ORIGINAL RESEARCH **Open access**

From Predictive Optimization to Responsible Intelligence: An Ethical AI Framework for Energy Management Systems

Mark Anderson[1*], Lisa Wong[2], Robert Smith[3], Sarah Lee[1], David Park[2]

Abstract

Energy management systems increasingly rely on AI-driven predictive optimization to reduce costs, enhance efficiency, and balance supply and demand. As these systems automate decisions affecting building occupants, industrial workers, and electricity consumers, the ethical dimensions of algorithmic energy control become impossible to ignore. Current AI architectures for energy management prioritize predictive accuracy and economic optimization while neglecting fairness, accountability, transparency, human agency, and occupant well-being. This narrow focus risks creating systems that are technically capable but ethically impoverished and socially insensitive. This article proposes an ethical AI framework for energy management systems that transitions from predictive optimization to responsible intelligence, embedding ethical principles into the AI design and operation lifecycle. The framework positions ethical reasoning as part of system architecture rather than an external compliance supplement. The framework is developed through conceptual synthesis of AI ethics principles, responsible AI scholarship, energy justice literature, occupant-centric building science, and energy management system governance. No experiments or empirical analyses are conducted. The proposed framework is organized into three integrated components: an ethical reasoning layer that translates values into operational constraints, a transparency and explainability layer that makes AI decisions auditable, and an accountability and oversight layer that allocates responsibility and enables contestation. Together, these components reframe predictive optimization as a morally significant act rather than a purely technical task. Moving from predictive optimization to responsible intelligence requires treating ethical principles as design requirements, not optional additions. The ethical AI framework provides a structured approach for developing energy management systems that are not only efficient but fair, transparent, accountable, and aligned with human flourishing.

Keywords Fairness, Ethical artificial intelligence, Responsible intelligence, Energy management systems, Predictive optimization, AI ethics

*Correspondence:
Mark Anderson
mark.anderson@gatech.edu
[1] Department of Artificial Intelligence for Resilient Power Systems, College of Engineering, Georgia Institute of Technology, Atlanta, United States
[2] Department of Intelligent Energy Analytics and Machine Learning, Faculty of Applied Science, University of British Columbia, Vancouver, Canada
[3] Department of Smart Grid Optimization, College of Engineering, University of Michigan, Ann Arbor, United States

Introduction

Artificial intelligence has become increasingly central to energy management systems, where predictive models are used to anticipate demand, optimize control actions, coordinate distributed assets, and reduce operational costs. Data-driven building energy research has shown how forecasting and control methods can support more adaptive energy use across

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

buildings and grid-interactive environments [1]. Reviews of building energy prediction further demonstrate that AI-enabled systems can improve the technical sophistication of energy demand management, particularly when large-scale operational data are available [2]. Yet, when these systems are framed only around efficiency, cost minimization, or predictive accuracy, they risk overlooking the ethical consequences of automated energy control.

The ethical gap in AI-driven energy management arises because optimization decisions can distribute benefits and burdens unevenly across occupants, tenants, consumers, and communities. A system that minimizes aggregate energy cost may still generate unfair indoor temperature conditions, reduce comfort for less powerful users, or impose automated demand response burdens on households with limited flexibility. Energy justice scholarship has emphasized that energy systems are never purely technical because access, affordability, vulnerability, and recognition shape how energy decisions affect different groups [3]. Therefore, efficiency alone cannot provide a sufficient normative foundation for AI in energy management systems.

Responsible intelligence offers an alternative paradigm in which AI systems are designed not merely to predict and optimize, but to act within ethically justified boundaries. Broader AI ethics scholarship has argued that responsible AI requires more than abstract principles; it requires translation into practical governance, design methods, and accountability mechanisms [4]. In the EMS domain, this means adapting fairness, transparency, accountability, and human well-being to decisions about heating, cooling, storage dispatch, load shifting, and demand response. Such adaptation is necessary because energy management decisions affect comfort, health, autonomy, and participation in increasingly automated built environments.

This article proposes an ethical AI framework for energy management systems that systematically moves from predictive optimization toward responsible intelligence. The framework builds on the argument that ethical AI cannot be achieved through principles alone unless those principles are embedded into design practices, institutional procedures, and operational decision rules [5]. It conceptualizes EMS intelligence as a socio-technical capability that must integrate ethical reasoning, explainability, accountability, and stakeholder inclusion. The central thesis is that responsible EMS AI should optimize only within ethically defensible constraints and should remain contestable by the humans and communities it affects.

Background

Energy management systems and predictive optimization

Modern energy management systems coordinate sensing, forecasting, control, and optimization across buildings, industrial facilities, distributed energy resources, and grid-edge environments. AI methods have expanded this capability by enabling load forecasting, occupancy prediction, model predictive control, and reinforcement learning for demand response and automated energy scheduling [6]. Studies of large-scale building energy forecasting show that predictive systems can incorporate weather, occupancy, historical consumption, and operational variables to guide control decisions [7]. However, these systems often operate with limited direct human oversight, which makes their ethical assumptions difficult to inspect once predictions are translated into automated actions.

The ethical turn in AI and responsible intelligence

The ethical turn in AI has shifted attention from technical performance alone toward questions of trust, reliability, human flourishing, governance, and social legitimacy. Surveys of AI ethics guidelines show broad convergence around principles such as transparency, justice, non-maleficence, responsibility, and privacy, but also reveal challenges in implementation across domains [8]. Responsible AI scholarship further argues that ethical commitments must be translated into methods, tools, and organizational practices rather than remaining aspirational declarations [9]. Energy management warrants a dedicated ethical framework because EMS decisions influence thermal comfort, electricity access, health-related indoor conditions, and the distribution of energy flexibility burdens.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Principles of AI ethics applied to energy management

Classic AI ethics principles can be translated into EMS-specific obligations when energy control is understood as a morally consequential practice. Beneficence requires EMS AI to support human and environmental well-being, while non-maleficence requires avoiding harmful comfort deprivation, unsafe automation, or excessive curtailment of essential energy services [10]. Autonomy requires meaningful user awareness and override options, justice requires fair distribution of costs and comfort, and explicability requires that automated energy decisions be understandable to affected stakeholders. These principles are especially important because energy transitions can intensify justice concerns when technical innovation fails to protect vulnerable users [11].

Fairness, accountability, and transparency in automated energy decisions

Fairness, accountability, and transparency are interdependent ethical requirements for automated energy decisions. In building EMS, fairness may involve avoiding systematically colder or warmer conditions for specific zones, workgroups, tenants, or vulnerable occupants, while accountability requires that responsibility for AI-mediated control remain traceable to human and institutional actors. Transparency is also crucial because explainable AI research in energy and power systems shows that black-box decision systems can undermine trust when users cannot understand why curtailment, dispatch, or comfort adjustments occur [12]. Without these requirements, predictive EMS optimization could become technically effective but ethically opaque.

Socio-technical governance of AI in energy management

Socio-technical governance treats EMS AI as part of an institutional and human system rather than as a self-contained optimization engine. Responsible research and innovation scholarship emphasizes anticipation, inclusion, reflexivity, and responsiveness as governance requirements for emerging digital technologies [13]. In energy management, such governance may include impact assessments, audit logs, human-in-the-loop review, regulatory alignment, vendor accountability, and participatory design with affected stakeholders. This approach is consistent with energy justice work that links technical energy transitions to recognition, procedural fairness, and equitable distribution of benefits and harms [14].

Recent scholarship has further emphasized that ethical AI in energy systems must be governed throughout the entire system lifecycle rather than only during deployment or post-hoc auditing. A lifecycle-oriented ethics approach for energy systems argues that trustworthy AI requires continuous assessment of fairness, transparency, accountability, privacy, and stakeholder impact across design, implementation, operation, and revision stages [15]. Similarly, emerging research on fairness in local energy systems demonstrates that user-centered energy transitions increasingly depend on procedural justice, equitable participation, and fair allocation of energy-related benefits and burdens within decentralized and intelligent energy infrastructures. These perspectives reinforce the argument that responsible EMS intelligence should integrate ethical governance directly into operational and institutional processes rather than treating ethics as an external compliance exercise. As shown in **Figure 1**, responsible AI governance in energy management should be understood as a lifecycle-based socio-technical process in which technical optimization, institutional accountability, stakeholder participation, and energy justice principles are continuously integrated across design, implementation, operation, and revision stages.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

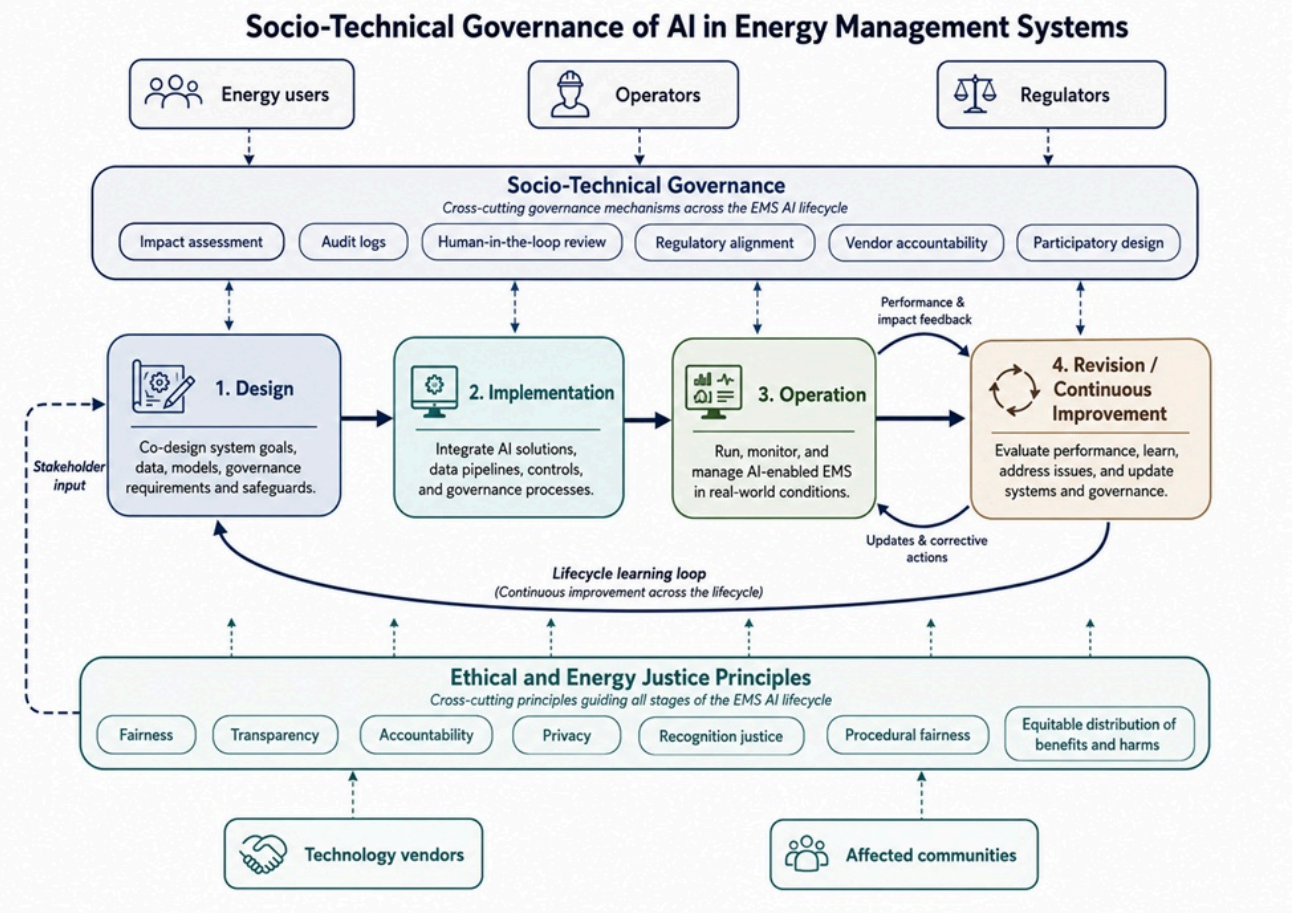

Figure 1. Lifecycle-Based Socio-Technical Governance Framework for AI-Enabled Energy Management Systems

Conceptual Foundation of The Ethical Ai Framework

From prediction to responsible action

The proposed framework begins by distinguishing prediction from responsible action. Predictive optimization asks what is likely to happen and how energy cost, peak load, or operational inefficiency can be minimized, whereas responsible intelligence asks what should be done when technical objectives interact with ethical duties. Reinforcement learning for demand response illustrates this distinction because an algorithm may learn efficient control policies while still requiring normative constraints on comfort, consent, and fairness [6]. Therefore, EMS predictions should be evaluated not only by accuracy or control performance, but by the moral quality of the actions they enable.

Ethical principles as design constraints

Ethical principles should be operationalized as design constraints within the EMS optimization problem rather than treated as post-hoc review criteria. Value sensitive design provides a useful conceptual bridge because it maps human values into technical design requirements and system decision structures [16]. In EMS AI, fairness constraints could prevent discriminatory setpoint allocation, non-maleficence constraints could limit comfort disruption, and autonomy constraints could require user override pathways before automated curtailment. This design-based interpretation helps move ethical AI from general principle statements toward operationally meaningful responsible intelligence.

Table 1 translates broad AI ethics principles into EMS-specific design requirements, showing how values such as fairness, autonomy, transparency, and accountability can become auditable operational commitments rather than abstract

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

declarations.

Table 1. Conceptual Conversion Matrix from AI Ethics Principles to EMS Design Requirements

Ethical principle	EMS-specific ethical risk	Design requirement	Operational mechanism	Evidence or audit indicator
Fairness	Unequal comfort, cost, or demand-response burdens across occupants, tenants, or consumers	Prevent systematic disadvantage in control outcomes	Fairness-aware optimization constraints; subgroup impact testing; zone-level comfort equity checks	Distribution of temperature deviations, curtailment frequency, override requests, and cost impacts across user groups
Non-maleficence	Harmful discomfort, unsafe curtailment, reduced access to essential energy services	Define minimum safety, comfort, and service-protection thresholds	Hard constraints on indoor environmental quality, critical-load protection, and vulnerable-user safeguards	Frequency and duration of threshold violations; incident logs; vulnerable-user impact reviews
Autonomy	Silent automation overriding user preferences or needs	Preserve meaningful awareness, consent, and override capacity	User preference settings, opt-out pathways, escalation rules, manual override options	Override availability, override success rate, consent records, user notification logs
Transparency	Black-box decisions that occupants, managers, or regulators cannot understand	Make EMS decisions explainable to different stakeholders	Technical explanations for operators; plain-language explanations for occupants; model documentation	Explanation availability, explanation completeness, stakeholder comprehension review
Accountability	Responsibility gaps between vendors, operators, facility owners, and algorithms	Assign responsibility for AI-mediated energy decisions	Audit trails, role-based approval procedures, vendor documentation, governance logs	Traceability of each control action to model, rule, operator, or institutional policy
Justice	Efficiency gains achieved by shifting burdens to less powerful users	Include distributional, procedural, and recognition concerns in EMS governance	Participatory design, stakeholder consultation, benefit-sharing mechanisms	Stakeholder participation records, equity impact assessments, complaint and remedy outcomes
Privacy and dignity	Intrusive occupancy sensing or behavioral profiling	Minimize data collection and protect contextual integrity	Data minimization, purpose limitation, privacy-preserving sensing, access controls	Data inventory, retention records, access logs, privacy impact assessment
Human well-being	Optimization of energy performance at the expense of comfort, health, or agency	Treat well-being as a primary EMS performance dimension	Comfort-aware control, health-sensitive thresholds, adaptive preference learning	Comfort satisfaction, indoor environmental quality metrics, health-related complaint trends

Human agency and occupant well-being

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Human agency and occupant well-being must be central commitments in ethical EMS AI because energy management decisions directly shape lived indoor experience. Occupant behavior research has shown that building performance depends not only on physical systems but also on social, psychological, and contextual factors that influence comfort and energy interaction [17]. AI-based thermal comfort control further demonstrates that predictive methods can personalize comfort, but ethical design must ensure that personalization does not become coercive, opaque, or inequitable [18]. Responsible EMS intelligence should therefore support informed consent, meaningful override, adaptive preference learning, and protection from rigid automation.

Framework Components

Ethical reasoning layer

The ethical reasoning layer translates principles such as fairness, non-maleficence, autonomy, transparency, and justice into computable constraints, objectives, and decision rules for EMS operation. It can be conceptualized as a normative interface between predictive models and control decisions, ensuring that forecasts are not converted into actions unless they satisfy ethically defined conditions. Multi-objective EMS optimization can incorporate energy cost, emissions, comfort, resilience, and equity as simultaneously relevant objectives rather than reducing system intelligence to a single efficiency metric [19]. This layer would require further normative validation because the relative priority of values may differ across building types, communities, and regulatory contexts.

The proposed transition from predictive optimization to responsible intelligence is illustrated in **Figure 2**, which positions ethical reasoning, explainability, and accountability as architectural layers between AI prediction and EMS control action.

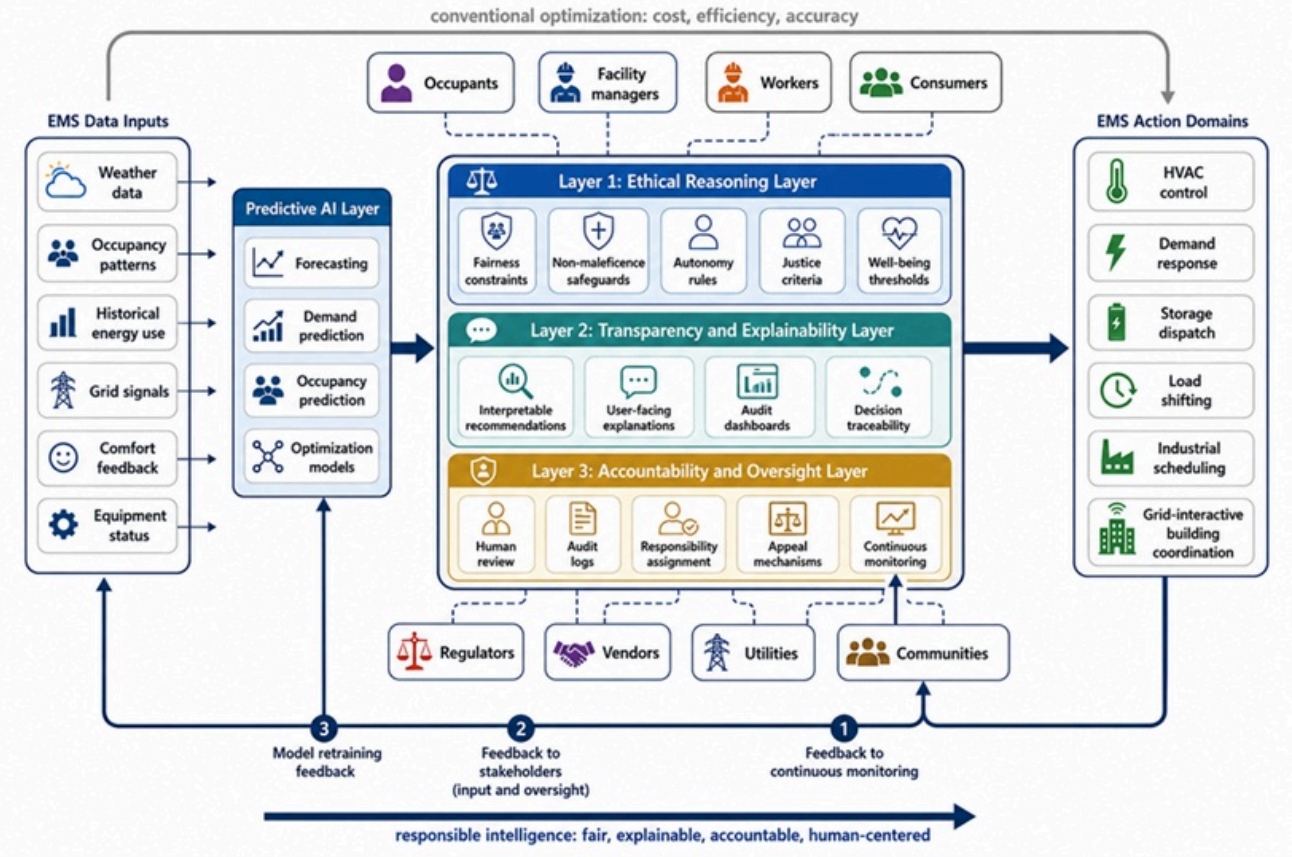

Figure 2. Ethical AI architecture for responsible energy management systems

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Transparency and explainability layer

The transparency and explainability layer makes EMS decisions interpretable to facility managers, occupants, auditors, and regulators. Explainable AI techniques in energy systems can help clarify why a model recommends load curtailment, modifies HVAC schedules, dispatches storage, or prioritizes certain control actions under grid stress [12]. However, interpretability must be tailored to the user because an engineer may need technical feature-level explanations, while an occupant may need plain-language reasons and options for contestation. This layer therefore connects technical explainability with ethical explicability, ensuring that affected users can understand how AI reasoning shapes energy conditions.

Accountability and oversight layer

The accountability and oversight layer assigns responsibility, records decisions, enables human review, and provides mechanisms for appeal and correction. Ethical AI literature warns that reliability alone is insufficient when systems operate in ways that obscure responsibility for consequential decisions [10]. In EMS, accountability may require decision logs showing which model, rule, operator, vendor setting, or institutional policy produced a particular control action. This layer ensures that no automated energy decision remains beyond human and organizational responsibility, especially when comfort, access, cost, or safety are affected.

Responsible Intelligence in Predictive Ems Operations

Fair load forecasting and bias mitigation

Fair load forecasting requires attention to the data conditions under which EMS models are trained, validated, and deployed. AI-based building energy prediction studies often rely on historical consumption, weather, occupancy, and operational datasets, but these data may reflect unequal building quality, uneven access to efficient technologies, or behavior patterns shaped by social disadvantage [20]. If models trained on privileged buildings are transferred without scrutiny, they may normalize comfort expectations, flexibility assumptions, or consumption baselines that do not apply fairly to vulnerable users. Responsible intelligence would therefore require bias detection, contextual data interpretation, and fairness-aware learning without reducing ethical evaluation to technical model performance alone.

Ethical optimization objectives and trade-offs

Ethical EMS optimization must make trade-offs explicit rather than hiding them inside objective functions that privilege energy cost or peak reduction. Demand-side response systems can support grid flexibility, but they may also shift inconvenience, discomfort, or behavioral burdens onto consumers who have unequal capacity to adapt [21]. Energy justice scholarship shows that low-carbon and smart energy transitions can generate inequitable outcomes when distributional, procedural, and recognition dimensions are not incorporated into governance [22]. The proposed framework therefore treats trade-offs among cost, comfort, grid reliability, environmental benefit, and justice as governable ethical choices rather than purely technical calculations.

Table 2 summarizes the major ethical trade-offs that emerge in AI-enabled EMS optimization and the corresponding governance responses proposed within the responsible intelligence framework.

Table 2. Ethical Trade-offs and Governance Responses in AI-Enabled Energy Management Systems

Optimization objective	Potential ethical trade-off	Associated ethical risk	Responsible governance response
Energy cost minimization	Reduced thermal comfort during peak reduction	Disproportionate discomfort for vulnerable occupants	Comfort protection thresholds and fairness-aware optimization

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Aggressive demand response	Automated curtailment without meaningful consent	Loss of occupant autonomy and trust	Opt-out mechanisms, informed consent, and override pathways
Predictive occupancy control	Extensive behavioral and sensing data collection	Privacy intrusion and surveillance concerns	Data minimization and privacy-preserving sensing protocols
Grid flexibility optimization	Unequal distribution of curtailment burdens	Procedural and distributional injustice	Stakeholder participation and subgroup equity assessment
Black-box predictive optimization	Limited interpretability of automated decisions	Accountability gaps and reduced transparency	Explainable AI interfaces and audit-ready decision logs
Autonomous control scaling	Reduced human review of operational decisions	Institutional responsibility ambiguity	Human-in-the-loop oversight and governance documentation

Human-in-the-loop override and consent mechanisms

Human-in-the-loop override and consent mechanisms are necessary because responsible EMS intelligence should assist human decision-making rather than silently replacing it. Occupancy prediction and sensing-based control can improve responsiveness, but these systems may also raise concerns about surveillance, consent, and user autonomy if occupants do not understand how their data are used [23]. Trustworthy EMS AI should therefore provide transparent recommendations, configurable preferences, escalation pathways, and meaningful override rights for facility managers and affected occupants. This approach aligns with broader responsible AI arguments that human flourishing requires systems designed around agency, dignity, and institutional responsibility rather than automation alone [24].

Accountability, Governance, and Stakeholder Inclusion

Algorithmic accountability in energy management

Algorithmic accountability in energy management means that automated decisions about energy use, comfort conditions, demand response, and operational prioritization must remain traceable to responsible human and institutional actors. The ethics of AI ethics literature warns that principles can become performative if organizations adopt them without concrete mechanisms for implementation, monitoring, and responsibility allocation [25]. In EMS contexts, accountability may require audit trails that document model inputs, control recommendations, override decisions, vendor settings, and institutional approval procedures. Such mechanisms would connect AI-mediated energy decisions to identifiable responsibilities rather than allowing predictive optimization to function as an opaque operational authority.

Regulatory compliance and standards

Ethical EMS design should anticipate regulatory and standards-based expectations for trustworthy, explainable, and accountable AI. Reviews of AI explainability and governance in smart energy systems emphasize that energy AI requires governance structures capable of linking technical transparency with regulatory oversight, organizational accountability, and stakeholder trust [26]. Although standards and compliance regimes may differ across jurisdictions, the ethical AI framework proposed here would require EMS vendors and operators to document design assumptions, risk controls, explainability provisions, and human oversight pathways. This anticipatory posture is important because energy management systems increasingly operate within critical infrastructure settings where failures can affect comfort, affordability, resilience, and public legitimacy.

Figure 3 illustrates how ethical EMS design can translate regulatory and standards-based expectations into practical documentation, explainability, risk-control, and human-oversight requirements for trustworthy AI operation in critical

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

energy infrastructure.

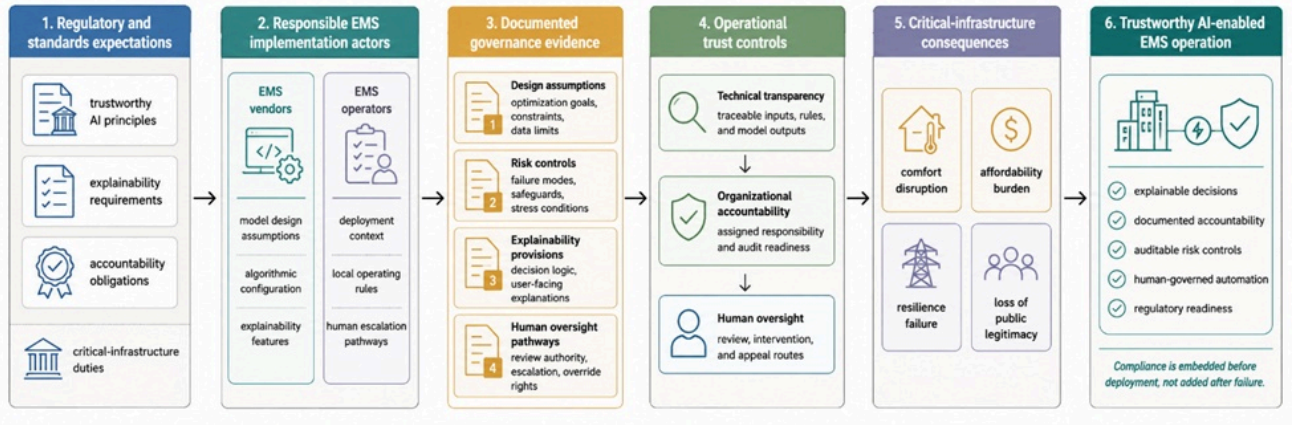

Figure 3. Regulatory Compliance Pathway for Trustworthy and Accountable AI-Enabled Energy Management Systems

Stakeholder engagement and participatory design

Stakeholder engagement is necessary because ethical values in EMS cannot be fully defined by engineers, vendors, or facility owners alone. AI and energy justice scholarship argues that automated energy systems can intensify existing vulnerabilities if affected communities are excluded from decisions about data use, optimization priorities, and system governance [27]. Participatory design could involve building occupants, tenant representatives, facility managers, worker groups, energy communities, and regulators in identifying acceptable comfort ranges, override rights, data boundaries, and fairness priorities. By incorporating stakeholder experience into EMS design, responsible intelligence becomes more legitimate, context-sensitive, and responsive to lived energy needs.

Auditing and continuous monitoring

Auditing and continuous monitoring are essential because ethical EMS performance cannot be assumed at the moment of deployment and then left unchecked. Interpretable models are especially important in high-stakes settings because black-box systems may conceal errors, unfair patterns, or unsafe decision pathways that are difficult to challenge after harm occurs [28]. In EMS, periodic audits could review whether thermal comfort burdens, load curtailment decisions, data collection practices, and override outcomes remain consistent with fairness, transparency, and accountability commitments. Continuous monitoring would therefore transform ethics from a design-time aspiration into an ongoing operational responsibility across the EMS lifecycle.

Table 3 consolidates the manuscript's central theoretical contribution into a maturity model, clarifying how EMS AI can progress from basic automation and predictive optimization toward accountable responsible intelligence.

Table 3. Responsible Intelligence Maturity Model for AI-Enabled Energy Management Systems

Maturity level	EMS intelligence model	Primary optimization logic	Ethical integration	Governance capacity	Typical failure mode	Target advancement
Level 1: Instrumented automation	Rule-based EMS with limited sensing and fixed	Maintain operational stability and predefined setpoints	Ethics implicit or absent	Manual supervision by facility staff	Rigid control ignores diverse occupant needs	Add feedback channels and basic user override

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

	control schedules					
Level 2: Predictive optimization	AI forecasting and optimization for energy cost, demand, or efficiency	Minimize cost, peak load, or energy waste	Ethics treated as external compliance	Technical performance monitoring	Efficient outcomes may create unfair comfort or burden distributions	Introduce ethical performance indicators
Level 3: Constrained optimization	Predictive models operate within comfort, safety, and fairness limits	Optimize only within predefined ethical thresholds	Ethical principles translated into constraints	Logs, thresholds, and exception handling	Constraints may be narrow, static, or poorly justified	Expand stakeholder-informed ethical constraints
Level 4: Explainable responsible EMS	AI decisions are interpretable and contestable by operators and affected users	Balance efficiency, comfort, emissions, and fairness	Ethics embedded in explanations and decision interfaces	Audit dashboards, user-facing explanations, review procedures	Explanations may exist without meaningful remedy	Link explanations to appeal and correction mechanisms
Level 5: Accountable responsible intelligence	EMS AI functions as a socio-technical system governed across its lifecycle	Optimize under ethical, institutional, and community accountability	Ethics embedded into design, deployment, monitoring, and revision	Continuous auditing, responsibility assignment, participatory review, regulatory alignment	Governance complexity may slow deployment or create institutional burden	Develop standardized audit templates, procurement criteria, and regulatory guidance

Application Domains in Energy Management Systems

Building energy management and occupant-centric control

In building energy management, responsible intelligence requires balancing energy efficiency with occupant comfort, health, autonomy, and indoor environmental quality. AI-based thermal comfort prediction can support adaptive control, but it also raises ethical questions about whose comfort is learned, whose preferences are prioritized, and whether vulnerable occupants are protected from discomfort imposed for efficiency gains [18]. Occupant-centric EMS should therefore treat thermal comfort and indoor environmental quality as justice-relevant design concerns rather than as secondary constraints after cost optimization. The framework would support building control strategies that remain explainable, adjustable, and accountable to the people who experience their effects.

Industrial EMS and worker welfare

Industrial EMS applications can optimize energy-intensive processes, equipment scheduling, peak demand management, and production-energy coordination, but such optimization should not undermine worker welfare. Human flourishing approaches to AI suggest that responsible systems must be evaluated according to their effects on dignity, safety, agency,

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

and social well-being, not simply by operational efficiency [24]. In industrial settings, ethical EMS AI should avoid algorithmic speed-ups that intensify work pressure, reduce safe thermal conditions, or obscure the rationale for automated changes in production-energy schedules. The framework therefore requires transparency for workers and managers, institutional responsibility for control policies, and safeguards against efficiency gains achieved through hidden human costs.

Demand response and consumer protection

Demand response programs illustrate why ethical EMS AI must include consumer protection and justice safeguards. AI-enabled demand-side response can coordinate flexible loads and support grid stability, but it may also impose disproportionate inconvenience on households with limited ability to shift electricity use [12]. The justice and equity implications of clean energy transitions show that low-income and vulnerable groups can be harmed when transition benefits and burdens are distributed unevenly [29]. Responsible EMS intelligence should therefore ensure informed participation, fair compensation, opt-out protections, transparent control logic, and special safeguards for consumers who depend on uninterrupted energy services.

Grid-interactive efficient buildings and community impact

Grid-interactive efficient buildings extend EMS decision-making beyond individual facilities by allowing buildings to provide flexibility, load shifting, and distributed support to the wider grid. This creates community-level ethical questions about who benefits from flexibility revenues, who bears comfort adjustments, and how data are shared among building operators, aggregators, utilities, and public authorities. Energy citizenship scholarship emphasizes that participation, empowerment, and inclusion are central to fair energy transitions, especially for energy-poor households and marginalized groups [30]. The framework would therefore require community impact awareness, transparent benefit sharing, and accountable governance when building-level intelligence contributes to grid-level optimization.

Governance Integration and Policy Implications

Policy mechanisms for ethical EMS AI

Policymakers could encourage ethical EMS AI by embedding fairness, accountability, transparency, and human oversight requirements into building codes, energy efficiency standards, incentive programs, and public procurement rules. Energy justice research shows that transitions become more legitimate when procedural justice and recognition are integrated into policy design rather than added after technical deployment [14]. For EMS AI, this could mean requiring documented impact assessments, explanation interfaces, occupant override provisions, and audit-ready control logs as conditions for public subsidies or smart-building certification. Such policy mechanisms would help ensure that intelligent energy management contributes to both decarbonization and social legitimacy.

Industry self-governance and best practices

Industry self-governance can complement public regulation when EMS vendors, facility management firms, and energy service companies adopt responsible AI practices internally. Tools and methods for translating AI ethics into practice show that organizations need concrete procedures, design checklists, risk assessments, and review mechanisms to move from principles to implementation [9]. EMS firms could establish internal ethics review boards, vendor documentation standards, explainability requirements, and post-deployment monitoring protocols for AI-enabled control systems. However, self-governance should not replace external accountability because organizations may face commercial incentives to prioritize optimization performance over fairness, transparency, and user agency.

Evaluation Strategy for The Ethical AI Framework

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Conceptual coherence and ethical soundness

The ethical AI framework should first be evaluated for conceptual coherence, ethical soundness, and internal consistency. AI ethics scholarship cautions that ethical principles can conflict unless their relationships are carefully specified, operationalized, and connected to concrete decision contexts [5]. In EMS, this means examining whether fairness, autonomy, transparency, non-maleficence, and environmental benefit are defined clearly enough to guide design without collapsing into vague aspirations. Ethical soundness would also require assessing whether the framework adequately distinguishes predictive accuracy from morally justified action.

Stakeholder acceptance and legitimacy

Stakeholder acceptance and legitimacy should be evaluated through deliberative processes rather than assumed from technical system performance. The global landscape of AI ethics guidelines shows that trust, accountability, fairness, and transparency recur across ethical frameworks, but their practical meaning depends on institutional and cultural context [8]. In EMS applications, ethical review panels, occupant workshops, facility manager consultations, and regulator discussions could examine whether the framework reflects the values of those affected by automated energy decisions. No empirical results are reported here; rather, the framework suggests deliberative evaluation as a necessary step before practical operationalization.

Transferability across EMS contexts and scales

The framework's transferability should be assessed across single buildings, building portfolios, campuses, districts, industrial facilities, and grid-interactive energy communities. Data-driven energy management research shows that AI methods vary across building types, data environments, control objectives, and operational scales, which means ethical requirements may also require contextual adaptation [1]. A fairness rule appropriate for a multi-tenant office building may not directly apply to an industrial plant, residential demand response program, or energy-poor community. The framework is therefore best understood as a transferable ethical architecture whose specific constraints, oversight mechanisms, and stakeholder processes must be customized to context.

Limitations

Conceptual abstraction and need for empirical operationalization

The main limitation of this article is that the proposed framework remains conceptual and has not been implemented in real EMS software, building automation hardware, or grid-interactive control environments. Responsible AI research emphasizes that moving from ethical principle to practice requires tools, methods, organizational processes, and iterative validation [9]. Future research should translate the ethical reasoning, transparency, and accountability layers into concrete design artifacts, such as EMS procurement criteria, audit templates, explanation interfaces, and fairness-aware control specifications. Such work should test feasibility without reducing ethical adequacy to technical performance metrics alone.

Contextual specificity of ethical norms

A second limitation is that ethical expectations regarding comfort, privacy, fairness, autonomy, and participation vary across cultures, building types, climates, ownership models, and regulatory systems. Energy justice scholarship demonstrates that justice claims are shaped by local vulnerabilities, institutional arrangements, and historically uneven access to energy benefits [3]. Therefore, the framework should not be treated as a universal checklist that automatically resolves every EMS ethics problem. It provides a general structure for responsible intelligence, but its practical interpretation requires contextual deliberation, stakeholder engagement, and normative sensitivity.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Conclusion

This article has proposed an ethical AI framework for energy management systems that guides the transformation from predictive optimization toward responsible intelligence. The framework reframes EMS AI as a socio-technical and morally consequential system rather than a neutral optimization tool.

This shift is necessary because AI-driven energy decisions directly affect human well-being, fairness, autonomy, comfort, and access to essential energy services. Energy management cannot be governed by efficiency metrics alone when automated decisions shape the everyday conditions under which people live, work, and participate in energy transitions.

The framework integrates ethical reasoning, transparency, accountability, and stakeholder inclusion into a coherent response to the ethical deficit in current EMS AI. By embedding values into design constraints, explanation practices, audit mechanisms, and governance processes, it provides a conceptual pathway for developing energy systems that are intelligent and responsible.

Researchers, industry actors, and policymakers should collaborate to operationalize, pilot, and govern ethical AI in energy management. The intelligent buildings and grids of the future should not only optimize energy flows, but also serve people responsibly, fairly, transparently, and with sustained accountability.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 28 Jun 2022 Revised: 02 Sep 2022 Accepted: 19 Sep 2022
Published online: 15 June 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

1. Molina-Solana M, Ros M, Ruiz MD, Gómez-Romero J, Martin-Bautista MJ. Data science for building energy management: a review. Renew Sustain Energy Rev. 2017;70:598-609.

2. Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev. 2018;81:1192-205.

3. Sovacool BK, Burke M, Baker L, Kotikalapudi CK, Wlokas H. New frontiers and conceptual frameworks for energy justice. Energy Policy. 2017;105:677-91.

4. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. 2018;28(4):689-707.

5. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nat Mach Intell. 2019;1(11):501-7.

6. Vázquez-Canteli JR, Nagy Z. Reinforcement learning for demand response: a review of algorithms and modeling techniques. Appl Energy. 2019;235:1072-89.

7. Ahmad T, Chen H, Guo Y, Wang J. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: a review. Energy Build. 2018;165:301-20.

8. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389-99.

9. Morley J, Floridi L, Kinsey L, Elhalal A. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. In: Dubber M, Pasquale F, Das S, editors. Ethics, governance, and policies in artificial intelligence. Cham: Springer; 2021. p. 153-83.

10. Ryan M. In AI we trust: ethics, artificial intelligence, and reliability. Sci Eng Ethics. 2020;26(5):2749-67.

11. McCauley D, Heffron R. Just transition: integrating climate, energy and environmental justice. Energy Policy. 2018;119:1-7.

12. Machlev R, Heistrene L, Perl M, Levy KY, Belikov J, Mannor S, et al. Explainable artificial intelligence (XAI) techniques for energy and power systems: review, challenges and opportunities. Energy AI. 2022;9:100169.

13. Jirotka M, Grimpe B, Stahl B, Eden G, Hartswood M. Responsible research and innovation in the digital age. Commun ACM. 2017;60(5):62-8.

14. Jenkins K, Sovacool BK, McCauley D. Humanizing sociotechnical transitions through energy justice: an ethical framework for global transformative change. Energy Policy. 2018;117:66-74.

15. Limmer S. Empirical study of stability and fairness of schemes for benefit distribution in local energy communities. Energies. 2023;16(4):1756.

16. Umbrello S, Van de Poel I. Mapping value sensitive design onto AI for social good principles. AI Ethics. 2021;1(3):283-96.

17. D'Oca S, Chen CF, Hong T, Belafi Z. Synthesizing building physics with social psychology: an interdisciplinary framework for context and occupant behavior in office buildings. Energy Res Soc Sci. 2017;34:240-51.

18. Ngarambe J, Yun GY, Santamouris M. The use of artificial intelligence (AI) methods in the prediction of thermal comfort in buildings: energy implications of AI-based thermal comfort controls. Energy Build. 2020;211:109807.

19. Ahmad T, Madonski R, Zhang D, Huang C, Mujeeb A. Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: key developments, challenges, and future research opportunities in the context of smart grid paradigm. Renew Sustain Energy Rev. 2022;160:112128.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

20. Wang Z, Srinivasan RS. A review of artificial intelligence based building energy use prediction: contrasting the capabilities of single and ensemble prediction models. Renew Sustain Energy Rev. 2017;75:796-808.

21. Antonopoulos I, Robu V, Couraud B, Kirli D, Norbu S, Kiprakis A, et al. Artificial intelligence and machine learning approaches to energy demand-side response: a systematic review. Renew Sustain Energy Rev. 2020;130:109899.

22. Sareen S, Haarstad H. Bridging socio-technical and justice aspects of sustainable energy transitions. Appl Energy. 2018;228:624-32.

23. Wang W, Chen J, Hong T. Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings. Autom Constr. 2018;94:233-43.

24. Stahl BC, Andreou A, Brey P, Hatzakis T, Kirichenko A, Macnish K, et al. Artificial intelligence for human flourishing—beyond principles for machine learning. J Bus Res. 2021;124:374-88.

25. Hagendorff T. The ethics of AI ethics—an evaluation of guidelines. arXiv [Preprint]. 2019 [cited 2026 Aug 25]. Available from: arXiv:1903.03425.

26. Alsaigh R, Mehmood R, Katib I. AI explainability and governance in smart energy systems: a review. Front Energy Res. 2023;11:1071291.

27. Noorman M, Espinosa Apráez B, Lavrijssen S. AI and energy justice. Energies. 2023;16(5):2110.

28. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206-15.

29. Carley S, Konisky DM. The justice and equity implications of the clean energy transition. Nat Energy. 2020;5(8):569-77.

30. DellaValle N, Czako V. Empowering energy citizenship among the energy poor. Energy Res Soc Sci. 2022;89:102654.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press