ORIGINAL RESEARCH **Open access**

Foundation Models in Energy Systems: A Risk-Governance Framework for Hallucination, Data Sensitivity, Operational Dependence, and Decision Liability

Mohamed Saif[1*], Aisha Ahmed[1], Fatima Al-Mansoori[2], Khalid Al-Hassan[3]

Abstract

Foundation models are rapidly entering energy systems, offering advanced language, reasoning, summarization, and decision-support capabilities. Their adoption could reshape energy planning, operational analysis, customer interaction, and policy research. No dedicated risk-governance framework currently addresses the specific threats of hallucination, data sensitivity, operational dependence, and decision liability in energy-system foundation model deployment. This gap is especially important because energy systems are critical infrastructures with high reliability, privacy, and accountability requirements. This article develops a risk-governance framework tailored to foundation models in energy systems. It maps four core risk categories, identifies affected stakeholders, and links each risk to governance interventions. The framework is developed through conceptual synthesis across foundation model research, energy-system applications, AI governance, privacy scholarship, and liability analysis. No empirical data, experiments, simulations, or performance measurements are used. The framework characterizes hallucination, data sensitivity, operational dependence, and decision liability as interlocking governance problems rather than isolated technical issues. Two analytical tables support risk characterization and governance design. Proactive governance of foundation models is necessary to preserve energy-system reliability, data privacy, operational autonomy, and legal accountability. Energy-sector actors should treat foundation model deployment as a socio-technical governance challenge rather than a purely computational upgrade.

Keywords Foundation models, Energy systems, Hallucination, Data sensitivity, Operational dependence, Decision liability

*Correspondence:
Mohamed Saif
mohamed.saif@ku.ac.ae
[1] Department of Artificial Intelligence for Desert Solar Energy and Cooling Systems, College of Engineering, Khalifa University, Abu Dhabi, United Arab Emirates
[2] Department of Intelligent Energy Analytics for Smart Cities, College of Engineering, United Arab Emirates University, Al Ain, United Arab Emirates
[3] Department of Smart Grid Machine Learning, Faculty of Engineering, American University of Sharjah, Sharjah, United Arab Emirates

Introduction

Foundation models are beginning to influence energy systems through their capacity to summarize complex documents, support policy research, assist technical modeling, and generate operational knowledge across heterogeneous data sources. In wind energy policy research, large language models have already been explored as tools for processing siting ordinances and supporting regulatory analysis [1]. Generative AI is also being framed as a new paradigm for smart

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

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

energy research and education, suggesting that these models may soon affect how energy professionals learn, plan, and coordinate decisions [2]. Surveys of large models for power systems further indicate that foundation-model capabilities are being linked to forecasting, control support, knowledge extraction, and decision assistance [3].

The governance problem emerges because these capabilities are entering a sector where incorrect information, privacy leakage, or poorly allocated responsibility may have consequences beyond ordinary digital-service failure. Reviews of large language models in energy systems highlight substantial promise but also emphasize challenges related to reliability, interpretability, integration, security, and trust [4]. Agentic workflows for building energy model development illustrate how foundation models may be embedded into technical processes rather than used only for text generation [5]. As generative AI becomes part of future energy-system design workflows, governance must address how model outputs are validated before they influence planning or operational choices [6].

A dedicated risk-governance framework is needed because general AI governance principles do not automatically translate into energy-system controls. Sustainable energy AI research has already emphasized that algorithmic design must be aligned with environmental, operational, and social goals rather than treated as neutral optimization [7]. At the same time, the growing energy footprint of AI creates a paradox in which tools proposed for energy-system improvement may themselves impose sustainability burdens [8]. Earlier work on AI in the sustainable energy industry also shows that technical opportunity is inseparable from barriers involving data quality, implementation readiness, regulatory uncertainty, and institutional capacity [9].

This article contributes a conceptual risk-governance framework for foundation models in energy systems, organized around hallucination, data sensitivity, operational dependence, and decision liability. The framework builds from demand-side AI applications, where algorithmic decisions already influence consumer participation, flexibility, and system balancing [10]. It also incorporates smart-grid cybersecurity and privacy concerns, because foundation models may interact with sensitive operational and consumer data in ways that conventional information-security controls were not designed to manage [11]. The article proceeds from background and problem definition to risk categorization, stakeholder mapping, governance architecture, policy implications, research directions, limitations, and conclusion.

Background: Foundation Models in Energy Systems

Capabilities and current use cases

Foundation models can support energy systems by reading policy documents, generating technical explanations, assisting building-energy-model debugging, enabling smart-grid education, and helping analysts reason across fragmented data sources. The use of large language models for wind siting ordinances shows how textual regulatory corpora can be transformed into structured policy insight [1]. In smart energy systems, generative AI has been positioned as a tool for research acceleration, pedagogical support, and interdisciplinary knowledge integration [2]. Surveys of power-system large models and energy-system language models further suggest applications in grid operation support, maintenance reasoning, customer interaction, planning assistance, and domain-specific knowledge management [3, 4].

Early risk signals and failures

Early risk signals arise from the fact that foundation models can generate fluent but incorrect outputs, expose or infer sensitive data, and become embedded in technical workflows where human users may over-trust their recommendations. Hallucination has been described as a systematic challenge in large language models, involving outputs that appear plausible while lacking factual grounding [12]. Foundation models used for decision-making introduce additional risks because hallucinated outputs may influence actions rather than merely misinform readers [13]. Legal research on hallucinations shows that even high-stakes professional domains can be affected by fabricated or distorted reasoning, a warning that is relevant to energy regulation and critical-infrastructure decision support [14].

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

Existing governance gaps

Existing governance approaches provide important principles but often remain too general to govern foundation models in energy-specific settings. Responsible AI governance research emphasizes accountability, transparency, and organizational control, yet it also shows that governance remains fragmented across technical, managerial, and regulatory domains [15]. Global AI ethics reviews identify recurring principles such as fairness, transparency, privacy, and responsibility, but these principles require sector-specific translation before they can guide infrastructure operations [16, 17]. The limitation of principles-only governance is especially important in energy systems, where ethical commitments must be converted into procurement rules, audit procedures, verification requirements, and liability arrangements [18].

Problem Definition

The regulatory and operational void

The regulatory and operational void arises because foundation models are being introduced faster than energy-sector institutions can define their permissible uses, assurance requirements, and accountability boundaries. The draft EU Artificial Intelligence Act has been analyzed as a risk-based regulatory approach, but its sectoral implementation challenges remain significant for complex systems that combine public infrastructure, private vendors, and automated decision support [19]. AI governance in rapidly changing regulatory landscapes is difficult because legal requirements, technical architectures, and organizational practices evolve at different speeds [20]. For energy systems, this mismatch could leave utilities and regulators without clear rules for deciding when foundation model outputs may be used, audited, overridden, or excluded.

The four interlocking risk dimensions

The four interlocking risk dimensions are hallucination, data sensitivity, operational dependence, and decision liability. Hallucination matters because large language models may produce unsupported claims, false summaries, or misleading procedural guidance that appears authoritative [12]. Data sensitivity matters because smart meters and smart-grid systems generate granular information about households, facilities, behavior, and infrastructure conditions that can raise privacy and security concerns [21, 22]. Operational dependence and decision liability matter because external model infrastructures can shape internal energy decisions while distributing responsibility across model providers, system integrators, utilities, and human operators [23, 24].

Consequences of inaction

If these risks remain ungoverned, foundation models could reinforce brittle decision processes, undermine consumer trust, and create accountability gaps in critical infrastructure management. AI-powered cybersecurity research in smart-grid communication shows that digital threats are already a central concern for energy systems, and foundation models could intensify both defensive and offensive complexity [25]. Privacy-preservation studies in smart meters indicate that consumer data protection is not merely a technical issue but a condition for public acceptance and regulatory legitimacy [21]. Liability scholarship further suggests that when AI contributes to harmful decisions, conventional responsibility models may struggle to allocate fault proportionately among multiple actors [26].

Risk Categories

Hallucination

Hallucination refers to the generation of outputs that are syntactically fluent and contextually plausible but factually incorrect, incomplete, unsupported, or misleading. In energy systems, hallucination could manifest as fabricated policy requirements, incorrect grid-procedure summaries, false equipment interpretations, or unsupported recommendations in

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

planning and customer-service contexts. Research on hallucination taxonomies shows that the problem can involve factual inconsistency, source misattribution, reasoning errors, and ungrounded generation [12]. Because foundation models may be used in decision-support settings, hallucination detection for decision-making is especially relevant where outputs could shape dispatch planning, outage communication, compliance interpretation, or investment prioritization [13].

Data sensitivity

Data sensitivity refers to the exposure, inference, misuse, or insecure processing of information that reveals consumer behavior, operational vulnerabilities, grid topology, commercial strategy, or critical-infrastructure conditions. Smart-meter privacy research shows that energy-consumption data can reveal household routines and sensitive behavioral patterns, making consumer-facing foundation model applications a potential privacy risk if prompts, logs, or retrieval systems are poorly governed [21]. Evidence on smart-meter data privacy concerns and dynamic electricity tariffs also suggests that privacy perceptions can affect consumer acceptance of advanced energy services [22]. Cybersecurity standards for smart grids further indicate that privacy and security must be addressed together because operational data exposure can create both individual and system-level vulnerabilities [11].

Operational dependence

Operational dependence refers to the risk that energy organizations become reliant on foundation model infrastructures, vendors, APIs, or automated workflows in ways that reduce internal expertise, weaken resilience, or create single points of failure. Agentic workflows for building energy models show how large language models may become embedded in iterative technical processes, which can improve productivity but also create dependence on model availability and correctness [5]. Generative AI visions for future energy-system design similarly imply deeper integration into modeling, planning, and coordination practices [6]. If utilities use externally hosted models for operational reasoning without redundancy, fallback procedures, or human expertise preservation, model outages or vendor changes could become infrastructure-governance risks rather than ordinary software disruptions.

Decision liability

Decision liability refers to the allocation of responsibility when foundation model outputs influence decisions that cause harm, regulatory noncompliance, consumer loss, or infrastructure disruption. AI liability scholarship shows that responsibility becomes difficult to assign when harm emerges from interactions among developers, deployers, users, and automated systems [23]. European AI liability analysis further indicates that human oversight does not automatically solve accountability problems if oversight is symbolic, under-informed, or institutionally constrained [24]. In energy systems, liability diffusion could arise when a model provider supplies the foundation model, a vendor builds the interface, a utility deploys it, and an operator relies on its output during planning or response decisions.

Figure 1 illustrates how hallucination, data sensitivity, operational dependence, and decision liability emerge across the socio-technical pathway from energy data ingestion to model-assisted decision-making and institutional accountability

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

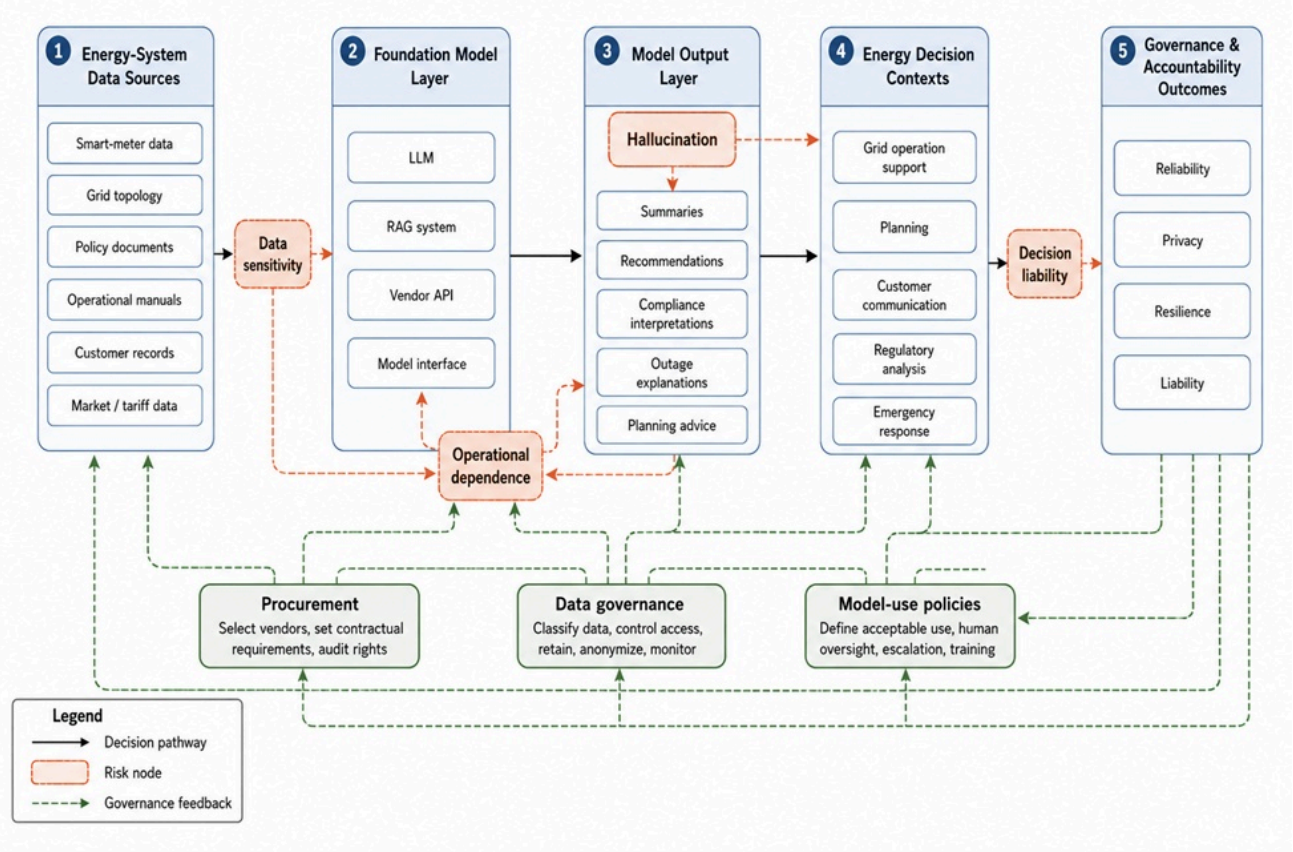


Figure 1. Socio-Technical Risk Pathways of Foundation Models in Energy Systems

Table 1 presents a Risk Characterisation Matrix for foundation models in energy systems, summarising key risk categories alongside their definitions, energy-sector manifestations, severity indicators, and the primary stakeholders affected.

Table 1. Risk Characterisation Matrix for Foundation Models in Energy Systems

Risk Category	Definition	Energy Manifestations	Severity Indicators	Primary Affected Stakeholders
Hallucination	Generation of plausible but incorrect, unsupported, incomplete, or misleading outputs.	Fabricated regulatory requirements, inaccurate outage explanations, incorrect equipment guidance, misleading summaries of grid procedures, unsupported planning recommendations.	Output used without verification, absence of source grounding, conflict with authoritative operational documents, high criticality of decision context, limited human domain expertise.	Utilities, system operators, regulators, consumers, emergency-response teams.
Data Sensitivity	Exposure, inference, leakage, or misuse of sensitive consumer, operational, commercial, or infrastructure data.	Leakage of smart-meter patterns, disclosure of grid topology, exposure of billing or tariff data, retention of sensitive prompts, insecure	Granularity of data, identifiability of consumers, criticality of infrastructure data, third-party model hosting, weak access controls,	Consumers, prosumers, utilities, regulators, model providers, society.

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

		retrieval over internal documents.	unclear data-retention rules.	
Operational Dependence	Reliance on external or opaque foundation model systems in ways that weaken resilience, internal capability, or continuity of operations.	Dependence on vendor APIs, reduced operator expertise, automated workflow lock-in, unavailable model services during incidents, lack of manual fallback procedures.	Absence of redundancy, high workflow integration, vendor lock-in, limited explainability, inadequate operator training, insufficient contingency planning.	Utilities, grid operators, energy-service providers, regulators, critical-infrastructure planners.
Decision Liability	Diffusion of legal, organizational, and professional responsibility when model-influenced decisions cause harm.	Unclear responsibility for incorrect dispatch advice, flawed compliance interpretation, customer harm from automated communication, disputed accountability between utility and vendor.	Multiple actors in decision chain, inadequate audit logs, weak human oversight, unclear contractual allocation, high consequence of decision, regulatory ambiguity.	Utilities, model providers, vendors, regulators, consumers, courts, insurers.

Stakeholder Mapping

Direct stakeholders: utilities, prosumers, and operators

Direct stakeholders include utilities, grid operators, energy-service companies, prosumers, and operational staff who may use or be affected by foundation model systems. Demand-side response research shows that AI and machine learning can shape how consumers participate in flexibility programs, pricing schemes, and energy-management decisions [10]. Prosumers may benefit from more personalized advice and automated coordination, but they may also face privacy risks, opaque recommendations, or unequal access to AI-enabled energy services. Operators and utilities may gain decision-support capacity, yet they also carry responsibility for validating outputs and preserving safe operational control.

Indirect stakeholders: regulators, society, and environment

Indirect stakeholders include regulators, policymakers, courts, insurers, civil society, and the wider environment affected by energy-system reliability and sustainability. AI ethics frameworks emphasize principles such as beneficence, non-maleficence, autonomy, justice, and explicability, which are relevant when foundation models influence essential services [27]. Environmental stakeholders are also implicated because AI's growing energy footprint raises questions about whether foundation model deployment aligns with decarbonization and sustainability goals [8]. Regulators must therefore assess not only whether these models work, but whether they support public value, consumer protection, system resilience, and environmental responsibility.

Stakeholder conflicts and alignments

Stakeholder conflicts arise when efficiency, innovation, privacy, reliability, and accountability point in different directions. Utilities may prefer scalable model services, while regulators may require auditability, consumers may demand privacy, and model providers may protect proprietary architectures. Transparency and accountability research in AI and robotics shows that explainability is not merely a technical feature but a condition for responsible delegation and oversight [28]. At the same time, proportional liability approaches suggest that accountability can be distributed according to risk contribution, which may help align vendors, deployers, and operators around shared governance duties [26].

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

Risk-governance framework

Framework architecture and principles

The proposed risk-governance framework treats foundation models in energy systems as socio-technical components whose risks emerge through interaction among model behavior, data environments, institutional routines, and legal responsibility. Responsible AI governance research suggests that effective governance requires coordination across strategy, operations, accountability, and technical assurance rather than isolated ethical statements [15]. The framework is therefore organized around four risk categories—hallucination, data sensitivity, operational dependence, and decision liability—and three governance layers: technical controls, organizational policies, and regulatory instruments. This architecture reflects the view that AI ethics principles require institutional translation before they can shape real-world infrastructure decisions [16, 18].

Linking risks to governance interventions

The framework links each risk category to governance interventions according to the pathway through which harm may arise. Hallucination requires grounding, retrieval, verification, and expert review because retrieval-augmented generation can help constrain outputs to authoritative documents when properly designed [29]. More recent work on retrieval-augmented text generation emphasizes that retrieval does not eliminate error by itself, so energy organizations should combine RAG with source ranking, answer abstention, traceable citations, and domain validation [30]. Data sensitivity requires privacy-preserving architecture, while operational dependence requires redundancy and human capability preservation, and decision liability requires contractual, audit, and oversight structures that clarify responsibility before deployment.

Responsibility and oversight layers

Responsibility in the framework is distributed across model providers, system integrators, utilities, operators, regulators, and affected publics. The law and economics of AI liability shows that harm allocation should consider the incentives and control capacities of different actors rather than assuming that one party always bears responsibility [23]. European AI liability analysis further indicates that human oversight must be meaningful, informed, and institutionally supported if it is to prevent accountability gaps [24]. For energy systems, oversight should therefore include documented model-use boundaries, audit logs, escalation routes, procurement obligations, incident-reporting duties, and regulatory review for high-risk use cases.

Figure 2 presents the proposed layered governance architecture, showing how technical controls, organizational policies, and regulatory instruments jointly mitigate the four foundation-model risk categories across the energy-system deployment lifecycle.

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

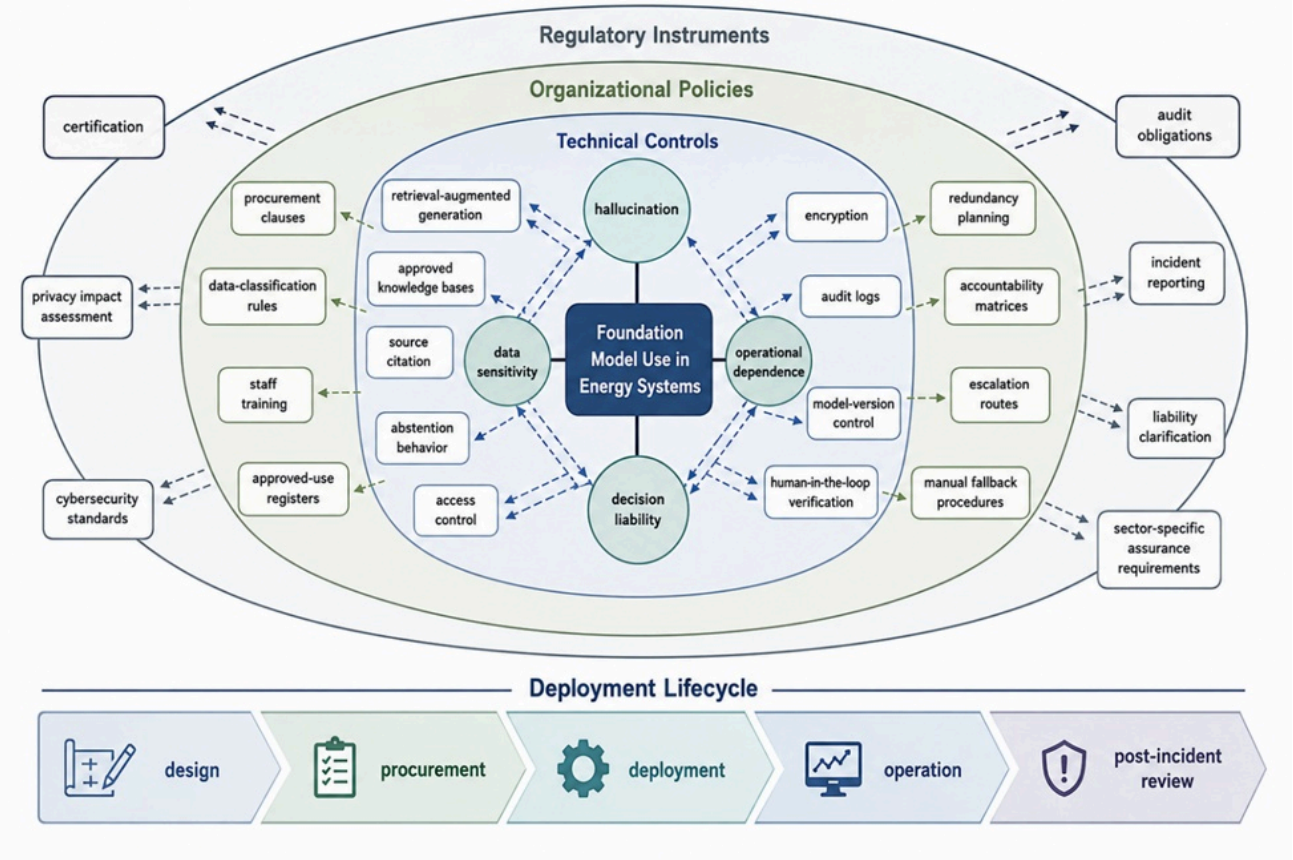


Figure 2. Layered Governance Architecture for Foundation Model Deployment in Energy Systems

Governance Mechanisms

Technical mitigations: RAG, guardrails, and verification

Technical mechanisms should begin with strict use-case classification, because not every foundation model application carries the same risk profile. Hallucination research shows that models may produce unsupported content even when outputs appear coherent, so energy applications should require retrieval from approved knowledge bases, automated consistency checks, uncertainty signaling, and refusal behavior for unsupported claims [12]. RAG systems can reduce ungrounded generation by connecting responses to relevant documents, but their effectiveness depends on the quality, scope, freshness, and access controls of the retrieval corpus [29, 30]. For safety-relevant contexts, technical mitigation should also include human-in-the-loop verification, adversarial testing, red-team evaluation, prompt logging, and separation between advisory outputs and executable operational commands.

Organizational policies: procurement, training, and redundancy

Organizational governance should define who may use foundation models, for what purposes, with which data, and under what review obligations. Energy organizations deploying large language models should require procurement clauses covering data retention, model updates, service continuity, auditability, cybersecurity, and incident notification, because energy-system LLM reviews highlight integration and trust as major implementation challenges [4]. Training policies should prevent automation bias by teaching operators to treat model outputs as advisory artifacts rather than authoritative decisions. Redundancy policies are equally important because agentic energy workflows can create hidden dependencies if organizations remove manual expertise or fallback procedures too quickly [5].

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

Regulatory instruments: certification, audits, and liability rules

Regulatory governance should require assurance proportional to the criticality of the use case, with stricter certification and audit duties for applications affecting grid reliability, consumer rights, tariff interpretation, or infrastructure planning. The EU risk-based regulatory approach illustrates the value of classifying AI systems according to potential harm, but sector-specific implementation is needed for energy contexts involving critical infrastructure and public-service obligations [19]. Liability rules should clarify whether responsibility lies with the model provider, application developer, system integrator, utility, operator, or a shared arrangement based on control and foreseeability [23, 26]. Certification, mandatory documentation, independent audits, incident disclosure, and proportional liability could help prevent foundation model governance from becoming a voluntary compliance exercise.

Table 2 summarises a multi-layer governance framework for foundation model risks in energy systems, mapping key risk categories to corresponding technical mitigations, organisational policies, regulatory instruments, and liability arrangements to support proportionate and accountable deployment across critical infrastructure contexts.

Table 2. Governance Interventions for Foundation Model Risks in Energy Systems

Risk Category	Technical Mitigation	Organisational Policy	Regulatory Instrument	Liability Arrangement
Hallucination	Retrieval-augmented generation using approved sources; source citation; answer abstention; consistency checking; domain-expert verification; restricted deployment in safety-critical contexts.	Mandatory review before operational use; staff training on automation bias; clear distinction between advisory and executable outputs; documentation of accepted and rejected model recommendations.	Sector-specific assurance requirements; audit obligations for high-risk deployments; mandatory incident reporting for harmful or misleading outputs.	Utility remains responsible for operational decisions, while vendors share responsibility when model defects, misleading documentation, or inadequate safeguards contribute to harm.
Data Sensitivity	Data minimization; secure retrieval environments; encryption; access control; prompt filtering; local or private model deployment for sensitive data; privacy-preserving logging.	Data-classification policy; prohibition on entering restricted grid or consumer data into public models; procurement rules on retention and reuse; privacy training for users.	Privacy impact assessments; smart-meter data protection rules; cybersecurity certification; restrictions on cross-border processing of critical infrastructure data.	Deployers are liable for unauthorized data use, while providers and processors share liability where contractual safeguards, retention practices, or security failures cause exposure.
Operational Dependence	Redundant systems; fallback procedures; local copies of critical knowledge bases; service monitoring; model-version control; separation of model advice from control execution.	Vendor-risk management; continuity planning; preservation of internal expertise; periodic manual drills; dependency review before scaling model use.	Critical-infrastructure resilience standards; third-party service continuity obligations; audit of vendor lock-in and systemic dependency risks.	Shared liability based on failure source, with utilities accountable for continuity planning and providers accountable for outages, undisclosed changes, or service failures within their control.

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

Decision Liability	Audit trails; model-output logging; explainability records; role-based approval workflows; traceable handoff from model suggestion to human decision.	Accountability matrix; named decision owners; escalation procedures; legal review of high-impact use cases; insurance and contractual allocation of risk.	Mandatory documentation; liability clarification rules; certification of oversight processes; requirements for meaningful human review.	Proportional liability based on control, foreseeability, and contribution to harm, distributed among provider, integrator, deployer, and human decision authority.

Policy and Industry Implications

Policy recommendations

Policy should treat foundation models in energy systems as critical-infrastructure governance objects rather than generic software tools. AI governance scholarship shows that fast-changing regulatory environments require adaptive rules that can evolve with technical capabilities while preserving accountability and public value [20]. Policymakers should require risk classification, approved-use registers, privacy impact assessments, auditability standards, incident reporting, and clear documentation for high-risk foundation model applications. Because energy systems involve public welfare, environmental goals, and essential services, policy should also ensure that foundation model deployment does not undermine consumer protection, democratic oversight, or system resilience.

Industry best practices and standards

Industry best practice should combine cautious experimentation with strict boundaries around sensitive data, operational authority, and liability allocation. Smart-grid cybersecurity standards demonstrate that security and privacy controls must be embedded into system design rather than added after deployment [11]. Energy organizations should create internal foundation model governance boards, maintain approved model inventories, classify permitted use cases, test outputs against authoritative documents, and require fallback procedures for any workflow that becomes model-assisted. Standards bodies and industry associations could also develop shared assurance templates for energy-specific RAG systems, vendor contracts, audit logs, and human oversight protocols.

Future Research Directions

Empirical validation of risks

Future research should empirically examine how hallucination, data sensitivity, operational dependence, and decision liability appear in real energy-sector foundation model deployments. Current energy LLM scholarship identifies applications and challenges, but further studies are needed to test how users interpret model outputs, when they over-trust generated recommendations, and how governance controls affect behavior [3, 4]. Research on legal hallucinations suggests that professional settings remain vulnerable to fabricated reasoning, making energy-specific evaluation necessary for policy interpretation, compliance support, and operational documentation [14]. Empirical work should avoid treating accuracy alone as sufficient and should instead examine risk pathways, oversight quality, and institutional consequences.

Dynamic governance for evolving models

Dynamic governance is needed because foundation models change through updates, fine-tuning, retrieval-corpus modification, vendor policies, and shifting regulatory expectations. Research on generative AI in smart energy systems suggests that foundation models may become more capable and more deeply embedded in research, education, and operational support, which means governance must be designed for adaptation rather than one-time approval [2]. AI

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

governance in complex regulatory landscapes also shows that institutions must manage uncertainty as rules, technologies, and social expectations evolve [20]. Future research should therefore explore model-change documentation, post-deployment monitoring, version-specific certification, and governance mechanisms for retiring or restricting unsafe model configurations.

Cross-sectoral learning

Energy-system governance can learn from other high-stakes domains where explainability, human oversight, and liability have already been debated. Work on transparent, explainable, and accountable AI for robotics shows that accountability must be connected to how decisions are delegated, interpreted, and controlled in practice [28]. AI liability research also indicates that proportional and risk-sharing approaches may be useful where several actors jointly contribute to system behavior [26]. Future cross-sectoral research should compare energy with healthcare, finance, transportation, and robotics to identify which governance mechanisms transfer well and which require adaptation to the physical, environmental, and public-service characteristics of energy infrastructure.

Limitations

Conceptual nature and data scarcity

This article is conceptual and does not test the proposed framework through experiments, interviews, case studies, surveys, or operational incident analysis. The available literature on foundation models in energy systems is still emerging, with many contributions focused on opportunities, surveys, and prospective applications rather than mature governance evidence [3, 4, 6]. As a result, the framework should be understood as an analytical structure for guiding future research, policy design, and organizational governance rather than as a validated empirical model. Further evidence will be needed to determine how specific controls perform across different energy subsectors, organizational sizes, and regulatory jurisdictions.

Rapid technological change

A second limitation is that foundation model capabilities, deployment architectures, and regulatory expectations are changing rapidly. Generative AI research in smart energy systems points toward expanding use cases, while work on sustainable AI warns that the computational footprint of AI itself must remain part of governance evaluation [2, 8]. Legal and regulatory scholarship also shows that liability and risk regulation are moving targets, particularly as lawmakers attempt to address automated decision-making and human oversight [19, 24]. The framework therefore requires periodic revision as models become more multimodal, agentic, integrated with control systems, or subject to new statutory duties.

Conclusion

Foundation models in energy systems offer meaningful opportunities for knowledge synthesis, planning support, technical assistance, and institutional learning. Yet these opportunities are inseparable from risks that affect reliability, privacy, operational autonomy, and legal accountability. This article has developed a risk-governance framework centered on hallucination, data sensitivity, operational dependence, and decision liability. The framework shows that these risks should be governed together rather than treated as separate technical problems.

The urgency of proactive governance lies in the critical nature of energy systems. Once foundation models become embedded in planning, compliance, customer communication, or operational decision support, weak controls may become difficult to reverse. Energy organizations should therefore establish governance before model use becomes normalized across sensitive workflows. Waiting for harmful incidents would be a poor strategy for a sector responsible for essential services and public trust.

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

Effective governance will require collaboration among researchers, utilities, model providers, regulators, legal experts, consumer advocates, and standards bodies. Foundation models should not be rejected simply because they introduce risk, but neither should they be adopted as neutral productivity tools. The appropriate path is governed experimentation, clear accountability, rigorous validation, and sector-specific regulation. Energy systems need foundation models that are not only powerful, but reliable, privacy-preserving, resilient, and accountable.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 08 Jul 2024 Revised: 13 Sep 2024 Accepted: 30 Sep 2024
Published online: 15 June 2026

Rights and permissions

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

References

1. Buster G, Pinchuk P, Barrons J, McKeever R, Levine A, Lopez A. Supporting energy policy research with large language models: a case study in wind energy siting ordinances. Energy AI. 2024;18:100431. https://doi.org/10.1016/j.egyai.2024.100431.

2. Lin X, Luo Z, Du-Ikonen L, Lin X, Mao Y, Jiang H, et al. Generative artificial intelligence: pioneering a new paradigm for research and education in smart energy systems. Energy AI. 2025;22:100610. https://doi.org/10.1016/j.egyai.2025.100610.

3. Yao Q, Fang F, Chen Y, Liu J, Mo H, Ao Y. AI large models for power system: a survey and outlook. IET Smart Energy Syst. 2025;1(1):3-21. https://doi.org/10.1049/ses2.70000.

4. Mirshekali H, Shadi MR, Ghanadi Ladani F, Shaker HR. A review of large language models for energy systems: applications, challenges, and future prospects. IEEE Access. 2025;13:163162-88. https://doi.org/10.1109/ACCESS.2025.3610994.

5. Zhang L, Ford V, Chen Z, Chen J. Automatic building energy model development and debugging using large language models agentic workflow. Energy Build. 2025;327:115116. https://doi.org/10.1016/j.enbuild.2024.115116.

6. Glaws A, King RN, Emami P, Buster G, Benton BN, Zhang X, et al. Designing future energy systems with generative AI. Comput Sci Eng. 2025;27(4):8-18. https://doi.org/10.1109/MCSE.2025.3567208.

7. Chen D, Lin X, Qiao Y. Perspectives for artificial intelligence in sustainable energy systems. Energy. 2025;318:134711. https://doi.org/10.1016/j.energy.2025.134711.

8. De Vries A. The growing energy footprint of artificial intelligence. Joule. 2023;7(10):2191-4. https://doi.org/10.1016/j.joule.2023.09.004.

9. 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.

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

11. Leszczyna R. Cybersecurity and privacy in standards for smart grids—a comprehensive survey. Comput Stand Interfaces. 2018;56:62-73.

12. Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, et al. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. ACM Trans Inf Syst. 2025;43(2):42:1-55. https://doi.org/10.1145/3703155.

13. Chakraborty N, Ornik M, Driggs-Campbell K. Hallucination detection in foundation models for decision-making: a flexible definition and review of the state of the art. ACM Comput Surv. 2025;57(7):188:1-35. https://doi.org/10.1145/3716846.

14. Dahl M, Magesh V, Suzgun M, Ho DE. Large legal fictions: profiling legal hallucinations in large language models. J Leg Anal. 2024;16(1):64-93.

15. Papagiannidis E, Mikalef P, Conboy K. Responsible artificial intelligence governance: a review and research framework. J Strateg Inf Syst. 2025;34(2):101885. https://doi.org/10.1016/j.jsis.2024.101885.

16. Corrêa NK, Galvão C, Santos JW, Del Pino C, Pontes Pinto E, Barbosa C, et al. Worldwide AI ethics: a review of 200 guidelines and recommendations for AI governance. Patterns (N Y). 2023;4(10):100857. https://doi.org/10.1016/j.patter.2023.100857.

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

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

19. Veale M, Borgesius FZ. Demystifying the draft EU artificial intelligence act. arXiv. 2021. arXiv:2107.03721.

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

20. Zaidan E, Ibrahim IA. AI governance in a complex and rapidly changing regulatory landscape: a global perspective. Humanit Soc Sci Commun. 2024;11:1121. https://doi.org/10.1057/s41599-024-03560-x.

21. Kua J, Hossain MB, Natgunanathan I, Xiang Y. Privacy preservation in smart meters: current status, challenges and future directions. Sensors. 2023;23(7):3697. https://doi.org/10.3390/s23073697.

22. von Loessl V. Smart meter-related data privacy concerns and dynamic electricity tariffs: evidence from a stated choice experiment. Energy Policy. 2023;180:113645. https://doi.org/10.1016/j.enpol.2023.113645.

23. Buiten M, De Streel A, Peitz M. The law and economics of AI liability. Comput Law Secur Rev. 2023;48:105794. https://doi.org/10.1016/j.clsr.2023.105794.

24. Arcila BB. AI liability in Europe: how does it complement risk regulation and deal with the problem of human oversight? Comput Law Secur Rev. 2024;54:106012. https://doi.org/10.1016/j.clsr.2024.106012.

25. Afrin S, Al Muttaki MR, Anil AI, Hasan S. AI-powered cybersecurity for smart grid communication: a systematic review of intrusion detection and threat mitigation systems. Energy Convers Manag X. 2026;29:101416. https://doi.org/10.1016/j.ecmx.2025.101416.

26. Bashayreh MH, Sibai FN, Tabbara A. Artificial intelligence and legal liability: towards an international approach of proportional liability based on risk sharing. Inf Commun Technol Law. 2021;30(2):169-92. https://doi.org/10.1080/13600834.2020.1856025.

27. Floridi L, Cowls J. A unified framework of five principles for AI in society. In: Carta S, editor. Machine learning and the city: applications in architecture and urban design. Hoboken (NJ): Wiley; 2022. p.535-45.

28. Wachter S, Mittelstadt B, Floridi L. Transparent, explainable, and accountable AI for robotics. Sci Robot. 2017;2(6):eaan6080. https://doi.org/10.1126/scirobotics.aan6080.

29. Arslan M, Ghanem H, Munawar S, Cruz C. A survey on RAG with LLMs. Procedia Comput Sci. 2024;246:3781-90. https://doi.org/10.1016/j.procs.2024.09.178.

30. Huang Y, Huang JX. A survey on retrieval-augmented text generation for large language models. ACM Comput Surv. 2024;56(12).

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