ORIGINAL RESEARCH **Open access**

A Socio-Technical Governance Framework for Explainable Artificial Intelligence in Decentralized Energy Systems: Conceptualizing Trust, Accountability, and Human Oversight

William Clark[1], Maria Rossi[2*], James Anderson[1]

Abstract

Decentralized energy systems increasingly depend on artificial intelligence to coordinate distributed energy resources, renewable generation, flexible loads, microgrids, prosumers, and peer-to-peer energy exchanges. While AI can improve forecasting, optimization, and real-time coordination, its growing role in energy decision-making also raises significant governance concerns related to opacity, accountability, human oversight, stakeholder trust, and contestability. This article develops a socio-technical governance framework for explainable artificial intelligence in decentralized energy systems. Rather than treating explainability as a purely technical feature, the framework conceptualizes it as a governance mechanism that can support transparency, auditability, responsible intervention, and institutional accountability. Using conceptual synthesis and theory-driven framework development, the article integrates insights from explainable AI, responsible AI, socio-technical systems theory, smart grid governance, decentralized energy systems, and energy justice. The proposed framework is organized around five interrelated dimensions: technical interpretability, human oversight, institutional accountability, stakeholder trust, and socio-energy legitimacy. It shows how AI explanations can help actors understand algorithmic decisions while also enabling review, appeal, audit, and corrective action across distributed energy decision chains. The article argues that trustworthy AI-enabled energy transitions require explanations that are context-sensitive, role-specific, accessible to affected stakeholders, and embedded within accountable institutional arrangements. By positioning explainable AI as part of the broader governance architecture of decentralized energy systems, the framework contributes to debates on responsible AI, smart grid governance, and just energy transitions. It offers a conceptual foundation for designing AI-supported energy systems that remain transparent, accountable, inclusive, and socially legitimate.

Keywords Smart grids, Explainable artificial intelligence, Decentralized energy systems, Socio-technical governance, Trustworthy AI, Human oversight

*Correspondence:
Maria Rossi
maria.rossi@polimi.it
[1] Department of AI for Renewable Energy Forecasting, Faculty of Engineering, University of Manchester, Manchester, United Kingdom
[2] Department of Intelligent Smart Grids and Machine Learning, Faculty of Engineering, Politecnico di Milano, Milan, Italy

Introduction

Decentralized energy systems are increasingly shaped by artificial intelligence because distributed energy resources, renewable generation, microgrids, flexible demand, and prosumer participation require continuous coordination across technically complex environments. AI applications in sustainable energy already support forecasting, demand-side

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

response, energy management, and system optimization, making algorithmic decision-making increasingly relevant to operational planning and real-time coordination [1]. Peer-to-peer and community-based energy models further intensify this complexity because energy decisions are no longer confined to centralized utilities but are distributed across platforms, households, aggregators, and local markets [2]. As these systems become more digitally mediated, the reasoning behind AI-supported decisions can become difficult for human actors to interpret, contest, and govern.

The governance problem arises because opaque AI systems may influence energy access, flexibility participation, pricing, dispatch, curtailment, and market coordination without making their underlying decision logic visible to affected actors. Explainable AI research has shown that black-box models create practical barriers to understanding model behavior, especially when decisions require justification to non-specialist users or institutional decision-makers [3]. In decentralized energy systems, opacity can weaken trust in platform operators, blur responsibility between software vendors and utilities, and create automation dependency among operators who rely on algorithmic recommendations without fully understanding them [4]. These risks are not merely technical, because they affect the legitimacy of energy governance and the ability of regulators, consumers, and communities to challenge automated decisions.

Explainable AI offers a potential response to opacity, but it should not be reduced to post hoc model interpretation or isolated visualization tools. General XAI literature distinguishes explanation, interpretability, transparency, and accountability, indicating that technical explanations are meaningful only when they support human understanding and decision-making in context [5]. In energy systems, explanation must therefore be connected to institutional responsibility, regulatory accountability, human oversight, and stakeholder participation rather than being treated only as a machine learning feature [6]. This means that an explanation is not simply an answer to why a model produced an output; it is also a governance instrument that may enable audit, review, contestation, and responsible intervention.

This article proposes a socio-technical governance framework for explainable AI in decentralized energy systems, organized around trust, accountability, human oversight, technical interpretability, and institutional legitimacy. The framework builds from responsible AI principles that emphasize fairness, transparency, accountability, and human-centered design [7]. It also draws on energy democracy and energy justice scholarship, which shows that decentralized energy transitions should be assessed not only by technical efficiency but also by participation, inclusion, institutional responsibility, and distributional consequences [8]. The central thesis is that explainable AI can support decentralized energy governance only when technical explanations are embedded within accountable institutions, meaningful human oversight arrangements, and participatory energy decision processes.

Background

Decentralized energy systems and the changing logic of grid governance

Energy systems are shifting from centralized generation and one-directional electricity flows toward decentralized, digitalized, renewable-rich, and prosumer-oriented arrangements. Peer-to-peer energy systems demonstrate how connected communities can exchange energy, coordinate flexible resources, and participate in local energy markets, but they also introduce new governance questions concerning platform design, market rules, data access, and responsibility allocation [2]. Microgrid market design similarly shows that decentralized systems require coordination mechanisms that connect technical operation with economic incentives and institutional rules [9]. As distributed energy resources, energy communities, virtual power plants, and flexible demand become more significant, grid governance must address not only physical reliability but also the legitimacy, transparency, and accountability of digitally mediated coordination.

Artificial intelligence in smart and decentralized energy systems

Artificial intelligence plays a growing role in decentralized energy systems by supporting forecasting, optimization, renewable integration, fault detection, demand response, market coordination, and autonomous energy management. Demand-side response research shows that AI and machine learning can help manage flexible consumption, but such

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

applications may also influence consumer participation, comfort, cost allocation, and perceived fairness [10]. In smart grid contexts, machine learning systems can process complex operational and market data, yet their recommendations may carry operational, economic, and social consequences for operators, aggregators, utilities, and consumers [11]. For this reason, AI in decentralized energy systems should be understood as part of a socio-technical decision environment rather than as a neutral computational layer.

Explainable artificial intelligence and interpretability

Explainable artificial intelligence can be defined as a set of concepts, methods, and design practices intended to make AI behavior understandable to human users, decision-makers, and affected stakeholders. XAI literature identifies techniques such as feature attribution, local surrogate explanations, counterfactual reasoning, rule-based explanations, inherently interpretable models, and explanation interfaces as ways to clarify how model inputs relate to model outputs [3]. Surveys of interpretable machine learning emphasize that interpretability differs from broader transparency and accountability because a model can be technically explainable while still being institutionally unaccountable or socially unintelligible [12]. In decentralized energy systems, this distinction matters because an explanation must be appropriate to the decision context, the actor receiving it, and the governance purpose it is expected to serve.

Recent advances in explainable machine learning highlight that explanation quality is not only a property of the model but also a function of user context, task requirements, and decision stakes. Research on local surrogate explanations such as LIME demonstrates how complex model behavior can be approximated through interpretable local models, enabling stakeholders to understand individual predictions in high-dimensional systems [13]. Complementary work on counterfactual explanations emphasizes that meaningful explanations should identify minimal changes to inputs that would alter outcomes, thereby supporting actionable understanding and decision recourse in socio-technical environments [14]. Social science perspectives further argue that explanations must be evaluated in terms of their communicative adequacy for different audiences rather than purely technical fidelity, reinforcing the need for context-sensitive explanation design in critical infrastructure domains such as energy systems [15].

Trustworthy and responsible AI in critical infrastructure

AI used in critical infrastructure requires higher standards of reliability, auditability, safety, fairness, security, and human oversight because failures can affect public welfare rather than only digital user experience. Ethical AI frameworks emphasize that trustworthy AI must align with principles such as beneficence, non-maleficence, autonomy, justice, and explicability, which are especially relevant when algorithmic decisions shape essential services [7]. Human-centered AI further argues that reliable, safe, and trustworthy systems must preserve meaningful human control rather than simply maximizing automation [16]. Energy systems differ from ordinary digital platforms because AI-supported decisions can affect electricity access, affordability, reliability, and social vulnerability, making responsible governance a core design requirement.

Socio-technical systems theory and energy governance

Socio-technical systems theory provides a foundation for understanding energy governance as an interaction between technologies, infrastructures, institutions, rules, markets, human actors, and social values. Energy democracy scholarship argues that energy transitions should be evaluated through participation, accountability, ownership, and democratic control rather than only through decarbonization or efficiency outcomes [17]. This perspective is important for explainable AI because algorithmic transparency alone cannot resolve governance problems if institutional roles, stakeholder rights, and mechanisms for contesting decisions remain unclear. Therefore, explainable AI governance in decentralized energy systems must integrate technical model transparency with the institutional settings in which AI decisions are produced, interpreted, authorized, and challenged.

Conceptual Foundation of the Framework

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

Explainability as a governance function

Explainability can be conceptualized not only as a technical method for interpreting model outputs but also as a governance function that enables accountability, contestability, informed oversight, and institutional learning. The XAI literature shows that explanation methods can make black-box models more understandable, but their value depends on whether they help users make justified decisions in practice [18]. In decentralized energy systems, explanations must be meaningful to different actors, including grid operators who need operational clarity, regulators who need auditability, utilities and aggregators who need responsibility mapping, and prosumers who need accessible reasons for decisions that affect them [6]. Explainability therefore becomes a bridge between algorithmic decision-making and socio-technical governance.

Trust as a multi-level socio-technical relationship

Trust in AI-enabled decentralized energy systems should be understood as a multi-level socio-technical relationship rather than as simple confidence in model accuracy. Stakeholders may need to trust the model's logic, the quality and governance of data infrastructure, the organization deploying the system, the fairness of market outcomes, and the capacity of regulators to intervene when harm occurs [19]. Human-centered AI scholarship suggests that trust depends on reliability, safety, transparency, and meaningful user control, which are institutional and social conditions as much as technical features [16]. In this framework, trust emerges when explanations are understandable, oversight is credible, accountability is visible, and energy outcomes are perceived as legitimate.

Accountability across distributed decision chains

Accountability becomes difficult in decentralized energy systems because decision-making is distributed across data providers, software developers, platform operators, utilities, aggregators, grid operators, regulators, communities, and consumers. Blockchain-enabled and peer-to-peer energy systems illustrate how decentralized architectures can create new opportunities for coordination while also complicating responsibility when automated platforms mediate transactions and operational decisions [20]. Responsible AI scholarship warns that ethical principles must be translated into concrete tools, practices, and institutional processes if accountability is to move beyond abstract commitment [21]. A governance framework for explainable AI must therefore clarify responsibility across the full decision chain, from data collection and model design to deployment, monitoring, appeal, and corrective action.

Framework Components

Technical interpretability layer

The technical interpretability layer concerns the capacity of an AI system to provide intelligible reasons for its recommendations, classifications, forecasts, or control actions in an energy decision context. This layer may include model transparency, feature attribution, counterfactual reasoning, uncertainty communication, interpretable model architectures, and explanation interfaces tailored to operational tasks [22]. In smart grid applications, interpretable machine learning can help clarify which load, network, market, weather, or distributed resource variables influence model outputs, thereby supporting more informed decision-making by operators and energy managers [11]. Conceptually, this layer answers the question of why an AI system produced a particular recommendation, but it does not by itself determine who is responsible for accepting, rejecting, or acting on that recommendation.

Figure 1 presents the proposed socio-technical governance architecture linking AI-supported energy decisions to explainability, human oversight, institutional accountability, stakeholder trust, and socio-energy legitimacy.

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

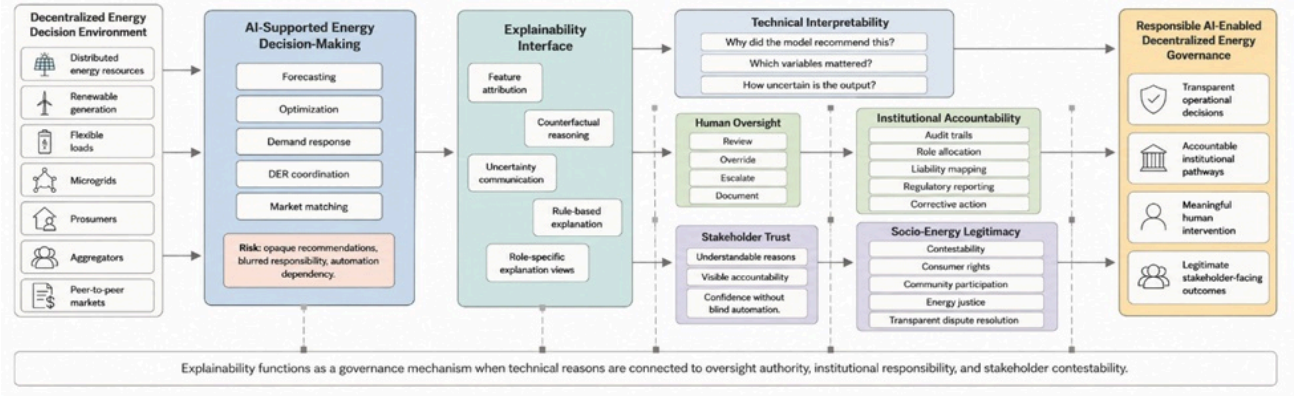


Figure 1. Socio-Technical Governance Architecture for Explainable AI in Decentralized Energy Systems

Human oversight layer

The human oversight layer defines how people remain meaningfully involved in AI-supported energy decisions through human-in-the-loop, human-on-the-loop, and human-in-command arrangements. Human-centered AI emphasizes that reliable and trustworthy systems should support human judgment, intervention, and control rather than replacing these functions with unchecked automation [16]. In decentralized energy systems, operators, regulators, energy managers, and community representatives may use explanations to review, validate, override, or contest AI recommendations when decisions affect reliability, affordability, participation, or access [10]. This layer therefore transforms explanations into practical oversight resources rather than leaving them as passive technical artifacts.

Institutional accountability layer

The institutional accountability layer connects AI explanations to organizational responsibility through audit trails, governance protocols, role allocation, liability structures, documentation requirements, escalation mechanisms, and regulatory reporting. AI ethics research has shown that principles often remain ineffective when they are not translated into operational governance tools and accountable institutional practices [23]. In energy systems, documentation and auditability are especially important because AI-supported decisions can be embedded within complex relationships among utilities, aggregators, vendors, regulators, and consumers [4]. This layer ensures that explanations are not isolated outputs but part of a traceable governance process that identifies who authorized a model, who monitored its use, who intervened when needed, and who remains accountable for outcomes.

Explainability and Human Oversight Architecture

Explanation design for energy decision contexts

AI explanations should be designed differently across energy decision contexts because demand response, distributed energy resource coordination, microgrid control, peer-to-peer trading, and renewable forecasting involve different actors, risks, and governance questions. Demand response explanations may need to clarify why certain consumers or devices are selected for flexibility actions, while peer-to-peer trading explanations may need to clarify pricing, matching, settlement, and market participation decisions [24]. In technical operations, explanation interfaces may support grid operators by showing influential variables and uncertainty, whereas regulators may require policy-oriented explanations that connect decisions to rules, obligations, and accountability pathways [25]. For prosumers and communities, explanations should be accessible, timely, and relevant to participation, cost, fairness, and local energy priorities.

Table 1 translates explainability from a technical model feature into distinct governance functions across decentralized energy decision contexts.

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

Table 1. Governance Translation Matrix for Explainable AI in Decentralized Energy Decision Contexts

Decentralized energy decision context	Primary AI-supported decision	Core explanation question	Main affected actors	Governance risk if explanation is absent	Explanation form with highest governance value	Human oversight requirement
Demand response and flexible load management	Selection of consumers, buildings, devices, or loads for flexibility activation	Why was this participant or device selected for flexibility action?	Consumers, prosumers, aggregators, utilities, grid operators	Perceived unfairness, reduced participation, discomfort, loss of autonomy, hidden cost shifting	Plain-language reason codes linking activation to peak demand, congestion, renewable variability, or price signals	Ability to review, defer, override, or appeal flexibility requests when welfare or fairness concerns arise
Distributed energy resource coordination	Dispatch, curtailment, storage use, EV charging, battery cycling, or local balancing	Why was this asset dispatched, curtailed, charged, or reserved?	Prosumers, aggregators, microgrid operators, utilities, grid operators	Asset misuse, unclear benefit allocation, disputed curtailment, distrust in aggregators	Feature-based and counterfactual explanations showing congestion, forecast uncertainty, local price, resilience, or network constraints	Operator or asset-owner review for high-impact recommendations and abnormal system conditions
Peer-to-peer and transactive energy markets	Pricing, matching, settlement, flexibility valuation, and participation access	Why did this transaction, price, settlement, or matching outcome occur?	Prosumers, platform operators, market participants, regulators, communities	Opaque market power, unfair matching, settlement disputes, exclusion from participation	Transaction-level explanation linking matching logic, pricing rules, constraints, and market conditions	Review of disputed transactions and capacity to suspend or correct automated market outcomes
Microgrid and local energy community management	Local dispatch, shared-resource allocation, resilience prioritization, and community energy scheduling	Why was this local energy decision prioritized over another option?	Community members, microgrid operators, local authorities, vulnerable consumers	Exclusion, perceived illegitimacy, unequal benefit distribution, weak	Community-facing explanation connecting technical constraints to collective priorities and	Community or governance-board review for decisions affecting shared benefits, access, or resilience

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

				democratic control	distributional consequences	
Renewable forecasting and balancing	Forecast-informed scheduling, reserve activation, curtailment, or balancing recommendation	What uncertainty or system condition drove the recommendation?	Grid operators, utilities, aggregators, regulators, energy managers	Automation dependency, poor response to uncertainty, reliability risk, unclear responsibility for forecast-driven actions	Uncertainty-aware explanations showing forecast confidence, influential weather variables, and operational consequences	Human-on-the-loop monitoring with escalation during high-uncertainty or high-impact conditions
Regulatory and consumer-protection supervision	Review of automated energy decisions for fairness, reliability, transparency, and compliance	Can the decision pathway be reconstructed and justified?	Regulators, utilities, platform operators, consumers, communities	Unreviewable automation, weak enforcement, regulatory blind spots, limited consumer rights	Audit-oriented explanation records linking model inputs, outputs, actions, human interventions, and final outcomes	Regulatory review authority and power to request documentation, correction, or suspension

Human-in-the-loop governance of AI decisions

Human-in-the-loop governance can help prevent excessive automation dependency in decentralized energy systems by ensuring that AI recommendations remain reviewable, contestable, and overrideable. Explainable AI techniques may support this process by presenting reasons for recommendations in ways that allow human decision-makers to examine assumptions, detect anomalies, and assess whether automated outputs are appropriate for the operational context [5]. In energy management, this is especially important when AI decisions affect grid reliability, consumer welfare, flexibility participation, market access, or the distribution of costs and benefits [1]. Human oversight should therefore be designed as a substantive governance process with authority, competence, and escalation pathways, not as a symbolic compliance step.

Auditability, traceability, and explanation records

Auditable explanation records are needed to document model inputs, decision pathways, explanation outputs, human interventions, and final decisions in AI-enabled decentralized energy systems. XAI research indicates that explanations can support accountability only when they are preserved in forms that allow later review, comparison, and institutional learning [26]. Traceability is particularly important in local energy markets and transactive energy systems because automated decisions may involve multiple parties, contractual relationships, platform rules, and settlement processes [27]. By linking explanations to logs, review procedures, and escalation records, decentralized energy organizations can support internal governance, regulatory review, post-incident analysis, and public accountability.

Figure 2 illustrates how auditable explanation records connect AI model inputs, decision pathways, explanation outputs, human interventions, and final decisions to governance logs, escalation records, settlement review, and regulatory accountability in decentralized and transactive energy systems.

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

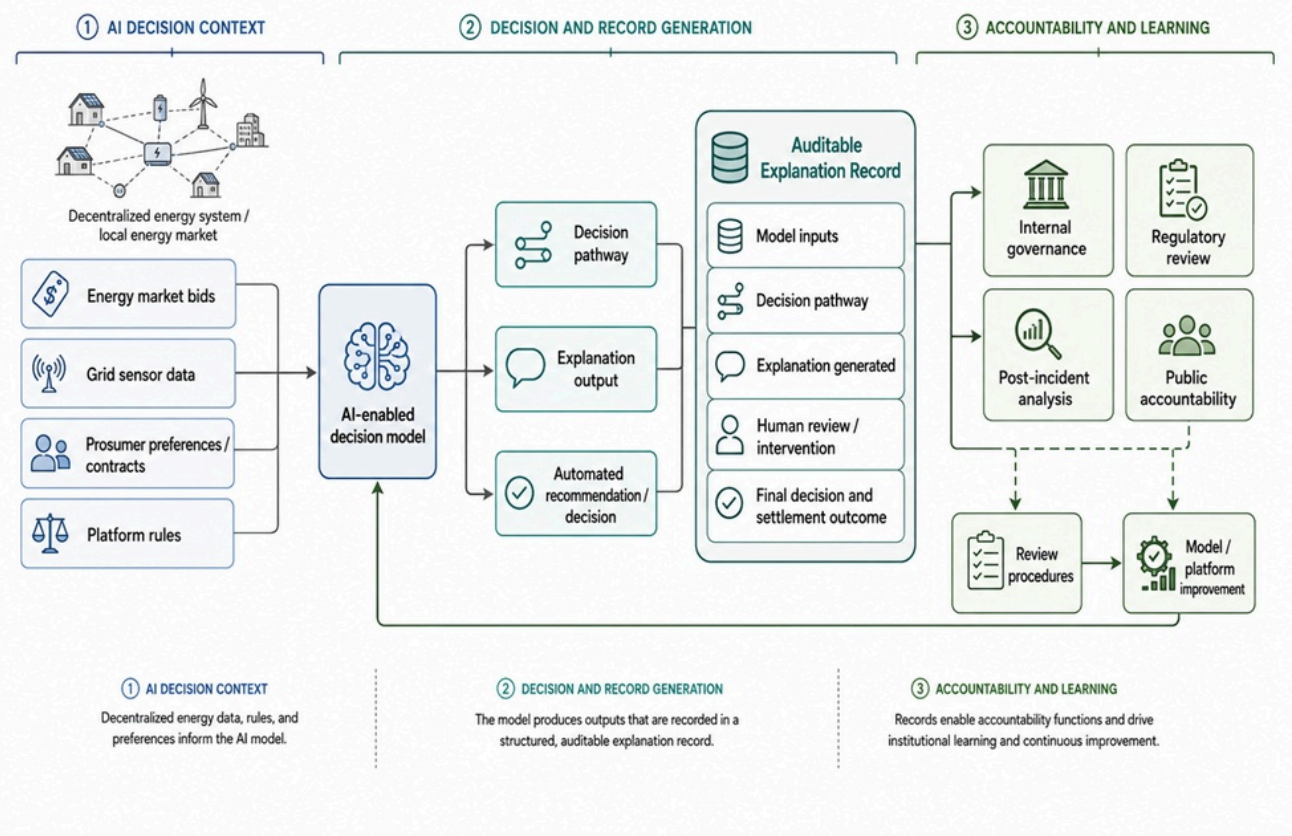


Figure 2. Auditable Explanation Records for AI-Enabled Decentralized Energy Decisions

Trust, Accountability, and Institutional Responsibility

Building trust through meaningful explanations

Meaningful explanations can support trust by allowing stakeholders to understand how AI-supported energy decisions are made, why a recommendation was produced, and what options remain available for human review. In smart grid stability and energy forecasting contexts, explainable AI can help clarify which system variables contribute to a recommendation, but such technical transparency becomes trustworthy only when it is understandable and actionable for the relevant actor [28]. For operators, meaningful explanation may involve causal or feature-based reasoning, while for consumers and prosumers it may involve accessible reasons for pricing, curtailment, flexibility requests, or participation outcomes. Therefore, trust should be built through explanations that are timely, relevant, role-specific, and connected to mechanisms for questioning or correcting automated decisions.

Accountability gaps in decentralized AI decision-making

Accountability gaps emerge when AI-supported decisions are distributed across utilities, aggregators, platform providers, prosumers, regulators, and automated control systems without clear responsibility pathways. Local energy market research shows that peer-to-peer, community self-consumption, and transactive energy models often involve multiple interacting actors, making it difficult to determine who is responsible for market design, algorithmic matching, settlement, and dispute resolution [27]. Emerging business models in decentralized energy markets further complicate accountability because technical platforms, contractual arrangements, and community participation structures may not align neatly with existing regulatory categories [29]. A socio-technical governance framework must therefore map responsibility across technical design, operational deployment, organizational decision-making, regulatory supervision, and user-facing consequences.

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

Human oversight and the limits of automation

Fully automated decision-making has limits in decentralized energy systems because AI-supported decisions may affect reliability, affordability, access, equity, safety, and public confidence. Interpretable forecasting and smart grid models can support operational decision-making, but they should not remove human responsibility for reviewing high-impact recommendations or responding to abnormal system conditions [30]. Human oversight should therefore be designed as a substantive governance process that includes authority to intervene, competence to interpret explanations, and institutional procedures for escalation. In this sense, human oversight is not merely a safeguard against technical failure but a condition for maintaining accountable energy governance under increasing automation.

Contestability and stakeholder rights

Contestability is essential in AI-enabled energy systems because consumers, prosumers, and communities may be affected by automated pricing, demand response, grid access, flexibility activation, or market participation decisions. Energy democracy scholarship suggests that legitimate energy transitions require participation, transparency, and democratic control, which implies that affected actors should have meaningful ways to question and challenge algorithmic decisions [31]. Explainable AI can support contestability by providing reasons that make appeal, correction, review, and participatory governance possible rather than leaving stakeholders dependent on opaque platform authority. In decentralized energy settings, stakeholder rights should therefore include access to understandable explanations, clear complaint channels, and accountable review procedures.

Application Domains in Decentralized Energy Systems

Explainable AI for demand response and flexible load management

Explainable AI can support demand response by clarifying why certain consumers, devices, buildings, or loads are selected for flexibility actions. Machine learning approaches to demand-side response can coordinate flexible consumption, but participation may depend on whether consumers understand how decisions affect comfort, cost, autonomy, and fairness [10]. Explanations could show whether a flexibility request is linked to peak demand, renewable variability, network constraints, price signals, or local balancing needs. By making these reasons visible, explainable AI may improve trust, participation, perceived fairness, and consumer acceptance in decentralized demand response programs.

Explainable AI for distributed energy resource coordination

Explainable AI can support coordination among rooftop solar, batteries, electric vehicles, heat pumps, microgrids, and virtual power plants by clarifying the reasoning behind dispatch, curtailment, storage, charging, or balancing recommendations. Artificial intelligence in sustainable energy systems already contributes to renewable integration and operational optimization, but decentralized resource coordination requires explanations that help human actors understand system-level trade-offs [1]. For example, an explanation may indicate whether a battery dispatch recommendation is driven by network congestion, forecast uncertainty, local price signals, or resilience needs. Such explanations could support operators and aggregators while also helping prosumers understand how their assets are being used within wider grid governance arrangements.

Table 2 summarizes how different XAI methods align with governance needs across distributed energy resource coordination tasks.

Table 2. Mapping of Explainable AI Methods to Governance Requirements in Distributed Energy Resource Coordination

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

XAI Method	Core Mechanism	Strength in Energy Systems	Governance Value	Limitation
LIME [13]	Local surrogate models	Explains individual dispatch or prediction decisions	Supports operator-level interpretability and debugging	Local fidelity may not reflect global system behavior
SHAP	Feature attribution	Quantifies feature influence in forecasts and control decisions	Enhances transparency for regulators and auditors	Computationally expensive in real-time systems
Counterfactual explanations [14]	Minimal input changes for different outcomes	Shows how demand response or dispatch decisions could change	Supports contestability and user recourse	May propose unrealistic operational scenarios
Rule-based explanations	Symbolic decision rules	Maps grid constraints and pricing logic	High interpretability for non-technical stakeholders	Limited scalability in complex grids
Intrinsic interpretable models	Transparent model structures	Directly understandable forecasting and control logic	Strong baseline accountability	May underperform complex deep learning models

Explainable AI for peer-to-peer and transactive energy markets

Peer-to-peer and transactive energy markets require transparent governance because automated systems may shape pricing, matching, settlement, flexibility valuation, and market participation opportunities. Reviews of peer-to-peer and community-based markets show that local energy trading depends on technical architectures, market rules, participant incentives, and institutional arrangements that must be understandable to users and regulators [24]. Blockchain-enabled energy systems may increase traceability, but technological decentralization does not automatically guarantee fairness, accountability, or meaningful explanation [20]. Explainable AI can therefore function as a governance layer that makes market decisions more transparent, supports dispute resolution, and links automated trading logic to institutional accountability.

Explainable AI for energy communities and prosumer governance

Explainable AI can support energy communities by making collective energy decisions more transparent, inclusive, and aligned with local priorities. Energy justice research warns that low-carbon transitions can reproduce exclusion, unequal access, and distributional burdens when governance does not explicitly address fairness and vulnerability [32]. In community energy settings, explanations may help residents understand how shared assets are dispatched, how benefits are allocated, why certain flexibility actions are requested, or how local priorities are balanced against grid constraints. This makes explainability relevant not only to technical performance but also to democratic participation, social legitimacy, and the fair governance of prosumer communities. **Figure 3** illustrates how explainable AI can translate technical energy-community decisions into transparent, resident-facing governance information that supports fairness, participation, and prosumer trust.

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



Figure 3. Explainable AI–Enabled Governance Pathway for Fair and Participatory Energy Communities

Governance Integration and Policy Implications

Embedding explainable AI into energy regulation

Regulators could embed explainability requirements into energy market rules, grid codes, consumer protection frameworks, and AI governance standards for decentralized energy systems. Global AI ethics research shows that principles such as transparency, accountability, fairness, and human oversight are widely recognized, but they require sector-specific translation to become operationally meaningful [19]. In energy regulation, this translation could connect explanation duties to auditability, reliability, cybersecurity, non-discrimination, consumer rights, and dispute resolution. Explainability should therefore be treated as part of regulatory infrastructure for AI-enabled energy governance rather than as an optional technical feature selected by system developers.

Organizational governance for utilities, aggregators, and platform operators

Utilities, aggregators, and energy platform operators can institutionalize explainable AI through governance boards, model documentation, human review protocols, accountability registers, stakeholder engagement processes, and post-deployment monitoring. Responsible AI tools and methods are most useful when they translate high-level ethical principles into everyday organizational practices, including documentation, review, evaluation, and accountability procedures [21]. In decentralized energy contexts, organizational governance should define who approves AI models, who interprets explanations, who can override automated recommendations, and who responds to stakeholder complaints. This can help translate technical explainability into responsible operational practice across utilities, aggregators, vendors, regulators, and communities.

Table 3 consolidates the framework's accountability logic by connecting governance layers, responsible actors, evidence artifacts, and failure modes.

Table 3. Socio-Technical Accountability Architecture for Explainable AI Governance

Governance layer	Core governance purpose	Responsible actors	Required evidence artifact	Decision authority that must be explicit	Failure mode if poorly governed	Practical governance safeguard	Co ac ar
Technical interpretability	Make AI outputs intelligible in	AI developers, data scientists,	Model card, explanation	Who validates	Technically accurate but	Pre-deployment	epi

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

	relation to model inputs, uncertainty, assumptions, and operational constraints	energy system engineers, model owners	report, feature-attribution record, uncertainty statement, interface documentation	whether the explanation is technically adequate for the energy decision context	unusable explanations; misleading simplification; hidden uncertainty	explanation testing with operators, regulators, and stakeholder representatives	ma rec
Human oversight	Ensure that AI recommendations remain reviewable, overrideable, and connected to competent human judgment	Grid operators, energy managers, utility teams, aggregators, community representatives, regulators	Human review log, override record, escalation protocol, intervention justification	Who can accept, reject, pause, override, or escalate an AI-supported decision	Symbolic human-in-the-loop design; automation dependency; unclear intervention responsibility	Tiered oversight rules based on risk, impact, uncertainty, and stakeholder consequences	h a a d
Institutional accountability	Translate explanations into organizational responsibility, liability, documentation, and corrective action	Utilities, platform operators, vendors, aggregators, regulators, governance boards	Accountability register, audit trail, liability map, regulatory report, corrective-action file	Who is accountable for model approval, deployment, monitoring, and harm response	Responsibility gaps across vendors, operators, utilities, and regulators	Named role allocation, audit requirements, incident review, and post-deployment accountability review	r v de
Stakeholder contestability	Enable affected actors to question, appeal, correct, or challenge AI-supported energy decisions	Consumers, prosumers, community organizations, regulators, ombuds bodies, platform operators	Explanation notice, appeal record, complaint pathway, dispute-resolution file	Who receives complaints, who reviews them, and who can remedy the decision	Explanations without rights; inaccessible complaint channels; procedural exclusion	Plain-language explanations, accessible appeal mechanisms, transparent review timelines	exp pr
Socio-energy legitimacy	Align AI-enabled energy governance with fairness, participation, inclusion, and democratic accountability	Regulators, community energy organizations, local authorities, utilities, civil society actors	Participation record, energy justice assessment, distributional-impact review, community consultation report	Who defines acceptable trade-offs among reliability, efficiency, affordability, autonomy, and fairness	Technically optimized but socially contested decisions; unequal distribution of burdens and benefits	Participatory governance, equity review, community reporting, and transparent benefit-allocation rules	pu g e ar
Continuous governance learning	Improve AI governance over time through monitoring,	Model owners, operators, regulators, auditors,	Monitoring dashboard, post-incident review, model-	Who decides when explanations, models, or	Static governance; outdated explanations;	Scheduled audits, model drift review, stakeholder	En a re

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

	incident analysis, stakeholder feedback, and institutional adaptation	community representatives, researchers	update record, stakeholder feedback log	governance procedures must be revised	repeated failures without institutional learning	feedback cycles, governance revision procedures	as er

Evaluation Strategy for the Conceptual Framework

Conceptual coherence and theoretical contribution

The proposed framework should be evaluated for conceptual coherence, theoretical clarity, internal consistency, and contribution to the literature on AI governance, socio-technical systems, and decentralized energy transitions. Explainable AI taxonomies provide a basis for assessing whether the framework distinguishes technical interpretability, transparency, explainability, and accountability with sufficient precision [15]. Energy democracy scholarship can also help evaluate whether the framework connects AI governance to participation, institutional legitimacy, and democratic control rather than treating explanation as a purely computational problem [27]. A coherent framework would show how technical explanations, human oversight, institutional responsibility, stakeholder trust, and socio-energy legitimacy operate as interdependent governance dimensions.

Stakeholder relevance and practical applicability

The framework could be assessed through expert review, stakeholder deliberation, regulatory workshops, or design-oriented evaluation without claiming empirical results at this conceptual stage. Human-centered AI emphasizes that trustworthy systems should be evaluated in relation to users, contexts, risks, and institutional settings rather than only through technical performance metrics [16]. Stakeholder-oriented assessment could examine whether the framework addresses the needs of operators, regulators, utilities, aggregators, communities, and prosumers who interpret AI explanations for different purposes. Such evaluation would help determine whether the framework can guide practical governance design while remaining adaptable to diverse decentralized energy settings.

Governance robustness and transferability

The framework could be evaluated for governance robustness across microgrids, energy communities, virtual power plants, demand response platforms, and peer-to-peer energy markets. Reviews of decentralized and transactive energy models show that local energy systems differ in ownership structure, market design, platform architecture, and regulatory context, which means that explainable AI governance must be transferable without becoming overly generic [29]. Robustness would depend on whether the framework can clarify responsibility, preserve human oversight, support contestability, and maintain meaningful explanation across different institutional environments. Transferability should therefore be assessed across jurisdictions, ownership models, community arrangements, and regulatory regimes.

Limitations

Conceptual scope and lack of empirical validation

This article develops a conceptual framework and does not provide empirical validation, technical implementation, or quantitative testing. Although explainable AI and interpretable machine learning research offers methods that could be applied in smart grid and energy management contexts, this framework does not test any model, dataset, deployment, or organizational process [12]. Future research should examine how the framework performs in real-world AI-enabled energy systems, regulatory settings, and organizational contexts. Such work could explore whether the proposed layers actually support trust, accountability, oversight, and contestability when implemented in operational energy environments.

Context dependence and governance complexity

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

The framework may require adaptation across regulatory systems, market structures, grid architectures, ownership models, and cultural expectations of trust and accountability. Decentralized energy systems differ significantly across regions and institutions, and peer-to-peer or community-based energy models may involve different forms of participation, platform governance, and regulatory supervision [2]. Explainable AI governance is therefore context-dependent because explanations that are meaningful for one actor, market, or jurisdiction may be insufficient for another. Future conceptual and empirical research should refine the framework in relation to specific governance contexts while preserving its core commitment to transparency, accountability, human oversight, and socio-energy legitimacy.

Conclusion

This article has developed a socio-technical governance framework for explainable artificial intelligence in decentralized energy systems. The framework centers on trust, accountability, human oversight, technical interpretability, and institutional responsibility as interdependent dimensions of responsible AI-enabled energy governance. Its main contribution is to conceptualize explainability not only as a model property but as part of the governance architecture required for decentralized, digitalized, and prosumer-oriented energy systems.

Treating explainability as a governance mechanism expands its role beyond technical model interpretation. AI explanations can support transparency by making algorithmic reasoning visible, oversight by helping human actors review and intervene, contestability by enabling affected stakeholders to question decisions, and regulatory accountability by creating auditable decision pathways. This broader understanding is especially important where AI systems influence energy access, flexibility participation, market outcomes, and local energy coordination.

Decentralized energy contexts make this governance challenge particularly urgent because AI decisions are distributed across utilities, aggregators, platform operators, vendors, communities, regulators, and prosumers. In such settings, accountability cannot be assumed to reside in a single organization or technical component. Clear responsibility pathways, meaningful human oversight, and accessible explanations are needed to ensure that automated energy decisions remain institutionally accountable and socially legitimate.

Future work should advance interdisciplinary research on explainable AI governance in real-world decentralized energy systems. This will require collaboration among AI researchers, energy system engineers, regulators, utilities, social scientists, community organizations, and affected stakeholders. Regulatory engagement, participatory design, and practical evaluation will be essential for determining how socio-technical governance frameworks can support transparent, accountable, inclusive, and trustworthy AI-enabled energy transitions.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

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

None

Received: 14 Jun 2022 Revised: 19 Aug 2022 Accepted: 05 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

1. Ahmad T, Zhang D, Huang C, Zhang H, Dai N, Song Y, et al. Artificial intelligence in sustainable energy industry: status quo, challenges and opportunities. J Clean Prod. 2021;289:125834.

2. Tushar W, Yuen C, Saha TK, Morstyn T, Chapman AC, Alam MJ, et al. Peer-to-peer energy systems for connected communities: a review of recent advances and emerging challenges. Appl Energy. 2021;282:116131.

3. 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.

4. 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.

5. 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.

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

7. 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.

8. Burke MJ, Stephens JC. Energy democracy: goals and policy instruments for sociotechnical transitions. Energy Res Soc Sci. 2017;33:35-48.

9. Mengelkamp E, Gärttner J, Rock K, Kessler S, Orsini L, Weinhardt C. Designing microgrid energy markets: a case study: the Brooklyn Microgrid. Appl Energy. 2018;210:870-80.

10. 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.

11. Xu C, Liao Z, Li C, Zhou X, Xie R. Review on interpretable machine learning in smart grid. Energies. 2022;15(12):4427.

12. Carvalho DV, Pereira EM, Cardoso JS. Machine learning interpretability: a survey on methods and metrics. Electronics. 2019;8(8):832.

13. Zhang C, Han J. Data mining and knowledge discovery. In: Urban Informatics. Singapore: Springer; 2021. p. 797-814.

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

14. Wachter S, Mittelstadt B, Russell C. Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv J Law Technol. 2017;31:841.

15. Miller T. Explanation in artificial intelligence: insights from the social sciences. Artif Intell. 2019;267:1-38.

16. Shneiderman B. Human-centered artificial intelligence: reliable, safe & trustworthy. Int J Hum Comput Interact. 2020;36(6):495-504.

17. Szulecki K. Conceptualizing energy democracy. Environ Polit. 2018;27(1):21-41.

18. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138-60.

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

20. Andoni M, Robu V, Flynn D, Abram S, Geach D, Jenkins D, et al. Blockchain technology in the energy sector: a systematic review of challenges and opportunities. Renew Sustain Energy Rev. 2019;100:143-74.

21. 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: Ethics, governance, and policies in artificial intelligence. Cham: Springer; 2021. p. 153-83.

22. Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. Interpretable machine learning: fundamental principles and 10 grand challenges. Stat Surv. 2022;16:1-85.

23. Hagendorff T. The ethics of AI ethics—an evaluation of guidelines. arXiv. 2019. arXiv:1903.03425.

24. Sousa T, Soares T, Pinson P, Moret F, Baroche T, Sorin E. Peer-to-peer and community-based markets: a comprehensive review. Renew Sustain Energy Rev. 2019;104:367-78.

25. Luo Y, Lu C, Zhu L, Song J. Graph convolutional network-based interpretable machine learning scheme in smart grids. IEEE Trans Autom Sci Eng. 2023;20(1):47-58.

26. Gunning D, Aha D. DARPA's explainable artificial intelligence (XAI) program. AI Mag. 2019;40(2):44-58.

27. Capper T, Gorbatcheva A, Mustafa MA, Bahloul M, Schwidtal JM, Chitchyan R, et al. Peer-to-peer, community self-consumption, and transactive energy: a systematic literature review of local energy market models. Renew Sustain Energy Rev. 2022;162:112403.

28. Ucar F. A comprehensive analysis of smart grid stability prediction along with explainable artificial intelligence. Symmetry. 2023;15(2):289.

29. Schwidtal JM, Piccini P, Troncia M, Chitchyan R, Montakhabi M, Francis C, et al. Emerging business models in local energy markets: a systematic review of peer-to-peer, community self-consumption, and transactive energy models. Renew Sustain Energy Rev. 2023;179:113273.

30. Xu C, Li C, Zhou X. Interpretable LSTM based on mixture attention mechanism for multi-step residential load forecasting. Electronics. 2022;11(14):2189.

31. Szulecki K, Overland I. Energy democracy as a process, an outcome and a goal: a conceptual review. Energy Res Soc Sci. 2020;69:101768.

32. Sovacool BK, Martiskainen M, Hook A, Baker L. Decarbonization and its discontents: a critical energy justice perspective on four low-carbon transitions. Clim Change. 2019;155(4):581-619.

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