ORIGINAL RESEARCH **Open access**

Conceptualizing AI-Based Energy System Autonomy: A Layered Model of Perception, Prediction, Decision, and Governance

Kristian Eriksen[1*], Solveig Berg[1]

Abstract

Energy systems are evolving toward higher degrees of AI-driven autonomy, yet the architecture of such autonomy remains fragmented and poorly conceptualized. A structured understanding of how perception, prediction, decision, and governance interrelate is missing. Without a layered model that integrates governance into the autonomy stack, AI-based energy systems risk developing opaque, unaccountable, and brittle autonomous capabilities that conflict with regulatory and societal expectations. This problem is especially significant in energy systems because autonomous actions can affect reliability, safety, equity, and institutional accountability. This article proposes a layered model of AI-based energy system autonomy comprising four interdependent layers: perception, prediction, decision, and governance. The model clarifies how autonomy capabilities accumulate across technical and institutional functions. The model is constructed through conceptual synthesis of autonomous systems architectures, AI capabilities for energy, and governance frameworks for critical infrastructure. No empirical tests are performed, and the article does not report quantitative performance results. The model defines the functions and outputs of each layer, specifies the interactions between layers, and demonstrates how governance constraints must be embedded within the autonomy architecture. Two analytical tables illustrate the layer-capability mapping and the relationship between autonomy levels and governance requirements. A layered governance-integrated model of autonomy can guide the design, regulation, and responsible deployment of increasingly autonomous energy systems. The model suggests that technical autonomy in energy systems cannot be separated from institutional responsibility, safety assurance, and public accountability.

Keywords Prediction, Energy systems, AI-based autonomy, Layered model, Perception, Decision

*Correspondence:
Kristian Eriksen
kristian.eriksen@ntnu.no
[1] Department of Artificial Intelligence for Hydroelectric and Wind Systems, Faculty of Engineering, Norwegian University of Science and Technology, Trondheim, Norway

Introduction

Energy systems are moving from conventional automation toward AI-supported autonomy, where sensing, forecasting, optimization, and control can increasingly be performed with limited direct human intervention. This shift is driven by renewable variability, distributed energy resources, electrified demand, and digital monitoring infrastructures that make manual coordination increasingly difficult [1]. AI methods applied to smart grids already support functions ranging from state awareness to predictive analytics and operational decision support [2]. As Cheng and Yu argue, the integration of machine learning into power and energy systems signals a broader transition in which computational intelligence becomes part of operational architecture rather than a peripheral analytical tool [3].

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

Despite this transition, the architecture of AI-based energy autonomy remains conceptually underdeveloped. Governance is often treated as an external compliance activity, even though autonomous energy decisions can affect safety, accountability, and social acceptability [4]. General AI governance research has emphasized that accountability, risk management, and oversight must be built into AI systems rather than merely reviewed after deployment [5]. Ethical AI frameworks similarly show that autonomy without responsibility structures risks producing systems that are technically capable but institutionally fragile [6].

A layered model can clarify the functional structure of AI-based energy autonomy by distinguishing perception, prediction, decision, and governance as interdependent dimensions. Perception provides state awareness through data interpretation, prediction anticipates future system conditions, and decision layers translate intelligence into operational action [7]. Reinforcement learning and control-oriented AI studies show how prediction and decision functions increasingly merge in energy applications, but these capabilities require constraint structures that define permissible action [8]. Governance therefore should be conceptualized as a constitutive layer that shapes the operation of technical layers rather than as an external approval process [9].

This article develops a conceptual model of AI-based energy system autonomy structured around four layers: perception, prediction, decision, and governance. The objective is not to propose a specific algorithm, simulation, or implementation platform, but to clarify how autonomy can be architected in a safety-critical and socially accountable energy context. The model is informed by research on autonomous grids, smart grid AI, reinforcement learning, digital twins, and trustworthy AI governance [1]. It contributes a conceptual vocabulary for distinguishing autonomy capabilities from mere automation and for mapping how governance requirements intensify as decision authority shifts toward AI systems [10].

Background

The evolution from automation to autonomy in energy systems

Energy automation has historically relied on predefined control rules, operator procedures, and deterministic protection schemes, whereas autonomy implies adaptive perception, anticipatory reasoning, and context-sensitive action selection. The autonomous grid literature describes this transition as a movement toward systems that can monitor, diagnose, and coordinate operations with reduced human intervention [1]. Smart grid AI surveys show that the expansion of distributed sensing and computational intelligence enables broader automation of monitoring, forecasting, and control functions [2]. However, autonomy differs from automation because it involves delegated decision scope, changing responsibility boundaries, and a need for explicit governance over system behavior [11].

Foundational AI capabilities: perception, prediction, and decision

AI-based energy autonomy depends on a sequence of capabilities that can be analytically grouped into perception, prediction, and decision. Perception includes state estimation, sensor interpretation, anomaly detection, computer vision inspection, and event classification, as shown in research on distribution system state estimation, wind turbine fault detection, and smart grid anomaly classification [7]. Prediction includes load forecasting, renewable generation forecasting, building energy prediction, and risk anticipation, which are central to probabilistic and adaptive energy management [12]. Decision functions include reinforcement learning, model-free control, energy scheduling, voltage regulation, and real-time operational optimization, where AI systems translate system intelligence into control recommendations or actions [13].

Beyond these core capabilities, emerging research highlights that perception-prediction integration shapes overall system robustness, particularly under data scarcity or cyber-physical stress [14]. Prediction-layer uncertainty quantification has been shown to directly influence decision-layer safety margins in grid control applications [15]. Furthermore, the coupling between perception fidelity and prediction horizon length determines the feasible degree of autonomous action in time-critical energy operations [16].

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

Governance of autonomous systems in critical infrastructure

Autonomous energy systems raise governance questions because technical actions can have consequences for reliability, affordability, safety, and public accountability. Energy-specific AI governance work has argued that electricity systems require oversight structures that account for operational risk and regulatory responsibility [4]. Broader AI governance research emphasizes that trustworthy autonomy requires auditable decision processes, human oversight, and risk-based accountability mechanisms [5]. Ethical AI principles are especially relevant in critical infrastructure because the acceptability of autonomous action depends not only on technical performance but also on transparency, fairness, contestability, and institutional legitimacy [6].

Layered architectures in AI and complex systems

Layered architectures are useful for understanding complex AI systems because they separate functions while preserving the dependencies between them. In energy contexts, digital twin research shows how sensing, modeling, simulation, and operational control can be organized into structured architectures that support monitoring and decision support [10]. Power system analytics surveys similarly suggest that machine learning capabilities become more manageable when differentiated according to data acquisition, modeling, inference, and operational use [11]. A layered approach is therefore appropriate for AI-based energy autonomy because it can distinguish what the system perceives, what it predicts, what it decides, and how those decisions are governed [9].

Gaps in current conceptualizations of energy system autonomy

Current discussions of energy autonomy often conflate higher automation with genuine autonomy, even though the latter involves delegated decision authority and responsibility allocation. Autonomous grid research identifies the importance of intelligence and self-management, but the relationship between technical autonomy and governance architecture remains insufficiently specified [1]. Governance-focused studies show that regulatory oversight is frequently discussed as a separate institutional layer rather than as an embedded feature of system design [4]. This creates a conceptual gap: energy autonomy requires a model that links AI capabilities to governance constraints, rather than treating prediction accuracy, control sophistication, or automation scope as sufficient indicators of autonomy [5].

Conceptual Foundation of the Layered Model

Defining energy system autonomy as a multi-layer phenomenon

Energy system autonomy can be conceptualized as a multi-layer phenomenon rather than a single technical property. At the perception layer, AI systems construct operational awareness from measurements, images, logs, and event signals; at the prediction layer, they generate anticipatory representations of demand, supply, risk, or system stress [7]. At the decision layer, AI systems select or recommend actions such as voltage control, dispatch adjustment, or energy scheduling, drawing on methods that include reinforcement learning and optimization [13]. Governance forms the fourth layer because autonomous decisions in energy systems require accountability, permissible action boundaries, and institutional authorization, not merely technical capability [4].

The principle of governance-by-design

Governance-by-design means that accountability, oversight, safety, and compliance should be designed into the autonomy architecture from the beginning. Ethical AI scholarship argues that trustworthy systems require principles such as transparency, responsibility, and human agency to be operationalized through system design rather than stated only as external values [6]. The European risk-based approach to AI regulation further indicates that high-impact autonomous systems require governance mechanisms proportional to their potential harm and decision authority [17]. In energy

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

systems, this means that governance should constrain perception quality, prediction use, decision boundaries, and fallback procedures as part of the architecture itself [4].

Table 1 contrasts governance-integrated autonomy with conventional automation across the four layers.

Table 1. Comparative Characteristics of Conventional Automation versus Governance-Integrated AI Autonomy in Energy Systems

Functional Layer	Conventional Automation (Rule-Based)	Governance-Integrated AI Autonomy
Perception	Threshold-based alerting from calibrated sensors	Probabilistic state estimation with uncertainty quantification and data trust scoring
Prediction	Deterministic look-up tables or simple time-series heuristics	Probabilistic forecasting with confidence envelopes, scenario generation, and uncertainty propagation
Decision	Pre-programmed if-then-else control logic	Policy learning (e.g., reinforcement learning) constrained by safety boundaries and authorization rules
Governance	External operator procedures and manual overrides	Embedded accountability logs, real-time boundary enforcement, audit trails, and fallback triggers

Layer dependencies and information flows

The layered model assumes that autonomy emerges through bidirectional dependencies between technical intelligence and governance constraints. Upward information flows convert raw data into state awareness, anticipated futures, and actionable choices, while downward constraints define what data quality is acceptable, which forecasts may be acted upon, and which decisions remain within authorized boundaries [9]. Reinforcement learning in energy systems illustrates the importance of this relationship because action selection depends on both learned system behavior and externally defined safety constraints [13]. Digital twin architectures further suggest that autonomy requires continuous interaction between monitored reality, modeled futures, and governed intervention pathways [10].

Figure 1 illustrates the governance-integrated layered architecture of AI-based energy system autonomy, showing how perception, prediction, decision, and governance layers interact through upward information flows and downward governance constraints to support safe and accountable autonomous operation.

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

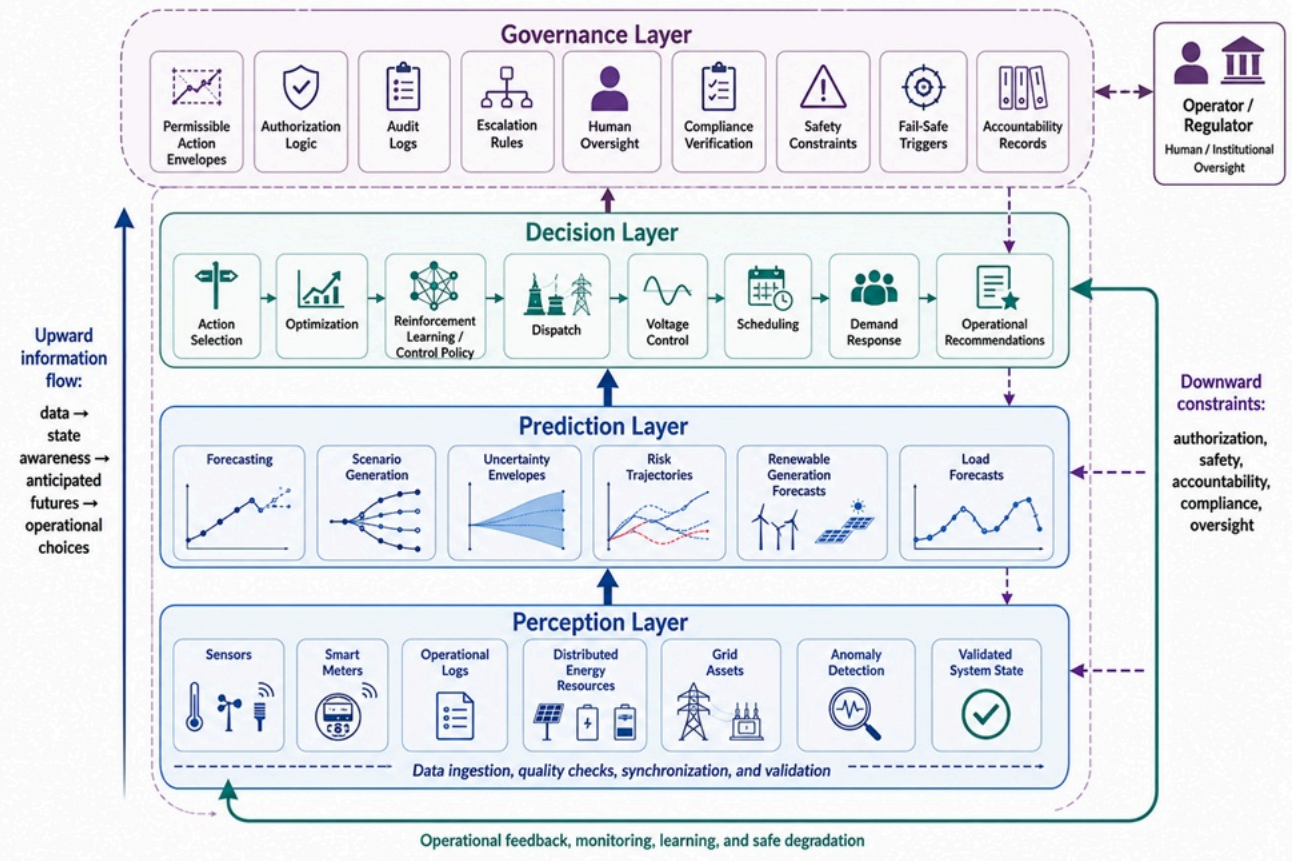


Figure 1. Governance-Integrated Layered Architecture for AI-Based Energy System Autonomy

The Layered Model Architecture

Perception layer: sensing and state awareness

The perception layer provides the informational foundation for autonomy by converting operational signals into usable representations of system state. In distribution systems, state estimation techniques support situational awareness by inferring network conditions from incomplete, uncertain, or heterogeneous measurements [7]. Deep learning approaches to state estimation, wind turbine fault detection, and distribution-grid event classification extend this layer by enabling pattern recognition across complex operational data [18]. Perception also includes anomaly and false-data attack detection, because autonomous decisions depend on the ability to distinguish valid system signals from corrupted, adversarial, or abnormal inputs [19].

Prediction layer: anticipation and scenario generation

The prediction layer transforms perceived system states into anticipatory knowledge about future demand, supply, risk, and operational stress. Deep learning surveys on load and renewable forecasting show how AI can support short-term and longer-horizon anticipation in smart microgrids and energy systems [12]. Building energy prediction research demonstrates that both single and ensemble models can contribute to operational foresight, although prediction outputs remain dependent on data quality and contextual assumptions [20]. For autonomy, prediction is not simply a forecasting function; it becomes a scenario-generation layer that enables decision systems to evaluate possible futures before selecting actions [21].

Decision layer: action selection and optimization

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

The decision layer converts perception and prediction into operational choices, including dispatch, voltage control, scheduling, demand response, and grid reconfiguration. Reinforcement learning has become especially important in this layer because it allows energy systems to learn action policies under changing conditions, while still requiring safeguards for real-time operational use [13]. Applications in voltage control, energy hubs, building optimization, electric vehicle charging, and demand response show that AI decision-making can be structured around control objectives, constraints, and adaptive learning [21]. However, decision autonomy in energy systems should always be interpreted through governance constraints, because technically optimal actions may be unsafe, inequitable, non-compliant, or unacceptable without institutional authorization [9].

Table 2 summarizes the four-layer functional architecture of AI-enabled energy autonomy, illustrating how perception, prediction, decision-making, and governance collectively support adaptive, safe, and institutionally accountable operation in modern energy systems.

Table 2. Functional Layers of AI-Enabled Energy System Autonomy and Their Operational Capabilities

Layer	Core AI Functions	Key Outputs	Autonomy Capabilities Enabled
Perception	Multi-source signal interpretation, state estimation, event recognition, anomaly discrimination	Validated system state, detected events, asset condition indicators, data trust flags	Situational awareness, fault visibility, input reliability screening, operational context formation
Prediction	Forecasting, probabilistic inference, scenario construction, risk anticipation	Demand and generation forecasts, stress trajectories, uncertainty envelopes, probable system futures	Anticipatory coordination, early warning, risk-sensitive planning, future-oriented operating posture
Decision	Policy learning, optimization, control selection, resource scheduling	Dispatch instructions, voltage actions, routing priorities, control recommendations	Automated action selection, adaptive control, constrained optimization, real-time operational response
Governance	Boundary definition, authorization logic, audit control, oversight coordination	Permissible action envelopes, escalation rules, accountability records, compliance evidence	Responsible autonomy, traceable delegation, safety containment, institutionally authorized operation

Layer Interactions and Autonomy Dynamics

Upward information flow: from data to decisions

Upward information flow begins when sensor measurements, operational logs, and external signals are translated into a coherent representation of system state. State estimation and anomaly detection studies show that this first transformation is not neutral, because errors, uncertainty, and adversarial data can shape all subsequent predictions and decisions [7]. Prediction then abstracts perceived conditions into future-oriented representations, such as load forecasts, risk trajectories, or renewable variability scenarios [12]. Decision systems use these representations to select actions, but their reliability depends on whether earlier layers have produced trustworthy and contextually meaningful intelligence [11].

Downward constraints: governance embedding

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

Downward constraints operate in the opposite direction by defining the boundaries within which perception, prediction, and decision layers may function. Energy AI governance research suggests that autonomous systems must be constrained by safety expectations, regulatory obligations, and human oversight requirements before decisions are executed [4]. General AI governance work further indicates that accountability mechanisms should define who is responsible for system design, operational monitoring, intervention, and post-event review [5]. In the layered model, governance therefore does not merely evaluate decisions after they occur; it shapes data admissibility, prediction use, decision authority, and escalation thresholds throughout the autonomy stack [17].

Feedback loops and learning across layers

AI-based energy autonomy is dynamic because decisions generate new system states, new data, and new learning opportunities. Safe reinforcement learning research shows that energy control systems must learn from operational feedback while remaining inside predefined safety boundaries [21]. Multi-agent and voltage-control studies demonstrate that autonomous decision processes can adapt across distributed grid environments, but this adaptation requires coordination mechanisms that prevent local optimization from undermining system-level reliability [22]. Digital twin research also suggests that feedback between monitored operations and modeled representations can support continuous improvement, provided that governance mechanisms determine when learning, updating, or autonomous action is permissible [10].

Autonomy Levels and Decision Authority

A Spectrum of autonomy for energy systems

AI-based energy system autonomy can be structured as a spectrum rather than a binary distinction between human control and machine control. At lower levels, AI may assist operators by detecting anomalies, forecasting risks, or recommending actions, while authority remains primarily human [23]. At intermediate levels, AI may perform selected control functions under supervision, as seen in reinforcement learning approaches to voltage control, energy scheduling, and demand response [24]. At higher levels, AI could operate within bounded envelopes where actions are pre-authorized, continuously monitored, and reversible when system conditions exceed approved limits [25].

Mapping autonomy levels to layer responsibility

The layered model clarifies that autonomy levels differ according to which layers are AI-dominated and which remain human-dominated. In operator-assisted autonomy, AI may dominate perception and partially support prediction, but decision authority and governance remain strongly human-centered [26]. In shared control, AI systems may generate forecasts and candidate actions, while operators retain approval authority over operationally significant decisions [27]. In supervised autonomy, AI may execute defined decisions within governance limits, but human actors retain supervisory, intervention, and accountability roles [28].

Governance shifts with increasing autonomy

As autonomy increases, governance must shift from direct human approval of individual actions toward pre-authorized boundaries, audit-based oversight, and dynamic condition monitoring. Model-free electric vehicle scheduling and demand response systems illustrate how AI can make operational decisions in complex environments, but such decisions require explicit constraints on acceptable action, user impact, and system reliability [28]. Testbed work on reinforcement learning-based energy management also indicates that autonomous decision support must be evaluated in relation to operational feasibility and control responsibility rather than algorithmic sophistication alone [29]. The governance challenge is therefore not merely whether AI can decide, but how decision authority is delegated, bounded, monitored, and withdrawn when conditions change [17]. **Figure 2** illustrates how increasing AI autonomy in EV scheduling and demand response requires governance to move from case-by-case human approval toward bounded delegation, continuous condition monitoring, audit-based oversight, and authority withdrawal when operational or social conditions change.

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

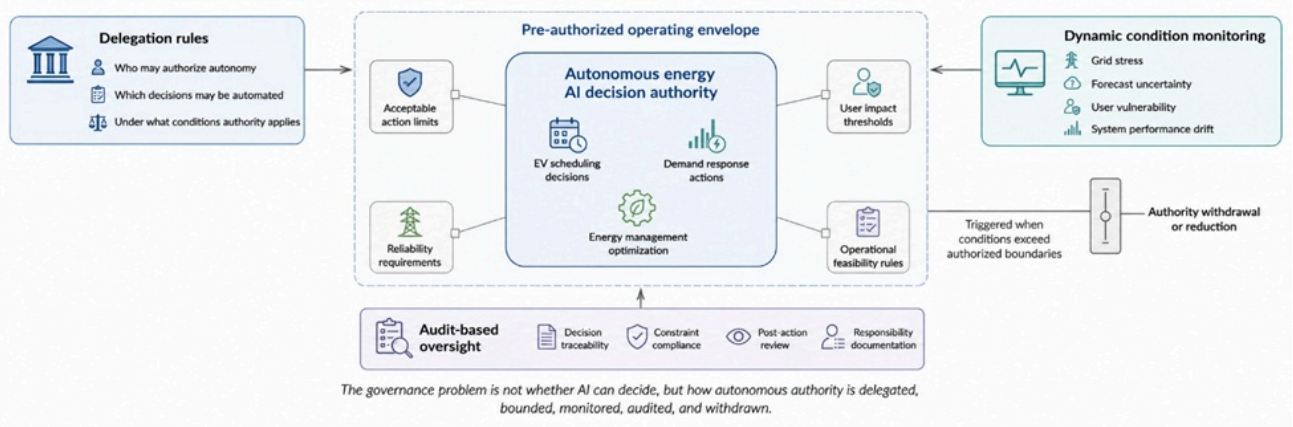


Figure 2. Bounded Delegation Governance for Autonomous Energy AI

Transition protocols and safe degradation

Autonomy levels should be designed to shift safely when operating conditions become abnormal, uncertain, or outside the validated domain of the system. Deep reinforcement learning methods for voltage control show that AI may support distributed and rapid action, but safe operation requires mechanisms for reverting, limiting, or escalating control when system conditions become unstable [25]. Multi-agent voltage-control research similarly demonstrates that distributed autonomy can introduce coordination challenges if local agents act without adequate system-wide constraints [22]. Safe degradation therefore requires predefined fallback modes, human takeover pathways, degraded-service procedures, and governance triggers that reduce autonomy before failures cascade [9].

Governance Layer: Accountability, Safety, and Regulation

Embedding accountability in the autonomy architecture

The governance layer should include explicit accountability structures that define who is responsible for system design, authorization, monitoring, intervention, and post-event review. AI governance scholarship emphasizes that responsibility cannot be assigned only after harm occurs, because accountable autonomy requires traceable decisions, documented authority, and clear institutional ownership [5]. In the layered model, decision logs, explanation records, and authorization histories become operational artifacts rather than administrative afterthoughts. This accountability architecture is especially important where autonomous control systems may affect grid reliability, consumer access, or market outcomes [4].

Safety assurance and risk containment

Safety assurance in autonomous energy systems requires the governance layer to monitor, constrain, and, when necessary, override technical decision processes. Safe reinforcement learning for smart energy hubs shows that AI control can be designed around safety boundaries, but these boundaries must be explicit rather than assumed [21]. Fault detection research in wind turbines and active distribution grids demonstrates that early detection can support safer intervention when abnormal conditions emerge [30]. A governance-integrated autonomy architecture would therefore treat risk containment as a continuous function that supervises perception validity, prediction uncertainty, and decision permissibility [31].

Figure 3 illustrates how governance-integrated autonomy can continuously supervise an autonomous energy system by monitoring perception validity, prediction uncertainty, and decision permissibility before allowing, constraining, or

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

overriding AI control actions.

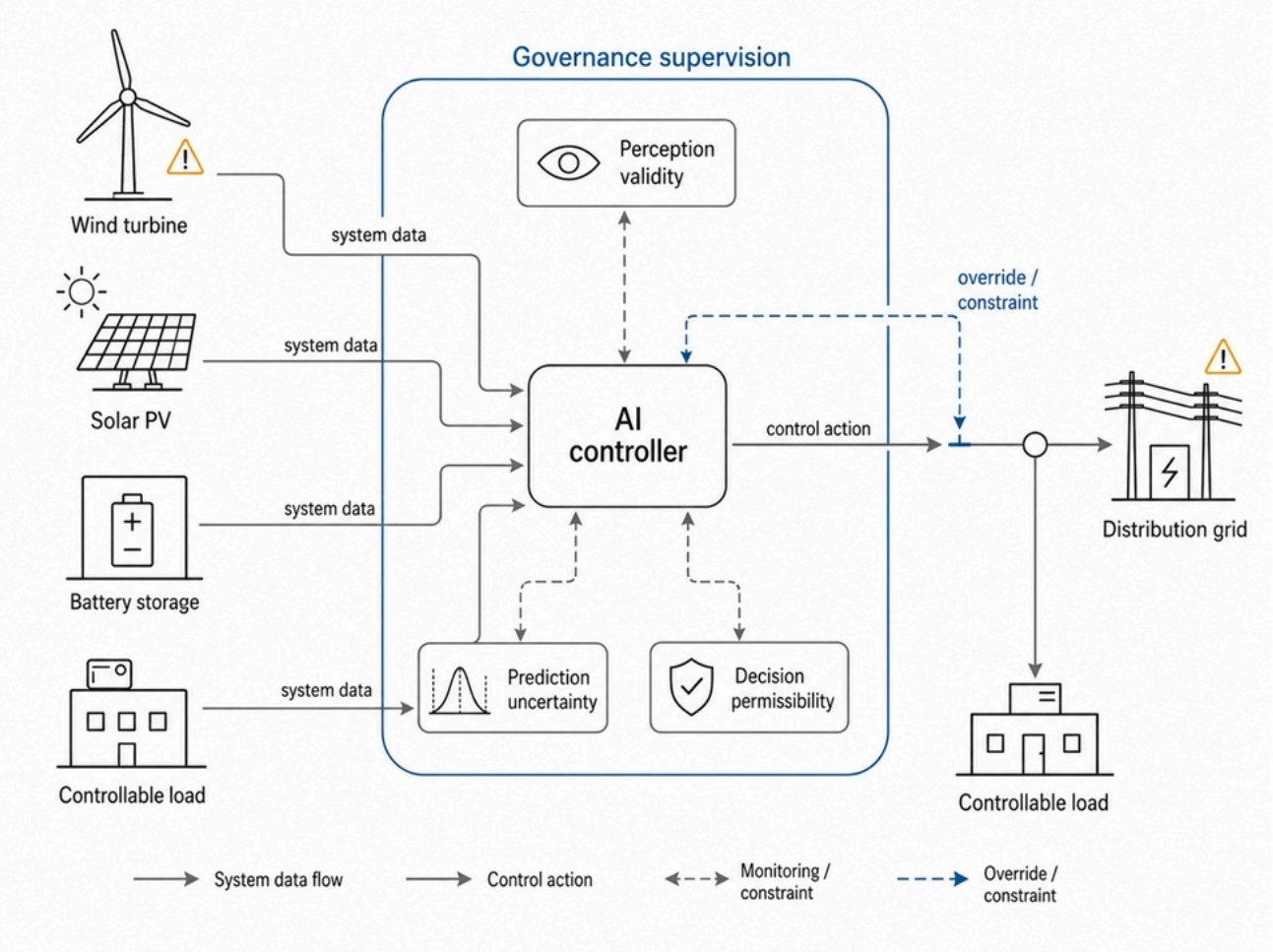


Figure 3. Governance-Supervised Safety Assurance in Autonomous Energy Systems

Regulatory alignment and compliance verification

Regulatory alignment requires autonomous energy systems to produce evidence that their decisions remain within approved technical and institutional boundaries. Discussions of AI governance in electricity systems show that energy-sector AI cannot be governed only through generic digital regulation because electricity systems have distinctive reliability, safety, and public-service obligations [4]. Trustworthy AI debates around risk acceptability further indicate that compliance should not be confused with social legitimacy or operational responsibility [17]. In the layered model, compliance verification can be structured through auditable prediction inputs, decision constraints, escalation rules, and records of human or institutional authorization [5].

Multi-stakeholder governance and public trust

Autonomous energy systems can affect consumers, utilities, markets, regulators, and communities, so governance should include multi-stakeholder accountability rather than only technical risk control. Ethical AI frameworks emphasize that transparency, justice, beneficence, and human agency are central to the public acceptability of AI systems [6]. Smart grid anomaly detection and false-data attack research further shows that technical security is also a public trust issue because compromised perception or decision layers can affect service continuity and institutional confidence [32]. Public

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

trust therefore requires that governance mechanisms make autonomous energy actions explainable, contestable, and aligned with social obligations as well as technical objectives [19].

Implementation Pathways and Sector Implications

Applying the layered model in utility and market contexts

The layered model can guide utilities, grid operators, aggregators, and market actors in specifying what kind of autonomy is being procured or deployed. For example, building energy optimization, electric vehicle charging, and demand response systems may all involve AI decision-making, but they differ in scale, risk, user impact, and required oversight [27]. In market contexts, AI may support bidding, scheduling, or flexibility aggregation, but governance must determine when automated action affects fairness, transparency, or system reliability [29]. The model therefore helps distinguish between AI as an analytical tool, AI as a control mechanism, and AI as a delegated operational actor [23].

Standards and certification for autonomous energy systems

The layered model could inform future standards and certification processes by defining layer-specific requirements for data validity, forecasting reliability, decision constraints, and governance evidence. Digital twin research suggests that cyber-physical energy systems increasingly depend on architectures that link real-time monitoring, simulation, and operational decision support [10]. Machine learning analytics surveys also indicate that energy AI applications are diverse, which makes cross-cutting certification difficult unless requirements are organized by function and risk [11]. A layered certification approach could therefore assess perception quality, prediction uncertainty, decision boundaries, and governance readiness separately before approving integrated autonomy [17].

Evaluation Strategy for the Layered Model

Conceptual clarity and explanatory utility

The layered model should be evaluated first by its conceptual clarity and explanatory utility rather than by predictive performance or operational metrics. It clarifies that perception, prediction, decision, and governance are analytically distinct, even though they interact in deployed systems [3]. Compared with non-layered descriptions of smart grid intelligence, the model makes visible where autonomy is located and where responsibility must be assigned [2]. This explanatory value is important because energy autonomy is often discussed through specific techniques, such as reinforcement learning or forecasting, without a broader architecture of delegated authority [8].

Expert validation and design workshops

Further development of the model would require structured expert validation with grid operators, AI developers, regulators, safety specialists, and energy governance scholars. Such validation should examine whether the four-layer structure captures real operational decision pathways and whether the autonomy levels correspond to plausible responsibility arrangements [1]. Co-design workshops could test whether operators can use the model to identify handoff points, escalation thresholds, and governance gaps in proposed AI deployments [4]. This would not constitute empirical performance evaluation, but it would support conceptual refinement and practical operationalization [5].

Generality across energy system types and scales

The model is intended to be general enough to apply across transmission systems, distribution networks, microgrids, building energy systems, electric vehicle coordination, and multi-energy platforms. Research on reinforcement learning across sustainable energy and electric systems shows that AI decision-making is already being explored across multiple operational scales [23]. Forecasting and state-estimation studies further indicate that perception and prediction

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

challenges vary by system type, data availability, and temporal resolution [12]. The model's generality depends on its ability to preserve the same four-layer logic while allowing each layer to be specified differently across technical domains and governance contexts [7].

Limitations

Conceptual abstraction and need for operational tools

The model remains a conceptual abstraction and does not provide detailed engineering procedures, regulatory templates, or validation metrics. Although conceptual abstraction is useful for clarifying autonomy architecture, energy AI deployment ultimately requires operational tools that translate layer functions into design specifications and assurance practices [11]. For example, reinforcement learning control systems require domain-specific constraints, monitoring procedures, and fallback protocols that cannot be fully specified by a general model alone [9]. The layered model should therefore be treated as a structuring framework requiring further operationalization rather than as a complete deployment methodology [21].

The challenge of rapid AI and regulatory evolution

A second limitation is that both AI capabilities and regulatory expectations are changing rapidly, which may require the model to be updated over time. Governance debates around trustworthy AI and the European Union AI Act show that the meaning of risk, oversight, and accountability continues to evolve as regulatory systems respond to new forms of autonomy [17]. At the same time, advances in digital twins, distributed learning, and autonomous control may blur the boundaries between perception, prediction, and decision layers [10]. The model therefore should be understood as a flexible conceptual architecture that can be revised as energy systems and AI governance mature [4].

Conclusion

AI-based energy system autonomy should be understood as a layered architecture rather than as a single level of automation. This article has proposed a model structured around perception, prediction, decision, and governance, showing how technical autonomy depends on both computational capability and institutional authorization. The central contribution is the integration of governance into the core architecture of autonomy rather than treating it as an external compliance process.

Autonomous energy systems without governance-by-design are unlikely to be trustworthy, safe, or publicly acceptable. Energy systems are critical infrastructures, and their autonomous functions must remain accountable to reliability obligations, social expectations, and regulatory responsibilities. The model therefore reframes autonomy as a socio-technical property shaped by the interaction between AI capabilities and governance constraints.

Future work should refine the model through interdisciplinary collaboration among energy engineers, AI researchers, system operators, regulators, and social scientists. Such collaboration can support the development of implementation guidelines, assurance methods, and governance-by-design practices for the next generation of autonomous energy systems. The long-term challenge is not simply to make energy systems more autonomous, but to ensure that autonomy remains safe, explainable, accountable, and aligned with public responsibility.

Acknowledgements

None

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

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 19 Dec 2022 Revised: 24 Feb 2023 Accepted: 13 Mar 2023
Published online: 15 June 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. Biagini V, Subasic M, Oudalov A, Kreusel J. The autonomous grid: automation, intelligence and the future of power systems. Energy Res Soc Sci. 2020;65:101460. https://doi.org/10.1016/j.erss.2020.101460.

2. Omitaomu OA, Niu H. Artificial intelligence techniques in smart grid: a survey. Smart Cities. 2021;4(2):548-68. https://doi.org/10.3390/smartcities4020029.

3. Cheng L, Yu T. A new generation of AI: a review and perspective on machine learning technologies applied to smart energy and electric power systems. Int J Energy Res. 2019;43(6):1928-73. https://doi.org/10.1002/er.4333.

4. Niet I, van Est R, Veraart F. Governing AI in electricity systems: reflections on the EU artificial intelligence bill. Front Artif Intell. 2021;4:690237. https://doi.org/10.3389/frai.2021.690237.

5. Taeihagh A. Governance of artificial intelligence. Policy Soc. 2021;40(2):137-57. https://doi.org/10.1080/14494035.2021.1928377.

6. 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.

7. Dehghanpour K, Wang Z, Wang J, Yuan Y, Bu F. A survey on state estimation techniques and challenges in smart distribution systems. IEEE Trans Smart Grid. 2019;10(2):2312-22.

8. Perera AT, Kamalaruban P. Applications of reinforcement learning in energy systems. Renew Sustain Energy Rev. 2021;137:110618. https://doi.org/10.1016/j.rser.2020.110618.

9. Chen X, Qu G, Tang Y, Low S, Li N. Reinforcement learning for selective key applications in power systems: recent advances and future challenges. IEEE Trans Smart Grid. 2022;13(4):2935-58. https://doi.org/10.1109/TSG.2022.3154718.

10. Jafari M, Kavousi-Fard A, Chen T, Karimi M. A review on digital twin technology in smart grid, transportation system and smart city: challenges and future. IEEE Access. 2023;11:17471-84. https://doi.org/10.1109/ACCESS.2023.3241588.

11. Miraftabzadeh SM, Longo M, Foiadelli F, Pasetti M, Igual R. Advances in the application of machine learning techniques for power system analytics: a survey. Energies. 2021;14(16):4776. https://doi.org/10.3390/en14164776.

12. Aslam S, Herodotou H, Mohsin SM, Javaid N, Ashraf N, Aslam S. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. Renew Sustain Energy Rev. 2021;144:110992. https://doi.org/10.1016/j.rser.2021.110992.

13. Cao D, Hu W, Zhao J, Zhang G, Zhang B, Liu Z, et al. Reinforcement learning and its applications in modern power and energy systems: a review. J Mod Power Syst Clean Energy. 2020;8(6):1029-42.

14. Zahid M, Ahmed F, Javaid N, Abbasi RA, Zainab Kazmi HS, Javaid A, et al. Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids. Electronics. 2019;8(2):122. https://doi.org/10.3390/electronics8020122.

15. Yang Y, Li W, Gulliver TA, Li S. Bayesian deep learning-based probabilistic load forecasting in smart grids. IEEE Trans Ind Inform. 2020;16(7):4703-13.

16. Al-Ali R, Bulej L, Kofroň J, Bureš T. A guide to design uncertainty-aware self-adaptive components in cyber-physical systems. Future Gener Comput Syst. 2022;128:466-89.

17. Laux J. Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act. AI Soc. 2024;39(6):2853-66. https://doi.org/10.1007/s00146-023-01777-z.

18. Raghuvamsi Y, Teeparthi K. A review on distribution system state estimation uncertainty issues using deep learning approaches. Renew Sustain Energy Rev. 2023;187:113752. https://doi.org/10.1016/j.rser.2023.113752.

19. Cui L, Qu Y, Gao L, Xie G, Yu S. Detecting false data attacks using machine learning techniques in smart grid: a survey. J Netw Comput Appl. 2020;170:102808.

20. Wang Z, Srinivasan RS. A review of artificial intelligence based building energy use prediction: contrasting the capabilities of single and ensemble prediction models. Renew Sustain Energy Rev. 2017;75:796-808.

21. Qiu D, Dong Z, Zhang X, Wang Y, Strbac G. Safe reinforcement learning for real-time automatic control in a smart energy-hub. Appl Energy. 2022;309:118403.

22. Gao Y, Wang W, Yu N. Consensus multi-agent reinforcement learning for volt-var control in power distribution networks. IEEE Trans Smart Grid. 2021;12(4):3594-604. https://doi.org/10.1109/TSG.2021.3058996.

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

23. Yang T, Zhao L, Li W, Zomaya AY. Reinforcement learning in sustainable energy and electric systems: a survey. Annu Rev Control. 2020;49:145-63.

24. Glavic M, Fonteneau R, Ernst D. Reinforcement learning for electric power system decision and control: past considerations and perspectives. IFAC-PapersOnLine. 2017;50(1):6918-27.

25. Cao D, Zhao J, Hu W, Yu N, Ding F, Huang Q, et al. Deep reinforcement learning enabled physical-model-free two-timescale voltage control method for active distribution systems. IEEE Trans Smart Grid. 2022;13(1):149-65. https://doi.org/10.1109/TSG.2021.3113085.

26. Huang B, Wang J. Applications of physics-informed neural networks in power systems: a review. IEEE Trans Power Syst. 2023;38(1):572-88. https://doi.org/10.1109/TPWRS.2022.3162473.

27. Wang W, Yu N, Gao Y, Shi J. Safe off-policy deep reinforcement learning algorithm for volt-var control in power distribution systems. IEEE Trans Smart Grid. 2020;11(4):3008-18. https://doi.org/10.1109/TSG.2019.2962625.

28. Hirth L, Mühlenpfordt J, Bulkeley M. The ENTSO-E Transparency Platform—a review of Europe's most ambitious electricity data platform. Appl Energy. 2018;225:1054-67.

29. Yayla A, Haghnegahdar L, Dincelli E. Explainable artificial intelligence for smart grid intrusion detection systems. IT Prof. 2022;24(5):18-24.

30. Helbing G, Ritter M. Deep learning for fault detection in wind turbines. Renew Sustain Energy Rev. 2018;98:189-98. https://doi.org/10.1016/j.rser.2018.09.012.

31. Rizeakos V, Bachoumis A, Andriopoulos N, Birbas M, Birbas A. Deep learning-based application for fault location identification and type classification in active distribution grids. Appl Energy. 2023;338:120932.

32. Siniosoglou I, Radoglou-Grammatikis P, Efstathopoulos G, Fouliras P, Sarigiannidis P. A unified deep learning anomaly detection and classification approach for smart grid environments. IEEE Trans Netw Serv Manag. 2021;18(2):1137-51

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