ORIGINAL RESEARCH **Open access**

AI-Mediated Energy Justice in Smart Grids: A Conceptual Model for Algorithmic Decision-Making, Demand Response, and Vulnerable Consumer Protection

Kenji Tanaka[1*], Yui Kobayashi[1], Takeshi Nakamura[2], Rina Sato[3]

Abstract

Smart grids increasingly rely on AI for demand response, load forecasting, and automated decision-making. These algorithmic processes directly shape who benefits from, pays for, and participates in smart energy services, raising urgent questions of energy justice. Current AI designs for demand response often prioritize system-level optimization without adequately accounting for distributive fairness, procedural inclusion, or the protection of vulnerable consumers. Algorithmic decision-making may inadvertently deepen energy inequalities and exclude those most in need of protection. This article develops a conceptual model for AI-mediated energy justice in smart grids, with a specific focus on algorithmic decision-making, demand response, and the safeguarding of vulnerable consumers. The model is constructed through conceptual synthesis of energy justice theory, algorithmic fairness principles, smart grid governance literature, and vulnerable consumer protection frameworks. No empirical experimentation is conducted. The proposed model organizes justice-oriented AI governance into three integrated components: a fairness-aware algorithmic design layer, a demand response equity assessment layer, and a vulnerability identification and safeguards layer. These components are connected through accountability, transparency, and redress mechanisms that embed justice into the entire AI lifecycle. AI-mediated energy justice requires moving beyond technical optimization to institutionalize fairness, recognition, and procedural rights within smart grid decision-making. The conceptual model offers a structured pathway for aligning AI-supported demand response with the protection of all consumers, especially the most vulnerable.

Keywords Demand response, Smart grids, Energy justice, Algorithmic fairness, Vulnerable consumer protection, AI governance

*Correspondence:
Kenji Tanaka
tanaka.kenji@eng.u-tokyo.ac.jp
[1] Department of Intelligent Energy Systems and AI, Faculty of Engineering, University of Tokyo, Tokyo, Japan
[2] Department of Smart Grid Data Science, Graduate School of Engineering, Kyoto University, Kyoto, Japan
[3] Department of AI for Distributed Energy Resources, Faculty of Engineering, Osaka University, Osaka, Japan

Introduction

AI-driven decision-making is becoming central to smart grid operation, particularly in demand response, automated load control, dynamic pricing, and household-level energy management. Reviews of AI-supported demand-side response show that machine learning systems can coordinate consumption patterns, optimize flexibility, and support grid stability, but these same capabilities also determine which consumers are asked to shift, reduce, or pay differently for energy use [1]. Because energy is an essential service rather than an ordinary commodity, algorithmic choices in smart grids may

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

distribute costs, inconvenience, and opportunities unevenly across households. This makes AI-mediated energy management a justice issue as much as a technical optimization problem.

The central justice gap is that algorithmic optimization can prioritize efficiency while overlooking households with fewer resources, limited digital access, medical energy dependence, or restricted flexibility. Demand response and time-of-use tariff studies indicate that flexibility is socially uneven, because households differ in work schedules, appliance ownership, care responsibilities, and capacity to shift consumption [2]. Vulnerable consumers may therefore be excluded from smart grid benefits or exposed to higher burdens when AI systems treat non-participation as inefficiency rather than constraint. Without justice-oriented design, AI can become a vector through which existing energy inequalities are automated and intensified.

Energy justice provides a tripartite framework for analyzing these risks through distributive, procedural, and recognition justice. Conceptual and review scholarship has shown that energy transitions must attend not only to aggregate decarbonization or efficiency but also to who benefits, who participates, and whose needs are recognized [3]. In AI-mediated smart grids, these justice dimensions require operational translation into algorithmic objectives, consumer rights, vulnerability safeguards, and governance obligations. Vulnerable consumer protection should therefore be treated as a constitutive design principle rather than a compensatory measure added after automation has already produced harm.

This article proposes a conceptual model that integrates energy justice into AI-mediated demand response and smart grid decision-making. Building on work that links smart grids with justice concerns, the model conceptualizes fairness-aware algorithmic design, demand response equity assessment, and vulnerability safeguards as mutually reinforcing layers of governance [4]. It also connects these layers to accountability, transparency, contestability, and regulatory embedding so that justice is not reduced to a technical fairness metric. The article is therefore theory-driven and model-oriented, aiming to specify how AI-supported energy systems could be designed and governed to protect vulnerable consumers while enabling flexibility.

Background

Smart grids and algorithmic decision-making in demand response

Smart grids use digital sensors, smart meters, automated controls, and predictive analytics to make increasingly granular decisions about energy consumption and demand response participation. AI-supported demand-side response can forecast load, classify flexible resources, schedule appliances, and automate energy management in ways that move beyond manual load shedding toward real-time optimization [1]. Smart home scheduling studies illustrate how algorithmic systems can coordinate household devices for energy management, but they also show how control decisions may become embedded in technical infrastructures that consumers do not fully understand [5]. This transition has important implications for autonomy and fairness because automated systems may decide when flexibility is requested, whose consumption is interrupted, and how benefits are allocated.

Table 1 summarizes the principal AI-enabled smart grid functions together with their associated justice risks and governance implications.

Table 1. AI-Enabled Smart Grid Functions and Associated Energy Justice Risks

AI-enabled smart grid function	Primary operational purpose	Potential justice risk	Vulnerable consumer implication	Governance consideration

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

Dynamic pricing algorithms	Optimize electricity demand across peak periods	Unequal exposure to high prices	Low-income households may face affordability stress	Affordability safeguards and tariff oversight
Automated load control	Reduce grid stress during peak demand	Reduced consumer autonomy	Medically dependent households may experience harmful interruptions	Human override and protected-customer exemptions
Predictive demand forecasting	Improve load balancing and energy planning	Biased consumption profiling	Misclassification of households with irregular usage patterns	Fairness-aware model auditing
Smart home energy scheduling	Automate appliance coordination	Unequal access to smart technologies	Digitally excluded consumers may not benefit from optimization	Inclusive technology access policies
AI-based flexibility markets	Allocate incentives for demand response participation	Uneven distribution of financial rewards	Consumers with limited flexibility may be excluded from benefits	Equity-focused participation criteria

Energy justice as a framework for energy transitions

Energy justice frames energy systems through distributive justice, procedural justice, and recognition justice. Distributive justice asks how benefits and burdens are allocated, procedural justice concerns participation and decision-making rights, and recognition justice requires attention to diverse identities, needs, and forms of disadvantage [6]. Clean energy transition scholarship argues that low-carbon transformation can reproduce inequality when justice is not integrated into system design and governance [3]. In smart grids, this means that digital innovation should be judged not only by efficiency or emissions performance but also by whether it produces fair, inclusive, and accountable energy services.

Vulnerable consumers in the energy system

Vulnerability in energy systems includes energy poverty, health-related dependence on stable energy supply, digital exclusion, socio-economic disadvantage, insecure housing, and limited ability to respond to price or control signals. Research on vulnerable consumers and smart technologies shows that smart meter and smart grid participation may be difficult for households facing financial stress, low trust, or limited digital literacy [7]. Studies of neglected and vulnerable households in smart grids further suggest that automation can create new forms of disadvantage when systems assume universal access, capacity, and willingness to participate [8]. Vulnerability should therefore be understood as dynamic and relational rather than as a fixed consumer category.

Algorithmic fairness and AI-mediated decisions

Algorithmic fairness provides tools for identifying and mitigating bias in automated decision-making, commonly distinguishing group fairness, individual fairness, and counterfactual fairness. General machine learning fairness scholarship shows that bias can enter through data, model design, objective functions, and deployment contexts, making fairness both a technical and institutional challenge [9]. In energy systems, fairness cannot be reduced to formal parity because energy access, affordability, comfort, and safety are socially embedded goods. A fairness-aware smart grid must therefore connect metrics and constraints to legal duties, consumer protection, and energy justice principles.

Integrating justice and AI governance in smart grids

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

AI governance scholarship increasingly calls for the integration of ethics, accountability, and human rights into automated systems that affect essential services. Analyses of AI ethics guidelines identify recurring principles such as transparency, fairness, accountability, privacy, and human oversight, but also warn that principles alone may not guarantee ethical outcomes without implementation mechanisms [10]. Human rights approaches to algorithmic accountability emphasize the need for explainability, remedy, and institutional responsibility when automated decisions affect people's lives [11]. Smart grids represent a distinctive governance challenge because they combine AI-mediated decisions with essential infrastructure, market incentives, and consumer vulnerability.

Recent scholarship further argues that AI-assisted energy transitions may disproportionately affect already marginalized and vulnerable communities if governance systems fail to embed participatory safeguards, representative datasets, and community-centered accountability mechanisms. This perspective reinforces the importance of integrating procedural inclusion and social protection directly into AI-supported smart grid governance rather than treating vulnerable consumer protection as an external policy add-on [12].

Conceptual Foundation of the Model

Energy justice dimensions and algorithmic mediation

The proposed model begins from the premise that algorithmic mediation changes how energy justice must be operationalized. Distributive justice maps onto the allocation of demand response burdens, dynamic pricing exposure, compensation, and reliability impacts, while procedural justice maps onto consumer consent, voice, opt-out rights, and access to explanations [13]. Recognition justice maps onto whether AI systems identify diverse household circumstances, including care obligations, disability, medical dependence, and culturally specific energy practices. A justice-oriented model must address all three dimensions together because fair allocation without participation or recognition can still reproduce algorithmic exclusion.

Table 2 translates distributive, procedural, and recognition justice into concrete AI design requirements for demand response, algorithmic decision-making, and vulnerable consumer protection.

Table 2. Energy Justice Translation Matrix for AI-Mediated Smart Grid Decision-Making

Energy justice dimension	Core justice question in smart grids	AI-mediated risk	Design requirement for the conceptual model	Governance implication
Distributive justice	Who receives the benefits of AI-enabled demand response, and who carries the burdens of load reduction, price exposure, or reduced comfort?	Optimization may repeatedly select low-income, low-flexibility, or poorly represented households for demand reduction because they appear cheaper or easier to control.	Demand response algorithms should include fairness constraints that prevent systematic concentration of burdens on vulnerable consumers or disadvantaged neighborhoods.	Regulators and utilities should require distributional impact assessments before deploying AI-mediated demand response programs.
Procedural justice	Are consumers informed, represented, and able to influence or challenge	Automated load control may occur without meaningful notice, opt-out	AI systems should include consent pathways, accessible explanations, opt-out mechanisms, and	Demand response program rules should define minimum procedural

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

	algorithmic energy decisions?	options, explanation, or human review.	human-in-the-loop review for consequential decisions.	rights for affected consumers.
Recognition justice	Do AI systems recognize diverse household needs, vulnerabilities, cultural practices, health dependencies, and digital access barriers?	Algorithmic categories may misclassify households by treating energy flexibility as uniform across consumers.	Vulnerability-aware modeling should recognize medical dependence, care responsibilities, energy poverty, digital exclusion, and temporary hardship without stigmatization.	Utilities should institutionalize vulnerability registries, privacy safeguards, and community-informed criteria for consumer protection.
Corrective justice	What remedies exist when AI-mediated energy decisions cause harm, exclusion, or unfair treatment?	Consumers may be unable to identify, contest, or reverse harmful algorithmic decisions.	The model should include contestability, redress pathways, case review, compensation mechanisms, and escalation to accountable human decision-makers.	Energy regulators should require auditable complaint-resolution procedures for algorithmic demand response harms.
Intergenerational justice	Do AI-enabled smart grids support decarbonization without transferring present costs to disadvantaged or future consumers?	Efficiency-focused automation may prioritize short-term grid savings while neglecting long-term affordability and social legitimacy.	The model should balance flexibility, decarbonization, affordability, and protection of essential energy access.	Smart grid governance should align AI procurement and demand response policy with just transition goals.
Data justice	Are the data used to train, optimize, and evaluate AI systems representative, privacy-preserving, and socially legitimate?	Biased, incomplete, or intrusive data may reproduce structural inequalities or expose vulnerable households to surveillance.	Data governance should include minimization, consent, bias review, representativeness checks, and privacy-preserving vulnerability indicators.	Independent audits should examine both algorithmic outputs and the social conditions embedded in training and operational data.

Vulnerable consumer protection as a design principle

Vulnerable consumer protection should be treated as a non-negotiable design principle for AI-supported smart grids. Research on marginalization in smart grids suggests that household users can be sidelined when innovation focuses on system functionality and market flexibility rather than lived energy needs [14]. The model therefore conceptualizes vulnerability as context-dependent, potentially temporary, and shaped by interactions between household circumstances and automated energy programs. AI systems should be designed to recognize changing vulnerability conditions and adapt safeguards without stigmatizing consumers or forcing intrusive disclosure.

Algorithmic accountability in energy decision-making

Algorithmic accountability in the energy sector can be defined as the obligation of utilities, aggregators, technology vendors, and regulators to justify, explain, audit, and remedy automated decisions affecting consumers. Work on AI ethics implementation stresses that accountability requires practical tools, organizational processes, and governance mechanisms rather than abstract commitments alone [15]. In smart grids, accountability should include auditability of

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

demand response algorithms, transparency about decision criteria, and contestability when consumers experience unfair treatment. The model therefore links technical design choices to institutional duties of explanation, oversight, and redress.

Model Components

Fairness-aware algorithmic design layer

The fairness-aware algorithmic design layer embeds justice requirements directly into the AI systems used for demand response, load control, and energy management. It could incorporate bias detection, protected-consumer constraints, distributional fairness objectives, and model monitoring so that vulnerable households are not systematically selected for load reduction or excluded from benefits. Fairness research shows that algorithmic systems require attention to both data-driven bias and the normative choices embedded in objective functions [9]. In this model, fairness-aware design is not a purely mathematical exercise but a way of translating distributive and recognition justice into operational rules.

Demand response equity assessment layer

The demand response equity assessment layer evaluates how AI-mediated demand response decisions may affect different consumer groups before, during, and after deployment. Studies of time-of-use tariffs demonstrate that pricing and flexibility mechanisms can have unequal effects because households vary in their ability to shift activities and absorb costs [16]. This layer would therefore assess whether automated interventions are likely to be regressive, whether compensation is adequate, and whether participation conditions exclude households with limited flexibility. It acts as a conceptual checkpoint that constrains efficiency-driven optimization before harmful distributions become embedded in routine grid operation.

Table 3 presents the principal dimensions through which demand response equity can be evaluated in AI-mediated smart grid systems.

Table 3. Equity Assessment Dimensions for AI-Mediated Demand Response Programs

Equity assessment dimension	Key evaluation question	Example indicator	Justice concern addressed
Participation equity	Which households are able to participate in demand response programs?	Participation rate by income or demographic group	Recognition justice
Burden distribution	Which households experience repeated curtailment or inconvenience?	Frequency of load reduction events per household	Distributive justice
Financial fairness	Are savings and incentives distributed equitably?	Net bill savings across household categories	Distributive justice
Accessibility and comprehension	Can consumers understand automated decisions and program rules?	Availability of multilingual and accessible explanations	Procedural justice
Consumer contestability	Can affected users challenge algorithmic decisions?	Presence of appeal and human-review mechanisms	Procedural justice
Vulnerability protection	Are vulnerable consumers shielded from harmful outcomes?	Protected-customer exemption rates	Recognition justice

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

Vulnerability identification and safeguards layer

The vulnerability identification and safeguards layer identifies consumers who may require protection from certain AI-mediated decisions and activates tailored safeguards. Vulnerability could be recognized through privacy-preserving indicators, self-declaration, trusted intermediaries, or community-informed criteria, but the process should preserve dignity and avoid intrusive surveillance. Research on smart technologies among vulnerable consumers shows that protection must consider trust, comprehension, affordability, and the risk that digital systems place responsibility on households least able to bear it [7]. Safeguards could include exemption from automated load reduction, enhanced notification, alternative tariff options, targeted compensation, and human review.

Figure 1 presents the proposed conceptual architecture for embedding energy justice, fairness-aware algorithmic design, vulnerability safeguards, and accountability mechanisms into AI-mediated smart grid demand-response decisions.

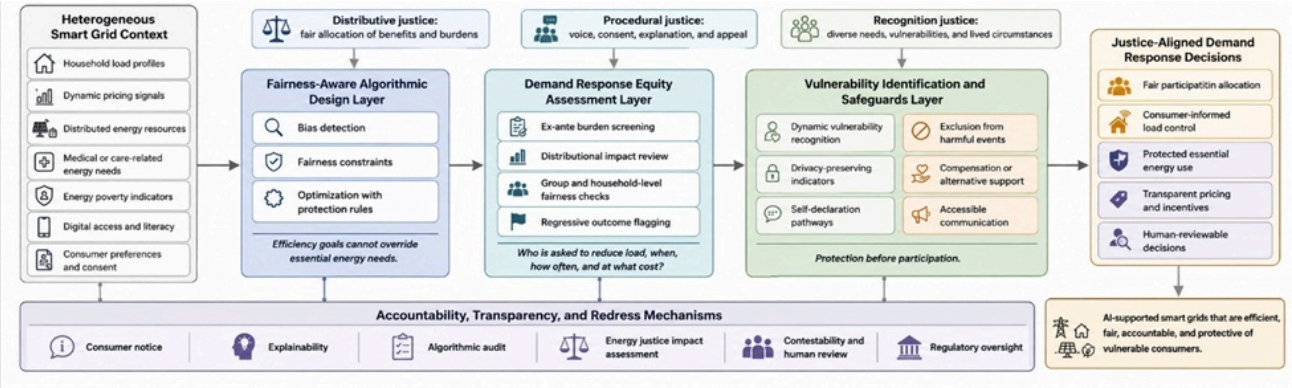


Figure 1. Conceptual Architecture for AI-Mediated Energy Justice in Smart Grids

Algorithmic Decision-Making and Demand Response Equity

Distributional effects of AI-optimized demand response

AI-optimized demand response can inadvertently concentrate inconvenience, risk, or financial penalties on households and neighborhoods that appear technically flexible but are socially constrained. Distributional effects may arise when algorithms optimize aggregate load reduction without considering who repeatedly curtails consumption, who lacks appliance automation, or who faces higher welfare losses from reduced comfort. Research on the distributional effects of time-of-use tariffs shows that demand-side policies can produce uneven household impacts, especially when flexibility is assumed rather than assessed [2]. Equity therefore requires deliberate constraints on pure efficiency optimization and ongoing assessment of how burdens and benefits are shared.

Procedural justice in algorithmic load control

Procedural justice in algorithmic load control concerns whether consumers are informed, whether they can meaningfully consent or opt out, whether they have a voice in program design, and whether explanations are accessible. Smart grid justice research indicates that public acceptance and legitimacy depend on more than technical performance; they also require fair processes and recognition of affected users [17]. In AI-mediated demand response, procedural justice should include transparency about automated decision criteria, opportunities for consumer input, and human review when decisions affect comfort, affordability, or essential energy use. Algorithms should therefore be governed as decision-making institutions rather than neutral technical tools.

Recognition justice and vulnerable consumer inclusion

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

Recognition justice requires AI systems to acknowledge that consumers have diverse energy needs, capacities, and constraints that cannot be captured by average consumption profiles alone. Household rhythms, care duties, health needs, religious or cultural practices, work schedules, and housing conditions can all shape whether demand response participation is fair or feasible. Research on dynamic pricing and infrastructure services emphasizes that temporal flexibility is not evenly distributed and that fairness depends on whose daily rhythms are privileged by pricing or control regimes [18]. Misrecognition can therefore be built into algorithmic categories when systems classify consumers as flexible, inefficient, or non-compliant without understanding their circumstances.

Vulnerable Consumer Protection and Fairness Safeguards

Identifying and representing vulnerability in AI models

Representing vulnerability in AI models requires a careful balance between protection, privacy, dignity, and non-stigmatization. AI and energy justice scholarship argues that smart grid systems should not treat vulnerability as a simple demographic attribute, because vulnerability may emerge from changing interactions among income, health, housing, digital access, and energy dependence [19]. The model could therefore combine voluntary self-identification, community-informed indicators, trusted intermediary verification, and dynamic reassessment rather than relying exclusively on opaque proxy variables. Such representation should be ethically governed so that protective classification does not become a basis for exclusion, surveillance, or discriminatory pricing.

Protective constraints and fairness metrics

Protective constraints and fairness metrics can translate vulnerable consumer protection into the optimization logic of AI-mediated demand response. For example, hard constraints could prevent medically dependent households from being selected for automated load shedding, while softer fairness metrics could monitor whether participation rates, incentives, and burdens are distributed equitably across consumer groups. Emergency demand response research shows that fairness-aware algorithms can be designed to allocate curtailment responsibilities more equitably, although such designs require explicit normative choices about what counts as fair treatment [20]. The model therefore treats efficiency and protection as governed trade-offs, not as purely technical parameters.

Dynamic protections and adaptive safeguards

Vulnerability should be conceptualized as dynamic because households may become temporarily vulnerable through illness, income shock, extreme weather, equipment failure, or care responsibilities. Data-driven demand response systems show that AI can adapt to changing consumption behavior, but justice-oriented adaptation would also require sensitivity to changing household capacity and risk exposure [21]. The model could therefore include safeguards that are activated, suspended, or intensified as circumstances change, subject to consumer consent and regulatory oversight. Adaptive protection would be expected to reduce the risk that automated systems treat temporary hardship as ordinary inflexibility or non-compliance.

Transparency and explanation for vulnerable consumers

Transparency for vulnerable consumers must be designed around practical comprehension rather than formal information disclosure alone. AI ethics frameworks emphasize transparency and human oversight, but critics note that ethical principles only become meaningful when translated into usable rights, accessible explanations, and accountable practices [22]. In smart grids, explanations should clarify why a household was selected for a demand response event, how a tariff change affects bills, what opt-out rights exist, and how to request human review. Such explanations should be available through multiple channels and languages so that digital literacy, disability, or distrust does not prevent consumers from challenging unfair decisions.

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

Accountability, Transparency, and Redress Mechanisms

Algorithmic auditing and impact assessments

The model supports algorithmic auditing and energy justice impact assessments as core accountability mechanisms for AI-mediated smart grids. Audits should examine whether demand response algorithms distribute burdens unfairly, misclassify vulnerable consumers, or embed biased assumptions about flexibility, affordability, and acceptable discomfort. Work on AI ethics tools suggests that impact assessments, audit protocols, and organizational implementation methods are necessary to move from broad principles to operational governance [15]. In energy systems, these assessments should be regulator-mandated, independently reviewable, and connected to public reporting where consumer rights or essential service access may be affected.

Figure 2 illustrates how algorithmic auditing, energy justice impact assessment, independent review, public reporting, and redress mechanisms can be connected into an accountability pathway for AI-mediated smart-grid governance.

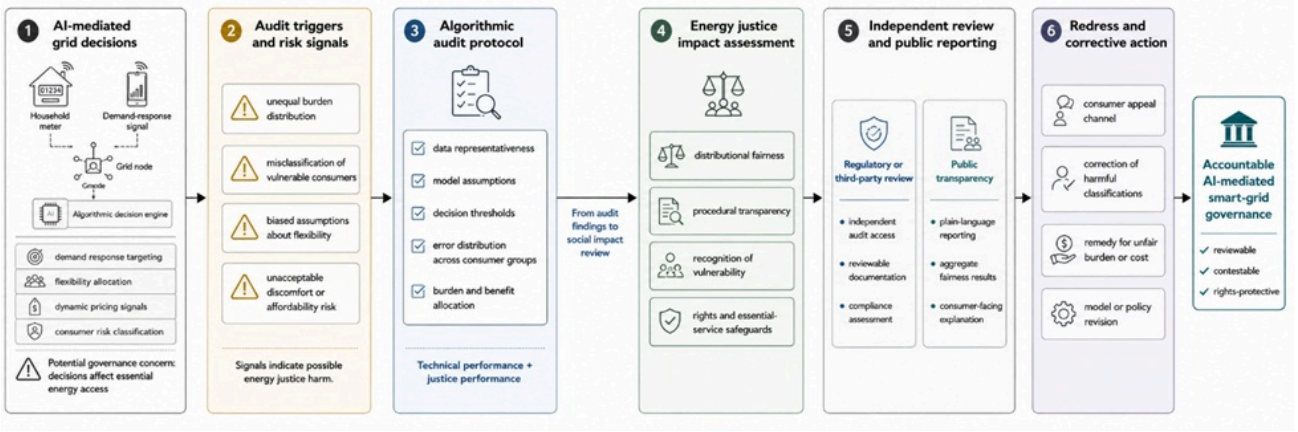


Figure 2. Accountability pathway for auditing AI-mediated smart-grid decisions and enabling consumer redress

Transparency and explainability for affected consumers

Affected consumers should have access to understandable explanations when AI decisions influence their energy supply, comfort, participation opportunities, or bills. The fallacy of algorithmic inscrutability is that complex systems are treated as inherently beyond explanation, even though governance can require documentation, traceability, and meaningful accounts of decision logic [23]. Consumer-centric transparency should therefore be distinguished from model-centric explainability: the former asks what consumers need to know to understand, challenge, or avoid harm, while the latter often focuses on technical interpretability for experts. The model prioritizes explanations that support consumer agency, especially where automated decisions affect essential energy needs.

Contestability and redress pathways

Contestability is essential because even well-designed AI systems may produce unfair, mistaken, or harmful decisions. Human rights approaches to algorithmic accountability emphasize that people affected by automated decisions should have access to review, remedy, and institutional responsibility rather than being left to confront technical systems alone [11]. In smart grids, redress pathways could include human review of automated load-control decisions, correction of vulnerability classifications, reimbursement for unfair penalties, and escalation to regulators or ombudsman bodies. These mechanisms operationalize procedural justice by ensuring that consumers are not merely informed about algorithmic decisions but can challenge and remedy them.

Institutional accountability and regulatory oversight

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

Accountability cannot rest solely with technical systems, because AI-mediated energy decisions are shaped by utilities, aggregators, platform vendors, regulators, and market rules. Critiques of AI ethics argue that principles may remain weak unless institutions define obligations, assign responsibility, and enforce compliance when harms occur [24]. The model therefore locates accountability within governance structures that include regulatory supervision, consumer advocacy, public-interest auditing, and utility-level responsibility for outsourced algorithmic systems. Institutional oversight is especially important in smart grids because consumers may not know whether a harmful decision came from a utility policy, an aggregator platform, a tariff rule, or an AI model.

Table 4 consolidates the proposed safeguard architecture by linking vulnerability risks, algorithmic controls, institutional responsibilities, and redress mechanisms across the AI lifecycle.

Table 4. Governance and Safeguard Architecture for Protecting Vulnerable Consumers in AI-Mediated Demand Response

AI lifecycle stage	Vulnerability risk	Required safeguard	Responsible actor	Evidence of accountability
Problem formulation	Vulnerable consumers may be framed as optimization constraints rather than rights-bearing participants.	Define demand response objectives to include affordability, protection of essential energy use, and avoidance of regressive burdens.	Utility leadership, regulators, consumer advocates, AI governance boards	Published program objectives that explicitly include energy justice and vulnerable consumer protection.
Data collection and representation	Energy poverty, medical dependence, care needs, or digital exclusion may be invisible in operational data.	Use privacy-preserving vulnerability indicators, voluntary self-declaration, community-informed criteria, and dynamic reassessment.	Utilities, data stewards, consumer protection agencies	Data governance documentation, privacy assessments, and vulnerability representation protocols.
Model development	Training data and optimization goals may reproduce unequal participation or exclusion patterns.	Include bias detection, fairness-aware optimization, protected-consumer constraints, and testing for unequal burden allocation.	AI developers, utility analytics teams, external auditors	Model cards, audit reports, fairness review logs, and documented mitigation actions.
Demand response event selection	AI systems may repeatedly target households with limited flexibility or limited ability to contest participation.	Apply pre-event equity screening, participation caps, essential-use exclusions, and household-level burden monitoring.	Demand response operators, aggregators, utility control centers	Event-level equity reports showing who was selected, who was excluded, and why.
Consumer communication	Vulnerable consumers may not understand automated decisions, pricing changes, opt-out rights, or available protections.	Provide plain-language explanations, multilingual communication, non-digital alternatives, and accessible notice before consequential actions.	Utilities, customer service teams, consumer advocacy bodies	Communication templates, accessibility checks, consumer comprehension review, and service-channel records.

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

Human oversight	Fully automated decisions may override context-sensitive judgment about hardship, medical need, or care responsibilities.	Require human review for high-impact or disputed decisions and maintain escalation pathways for protected consumers.	Utility governance officers, ombudsman services, regulators	Human-review logs, escalation records, and decision reversal documentation.
Post-event assessment	Harms may remain invisible if evaluation focuses only on grid performance or cost savings.	Conduct energy justice impact assessments that examine comfort loss, bill effects, participation frequency, complaints, and exclusion patterns.	Regulators, independent auditors, utilities	Public impact assessment summaries and corrective-action plans.
Contestability and redress	Consumers may lack practical routes to challenge unfair algorithmic treatment or obtain remedies.	Establish appeal channels, rapid human review, compensation procedures, and protection from retaliation or service penalty.	Regulators, ombudsman services, utilities, consumer protection agencies	Complaint outcomes, remedy statistics, appeal timelines, and regulatory compliance reviews.
Policy embedding	Safeguards may remain voluntary, inconsistent, or dependent on individual utility practices.	Incorporate fairness audits, vulnerability protections, transparency duties, and redress requirements into regulation and procurement standards.	Energy regulators, policymakers, standard-setting bodies	Binding regulatory rules, procurement criteria, compliance audits, and enforcement mechanisms.

Table 5 illustrates the institutional responsibilities required to operationalize accountability and transparency in AI-mediated smart grid governance.

Table 5. Institutional Responsibilities in AI-Mediated Smart Grid Governance

Governance actor	Core responsibility	Accountability mechanism	Consumer protection role
Energy regulators	Establish mandatory fairness and transparency requirements	Regulatory audits and compliance reviews	Enforce protected-consumer safeguards
Utilities	Operate AI-mediated demand response systems responsibly	Public reporting and internal governance review	Provide accessible communication and redress
AI developers and vendors	Design fairness-aware and explainable systems	Algorithmic testing and documentation	Reduce bias and discriminatory outcomes
Consumer advocacy organizations	Represent affected households and vulnerable groups	Independent monitoring and public consultation	Support complaints and participation rights

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

Ombudsman and review bodies	Resolve disputes and appeals related to automated decisions	Case review and remedy procedures	Ensure procedural fairness and redress
Policymakers and standards bodies	Embed justice principles into energy policy and procurement rules	Legislative oversight and standards enforcement	Institutionalize long-term consumer protections

Governance Integration and Policy Embedding

Embedding justice into energy AI regulation

Energy regulators could embed justice requirements into smart grid codes, AI procurement rules, tariff approvals, and demand response program standards. Artificial intelligence in sustainable energy systems is often framed as an opportunity for optimization and decarbonization, but governance must ensure that such innovation does not compromise fairness, affordability, or consumer protection [25]. Regulatory tools could include mandatory fairness audits, vulnerability impact assessments, explainability requirements, protected-consumer exemptions, and reporting duties for AI-mediated demand response. Embedding these requirements into regulation would make justice a condition of smart grid deployment rather than a voluntary ethical aspiration.

Utility and aggregator governance for equity

Utilities, aggregators, and energy service providers can institutionalize the conceptual model through internal governance structures that connect technical teams with justice, consumer protection, and community expertise. Dynamic price offering research suggests that retailers' assumptions about consumer engagement shape how demand response opportunities are designed and communicated, making organizational governance central to equity [26]. Justice boards, equity officers, community advisory panels, human-in-the-loop oversight protocols, and consumer complaint review teams could help ensure that algorithmic decisions remain accountable to affected households. Such arrangements would also support recognition justice by making lived energy constraints visible within institutional decision-making.

Evaluation Strategy for the Conceptual Model

Conceptual coherence and justice theory alignment

The conceptual model should first be evaluated for internal coherence and alignment with established energy justice theory. Energy justice research on low-carbon transitions shows that decarbonization policies can generate new injustices when distributive, procedural, and recognition dimensions are not jointly considered [27]. The model is coherent only if its fairness-aware design layer, equity assessment layer, and vulnerability safeguards reinforce rather than contradict these three dimensions. Such evaluation would examine whether algorithmic objectives, governance mechanisms, and consumer protections form a consistent justice architecture.

Stakeholder relevance and pro-consumer assessment

The model could also be assessed through structured stakeholder engagement without claiming empirical validation or reporting findings. Large-household participation research indicates that demand response acceptance depends on perceived personal and collective benefits, flexibility, and program design, which makes consumer perspectives important for judging relevance [28]. Consumer panels, advocacy organizations, regulators, utilities, and aggregators could review whether the model's safeguards, explanations, and redress mechanisms are understandable and practically meaningful.

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

This form of pro-consumer assessment would test conceptual acceptability while preserving the article's non-empirical character.

Transferability across smart grid contexts

Transferability should be evaluated by asking whether the model can be adapted across different levels of smart grid maturity, regulatory capacity, market design, and consumer protection regimes. AI ethics guideline scholarship shows that ethical principles often travel globally, but their meaning and implementation vary across institutional contexts [29]. The model may therefore require adaptation for community energy systems, urban smart grids, off-grid networks, or highly liberalized markets. Its core contribution is not a fixed technical template but a justice-oriented governance structure that can be translated into varied smart energy contexts.

Limitations

Conceptual nature and need for empirical validation

The model is conceptual and has not been empirically tested in pilot projects, regulatory sandboxes, or operational smart grid environments. Research on AI ethics guidelines for developers and users emphasizes that principles and frameworks require practical interpretation, implementation, and evaluation before their real-world effects can be known [30]. Future work should operationalize the model in controlled settings to examine whether its fairness-aware design, equity assessment, vulnerability safeguards, and redress mechanisms are feasible and effective. Such validation should remain attentive to the risk that measurement tools may oversimplify justice or create new forms of exclusion.

Contextual specificity of vulnerability and justice

Definitions of vulnerability, fairness, and consumer protection vary across cultural, legal, economic, and infrastructural contexts. Smart grid experiments show that justice concerns are shaped by national institutions, regulatory arrangements, and expectations about household participation, meaning that a model developed at a conceptual level cannot be transferred mechanically across jurisdictions [17]. The proposed framework may therefore require significant adaptation when applied to different markets, welfare systems, energy poverty conditions, or data protection regimes. This limitation does not weaken the need for AI-mediated energy justice, but it does caution against treating any single model as universally complete.

Conclusion

This article has developed a conceptual model for embedding energy justice into AI-mediated demand response and algorithmic decision-making in smart grids. Its central contribution is to connect distributive, procedural, and recognition justice with fairness-aware algorithmic design, demand response equity assessment, and vulnerable consumer safeguards.

Without justice-oriented design, AI-driven smart grids risk amplifying existing energy inequalities through automated load control, dynamic pricing, exclusionary participation requirements, and opaque decision-making. The model provides a structured response to this risk by treating vulnerable consumer protection as a constitutive design and governance requirement.

The proposed framework integrates algorithmic fairness, equity assessment, vulnerability identification, transparency, contestability, and institutional accountability into a single justice-oriented architecture. It therefore moves beyond the view that technical optimization alone can deliver socially legitimate smart energy systems.

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

Regulators, utilities, researchers, aggregators, and consumer advocates should co-develop, pilot, and refine justice-driven AI governance models for the next generation of smart energy systems. Doing so would help ensure that smart grids advance decarbonization and flexibility while protecting the rights, dignity, and needs of all consumers.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 21 May 2022 Revised: 27 Jul 2022 Accepted: 12 Aug 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. 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.

2. Yunusov T, Torriti J. Distributional effects of Time of Use tariffs based on electricity demand and time use. Energy Policy. 2021;156:112412.

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

4. 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.

5. Rocha HR, Honorato IH, Fiorotti R, Celeste WC, Silvestre LJ, Silva JA. An artificial intelligence based scheduling algorithm for demand-side energy management in smart homes. Appl Energy. 2021;282:116145.

6. Jenkins KE, Sovacool BK, Mouter N, Hacking N, Burns MK, McCauley D. The methodologies, geographies, and technologies of energy justice: a systematic and comprehensive review. Environ Res Lett. 2021;16(4):043009.

7. Shirani F, Groves C, Henwood K, Pidgeon N, Roberts E. 'I'm the smart meter': perceptions of smart technology amongst vulnerable consumers. Energy Policy. 2020;144:111637.

8. Tarasova E, Wallsten A. Stakeholder perspectives on neglected and vulnerable households in smart grids. Environ Sociol. 2023;9(3):257-68.

9. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Comput Surv. 2021;54(6):1-35.

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

11. McGregor L, Murray D, Ng V. International human rights law as a framework for algorithmic accountability. Int Comp Law Q. 2019;68(2):309-43.

12. Necula SC. Assessing the potential of artificial intelligence in advancing clean energy technologies in Europe: A systematic review. Energies. 2023;16(22):7633.

13. Calver P, Simcock N. Demand response and energy justice: A critical overview of ethical risks and opportunities within digital, decentralised, and decarbonised futures. Energy Policy. 2021;151:112198.

14. Tarasova E, Rohracher H. Marginalising household users in smart grids. Technol Soc. 2023;72:102185.

15. 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 International Publishing; 2021. p.153-83.

16. Torriti J, Yunusov T. It's only a matter of time: Flexibility, activities and time of use tariffs in the United Kingdom. Energy Res Soc Sci. 2020;69:101697.

17. Milchram C, Künneke R, Doorn N, van de Kaa G, Hillerbrand R. Designing for justice in electricity systems: A comparison of smart grid experiments in the Netherlands. Energy Policy. 2020;147:111720.

18. Jalas M, Numminen S. Prime-time access for whom? Rhythms fairness and the dynamic pricing of infrastructure services. Local Environ. 2022;27(10-11):1355-71.

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

20. Jacubowicz Y, Raz D, Beck Y. Fairness algorithm for emergency demand response operation in distribution networks. Int J Electr Power Energy Syst. 2022;138:107871.

21. Antonopoulos I, Robu V, Couraud B, Flynn D. Data-driven modelling of energy demand response behaviour based on a large-scale residential trial. Energy AI. 2021;4:100071.

22. 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.

23. Kroll JA. The fallacy of inscrutability. Philos Trans R Soc A Math Phys Eng Sci. 2018;376(2133):20180084.

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

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

25. 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.

26. Numminen S, Ruggiero S, Jalas M. Locked in flat tariffs? An analysis of electricity retailers' dynamic price offerings and attitudes to consumer engagement in demand response. Appl Energy. 2022;326:120002.

27. 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.

28. Parag Y. Which factors influence large households' decision to join a time-of-use program? The interplay between demand flexibility, personal benefits and national benefits. Renew Sustain Energy Rev. 2021;139:110594.

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

30. Ryan M, Stahl BC. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. J Inf Commun Ethics Soc. 2021;19(1):61-86.

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