REVIEW **Open access**

Responsible, Explainable, and Sustainable AI in Energy Systems: A Critical Review of Ethical Challenges, Governance Frameworks, and Theory Development

Linus Chen[1*], Yuki Sato[2], Olivia Harris[1], Takeshi Mori[2]

Abstract

Artificial intelligence in energy systems is increasingly evaluated against normative criteria of responsibility, explainability, and sustainability. These dimensions shape debates on how AI should be designed, deployed, governed, and theoretically understood in energy transitions. The literature remains fragmented, often treating responsible AI, explainable AI, and sustainable AI as separate research agendas. This fragmentation obscures their interactions, ethical tensions, and shared governance requirements. This critical review synthesises research on responsible, explainable, and sustainable AI in energy systems. It identifies cross-cutting ethical challenges, evaluates governance frameworks, and advances the need for theoretical integration. A targeted critical review of 27 peer-reviewed journal articles published between 2017 and 2025 is conducted. The review is organised thematically around responsibility, explainability, sustainability, ethical challenges, governance frameworks, and theory development. The review reveals recurring tensions between explainability and computational sustainability, fairness and energy efficiency, and technical optimisation and democratic accountability. Existing governance frameworks remain uneven in their ability to address the three dimensions simultaneously. A unified approach to responsible, explainable, and sustainable AI is urgently needed. This review provides a conceptual foundation and research agenda for integrating ethical, technical, environmental, and governance concerns in energy AI.

Keywords Explainable AI, Energy systems, Responsible AI, Sustainable AI, Ethical challenges, Governance frameworks

*Correspondence:
Linus Chen
linus.chen@stanford.edu
[1] Department of Artificial Intelligence for Smart Power Systems and Electric Vehicles, Faculty of Engineering, Stanford University, Stanford, United States
[2] Department of Intelligent Energy Networks and AI Optimization, Graduate School of Engineering, University of Tokyo, Tokyo, Japan

Introduction

Artificial intelligence has become central to the transformation of energy systems, supporting forecasting, demand response, grid optimisation, renewable integration, and operational decision-making. Early reviews of AI in renewable energy emphasised its potential to improve prediction, planning, and system efficiency, yet they also revealed a tendency to frame AI primarily as a technical instrument rather than a normative intervention in energy governance [1]. As energy systems become more data-intensive and automated, responsible AI, explainable AI, and sustainable AI have emerged as three distinct but increasingly interdependent criteria for evaluating whether AI can support just, transparent, and environmentally aligned energy transitions.

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.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

Responsible AI in energy systems concerns fairness, accountability, transparency, human oversight, and institutional responsibility, but these values are not always translated into sector-specific design or governance mechanisms. General AI ethics frameworks have articulated principles such as beneficence, autonomy, justice, explicability, and accountability, yet energy systems raise distinctive issues because AI decisions may affect access to electricity, market participation, infrastructure reliability, and distributional justice [2]. Studies of energy justice and smart grid governance show that digitalisation can reproduce exclusion if vulnerable users are not represented in system design, data collection, or decision-making processes [3].

Explainability and sustainability further complicate the responsible AI agenda because transparent AI models may require additional computation, while computationally efficient models may not always provide interpretable reasoning. Broad surveys of explainable AI show that interpretability methods can make black-box models more intelligible, but the social usefulness of explanations depends on the needs of operators, consumers, regulators, and affected communities [4]. This review therefore examines responsible, explainable, and sustainable AI together, following a critical structure that moves from conceptual foundations to ethical challenges, governance frameworks, theory development, and future research directions [5].

Motivation for Critical Review

A critical review is necessary because energy AI research is expanding faster than its ethical and theoretical integration. Global surveys of AI ethics guidelines show convergence around high-level principles, but they also reveal substantial variation in how such principles are operationalised across domains, institutions, and regulatory contexts [6]. In energy systems, this gap is especially consequential because AI may influence grid reliability, electricity markets, carbon management, infrastructure planning, household demand, and the allocation of energy benefits and burdens.

This review differs from technical AI surveys by asking how responsibility, explainability, and sustainability interact as normative dimensions rather than treating them as separate performance criteria. Mittelstadt argues that principles alone cannot guarantee ethical AI because principles require institutional translation, accountability mechanisms, and context-sensitive interpretation [7]. In the energy domain, this means that AI governance must move beyond generic ethical checklists and address the socio-technical conditions under which AI systems are designed, justified, contested, and revised.

Review Scope and Method

The review focuses on peer-reviewed journal articles published between 2017 and 2025 that address responsible AI, explainable AI, sustainable AI, energy systems, governance frameworks, and theory development. The selected literature includes foundational AI ethics work, general explainability and sustainability studies, and energy-specific reviews on demand response, forecasting, smart grids, maintenance, local energy systems, and building energy modelling. The inclusion of both general AI governance literature and energy-specific research is necessary because energy AI ethics draws on cross-domain concepts while facing sector-specific implications for reliability, justice, and decarbonisation [8].

The thematic analysis was organised around three normative dimensions: responsibility, explainability, and sustainability. Demand-side response research demonstrates how AI can reshape user participation and operational flexibility, making responsibility central to system design rather than an afterthought [9]. Sustainability-oriented AI literature similarly frames AI as both a tool for achieving sustainability objectives and a source of new environmental burdens, which requires examining whether AI contributes to or undermines broader energy transition goals [10].

The review is limited by its conceptual and critical orientation, which means it does not conduct a meta-analysis, statistical comparison, or empirical validation of AI systems. Instead, it synthesises how the literature frames ethical problems, governance mechanisms, and theoretical gaps. This approach is appropriate because Green AI research has shown that

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

computational efficiency and environmental cost are often under-theorised in mainstream AI evaluation, requiring critical interpretation rather than only technical benchmarking [11].

Responsible AI in Energy Systems

Fairness and non-discrimination in energy AI

Fairness in energy AI involves more than avoiding biased predictions; it concerns how algorithmic systems distribute access, risk, cost, and benefit across households, communities, firms, and infrastructures. Research on AI for sustainability warns that AI systems may reinforce existing inequalities when data availability, institutional capacity, or digital access are unevenly distributed [12]. In energy systems, this problem is intensified because smart-grid participation, demand response incentives, and automated energy services may favour users with better connectivity, more flexible consumption, or greater capacity to respond to algorithmic signals.

Accountability and liability mechanisms

Accountability in energy AI requires clarity about who is responsible when algorithmic decisions contribute to service disruption, unfair market outcomes, unsafe operational recommendations, or opaque infrastructure planning. Sustainable AI scholarship argues that responsibility must include both the societal benefits of AI and the environmental costs of designing, training, deploying, and maintaining AI systems [13]. In the energy sector, accountability therefore extends across model developers, utilities, regulators, platform providers, equipment manufacturers, and public institutions that rely on AI-supported decisions.

Human oversight and meaningful control

Human oversight is frequently invoked as a safeguard for responsible AI, yet the literature reveals that oversight can become symbolic if humans lack the expertise, authority, or timely information needed to intervene. Contextual analyses of AI for sustainable energy show that AI systems are often embedded in complex technical and organisational environments where decision rights are distributed across multiple actors [14]. Meaningful control in energy AI must therefore be designed as an institutional capability, not merely as a requirement that a human operator remain nominally “in the loop.”

Responsible AI governance in practice

Responsible AI governance in practice requires translating ethical principles into procurement rules, documentation requirements, accountability procedures, stakeholder participation, and monitoring routines. Energy-specific XAI reviews show that explainability can support responsible governance by making model outputs more understandable, but they also caution that explanation tools must be adapted to the operational context and decision stakes of power and energy systems [15]. A recurring challenge is that governance remains fragmented across technical validation, ethical review, regulatory compliance, and organisational risk management rather than being integrated into a single life-cycle approach.

Explainable AI in Energy Systems

Post-hoc and intrinsic explainability methods

Explainable AI in energy systems includes post-hoc techniques that explain complex models after training and intrinsic approaches that favour interpretable model structures from the outset. Reviews of AI in renewable energy and energy systems show that machine learning has become increasingly important for forecasting, fault detection, optimisation, and control, but model complexity can reduce interpretability for engineers and decision-makers [16]. The critical issue is not

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

only whether an explanation can be generated, but whether it is technically faithful, contextually meaningful, and actionable for the energy decision being supported.

Explainability for operators, consumers, and regulators

Different energy actors require different forms of explanation because operators, consumers, regulators, and system planners face distinct responsibilities and risks. Reviews of artificial intelligence in energy systems indicate that AI applications span operational, strategic, and policy-relevant domains, which means that a single explanation standard is unlikely to be sufficient [17]. Operators may need causal or diagnostic explanations during abnormal events, consumers may need understandable reasons for automated pricing or demand response recommendations, and regulators may need audit-ready documentation that supports accountability.

The Link between explainability and trust

The literature often links explainability to trust, but this relationship is more complex than the assumption that more explanation automatically produces more legitimate AI. Research on explainability and governance in smart energy systems suggests that explanations can support trust when they clarify model logic, uncertainty, limitations, and human responsibility [18]. However, explanations may also create misplaced confidence if they are overly simplified, strategically selective, or disconnected from the institutional conditions that determine whether users can contest or override AI-supported decisions.

Limits and critiques of XAI in energy

The limits of XAI become apparent when explanation is treated as a technical add-on rather than a socio-technical practice embedded in governance. Energy justice scholarship shows that transparency alone cannot resolve exclusion, unequal participation, or asymmetrical control over smart energy infrastructures [19]. Critical review therefore requires asking not only whether AI models can be explained, but whose interests explanations serve, what forms of accountability they enable, and whether they make energy AI more democratic, sustainable, and responsible.

Sustainable AI in Energy Systems

Computational efficiency and green AI

Sustainable AI in energy systems requires attention to the computational resources consumed by model development, training, inference, storage, monitoring, and repeated updating. Green AI argues that model evaluation should include efficiency and environmental cost rather than only predictive performance, which is especially relevant when energy AI is deployed to support decarbonisation while consuming substantial computational resources [11]. In this context, sustainability is not simply an application domain for AI but an evaluation criterion that challenges the assumption that larger, more complex, and more data-intensive models are always preferable.

Carbon-aware and energy-proportional AI

Carbon-aware AI extends the sustainability agenda by asking whether computation can be aligned with low-carbon energy availability, efficient scheduling, and proportional use of resources. Sustainable AI scholarship distinguishes between AI for sustainability and the sustainability of AI itself, emphasising that energy systems research must evaluate both the benefits and burdens of algorithmic intervention [13]. A recurring challenge is that energy AI may support carbon reduction in grid operations while simultaneously increasing demand for data centres, specialised hardware, and continuous computational infrastructure.

Life-cycle sustainability of AI hardware

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

The life-cycle sustainability of AI hardware remains underdeveloped in energy AI research, despite its relevance to embodied emissions, material extraction, equipment turnover, and electronic waste. Lifecycle-oriented AI ethics in energy systems argues that responsibility must extend across design, development, deployment, operation, maintenance, and retirement rather than focusing only on model outputs [20]. This perspective is critical because sustainable AI cannot be achieved through efficient algorithms alone if the infrastructures that support those algorithms remain environmentally opaque.

Ethical Challenges

Tensions between explainability and sustainability

A central ethical tension concerns the relationship between explainability and computational sustainability. Explainable load forecasting research shows that explanation methods can improve transparency in energy prediction, but they may also add computational overhead through repeated model interrogation, feature attribution, scenario analysis, or visualisation layers [21]. The literature therefore reveals a trade-off between making AI more accountable through explanation and making AI more sustainable through computational restraint.

Responsibility vs. efficiency trade-offs

Responsibility and efficiency may conflict when AI systems optimise energy flows, market participation, or demand response without adequately considering fairness, accountability, or user agency. Reviews of fairness in local energy systems show that efficiency-oriented optimisation may create uneven benefits when households differ in flexibility, technology access, consumption patterns, or capacity to participate [22]. Responsible AI in energy must therefore evaluate not only whether systems reduce costs or emissions, but also whether efficiency gains are distributed fairly and governed transparently.

Justice and inclusion across the three dimensions

Justice and inclusion cut across responsible, explainable, and sustainable AI because affected communities must be able to understand, contest, and benefit from energy AI systems. Energy justice research on smart grids demonstrates that digital infrastructures can reproduce exclusion when vulnerable users are not involved in system design or when governance privileges technically capable actors [3]. Explainability can support inclusion only if explanations are accessible to non-experts, while sustainable AI can support justice only if environmental benefits do not come at the cost of digital exclusion or unequal energy burdens.

Emergent risks and unintended consequences

Emergent risks arise when AI systems interact with complex energy infrastructures, institutional rules, market incentives, and user behaviours in ways that cannot be fully anticipated at design time. Reviews of XAI for energy systems maintenance show that interpretability can support fault diagnosis and operational confidence, yet explanations may still be incomplete when models operate under shifting technical conditions or rare failure scenarios [23]. This review identifies unintended consequences as a cross-dimensional ethical challenge because responsibility, explainability, and sustainability can each fail when AI systems are deployed into dynamic socio-technical energy environments.

Table 1 presents a structured synthesis of cross-dimensional ethical challenges in responsible, explainable, and sustainable AI for energy systems, highlighting how fairness, accountability, and transparency tensions emerge across socio-technical deployment contexts.

Table 1. Cross-Dimensional Ethical Challenges in Responsible, Explainable, and Sustainable AI for Energy

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

Ethical challenge	Responsible AI dimension	Explainable AI dimension	Sustainable AI dimension	Energy system implications
Fairness-efficiency conflict	Optimisation may privilege flexible, digitally connected, or high-income users	Explanations may reveal unequal treatment without providing remedies	Efficiency gains may reduce emissions while shifting burdens to less flexible users	Demand response, local markets, and smart tariffs may deepen energy inequality if distributional effects are not governed
Explainability-computation trade-off	Accountability demands more documentation, monitoring, and justification	Post-hoc explanation may require additional model runs and interpretation layers	Added computation may increase energy use and carbon footprint	Transparent forecasting and control systems may become environmentally costly if explanation is not resource-aware
Human oversight overload	Operators may be formally accountable without enough time, authority, or expertise	Complex explanations may exceed the cognitive capacity of users under operational pressure	Continuous monitoring may require persistent computational infrastructure	Grid control, maintenance, and emergency response may rely on symbolic rather than meaningful oversight
Data exclusion and representational bias	Underrepresented users may receive poorer predictions or fewer benefits	Explanations may be less reliable for groups poorly represented in training data	Sustainability metrics may ignore social and geographic variation in energy need	AI-enabled planning and service allocation may overlook vulnerable households or peripheral communities
Accountability diffusion	Responsibility may be split across vendors, utilities, regulators, and platform providers	Explanations may not clarify legal or organisational responsibility	Environmental responsibility may be externalised to cloud, hardware, or supply-chain actors	Failures in automated energy decisions may be difficult to attribute, contest, or correct
Lifecycle opacity	Ethical review may focus only on deployment decisions	Model explanations may omit infrastructure dependencies	Hardware, data-centre, and disposal impacts may remain invisible	Energy AI may appear sustainable at the application level while producing hidden environmental costs
Trust without contestability	Users may be encouraged to accept AI outputs without procedural rights	Explanations may produce confidence without enabling challenge or appeal	Green claims may legitimise opaque systems	Public acceptance of smart grids may be weakened if trust is demanded rather than institutionally earned

Governance Frameworks

Regulatory instruments and standards

Regulatory instruments and standards for AI increasingly emphasise transparency, accountability, safety, fairness, and human oversight, but energy-sector adaptation remains uneven. General AI ethics mapping shows that many guidelines

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

converge around similar principles, yet they differ substantially in enforceability, institutional responsibility, and practical implementation [6]. For energy systems, governance must translate these principles into sector-specific requirements for grid reliability, market fairness, cybersecurity, consumer protection, decarbonisation, and environmental accountability.

Organisational governance mechanisms

Organisational governance mechanisms are essential because many AI risks emerge during procurement, model integration, data stewardship, operational monitoring, and post-deployment adaptation. A lifecycle approach to AI ethics in energy systems emphasises that governance should cover the full system trajectory rather than focusing only on pre-deployment ethical approval or technical validation [20]. Utilities, aggregators, energy platforms, regulators, and technology vendors therefore need governance routines that connect model documentation, human oversight, environmental assessment, and accountability escalation.

Certification and auditing

Certification and auditing can provide structured mechanisms for assessing whether AI systems meet responsibility, explainability, and sustainability requirements. Explainable AI research in urban building energy modelling shows that methods, applications, and evaluation practices vary widely, creating a need for more consistent assessment criteria across energy domains [24]. Auditing should therefore examine not only model performance and interpretability, but also data representativeness, environmental footprint, human oversight arrangements, and mechanisms for appeal or correction.

Gaps in current governance approaches

Current governance approaches remain limited because they often treat ethical, technical, and environmental questions as separate compliance domains. Research on AI-driven energy futures shows that sociotechnical imaginaries promoted by technology actors can shape expectations about automation, intelligence, and system transformation, yet such imaginaries may understate governance tensions and public accountability requirements [25]. Few frameworks integrate responsible AI, explainable AI, and sustainable AI into a single energy-specific governance architecture that addresses both operational performance and normative legitimacy. **Table 2** synthesises and contrasts leading AI governance frameworks in energy systems, highlighting that while most approaches address responsibility, explainability, or sustainability in isolation, only emerging integrated models attempt to unify these dimensions into a coherent governance architecture capable of supporting both operational effectiveness and normative legitimacy.

Table 2. Comparison of Governance Frameworks for Responsible, Explainable, and Sustainable AI in Energy

Framework	Coverage of Responsibility	Coverage of Explainability	Coverage of Sustainability	Energy-Sector Specificity	Key Strengths	Key Gaps
General AI ethics principles	Strong at the level of fairness, accountability, transparency, and human oversight	Moderate, usually framed as transparency or explicability	Weak to moderate, often indirect or environmental only at principle level	Low	Provides shared normative vocabulary across sectors	Often lacks enforceability, sector-specific translation, and operational detail
Responsible AI lifecycle governance	Strong, because it links responsibility to design, deployment,	Moderate to strong when documentation and interpretability	Moderate, especially when environmental impacts are	Moderate	Encourages continuous governance rather than	Requires organisational capacity and may not specify measurable

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

	monitoring, and retirement	are embedded across the lifecycle	included in lifecycle assessment		one-time ethical review	sustainability criteria
XAI-centred governance	Moderate, because explanations can support accountability and oversight	Strong, particularly for model transparency, feature attribution, and decision justification	Weak unless explanation cost is explicitly assessed	Moderate to high in forecasting, control, maintenance, and building energy contexts	Makes black-box energy AI more inspectable for operators and regulators	May overestimate the value of explanation without procedural rights or environmental assessment
Sustainable AI and Green AI frameworks	Moderate, when responsibility includes environmental stewardship	Weak to moderate, unless interpretability is treated as part of efficient model design	Strong, especially for computational efficiency and carbon-aware evaluation	Low to moderate	Reframes AI performance around resource use and environmental impact	May neglect fairness, contestability, and sector-specific accountability
Energy justice and smart grid governance	Strong for inclusion, fairness, participation, and distributional effects	Moderate, when transparency is connected to user agency	Moderate, when sustainability is linked to just energy transitions	High	Grounds AI governance in social consequences and affected communities	Often less detailed on technical XAI methods and computational sustainability metrics
Certification and auditing models	Strong if audits include accountability, bias, documentation, and oversight	Strong if interpretability evidence is part of certification	Moderate if environmental reporting is included	Moderate	Creates repeatable assessment procedures and institutional accountability	Risks becoming checklist-based unless tied to enforcement, public participation, and post-deployment monitoring
Integrated responsible-explainable-sustainable AI governance	Potentially strong across all three dimensions	Potentially strong if explanations are audience-specific and resource-aware	Potentially strong if lifecycle and carbon-aware criteria are embedded	Still emerging	Offers the most coherent pathway for energy AI legitimacy	Underdeveloped theoretically and rarely operationalised in existing energy AI research

Theory Development

Existing theoretical contributions

Theory development in responsible, explainable, and sustainable AI remains uneven, with stronger conceptual foundations in AI ethics and energy justice than in integrated energy AI governance. AI4People provides a broad ethical foundation for trustworthy and socially beneficial AI, while energy justice frameworks highlight distributional, procedural,

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

and recognition concerns in smart energy systems [2]. However, the literature still lacks a mature theory explaining how responsibility, explainability, and sustainability should be balanced when they generate competing design and governance demands.

Towards an integrative conceptual framework

An integrative conceptual framework should treat responsibility, explainability, and sustainability as mutually conditioning dimensions rather than as separate evaluation categories. A domain-specific systematic review of explainable AI in energy forecasting shows that explanation needs vary across tasks, users, and decision contexts, which suggests that theory must connect technical interpretability with institutional accountability and environmental cost [26]. Such a framework would ask whether an AI system is accountable, intelligible, and resource-conscious across its full lifecycle, not merely whether it performs well on a narrow operational task.

Figure 1 presents an integrated responsible–explainable–sustainable AI framework for energy systems. The framework positions responsibility, explainability, and sustainability as mutually dependent dimensions rather than separate evaluation criteria. It shows how cross-dimensional ethical tensions, including fairness–efficiency conflicts, explainability–computation trade-offs, accountability diffusion, and lifecycle opacity, must be addressed through lifecycle governance mechanisms such as auditing, documentation, stakeholder participation, human oversight, and carbon-aware assessment. The framework therefore translates the review's theoretical argument into a practical model for evaluating AI systems across design, deployment, operation, monitoring, and retirement.

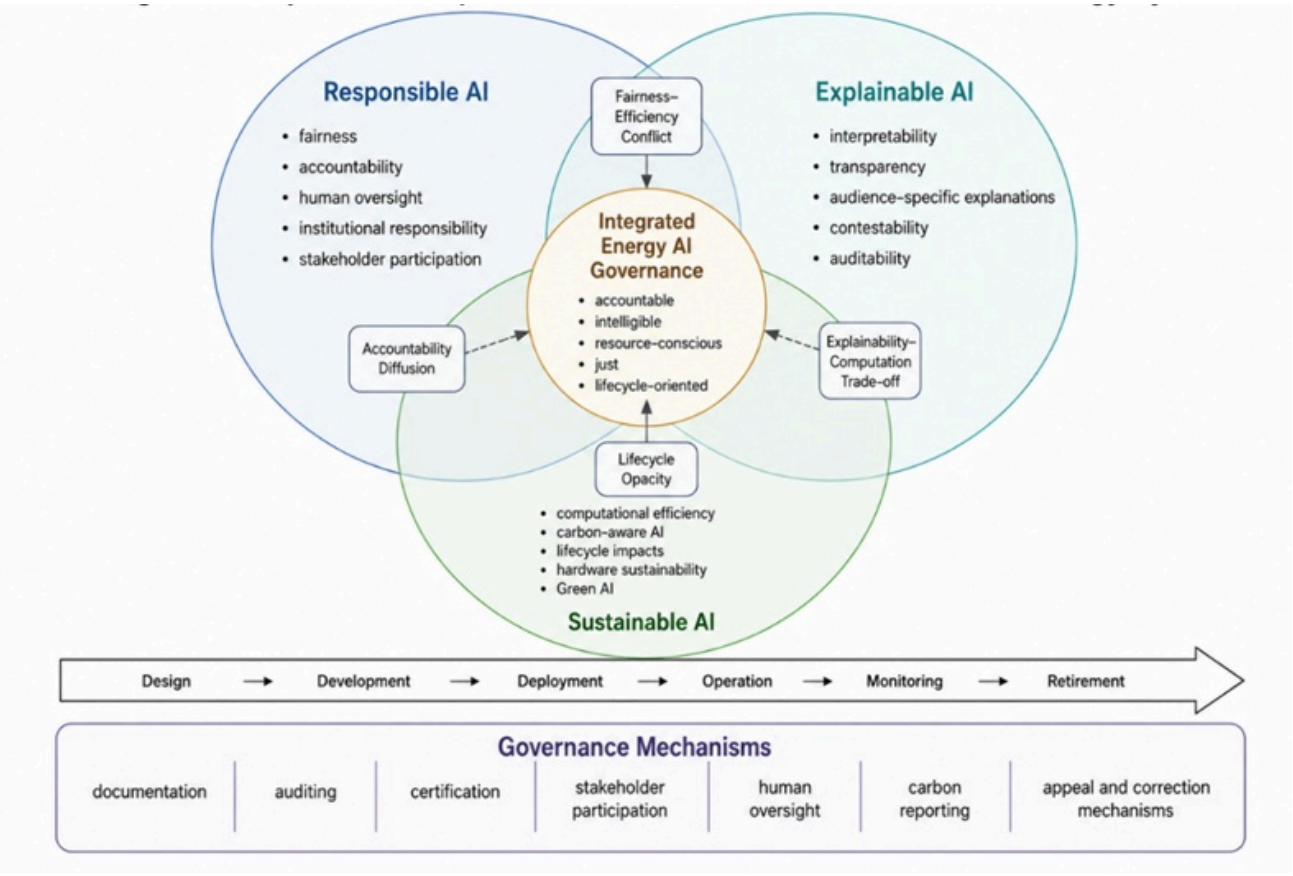

Figure 1. Integrated Responsible–Explainable–Sustainable AI Framework for Energy Systems

The role of interdisciplinary synthesis

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

Interdisciplinary synthesis is necessary because no single field can adequately address the ethical, technical, environmental, legal, and organisational complexity of AI in energy systems. Ethical assessment frameworks for smart grid AI security and ethics point toward the need to connect technical risk, governance accountability, and sector-specific operational constraints [27]. Future theory development should therefore bridge AI ethics, power systems engineering, sustainability science, law, public policy, human-computer interaction, and energy justice to produce concepts that are both analytically rigorous and institutionally usable.

Future Research Agenda

Empirical research priorities

Future empirical research should examine how responsible, explainable, and sustainable AI requirements are interpreted in real energy organisations, including utilities, regulators, aggregators, platform firms, and community energy initiatives. Building on demand-side response research, future studies could investigate whether AI-enabled flexibility programs distribute benefits fairly, provide understandable explanations, and avoid shifting computational or environmental costs into hidden infrastructures [9]. Such studies should be qualitative, comparative, and socio-technical rather than narrowly focused on model accuracy.

Developing integrated design principles

Future design-oriented research should develop integrated principles for energy AI systems that are accountable, explainable, and computationally sustainable from the outset. Explainable AI taxonomies show that interpretability can be built through multiple technical strategies, but energy AI design should also consider when simpler models, hybrid approaches, or resource-aware explanation methods are preferable [8]. The key research task is to develop design principles that make trade-offs visible before deployment rather than discovering ethical conflicts after systems are embedded in critical infrastructure.

Policy and regulatory experimentation

Future policy research should explore regulatory experimentation, controlled pilots, certification schemes, and public-interest audits for responsible, explainable, and sustainable AI in energy systems. The critique that principles alone cannot guarantee ethical AI suggests that governance innovation must test institutional mechanisms capable of enforcing accountability, transparency, and environmental responsibility in practice [7]. Policy experimentation should therefore examine not only compliance rules, but also rights to explanation, rights to contest automated decisions, carbon reporting for AI systems, and public participation in smart grid governance.

Conclusion

This critical review has shown that responsible, explainable, and sustainable AI represent three indispensable but insufficiently integrated dimensions of energy AI governance. Responsible AI foregrounds fairness, accountability, human oversight, and institutional responsibility. Explainable AI addresses the need for intelligible and contestable model behaviour. Sustainable AI broadens evaluation to include computational efficiency, carbon awareness, and life-cycle environmental responsibility.

The review has also shown that these dimensions generate ethical tensions rather than simply reinforcing one another. Explainability can improve accountability while increasing computational burden. Efficiency can support sustainability while intensifying inequity if benefits and burdens are unevenly distributed. Governance frameworks can promote transparency while failing to address lifecycle impacts, public participation, or sector-specific accountability.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

A unified research agenda is now needed to connect AI ethics, energy justice, explainability, sustainability science, policy design, and power systems governance. Future work should develop integrated design principles, empirical studies of implementation, and regulatory experiments that make AI in energy systems accountable, intelligible, and environmentally responsible. Without such integration, energy AI risks becoming technically impressive but normatively incomplete.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 27 Jul 2024 Revised: 02 Oct 2024 Accepted: 18 Oct 2024
Published online: 15 June 2026

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

1. Jha SK, Bilalovic J, Jha A, Patel N, Zhang H. Renewable energy: present research and future scope of artificial intelligence. Renew Sustain Energy Rev. 2017;77:297-317. https://doi.org/10.1016/j.rser.2017.04.018.

2. 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. https://doi.org/10.1007/s11023-018-9482-5.

3. Milchram C, Hillerbrand R, van de Kaa G, Doorn N, Künneke R. Energy justice and smart grid systems: evidence from the Netherlands and the United Kingdom. Appl Energy. 2018;229:1244-59. https://doi.org/10.1016/j.apenergy.2018.08.053.

4. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM Comput Surv. 2018;51(5):1-42. https://doi.org/10.1145/3236009.

5. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138-60. https://doi.org/10.1109/ACCESS.2018.2870052.

6. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389-99. https://doi.org/10.1038/s42256-019-0088-2.

7. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nat Mach Intell. 2019;1(11):501-7. https://doi.org/10.1038/s42256-019-0114-4.

8. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82-115. https://doi.org/10.1016/j.inffus.2019.12.012.

9. 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. https://doi.org/10.1016/j.rser.2020.109899.

10. Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, Domisch S, et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun. 2020;11(1):233. https://doi.org/10.1038/s41467-019-14108-y.

11. Schwartz R, Dodge J, Smith NA, Etzioni O. Green AI. Commun ACM. 2020;63(12):54-63. https://doi.org/10.1145/3381831.

12. Nishant R, Kennedy M, Corbett J. Artificial intelligence for sustainability: challenges, opportunities, and a research agenda. Int J Inf Manage. 2020;53:102104. https://doi.org/10.1016/j.ijinfomgt.2020.102104.

13. Van Wynsberghe A. Sustainable AI: AI for sustainability and the sustainability of AI. AI Ethics. 2021;1(3):213-8. https://doi.org/10.1007/s43681-021-00043-6.

14. Saheb T, Dehghani M, Saheb T. Artificial intelligence for sustainable energy: a contextual topic modeling and content analysis. Sustain Comput Inform Syst. 2022;35:100699. https://doi.org/10.1016/j.suscom.2022.100699.

15. 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. https://doi.org/10.1016/j.egyai.2022.100169.

16. Zhang L, Ling J, Lin M. Artificial intelligence in renewable energy: a comprehensive bibliometric analysis. Energy Rep. 2022;8:14072-88. https://doi.org/10.1016/j.egyr.2022.10.347.

17. Entezari A, Aslani A, Zahedi R, Noorollahi Y. Artificial intelligence and machine learning in energy systems: a bibliographic perspective. Energy Strateg Rev. 2023;45:101017. https://doi.org/10.1016/j.esr.2022.101017.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press

18. Alsaigh R, Mehmood R, Katib I. AI explainability and governance in smart energy systems: a review. Front Energy Res. 2023;11:1071291. https://doi.org/10.3389/fenrg.2023.1071291.

19. Noorman M, Espinosa Apráez B, Lavrijssen S. AI and energy justice. Energies. 2023;16(5):2110. https://doi.org/10.3390/en16052110.

20. El-Haber N, Burnett D, Halford A, Stamp K, De Silva D, Manic M, et al. A lifecycle approach for artificial intelligence ethics in energy systems. Energies. 2024;17(14):3572. https://doi.org/10.3390/en17143572.

21. Baur L, Ditschuneit K, Schambach M, Kaymakci C, Wollmann T, Sauer A. Explainability and interpretability in electric load forecasting using machine learning techniques—A review. Energy AI. 2024;16:100358. https://doi.org/10.1016/j.egyai.2024.100358.

22. Soares J, Lezama F, Faia R, Limmer S, Dietrich M, Rodemann T, et al. Review on fairness in local energy systems. Appl Energy. 2024;374:123933. https://doi.org/10.1016/j.apenergy.2024.123933.

23. Shadi MR, Mirshekali H, Shaker HR. Explainable artificial intelligence for energy systems maintenance: a review on concepts, current techniques, challenges, and prospects. Renew Sustain Energy Rev. 2025;216:115668. https://doi.org/10.1016/j.rser.2025.115668.

24. Darvishvand L, Kamkari B, Huang MJ, Hewitt NJ. A systematic review of explainable artificial intelligence in urban building energy modeling: methods, applications, and future directions. Sustain Cities Soc. 2025;128:106492. https://doi.org/10.1016/j.scs.2025.106492.

25. Guasselli F. AI-driven energy futures: emerging sociotechnical imaginaries in the tech industry. Energy Res Soc Sci. 2025;129:104407. https://doi.org/10.1016/j.erss.2025.104407.

26. Arabzadeh V, Frank R. A four-dimensional analysis of explainable AI in energy forecasting: a domain-specific systematic review. Mach Learn Knowl Extr. 2025;7(4):153. https://doi.org/10.3390/make7040153.

27. Zhang Y, Zhao C, Meng Z, Lai CS, Chen X. An ethical assessment framework for AI security and ethics in smart grid. Glob Energy Interconnect. 2026;9(2):298-314. https://doi.org/10.1016/j.gloei.2025.10.006.

© 2026 The Author(s). | Published by Institute of Smart Energy Systems Press