ORIGINAL RESEARCH **Open access**

An AI Ethics Maturity Model for Energy Systems: Transparency, Fairness, Safety, Sustainability, and Institutional Responsibility

Ana Rodrigues[1*], Tiago Martins[1]

Abstract

Energy systems increasingly embed artificial intelligence in safety-critical, consumer-facing, and environmentally significant decisions. Maturity in AI ethics is therefore no longer optional but a regulatory, organizational, and societal imperative. Existing AI ethics frameworks often remain abstract and principle-oriented. Energy organizations lack a structured, actionable maturity model that translates high-level ethical commitments into progressive capability levels across multiple operational domains. This article develops an AI ethics maturity model for energy systems. The model is structured around transparency, fairness, safety, sustainability, and institutional responsibility, and it defines five progressive maturity levels. The model is constructed through conceptual synthesis of AI ethics principles, maturity model design, energy governance, responsible innovation, and energy-sector AI literature. No empirical data, experiments, surveys, simulations, or quantitative assessments are used. The model provides a five-by-five grid in which five ethical dimensions are assessed across Nascent, Developing, Established, Advanced, and Leading maturity levels. Each level identifies indicative organizational practices and governance capabilities, while two tables present the maturity grid and its practical assessment logic. The maturity model offers a practical conceptual tool for energy stakeholders to self-assess, benchmark, and improve their AI ethics posture. It supports regulatory alignment, responsible innovation, institutional learning, and public trust in AI-enabled energy transitions.

Keywords Fairness, AI ethics, Transparency, Energy systems, Maturity model, Safety

*Correspondence:
Ana Rodrigues
ana.rodrigues@ulisboa.pt
[1] Department of AI for Power System Optimization and Renewables, Faculty of Engineering, University of Lisbon, Lisbon, Portugal

Introduction

Artificial intelligence is increasingly used across energy systems for demand forecasting, grid optimization, demand-side response, cybersecurity, operational planning, and sustainability-oriented decision support. This expansion creates ethical concerns because AI can shape essential-service access, consumer treatment, operational reliability, environmental priorities, and institutional accountability in ways that are not always visible to affected stakeholders [1]. Energy AI therefore cannot be governed only as a technical optimization tool; it must also be understood as part of a socio-technical system in which algorithmic decisions affect public welfare, energy justice, infrastructure resilience, and decarbonization trajectories [2]. General AI ethics frameworks have emphasized transparency, accountability, fairness, and human oversight, but their translation into energy-sector practice remains uneven and incomplete [3].

A central problem is that many energy organizations lack a clear roadmap for building ethical AI capabilities over time. AI ethics principles are widely discussed in cross-sector literature, yet principle lists often do not specify how utilities, system

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.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

operators, aggregators, or energy service providers should progress from fragmented awareness to mature institutional practice [4]. Existing tools for operationalizing AI ethics provide useful starting points, but they frequently remain generic rather than tailored to regulated, infrastructure-dependent sectors such as energy [5]. As a result, organizations may acknowledge responsible AI goals without knowing whether their transparency practices, fairness controls, safety safeguards, sustainability alignment, and accountability structures are sufficiently mature [6].

A maturity model provides a useful conceptual approach because it organizes capability development into progressive levels, dimensions, and observable characteristics. In organizational settings, maturity models can support self-assessment, benchmarking, capability sequencing, and governance improvement when complex practices cannot be implemented all at once [7]. Responsible AI governance literature similarly suggests that organizations need structured mechanisms for moving from abstract principles to accountable routines, decision rights, audits, and institutional learning [8]. Applying this logic to energy systems could help organizations evaluate not merely whether they endorse AI ethics, but whether they have developed reliable capabilities for embedding ethics into AI design, deployment, monitoring, and oversight [9].

This article contributes a five-dimensional, five-level AI ethics maturity model tailored to energy systems. The proposed model defines maturity across transparency, fairness, safety, sustainability, and institutional responsibility, while recognizing that these dimensions are interdependent rather than isolated compliance categories. The model is conceptual and assessment-oriented, meaning that it does not assign scores or report empirical validation, but instead specifies how ethical AI capability could be characterized at successive levels of organizational development [10]. By connecting responsible AI governance with energy-sector risks and responsibilities, the article offers a structured tool for self-assessment, capability planning, and future empirical refinement [11].

Background

The ethical imperative for AI in energy systems

The ethical stakes of AI in energy systems are unusually high because energy is an essential service, a critical infrastructure domain, and a central arena of climate governance. AI-enabled decisions in pricing, demand response, grid balancing, outage management, energy trading, and customer targeting may affect vulnerable consumers, environmental outcomes, and system reliability [12]. Demand-side response systems, for example, can support flexibility and efficiency, but they may also create uneven burdens if participation, automation, or incentives are distributed unequally across households [13]. These concerns make AI ethics in energy inseparable from questions of public accountability, social inclusion, and energy justice [14].

Core dimensions of AI ethics in the energy context

The five dimensions proposed in this article—transparency, fairness, safety, sustainability, and institutional responsibility—reflect both general AI ethics principles and energy-sector-specific responsibilities. Transparency is necessary because energy AI decisions may be technically complex and institutionally consequential, requiring explainability for operators, regulators, and affected customers [15]. Fairness is essential because algorithmic energy systems may reproduce or intensify social vulnerabilities, especially where data-driven tools influence service access, pricing, prioritization, or flexibility participation [12]. Safety, sustainability, and institutional responsibility complete the model because energy AI must protect physical infrastructure, align with decarbonization goals, and remain accountable within organizations that operate under public-interest obligations [2].

Maturity models: theory and practice

Maturity models are useful where organizations need to develop capabilities gradually across multiple dimensions rather than implement a single technical fix. In AI ethics, maturity thinking can help translate principles into practices by defining

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

staged expectations, capability indicators, governance routines, and evidence sources [6]. This staged logic is particularly valuable for organizations that face uneven readiness across technical teams, compliance units, executive leadership, and operational departments [7]. In energy systems, a maturity model can therefore function as both a diagnostic tool and a governance roadmap, helping organizations identify whether ethical AI practices are ad hoc, emerging, institutionalized, optimized, or sector-leading [8].

Existing AI ethics and maturity approaches

Existing AI ethics and AI governance approaches provide important foundations, but they are often insufficiently energy-specific. Cross-sector frameworks have identified high-level principles such as beneficence, non-maleficence, autonomy, justice, and explicability, yet such principles require organizational tools before they can shape AI procurement, deployment, monitoring, and accountability [3]. Reviews of AI ethics guidelines have also shown that fairness, transparency, and accountability are commonly endorsed but inconsistently translated into concrete implementation mechanisms [16]. AI governance scholarship has begun to address organizational responsibilities, barriers, and best practices, but energy organizations need a model that explicitly integrates ethics with essential-service obligations, grid safety, climate responsibility, and regulatory oversight [9].

The need for an energy-specific AI ethics maturity model

An energy-specific AI ethics maturity model is needed because the risks of AI in energy differ from those in many consumer or commercial domains. Energy AI systems may influence grid stability, decarbonization pathways, cybersecurity exposure, market behavior, and social equity, meaning that ethical maturity must include technical assurance, environmental alignment, and institutional responsibility [17]. The literature on sustainable AI emphasizes that AI should be evaluated not only for what it enables but also for the environmental costs and systemic effects of its design and deployment [18]. A tailored maturity model can therefore help energy organizations move beyond generic responsible AI commitments toward staged, context-sensitive capability development [19].

The urgency of this energy-specific approach is underscored by emerging evidence that AI systems in power grids exhibit measurable biases across socio-economic and geographic dimensions, affecting load forecasting accuracy and predictive maintenance reliability for underserved communities [20]. Furthermore, the governance of AI-enabled consumer-energy relationships requires explicit attention to institutional accountability structures that prevent algorithmic systems from silently reshaping household autonomy and access [21].

Recent regulatory and standards developments further strengthen the case for an energy-specific AI ethics maturity model. The European Commission's trustworthy AI guidance identifies human oversight, transparency, fairness, technical robustness, environmental well-being, and accountability as core requirements for trustworthy AI [22], while the NIST AI Risk Management Framework translates similar concerns into practical governance functions such as governing, mapping, measuring, and managing AI risks across the system lifecycle [23]. In Europe, the AI Act reinforces a risk-based approach to AI governance and creates stronger expectations for trustworthy, safe, and rights-respecting AI systems, which is especially relevant for critical infrastructure and energy-related applications [24]. At the organizational level, ISO/IEC 42001 provides a management-system approach for establishing, implementing, maintaining, and improving AI governance processes [25], while IEEE 7000-2021 emphasizes the traceable integration of ethical values into system design and stakeholder communication [26]. These governance instruments are particularly important because AI is becoming both a tool for energy optimization and a source of new energy demand, with the International Energy Agency highlighting the growing electricity implications of AI and data centres [27]. At the same time, recent power-systems research shows that AI bias can affect load forecasting, predictive maintenance, resource allocation, reliability, and socioeconomic equity, confirming that ethical maturity in energy systems must combine technical assurance with fairness, sustainability, and institutional accountability [28].

Table 1 synthesizes the primary social justice risks, bias manifestations, and governance deficits identified in recent energy-AI ethics research.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Table 1. Social Justice Risks and Governance Challenges in AI-Enabled Energy Systems

Risk Category	Bias/Issue Type	Affected Stakeholders	Governance Gap	Representative Source
Algorithmic Bias	Data representation bias in load forecasting; sampling bias in predictive maintenance	Low-income households; rural communities; renters	Absence of fairness-aware model validation in operational deployment	[20]
Consumer Autonomy Erosion	Opacity in algorithmic nudges and adaptive controls; behavioral shaping without explicit consent	Residential consumers; elderly; digitally excluded households	Lack of transparency requirements for AI-driven demand-response interfaces	[21]
Distributive Injustice	Unequal access to flexibility programs; benefit asymmetry favoring technically equipped consumers	Vulnerable consumers; energy-poor households	No systematic equity auditing for AI-mediated service eligibility	[20]
Accountability Deficit	Dispersed responsibility across platform operators, aggregators, and technology providers	All consumers; regulators	Unclear liability pathways for AI-mediated consumer harm	[21]

Conceptual Foundation of the Maturity Model

Defining AI ethics maturity for energy systems

AI ethics maturity in energy systems can be defined as the organizational capability to design, deploy, monitor, and govern AI systems in ways that progressively integrate transparency, fairness, safety, sustainability, and institutional responsibility. This definition treats maturity as more than compliance; it includes the ability to make ethical commitments operational through documentation, governance roles, stakeholder engagement, risk controls, and continuous improvement [5]. Because energy organizations operate in regulated and socially significant environments, maturity should also reflect how well AI practices align with public-interest responsibilities and long-term energy transition goals [1]. The proposed definition therefore connects internal capability development with external accountability and societal legitimacy [10].

The five maturity levels: nascent, developing, established, advanced, leading

The model uses five maturity levels: Nascent, Developing, Established, Advanced, and Leading. At the Nascent level, organizations would typically show fragmented awareness of AI ethics, limited documentation, and reactive responses to ethical concerns; at the Developing level, they begin creating policies, inventories, and early review practices [6]. Established maturity indicates that ethical AI processes are formalized and regularly applied, while Advanced maturity suggests integrated governance, cross-functional accountability, and proactive monitoring [8]. Leading maturity represents adaptive, externally accountable, and sector-shaping practice, where organizations not only govern their own AI systems but also contribute to responsible AI standards, public learning, and institutional trust [11].

Figure 1 illustrates the proposed maturity model as a progressive ladder from Nascent to Leading AI ethics capability across five interconnected dimensions.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

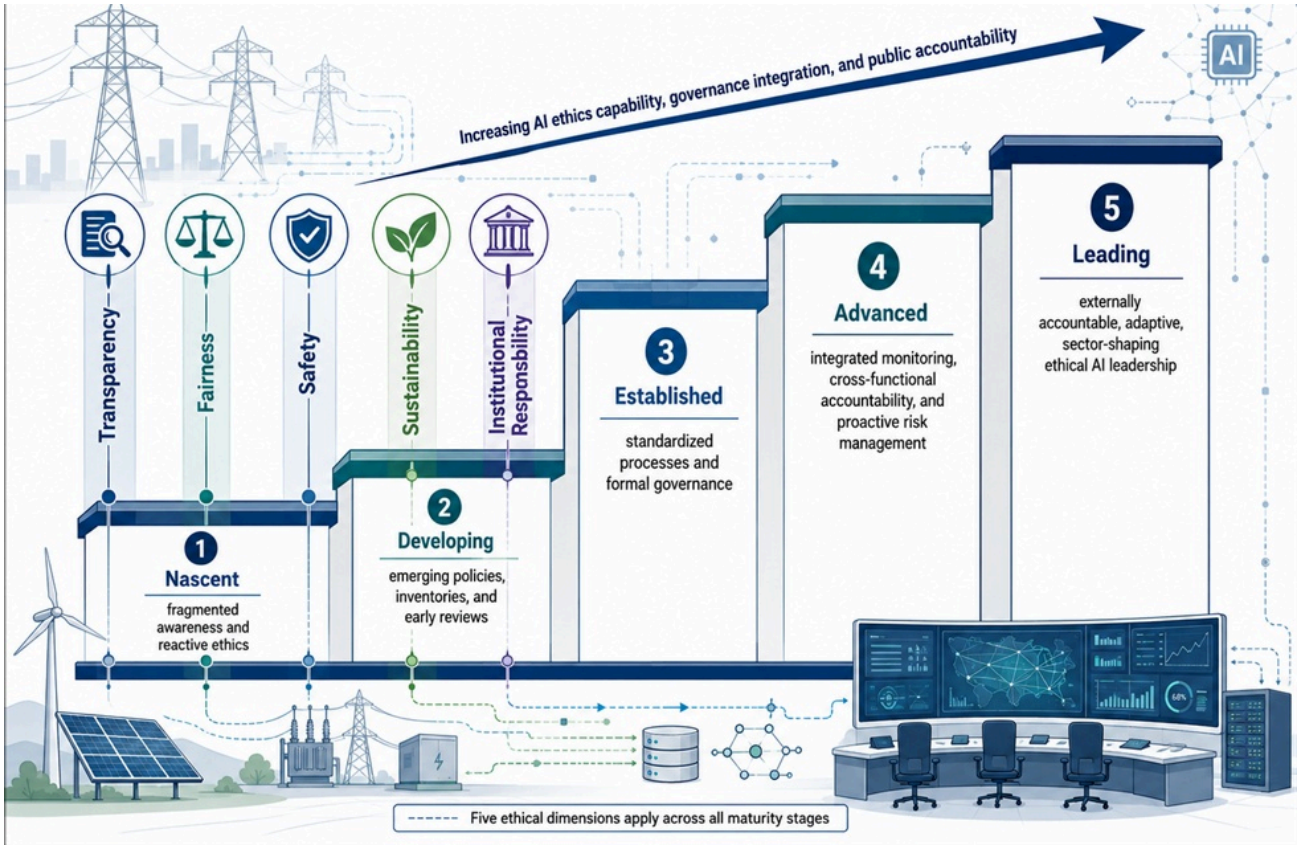


Figure 1. AI Ethics Maturity Model Ladder for Energy Systems

The five dimensions as interconnected capability areas

The five dimensions should be understood as interconnected capability areas rather than separate boxes. Transparency supports fairness because bias cannot be meaningfully identified or contested when model logic, data provenance, or decision pathways are hidden [29]. Safety depends partly on institutional responsibility because robust systems require clear ownership, escalation procedures, validation routines, and post-incident learning [30]. Sustainability also interacts with transparency and governance because organizations must be able to explain both how AI supports decarbonization and how AI's own resource demands are evaluated [31].

Maturity Model Dimensions and Levels

The five dimensions of AI ethics maturity for energy

The proposed maturity model is organized around five dimensions that reflect the main ethical pressures facing AI-enabled energy systems. Transparency concerns the explainability, traceability, and communicability of AI decisions; fairness concerns bias prevention, non-discrimination, procedural justice, and vulnerable consumer protection [12]. Safety concerns robustness, cybersecurity, reliability, fail-safe operation, and physical infrastructure protection, while sustainability concerns both AI for decarbonization and the environmental footprint of AI itself [2]. Institutional responsibility concerns governance ownership, accountability, human oversight, liability management, regulatory alignment, and public trust [17].

The maturity progression logic

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

The progression from Nascent to Leading maturity is not intended to imply that every organization must immediately reach the highest level across all dimensions. Instead, the model conceptualizes maturity as an incremental capability pathway in which organizations move from informal awareness toward embedded, adaptive, and externally accountable practice [7]. In energy systems, this progression is especially important because AI adoption often begins in isolated technical projects before becoming integrated into operational planning, customer programs, market participation, and governance routines [13]. A maturity progression therefore helps organizations identify which ethical capabilities must be strengthened before AI systems are scaled into higher-risk or more socially consequential applications [8].

The maturity matrix structure

The maturity matrix cross-references five ethical dimensions with five capability levels, producing a structured grid that energy organizations could use for self-assessment and strategic planning. Each cell describes indicative characteristics rather than fixed metrics, because the model is conceptual and should be adapted to different organizational roles, regulatory environments, and AI use cases [6]. As shown in **Table 2,** the matrix distinguishes between minimal, emerging, institutionalized, proactive, and sector-leading forms of ethical capability across transparency, fairness, safety, sustainability, and institutional responsibility. This structure enables organizations to identify uneven maturity patterns, such as strong technical safety practices but weak fairness auditing or limited stakeholder accountability [10].

Table 2. AI Ethics Maturity Model for Energy Systems — Dimension-Level Grid

Dimension	Nascent	Developing	Established	Advanced	Leading
Transparency	AI use is known only within technical teams, with limited documentation of model purpose, data lineage, or decision pathways.	Basic model registers and internal documentation exist, but explanations remain inconsistent across operational and customer-facing applications.	Standardized explainability requirements are embedded in AI procurement, deployment, and review processes.	Context-specific explanations are tailored for operators, regulators, customers, and auditors, with traceable decision records.	Proactive, externally reviewable transparency is integrated into real-time governance, public reporting, and sector learning.
Fairness	Bias is addressed only after complaints, incidents, or regulatory concerns arise.	Initial fairness checks are performed for selected models, but vulnerable groups and distributional effects are not systematically assessed.	Formal bias testing, mitigation procedures, and fairness review gates are applied across major AI systems.	Fairness monitoring is continuous, linked to energy justice objectives, and used to redesign data, models, and service processes.	The organization demonstrates externally accountable fairness leadership, including participatory review and transparent equity reporting.
Safety	AI safety is treated as a technical reliability issue without explicit ethical governance or	Safety risks are identified in project-level assessments, with early cybersecurity and	Validation, stress testing, human override, and incident-response protocols are	Anticipatory resilience practices integrate adversarial testing, operational simulation,	Safety governance is externally benchmarked, continuously updated, and

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

	escalation pathways.	fail-safe procedures.	institutionalized for safety-critical AI.	cybersecurity monitoring, and cross-functional accountability.	contributes to sector-wide reliability and certification practices.
Sustainability	AI sustainability is assumed through efficiency claims, with little assessment of environmental costs or rebound risks.	Energy use, computational intensity, or decarbonization relevance is considered for selected AI applications.	Sustainability criteria are included in AI design, procurement, model selection, and lifecycle review.	AI systems are optimized for net environmental value, balancing model resource use, grid flexibility, and decarbonization support.	The organization uses transparent, audited AI sustainability governance to shape industry norms and climate-aligned innovation.
Institutional Responsibility	Accountability is informal, dispersed, and dependent on individual project owners.	AI governance roles are emerging, but decision rights, liability pathways, and oversight mechanisms remain incomplete.	Clear ownership, human oversight, compliance review, and board-level reporting exist for significant AI systems.	Responsibility is integrated across governance, legal, operational, technical, and stakeholder engagement functions.	Institutional responsibility is publicly accountable, independently reviewable, and embedded in sector-level responsible AI collaboration.

Transparency and Explainability

Levels of transparency maturity

Transparency maturity progresses from limited internal awareness toward externally accountable explainability and traceability. At the Nascent level, an energy organization may know that AI is being used but lack a complete inventory of models, datasets, assumptions, and decision contexts [15]. At Established maturity, transparency would typically include documented model purpose, data provenance, explanation requirements, and internal review mechanisms for operational and customer-facing applications [29]. At Leading maturity, transparency becomes proactive and stakeholder-sensitive, meaning that explanations are designed not only for data scientists but also for grid operators, regulators, affected customers, and public accountability forums [32].

Explainability mechanisms and stakeholder communication

Explainability mechanisms in energy systems should be matched to the decision context, the affected stakeholder, and the level of operational risk. Technical methods such as interpretable modeling, post-hoc explanation, feature attribution, scenario-based reasoning, and audit trails can support transparency, but they do not automatically create meaningful understanding [29]. Energy organizations therefore need organizational practices that translate technical explanations into operationally usable and socially accountable communication [32]. For example, operators may require fast causal cues for reliability decisions, while customers and regulators may require understandable justification for demand response actions, service prioritization, or AI-mediated program eligibility [12].

Transparency challenges in proprietary and safety-critical systems

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Transparency in energy AI is complicated by proprietary models, cybersecurity concerns, market sensitivity, and safety-critical infrastructure constraints. Full disclosure may not always be possible where model details could expose system vulnerabilities or reveal commercially sensitive trading strategies, yet excessive opacity can undermine trust, fairness review, and regulatory accountability [17]. Mature transparency therefore should not be equated with unlimited public disclosure; it should involve layered access, secure auditability, role-specific explanations, and accountable documentation [10]. This balance is especially important in smart grids and energy markets, where AI systems may influence both physical reliability and institutional legitimacy [30].

Fairness and Non-Discrimination

Fairness maturity across levels

Fairness maturity in energy AI progresses from reactive complaint handling toward systematic, participatory, and externally accountable equity governance. At the Nascent level, organizations may treat fairness as a legal or reputational issue that arises only after consumer complaints, media attention, or regulatory scrutiny [4]. At the Developing and Established levels, fairness would typically involve structured bias identification, demographic impact review, procedural safeguards, and documentation of how AI systems affect different customer groups or service territories [16]. At Advanced and Leading levels, fairness becomes embedded in model design, data governance, stakeholder engagement, and energy justice strategy rather than being treated as a downstream compliance correction [12].

Addressing algorithmic bias in energy AI

Algorithmic bias in energy systems can emerge from historical consumption patterns, unequal access to smart technologies, incomplete customer data, geographic disparities, and assumptions embedded in optimization objectives. Fairness-aware energy AI should therefore assess not only whether a model performs accurately overall, but also whether its errors, benefits, burdens, and opportunities are distributed unevenly across social groups and locations [33]. In demand response, for example, AI systems may privilege households with flexible appliances, stable internet access, or predictable occupancy patterns, while excluding or penalizing consumers with fewer resources [13]. Mature organizations would typically address these risks through bias testing, representative data practices, fairness-aware model design, and structured review of downstream operational consequences [29].

Energy justice and vulnerable consumer protection

Fairness in energy AI should be connected to energy justice because algorithmic decisions can affect affordability, reliability, participation, and protection for vulnerable consumers. Energy justice scholarship emphasizes that digital and smart energy transitions may reproduce exclusion when benefits are captured by technologically equipped consumers while burdens fall on those with limited flexibility or lower bargaining power [14]. A mature fairness approach would therefore examine whether AI-enabled programs support inclusive access, procedural participation, and protection from discriminatory or exploitative outcomes [12]. This is especially important where AI influences disconnection risk, tariff personalization, customer segmentation, flexibility enrollment, or prioritization during constrained supply and outage recovery [34].

Fairness auditing and metrics

Fairness auditing in energy AI should combine technical indicators with contextual assessment rather than rely on a single universal fairness metric. Technical evidence may include subgroup performance, error distribution, feature sensitivity, adverse-impact analysis, and robustness of model behavior across geographic or demographic segments [29]. However, maturity also requires governance evidence, such as documented fairness thresholds, escalation pathways, affected-stakeholder consultation, and review by legal, regulatory, consumer protection, and operational teams [10]. A Leading organization would not only audit fairness internally but also make fairness governance reviewable by regulators, consumer representatives, and independent experts where appropriate [35].

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Safety, Robustness, and Reliability

Safety maturity levels in energy AI

Safety maturity in energy AI concerns the ability to ensure that AI systems remain reliable, controllable, secure, and fail-safe under normal, stressed, and unexpected conditions. At the Nascent level, safety may be interpreted narrowly as model accuracy or software uptime, while broader operational risks remain underdeveloped [30]. At Established maturity, organizations would typically require validation, stress testing, human override, incident response, and clear operational boundaries before AI systems influence critical energy functions [36]. At Advanced and Leading levels, safety becomes anticipatory, integrating adversarial resilience, cyber-physical risk analysis, post-deployment monitoring, and continuous learning from near misses and failures [17].

Adversarial robustness and cybersecurity

AI safety in energy systems is inseparable from cybersecurity because smart grids, digital platforms, and AI-enabled control systems can create new attack surfaces. Machine learning models may be vulnerable to adversarial manipulation, data poisoning, false-data injection, model extraction, or cascading failures when integrated into operational environments [30]. Smart grid security research shows that AI can support threat detection, but AI itself must also be secured against compromised data streams, malicious actors, and unreliable communication infrastructures [36]. A mature organization would therefore combine conventional cybersecurity controls with AI-specific assurance practices, including adversarial testing, anomaly monitoring, secure data governance, and predefined fallback procedures [19].

Figure 2 presents a cybersecurity–AI assurance architecture showing how adversarial robustness, secure data governance, anomaly monitoring, and fallback procedures jointly protect AI-enabled energy systems from compromised inputs, model-level attacks, and operational disruption.

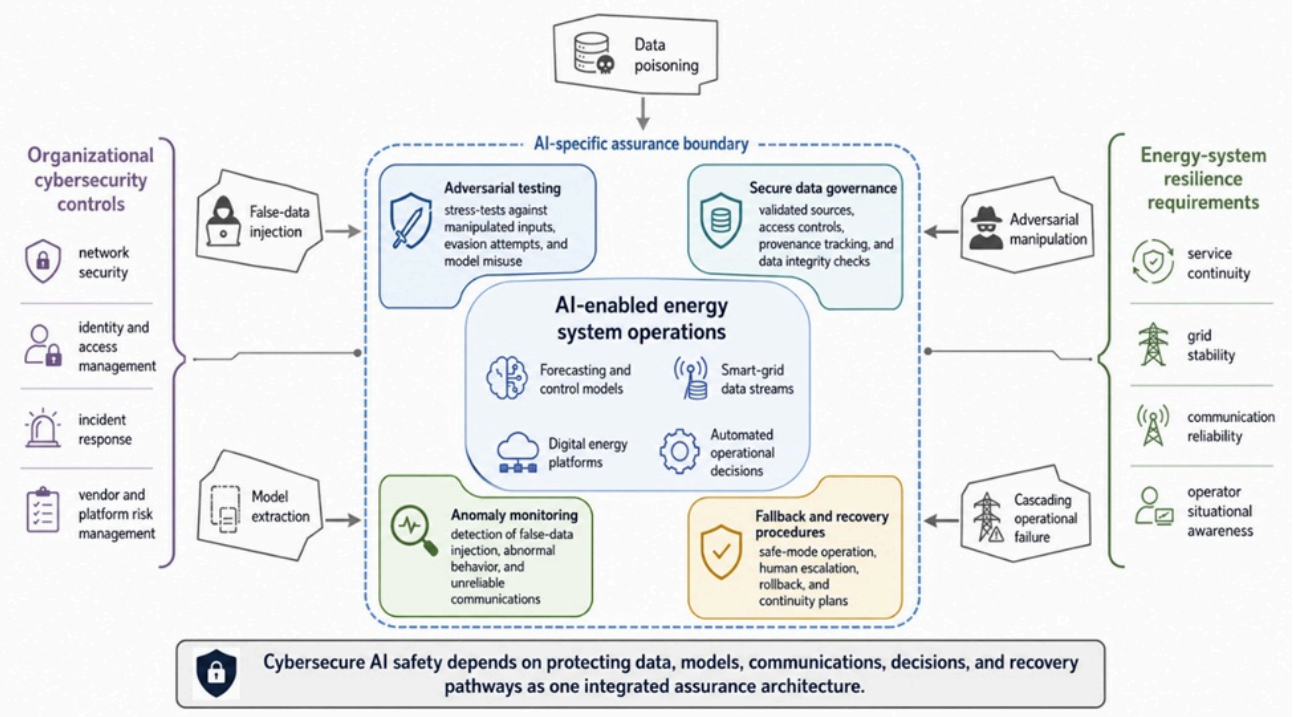


Figure 2. Adversarial Robustness and Cybersecurity Assurance Architecture for AI-Enabled Energy Systems

Certification, testing, and validation

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Testing and validation are central to safety maturity because energy AI systems may influence physical infrastructure and essential-service continuity. Unlike low-risk analytics tools, AI systems used in dispatch support, grid balancing, outage prediction, demand response, or energy trading require validation under plausible stress conditions, not only historical average performance [31]. Mature certification logic would typically include model documentation, operational boundary definition, scenario testing, human-in-the-loop review, cybersecurity assessment, and periodic revalidation when data, infrastructure, or market conditions change [8]. **Table 3** provides an assessment guide that energy organizations could use to connect ethical dimensions with indicators, evidence sources, stakeholder involvement, and practical assessment methods [6].

Table 3. Assessment Guide for AI Ethics Maturity in Energy Systems

Dimension	Key Indicators	Evidence Sources	Stakeholder Involvement	Assessment Method
Transparency	Model inventory completeness, data provenance records, explanation quality, audit-trail availability, role-specific communication procedures.	Model cards, data sheets, procurement records, operator manuals, regulator submissions, customer-facing explanation materials.	Data scientists, grid operators, compliance officers, regulators, customer service teams, consumer representatives.	Review whether explanations are understandable, traceable, timely, and appropriate for each stakeholder group and decision context.
Fairness	Bias testing coverage, subgroup error analysis, vulnerable consumer safeguards, fairness thresholds, complaint and appeal mechanisms.	Fairness audit reports, customer-impact assessments, demographic or geographic performance reviews, consumer complaints, tariff and program records.	Consumer advocates, vulnerable-customer representatives, regulators, legal teams, social policy experts, operational managers.	Assess whether AI benefits, burdens, errors, and access opportunities are equitably distributed and procedurally contestable.
Safety	Stress-test results, adversarial robustness checks, fail-safe design, human override procedures, incident-response readiness.	Validation reports, cybersecurity logs, incident registers, simulation protocols, control-room procedures, post-incident learning records.	System operators, cybersecurity teams, engineers, emergency planners, regulators, safety officers, technology vendors.	Evaluate whether AI remains reliable, controllable, secure, and reversible under normal, abnormal, and adversarial conditions.
Sustainability	AI energy-use tracking, lifecycle assessment, decarbonization relevance, model efficiency criteria, rebound-risk review.	Cloud and compute reports, procurement criteria, sustainability reports, carbon-accounting documentation, project approval records.	Sustainability officers, energy planners, procurement teams, climate experts, regulators, community stakeholders.	Examine whether AI use is environmentally justified, resource-efficient, and aligned with broader decarbonization objectives.
Institutional Responsibility	Clear accountability ownership, board reporting, human	Governance charters, AI policy documents, risk registers, meeting	Executives, board members, legal counsel, regulators,	Determine whether responsibility for AI decisions is explicit,

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

	oversight, regulatory compliance, external review, liability pathways.	minutes, audit reports, legal review files, oversight committee records.	ethics committees, affected communities, independent auditors.	enforceable, reviewable, and connected to institutional decision rights.

Incident response and continuous safety monitoring

Incident response and continuous safety monitoring are necessary because AI systems can degrade after deployment as data distributions, operating conditions, consumer behavior, cyber threats, and policy environments change. Energy organizations should therefore treat AI safety as a lifecycle governance responsibility rather than a one-time pre-deployment approval [9]. Mature safety governance would include real-time monitoring, threshold-based escalation, human override, independent incident review, and structured learning from near misses [30]. In high-maturity settings, these practices would be linked to institutional accountability so that responsibility for AI failures, corrective actions, and public communication is not displaced onto technical teams alone [35].

Sustainability Alignment and Environmental Responsibility

Integrating environmental goals into AI ethics maturity

Sustainability is a distinct ethical dimension because AI can both support and burden the energy transition. AI may contribute to climate mitigation through improved forecasting, flexibility, optimization, and infrastructure planning, but it also consumes computational resources and may create rebound effects if efficiency gains encourage expanded energy use elsewhere [2]. Sustainable AI scholarship therefore distinguishes between AI for sustainability and the sustainability of AI itself, both of which are relevant to energy organizations [18]. A mature energy AI ethics model should assess whether AI systems are environmentally justified, lifecycle-aware, and aligned with decarbonization rather than assumed to be beneficial because they are technologically advanced [37].

Maturity progression in sustainability

Sustainability maturity progresses from unexamined efficiency claims toward transparent environmental accounting and climate-aligned AI governance. At early maturity levels, organizations may adopt AI for operational efficiency without assessing model energy use, computational intensity, lifecycle impacts, or rebound risks [38]. At Established and Advanced levels, organizations would typically include sustainability criteria in AI procurement, model selection, cloud computing choices, and deployment approval, while evaluating whether AI applications genuinely support decarbonization goals [31]. At Leading maturity, energy organizations would use AI sustainability governance to shape sector norms, openly report environmental trade-offs, and prioritize AI applications that deliver credible net environmental value [2].

Institutional Responsibility and Governance Integration

Embedding accountability into organizational structures

Institutional responsibility refers to the organizational capability to assign, enact, review, and enforce accountability for AI decisions. In energy systems, responsibility cannot remain with isolated data science teams because AI systems often affect compliance, consumer protection, safety, sustainability, and public trust [8]. Mature organizations would typically define clear decision rights, escalation pathways, executive oversight, board-level reporting, and human accountability for AI-enabled outcomes [9]. This responsibility structure is especially important where automated or semi-automated systems influence essential services, since affected stakeholders need identifiable institutions that can explain, correct, and justify AI-mediated decisions [17].

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Regulatory compliance and beyond

Regulatory compliance is necessary but insufficient for AI ethics maturity in energy systems. Compliance-oriented governance may ensure that organizations meet formal obligations, but higher maturity requires anticipatory governance that identifies emerging ethical risks before they become legal failures or public controversies [16]. Energy AI regulation is evolving, and organizations that operate across electricity markets, infrastructure planning, consumer services, and digital platforms may face overlapping expectations from energy regulators, data protection authorities, cybersecurity agencies, and sustainability reporting regimes [17]. A mature institution should therefore move beyond minimal compliance toward integrated responsible AI governance that connects legal obligations with ethical review, stakeholder trust, and operational accountability [11].

Stakeholder engagement and public trust

Stakeholder engagement is a core feature of institutional responsibility because energy AI affects communities, customers, operators, regulators, technology vendors, and future generations. Mature organizations would typically involve stakeholders not only after AI deployment but also during problem formulation, risk assessment, explanation design, fairness review, and governance evaluation [14]. Public trust is strengthened when institutions can show that AI systems are explainable, fair, safe, environmentally justified, and accountable through credible evidence rather than general statements of principle [3]. Leading maturity would therefore include participatory review, independent auditing, transparent reporting, and willingness to revise AI systems when stakeholders identify unacceptable risks or harms [35].

Limitations

Conceptual abstraction and need for empirical validation

The proposed maturity model is conceptual and should not be interpreted as an empirically validated measurement instrument. Although it synthesizes AI ethics, energy governance, maturity model, sustainability, and responsible AI literature, its dimensions and levels require future testing in real organizational settings [6]. Empirical validation could examine whether the model captures meaningful differences among utilities, system operators, aggregators, and energy service providers, and whether its categories support practical improvement [7]. The model therefore provides a structured theoretical foundation but requires refinement through interviews, case studies, expert review, organizational piloting, and regulatory feedback [10].

Rapidly evolving AI technologies and ethical standards

A second limitation is that AI technologies, energy markets, regulatory expectations, and ethical standards are changing rapidly. New forms of generative AI, autonomous control, edge intelligence, digital twins, and market-facing AI may create ethical risks that require future extensions of the model [19]. Similarly, emerging expectations around AI auditability, environmental reporting, cybersecurity certification, and consumer protection may change what counts as Established, Advanced, or Leading maturity [17]. The model should therefore be treated as an adaptive governance tool that requires periodic review rather than a fixed framework for all future AI applications [8].

Conclusion

This article developed an AI ethics maturity model for energy systems structured around transparency, fairness, safety, sustainability, and institutional responsibility. The model defines five progressive maturity levels: Nascent, Developing, Established, Advanced, and Leading. Together, these dimensions and levels provide a practical conceptual grid for understanding how energy organizations can move from fragmented ethical awareness toward embedded and accountable responsible AI capability.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

The article has argued that ethical maturity is a journey rather than a destination. Energy organizations may be advanced in one dimension while remaining immature in another, and this unevenness is itself an important diagnostic finding. The model therefore encourages organizations to identify capability gaps, prioritize improvement pathways, and understand ethics as a continuous governance responsibility rather than a one-time compliance exercise.

Future work should test, refine, and adapt the model across different types of energy organizations and regulatory contexts. Utilities, system operators, aggregators, energy service providers, regulators, researchers, and civil society actors all have roles in shaping responsible AI for the energy transition. A mature AI ethics posture will be essential if AI is to support decarbonization, reliability, fairness, and public trust rather than deepen opacity, exclusion, risk, or institutional irresponsibility.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 29 Mar 2023 Revised: 04 Jun 2023 Accepted: 20 Jun 2023
Published online: 15 November 2024

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. https://doi.org/10.1016/j.jclepro.2021.125834.

2. Kaack LH, Donti PL, Strubell E, Kamiya G, Creutzig F, Rolnick D. Aligning artificial intelligence with climate change mitigation. Nat Clim Change. 2022;12(6):518-27. https://doi.org/10.1038/s41558-022-01377-7.

3. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. 2018;28(4):689-707. https://doi.org/10.1007/s11023-018-9482-5.

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

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

6. Krijger J, Thuis T, de Ruiter M, Ligthart E, Broekman I. The AI ethics maturity model: a holistic approach to advancing ethical data science in organizations. AI Ethics. 2023;3(2):355-67. https://doi.org/10.1007/s43681-022-00228-7.

7. Ferreira RM, Grilo A. Piloting a maturity model for responsible artificial intelligence: a Portuguese case study. J Responsib Technol. 2025;22:100117. https://doi.org/10.1016/j.jrt.2025.100117.

8. Papagiannidis E, Mikalef P, Conboy K. Responsible artificial intelligence governance: a review and research framework. J Strateg Inf Syst. 2025;34(2):101885.

9. Papagiannidis E, Enholm IM, Dremel C, Mikalef P, Krogstie J. Toward AI governance: identifying best practices and potential barriers and outcomes. Inf Syst Front. 2023;25(1):123-41.

10. Laine J, Minkkinen M, Mäntymäki M. Ethics-based AI auditing: a systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders. Inf Manag. 2024;61(5):103969. https://doi.org/10.1016/j.im.2024.103969.

11. Zhou J, Chen F. AI ethics: from principles to practice. AI Soc. 2023;38(6):2693-703.

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

13. 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.

14. Sareen S. Digitalisation and social inclusion in multi-scalar smart energy transitions. Energy Res Soc Sci. 2021;81:102251.

15. Xu C, Liao Z, Li C, Zhou X, Xie R. Review on interpretable machine learning in smart grid. Energies. 2022;15(12):4427. https://doi.org/10.3390/en15124427.

16. Hagendorff T. The ethics of AI ethics—an evaluation of guidelines. arXiv [Preprint]. 2019. arXiv:1903.03425.

17. Jørgensen BN, Ma ZG. Regulating AI in the energy sector: a scoping review of EU laws, challenges, and global perspectives. Energies. 2025;18(9):2359. (mdpi.com) https://doi.org/10.3390/en18092359.

18. Van Wynsberghe A. Sustainable AI: AI for sustainability and the sustainability of AI. AI Ethics. 2021;1(3):213-8.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

19. Worthmann C, Malik K. Comprehensive and comparative analysis of AI in the energy sector. AI Ethics. 2025;5(6):5941-53. (link.springer.com)
https://doi.org/10.1007/s43681-025-00728-2.

20. Wang B, Wang J, Dong K, Nepal R. How does artificial intelligence affect high-quality energy development? Achieving a clean energy transition society. Energy Policy. 2024;186:114010.

21. Nwakanma CI, Ahakonye LA, Njoku JN, Gyawali P, Srivastava AK. Explainable AI for interpretable and model-agnostic energy consumption prediction. In: 2025 IEEE Industry Applications Society Annual Meeting (IAS). IEEE; 2025. p.1-6.

22. High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019. (op.europa.eu)
https://doi.org/10.2759/346720.

23. Tabassi E. Artificial intelligence risk management framework (AI RMF 1.0). NIST AI 100-1. Gaithersburg (MD): National Institute of Standards and Technology; 2023. (nist.gov)
https://doi.org/10.6028/NIST.AI.100-1.

24. European Parliament, Council of the European Union. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Off J Eur Union. 2024.

25. International Organization for Standardization. ISO/IEC 42001:2023 Information technology—Artificial intelligence—Management system. Geneva: ISO; 2023.

26. IEEE Standards Association. IEEE 7000-2021: IEEE Standard Model Process for Addressing Ethical Concerns during System Design. New York: IEEE; 2021.

27. International Energy Agency. Energy and AI. Paris: IEA; 2025. (iea.org)

28. Eze C, Ezema A, Roth L, Pan Z, Ponci F, Monti A. AI bias in power systems domain—exemplary cases and approaches. Energies. 2025;18(18):4819. (mdpi.com)
https://doi.org/10.3390/en18184819.

29. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Inf Fusion. 2023;99:101805.

30. Berghout T, Benbouzid M, Muyeen SM. Machine learning for cybersecurity in smart grids: a comprehensive review-based study on methods, solutions, and prospects. Int J Crit Infrastruct Prot. 2022;38:100547. (sciencedirect.com)
https://doi.org/10.1016/j.ijcip.2022.100547.

31. Donti PL, Kolter JZ. Machine learning for sustainable energy systems. Annu Rev Environ Resour. 2021;46(1):719-47.

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

33. Chen L, Jiang N, Wang S. An impossible driver for energy justice? Exploring the impact of artificial intelligence on China's energy transition. Energy Policy. 2025;207:114839. (sciencedirect.com)
https://doi.org/10.1016/j.enpol.2025.114839.

34. Parrish B, Heptonstall P, Gross R, Sovacool BK. A systematic review of motivations, enablers and barriers for consumer engagement with residential demand response. Energy Policy. 2020;138:111221.

35. Birkstedt T, Minkkinen M, Tandon A, Mäntymäki M. AI governance: themes, knowledge gaps and future agendas. Internet Res. 2023;33(7):133-67.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

36. Chehri A, Fofana I, Yang X. Security risk modeling in smart grid critical infrastructures in the era of big data and artificial intelligence. Sustainability. 2021;13(6):3196.

37. Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, Domisch S, et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun. 2020;11(1):233.

38. Rolnick D, Donti PL, Kaack LH, Kochanski K, Lacoste A, Sankaran K, et al. Tackling climate change with machine learning. ACM Comput Surv. 2022;55(2):1-96

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press