

SYSTEMATIC REVIEW

Open access

A Systematic Review of Foundation Models for Energy-System Forecasting, Grid-Edge Control, and Smart-Grid Decision Support

Hassan Rahman^{1*}, Tariq Mahmood¹

Abstract

Foundation models are transforming artificial intelligence by enabling pre-training, adaptation, and reuse across heterogeneous tasks. Their application in energy systems is emerging rapidly, creating a need for systematic synthesis across forecasting, control, and decision-support settings. This systematic review identifies, categorizes, and evaluates the use of foundation models in energy-system forecasting, grid-edge control, and smart-grid decision support. It focuses on how large language models, pre-trained transformers, multi-modal models, and emerging time-series foundation models are adapted for energy tasks. A PRISMA-guided search of peer-reviewed literature from 2017 to 2025 resulted in 30 included studies. Models were classified by architecture, fine-tuning method, data modality, and application domain. The review finds that large language models dominate decision-support applications, while pre-trained transformers and time-series foundation models are most developed in forecasting. Grid-edge control applications remain nascent, with most studies addressing planning, diagnostics, or dispatch rather than closed-loop physical deployment. Foundation models hold significant promise for smart grids, but critical gaps remain in real-time control, safety assurance, privacy, and evaluation. A research agenda is proposed to guide rigorous, trustworthy, and operationally relevant deployment.

Keywords Systematic review, Decision support, Grid-edge control, Smart grid, Foundation models, Energy forecasting

*Correspondence:

Hassan Rahman
hassan.rahman@gmail.com

¹ Department of Energy Systems and Optimization, Faculty of Engineering, University of Karachi, Karachi, Pakistan

Introduction

Foundation models have become a defining development in artificial intelligence because they can be pre-trained on large corpora and adapted to multiple downstream tasks with limited task-specific data. In energy systems, this paradigm is beginning to influence applications ranging from electricity forecasting to building-energy modeling and operator support, as shown by early sector-level analyses of large language models in electric energy systems [1]. The energy domain is especially attractive for foundation models because it combines structured time series, engineering documents, geospatial assets, weather information, and operational constraints. Yet this same heterogeneity makes systematic synthesis necessary rather than optional.

The smart-grid literature now includes large language models for dispatch, building-energy automation, renewable-energy strategy, and human-facing decision support. For example, advanced power dispatch has been explored through a large-language-model approach [2], while building-energy modeling studies have used LLM-based workflows to automate

model construction, debugging, and prompt-guided simulation tasks [3]. These studies suggest that foundation models can reduce engineering effort and improve accessibility, but they also introduce risks associated with hallucination, opaque reasoning, and incorrect operational recommendations. A systematic review is therefore needed to separate demonstrated capability from speculative expectation.

Forecasting is the most mature application area because transformer-based architectures have already been tested for electricity load, wind power, and building-energy prediction. Transformer and temporal-fusion-transformer studies report strong performance in load forecasting [4], while newer work extends foundation-model concepts to wind-speed and wind-power forecasting using spatio-temporal large language models and cross-modal architectures [5]. These developments create continuity between earlier deep-learning forecasting research and newer foundation-model approaches. However, the degree to which such models outperform conventional baselines under robust cross-domain evaluation remains unevenly reported.

This review contributes a structured map of 30 peer-reviewed journal articles published between 2017 and 2025, focusing on foundation models for forecasting, grid-edge control, and smart-grid decision support. It follows PRISMA principles to make the search and screening process transparent, while also categorizing model families and application domains in a synthesis-oriented way [6]. The review does not present new empirical experiments; instead, it evaluates how the literature frames model adaptation, evaluation, safety, and operational readiness. Its central contribution is to clarify where foundation models are already useful, where evidence remains thin, and what research agenda is needed for trustworthy deployment in energy systems.

Figure 1 presents the review's evidence-to-application synthesis, showing how foundation-model families are distributed across energy forecasting, grid-edge control, and smart-grid decision-support use cases.

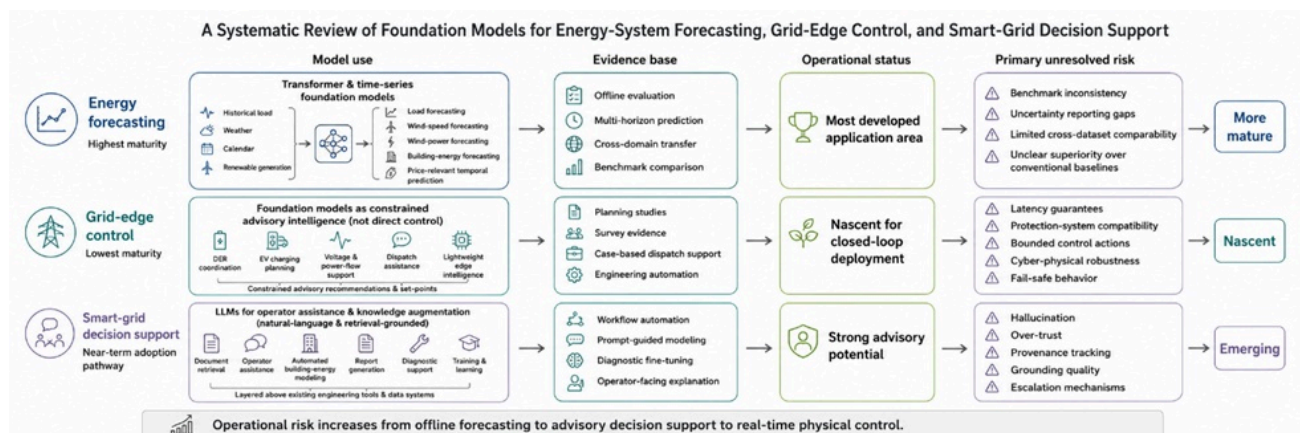


Figure 1. Evidence-to-Application Map of Foundation Models in Energy Systems

Research Questions

Primary research questions

The review asks three primary questions: how are foundation models being adapted for energy-system forecasting, how are they being explored for grid-edge control, and how are they used for smart-grid decision support? These questions reflect the observed spread of LLM-based energy studies from sector-wide capability assessments [1] to smart-grid integration reviews [7] and power-system surveys [8]. A fourth cross-cutting question examines how the studies evaluate trustworthiness, including hallucination, privacy, computational burden, and comparison with non-foundation-model baselines. Together, the questions define the review as an application-oriented and governance-aware synthesis.

Rationale for questions

The rationale for these questions is that forecasting, control, and decision support represent different levels of operational risk in energy systems. Forecasting studies such as transformer-based load prediction [4] and wind-speed foundation modeling [5] can often be evaluated offline, whereas dispatch and grid-edge control require stronger guarantees because incorrect recommendations may affect reliability. Decision-support studies add a further layer of concern because natural-language systems may persuade operators even when their outputs are incomplete or poorly grounded [6]. The questions therefore connect model capability to the operational context in which errors would matter.

Search Strategy

Databases and query strings

The search strategy targeted peer-reviewed journal literature in energy systems, smart grids, artificial intelligence, and applied engineering. Search strings combined foundation-model terms with energy-domain terms, including “foundation model” “energy forecasting” “grid,” “large language model” “smart grid” “control” OR “decision support,” and “pre-trained transformer” “electricity” “control.” These strings were designed to capture both explicit foundation-model studies and transformer-based forecasting work that forms the methodological bridge into foundation-model energy applications [9]. The target journals included energy, AI, engineering, smart-grid, and sustainability outlets aligned with the scope of the review.

Search execution and initial results

The search was executed for the 2017–2025 window, with the final cut-off set at 31 December 2025. Initial retrieval prioritized journal articles that directly addressed LLMs, transformers, multi-modal models, or generative AI in energy applications, including reviews of smart-grid LLM integration [6] and domain-specific renewable-energy LLM work [10]. Studies were retained for full screening when they contained a clear energy-system application rather than a generic AI discussion. The search also captured adjacent building-energy studies because automated building-energy modeling and HVAC diagnostics represent important grid-edge and demand-side intelligence contexts [11].

Data extraction plan

Data extraction recorded publication year, journal, model family, application domain, input data modality, adaptation strategy, evaluation method, and reported limitations. For forecasting studies, extraction emphasized whether the task concerned load, building energy, wind power, or multi-voltage demand, as reflected in studies on transfer learning for building-energy forecasting [12] and parameter-efficient cross-domain load forecasting [13]. For decision-support and control-oriented studies, extraction emphasized grounding, workflow integration, dispatch relevance, and safety concerns. This plan allowed the synthesis to compare not only performance claims but also methodological maturity and operational readiness.

Screening Criteria

Inclusion criteria

Studies were included when they were peer-reviewed journal articles published between 2017 and 2025 and addressed foundation models, large language models, pre-trained transformers, multi-modal models, or closely related transformer-based architectures in energy contexts. Articles on electricity load forecasting, wind forecasting, building-energy automation, power dispatch, smart-grid decision support, and AI governance were eligible when the energy application was explicit. For example, transformer-based load forecasting [14], LLM-based building-energy modeling [11], and large-

model power-system surveys [8] all met the inclusion logic. Reviews were included when they synthesized foundation-model, generative-AI, or LLM applications relevant to energy or smart grids.

Exclusion criteria and screening process

Studies were excluded if they were books, theses, reports, standards, websites, preprints without journal publication, or conference papers. Articles were also excluded when they used conventional deep learning without transformer or foundation-model relevance, addressed AI only at a conceptual level without an energy-system application, or fell outside the 2017–2025 publication window. The screening process prioritized full bibliographic completeness, including journal venue and DOI, and retained exactly 30 articles for synthesis. The resulting corpus includes broad reviews [15], application-specific forecasting studies [16], and decision-support or modeling studies that demonstrate how foundation models are entering energy workflows [17].

Table 1 shows a systematic PRISMA-guided screening process that refined 184 initial records down to 30 peer-reviewed studies focused on foundation models, transformers, LLMs, or multi-modal AI applications in energy systems.

Table 1. PRISMA Study Selection Flow

PRISMA stage	Records or studies	Count	Screening action	Main reasons for exclusion
Records identified through targeted database searches	Search hits from energy, AI, smart-grid, and engineering journals	184	Titles and abstracts collected using the prescribed query strings	Not applicable at identification stage
Duplicate and clearly irrelevant records removed	Records removed before screening	42	Duplicates, non-energy AI papers, and non-journal items removed	Duplicate indexing; generic AI without energy-system relevance
Records screened by title and abstract	Records screened	142	Relevance to foundation models, transformers, LLMs, or multi-modal AI in energy assessed	Conventional machine learning only; unrelated smart-city or ICT focus
Full-text articles assessed for eligibility	Full texts assessed	57	Journal status, year, DOI availability, and application-domain relevance checked	Conference papers, preprints, missing DOI, outside 2017–2025 window
Full-text articles excluded	Studies excluded after full-text screening	27	Exclusions recorded after detailed methodological review	No foundation-model relevance; insufficient energy-system specificity; non-peer-reviewed status
Studies included in systematic synthesis	Included studies	30	Final set used for thematic classification and synthesis	Not applicable because these studies met all criteria

Foundation Model Categories

Pre-trained transformers

Pre-trained transformers form the largest methodological bridge between conventional deep learning and foundation models in the reviewed forecasting literature. Electrical-load forecasting studies using transformer-based architectures [4]

and adaptive transformer frameworks with domain knowledge [9] show how attention mechanisms capture temporal dependencies and exogenous influences in power demand. Temporal fusion transformer approaches further add interpretability and multi-horizon forecasting capacity, which is valuable for operational planning in electricity systems [14]. These models are not always foundation models in the strictest general-purpose sense, but they establish the architectural basis for later energy-specific foundation models.

Large language models

Large language models appear most prominently in decision support, engineering automation, and dispatch-oriented applications. Sector-level assessments show that LLMs can assist with energy-system reasoning, code generation, documentation, and operator-facing explanation, while also exposing reliability and hallucination risks [1]. LLM-based dispatch research extends this capability toward operational decision support [2], and smart-grid integration reviews identify potential uses in control-room assistance, maintenance support, planning, and training [7]. Across the reviewed literature, LLMs are strongest where natural language, procedural knowledge, or engineering documentation must be connected to analytical workflows.

Multi-modal and emerging architectures

Multi-modal and emerging foundation architectures combine time-series, textual, spatial, and domain-specific signals to improve generalization across energy tasks. Cross-modal wind-power forecasting [18] and time-series-prompted foundation models for wind prediction [16] illustrate how renewable-generation forecasting is moving beyond single-stream numerical modeling. Building-energy studies also show emerging agentic and prompt-engineered workflows that integrate text instructions, simulation files, and diagnostic reasoning [19]. These architectures remain less standardized than LLMs or transformers, but they are central to future grid intelligence because energy operations depend on heterogeneous data rather than isolated time series.

Table 2 shows a comprehensive classification of foundation models in energy literature, highlighting model types, data sources, adaptation strategies, and their dominant applications across forecasting, grid management, and operator support.

Table 2. Classification of Foundation Models in Energy Literature

Foundation model category	Representative reviewed studies	Typical model scale or complexity	Main pre-training or input data	Fine-tuning or adaptation strategy	Dominant energy applications
Pre-trained transformers	L'Heureux, Grolinger, Capretz [4]; Gao, Chen, Hu, Zhang [9]; Huy, Minh, Tien, Anh [14]	Medium to large sequence models with attention layers	Historical electricity load, weather, calendar variables, operational covariates	Supervised task-specific training, transfer learning, temporal feature engineering	Short-term load forecasting, building-energy forecasting, demand prediction
Temporal fusion and interpretable transformers	Ferreira, Leite, Salvadeo [20]; Spencer, Ranathunga, Boulic, van Heerden, Susnjak [12]	Multi-horizon interpretable transformer architectures	Substation load, building-energy consumption, exogenous predictors	Transfer learning, interpretable attention, comparative baseline evaluation	Substation forecasting, building-energy prediction, explainable demand modeling

Large language models	Majumder, Dong, Doudi, Cai, Tian, Kalathil, Ding, Thatte, Li, Xie [1]; Cheng, Zhao, Zhou, Zhao, Cao, Yang, Cai [2]; Madani, Tavasoli, Khoshtarash Astaneh, Pineau [7]	Large general-purpose language models adapted to energy tasks	Technical documents, dispatch rules, energy-system prompts, operator queries	Prompt engineering, instruction tuning, retrieval augmentation, domain-specific fine-tuning	Dispatch support, operator assistance, smart-grid decision support
Domain-specific energy LLMs	Gabber, Hemied [10]; Zhang, Zhang, Lu, Zhao [21]; Jiang, Chen [22]	Fine-tuned or parameter-efficient LLM variants	Renewable-energy documents, HVAC fault data, building-energy modeling files	Supervised fine-tuning, parameter-efficient fine-tuning, labeled-data adaptation	Renewable-energy strategy, HVAC diagnostics, automated building-energy modeling
Multi-modal and cross-modal models	Wu, Ling [5]; Zhu, Jia, Xing, Xiang, Hu, Hao [18]; Fan, Song, Feng, Liu, Jiang [16]	Cross-modal or spatio-temporal large models	Wind speed, weather, temporal prompts, multi-source renewable-generation data	Cross-modal alignment, time-series prompting, exogenous-form tuning	Wind-speed forecasting, wind-power forecasting, renewable-generation prediction
Emerging graph and smart-grid foundation architectures	Saeed, Rehman, Shah, Diyan, Chen, Kang [23]; Banad, Sharif, Rezaei [24]	Graph-transformer or digitally integrated AI architectures	Grid topology, load signals, digital-twin data, smart-grid operational context	Graph representation learning, digital-twin integration, task-specific adaptation	Load forecasting, smart-grid intelligence, planning-oriented control support

Forecasting Applications

Load and demand forecasting

Load and demand forecasting is the most developed application domain in the reviewed corpus, with transformer-based and LLM-adapted models used to improve temporal representation, cross-domain transfer, and data efficiency. Zero-shot load forecasting for integrated energy systems demonstrates how LLM-based multi-task learning can support prediction when labeled target-domain data are limited [25]. Short-term electricity load forecasting studies also show that external factors can be weighted and optimized alongside large-language-model components to improve demand prediction [26]. More broadly, survey evidence indicates that deep-learning load forecasting is moving toward larger, more transferable architectures, although benchmark comparability remains inconsistent [27].

Renewable generation forecasting

Renewable generation forecasting is a second major area of foundation-model application because wind and solar output depend on complex temporal, meteorological, and spatial interactions. Spatio-temporal enhanced pre-trained large language models have been used for wind-speed forecasting, showing how energy-specific sequence modeling can extend the LLM paradigm beyond text [5]. Cross-modal large language models for wind-power forecasting further indicate that multi-source feature alignment may improve renewable prediction under variable weather conditions [18].

Generative-AI work on renewable energy forecasting and system optimization suggests that these approaches are increasingly being framed as part of broader renewable-energy intelligence systems rather than isolated predictors [28].

Electricity price forecasting

Electricity price forecasting is less directly represented than load and renewable forecasting in the selected corpus, but the reviewed studies still provide relevant methodological foundations. Transformer-based electricity forecasting research shows how attention models can process temporal volatility, calendar effects, and exogenous drivers that are also central to price dynamics [29]. Adaptive deep-learning frameworks that combine transformers with domain knowledge are particularly relevant because electricity prices are shaped by operational constraints, demand patterns, and market conditions [9]. The evidence base therefore supports price forecasting as a promising but underdeveloped foundation-model application area within the 2017–2025 literature.

Table 3 shows a synthesis of studies applying foundation models to diverse energy forecasting tasks, detailing model types, input modalities, reported performance gains, and key limitations for short-term load, building-energy, renewable, and price-relevant predictions.

Table 3. Foundation Models for Energy Forecasting: Summary of Studies

Forecasting task	Representative studies	Foundation model type	Input data and modality	Key performance outcome reported in the literature	Main limitation identified
Short-term load forecasting	L'Heureux, Grolinger, Capretz [4]; Huy, Minh, Tien, Anh [14]; Chan, Yeo [29]	Transformer and temporal fusion transformer	Historical load, calendar variables, weather, temporal covariates	Improved sequence representation and multi-horizon forecasting compared with conventional recurrent baselines	Limited cross-dataset comparability and uneven reporting of uncertainty
Cross-domain and zero-shot load forecasting	Li, Li, Yang, Xi, Yu, Xiao, Yang [25]; Liu, Bai, Zhao, Wen, Wang, Gu, Liu, Liang, Zhao, Dong [13]	LLM-based and parameter-efficient large model frameworks	Multi-voltage load data, integrated energy-system demand, auxiliary task signals	Better adaptation under limited labeled data and stronger transfer across domains	Real-world deployment evidence remains limited
Building-energy forecasting	Spencer, Ranathunga, Boulic, van Heerden, Susnjak [12]	Transfer-learning transformer	Building energy consumption, environmental variables, operational records	Transfer learning improves reuse across buildings and reduces full retraining burden	Generalization depends on building similarity and data quality
Wind-speed and wind-power forecasting	Wu, Ling [5]; Zhu, Jia, Xing, Xiang, Hu, Hao [18]; Fan, Song, Feng, Liu, Jiang [16]	Spatio-temporal LLM, cross-modal LLM, large time-series foundation model	Wind speed, meteorological data, temporal prompts, exogenous renewable signals	Cross-modal and prompt-based adaptation improves renewable forecasting flexibility	Multi-modal alignment and physical interpretability remain open challenges

Price-relevant temporal forecasting	Gao, Chen, Hu, Zhang [9]; Chan, Yeo [29]	Transformer with domain knowledge	Electricity time series, domain features, external covariates	Attention-based models provide transferable methods for volatile electricity-market prediction	Direct price-specific foundation-model studies remain scarce
-------------------------------------	--	-----------------------------------	---	--	--

Grid-Edge Control Applications

Voltage and power flow control

Grid-edge control is the least mature domain in the reviewed literature, particularly for real-time voltage and power-flow applications. Large-model surveys of power systems identify control and dispatch as promising application areas, but they also emphasize that most current studies remain closer to advisory intelligence than autonomous closed-loop control [8]. Smart-grid integration research similarly treats LLMs as tools for coordination, planning, and operator support rather than direct actuation at field devices [7]. This distinction is critical because voltage and power-flow control require bounded actions, latency guarantees, and physical safety constraints that are not yet consistently demonstrated.

DER and EV charging coordination

Distributed energy resource and electric-vehicle coordination appear mainly through planning, optimization, and digital-intelligence discussions rather than through field-validated LLM controllers. Artificial-intelligence and machine-learning reviews for smart grids describe how digital twins and LLM-driven intelligence may support DER coordination, scenario assessment, and operational planning [24]. LLM-based advanced power dispatch also provides evidence that language-model interfaces can assist complex operational reasoning when paired with structured power-system information [2]. However, the reviewed corpus does not yet establish that LLMs can safely replace conventional optimization or model-predictive control for DER or EV charging decisions.

Lightweight deployment at edge

Lightweight deployment at the grid edge is an emerging requirement because substations, buildings, EV chargers, and DER controllers often operate under latency, bandwidth, and privacy constraints. Parameter-efficient fine-tuning for automated building-energy modeling shows one route toward reducing adaptation cost while retaining domain capability [22]. Cross-domain multi-voltage load forecasting also illustrates how parameter-efficient large-model frameworks may support deployment across heterogeneous grid environments [13]. Nevertheless, edge deployment remains more a design objective than a proven operational capability in the foundation-model energy literature.

Table 4 shows a summary of studies leveraging foundation models for grid-edge control, highlighting model types, control tasks, evaluation approaches, and readiness levels across power dispatch, smart-grid integration, voltage management, and demand-side intelligence.

Table 4. Foundation Models for Grid-Edge Control: Summary of Studies

Grid-edge control area	Representative studies	Foundation model used	Control or coordination task	Evaluation metrics or evidence type	Readiness assessment
Power dispatch support	Cheng, Zhao, Zhou, Zhao, Cao, Yang, Cai [2]	Large language model	Advanced power dispatch	Case-based dispatch evaluation and	Promising for advisory dispatch,

			recommendation and reasoning	comparison with operational expectations	not yet sufficient for autonomous control
Smart-grid integration and coordination	Madani, Tavasoli, Khoshtarash Astaneh, Pineau [7]; Banad, Sharif, Rezaei [24]	LLMs, digital-twin-linked AI, emerging smart-grid foundation models	DER coordination, operational planning, and grid-intelligence integration	Review-based synthesis of opportunities, architectures, and risks	Conceptual to early applied stage
Voltage and power-flow control support	Yao, Fang, Chen, Liu, Mo, Ao [8]	AI large models for power systems	Control assistance, power-system operation, and planning support	Survey evidence and application mapping	Underdeveloped for closed-loop real-time control
Building and demand-side edge intelligence	Jiang, Chen [22]; Zhang, Ford, Chen, Chen [19]	Fine-tuned LLMs and agentic LLM workflows	Building-energy model automation, debugging, and demand-side operational support	Workflow accuracy, automation capability, and engineering-effort reduction	More mature for engineering automation than physical control
Multi-voltage and local forecasting for control context	Liu, Bai, Zhao, Wen, Wang, Gu, Liu, Liang, Zhao, Dong [13]	Parameter-efficient large language model framework	Forecasting support for local grid operation and voltage-level demand planning	Cross-domain forecasting performance	Useful for control inputs, but not a controller itself

Decision-Support Applications

Operator assistance and natural language interfaces

Decision support is the clearest application area for large language models because grid operators and engineers work with procedures, reports, logs, manuals, and planning documents. Sector-level analysis of LLMs in electric energy systems shows that these models can support explanation, summarization, coding, and information retrieval, while also requiring careful safeguards against inaccurate outputs [1]. Reviews of smart-grid LLM opportunities highlight operator-facing natural-language interfaces as a likely near-term use case because they can sit above existing control and analysis tools [6]. This makes decision support a comparatively lower-risk entry point than autonomous grid-edge actuation.

Automated planning and report generation

Automated planning and report generation are represented most strongly in the building-energy and renewable-energy subsets of the corpus. LLM-based automated building-energy modeling platforms demonstrate how natural-language systems can reduce manual work in simulation setup and engineering documentation [11]. Prompt-engineering studies further show that LLM outputs can be steered for building-energy modeling tasks when prompts encode domain context and expected computational steps [17]. In renewable energy, domain-specific LLM work suggests that strategy generation and deployment planning can benefit from models adapted to energy terminology and policy-relevant context [10].

Knowledge retrieval and training

Knowledge retrieval and training applications rely on the ability of foundation models to connect user questions with technical knowledge, manuals, and domain-specific procedures. Reviews of LLMs in electrical power and energy systems identify training, troubleshooting, documentation support, and knowledge access as recurring use cases [15]. HVAC fault diagnosis using domain-specific LLM fine-tuning shows how labeled engineering data can adapt language models for diagnostic support rather than generic conversation [21]. However, these applications require retrieval grounding, provenance tracking, and escalation mechanisms because persuasive natural-language answers may otherwise conceal uncertainty.

Table 5 shows that foundation models, especially LLMs, are increasingly applied in smart-grid decision support to aid operator reasoning, dispatch planning, building-energy modeling, and renewable-energy strategy, while their effectiveness depends critically on validation, data quality, and governance safeguards.

Table 5. Foundation Models for Smart-Grid Decision Support: Summary of Studies

Decision-support task	Representative studies	Model or architecture	User or operational context	Main contribution	Key risk
Operator assistance	Majumder, Dong, Doudi, Cai, Tian, Kalathil, Ding, Thatte, Li, Xie [1]; Shi, Fang, Chen, Gu, Ma, Zhang, Zhang, Gu, Lim [6]	General-purpose and energy-adapted LLMs	Control-room reasoning, explanation, and workflow assistance	Shows how LLMs can support operator-facing interpretation and knowledge access	Hallucination and over-trust
Dispatch decision support	Cheng, Zhao, Zhou, Zhao, Cao, Yang, Cai [2]	Large language model for power dispatch	Power-system dispatch planning and recommendation	Demonstrates the feasibility of LLM-assisted dispatch reasoning	Requires validation against physical and security constraints
Automated building-energy modeling	Zhang, Chen, Ford [3]; Jiang, Ma, Zhang, Chen [11]; Zhang, Ford, Chen, Chen [19]	LLM platform and agentic LLM workflow	Building-energy engineers and modelers	Automates model development, debugging, and workflow support	Simulation errors may propagate into planning decisions
Domain-specific diagnostics	Zhang, Zhang, Lu, Zhao [21]	Fine-tuned domain-specific LLM	HVAC fault diagnosis and maintenance decision support	Uses labeled data to specialize LLMs for engineering diagnostics	Data coverage and fault-label quality constrain reliability
Renewable-energy planning and education	Gabber, Hemied [10]; Erdiwansyah, Mamat, Syafrizal, Ghazali, Basrawi, Rosdi [28]	Domain-specific LLM and generative AI	Renewable deployment planning, forecasting interpretation, and system optimization	Frames generative AI as a support layer for renewable-energy strategy	Governance and evidence quality remain uneven

Figure 2 compares the operational readiness of foundation-model applications, highlighting the stronger evidence base for forecasting and decision support relative to the still-nascent status of closed-loop grid-edge control.

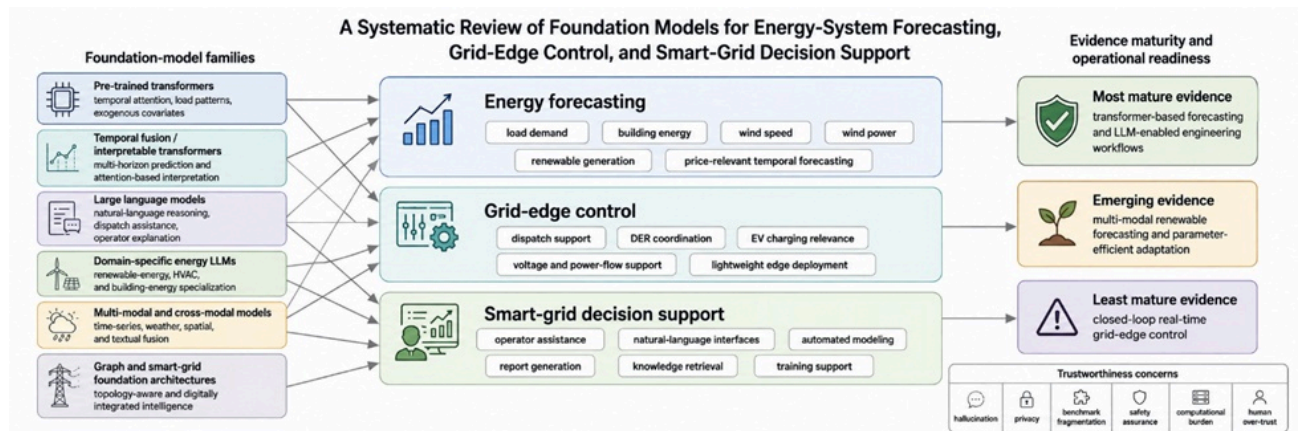


Figure 2. Comparative Operational Readiness of Foundation-Model Applications in Smart-Grid Contexts

Gaps

Real-time and safety gaps

The most important gap is the distance between offline model performance and safe real-time operation. Forecasting studies often report accuracy improvements, but grid-edge control requires latency guarantees, fail-safe behavior, and compatibility with protection systems [20]. LLM-based dispatch and smart-grid reviews acknowledge operational potential, yet they rarely demonstrate formal safety envelopes or adversarial robustness under stressed-grid conditions [2]. This gap limits the immediate use of foundation models in autonomous voltage, DER, or emergency-control settings.

Data privacy and hallucination

Data privacy and hallucination are cross-cutting risks because energy systems contain sensitive operational data, customer information, asset conditions, and security-relevant procedures. The energy hunger of AI also raises sustainability concerns because large models may increase computational demand while being proposed as tools for energy optimization [30]. LLM reviews in electrical power systems highlight hallucination, explainability, cybersecurity, and trust as unresolved barriers to adoption [15]. These risks are particularly serious in decision support because an incorrect answer may appear confident and operationally plausible.

Evaluation benchmarks and standards

Evaluation remains fragmented across datasets, metrics, baselines, and reporting conventions. Load-forecasting surveys show that deep-learning studies often differ in horizon, data preprocessing, train-test splits, and benchmark selection, making it difficult to determine whether larger models genuinely improve generalization [27]. Building-energy and transformer studies provide useful comparative evidence, but they do not yet form a standardized benchmark ecosystem for foundation models [12]. The field therefore needs common testbeds that evaluate accuracy, uncertainty, robustness, privacy, energy cost, and human factors together.

Future Agenda

Research priorities

Future research should prioritize energy-specific foundation models that combine physical constraints, temporal generalization, and safe human interaction. Forecasting work should move from isolated datasets toward cross-region, cross-voltage, and cross-asset evaluation, building on zero-shot and multi-voltage load-forecasting studies [25]. Control-

oriented research should treat LLMs as constrained reasoning components embedded within verified optimization and protection frameworks rather than as unconstrained autonomous agents [8]. Decision-support research should focus on grounded, auditable, and operator-calibrated tools.

Methodological advances needed

Methodological advances are needed in retrieval grounding, parameter-efficient adaptation, uncertainty quantification, and physics-aware model integration. Fine-tuning studies in building-energy modeling show that efficient adaptation can reduce the cost of domain specialization [22], while prompt-engineered building-energy workflows demonstrate that model behavior can be shaped through structured interaction design [17]. Renewable forecasting research should further explore multi-modal alignment, time-series prompting, and domain-aware pre-training, as suggested by recent wind-power foundation-model studies [16]. Across all domains, evaluation should include ablation studies that separate model scale from data quality, prompt design, and domain constraints.

Policy and industry implications

Policy and industry adoption will depend on whether foundation models can be governed as critical-infrastructure technologies rather than generic software tools. Smart-grid AI reviews emphasize that LLM-driven intelligence must be integrated with cybersecurity, privacy, reliability, and accountability requirements [24]. Industry pilots should begin with advisory decision support, automated documentation, and forecasting augmentation before moving toward safety-critical control. Regulators and system operators should require traceability, human override, model-update governance, and incident reporting for any foundation-model system used in grid operations.

Figure 3 translates the review's identified gaps into a research agenda for trustworthy foundation-model deployment across forecasting, decision support, and grid-edge control.

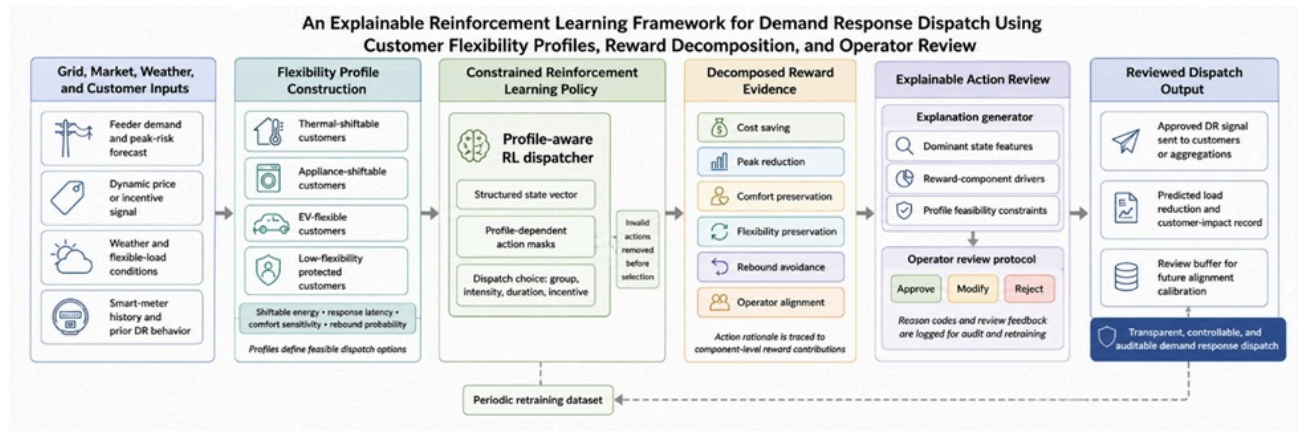


Figure 3. Research Agenda for Trustworthy Foundation-Model Deployment in Energy Systems

Conclusion

This systematic review shows that foundation models are entering energy systems through three main pathways: forecasting, grid-edge intelligence, and smart-grid decision support. The strongest evidence is currently found in transformer-based forecasting and LLM-enabled engineering workflows, while closed-loop grid-edge control remains comparatively immature.

The review also shows that foundation models are not a single technology class but a family of architectures with different strengths. Pre-trained transformers are best suited to structured temporal prediction, LLMs are strongest for language-

rich decision support, and multi-modal models are emerging for renewable forecasting and complex operational contexts.

The systematic map clarifies that the field is promising but not yet operationally settled. Many studies demonstrate feasibility, but fewer address real-time safety, privacy, adversarial robustness, benchmark standardization, or human factors in high-stakes grid environments.

A coordinated research agenda is therefore needed to move from demonstration to trustworthy deployment. Progress will depend on shared benchmarks, physics-aware adaptation, auditable decision-support systems, and governance frameworks that recognize the critical-infrastructure role of energy systems.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 13 Oct 2025

Revised: 20 Nov 2025

Accepted: 31 Dec 2025

Published online: 20 March 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

- Majumder S, Dong L, Doudi F, Cai Y, Tian C, Kalathil D, et al. Exploring the capabilities and limitations of large language models in the electric energy sector. *Joule*. 2024;8(6):1544-9.
- Cheng Y, Zhao H, Zhou X, Zhao J, Cao Y, Yang C, et al. A large language model for advanced power dispatch. *Sci Rep*. 2025;15(1):8925.
- Zhang L, Chen Z, Ford V. Advancing building energy modeling with large language models: Exploration and case studies. *Energy Build*. 2024;323:114788.

- L'Heureux A, Grolinger K, Capretz MA. Transformer-based model for electrical load forecasting. *Energies*. 2022;15(14):4993.
- Wu T, Ling Q. STELLM: Spatio-temporal enhanced pre-trained large language model for wind speed forecasting. *Appl Energy*. 2024;375:124034.
- Shi H, Fang L, Chen X, Gu C, Ma K, Zhang X, et al. Review of the opportunities and challenges to accelerate mass-scale application of smart grids with large-language models. *IET Smart Grid*. 2024;7(6):737-59.
- Madani S, Tavasoli A, Astaneh ZK, Pineau PO. Large language models integration in smart grids. *Energy Rep*. 2025;14:1562-77.
- 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.
- Gao J, Chen Y, Hu W, Zhang D. An adaptive deep-learning load forecasting framework by integrating transformer and domain knowledge. *Adv Appl Energy*. 2023;10:100142.
- Gabber HA, Hemied OS. Domain-specific large language model for renewable energy and hydrogen deployment strategies. *Energies*. 2024;17(23):6063.
- Jiang G, Ma Z, Zhang L, Chen J. EPlus-LLM: A large language model-based computing platform for automated building energy modeling. *Appl Energy*. 2024;367:123431.
- Spencer R, Ranathunga S, Boulic M, van Heerden AH, Susnjak T. Transfer learning on transformers for building energy consumption forecasting—A comparative study. *Energy Build*. 2025;336:115632.
- Liu G, Bai Y, Zhao H, Wen K, Wang X, Gu J, et al. MVLFLM: A parameter-efficient large language model framework for cross-domain multi-voltage load forecasting in smart grids. *Energy AI*. 2025;22:100651.
- Huy PC, Minh NQ, Tien ND, Anh TT. Short-term electricity load forecasting based on temporal fusion transformer model. *Ieee Access*. 2022;10:106296-304.
- Amjad F, Korotko T, Rosin A. Review of llms applications in electrical power and energy systems. *IEEE Access*. 2025;13:150951-69.
- Fan Y, Song T, Feng C, Liu C, Jiang D. Wind power prediction using foundation large time series models enhanced by time series prompt in exogenous and tuning forms. *Appl Energy*. 2025;400:126535.
- Jiang G, Ma Z, Zhang L, Chen J. Prompt engineering to inform large language model in automated building energy modeling. *Energy*. 2025;316:134548.
- Zhu G, Jia W, Xing Z, Xiang L, Hu A, Hao R. CMLLM: A novel cross-modal large language model for wind power forecasting. *Energy Convers Manag*. 2025;330:119673.
- 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.
- Ferreira AB, Leite JB, Salvadeo DH. Power substation load forecasting using interpretable transformer-based temporal fusion neural networks. *Electr Power Syst Res*. 2025;238:111169.
- Zhang J, Zhang C, Lu J, Zhao Y. Domain-specific large language models for fault diagnosis of heating, ventilation, and air conditioning systems by labeled-data-supervised fine-tuning. *Appl Energy*. 2025;377:124378.
- Jiang G, Chen J. Efficient fine-tuning of large language models for automated building energy modeling in complex cases. *Autom Constr*. 2025;175:106223.

Saeed F, Rehman A, Shah HA, Diyan M, Chen J, Kang JM. SmartFormer: Graph-based transformer model for energy load forecasting. *Sustain Energy Technol Assess*. 2025;73:104133.

Banad YM, Sharif SS, Rezaei Z. Artificial intelligence and machine learning for smart grids: From foundational paradigms to emerging technologies with digital twin and large language model-driven intelligence. *Energy Convers Manag X*. 2025;28:101329.

Li J, Li D, Yang Y, Xi H, Yu W, Xiao Y, et al. Zero-shot load forecasting for integrated energy systems: A large language model-based framework with multi-task learning. *Neurocomputing*. 2025;654:131288.
<https://doi.org/10.1016/j.neucom.2025.131288>.

Li P, Hu Z, Shen Y, Cheng X, Alhazmi M. Short-term electricity load forecasting based on large language models and weighted external factor optimization. *Sustain Energy Technol Assess*. 2025;82:104449.

Dong Q, Huang R, Cui C, Towe D, Zhou L, Tian J, et al. Short-Term Electricity-Load Forecasting by deep learning: A comprehensive survey. *Eng Appl Artif Intell*. 2025;154:110980.

Erdiwansyah, Mamat R, Syafrizal, Ghazali MF, Basrawi F, Rosdi SM. Emerging role of generative AI in renewable energy forecasting and system optimization. *Sustain Chem Clim Action*. 2025;7:100099.

Chan JW, Yeo CK. A transformer based approach to electricity load forecasting. *Electr J*. 2024;37(2):107370.

Hinov N. The energy hunger of AI: Large language models as challenges and enablers for sustainable energy. *Energies*. 2025;18(17):4701.