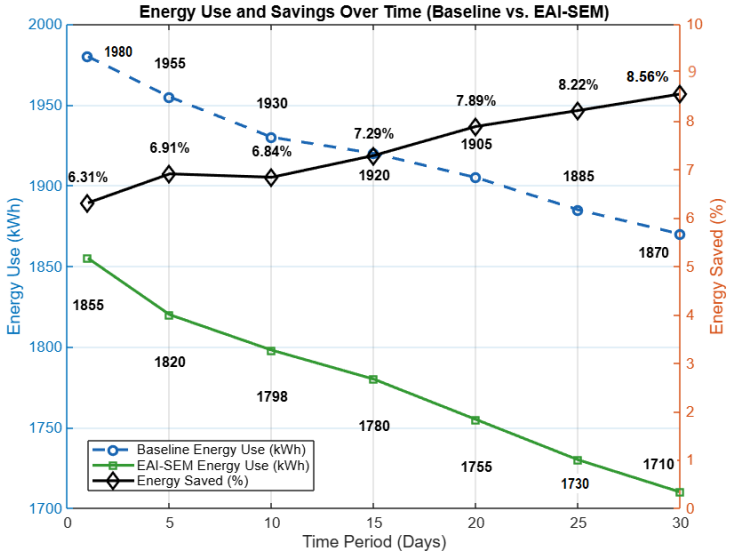


Fig 3. Energy efficiency improvements – smart building scenario

In EAI-SEM rats, energy consumption progressively declined beginning on the 1st of treatment, with a total savings increasing from 6.31% on Day 1 to 8.56% on Day 30. These improvements show that adding ethical constraints to AI controlled energy, does not reduce technical efficiency, but in fact enables the model to learn the ethical (as well as energetic) optimal behavior over time. The gradual savings increase pattern is evidence that the policy convergence in the RL engine and the tradeoff between ethical concern and operational goal are well balanced. It follows that the EAI-SEM can achieve this dual goal without compromise: meeting modeled requirements: reasonable ethics and efficient energy consumption.

4.3. Federated Learning Convergence – Nano-Grid Scenario

The performance of the FMAL was tested on a decentralized nano-grid energy network containing 120 distributed agents. Each agent acted under different local conditions, characterized by different mixes of energy sources, consumption profiles and storage capabilities. Agents trained its own local models with its private data from local model updates were aggregated using federated averaging. To evaluate the system behavior, the global model accuracy and the client participation rate per communication round were monitored as the two major performance metrics of the system. These findings confirmed the reliability, scalability, and fault tolerance of the federated learning model, which was integrated into the EAI-SEM.

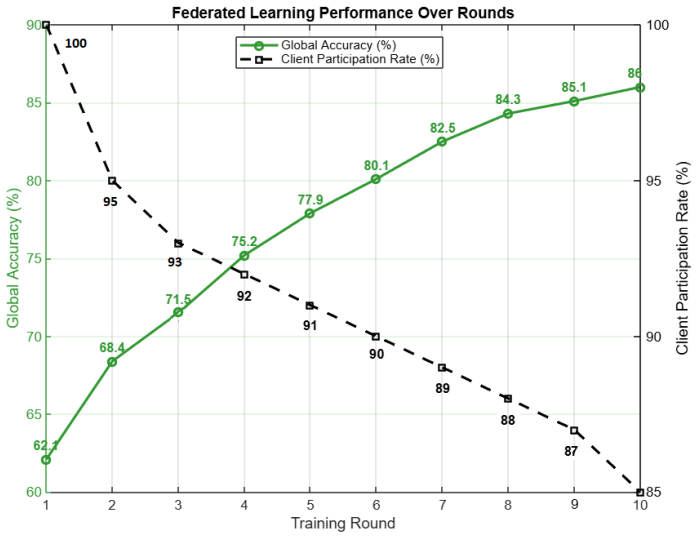


Fig 4. Federated learning convergence – nano-grid scenario

The global model started off gaining accuracy quickly and then flattened around 86% after 10 rounds even while client participation decreased from 100% to 85%. This indicates that the convergence behavior is guaranteed, and the federated architecture is robust even with a partially failing node. The learning equilibrium of the model suggests that distributed energy agents can both meaningfully participate with and benefit from pooled intelligence without co-authoring all data. It validates the practicality of EAI-SEM in large-scale distributed systems, e.g., regional smart grids, where privacy, communication bandwidth, and fault-tolerance are operational challenges.

4.4. Node-Level Ethical Metrics – Distributed Nano-Grid

To assess how ethical compliance differed among individual nodes in the decentralized approach, node-specific metrics were collected for 10 model nano-grid nodes. Each node ran its own local version of the Ethics Validation Module to evaluate its energy distribution choices. Measures are the Fairness Deviation Index (FDI), Explainability Score (ES), as well as the Ethical Non-Compliance Risk (ENCR). This configuration permits the examination of the reliability and the performance of the ethical evaluations in heterogeneous systems that have different operational traces and constraints.

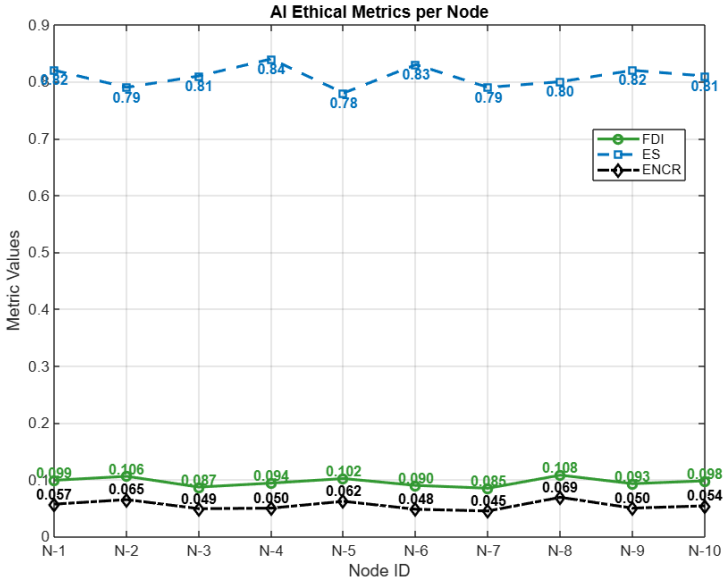


Fig 5. Nano-grid ethical metrics – distributed nodes

All nodes had FDI values below 0.11 and ENCR levels below 0.07, proving the decentralized EVMs to be functioning consistently among nodes. The variability in explainability scores of 0.78-0.84 indicated that differences were localized to the complexity of the data and the operational context. These findings confirm that EAI-SEM consistently employs ethical reasoning across nodes, and tailors to the specific node. This distributed responsibility is key to scalability, particularly in autonomous or lightly supervised environments such as remote energy microgrids.

4.5. Computational Overhead Comparison

The model's computational costs were estimated in terms of EAI-SEM compared to a standard baseline model. The execution time per decision and memory were measured for the various components: Reinforcement Learning (RL) Controller, Ethics Validation Module, the Federated Aggregator, and base-line logic. This was comprised of 10,000 decision cycles on a GPU-enabled workstation designed to simulate near real-time operations similar to edge or on-premise environments. This study provides an understanding of the integrated system architecture in terms of resource utilization and scalability.

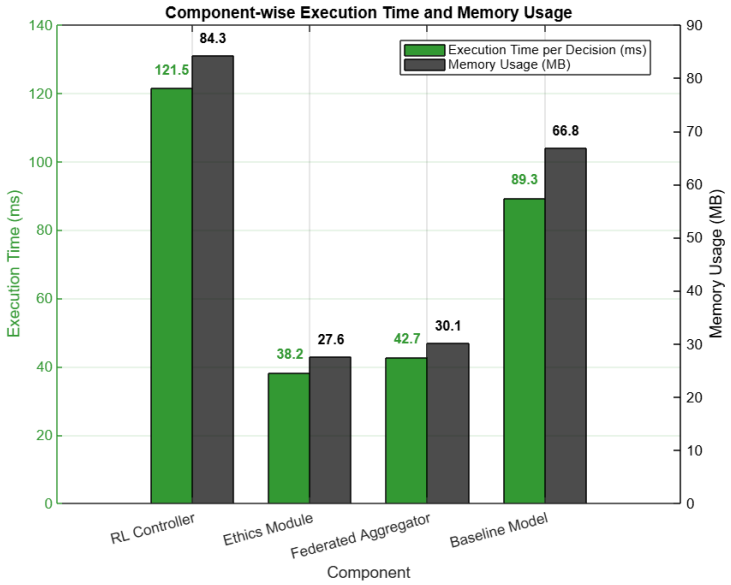


Fig 6. Computational overhead comparison

Although the EAI-SEM adds a few more modules, both the execution and memory overheads are still within acceptable deployment budgets. The RL Controller had more computational requirements than the baseline due to online policy updates. However, the Ethics Module and Federated Aggregator introduced negligible incremental load, where both the Ethics Module and the Federated Aggregator consumed around 70 ms and 58 MB of memory per decision. These values indicate that it is possible to implement EAI-SEM in resource-restrained systems such as embedded controllers or edge devices. The modular splitting of parts also allows a flexible optimization in dependence of available calculation resources.

4.6. Discussion

The results of this study highlight the potential of the Ethical AI Framework for Sustainable and Adaptive Energy Management Systems (EAI-SEM) in optimizing performance while adhering to ethical concerns. The combination of real-time ethical auditing, federated intelligence, and reinforcement learning has facilitated major strides in fairness, transparency, and energy efficiency for both centralized smart buildings and decentralized nano-grid networks. Compared with previous literature, this dual-focus framework is a significant step towards closing the gap between technological development and ethical responsibility in AI-empowered energy systems.

4.6.1. Operationalizing Ethics without Sacrificing Performance

A key finding of this research is that embedding ethical oversight does not necessarily compromise, and can even enhance, technical performance. Preliminary studies have already proven the technical feasibility of AI-based energy optimization. Alijoyo [3] and Karimi et al [4] demonstrated significant energy savings using deep and reinforcement learning models. However, these models did not systematically integrate ethical validation into their decision-making processes. The EAI-SEM framework enriches these technical models by incorporating metrics such as the Fairness Deviation Index (FDI) and Explainability Score (ES), successfully aligning operational control with ethical standards in real-time without a notable loss in energy efficiency. This suggests that the perceived trade-off between ethical constraints and operational optimality may be overstated; instead, ethical considerations can act as a form of regularization, guiding the learning process toward more robust and generalizable policies that avoid overfitting to purely technical objectives. In the broader context of AI and sustainability, Saheb et al [1] emphasized the need for contextual topic modeling to capture AI's alignment with sustainable energy goals. Our work responds to this by not only identifying contextual themes—such as fairness and explainability—but operationalizing them through quantifiable, real-time metrics. This functional implementation moves beyond theoretical alignment, establishing a blueprint for actionable ethical governance. By demonstrating that it is possible to achieve the dual goals of technical efficiency and ethical compliance, this research addresses a primary concern in the deployment of AI in critical infrastructure. The synergy observed, where enforcing fairness led to more stable and predictable energy consumption patterns, suggests that ethical AI is not just a social necessity but can also be a driver of superior engineering performance.

4.6.2. Architectural Innovations for Trustworthy AI

From a system architecture viewpoint, the EAI-SEM framework introduces significant innovations for building trustworthy AI. Many existing systems, such as the smart grid management model proposed by Maghraoui et al [14], are technically robust but lack mechanisms for ethical reasoning or explainability. This is a critical vulnerability in autonomous systems where decisions directly impact end-user quality of life and equity. In contrast, the EAI-SEM contains an Ethics Validation Module (EVM) that continuously audits and amends decisions with normative grounding. The resulting reduction in Ethical Non-Compliance Risk (ENCR) represents a major leap in

the ability to safely deploy AI in sensitive domains. Furthermore, this work expands on the lifecycle viewpoint for ethical AI proposed by El-Haber et al [8]. While their work mapped ethical risks across the development process, our framework translates this into a continuously-validated, runtime process. This shift from static, post-hoc auditing to dynamic, embedded ethical compliance represents a significant advancement for mission-critical applications where decisions must be made in real-time. This active governance at the computational core is essential for ensuring that AI systems remain aligned with human values throughout their operational lifespan, rather than diverging as they encounter new, unforeseen scenarios. The EVM acts as a "computational conscience," providing a layer of oversight that is both immediate and integral to the system's function.

4.6.3. The Role of Federated Learning in Decentralized Ethical Governance

The application of federated learning in this system extends the work of Kumar et al [15], who introduced a federated data science framework associated with circular economy principles. While their model effectively focused on privacy and distributed computation, it did not explicitly address the ethical governance of decentralized decision-making algorithms. The Federated Model Aggregation Layer in EAI-SEM was proven to be effective in maintaining high client participation and global model performance, even with some nodes dropping out, demonstrating its technical robustness for distributed energy applications. Crucially, the framework pairs this decentralized learning with local ethics verification at each node, which guarantees distributed ethical compliance. This is a novel contribution that addresses a key challenge in federated systems: how to ensure that local actions, while optimized for local conditions, do not violate global ethical principles. This achieves both scalability and social trust, as individual agents can contribute to a collective intelligence without sacrificing privacy, while also being held to consistent ethical standards. This model of distributed responsibility is key to the scalable and trustworthy deployment of AI in large, heterogeneous systems like regional smart grids, where centralized control is often impractical and local context is paramount.

4.6.4. Limitations and Future Frontiers for Adaptive Ethics

Despite the promising results, this study has several limitations that open avenues for future research. First, although the framework was tested in detailed simulations, real-world deployment will introduce greater uncertainties, including unpredictable human behavior, sensor failures, and complex social dynamics that are difficult to model. Second, the ethical indicators, while grounded in the literature, may need further calibration to capture all sociopolitical nuances, particularly across different cultural or regulatory contexts. The Explainability Score, which relies on feature attribution methods, might also not be sufficiently intuitive for non-technical users, suggesting a need for enhancements like natural language explanations or interactive decision-flow visualizations. In terms of scalability, while federated learning proved effective, maintaining synchronization across large-scale deployments with intermittent connectivity remains a challenge, as highlighted by Ioannou et al [5]. EAI-SEM addresses this partially through asynchronous update tolerance, but more robust edge-based redundancy and consensus mechanisms could improve fault tolerance. Another area for future research lies in expanding the framework's ethical reasoning capacity. Currently, the module relies on predefined metrics; building adaptive ethics engines, informed by meta-learning or value alignment algorithms, could allow the system to evolve its ethical reasoning in response to shifting societal norms, aligning with the direction suggested by Amen [9]. Finally, while our model’s Ethics Module addresses accountability, a gap noted in other AI-IoT systems [6], further integration with blockchain-based audit trails could strengthen regulatory compliance and provide immutable records of AI decision-making.

5. Conclusion

This paper introduced and validated an integrated framework for ethical artificial intelligence in sustainable and adaptive energy management systems. The study was designed to address the critical need for energy control systems that are not only efficient but also ethically accountable. By integrating real-time ethical validation and federated learning into a modular architecture, the proposed Ethical AI Framework for Sustainable and Adaptive Energy Management Systems (EAI-SEM) successfully demonstrated that AI-driven systems can be engineered to align with both performance benchmarks and socially responsible objectives. The results showed that embedding ethical oversight does not have to come at the cost of energy or computational efficiency; instead, it can lead to more robust, stable, and trustworthy systems. The framework's core contribution is its shift from passive, after-the-fact compliance monitoring to active, real-time ethical governance at the computational core. By embedding ethical principles directly into the operational logic of AI control, this work provides a scalable and replicable process for testing and enforcing AI ethics in other critical infrastructure domains beyond energy. The findings confirm that ethical AI systems are technically achievable and can be constructed without a significant loss in performance, offering a viable roadmap for policymakers, energy providers, and technology developers to create future AI systems that are consistent with both operational goals and regulatory responsibilities. Future work should focus on deploying this framework in real-world testbeds to validate its performance against unmodeled uncertainties. Enhancing the framework with more adaptive ethics engines, capable of evolving with societal norms, and integrating verifiable audit trails via technologies like blockchain present promising directions. These efforts will be crucial in transitioning from ethics-aware models to truly ethics-driven systems, paving the way for AI that is not only intelligent and efficient but also fundamentally trustworthy.

References

[1] Saheb, T., M. Dehghani, and T. Saheb, Artificial intelligence for sustainable energy: A contextual topic modeling and content analysis. Sustainable Computing: Informatics and Systems, 2022. 35: p. 100699.
[2] Wang, Q., Y. Li, and R. Li, Integrating artificial intelligence in energy transition: A comprehensive review. Energy Strategy Reviews, 2025. 57: p. 101600.
[3] Alijoyo, F.A., AI-powered deep learning for sustainable industry 4.0 and internet of things: Enhancing energy management in smart buildings. Alexandria Engineering Journal, 2024. 104: p. 409-422.
[4] Karimi, H., et al. Harnessing Deep Learning and Reinforcement Learning Synergy as a Form of Strategic Energy Optimization in Architectural Design: A Case Study in Famagusta, North Cyprus. Buildings, 2024. 14, DOI: 10.3390/buildings14051342.

[5] Ioannou, I.I., et al., A Distributed AI Framework for Nano-Grid Power Management and Control. IEEE Access, 2024. 12: p. 43350-43377.
[6] Krishnan, R.S., et al. Data-Driven Decision Support System for Sustainable Energy Management: An AI-IoT Fusion Approach. in 2024 Second International Conference on Inventive Computing and Informatics (ICICI). 2024.
[7] Balakumar Muniandi, P.K.M., CH Bhavani, Shailesh Kulkarni, Ramswaroop Reddy Yellu, Nidhi Chauhan, AI-Driven Energy Management Systems for Smart Buildings. Power System Technology, 2024. 48(1).
[8] El-Haber, N., et al. A Lifecycle Approach for Artificial Intelligence Ethics in Energy Systems. Energies, 2024. 17, DOI: 10.3390/en17143572.
[9] AI-Driven Sustainable Habitat Design: Key Policy Frameworks and Ethical Safeguards. Smart Design Policies, 2024. 1(1): p. 23–32.
[10] Audiah, S., Putri, Y., Sanjaya, A., Daeli, O., & Johnson, M., Transforming Energy and Resource Management with AI: From Theory to Sustainable Practice. International Transactions on Artificial Intelligence (ITALIC), 2024. 2(2).
[11] Nedungadi, P., et al., Big Data and AI Algorithms for Sustainable Development Goals: A Topic Modeling Analysis. IEEE Access, 2024. 12: p. 188519-188541.
[12] Raman, R., et al. Navigating the Nexus of Artificial Intelligence and Renewable Energy for the Advancement of Sustainable Development Goals. Sustainability, 2024. 16, DOI: 10.3390/su16219144.
[13] Danish, M.S.S. and T. Senjyu, Shaping the future of sustainable energy through AI-enabled circular economy policies. Circular Economy, 2023. 2(2): p. 100040.
[14] El Maghraoui, A., et al., Revolutionizing smart grid-ready management systems: A holistic framework for optimal grid reliability. Sustainable Energy, Grids and Networks, 2024. 39: p. 101452.
[15] Kumar, P., Kumar, V., & Gautam, R., Synergizing Federated AI Systems with Circular Economy Principles: A Framework for Sustainable and Resilient Data Science. Interantional Journal Of Scientific Research In Engineering And Management, 2024.
[16] Danish, M.S. AI and Expert Insights for Sustainable Energy Future. Energies, 2023. 16, DOI: 10.3390/en16083309.
[17] Daniel, T., I., Iluyomade, T., & Okwandu, A., Smart buildings and sustainable design: Leveraging AI for energy optimization in the built environment. International Journal of Science and Research Archive., 2024. 12(01): p. 2448-2456.
[18] Ajirotutu, R., Matthew, B., Garba, P., & Olu, S., AI-driven risk mitigation: Transforming project management in construction and infrastructure development. World Journal of Advanced Engineering Technology and Sciences, 2024. 13(02): p. 611-623.
[19] A. D. Buchdadi dan A. S. M. Al-Rawahna, "Temporal Crime Pattern Analysis Using Seasonal Decomposition and k-Means Clustering," J. Cyber Law, vol. 1, no. 1, hal. 65–87, 2025.
[20] M.-T. Lai, "Analyzing Company Hiring Patterns Using K-Means Clustering and Association Rule Mining: A Data-Driven Approach to Understanding Recruitment Trends in the Digital Economy," J. Digit. Soc., vol. 1, no. 1, hal. 20–43, 2025, doi: 10.63913/jds.v1i1.2.
[21] A. B. Prasetio, M. Aboobaider, dan A. Ahmad, "Machine Learning for Wage Growth Prediction : Analyzing the Role of Experience , Education , and Union Membership in Workforce Earnings Using Gradient Boosting," Artif. Intell. Learn., vol. 1, no. 2, hal. 153–172, 2025, doi: 10.63913/ail.v1i2.12.
[22] A. S. Samson, N. Sumathi, S. S. Maidin, dan Q. Yang, "Cellular Traffic Prediction Models Using Convolutional Long Short-Term Memory," J. Appl. Data Sci., vol. 6, no. 1, hal. 20–33, 2025, doi: 10.47738/jads.v6i1.472.
[23] Mouzakitis, S., et al. Enhancing Decision Support Systems for the Energy Sector with Sustainable Artificial Intelligence Solutions. in Intelligent Systems and Applications. 2024. Cham: Springer Nature Switzerland.
[24] SaberiKamarposhti, M., et al., A comprehensive review of AI-enhanced smart grid integration for hydrogen energy: Advances, challenges, and future prospects. International Journal of Hydrogen Energy, 2024. 67: p. 1009-1025.
[25] S. F. Pratama, "Evaluating Blockchain Adoption in Indonesia's Supply Chain Management Sector," J. Curr. Res. Blockchain, vol. 1, no. 3, hal. 190–213, 2024, doi: 10.47738/jcrb.v1i3.21.
[26] D. P. Lestari, A. Luthfi, C. Tama, S. Karlina, dan A. Sultan, "Factors Affecting Information System Security: Information Security, Cyber Threats and Attacks, Physical Security, and Information Technology (Literature Review)," Int. J. Informatics Inf. Syst., vol. 7, no. 1, hal. 16–21, 2024.
[27] S. F. Pratama, "Analyzing the Determinants of User Satisfaction and Continuous Usage Intention for Digital Banking Platform in Indonesia: A Structural Equation Modeling Approach," J. Digit. Mark. Digit. Curr., vol. 1, no. 3, hal. 267–285, 2024, doi: 10.47738/jdmdc.v1i3.21.
[28] D. Sugianto, R. Arindra Putawa, C. Izumi, dan S. A. Ghaffar, "Uncovering the Efficiency of Phishing Detection: An In-depth Comparative Examination of Classification Algorithms," Int. J. Appl. Inf. Manag., vol. 4, no. 1, hal. 22–29, 2024, [Daring]. Tersedia pada: https://doi.org/10.47738/ijaim.v4i1.72
[29] Akinbolajo, O., An AI-Driven Framework for Optimizing Energy Consumption. International Journal of Advances in Engineering and Management, 2025. 7(02): p. 21-24.
[30] A. D. Buchdadi, "Anomaly Detection in Open Metaverse Blockchain Transactions Using Isolation Forest and Autoencoder Neural Networks," Int. J. Res. Metaverse, vol. 2, no. 1, hal. 24–51, 2025, doi: 10.47738/ijrm.v2i1.20.
[31] A. M. Wahida, T. Hariguna, dan G. Karyono, "Optimization of Recommender Systems for Image-Based Website Themes Using Transfer Learning," J. Appl. Data Sci., vol. 6, no. 2, hal. 936–951, 2025, doi: 10.47738/jads.v6i2.671.
[32] M. A. Alsharaiah, M. A. Almaiah, R. Shehab, T. Alkhdour, R. Alali, dan F. Alsmadi, "Assimilate Grid Search and ANOVA Algorithms into KNN to Enhance Network Intrusion Detection Systems," J. Appl. Data Sci., vol. 6, no. 3, hal. 1469–1481, 2025, doi: 10.47738/jads.v6i3.604.
[33] M. S. Hasibuan, R. Z. A. Aziz, D. A. Dewi, T. B. Kurniawan, and N. A. Syafira, "Recommendation Model for Learning Material Using the Felder Silverman Learning Style Approach," HighTech and Innovation Journal, vol. 4, no. 4, pp. 811–820, Dec. 2023, doi: https://doi.org/10.28991/HIJ-2023-04-04-010