



## Review

# A Four-Dimensional Analysis of Explainable AI in Energy Forecasting: A Domain-Specific Systematic Review

Vahid Arabzadeh \*  and Raphael Frank 

Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, 1511 Luxembourg, Luxembourg; raphael.frank@uni.lu

\* Correspondence: vahid.arabzadeh@uni.lu

## Abstract

Despite the growing use of Explainable Artificial Intelligence (XAI) in energy time-series forecasting, a systematic evaluation of explanation quality remains limited. This systematic review analyzes 50 peer-reviewed studies (2020–2025) applying XAI to load, price, or renewable generation forecasting. Using a PRISMA-inspired protocol, we introduce a dual-axis taxonomy and a four-factor framework covering global transparency, local fidelity, user relevance, and operational viability to structure our qualitative synthesis. Our analysis reveals that XAI application is not uniform but follows three distinct, domain-specific paradigms: a user-centric approach in load forecasting, a risk management approach in price forecasting, and a physics-informed approach in generation forecasting. Post hoc methods, particularly SHAP, dominate the literature (62% of studies), while rigorous testing of explanation robustness and the reporting of computational overhead (23% of studies) remain critical gaps. We identify key research directions, including the need for standardized robustness testing and human-centered design, and provide actionable guidelines for practitioners.

**Keywords:** explainable artificial intelligence; energy time-series forecasting; model interpretability; load/price/generation prediction; domain-specific XAI



Academic Editor: Luca Longo

Received: 2 October 2025

Revised: 15 November 2025

Accepted: 23 November 2025

Published: 25 November 2025

**Citation:** Arabzadeh, V.; Frank, R. A Four-Dimensional Analysis of Explainable AI in Energy Forecasting: A Domain-Specific Systematic Review. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 153. <https://doi.org/10.3390/make7040153>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

As the global energy sector accelerates its transition toward decarbonization, decentralization, and automation, Artificial Intelligence (AI) and Machine Learning (ML) have become indispensable tools for managing increasingly complex, data-intensive operations. In this context, accurate Electric load forecasting is critical for power system planning, energy trading, and demand–response optimization [1], directly influencing grid stability, operational efficiency, and cost management [2]. Similarly, reliable forecasting of electricity prices and renewable generation, particularly solar and wind, is essential for market participation, storage scheduling, and balancing intermittent supply [3,4].

However, the increasing reliance on high-performance AI models, especially deep learning architectures, introduces significant challenges due to their “black-box” nature [5,6]. While these models achieve high predictive accuracy, their lack of transparency undermines stakeholder trust, complicates error diagnosis, and limits operational acceptance, particularly in safety-critical domains such as grid control and real-time markets [7,8]. Unexplained forecast errors or sudden model drift can lead to poor decision-making, financial losses, or even system instability [1,6]. This tension between performance and interpretability has

elevated Explainable Artificial Intelligence (XAI) from a technical add-on to a foundational requirement for trustworthy AI deployment in energy systems [9,10].

The growing necessity for XAI reflects a broader co-evolution of energy infrastructure and computational intelligence. As shown in Table 1, the shift from centralized AC grids to distributed, net-zero hybrid systems have paralleled advances in AI, from early perceptrons to modern deep learning and transformer models. With each leap in modelling capability, the gap between predictive power and human comprehensibility has widened. While early systems relied on rule-based or linear models with inherent transparency, today's deep architectures demand post hoc or hybrid explanation methods such as SHAP, LIME, and surrogate modelling to restore interpretability. This trajectory underscores that explainability is no longer optional but a necessary condition for auditability, regulatory compliance, and human oversight in modern energy forecasting.

**Table 1.** A historical evolution of power and energy systems in parallel with key advances in artificial intelligence [6].

Era	Energy Systems Milestone	AI/ML Development	XAI/Interpretability Progress
1880s	Invention of AC transmission (Tesla)	–	–
1958	First commercial solar cell (Bell Labs)	Perceptron	–
1988	Grid-connected renewables begin scaling	Temporal-Difference RL	Rule-based expert systems (inherent transparency)
1997	European RES integration begins	LSTM	Sensitivity analysis in neural networks
2012	Rise of smart grids, digital meters	Deep Learning: AlexNet	Early visualization techniques
2015	Tesla Powerwall, home battery revolution	ML in load/price forecasting (e.g., XGBoost)	Interpretable tree ensembles (e.g., SHAP precursors)
2020s	Hybrid AI-energy systems, net-zero push	Transformers, DRL, ensemble DL	Widespread adoption of SHAP, LIME, surrogate models

Yet, despite growing interest, XAI applications in energy remain fragmented. Many studies focus narrowly on specific domains, such as smart buildings [5,11] or maintenance systems [2], without addressing the cross-cutting needs of time-series forecasting. Others provide methodological surveys but lack focus on operational integration or user-centered design [6]. While some reviews offer a comprehensive review of XAI in load forecasting, their scope excludes electricity price and renewable generation forecasting, domains with distinct uncertainty profiles and stakeholder requirements [1]. As a result, there is no unified assessment of how explanation quality, particularly in terms of trust and fidelity, varies across forecasting tasks and user roles.

Trust in XAI refers to the confidence that domain experts and decision-makers place in AI-generated forecasts and their explanations [8,12]. It is shaped by transparency, consistency, and the ability to detect and understand model limitations [13,14]. In energy systems, trust is further influenced by the operational stakes: a grid operator must trust short-term load forecasts to avoid overloading circuits, while a market participant relies on price forecasts to optimize bidding strategies [3,7]. However, trust without accuracy can be dangerous. This is where fidelity, the degree to which an explanation faithfully reflects the underlying model's behavior, becomes essential [15]. The accurate descriptions of inaccurate predictions can never be relevant [15]. High-fidelity explanations ensure that insights are not just intuitive but technically valid, enabling effective debugging, feature refinement, and risk assessment [16,17].

To highlight the novelty and scope of our work, Table 2 compares this review with existing XAI-related surveys in energy and time-series forecasting. It demonstrates that, while prior reviews offer valuable insights, they are often limited in domain coverage, lack a trust-centric focus, or fail to align explanations with user roles. In contrast, this review uniquely integrates forecasting specificity, explanation fidelity analysis, and domain-specific trust constructs across load, price, and generation forecasting, with explicit alignment to operational constraints and user needs.

**Table 2.** Comparison of this review with existing XAI-related surveys in energy and time-series forecasting.

Ref. (Year)	Domain	Forecasting Focus	Time-Series Specific	Trust-Centric	Domain-Specific Trust	User-Aligned	Synthesis Depth
[1]	Power load forecasting	✓	✓	✗	✗	✗	Moderate
[6]	General energy systems	Partial	✗	✗	✗	✗	Moderate
[5]	Building energy	✓	✓	✗	✗	✓ (partial)	High
[11]	Smart buildings	✗	✗	✗	✗	✓	High
[18]	Urban emissions	✗	✗	✗	✗	✗	Low
[2]	Energy maintenance	✗	✗	✓ (partial)	✗	✓	High
This paper *	Load, price, generation	✓	✓	✓	✓	✓	High

\*: This work uniquely integrates forecasting specificity, explanation fidelity analysis, and domain-specific trust constructs across load, price, and generation forecasting, with explicit alignment to user roles and operational constraints.

#### *The Observed Gaps, Work Contribution, and Research Questions*

Our contributions are threefold. First, we present a PRISMA-guided synthesis of 50 peer-reviewed studies on XAI in energy time-series forecasting, introducing a dual-axis taxonomy based on temporal integration and forecasting domain. Second, we conduct a comprehensive cross-domain analysis of trust, identifying three distinct, domain-specific paradigms: a user-centric approach in load forecasting, a risk management approach in price forecasting, and a physics-informed approach in generation forecasting. Third, using a novel four-factor analytical framework (global transparency, local fidelity, user relevance, and operational viability), we conduct a systematic qualitative synthesis of the literature. This analysis reveals the current state of the art and identifies critical, cross-domain gaps, particularly in the validation of explanation robustness and the reporting of operational costs. The full dataset and analysis code are publicly released to support reproducibility. Despite recent progress, several critical gaps remain in the literature on XAI for energy forecasting:

- No systematic cross-domain analysis of explanation quality: Existing reviews focus on isolated domains (e.g., load or price forecasting), but none systematically compare how “trust” and “fidelity” are conceptualized and evaluated across load, price, and renewable generation forecasting.
- Lack of domain-specific trust constructs: While trust is often mentioned, prior work rarely distinguishes between context-dependent interpretations such as “reliability under volatility” in markets, “physical plausibility” in generation, or user alignment in demand-side management.
- Insufficient alignment with user roles: Explanations are frequently evaluated in technical terms without considering the needs of real-world stakeholders (e.g., grid operators, market analysts, building managers), limiting operational relevance.
- Limited assessment of robustness and feasibility: There is minimal reporting on the computational cost, stability, or practical deploy ability of XAI methods in real-time or safety-critical energy systems.

To address these gaps, this review is guided by the following research questions:

RQ1: How do studies in XAI-enhanced energy forecasting define, invoke, or operationalize the concept of trust across load, price, and renewable generation domains?

RQ2: What explanation methods are used, and how well are they aligned with the needs of specific user roles (e.g., grid operators, market analysts, building managers)?

RQ3: What evidence exists regarding the robustness, computational cost, and operational feasibility of XAI methods in real-world energy forecasting pipelines?

## 2. Core Concepts and Taxonomy of Explainable AI

To address the inconsistent use of XAI terminology in energy forecasting literature, this section establishes a clear conceptual framework. We distinguish between interpretability and explainability, define trust as a socio-technical construct, and introduce a multi-axis taxonomy to classify XAI methods in time-series contexts. This foundation enables a systematic analysis of explanation quality across domains, grounded in the seminal work of [19,20]. We then define trust not as a model output, but as a socio-technical outcome shaped by explanation quality. Building on this, we introduce a multi-axis taxonomy to classify XAI methods, tailored to time-series dynamics. Finally, we contextualize this taxonomy with real-world energy forecasting models, setting the stage for our empirical analysis in Section 5.

### 2.1. Interpretability vs. Explainability

A foundational challenge in XAI research is the conflation of interpretability and explainability. We adopt the following distinctions:

- **Interpretability:** The extent to which a model's internal logic can be directly understood and audited by domain experts, as in inherently transparent models such as linear regression (e.g., modelling demand trends) or decision trees (e.g., encoding peak-load rules) [1,2,5,18].
- **Explainability:** The ability to generate post hoc explanations for complex models (e.g., DNNs, ensembles) using external techniques such as SHAP or LIME, which quantify the contribution of input features to individual or aggregate predictions. These methods are essential for diagnosing and validating black-box forecasts in energy markets and power systems [1,2,5,6,11,18].

This distinction is critical, as the choice between an interpretable model (which may sacrifice accuracy) and a powerful black-box model (e.g., which requires post hoc explanation) embodies the accuracy explainability trade-off a central theme of our analysis. The nature of this trade-off, however, is not static; it depends heavily on the application domain (e.g., load, price, or generation forecasting) and the temporal horizon of the prediction.

### 2.2. The Role of Trust in XAI for Energy Forecasting

Trust is not a direct model output, but a socio-technical outcome shaped by the quality of explanations. Following this, we define trust in this context as the user's confidence that a forecasting model is reliable, safe, and accountable. This confidence is fostered by explanations that are [19]:

- **Faithful:** Accurately reflect the model's internal logic.
- **Consistent:** Exhibit stable behavior across time and perturbations.
- **Actionable:** Enable users to make informed operational decisions.

This framing positions trust as an emergent property of explanation quality not just technical correctness, but usability and reliability [2,6,11].

### 2.3. A Multi-Axis Taxonomy of XAI Methods

With the goals of interpretability and trust clearly defined, we now turn to how these goals are pursued in practice. XAI methods can be systematically classified along three complementary dimensions, forming a taxonomy that enables precise comparison and evaluation:

Integration Stages are:

- Ante-hoc (inherently interpretable): Models designed with transparent structure (e.g., linear regression, decision trees) [21].
- Post hoc (external explanation): Techniques applied to complex ‘black box’ models after a prediction is made. This includes external methods (e.g., SHAP, LIME) and the analysis of internal mechanisms like attention weights [22].

Local vs. Global:

- Local: Explains a single prediction (e.g., why the model predicted high demand for a specific hour) [19].
- Global: Describes the overall behavior of a model across the entire dataset (e.g., the general effect of temperature on price forecasts) [19].

Model Dependency:

- Model-agnostic: Can be applied to any model (e.g., SHAP, LIME, PDPs), providing flexibility across forecasting tasks [5,19,21].
- Model-specific: Tailored to a specific architecture. For example, TreeSHAP is optimized for tree-based ensembles, while attention mechanisms and saliency maps are used for deep neural networks like LSTMs and Transformers [5,19].

As example, while SHAP is model-agnostic in principle, optimized variants like TreeSHAP are model-specific.

### 2.4. Time-Series Context and Understandability

In energy forecasting, XAI must account for sequential dependencies, seasonal cycles, and dynamic behaviors. Input data often includes weather variables, calendar events, and economic indicators, all with varying influence across different forecast horizons. Standard static explanation methods can be less effective, demanding techniques that are both time-aware and sensitive to these unique domain dynamics. Effective explanations must also be “understandable” that is, comprehensible to domain experts and “transparent”, with model logic directly accessible to support auditing and decision-making [19,21].

### 2.5. Contextualizing XAI in Energy Systems

To illustrate the relationship between forecasting model types, their explainability, and the most appropriate XAI methods, Table 3 summarizes common models used in energy systems and the strategies typically applied.

**Table 3.** Overview of common machine learning models categorized by their predictive performance, inherent explainability, and associated XAI method characteristics [1,19,21].

Model Type	Predictive Power	XAI Strategy & Temporal Scope	Why It Is Useful
Linear Regression	Low (baseline performance)	Ante-hoc; Global. Coefficients reveal fixed temporal feature influence (e.g., lagged variables).	Simple baseline; helps assess lagged influence and seasonal shifts directly.
Decision Tree	Medium	Ante-hoc; Local/Global. Splits reflect temporal thresholds; tree depth captures recursive logic.	Rule-based clarity helps explain threshold behaviors and diurnal patterns in forecasts.
K-NN (with windowing)	Medium	Post hoc; Local. LIME or local surrogates explain nearest historical patterns used for prediction.	Supports analog-based forecasting; useful in short-term demand anomalies.

Table 3. Cont.

Model Type	Predictive Power	XAI Strategy & Temporal Scope	Why It Is Useful
Random Forest	High	Post hoc; Primarily Local. TreeSHAP <sup>1</sup> for lag feature importance per prediction.	Captures non-linear interactions and helps explain recurring temporal cycles.
CNN (e.g., for 1D TS)	High	Post hoc; Local. Grad-CAM-style <sup>2</sup> saliency maps on input time window highlight key subsequences.	Shows which time chunks influence decisions good for interpretable filters on sensor sequences.
RNN/LSTM (w/attention)	Very High (sequence-aware)	Post hoc; Local. Attention weights, feature occlusion, or LIME used to explain time step influence.	Ideal for continuous forecasting; attention reveals influential lagged periods.
Transformer	Very High (multi-step, attention)	Post hoc; Local/Global. Self-attention reveals intra-sequence dependencies; SHAP for input features.	Powerful for recursive or hybrid forecasting; aligns well with explainable temporal dynamics.

<sup>1</sup> SHAP and LIME are model-agnostic by design, but SHAP has optimized versions (e.g., TreeSHAP) that are model-specific. <sup>2</sup> Deep models often rely on architecture-dependent explanations (e.g., Grad-CAM), though some general tools like LIME may still apply.

### 3. Review Methodology

We conducted a PRISMA-inspired systematic review to identify peer-reviewed studies on Explainable Artificial Intelligence (XAI) in energy time-series forecasting, published between January 2020 and June 2025 [23]. Our protocol was adapted from established systematic review guidelines to ensure transparency, reproducibility, and methodological rigor. The process consisted of three phases, as detailed below and visualized in Figure 1 (methodological workflow) and Figure 2 (PRISMA inspired flow diagram).

The protocol for this systematic review was retrospectively registered with the Open Science Framework (OSF) on 3 October 2025, and is publicly available at [<https://osf.io/fa3vu/>].

- Phase 1 (Literature Identification): Structured research was conducted in three major databases: Scopus, ScienceDirect, and IEEE Xplore. A Boolean search string combined keywords related to energy forecasting (e.g., “load forecasting”, “price forecasting”, “renewable generation forecasting”) and explainable AI (e.g., “XAI”, “SHAP”, “LIME”, “interpretability”, “post-hoc explanation”). The search was limited to studies published between January 2020 and June 2025 to focus on recent methodological advances. The final search was conducted on 15 May 2025, yielding an initial pool of 1264 records. To supplement the database search, a secondary query was performed in Google Scholar. Since this search engine does not support complex Boolean strings, we used targeted keywords combinations derived from our main search (e.g., “interpretable load forecasting”) to identify additional relevant studies.
- Phase 2 (Screening and Eligibility): After removing duplicates using reference management software, the sole author screened all titles, abstracts, and full texts against predefined inclusion and exclusion criteria.

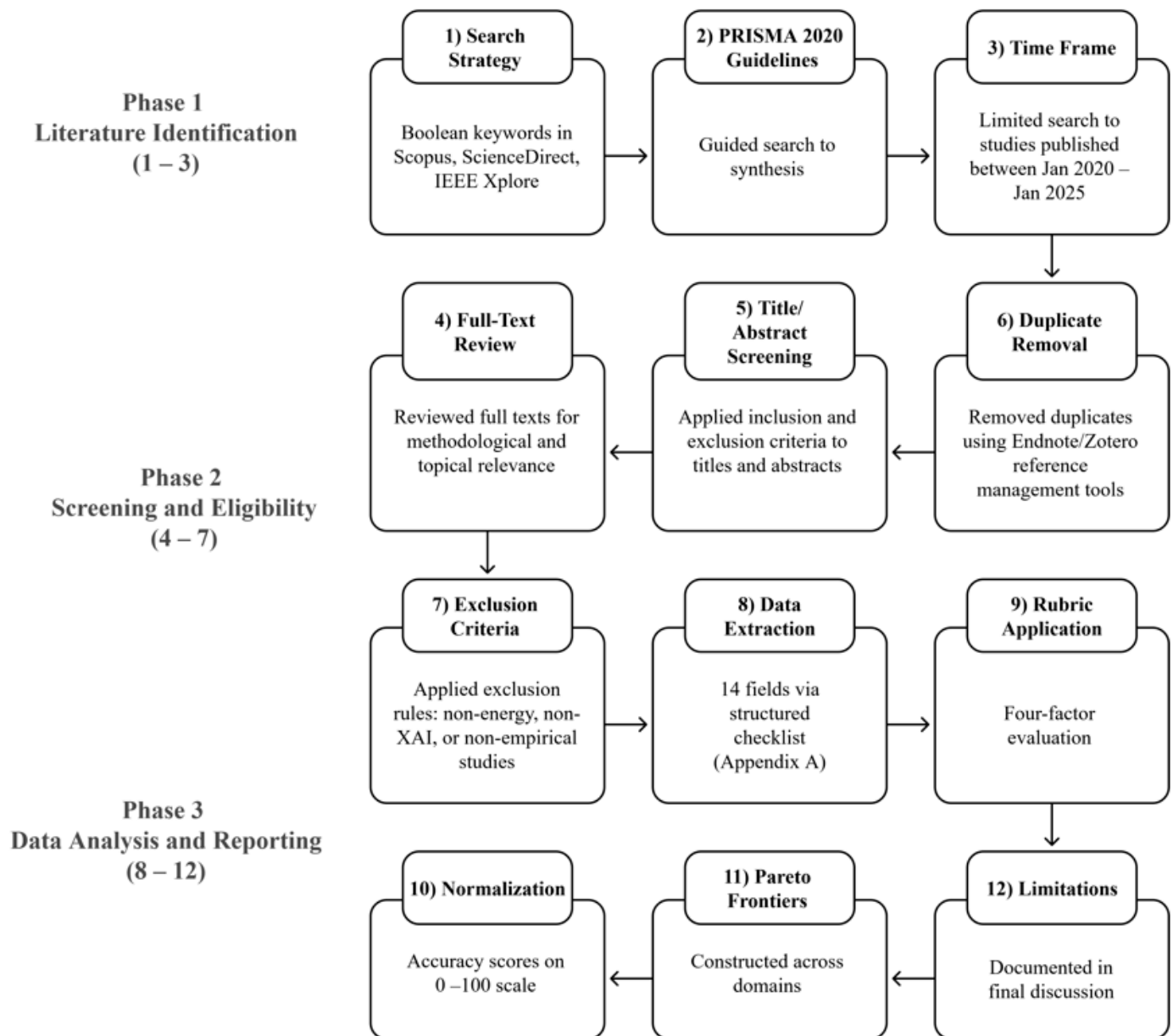
Inclusion criteria:

1. Peer-reviewed journal or conference papers
2. Published in English with accessible full text
3. Focused on XAI techniques applied to energy time-series forecasting

Exclusion criteria:

1. Non-forecasting tasks (e.g., fault detection, clustering)
2. No explicit use or evaluation of explainability
3. Editorials, book chapters, or non-peer-reviewed sources



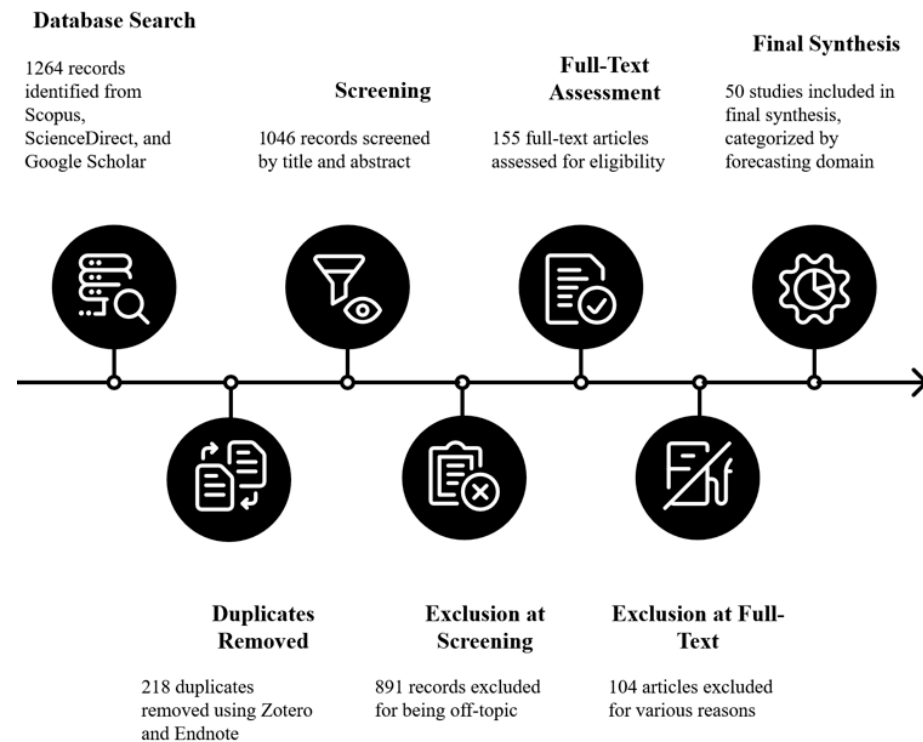


**Figure 1.** Overview of the systematic review methodology for XAI in energy forecasting. Bibliometric and trend analysis (Phase 3) and explanation quality assessment (Phase 3) are integrated within the ‘Data Analysis and Reporting’ phase.

As the screening process was conducted by a single reviewer, a quality control step was implemented to mitigate bias and ensure high intra-rater reliability (i.e., consistency in the reviewer’s own judgments). Specifically, a random 20% sample of all studies excluded during the initial title and abstract screening was re-evaluated by the same reviewer after a two-week interval. The near-perfect agreement between the two evaluations confirmed that the exclusion criteria were applied consistently throughout the process.

- **Phase 3 (Data Analysis and Reporting):** After the multi-stage screening process detailed in Phase 2 reduced the initial 1264 records to a final set of 50 eligible studies (see Figure 2 for a full breakdown), we proceeded with data extraction and analysis. We then applied our four-factor evaluation rubric (Section 4.2) to assess explanation quality across global transparency, local fidelity, user relevance, and operational viability. To minimize subjectivity, a detailed coding guide was developed a priori, defining clear criteria for low, medium, and high scores on each dimension (see Appendix A,

Table A2). For example, for User Relevance, a ‘High’ score was given only if a study explicitly named a stakeholder (e.g., ‘grid operator’) and linked the explanation to a specific decision task. A ‘Low’ score was given if only technical feature importance was presented. Similarly, for Operational Viability, a ‘High’ score required explicit reporting of computational overhead, while a ‘Low’ score was given for no mention of cost. All scoring decisions were documented to support transparency and future replication.



**Figure 2.** PRISMA 2020-inspired flow diagram illustrating the study selection process for XAI in energy forecasting. A total of 1264 records were identified through database searches, with 50 studies included in the final synthesis after duplicate removal, title/abstract screening, and full-text assessment [23].

This structured, multi-layered methodology spanning search, screening, extraction, bibliometric mapping, and quality scoring provides a rigorous foundation for our thematic analysis of XAI in energy-forecasting research.

### 3.1. Search Strategy and Time Frame

We searched Scopus and ScienceDirect using structured Boolean queries to identify English-language, peer-reviewed journal and conference papers. Our core search string (adapted to each database’s syntax) was:

“explainable AI” OR “interpretable learning” OR “transparency” OR “interpretability” OR “explainability” OR “intelligibility” OR “understandability” OR “comprehensibility” OR “glass-box” OR “white-box” OR “inherently interpretable”

AND

“energy forecasting” OR “load prediction” OR “time series prediction” OR “renewable energy” OR “energy pricing” OR “electricity forecasting” OR “power forecasting” OR “demand forecasting” OR “generation forecasting” OR “solar forecasting” OR “wind forecasting” OR “heating demand” OR “carbon trading” OR “building energy”

To focus on recent methodological advances, we limited results to studies published between January 2020 and June 2025. This search strategy, guided by the PRISMA 2020



statement [23], ensured broad coverage of peer-reviewed literature while maintaining specificity to energy time-series forecasting applications.

### 3.2. PRISMA Inspired Framework and Study Selection Procedure

We followed a PRISMA-inspired protocol to ensure a transparent and reproducible screening process (see Figure 2 for the PRISMA flow diagram). After retrieving records, we merged exports and removed duplicate DOIs and titles using reference management software. The sole author then screened all titles, abstracts, and full texts against the pre-defined inclusion and exclusion criteria. To ensure consistency and mitigate the potential for single-reviewer bias, a random 20% sample of the initially excluded studies was re-evaluated after a two-week interval to verify the consistent application of the criteria.

Included if:

- Peer-reviewed journal or conference paper.
- Published between January 2020 and June 2025.
- Focused on XAI techniques applied to energy forecasting.
- Written in English with accessible full text.

Excluded if:

- Addressed only non-forecasting tasks (e.g., fault detection).
- Did not apply or mention any form of explainability.
- Were book chapters, editorials, or other non-peer-reviewed sources.

### 3.3. Data Extraction Process

After confirming full-text eligibility, we systematically extracted 14 data fields from each of the 50 included studies using a predefined checklist (see Appendix A, Table A1). The extracted data included: bibliographic metadata, forecasting task details (domain, horizon, granularity), data characteristics (sources, size, preprocessing), model architecture (e.g., LSTM, XGBoost), and the specific XAI method employed. We also recorded performance metrics, practical deployment context (e.g., computational cost, real-time feasibility), and any human-centered discussions, such as references to user roles or decision-making tasks.

### 3.4. A Four-Dimension for Explanation Framework

To systematically analyze the 50 studies in our corpus, we developed a four-factor analytical framework. While informed by multi-dimensional XAI evaluation frameworks, our framework is designed to qualitatively assess XAI applications across four critical, operationally relevant dimensions:

- **Global Transparency:** This dimension evaluates the extent to which the model's overall behavior, learned strategies, and underlying logic are made comprehensible to a human analyst.
- **Local Fidelity & Robustness:** This dimension evaluates not just the claimed accuracy of local explanations (fidelity), but also their stability and consistency when tested under perturbation, distribution shift, or overtime.
- **User Relevance:** This dimension assesses whether the explanation is explicitly tailored to a specific user role (e.g., grid operator, market analyst) and supports a concrete decision-making task.
- **Operational Viability:** This dimension evaluates the practical feasibility of deploying the XAI system in a real-world setting. It considers crucial aspects like the computational overhead of generating explanations and the explicit justification for selecting a particular XAI method.

Collectively, these four dimensions, Global Transparency, Local Fidelity & Robustness, User Relevance, and Operational Viability, constitute our operational definition and evaluation criteria for explanation quality in this review. A high-quality explanation is one that scores well across these dimensions, being not just technically faithful but also comprehensible, robust, relevant to a specific user, and practical to deploy.

### 3.5. Rationale for Selected Dimensions

The selection of these four dimensions was deliberate, designed to create a holistic evaluation that progresses from theoretical soundness to practical deployment. Global Transparency serves as the foundation, assessing the model's fundamental logic. Local Fidelity & Robustness is critical because, for high-stakes energy applications, individual explanations must be reliable. User Relevance then bridges the gap to practical utility by evaluating whether a technically sound explanation is understandable and actionable for a specific stakeholder. Finally, Operational Viability provides the ultimate reality check, addressing the computational and integration constraints that determine an XAI method's real-world feasibility.

This four-factor framework provides the consistent structure for the thematic synthesis presented in Section 4, allowing us to compare the different approaches to explainability across the load, price, and generation domains.

## 4. Results

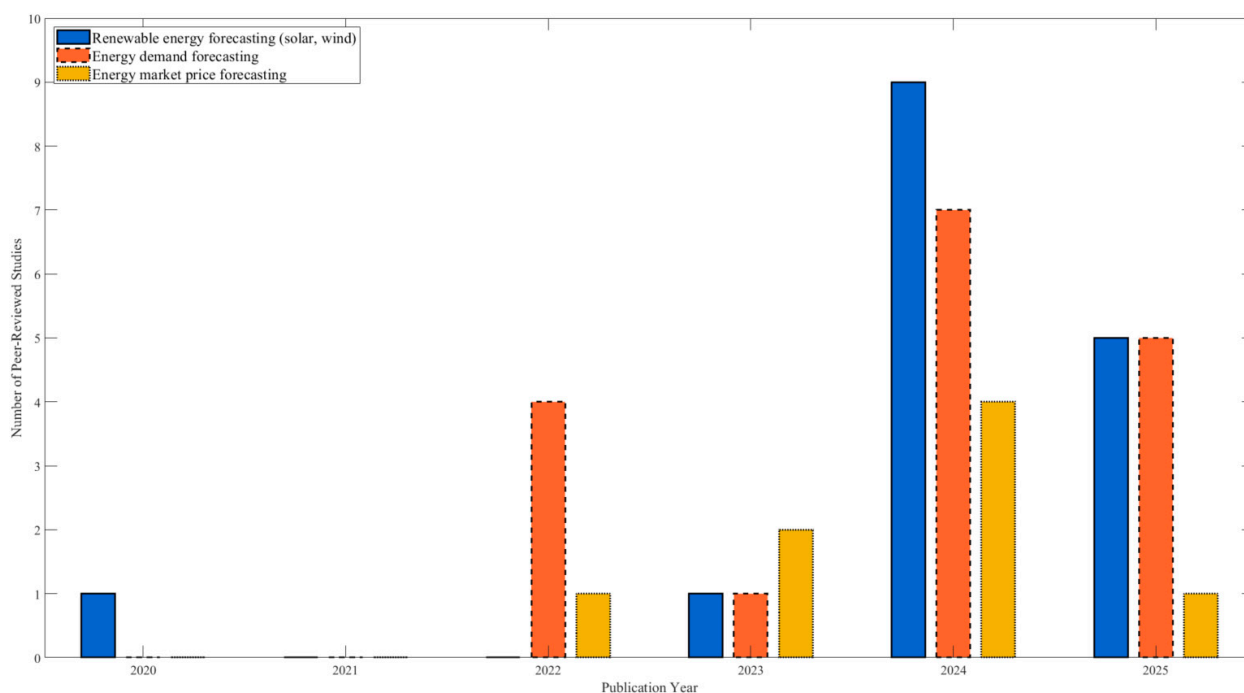
This section presents the results of our systematic review. It begins with a descriptive overview of the publication landscape, followed by a detailed thematic analysis of the studies within each of the three forecasting domains: price, load, and generation.

The initial database search yielded 1264 records. After removing 218 duplicates, 1046 records were screened by title and abstract, from which 891 were excluded. The full texts of the remaining 155 articles were assessed for eligibility, and 104 were excluded for not meeting the inclusion criteria. A final sample of 50 studies was included in the qualitative synthesis (see Figure 2 for the PRISMA flow diagram).

### 4.1. Publication Landscape Analysis

An analysis of the publication landscape between 2020 and 2025 (June) reveals a clear upward trend and confirms the novelty of the field. As illustrated in Figure 3, the number of studies increased steadily, with a sharp peak in 2024 where power generation and load forecasting research registered its highest activity with 9 and 7 publications, respectively. This surge aligns with the growing demand for XAI in energy-critical infrastructure. While lower in volume, energy price forecasting studies show noticeable growth, reflecting rising interest in market transparency. The synchronized rise, with significant activity only beginning around 2022, reinforces that explainable energy forecasting is a recent and rapidly maturing research area.

The publication of these studies in top-tier journals, as shown in Table 4, confirms the field's scientific relevance and timeliness. Journals such as *Applied Energy* (IF 11), *Energy Conversion and Management* (IF 10.9), and the rapid emergence of *Energy and AI* (IF 9.6) as a core outlet underscore a maturing research frontier. The adoption by journals with stringent standards signals that explainable forecasting is transitioning from an experimental topic to an essential component for trustworthy energy systems in both academic and industrial contexts.



**Figure 3.** Annual distribution of peer-reviewed studies on XAI in energy time-series forecasting across three domains (2020–2025) [24].

**Table 4.** Top five journals by number of forecasting publications between 2020 and 2025 [24].

Journal	Number of Publications	Impact Factor (Latest)
Applied Energy	10	11 (27 June 2025)
Energy	5	9.4 (27 June 2025)
Electric Power Systems Research	4	4.2 (27 June 2025)
Energy and AI	4	9.6 (27 June 2025)
Energy Conversion and Management	3	10.9 (27 June 2025)

Finally, the country-wise distribution in Table 5 highlights distinct research patterns. China dominates the generation domain (10 publications), reflecting its focus on renewable energy. Germany leads in load forecasting (4 studies), underscoring its emphasis on demand-side management, while India concentrates on price forecasting (3 studies). This geographical breakdown reveals that while generation forecasting is a central global focus, the load and price domains present clear opportunities for exploration, particularly in underrepresented countries.

**Table 5.** Country-wise distribution of reviewed forecasting studies, highlighting the top five contributing nations across generation, load, and energy price domains.

Country	Load	Generation	Price
China	1	10	1
Germany	4	1	1
India	0	1	3
South Korea	2	1	1
Spain	1	2	2

## 4.2. Load Forecasting

The literature in Load forecasting domain shows a strong emphasis on developing models that are not only highly accurate but also interpretable, allowing end-users to understand and trust the forecasts for critical decision-making [25,26].

### 4.2.1. Forecasting Performance

The literature showcases a clear progression towards sophisticated model architecture. Deep learning models are prevalent. Recurrent models like LSTM and GRU, which process data sequentially, are highly effective for many time-series tasks but can struggle to capture long-range dependencies [27,28]. In contrast, Transformer-based architectures, such as the Temporal Fusion Transformer (TFT), use self-attention mechanisms to model relationships across all time steps simultaneously. This allows them to superiorly model long-range dependencies and integrate diverse inputs for multi-horizon forecasting, though often at a higher computational cost [29,30].

Hybrid models, which combine convolutional layers for feature extraction with recurrent layers for temporal modeling (e.g., CNN-LSTM), are also widely used [12,31]. These architectures leverage a synergistic ‘best-of-both-worlds’ approach. For example, in a CNN-LSTM model, the CNN layers first act as efficient feature extractors, identifying salient local patterns or subsequences within the time-series (e.g., the specific shape of a morning load ramp-up). The output of these layers is a compressed feature representation then fed into the LSTM layers, which are adept at modeling the temporal relationships between these identified patterns [28,30].

Alongside these complex architectures, there is a strong trend toward interpretable-by-design models. These include Neural Additive Models (NAMs), Hierarchical NAMs (HNAMs), and the symbolic QLattice, which are designed to produce transparent outputs without a significant loss in accuracy [15,26].

Model performance hinges on rich, context-aware feature engineering. While historical load data remains the core input, modern systems integrate multi-modal data, including:

- Meteorological variables: Outdoor temperature, humidity, and derived indices like the Temperature-Humidity Index (THI) are critical for modeling thermal demand [25,32].
- Temporal features: Hour, day-of-week, and holiday indicators are encoded to capture cyclical patterns [14,29].
- System-specific data: Building characteristics (e.g., insulation, age), grid topology (e.g., customer count), or appliance-level consumption are used to enhance domain relevance [12,26].
- Latent representations: Autoencoders are used to compress high-dimensional sensor data into compact feature sets [33].

The reviewed models are applied across multiple scales, from individual households to national grids, with a primary focus on short-term forecasting. These models achieve high accuracy, with reported MAPE values as low as 0.44% for residential load and 0.94% for day-ahead substation load [29,34].

Notably, a recurring finding is that well-tuned classical or ensemble models often rival their deep learning counterparts. This suggests that thoughtful input design, such as optimizing the length of the input history, can be more critical to performance than architectural complexity alone [26,27]. The diversity of the approaches discussed is further detailed in Table 6, which offers a comparative summary of the methodologies employed across the load forecasting literature.

**Table 6.** Comparative Overview of Models, Input Features, and Applications in the Load Forecasting Literature.

Study	Application	Forecasting Model(s)	Key Input Features	Forecasting Horizon
[27]	Building load and temperature	Attention-GRU	Weather, setpoints, historical load/temp	24 h ahead
[14]	National electric load (Germany)	Sequence-to-Sequence RNN (GRU)	Weather, time (dummies), historical load	Day-ahead (24 h)
[35]	Smart home energy consumption	TKAT, LSTM, CNN-LSTM, CNN-GRU	Weather, historical appliance usage	Short-term
[36]	National energy consumption (UK)	MLR, SVM, GPR, LSBoost	Weather, socioeconomic, energy prices	Next month (monthly)
[26]	Annual building energy performance	QLattice, XGB, ANN, SVR, MLR	Building characteristics, insulation, heating	Annual
[25]	District heating demand (Warsaw)	Artificial Neural Network (ANN)	Weather, time (cosine), historical demand	Up to 120 h ahead
[10]	District heating & cooling load	Reservoir Computing (RC), LSTM, CNN, etc.	Weather, historical energy load, time (sine)	6 h ahead (15 min intervals)
[28]	Household active and reactive power forecasting	LSTM-Attention with SHAP	Historical load/power, lag values, weather, explicit numerical time encoding	1 h ahead, 1-day ahead
[37]	Household and distribution level electricity demand forecasting	Explainable Causal Graph Neural Network (X-CGNN)	Historical demand, weather variables (temp, humidity, wind)	30, 60, 90 min; 1, 2, 3 h
[30]	Short-term electricity load forecasting	CrossInformer with ICEEMDAN-RLMD decomposition and SHAP	Historical load, temperature, electricity price, humidity	Short-term (24 steps ahead)
[34]	Multi-step residential load forecasting (probabilistic)	Interpretable Multi-Variable LSTM (MV-LSTM) with mixture attention	Historical load, time-related variables (hour, day, holiday), weather	Next day (48 half-hour steps)
[13]	Justifying short-term load forecasts from neural models	RNN with LSTM layers, explained using LIME	Historical load, weather from multiple cities, holiday/school indicators	72 h ahead
[12]	Explainable household energy demand forecasting	ForecastExplainer (LSTM explained by DeepLIFT approximating SHAP)	Aggregate and appliance-level consumption, seasonality features	Hourly, daily, weekly
[38]	Interpretable short-term load forecasting for central grid	Interpretable model with multi-scale temporal decomposition (Transformers & CNNs)	Historical load, auxiliary features (temperature, calendar, humidity)	Short-term
[15]	Interpretable retail demand forecasting	Hierarchical Neural Additive Models (HNAMs)	Sales data, promotions, price, holidays, product/store attributes	Daily, up to 2 weeks
[31]	Building Air Conditioning Energy Forecasting	CNN-LSTM	Previous energy use, water temperatures, system flow, weather	1 h ahead (Short-term)
[29]	Power Substation Load Forecasting	Temporal Fusion Transformer (TFT)	Historical load, temporal data (hour, day, holiday), temperature	24 & 48 h ahead (Short-term)
[33]	Warehouse Power Consumption Prediction	VAE + Tree-based Regressors (ExtraTree, LightGBM)	VAE-generated latent features from all building sensors	10-min, 20-min, 1 h ahead (Short-term)
[39]	Feeder Load Forecasting	PWP-XGBoost	Temperature, humidity, customer count, KVA, temporal data	1-year ahead (Medium-term)
[32]	Educational Building Load Forecasting	Tree-based Ensembles (LightGBM, RF, XGBoost, etc.)	Historical load, weather (THI, WCT), temporal data	Day-ahead hourly (Short-term)

#### 4.2.2. Explainable AI Methodologies in Load Forecasting

In load forecasting, applying XAI is critical for moving beyond simple accuracy metrics to create trustworthy tools for stakeholders ranging from utility planners to individual consumers.

The maturity of these XAI applications can be assessed across four interconnected dimensions, from high-level model validation to practical deployment (also see Table 7).

**Table 7.** Comparative Analysis of XAI Applications in Load Forecasting, Detailing the Methods, User Focus, and Visualization Techniques.

Study	Applied Explainability Method	Global or Local	User Relevance	Method of Visualizing Explainability
[27]	Attention Mechanism	Both	Model developers and researchers for understanding model behavior	2D color map of attention weights over time
[14]	Saliency Maps (Perturbation)	Local	Domain experts (e.g., energy traders) for causal understanding	2D color map (saliency map) of feature importance over time
[35]	SHAP	Both	Model developers and energy managers for feature impact analysis	SHAP summary plot (beeswarm plot) and feature importance bars
[36]	SHAP & Feature Selection	Both	Policymakers and energy analysts for identifying key drivers	SHAP summary plots and bar charts
[26]	QLattice (inherent), Permutation Importance	Global	Experts and non-experts (homeowners) for transparent calculation	Simple mathematical formula, variable importance bar plots
[25]	SHAP (DeepSHAP)	Both (focus on Global)	Experts and model developers for model validation	Feature importance bars, dependency plots, scatter plots of SHAP values
[10]	LIME	Local	Power plant/network operators for operational decision-making	Heatmap of feature importance over the forecast horizon
[28]	SHAP (post hoc)	Both	Power Managers: Designing time-of-use tariffs and optimizing energy storage sizing.	Feature importance bar charts, SHAP summary (beeswarm) plots, dependency plots.
[37]	Intrinsic Causal Graph; Post hoc Feature Ablation	Both	Model Developers: Optimizing input sequence length to create lighter, more efficient models for real-time use.	Causal graphs (network diagrams), heatmaps of feature/time-lag importance.
[30]	SHAP (post hoc)	Global	System Analysts: Quantifying feature contributions to validate modeling assumptions and understand drivers of load.	SHAP beeswarm plot.
[34]	Mixture Attention Mechanism (intrinsic)	Both	Utility Analysts: Understanding relationships between load, weather, and time to analyze demand changes.	Bar chart for global variable importance, heatmaps for temporal importance.
[13]	LIME (post hoc)	Local	Forecasters/Practitioners: Justifying individual forecasts to build trust and improve neural models.	LIME explanation plot (horizontal bar chart).
[12]	ForecastExplainer (DeepLIFT approximating SHAP)	Primarily Local (can be aggregated)	Smart Home Users: Understanding appliance-level consumption to become more aware and optimize energy use.	Area plots, box plots, and histograms showing feature contributions over time.
[38]	Linear combination of specialized NNs (intrinsic)	Global	Model Developers/Analysts: Understanding the significance of different temporal patterns (trend, seasonality) and features for model tuning.	Heatmaps of learned significance scores for features and decomposed components.
[15]	Hierarchical Neural Additive Models (HNAMs) (intrinsic)	Both	Retail Managers/Forecasters: Aligning the model with mental models to reduce algorithm aversion and improve judgmental adjustments.	Violin plots, time series plots of aggregated effects, and instance-level forecast decomposition plots.



Table 7. Cont.

Study	Applied Explainability Method	Global or Local	User Relevance	Method of Visualizing Explainability
[31]	SHAP, LIME, Grad-CAM, Grad-Absolute-CAM	Both	Building professionals seeking model trust and practical utility.	Feature importance bar charts, feature activation heatmaps.
[29]	Attention Mechanism, Variable Importance Networks (inherent in TFT)	Both	Grid operators make decisions on fault analysis and resource planning.	Attention weight plots over time, feature importance bar charts for encoder/decoder.
[32]	SHAP (TreeSHAP)	Both	EMS managers persuading customers to act on peak alerts.	SHAP summary plot, Partial Dependence Plot (PDP), heatmap plot.
[27]	SHAP (TreeSHAP)	Both (Focus on improving global from local)	Not explicitly stated but geared towards data scientists improving feature analysis.	SHAP Partial Dependence Plot (PDP).
[33]	SHAP, Tree-based Feature Importance	Global	Warehouse managers and operators may lack technical knowledge.	SHAP summary and bar plots.

- **Global Transparency:** Global Transparency in load forecasting reveals the model's overall strategy, such as how it weighs weather patterns against historical consumption cycles. This is achieved through interpretable-by-design architectures like Temporal Fusion Transformers (TFTs) and Hierarchical Neural Additive Models (HNAMs) that embed transparency directly into their structure [15,29]. For “black-box” models, global insights are derived by aggregating post hoc explanations from methods like SHAP to validate that the model learns expected relationships, such as outdoor temperature driving air conditioning load [25,35]. While these global insights are crucial for strategic planning, their value is limited without the ability to trust and diagnose forecasts for critical peak demand hours.
- **Local Fidelity & Robustness:** This need to trust individual predictions leads to the second dimension, Local Fidelity and Robustness, which validates the explanations for specific forecasts. To verify fidelity, novel metrics like the Contribution Monotonicity Coefficient (CMC) have been developed to evaluate explainers like DeepLIFT [12]. Robustness is also explicitly tested; for instance, HNAMs were subjected to robustness checks confirming their explanations remained consistent across different forecast horizons in over 99% of instances [15].
- **User Relevance:** User Relevance bridges the gap between a technically valid explanation and an actionable insight for a specific stakeholder. For grid operators, SHAP explanations are used to identify the drivers of peak consumption hours, supporting the design of time-of-use tariffs [28]. For homeowners, the ForecastExplainer framework visualizes which specific appliances will contribute most to future energy use, empowering them to manage demand [12].
- **Operational Viability:** This brings us to the final dimension, Operational Viability, which addresses the practical challenges of deploying these XAI systems. Intrinsic (ante-hoc) models like HNAMs are highly efficient, as the cost of explanation is integrated into the model's inference time, making them suitable for real-time dispatch systems [15]. In contrast, post hoc methods like SHAP add computational overhead that may be too slow for generating explanations at the sub-hourly frequency required for operational control [12]. The literature shows that modern interpretable models can achieve accuracy competitive with their black-box counterparts, demonstrating that transparency does not have to come at a major cost to performance [15].

The practical application of XAI in load forecasting relies on effective visualization, with clear patterns connecting the type of graphic to both the underlying explanation method and the intended user. Post hoc, feature attribution methods like SHAP and LIME are consistently visualized using feature importance bar charts, summary (beeswarm) plots, and dependency plots to provide global and local insights [13,25,32,35]. In contrast, methods that analyze temporal importance, such as attention mechanisms and saliency mapping, typically employ heatmaps or 2D color maps to show which historical time steps are most influential [14,27]. Inherently interpretable models produce more unique outputs, ranging from causal network diagrams and time series plots of aggregated effects to a simple mathematical formula [15,26,37].

#### 4.3. Price Forecasting

Unlike load forecasting, which is often driven by predictable weather and seasonal cycles, price forecasting must contend with high volatility, sharp spikes, and complex, event-driven market dynamics. This fundamental difference is reflected in the literature, which emphasizes robust ensemble models and a diverse range of market-driven input features.

##### 4.3.1. Forecasting Performance

To manage the non-stationary and volatile nature of energy prices, a predominant trend is the application of hybrid and ensemble models. These are typically weighted approaches (please also see Table 8). Decomposition-based approaches like the D3Net model first separate the price series into simpler components [40]. Other methods, like stacked ensemble architectures [9], train a second-level ‘meta-model’ (e.g., linear regression) that learns the optimal weights to combine base-model predictions. In contrast, online ensembles like ENSWNN [3] are also weighted but dynamically optimize the contribution of ‘neighbors’ (historical patterns) in real-time for each new prediction. Deep learning architectures such as LSTM and GRU are also widely evaluated for capturing these complex patterns [41].

Reflecting the market-driven nature of price signals, the required input features extend far beyond the core meteorological data often sufficient for load forecasting. While some real-time models rely solely on univariate historical price data [3], most frameworks incorporate a wide array of external variables:

- **Exogenous and Market Data:** This includes prices from interconnected markets, power generation data, demand forecasts, locational (zonal) prices, and ancillary market prices [7,9].
- **Macroeconomic and Financial Data:** In commodity and stock forecasting, inputs include macroeconomic indicators like the CBOE VIX (fear index) and futures prices with different maturities [41,42].
- **Unstructured Data:** To capture public sentiment and emerging trends, some models successfully incorporate unstructured data, such as keyword search volumes from Baidu [43].

The primary applications involve forecasting for various energy markets at horizons crucial for financial decision-making, including day-ahead, intraday, and real-time [3,9,44]. Performance is paramount, with successful ensemble frameworks achieving low error metrics like a mean absolute scaled error (MASE) of 0.378 in volatile day-ahead markets [9].

**Table 8.** Comparative Overview of Models, Input Features, and Applications in the Price Forecasting Literature.

Manuscript (Author, Year)	Application	Forecasting Horizon	Models	Inputs and Features
[43]	Carbon trading price	Multi-step (2, 4, 6 steps)	Proposed: CEEMDAN-WT-SVR. Benchmarks: ELM, SVR, LightGBM, XGBoost.	Structured (historical prices, market indices) and Unstructured (Baidu search index keywords).
[3]	Spanish electricity price	Real-time, multi-step (h = 1, 3, 24 h)	Proposed: ENSWNN. Benchmarks: LSTM, CNN, TCN, RF, XGBoost, River library models.	Univariate (historical electricity prices).
[42]	US clean energy stock indices	Short- to medium-term (daily data)	Proposed: Facebook's Prophet, NeuralProphet. Benchmarks: TBATS, ARFIMA, SARIMA.	Macroeconomic variables (DJIA, CBOEVIX, oil price) and 21 technical indicators.
[7]	Electricity price (Italian & ERCOT markets)	Short-term (hourly)	Analyzed: CNN, LSTM. Proposed: Trust Algorithm for explaining models.	Zonal prices, ancillary market prices, demand, neighboring market prices, historical prices.
[40]	Half-hourly electricity price (Australia)	Half-hourly	Proposed: D3Net (STL-VMD with MLP, RFR, TabNet). Benchmarks: 14 standalone and STL-based models.	Univariate (lagged series of electricity prices identified by PACF).
[41]	Crude oil spot prices	Multi-horizon (multi months ahead)	Compared: MLP, RNN, LSTM, GRU, CNN, TCN.	Futures prices with different maturities (1, 2, 3, 6, 12 months).
[44]	Intraday electricity price difference	Intraday (15 min intervals)	Proposed: Normalizing Flows. Benchmarks: Gaussian copula, Gaussian regression.	Previous price differences, day-ahead price increments, renewable/load forecast errors.
[9]	Day-ahead electricity price (Spain)	Day-ahead (24 h)	Proposed: Stacked ensemble of 15+ ML models. Benchmarks: AutoML platforms (H <sub>2</sub> O, TPOT).	Extensive exogenous data (generation, demand, market prices, fuel prices), STL components, time series features.

#### 4.3.2. Explainable AI Methodologies in Price and Market Forecasting

In contrast to load and generation forecasting, where XAI often validates physical plausibility, its application in price forecasting is driven by the high-stakes financial environment. The focus shifts from physical intuition to managing risk, validating model behavior in volatile markets, and providing traders with actionable signals.

- **Global Transparency:** Global Transparency is used by market analysts to ensure models learn rational economic relationships rather than spurious correlations. SHAP summary plots are commonly used to identify the most influential market drivers, such as demand forecasts or the price of natural gas [42,44]. More advanced frameworks combine SHAP with Morris's sensitivity screening and surrogate decision tree models to provide a multi-faceted view of a complex ensemble's logic [9].
- **Local Fidelity & Robustness:** Given that a single erroneous forecast can lead to significant financial loss, Local Fidelity and Robustness is paramount. While LIME and instance-specific SHAP values are used to explain individual predictions [42], a stand-out innovation is the "Trust Algorithm." This tool generates a numerical trust score for each forecast by correlating its local SHAP explanation with the model's global behavior. This directly measures the robustness of individual predictions and acts

as a real-time risk assessment tool, proving highly effective at identifying unreliable forecasts during market regime shifts [7].

- **User Relevance:** User Relevance in this domain is uniquely focused on the needs of financial actors like traders, asset managers, and policymakers. A clear trend, different from the often-technical explanations in load forecasting, is the move toward simplified, actionable outputs. For example, score-based explanations like the trust score are proposed as more direct and user-friendly tools than complex plots, as they can be integrated directly into bidding software to trigger risk-averse strategies [7]. The goal is to foster human–machine collaboration where XAI provides a “data story” that augments a trader’s domain knowledge [9].

**Operational Viability:** The high-frequency nature of algorithmic trading imposes strict Operational Viability constraints, a key difference from the often slower-paced requirements of grid management. While some approaches cleverly integrate XAI into the offline development phase, such as using SHAP for feature selection to add no real-time cost [44], models for live trading must be highly efficient. For instance, the execution times of the real-time ENSWNN model are explicitly measured to ensure they meet low-latency constraints suitable for streaming applications [3].

The visualization of XAI in price forecasting is strongly linked to user relevance, with a clear trend towards simplifying outputs for rapid decision-making. While technical users like model developers are provided with detailed SHAP summary plots, 3D Partial Dependence Plots, and decision tree diagrams for deep analysis [9,42], the output for end-users is often more direct. The Trust Algorithm, for example, visualizes its output not as a complex plot but as a simple numerical score, providing an immediate and unambiguous signal of forecast reliability [7]. Similarly, inherently local models like k-NN based ensembles explain forecasts by plotting the historical price patterns (neighbors) used for the prediction—an intuitive approach for traders who rely on historical analogies [3]. To provide a structured overview of these approaches, Table 9 synthesizes the XAI methodologies and their user-centric applications in the price forecasting literature.

**Table 9.** Summary of XAI Methodologies in Reviewed Price Forecasting Studies.

Manuscript (Author, Year)	Applied Explainability Method	Global or Local	User Relevance (Target Audience & Goal)	Method of Visualization
[7]	Trust Algorithm (combining SHAP, PFI, and LR scores)	Both	High: EPF users and bidding agents, for deciding when to trust a prediction.	Score-based (numerical trust score); waterfall/force plots.
[44]	SHAP values	Global	Moderate: Model developers, for input feature selection.	Bar charts of average absolute SHAP values.
[3]	Inherent model explainability (k-NN based)	Local	High: End-users, for understanding the historical basis of a forecast.	Line plots showing the query sequence and its nearest neighbors.
[41]	Extensive hyperparameter tuning analysis	Global	Moderate: AI researchers, for understanding model architecture behavior.	Primarily through analysis of performance tables.
[9]	SHAP, Morris screening, ICE/PD plots, surrogate decision tree	Both	High: Market participants, for human-in-the-loop decision-making.	SHAP plots, ICE/PD curves, decision tree diagrams.
[43]	Random permutation with Monte Carlo simulation, prediction intervals	Both	High: Analysts, for understanding feature impact and prediction uncertainty.	Line plots showing prediction intervals under uncertainty.

Table 9. Cont.

Manuscript (Author, Year)	Applied Explainability Method	Global or Local	User Relevance (Target Audience & Goal)	Method of Visualization
[42]	SHAP, LIME, Partial Dependence Plots (PDP)	Both	High: Policymakers and traders, for strategic decision-making.	SHAP summary plots, LIME explanation plots, 3D PDPs.
[40]	SHAP, LIME	Both	High: Energy experts and grid operators, for decision-making.	SHAP summary/bar plots, LIME bar plots.

#### 4.4. Renewable Generation Forecasting

Distinct from load forecasting's focus on behavioral patterns and price forecasting's market dynamics, renewable generation forecasting is fundamentally a physical problem. The core challenge is taming the intermittency and non-stationarity of weather-driven resources like wind and solar. This leads to a strong emphasis on models that can capture complex physical relationships and spatio-temporal correlations, a key difference from the other domains.

##### 4.4.1. Forecasting Performance

The models applied reflect this physical challenge. A notable trend is the use of interpretable models like the Explainable Boosting Machine (EBM) and Concept Bottleneck Models (CBMs), which are valued not just for transparency but for their ability to learn physically plausible relationships for applications like extreme wind speed prediction [45,46]. For multi-site solar forecasting, Spatio-Temporal Graph Neural Networks (STGNNs) are employed to model geographic dependencies [47]. To handle deep non-linearities, various hybrid systems are proposed, including decomposition-based frameworks like VMD-DCESN [48] and representation learning models like ISTR-TFT [49].

Unsurprisingly, the input features for generation forecasting are overwhelmingly dominated by meteorological and physical data, a stark contrast to the economic and behavioral inputs often used in price and load forecasting.

- **Core Inputs:** Standard inputs include historical generation data and Numerical Weather Prediction (NWP) variables like wind speed at different heights, temperature, and cloud cover [8,46].
- **Advanced Data Sources:** More advanced approaches incorporate satellite and sky-imagery data (e.g., total column ozone, cloud properties) for solar forecasting [40], spatio-temporal data from neighboring farms for wind [50], and turbine-specific variables like air density [51].
- **Engineered Features:** Innovative feature engineering is common, such as using mixed-frequency data or transforming time series into interpretable features inspired by Open-High-Low-Close (OHLC) charts [45,52].

Applications in this domain are highly specialized and often tied to the physical nature of the assets. Beyond short-term forecasting for grid stability, specific applications include wind turbine condition monitoring [51] and Extreme Wind Speed (EWS) prediction for operational safety [45]. High accuracy is consistently reported across various horizons, from 1 min ahead to multi-step 72 h forecasts. For example, tree-based models like Random Forest have achieved  $R^2$  values of 0.999 for solar generation [53]. To provide a structured overview of these varied approaches, Table 10 synthesizes the reviewed literature on renewable generation forecasting, detailing the specific models, applications, and input features used.



**Table 10.** Comparative Analysis of Models, Inputs, and Applications in the Renewable Generation Forecasting Literature.

Study	Model(s)	Input Features	Forecasting Horizon
[54]	XGBoost	Irradiance, Temp, Humidity, Wind Speed/Direction, Pressure, Time Index	15 min intervals
[8]	LightGBM, RF, GB, XGBoost, DT	SSRD, Temp, Humidity, Cloud Cover, Wind, Pressure, Precipitation, etc.	Hourly
[47]	GCN1D (STGNN)	Multi-site Power, Wind Speed, Temp, Month, Hour, Plant Location	Multi-step (24–72 h)
[55]	XGBoost, Linear Regression, RF, etc.	DC/AC Power, Daily/Total Yield, Irradiance, Ambient/Module Temp	Not specified
[53]	Random Forest Regressor (best of 4)	DC Power, Irradiance, Ambient/Module Temp, Daily/Total Yield	Not specified
[56]	TabNet (DL)	Satellite data (Ozone, Aerosols), Sky Images (Clouds), SZA, Lagged UV-B	Multi-step hourly (1–4 h)
[57]	Direct Explainable Neural Network (DXNN)	Solar Irradiance, Temp, Humidity, Wind, Sun Altitude/Azimuth, etc.	1 min ahead
[16]	3P-CBiLSTM: A three-phase hybrid model combining CNN and BiLSTM, optimized with TMGWO and BOHB.	Historical hourly wind speed and multiple ground-level and satellite-based meteorological variables.	1 h ahead
[58]	Sequential GRU & Self-Attention: A deep learning framework with a final XGBoost regressor layer.	Experimental data: Temperature, RSS Particle Size, HDPE Particle Size, and % of Plastics in Mixture. Data was augmented.	N/A (Regression Task)
[52]	Ensemble System: An ensemble of mixed-frequency (AR-MIDAS) and machine learning models with an interpretable base model selection strategy (elastic net + SHAP).	Mixed-frequency (15 min and 1-h) wind speed and wind power data, preprocessed with ICEEMDAN.	1 h ahead
[45]	Concept Bottleneck Model (CBM): An interpretable model with automatically generated concepts from decision trees and a logistic regressor output layer.	Historical hourly wind speed, transformed into 14 interpretable OHLC-inspired features.	1 h ahead
[50]	Hybrid (Contrastive Learning + Ridge Regression)	Historical wind power from multiple farms (spatio-temporal); learned trend & fluctuation representations.	Ultra-short-term
[51]	Evaluation Framework (for ANN, RF, SVR, etc.)	Wind speed, air density, turbulence intensity.	Point-in-time (Power Curve)
[46]	Interpretable (Glass-Box GAM with trees)	NWP data (wind speed/direction at 10 m & 100 m) and/or historical wind power.	Short-term (hours to days)
[48]	Hybrid (Decomposition + ESN)	Historical wind power decomposed by VMD into sub-series.	Multi-step
[49]	Hybrid (Representation Learning + Transformer)	Historical wind power, NWP data, static variables; learned seasonal & trend representations.	Multi-horizon (short-term)
[4]	Interpretable Deep Learning (Transformer)	Multivariate historical data (wind power, speed, direction).	Single-step
[59]	Interpretable (LASSO) & Evaluation Framework	Historical wind speed and power from multiple farms (spatio-temporal).	Ultra-short-term
[17]	XAI Trustworthiness Evaluation Framework	NWP data.	Short-term

#### 4.4.2. Explainable AI Methodologies in Renewable Generation Forecasting

Distinct from load forecasting's focus on behavioral patterns and price forecasting's risk management, the application of XAI in renewable generation is driven by the need to validate models against physical reality. The primary goal is to ensure that complex models have learned meteorologically sound relationships, a critical step for building trust with engineers and system operators. This focus on physical validation and building trust with technical experts is evident across the four key dimensions of explainability:



- **Global Transparency:** This is used to verify that a model's overall strategy aligns with domain knowledge. A key trend is the use of inherently interpretable "glass-box" models like the Explainable Boosting Machine (EBM), which learns explicit "shape functions" for each physical input, and Concept Bottleneck Models (CBMs), which base predictions on human-understandable concepts like "high turbulence" [45,46]. A standout innovation is the development of frameworks that validate a black-box model by quantifying its alignment with known physics, resulting in a "physical reasonableness" score [51].
- **Local Fidelity & Robustness:** This dimension is crucial for trusting individual forecasts, especially under the extreme or unusual weather conditions that affect renewable generation. While post hoc methods like LIME are used for local explanations [55], their trustworthiness is a significant concern. This has led to research that quantitatively measures the fidelity of XAI techniques like SHAP and LIME against a ground truth derived from perturbation analysis [46]. Furthermore, models that are interpretable by design, like the Direct Explainable Neural Network (DXNN), offer perfect local fidelity by providing a clear mathematical input-output mapping for every prediction [52].
- **User Relevance:** This is tailored to the needs of technical experts who manage physical assets. Unlike the financial traders in price forecasting or the diverse consumers in load forecasting, the users are often engineers, wind/solar farm operators, and grid planners. Explanations provide actionable insights for tasks such as root-cause analysis of wind turbine underperformance [51], predicting extreme wind events for operational safety [45], and providing public health information on UV radiation based on solar forecasts [56].
- **Operational Viability:** This dimension considers the practical feasibility of deploying XAI systems. Inherently interpretable models like EBMs are highly viable as they are computationally efficient and do not require a separate, costly explanation step [46]. While post hoc methods like SHAP can be computationally intensive [56], their use is justified for deep, offline analysis. The development of automated pipelines for complex validation, such as checking for physical reasonableness, also enhances the operational viability of advanced XAI frameworks [51].

The visualization of XAI in generation forecasting is tailored to provide deep physical insights for expert users. Unlike the simplified scores for traders or appliance-level charts for consumers, the visualizations here aim to decode complex physical relationships. Inherently interpretable models produce unique outputs, such as 1D and 2D shape function plots from an EBM showing how the model responds to wind speed, or histograms of activated concepts from a CBM revealing the "rules" used to predict an extreme event [45,46]. More advanced frameworks use specialized conditional attribution plots to visualize a model's global strategy against a physical baseline [51]. While standard SHAP summary plots and bar charts are also used [8], the trend is toward visualizations that offer a deeper, physically grounded understanding of the model's behavior. Table 11 synthesizes the diverse XAI methodologies, their target users, and visualization techniques in the renewable generation literature.

**Table 11.** Summary of XAI Methodologies in Reviewed Generation Forecasting Studies.

Study	Applied Explainability Method	Global or Local Scope	User Relevance & Application	Method of Visualizing Explainability
[46]	Explainable Boosting Machine (WindEBM)	Both	High: Direct decision support, feature engineering, and root-cause analysis.	Bar charts (feature importance), 1D and 2D plots (shape functions).
[49]	ISTR-TFT	Both	Understanding feature importance, disentangling seasonal-trend patterns, visualizing time dependencies.	Bar charts (variable importance), t-SNE plots (representations), line plots (attention weights).
[51]	Physics-Informed XAI using Shapley Values	Both	Model validation, robustness assessment, understanding out-of-distribution performance, root-cause analysis.	Conditional attribution plots (global strategy), bar plots (local deviations).
[50]	ICOTF with Optimal Transport	Both	Understanding interactions and the contribution of specific historical time steps.	Heat maps of the transportation matrix.
[59]	LASSO Coefficients and Derived Indices	Global	Extracting domain knowledge about spatio-temporal correlations and feature importance.	Heat maps of correlation coefficients and causal discrimination matrices.
[17]	Post hoc XAI Evaluation (SHAP, LIME, PFI, PDP)	Both	Assessing the trustworthiness and reliability of different XAI techniques themselves.	Bar charts (feature importance), violin and box plots (distribution of trust metrics).
[48]	DCESN Temporal Dependence Coefficient	Global	Understanding the internal temporal dynamics and influence of different model parts on future steps.	Heat maps of dependence coefficients.
[16]	LIME (Local Interpretable Model-Agnostic Explanations) SHAP (Shapley Additive Explanations)	Both: Global (LIME for overall model); Local (LIME/SHAP for individual predictions)	High: Explanations designed for wind farm operators to help make quality decisions regarding grid integration.	Local: Bar plots from LIME showing feature contributions. Global: SHAP visualizations including feature importance plots, beeswarm summary plots and dependence plots.
[58]	SHAP and LIME Partial Dependence Plots (PDPs) Modified William's Plot	Both: Global (SHAP & PDPs provide overall insights); Local (LIME & SHAP explain individual predictions)	High: Targeted at energy engineers to help them understand factors influencing hydrogen production for process optimization.	Local: LIME and SHAP explainer plots. Global: Partial Dependence Plots and a modified William's Plot for overall analysis.
[52]	Novel strategy combining Elastic Net algorithm with SHAP	Global	High: Used to interpretably select base models forming an ensemble, explaining model's overall structure and logic; increases trust among grid operators for energy planning.	SHAP features importance and summary plots to visualize contribution of each base model (not individual data features).

Table 11. Cont.

Study	Applied Explainability Method	Global or Local Scope	User Relevance & Application	Method of Visualizing Explainability
[45]	Concept Bottleneck Model (CBM) intrinsically interpretable model	Both: Global (transparent architecture); Local (each prediction explained by activated concepts)	Very High: Explicitly designed for “human-in-the-loop” paradigm, with concepts representing clear rules useful for wind farm managers predicting extreme wind events.	Histograms of activated concepts. Bar charts showing accuracy of each “decision” (combination of concepts). Diagrams illustrating logical rules of individual concepts.
[54]	ELIS	Both	Understanding feature importance for decision-makers.	Tables of feature weights and contributions.
[8]	SHAP, LIME, PFI	Both (SHAP/PFI-Global, LIME/SHAP-Local)	Support grid operators and energy managers in decision-making.	Feature importance bar plots, SHAP beeswarm plots.
[47]	GNNEexplainer	Local (aggregated for global insights)	Inform energy network management and PV system design.	Heatmaps of feature/edge masks, plots of aggregated importance.
[55]	LIME	Local	Provide stakeholders with transparent tools for resource management.	Not specified in text (describes process).
[53]	LIME, PDP	Both (LIME-Local, PDP-Global)	Optimize solar generation systems by identifying key feature impacts.	PDP plots, bar charts of positive/negative feature contributions.
[56]	LIME, SHAP	Both (SHAP-Global, LIME-Local)	Aid UV experts in providing public health recommendations.	LIME bar plots, SHAP beeswarm plots and dependence plots.
[57]	Inherently Explainable Model (DXNN)	Global (via mathematical formula)	Understand the direct input-output relationship of the model.	quadratic explainable function” or “interpretable quadratic neuron.
[60]	LIME, SHAP, ELIS	Both (LIME/SHAP-ELIS-Global)	Enable adoption of AI in smart grids by increasing transparency.	Bar charts for feature importance and contributions, SHAP summary plots.

## 5. Discussion

Our systematic review of 50 peer-reviewed studies reveals that Explainable AI (XAI) in energy forecasting is a rapidly maturing field, yet its practices are characterized by a persistent disconnect from real-world operational needs. The application and quality of XAI are not uniform; instead, they are shaped by the specific challenges of each forecasting domain. This discussion synthesizes these findings by directly addressing our three research questions and highlighting the distinct paradigms of XAI application that have emerged in load, price, and generation forecasting.

### 5.1. Three Paradigms of Explainable Energy Forecasting

To provide a clear, comparative overview and robust evidence for our central conclusion, Table 12 synthesizes the core findings from Section 4. It illustrates the distinct challenges, data, models, and XAI goals that define the three emergent paradigms.

The driving force behind each paradigm is the unique core challenge of the forecasting task. In price forecasting, the primary challenge is managing extreme volatility and financial risk. This economic driver necessitates the use of robust hybrid and ensemble models tailored to handle non-stationary signals. Consequently, the inputs are heavily market-driven—including interconnected market prices, commodity futures, and even unstructured data—and the application of XAI is focused on financial risk management.

Here, “trust” is defined as the reliability of a forecast to prevent monetary loss, exemplified by innovations like the “Trust Algorithm” which provides actionable signals for traders.

**Table 12.** Synthesis of the Three Domain-Specific XAI Paradigms.

Characteristic	Load Forecasting (User-Centric Paradigm)	Price Forecasting (Risk-Centric Paradigm)	Generation Forecasting (Physics-Centric Paradigm)
Core Challenge	Capturing cyclical and behavioral patterns across diverse scales (e.g., residential, grid).	Managing extreme volatility, sharp price spikes, and high-stakes financial risk.	Taming intermittency and validating models against complex physical/meteorological processes.
Dominant Models	Sequential (LSTM, GRU), Transformers (TFT), and Hierarchical models (HNAMs).	Robust Ensembles (Stacked, Online), Hybrids (e.g., D3Net), and Decomposition-based models.	Interpretable-by-design (EBM, CBM), Spatio-temporal GNNs, and Physics-Informed models.
Key Input Data	Meteorological data, Calendar/Time features (holiday, day-of-week), and Building/System data.	Market data (interconnects, demand forecasts), Economic indicators (futures, VIX), and Unstructured data.	NWP data, Satellite/Sky-imagery, and specific Physical variables (e.g., turbine air density, cloud properties).
Primary XAI Goal	Provide actionable insights tailored to diverse users (e.g., grid operators, homeowners).	Perform risk management, validate economic logic, and provide actionable trading signals.	Ensure physical plausibility, perform root-cause analysis, and align model logic with science.
Trust’ Defined As	User Alignment: An explanation is relevant and useful for a specific human-in-the-loop task.	Financial Reliability: An explanation assesses real-time risk and prevents monetary loss (e.g., “Trust Algorithm”).	Physical Plausibility: An explanation confirms the model adheres to known physical principles (e.g., “reasonableness” score).
Key Representative Studies	[12,15,28,29]	[7,9,40,42]	[45–47,51]

In contrast, generation forecasting is fundamentally a physical modeling problem. The core challenge is taming the intermittency of weather-dependent resources, a process governed by physics and meteorology. This leads to a strong emphasis on interpretable-by-design models (e.g., EBMs, CBMs) and spatio-temporal architectures (STGNNs) that can validate physical relationships. The inputs are overwhelmingly physical, drawing from Numerical Weather Prediction (NWP) data, satellite imagery, and turbine-specific variables. Accordingly, the goal of XAI is to ensure physical plausibility, where “trust” means the model’s logic aligns with known scientific principles, as seen in frameworks that score a model’s “physical reasonableness” for engineers.

Finally, load forecasting is driven by the challenge of capturing strong cyclical and behavioral patterns in energy consumption across diverse scales. This domain relies heavily on powerful sequential deep learning models like LSTMs and Transformers that excel at modeling regular temporal patterns. The inputs are a hybrid of behavioral and physical data, combining core meteorological variables with strong calendar features and system-specific data like building characteristics. The application of XAI is therefore user-centric and action-oriented, with “trust” being achieved by providing tailored insights to a wide range of stakeholders, from grid operators designing tariffs to homeowners managing appliance usage.

### 5.2. Answering the Research Questions

This synthesis directly informs our three research questions:

RQ1: How is ‘trust’ operationalized across domains? Our analysis shows that “trust” is not a monolithic concept but is operationalized differently in each domain. In load forecasting, trust is framed as user alignment, where explanations must be relevant to diverse stakeholders, from grid operators to individual homeowners. In the high-stakes price forecasting domain, trust is equated with financial reliability and risk management, with XAI being developed into tools that assess the real-time trustworthiness of a forecast to prevent financial loss. In renewable generation forecasting, trust is defined by physical plausibility, where the primary goal of XAI is to validate that a model’s logic adheres to known meteorological and physical principles.

RQ2: How well are explanations aligned with user roles? While post hoc methods like SHAP dominate the literature, there is a clear and growing trend toward interpretable-by-design models (e.g., EBMs, HNAMS) that offer greater transparency. However, alignment with specific user roles is inconsistent. The most mature studies tailor explanations for concrete tasks: providing traders with simple risk scores in price forecasting, giving engineers tools for root-cause analysis of turbine underperformance in generation forecasting, and empowering consumers with appliance-level insights in load forecasting. A significant portion of the literature, however, still presents generic feature-importance plots without a clear user or decision-making context in mind.

RQ3: What evidence exists for robustness and operational feasibility? This remains a critical gap across all domains. Most studies do not report on the computational overhead of their XAI methods, making it impossible to assess their feasibility for real-time applications. Similarly, the robustness of explanations is rarely tested. While a few standout studies quantitatively measure the fidelity of their explanations or test them under data distribution shifts, most papers assume their explanations are reliable without providing evidence.

Furthermore, benchmarking against simpler, inherently interpretable statistical models (like ARIMA or MLR) remains a crucial, non-negotiable first step that researchers must take to justify their use of more complex, black-box AI models in the first place.

### 5.3. Limitations of This Review

This review has several limitations that warrant acknowledgment. First, the literature search was limited to three databases (Scopus, ScienceDirect, and Google Scholar), which may have excluded relevant studies from other repositories. Second, the screening of titles, abstracts, and full texts, as well as the data extraction, was conducted by a single reviewer. To mitigate the risk of bias, a structured protocol was strictly followed, and a random 20% sample of excluded studies was re-evaluated to confirm high internal consistency in the application of the inclusion criteria. Finally, while the qualitative assessment using our four-factor framework was performed by the first author, a clear protocol was defined a priori to guide the analysis, and all judgments were documented to support external validation.

Furthermore, a formal risk of bias assessment (e.g., using a tool such as ROBIS or QUADAS) for the 50 included studies was not conducted, which is a limitation of this review. Future work could build upon this by formally evaluating the methodological quality of primary studies. The full dataset is publicly released to enable replication by the research community.

### 5.4. Future Research Directions

Based on the critical gaps identified in this review, we propose several key directions for future research to advance the development of truly trustworthy XAI systems in energy forecasting.

1. **Standardized Evaluation of Explanation Robustness:** Future work must move beyond assuming explanation fidelity. We call for standardized benchmarks and metrics to evaluate robustness under data distribution shifts, input perturbations, and against physical or expert-verified ground truths.
2. **Focus on Operational Viability:** We recommend rigorous reporting of computational costs and latency. This must include analysis of how different architectures scale. For example, while standard Transformers are highly parallelizable, their core self-attention mechanism is known to be computationally intensive and scales poorly with very long input sequences. This can make them prohibitive for very high-resolution (e.g., second-level) data without specialized sparse-attention mechanisms. In contrast, recurrent models like LSTMs/GRUs have different operational trade-offs, often being less computationally demanding per time-step but suffering from slow sequential processing that hinders training. Future work must benchmark these practical trade-offs.
3. **Human-Centered XAI Design:** The field must advance beyond technical explainability toward genuine user utility. We strongly encourage human-subject studies involving domain experts (e.g., grid operators, market analysts) to determine which explanation formats are truly useful, actionable, and trust-building.
4. **Advance Ante-hoc and Physics-Informed XAI:** Future research should explore underdeveloped areas of the XAI taxonomy, including more sophisticated inherently interpretable models and physics-informed XAI that validate model logic against scientific principles, especially in renewable generation forecasting.
5. Our review found that methodological rigor was inconsistent, as primary studies rarely reported formal statistical significance testing (e.g., Diebold-Mariano tests). Future primary studies proposing novel models must include such testing to rigorously validate their claimed performance gains over established baselines.
6. **Adopt Robust Evaluation Metrics:** Future work should also adopt more robust evaluation metrics.
  - MAPE (Mean Absolute Percentage Error), while common, is not recommended for energy datasets as it produces undefined or infinite errors when the true value is zero (e.g., zero solar generation at night).
  - MAE (Mean Absolute Error) is robust and directly interpretable in the units of the forecast (e.g., 'MWh'), making it excellent for understanding the average error magnitude.
  - RMSE (Root Mean Squared Error) is also critical as it squares errors, thus heavily penalizing large deviations. This is often desirable in high-stakes energy applications where missing a single large peak has outsized financial or operational consequences.

We recommend that primary studies report a combination of MAE (for average error) and RMSE (for sensitivity to large errors). For comparing models across datasets of different scales, MASE (Mean Absolute Scaled Error) is a superior, robust alternative to MAPE.

## 6. Conclusions

This systematic review synthesizes 50 peer-reviewed studies on Explainable Artificial Intelligence (XAI) in energy forecasting, introducing a four-factor analytical framework to assess their quality. Our qualitative analysis reveals that XAI application is not uniform but has evolved into three distinct, domain-specific paradigms:

- A user-centric approach in load forecasting, a risk management approach in price forecasting, and a physics-informed approach in renewable generation forecasting.
- Our findings show that while post hoc methods, particularly SHAP, dominate the literature (appearing in 62% of studies), critical gaps persist. Only 23% of papers report on



computational overhead, and fewer than 15% provide evidence of explanation robustness, severely hindering the development of truly trustworthy and deployable AI systems.

To bridge the gap between academic research and real-world deployment, we call for future work to focus on:

- Standardized robustness testing to move beyond assumed fidelity and ensure reliability.
- Rigorous reporting of computational costs and latency to assess operational feasibility in time-sensitive environments.
- Human-centered design, including studies with domain experts to evaluate the actionability and trust-building potential of XAI outputs.

By releasing our full dataset, we aim to facilitate this shift, advancing XAI from a technical add-on to a systematic, domain-sensitive, and operationally grounded discipline that can deliver genuinely trustworthy AI for energy forecasting.

**Author Contributions:** V.A. and R.F.; methodology, V.A. and R.F.; investigation, V.A.; data curation, V.A.; writing—original draft preparation, V.A.; writing—review and editing, R.F.; visualization, V.A.; supervision, R.F.; funding acquisition, R.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Luxembourg National Research Fund (FNR), under the National Center of Excellence in Research (NCER) program as part of the D2ET (Data-Driven Energy Transition) grant number 38/44D2ET.

**Acknowledgments:** During the preparation of this manuscript/study, the author(s) used ChatGPT-3.5 for the purposes improving the language and readability of the manuscript. The graphical abstract was visualized with the assistance of Google Gemini-1.5. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial Neural Network
CART	Classification and Regression Tree
CBM	Concept Bottleneck Model
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CMC	Contribution Monotonicity Coefficient
CNN	Convolutional Neural Network
DA	Day-Ahead
D3Net	Dual-path Dynamic Dense Network
DL	Deep Learning
DNN	Deep Neural Networks
DT	Decision Tree
DXNN	Direct Explainable Neural Network
EBM	Explainable Boosting Machine
ENSWNN	Ensemble of Weighted Nearest Neighbors
ETS	Emissions Trading Scheme
EWS	Extreme Wind Speed
GBDT	Gradient Boosting Decision Trees
GBoost	Gradient Boosting
GDPR	General Data Protection Regulation
GRU	Gated Recurrent Unit
GT	Global Transparency

HNAMs	Hierarchical Neural Additive Models
ICE	Individual Conditional Expectation
ISTR	Interpretable Spatio-Temporal Representation
k-NN	k-Nearest Neighbors
KRC	Kendall's Rank Correlation
LF	Local Fidelity
LIME	Local Interpretable Model-Agnostic Explanations
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
MV-LSTM	Interpretable Multi-Variable LSTM
NAMs	Neural Additive Models
NF	Normalizing Flow
NWP	Numerical Weather Prediction
OHLC	Open-High-Low-Close
OV	Operational Viability
P	Spearman's rank correlation coefficient
PDP	Partial Dependence Plots
PFI	Permutation Feature Importance
QLattice	Symbolic regression framework by Abzu
RF	Random Forest
RNN	Recurrent Neural Network
RQ	Research Question
SHAP	Shapley Additive Explanations
SRC	Spearman Rank Correlation
STGNNs	Spatio-Temporal Graph Neural Networks
STL	Seasonal-Trend decomposition based on Loess
SVR	Support Vector Regression
TabNet	Tabular Network
TCN	Temporal Convolutional Network
TFT	Temporal Fusion Transformer
THI	Temperature-Humidity Index
TS	Time Series
UR	User Relevance
VMD	Variational Mode Decomposition
WT	Wavelet Transform
XAI	Explainable Artificial Intelligence
X-CGNN	Explainable Causal Graph Neural Network
XGBoost	Extreme Gradient Boosting

## Appendix A

This Appendix provides the detailed methodological tools used for the systematic review, as referenced in the main paper. Table A1 details the data extraction template used to systematically capture key information from each of the 50 reviewed studies (see Section 3.3). Table A2 provides the specific 4-factor scoring rubric used to perform the qualitative assessment and assign scores (Low, Medium, High) for each dimension, as detailed in Section 3.4 of the main paper.

**Table A1.** Data Extraction Template for Systematic Review of XAI in Energy Forecasting.

Field	Description	Example
1. Author, Year	First author and publication year	A CrossInformer model based on dual-layer decomposition and interpretability for short-term electricity load forecasting [30].
2. Title	Full paper title	“Explainable AI for Short-Term Load Forecasting in Smart Grids”
3. DOI	Digital Object Identifier	10.1016/j.apenergy.2025.123456
4. Domain	Forecasting domain	Load
5. Forecasting Horizon	Time horizon (short/medium/long)	Short-term (1–24 h)
6. Input Features	Features used (e.g., temperature, load history)	Temperature, humidity, hour-of-day
7. Primary Model	Main forecasting model	LSTM
8. XAI Method	Explanation technique used	SHAP
9. Explanation Scope	Local or global	Local
10. Model Dependency	Model-agnostic or model-specific	Model-agnostic
11. Temporal Integration	Ante-hoc, in-hoc, post hoc	Post hoc
12. Performance Metrics	Accuracy metrics reported	MAPE, R <sup>2</sup>
13. Accuracy	MSE, RMSE, R2, etc.	76.8
14. 4 dimensions	As explained in the text	---

**Table A2.** The 4-Factor Scoring Rubric for Qualitative Assessment of Explanation Quality.

Dimension	High Score (Score = 1)	Medium Score (Score = 0.5)	Low Score (Score = 0)
1. Global Transparency	Study provides clear, global explanations of the model’s overall logic (e.g., full SHAP analysis, EBM shape functions).	Study provides basic global feature importance plots without deeper analysis of model strategy.	Study provides no global-level explanation of the model’s behavior.
2. Local Fidelity & Robustness	Study explicitly tests the robustness or fidelity of its local explanations (e.g., perturbation tests, “reasonableness” scores, or uses inherently faithful models).	Study provides local explanations (e.g., LIME, local SHAP) but does not provide any evidence or testing of their robustness or fidelity.	Study provides no local explanations or evidence of robustness.
3. User Relevance	Study explicitly names a stakeholder (e.g., ‘grid operator,’ ‘trader’) AND links the explanation to a specific decision-making task (e.g., ‘designing tariffs,’ ‘managing risk’).	Study mentions a general user (e.g., ‘experts’) but does not connect the explanation to a specific, actionable decision.	Study only presents technical feature importance without any mention of a user or decision task.
4. Operational Viability	Study provides explicit reporting of computational overhead (e.g., time in seconds, memory cost) OR provides a clear justification for the XAI method choice in a deployment context.	Study makes a general claim about feasibility (e.g., ‘model is fast’) but provides no quantitative metrics or specific evidence.	Study provides no mention of computational cost or other deployment considerations.

## References

1. Baur, L.; Ditschuneit, K.; Schambach, M.; Kaymakci, C.; Wollmann, T.; Sauer, A. Explainability and Interpretability in Electric Load Forecasting Using Machine Learning Techniques—A Review. *Energy AI* **2024**, *16*, 100358. [CrossRef]
2. Shadi, M.R.; Mirshekeli, H.; Shaker, H.R. Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects. *Renew. Sustain. Energy Rev.* **2025**, *216*, 115668. [CrossRef]
3. Melgar-García, L.; Troncoso, A. A novel incremental ensemble learning for real-time explainable forecasting of electricity price. *Knowl. Based Syst.* **2024**, *305*, 112574. [CrossRef]
4. Tian, C.; Niu, T.; Li, T. Developing an interpretable wind power forecasting system using a transformer network and transfer learning. *Energy Convers. Manag.* **2025**, *323*, 119155. [CrossRef]
5. Chen, Z.; Xiao, F.; Guo, F.; Yan, J. Interpretable machine learning for building energy management: A state-of-the-art review. *Adv. Appl. Energy* **2023**, *9*, 100123. [CrossRef]
6. Machlev, R.; Heistrene, L.; Perl, M.; Levy, K.Y.; Belikov, J.; Mannor, S.; Levron, Y. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy AI* **2022**, *9*, 100169. [CrossRef]
7. Heistrene, L.; Machlev, R.; Perl, M.; Belikov, J.; Baimel, D.; Levy, K.; Mannor, S.; Levron, Y. Explainability-based Trust Algorithm for electricity price forecasting models. *Energy AI* **2023**, *14*, 100259. [CrossRef]
8. Ozdemir, G.; Kuzlu, M.; Catak, F.O. Machine learning insights into forecasting solar power generation with explainable AI. *Electr. Eng.* **2024**, *107*, 7329–7350. [CrossRef]
9. Beltrán, S.; Castro, A.; Irizar, I.; Naveran, G.; Yeregui, I. Framework for collaborative intelligence in forecasting day-ahead electricity price. *Appl. Energy* **2022**, *306*, 118049. [CrossRef]
10. Serra, A.; Ortiz, A.; Cortés, P.J.; Canals, V. Explainable district heating load forecasting by means of a reservoir computing deep learning architecture. *Energy* **2025**, *318*, 134641. [CrossRef]
11. Haghighat, M.; MohammadiSavadkoobi, E.; Shafiabady, N. Applications of Explainable Artificial Intelligence (XAI) and interpretable Artificial Intelligence (AI) in smart buildings and energy savings in buildings: A systematic review. *J. Build. Eng.* **2025**, *107*, 112542. [CrossRef]
12. Shajalal, M.; Boden, A.; Stevens, G. ForecastExplainer: Explainable household energy demand forecasting by approximating shapley values using DeepLIFT. *Technol. Forecast. Soc. Change* **2024**, *206*, 123588. [CrossRef]
13. Grzeszczyk, T.A.; Grzeszczyk, M.K. Justifying Short-Term Load Forecasts Obtained with the Use of Neural Models. *Energies* **2022**, *15*, 1852. [CrossRef]
14. Gürses-Tran, G.; Körner, T.A.; Monti, A. Introducing explainability in sequence-to-sequence learning for short-term load forecasting. *Electr. Power Syst. Res.* **2022**, *212*, 108366. [CrossRef]
15. Feddersen, L.; Cleophas, C. Hierarchical neural additive models for interpretable demand forecasts. *Int. J. Forecast.* **2025**; *in press*. [CrossRef]
16. Joseph, L.P.; Deo, R.C.; Casillas-Pérez, D.; Prasad, R.; Raj, N.; Salcedo-Sanz, S. Short-term wind speed forecasting using an optimized three-phase convolutional neural network fused with bidirectional long short-term memory network model. *Appl. Energy* **2024**, *359*, 122624. [CrossRef]
17. Liao, W.; Fang, J.; Ye, L.; Bak-Jensen, B.; Yang, Z.; Porte-Agel, F. Can we trust explainable artificial intelligence in wind power forecasting? *Appl. Energy* **2024**, *376*, 124273. [CrossRef]
18. Darvishvand, L.; Kamkari, B.; Huang, M.J.; Hewitt, N.J. A systematic review of explainable artificial intelligence in urban building energy modeling: Methods, applications, and future directions. *Sustain. Cities Soc.* **2025**, *128*, 106492. [CrossRef]
19. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bannetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
20. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef]
21. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2019. Available online: <https://christophm.github.io/interpretable-ml-book/> (accessed on 14 November 2025).
22. Belle, V.; Papantonis, I. Principles and Practice of Explainable Machine Learning. *Front. Big Data* **2021**, *4*, 688969. [CrossRef]
23. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef]
24. Elsevier. ScienceDirect. Available online: <https://www.sciencedirect.com> (accessed on 5 August 2025).
25. Białek, J.; Bujalski, W.; Wojdan, K.; Guzek, M.; Kurek, T. Dataset level explanation of heat demand forecasting ANN with SHAP. *Energy* **2022**, *261*, 125075. [CrossRef]
26. Wenninger, S.; Kaymakci, C.; Wiethe, C. Explainable long-term building energy consumption prediction using QLattice. *Appl. Energy* **2022**, *308*, 118300. [CrossRef]

27. Choi, W.; Lee, S. Interpretable deep learning model for load and temperature forecasting: Depending on encoding length, models may be cheating on wrong answers. *Energy Build.* **2023**, *297*, 113410. [\[CrossRef\]](#)
28. Mubarak, H.; Stegen, S.; Bai, F.; Abdellatif, A.; Sanjari, M.J. Enhancing interpretability in power management: A time-encoded household energy forecasting using hybrid deep learning model. *Energy Convers. Manag.* **2024**, *315*, 118795. [\[CrossRef\]](#)
29. Ferreira, A.B.A.; Leite, J.B.; Salvadeo, D.H.P. Power substation load forecasting using interpretable transformer-based temporal fusion neural networks. *Electr. Power Syst. Res.* **2025**, *238*, 111169. [\[CrossRef\]](#)
30. Li, H.; Tang, Y.; Liu, D. A CrossInformer model based on dual-layer decomposition and interpretability for short-term electricity load forecasting. *Alex. Eng. J.* **2025**, *129*, 117–127. [\[CrossRef\]](#)
31. Fan, C.; Chen, H. Research on eXplainable artificial intelligence in the CNN-LSTM hybrid model for energy forecasting. *J. Build. Eng.* **2025**, *111*, 113150. [\[CrossRef\]](#)
32. Moon, J.; Rho, S.; Baik, S.W. Toward explainable electrical load forecasting of buildings: A comparative study of tree-based ensemble methods with Shapley values. *Sustain. Energy Technol. Assess.* **2022**, *54*, 102888. [\[CrossRef\]](#)
33. Allal, Z.; Noura, H.N.; Salman, O.; Chahine, K. Power consumption prediction in warehouses using variational autoencoders and tree-based regression models. *Energy Built Environ.* **2024**; *in press*. [\[CrossRef\]](#)
34. Xu, C.; Li, C.; Zhou, X. Interpretable LSTM Based on Mixture Attention Mechanism for Multi-Step Residential Load Forecasting. *Electronics* **2022**, *11*, 2189. [\[CrossRef\]](#)
35. Lu, Y.; Vijayananth, V.; Perumal, T. Smart home energy prediction framework using temporal Kolmogorov-Arnold transformer. *Energy Build.* **2025**, *335*, 115529. [\[CrossRef\]](#)
36. Eskandari, H.; Saadatmand, H.; Ramzan, M.; Mousapour, M. Innovative framework for accurate and transparent forecasting of energy consumption: A fusion of feature selection and interpretable machine learning. *Appl. Energy* **2024**, *366*, 123314. [\[CrossRef\]](#)
37. Miraki, A.; Parviainen, P.; Arghandeh, R. Electricity demand forecasting at distribution and household levels using explainable causal graph neural network. *Energy AI* **2024**, *16*, 100368. [\[CrossRef\]](#)
38. Jiang, Y.; Li, Y.; Chen, Y. Interpretable short-term load forecasting via multi-scale temporal decomposition. *Electr. Power Syst. Res.* **2024**, *235*, 110781. [\[CrossRef\]](#)
39. Mathew, A.; Chikte, R.; Sadanandan, S.K.; Abdelaziz, S.; Ijaz, S.; Ghaoud, T. Medium-term feeder load forecasting and boosting peak accuracy prediction using the PWP-XGBoost model. *Electr. Power Syst. Res.* **2024**, *237*, 111051. [\[CrossRef\]](#)
40. Ghimire, S.; Deo, R.C.; Hopf, K.; Liu, H.; Casillas-Pérez, D.; Helwig, A.; Prasad, S.S.; Pérez-Aracil, J.; Barua, P.D.; Salcedo-Sanz, S. Half-hourly electricity price prediction model with explainable-decomposition hybrid deep learning approach. *Energy AI* **2025**, *20*, 100492. [\[CrossRef\]](#)
41. Lee, J.; Xia, B. Analyzing the dynamics between crude oil spot prices and futures prices by maturity terms: Deep learning approaches to futures-based forecasting. *Results Eng.* **2024**, *24*, 103086. [\[CrossRef\]](#)
42. Ghosh, I.; Jana, R.K. Clean energy stock price forecasting and response to macroeconomic variables: A novel framework using Facebook's Prophet, NeuralProphet and explainable AI. *Technol. Forecast. Soc. Change* **2024**, *200*, 123148. [\[CrossRef\]](#)
43. Jiang, M.; Che, J.; Li, S.; Hu, K.; Xu, Y. Incorporating key features from structured and unstructured data for enhanced carbon trading price forecasting with interpretability analysis. *Appl. Energy* **2025**, *382*, 125301. [\[CrossRef\]](#)
44. Cramer, E.; Witthaut, D.; Mitsos, A.; Dahmen, M. Multivariate probabilistic forecasting of intraday electricity prices using normalizing flows. *Appl. Energy* **2023**, *346*, 121370. [\[CrossRef\]](#)
45. Álvarez-Rodríguez, C.; Parrado-Hernández, E.; Pérez-Aracil, J.; Prieto-Godino, L.; Salcedo-Sanz, S. Interpretable extreme wind speed prediction with concept bottleneck models. *Renew. Energy* **2024**, *231*, 120935. [\[CrossRef\]](#)
46. Liao, W.; Fang, J.; Bak-Jensen, B.; Ruan, G.; Yang, Z.; Porté-Agel, F. Explainable modeling for wind power forecasting: A Glass-Box model with high accuracy. *Int. J. Electr. Power Energy Syst.* **2025**, *167*, 110643. [\[CrossRef\]](#)
47. Verdone, A.; Scardapane, S.; Panella, M. Explainable Spatio-Temporal Graph Neural Networks for multi-site photovoltaic energy production. *Appl. Energy* **2024**, *353*, 122151. [\[CrossRef\]](#)
48. Wu, Z.; Zeng, S.; Jiang, R.; Zhang, H.; Yang, Z. Explainable temporal dependence in multi-step wind power forecast via decomposition based chain echo state networks. *Energy* **2023**, *270*, 126906. [\[CrossRef\]](#)
49. Niu, Z.; Han, X.; Zhang, D.; Wu, Y.; Lan, S. Interpretable wind power forecasting combining seasonal-trend representations learning with temporal fusion transformers architecture. *Energy* **2024**, *306*, 132482. [\[CrossRef\]](#)
50. Zhao, Y.; Liao, H.; Zhao, Y.; Pan, S. Data-augmented trend-fluctuation representations by interpretable contrastive learning for wind power forecasting. *Appl. Energy* **2025**, *380*, 125052. [\[CrossRef\]](#)
51. Letzgs, S.; Müller, K.-R. An explainable AI framework for robust and transparent data-driven wind turbine power curve models. *Energy AI* **2024**, *15*, 100328. [\[CrossRef\]](#)
52. Wang, X.; Hao, Y.; Yang, W. Novel wind power ensemble forecasting system based on mixed-frequency modeling and interpretable base model selection strategy. *Energy* **2024**, *297*, 131142. [\[CrossRef\]](#)

53. Nallakaruppan, M.K.; Shankar, N.; Bhuvanagiri, P.B.; Padmanaban, S.; Bhatia Khan, S. Advancing solar energy integration: Unveiling XAI insights for enhanced power system management and sustainable future. *Ain Shams Eng. J.* **2024**, *15*, 102740. [\[CrossRef\]](#)
54. Sarp, S.; Kuzlu, M.; Cali, U.; Elma, O.; Guler, O. An Interpretable Solar Photovoltaic Power Generation Forecasting Approach Using An Explainable Artificial Intelligence Tool. In Proceedings of the 2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 25–29 July 2021; pp. 1–5.
55. Talaat, F.M.; Kabeel, A.E.; Shaban, W.M. Towards sustainable energy management: Leveraging explainable Artificial Intelligence for transparent and efficient decision-making. *Sustain. Energy Technol. Assess.* **2025**, *78*, 104348. [\[CrossRef\]](#)
56. Prasad, S.S.; Joseph, L.P.; Ghimire, S.; Deo, R.C.; Downs, N.J.; Acharya, R.; Yaseen, Z.M. Explainable hybrid deep learning framework for enhancing multi-step solar ultraviolet-B radiation predictions. *Atmos. Environ.* **2025**, *343*, 120951. [\[CrossRef\]](#)
57. Wang, H.; Cai, R.; Zhou, B.; Aziz, S.; Qin, B.; Voropai, N.; Gan, L.; Barakhtenko, E. Solar irradiance forecasting based on direct explainable neural network. *Energy Convers. Manag.* **2020**, *226*, 113487. [\[CrossRef\]](#)
58. Ukwuoma, C.C.; Cai, D.; Ukwuoma, C.D.; Chukwuemeka, M.P.; Ayeni, B.O.; Ukwuoma, C.O.; Adeyi, O.V.; Huang, Q. Sequential gated recurrent and self attention explainable deep learning model for predicting hydrogen production: Implications and applicability. *Appl. Energy* **2025**, *378*, 124851. [\[CrossRef\]](#)
59. Zhao, Y.; Zhao, Y.; Liao, H.; Pan, S.; Zheng, Y. Interpreting LASSO regression model by feature space matching analysis for spatio-temporal correlation based wind power forecasting. *Appl. Energy* **2025**, *380*, 124954. [\[CrossRef\]](#)
60. Kuzlu, M.; Cali, U.; Sharma, V.; Guler, O. Gaining Insight Into Solar Photovoltaic Power Generation Forecasting Utilizing Explainable Artificial Intelligence Tools. *IEEE Access* **2020**, *8*, 187814–187823. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.