Renewable and Sustainable Energy Reviews 239 (2026) 117171

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

# Harnessing AI-driven large language models (LLMs) for enhanced forecasting and optimisation of wind and solar energy generation and load demand profiles

Omar Aleassa [a], Mahmoud Bekhit [b], Walayat Hussain [b,c], I.M.R. Fattah [a,*]

[a] School of Civil and Environmental Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, 2007, Australia
[b] Peter Faber Business School, Australian Catholic University (ACU), North Sydney, NSW, 2060, Australia
[c] School of Computer Science, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, 2007, Australia

ARTICLE INFO

*Keywords:*
Large language models
Renewable energy forecasting
Wind power
Solar power
Load demand forecasting

ABSTRACT

The accelerated transition to renewable energy requires reliable forecasting and optimisation to support stable, efficient grid operations. In recent years, foundation models, including large language models (LLMs) and time-series foundation models, have been investigated for renewable energy forecasting tasks that integrate diverse inputs, such as sensor telemetry, maintenance logs, weather narratives, and geospatial descriptors. This paper presents a structured review of recent studies published from 2020 to 2025 on the application of language-focused foundation models and related transformer-based foundations to wind, solar, and load demand forecasting. Instead of asserting universal performance advantages, we synthesise evidence showing where studies report improvements compared with conventional machine learning (ML) baselines, most often in contexts involving multimodal fusion, longer forecasting horizons, or text-enriched decision support. We also highlight situations in which the reported benefits are limited or strongly influenced by the dataset and evaluation protocol. In addition, we examined practical constraints affecting deployment, including integration with Supervisory Control and Data Acquisition (SCADA) and Internet of Things (IoT) systems, latency, reliability, and governance. We summarise available interpretability approaches, including Shapley Additive Explanations (SHAP), attention visualisation, and post hoc explanations, and identify key open challenges such as computational cost, hallucination risk, and uncertainty quantification in safety-critical forecasting.

## 1. Introduction

The large-scale integration of renewable energy is essential for decarbonisation, yet it introduces operational volatility that complicates power system stability. Wind and solar generation are inherently intermittent and strongly driven by exogenous meteorological conditions. Consequently, short-term variability and forecast errors can amplify demand–supply mismatch, increase reserve requirements, and degrade scheduling and dispatch decisions, ultimately affecting grid reliability and operating cost [1–3].

Accurate forecasting and optimisation, therefore, remain central to reliable and efficient grid operation under high renewable penetration. The spatial heterogeneity of weather patterns, rapid ramps, and regime shifts in wind speed and solar irradiance can substantially alter generation within minutes to hours, forcing frequent updates to unit commitment, storage control, and demand response policies [4,5]. While classical statistical and ML approaches have delivered strong performance in many settings, their effectiveness can degrade under distribution shifts, sparsity of labels for rare events, and limited ability to exploit heterogeneous data sources beyond numerical time-series. These limitations are particularly relevant as modern energy systems increasingly produce multimodal data streams, including SCADA telemetry, maintenance logs, operator notes, incident reports, and weather narratives.

In this context, language-centric LLMs have emerged as a promising component within next-generation forecasting workflows. Beyond modelling numerical sequences, LLMs can ingest and reason over unstructured or semi-structured information (e.g., textual maintenance

* Corresponding author.
*E-mail addresses:* omar.aleassa@student.uts.edu.au (O. Aleassa), mahmoud.bekhit@acu.edu.au (M. Bekhit), walayat.hussain@acu.edu.au (W. Hussain), islamdrizwanul.fattah@uts.edu.au (I.M.R. Fattah).

https://doi.org/10.1016/j.rser.2026.117171
Received 27 August 2025; Received in revised form 9 May 2026; Accepted 5 June 2026
Available online 9 June 2026

reports, fault descriptions, curtailment notices, weather bulletins, and regulatory documents) and can support task orchestration, retrieval, and explanation through instruction-following and tool use [6–8]. Importantly, not every transformer-based forecaster should be treated as an LLM. A growing body of work also studies transformer architectures trained primarily on numerical time-series (often referred to as time-series foundation models), which are conceptually distinct from language-pretrained LLMs even when both use attention mechanisms. Clarifying these boundaries is necessary for interpreting reported gains and for identifying when language pretraining specifically contributes value in renewable energy forecasting [9–11].

Despite rapid progress, the literature remains fragmented across communities that emphasise (i) conventional ML and deep learning for time-series prediction, (ii) transformer-based forecasters and time-series foundation models, and (iii) language-centric approaches that integrate text, metadata, and decision support. Existing reviews often emphasise statistical and ML forecasting pipelines without a focused synthesis of how LLMs are being used (e.g., instruction-based prompting, retrieval augmentation, finetuning for energy-domain language, and multimodal fusion with time-series signals). For instance, Saberi et al. [12] review AI approaches for wind and solar forecasting, but do not provide a taxonomy of LLM-specific adaptation strategies or the role of text-augmented modelling. Lee and Zhao [13] discuss NLP applications in energy systems, yet they do not consolidate how LLMs interact with forecasting workflows, evaluation protocols, or deployment constraints in operational settings. These gaps motivate a structured synthesis that is explicit about what counts as "LLM-enabled" forecasting and how it differs from other transformer-based time-series methods.

Fig. 1 summarises a task-specific workflow for using LLM-enabled components in renewable energy forecasting. Heterogeneous inputs such as SCADA telemetry, weather forecasts, maintenance logs, operator notes, and market signals are first pre-processed and aligned, then fused through time-series, text, and, where relevant, spatial or image-based representations. Conventional baselines, time-series foundation models, LLM-enabled modules, or hybrid decision-fusion strategies may perform forecasting. The role of LLMs is therefore framed as contextual reasoning, explanation, and decision-support augmentation rather than as a universal replacement for numerical forecasting models. Outputs include point forecasts, uncertainty estimates, alerts, explanations, and operational recommendations, supported by monitoring of accuracy, calibration, latency, drift, and robustness [14,15].

Overall, this article analyses prominent LLM-enabled strategies and forecasting baselines for renewable generation and load demand prediction, with an emphasis on wind, solar, and load forecasting. We synthesise recent research, identify methodological gaps (e.g., inconsistent baselines, limited reporting of compute and latency, and sparse uncertainty evaluation), and summarise opportunities for building more reliable and deployable forecasting systems. Accordingly, this paper addresses the following research question: How can language-centric modelling and LLM-enabled techniques improve renewable energy forecasting pipelines, particularly when heterogeneous data sources and operational constraints are considered?

To position our contribution more precisely relative to prior surveys, Table 1 compares representative recent review papers across scope, review methodology, and analytical emphasis. Existing studies already make important contributions to the broader intersection of LLMs and energy systems, including cross-domain reviews spanning power systems, buildings, and forecasting, as well as power-systems-focused reviews covering forecasting, operation, security, and deployment challenges. Accordingly, we do not claim to be the first review of LLMs in energy. Rather, this paper provides a more targeted synthesis for renewable forecasting and optimisation by focusing specifically on

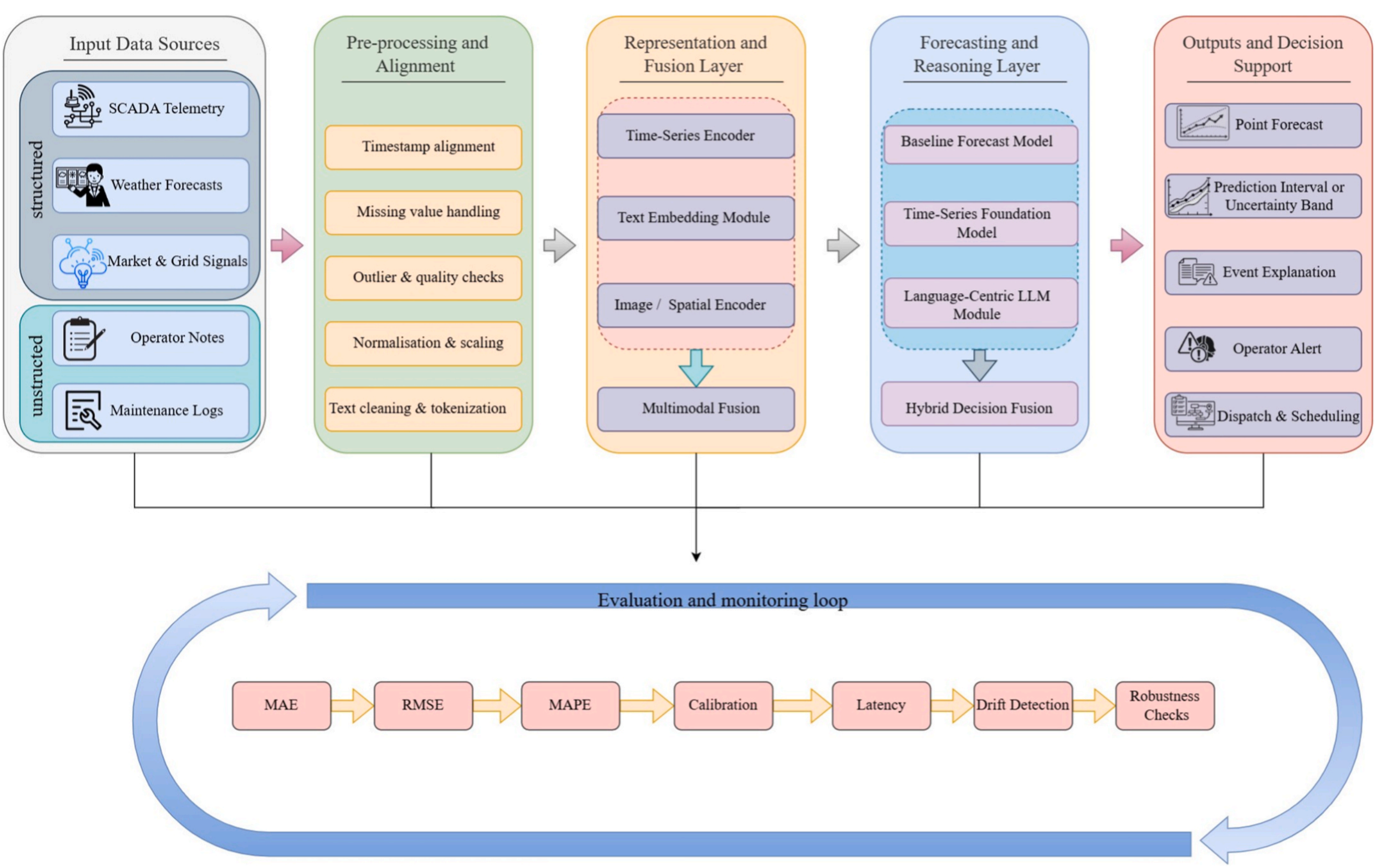


**Fig. 1.** Task-specific workflow for integrating multimodal data, forecasting models, language-centric LLM modules, and evaluation-aware decision support in renewable energy forecasting.

**Table 1**
Comparison of this review with representative prior surveys.

| Ref. | Primary scope | Review method | Renewable forecasting focus | Explicit LLM vs TSFM separation | Operational and evaluation cover-age | What this review adds beyond it |
|---|---|---|---|---|---|---|
| [16] | Cross-domain LLM review covering power systems, buildings, and forecasting | Structured review | Partial | No | Strong coverage of evaluation, finetuning, deployment constraints, and reporting considerations | Focuses specifically on renewable forecasting across wind, solar, and load demand, with the LLM versus TSFM distinction treated as a central analytical boundary |
| [17] | Electrical power and energy systems, including load forecasting, diagnosis, question answering, compliance, and analytics | Systematic analysis | Partial | Not explicit | Covers multimodal inputs, fine-tuning, deployment challenges, and benchmark needs | Provides a renewable forecasting-centred synthesis and more clearly separates language-centric LLM pipelines from generic transformer forecasting models |
| [18] | Power systems applications, challenges, and future directions, including forecasting and optimisation | Comprehensive review | Partial | Not explicit | Strong discussion of applications, interpretability, flexibility, and reliability | Adds a tighter renewable forecasting scope and a clearer methodological boundary between LLM-enabled forecasting and non-language transformer forecasters |
| [19] | Power systems review organised by application domains and strategies, including power forecasting | Comprehensive review with explicit domain framework | Partial | Not explicit | Strong coverage of forecasting, operation, security, demand-side applications, and domain-specific framework design | Moves from a broad power systems framework to a dedicated renewable forecasting synthesis across wind, solar, and load demand, with stronger emphasis on PRISMA-informed screening and forecasting-specific benchmarking and reporting |
| [20] | Building energy applications, opportunities, challenges, and roadmap | Perspective paper | No | Not explicit | Discusses domain customisation, retrieval-augmented generation, multimodal integration, and reliability | Extends beyond building energy to wind, solar, and load forecasting, with PRISMA-informed review reporting and forecasting specific benchmarking discussion |
| **This review** | **Language-centric foundation models and related transformer foundations for wind, solar, and load demand forecasting and optimisation** | **PRISMA informed structured review** | **Yes** | **Yes** | **Benchmarking and reporting guidance, SCADA and IoT integration, latency, reliability, interpretability, hallucination risk, and uncertainty quantification** | **Provides a more targeted and forecasting specific synthesis without claiming to be the first review in the wider LLM-energy literature** |

wind, solar, and load demand applications, by explicitly distinguishing language-centric LLM pipelines from time-series foundation models and generic transformer forecasters, and by linking the literature to forecasting-oriented benchmarking, reporting, SCADA and IoT integration, latency, reliability, interpretability, hallucination risk, and uncertainty quantification.

This paper presents a structured, taxonomy-driven survey of LLM-enabled and related foundation-model approaches for renewable energy forecasting (wind, solar, and load demand). Guided by the gaps highlighted in Table 1, this review makes the following contributions:

- We propose a clear taxonomy that distinguishes (a) language-centric LLM usage (e.g., instruction-following, retrieval-augmented reasoning over text, and text– time-series fusion) from (b) transformer-based time-series forecasters and time-series foundation models, enabling more precise claims and comparisons.
- We synthesise and categorise recent studies (2020–2025) across wind, solar, and load forecasting, highlighting recurring design patterns (multimodal fusion, spatiotemporal modelling, text-augmented decision support) and documenting common threats to validity in evaluation and reporting.
- We consolidate practical considerations for operational adoption, including data interface requirements (e.g., SCADA/IoT streams and maintenance logs), latency and reliability constraints, monitoring, and governance issues that influence deployability.
- We provide a benchmarking and reporting checklist to support fair, reproducible comparison against strong forecasting baselines (including modern transformer forecasters), without asserting universal performance superiority of LLM-based methods.

The remainder of the paper is organised as follows. Section 2 describes the survey methodology and screening procedure. Section 3 reviews forecasting model families and clarifies taxonomy boundaries between LLMs and other foundation models. Section 4 reviews applications of LLM-enabled and hybrid approaches in wind, solar, and load demand forecasting. Section 5 discusses the cross-domain roles of LLM-enabled and hybrid forecasting pipelines in energy forecasting and operations. Section 6 summarises open challenges, including reliability, uncertainty, interpretability, and deployment constraints. Finally, Section 7 concludes the paper and outlines future research directions.

## 2. Review methodology

This article reports a structured literature review designed to map and synthesise recent research on language-centric foundation models and related transformer foundations for renewable energy forecasting

**Table 2**
Exact query strings used for the database searches.

| Source | Query string (paste verbatim) |
|---|---|
| Scopus | TITLE-ABS-KEY((“large language model” OR LLM OR “foundation model” OR “instruction-tuned” OR prompting OR “in-context learning” OR ICL OR “retrieval-augmented generation” OR RAG) AND (forecast* OR predict* OR “time series” OR “load forecasting” OR “load demand” OR wind OR “wind power” OR solar OR photovoltaic OR PV OR “renewable energy” OR “power generation” OR “energy forecasting” OR “energy management” OR grid OR “smart grid” OR SCADA OR IoT OR optimis* OR “decision support”)) AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND (LIMIT-TO (LANGUAGE, “English”)) |
| Google Scholar | (“large language model” OR LLM OR “foundation model” OR “instruction tuned” OR prompting OR “in-context learning” OR ICL OR “retrieval augmented” OR RAG) AND (wind OR “wind power” OR solar OR photovoltaic OR PV OR “load forecasting” OR “load demand” OR “energy forecasting” OR “renewable energy” OR grid OR SCADA OR IoT OR “energy management” OR optimisation) |

and optimisation. We follow key PRISMA-informed reporting elements to document identification, screening, and inclusion decisions, while positioning the study as a systematic-style structured review rather than a full systematic review with comprehensive risk-of-bias scoring and meta-analysis. This positioning reflects the heterogeneity of datasets, baselines, and reporting standards across the literature, which limits the feasibility of pooled effect-size estimation.

### 2.1. *Search strategy and information sources*

The literature search was conducted using two primary sources: Scopus and Google Scholar. Scopus was used as the primary bibliographic database for its structured indexing and export functionality, while Google Scholar was used as a complementary source to improve recall and capture influential preprints and technical reports that may not yet be fully indexed. The search covered publications from 2020 to 2025 and was executed iteratively during manuscript preparation, with the final query run in mid-2025. For Scopus, Boolean queries were applied to titles, abstracts, and keywords. A representative Scopus query is shown in Table 2. For Google Scholar, we used simplified keyword combinations (due to limited Boolean control) and screened results using relevance ranking, focusing on highly cited or clearly empirical studies and applying the same eligibility criteria as used for Scopus.

### 2.2. *Screening and eligibility*

In this section, we specify the eligibility framework used to select the studies retained for the survey. We first define the inclusion criteria, covering the publication window and language constraints, the requirement for explicit LLM involvement, the target renewable energy forecasting and optimisation contexts (wind, solar/PV, and load demand), and the minimum expectations for methodological contribution and empirical evaluation (metrics and baseline comparisons where available). We then outline the exclusion criteria, describing how non-LLM studies, out-of-scope energy domains, non-scholarly or insufficiently substantiated sources, duplicates, and inaccessible or non-English full texts were removed to maintain focus and methodological consistency.

#### 2.2.1. *Inclusion criteria*

Studies were included if published between 2020 and 2025, written in English, and available in full text. Eligible papers explicitly involved large language models or language-centric foundation models, such as GPT-family models, LLaMA, or instruction-following variants, in renewable energy forecasting, prediction, optimisation, or decision support. The scope required that each study address at least one target domain: wind power, solar PV generation, or electrical load demand forecasting and management. We retained studies that reported a clear

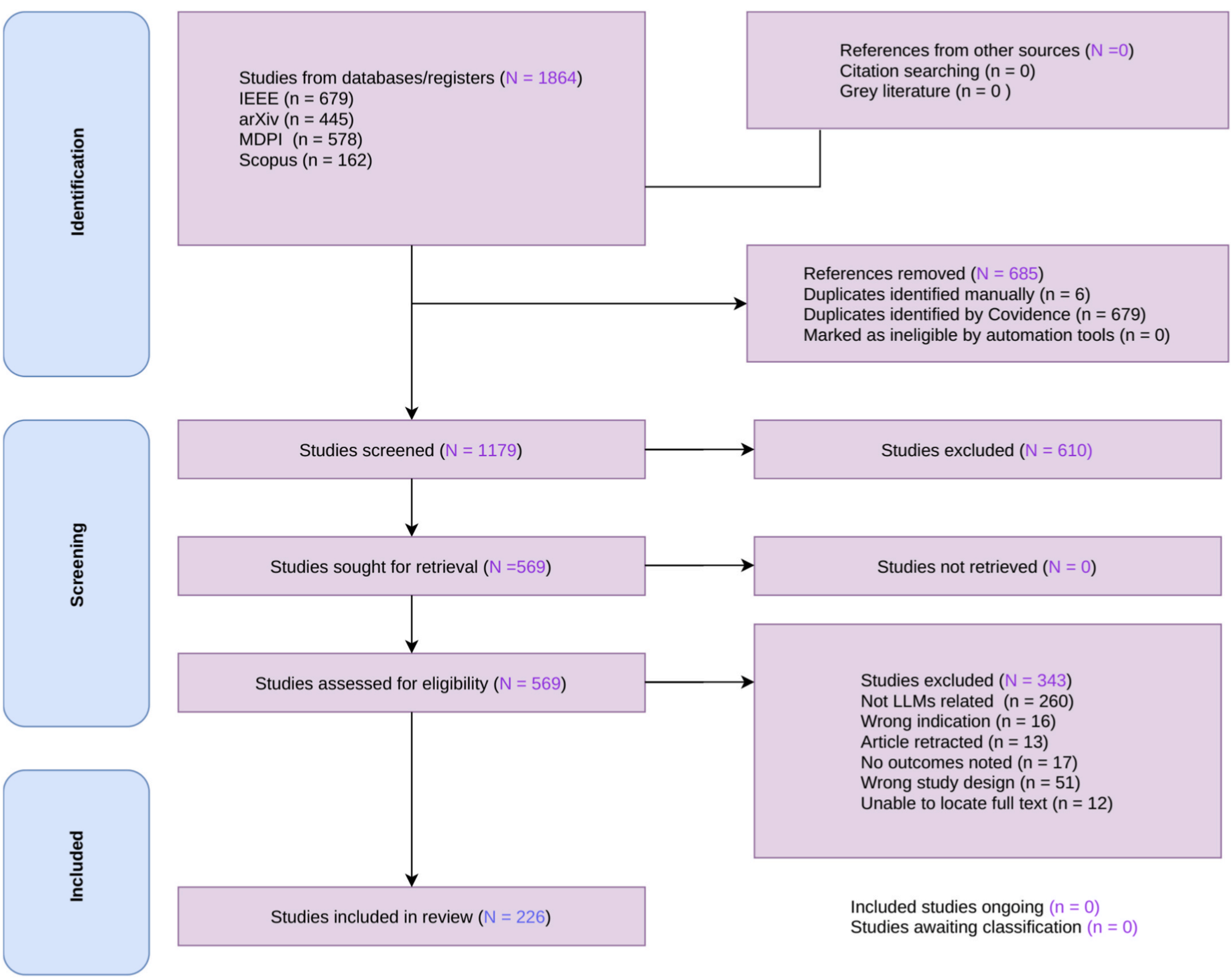


**Fig. 2.** PRISMA style flow diagram for study screening and selection.

methodological contribution, including model adaptation strategies, prompt design and in-context learning configurations, fine-tuning procedures, tool use, multimodal fusion, or integration into operational forecasting pipelines, and we also included works that offered actionable insights related to scalability, interpretability, reliability, or deployment constraints. For empirical studies, eligibility required evaluation using simulations or real-world datasets and reporting quantitative performance using metrics such as RMSE, MAE, R-squared, or MAPE, preferably alongside comparisons against one or more baseline models. Interdisciplinary hybrid approaches were also included when LLMs were combined with complementary methods such as reinforcement learning, physics-informed modelling, or federated learning, provided that the LLM component played a substantive role in the forecasting or optimisation workflow.

#### 2.2.2. Exclusion criteria

Studies were excluded when they did not explicitly involve LLMs, such as papers focused only on classical machine learning, deep learning (DL), statistical forecasting, or transformer forecasting architectures that lacked language-centric pretraining or instruction-following capabilities. We also excluded work that was not aligned with renewable energy or load demand forecasting and optimisation, meaning studies outside wind power, solar or photovoltaic generation, and load demand contexts. Non-scholarly or insufficiently substantiated sources were removed, including blog posts, brief opinion pieces, and position statements that lacked adequate methodological detail or evaluative evidence. Finally, duplicate records were discarded, and we excluded papers that were not written in English or whose full text could not be accessed.

### 2.3. Screening and selection procedure

The initial database search yielded 1864 records from four databases: IEEE (n = 679), arXiv (n = 445), MDPI (n = 578), and Scopus (n = 162), with no additional references identified through citation searching or grey literature. Following duplicate removal — comprising 6 duplicates identified manually and 679 identified via Covidence (total N = 685) — 1179 records remained and were screened at the title and abstract level to assess their alignment with the scope of this review, with particular emphasis on explicit large language model involvement and relevance to renewable energy forecasting and optimisation. Of these, 610 were excluded at the screening stage, leaving 569 studies sought for full-text retrieval, all of which were successfully retrieved (N = 0 not retrieved). Studies that appeared potentially relevant were then examined in full text to confirm eligibility based on the inclusion and exclusion criteria defined in Section 2.2; a further 343 studies were excluded at this stage due to not being LLMs-related (n = 260), wrong study design (n = 51), no outcomes noted (n = 17), wrong indication (n = 16), article retraction (n = 13), or inability to locate the full text (n = 12). The overall filtering workflow and final study set are summarised in the PRISMA-style flow diagram shown in Fig. 2. In total, 226 studies satisfied the eligibility criteria and were retained for synthesis and taxonomy-driven analysis.

Screening was performed in two stages (title/abstract review followed by full-text assessment. During this process, each record was examined for three core aspects: explicit use of large language models or language-centric foundation models; substantive relevance to renewable energy forecasting or optimisation; and sufficient methodological and evaluative detail to support comparative synthesis. Because this is a PRISMA-informed structured review rather than a full systematic review and meta-analysis, a formal inter-rater agreement statistic was not computed; nor was a standardised risk of bias assessment tool applied. Instead, methodological transparency was strengthened by explicit reporting of information sources, search strings, eligibility criteria, and screening decisions, while study quality considerations were addressed qualitatively during the synthesis stage, particularly regarding dataset scope, baseline strength, reporting completeness, and external validity.

### 2.4. Data extraction and synthesis framework

For each included study, we extracted a consistent set of descriptors to enable structured comparison across wind, solar, and load forecasting:

- Task definition: point forecasting, probabilistic forecasting, anomaly detection, predictive maintenance, or optimisation support.
- Forecast horizon and resolution: minutes to hours, days to weeks, seasonal horizons.
- Model class: language-centric LLM, time series foundation model, hybrid pipeline (LLM plus spatiotemporal or graph encoder), or transformer forecaster without language pretraining.
- Data modalities: time series, SCADA and IoT telemetry, weather variables, imagery, text logs and reports, and mixed modality configurations.
- Adaptation method: fine-tuning, parameter-efficient tuning (e.g., LoRA), in-context learning, retrieval augmentation, embedding-based feature extraction.
- Evaluation protocol: datasets, baselines, metrics, and whether uncertainty, calibration, or statistical testing was reported.
- Practical considerations: compute/cost indicators when reported, latency constraints, deployment context, interpretability method, and reliability considerations.
- The synthesis is primarily narrative and tabular, emphasising evidence-linked claims and recurring design patterns rather than pooled quantitative meta-analysis.

### 2.5. Feature engineering basis

To describe how renewable forecasting studies typically represent temporal structure and weather-driven variability, we summarise common feature-engineering practices reported in the literature. Frequently used constructs include time-trend features (e.g., smoothing and regression-based trend extraction), seasonal components (e.g., STL decomposition and Fourier-based periodicity proxies), and weather–energy association measures (e.g., correlation and mutual information between meteorological variables and outputs). We use these feature categories as an organising lens when comparing pipelines that integrate structured telemetry with additional modalities, including text-derived signals.

## 3. Background on forecasting models

Accurate forecasting is fundamental to the reliable operation of renewable-integrated power systems. Fast-changing meteorological conditions drive wind and solar generation, and load demand reflects both weather and socio-economic dynamics [21]. Consequently, forecasting pipelines must handle heterogeneous inputs (telemetry, numerical weather prediction, exogenous variables) and deliver predictions at multiple horizons for dispatch, storage scheduling, and planning. Building on the structured review in the previous section, this section summarises model families commonly used in renewable forecasting and clarifies terminology to avoid conflating (i) language-centric LLMs, (ii) time-series foundation models, and (iii) transformer forecasters that are not language-centric.

### 3.1. Non-LLM machine learning and time-series forecasting models

Before discussing LLMs, it is necessary to review the broader family of non-LLM forecasting approaches that have shaped renewable energy prediction. Renewable energy forecasting relies on heterogeneous inputs, including meteorological variables, historical generation records, numerical weather prediction (NWP) outputs, satellite imagery, maintenance logs, and market or operational signals [22]. These non-LLM approaches provide the methodological foundation for many current forecasting systems and serve as essential baselines for evaluating the added value of LLM-based frameworks.

#### 3.1.1. Classical ML and deep learning models

Classical machine learning (ML) methods have been widely used in renewable energy forecasting because they can learn predictive relationships from structured data and support short-term and medium-term forecasting tasks for solar, wind, and load prediction [23]. Early and still relevant approaches such as linear regression, support vector machines (SVMs), random forests, and gradient boosting remain useful because they are relatively interpretable, computationally manageable, and often effective when datasets are moderate in size, and feature engineering is strong [24,25]. These methods helped establish data-driven renewable forecasting by showing that weather variables, historical production data, and environmental measurements can be mapped to future energy outputs with reasonable accuracy.

As forecasting tasks became more nonlinear and data-rich, deep learning models gained attention for their ability to capture complex temporal and spatial dependencies. Artificial neural networks (ANNs) can model nonlinear relationships in renewable generation data, while long short-term memory (LSTM) networks are well-suited to sequential forecasting because they can capture temporal dependencies across time steps. Convolutional neural networks (CNNs) are especially useful for forecasting when the inputs contain spatial structure, such as satellite imagery, sky images, LiDAR fields, or gridded meteorological maps [26]. Their utility, however, is not restricted to visual inputs. Sequential or textual information, such as meteorological reports or maintenance records, can be transformed into numerical matrices via embeddings, term-frequency–inverse-document-frequency representations, or related feature-mapping procedures, enabling CNNs to extract local patterns from non-visual data as well. Hybrid methods further combine complementary modelling strengths, for example, by integrating statistical decomposition, signal processing, or multiple learning modules to improve predictive robustness across diverse operating conditions.

Although these models remain important, their limitations have also become clear. Classical ML methods often depend heavily on feature engineering and may struggle with highly nonlinear or long-range temporal dependencies [27]. Deep learning models can improve predictive accuracy, but they typically require larger datasets, careful hyperparameter tuning, and substantial computational resources [28]. These limitations motivated the development of more specialised time-series forecasting architectures designed for long-horizon prediction and more complex temporal dynamics.

#### 3.1.2. Modern transformer and mixer-based time-series forecasters

Renewable energy forecasting has increasingly adopted modern time-series architectures such as Informer, FEDformer, NSTransformer, TimesNet, PatchTST, TSMixer, TimeMixer, and TimeXer. These models are specialised time-series forecasting architectures designed to improve long-horizon prediction, multiscale temporal pattern extraction, computational efficiency, or robustness under non-stationary conditions [29].

Their growing popularity is mainly due to several practical advantages. Informer reduces the computational burden of self-attention for long sequences, making long-horizon forecasting more tractable. FEDformer introduces frequency-enhanced decomposition that helps capture seasonal and long-range temporal structure [30]. NSTransformer explicitly addresses non-stationarity and regime variation, which are common in renewable generation and demand data [31]. TimesNet and PatchTST have become strong baselines because they capture multi-period temporal structure and patch-level sequence information effectively [32], while TSMixer, TimeMixer, and TimeXer provide efficient alternatives for modelling temporal and cross-variable interactions in multivariate forecasting tasks [33]. Because of these properties, such models are now widely used as competitive baselines in solar, wind, and load forecasting studies.

Table 3 summarises representative non-LLM forecasting models used in renewable-energy applications, including classical ML, deep-learning, hybrid, transformer-based, and mixer-based time-series models. The table is retained in the main text because these models provide essential baseline context for evaluating whether LLM-enabled methods add value beyond established numerical forecasting approaches.

### 3.2. Taxonomy boundaries between LLMs and related foundation models

Transformer-based forecasting approaches are increasingly discussed together in renewable-energy research, but they should not be treated as a single model class. In this review, LLMs refer to language-centric models, such as GPT-family models, LLaMA, and instruction-tuned derivatives, that are pretrained primarily on natural-language corpora and adapted through prompting, in-context learning, retrieval-augmented generation, or fine-tuning. Their value is most relevant when forecasting workflows involve textual or contextual information, such as maintenance logs, operator notes, regulatory documents, market reports, or weather narratives [52,53].

By contrast, time-series foundation models such as Chronos and TimesFM are pretrained primarily on numerical temporal sequences [54]. Although they may use Transformer architectures, they are not LLMs because their pretraining objectives, input representations, and transfer mechanisms are time-series-centric rather than language-centric. Graph, spatiotemporal, and multimodal foundation models should also be treated as adjacent model families unless language-pretrained or instruction-following components play a substantive role in the forecasting workflow [14].

Accordingly, this review attributes reported benefits to the appropriate model class. Gains from prompting, retrieval over text, instruction-following, or text-aware fusion are treated as LLM-enabled contributions. Gains from numerical sequence pretraining, spatiotemporal modelling, or graph-based transfer are attributed to time-series, graph, or hybrid forecasting models rather than to LLMs by default. Fig. 3 summarises this taxonomy by distinguishing language-centric LLMs, time-series foundation models, graph/spatiotemporal foundation models, and hybrid multimodal systems.

### 3.3. Comparative dimensions of forecasting models

To illustrate how LLM-enabled forecasting may interact with operational energy systems, Fig. 4 presents an implementation-oriented view of renewable forecasting in SCADA and IoT environments. The blueprint highlights data ingestion, validation, feature extraction, retrieval, model serving, decision support, fallback control, and monitoring. It is included to show that operational forecasting depends not only on predictive accuracy, but also on latency, auditability, fallback mechanisms, and integration with operator-facing systems.

While both conventional ML models and LLM-enabled approaches can support renewable energy forecasting, their relative value depends strongly on task characteristics, data modality, interpretability requirements, and deployment constraints. Rather than assuming that one model family is universally superior, Fig. 5 conceptually illustrates the conditions under which each approach is more appropriate, and Table 4 summarises the main trade-offs reported in the reviewed literature. In general, conventional ML and specialised time-series models remain strong baselines when the forecasting task is narrowly defined, the prediction horizon is short to moderate, and the inputs are primarily structured numerical variables. By contrast, LLM-enabled approaches become more useful when forecasting must incorporate text-rich, semistructured, or multimodal context, including maintenance logs, metadata, retrieval outputs, explanations, and operator-facing decision support.

Based on this synthesis, the literature reviewed suggests that model

**Table 3**
Summary of representative non-LLM forecasting models for renewable energy forecasting, including classical ML, deep learning, hybrid models, and recent transformer or mixer-based time-series forecasters.

| Ref. | Model Type | Application in Renewable Energy Forecasting | Advantages | Limitations |
|---|---|---|---|---|
| [34] | Linear Regression | Predicting solar and wind power generation based on historical data and weather variables. | Simple to implement, interpretable results. | Limited to linear relationships, sensitive to outliers. |
| [35] | Support Vector Machines (SVM) | Regression and classification tasks for energy demand and generation prediction. | Effective for small- to medium-sized datasets with clear margins. | Computationally intensive for large datasets. |
| [36] | Random Forest | Modelling nonlinear relationships in energy production data. | Handles nonlinear relationships and is robust to overfitting. | Can become complex and slow with large datasets. |
| [37] | Artificial Neural Networks (ANN) | Capturing complex patterns in solar, wind, and hydro energy generation. | Can model complex and nonlinear relationships. | Requires large datasets and is prone to overfitting. |
| [38] | Long Short-Term Memory (LSTM) Networks | Time-series forecasting for solar and wind power generation. | Captures long-term dependencies in time-series data. | High computational cost, requires large datasets. |
| [39] | Gradient Boosting Machines (GBM) | Enhancing energy forecasting accuracy with boosting techniques. | High accuracy through ensemble learning. | Prone to overfitting if not properly tuned. |
| [40] | Convolutional Neural Networks (CNN) | Extracting spatial and temporal features from satellite imagery, and from text or timeseries data converted into two-dimensional matrices (e.g., embedded word vectors, spectrograms) for forecasting. | Good at capturing spatial features. | High computational requirements. |
| [41] | Hybrid Models (e.g., ANN + ARIMA) | Combining linear and nonlinear models to improve prediction accuracy for energy output. | Combines the strengths of different models for improved performance. | Complex implementation and parameter tuning. |
| [42, 43] | Informer | Long-horizon wind, solar, and load forecasting using efficient attention to reduce complexity for long sequences. | Scales to long sequences and performs strongly in long-horizon forecasting. | Sensitive to tuning and preprocessing; not language-centric; may degrade under distribution shift. |
| [30, 44] | FEDformer | Renewable generation and load forecasting leveraging frequency-domain decomposition for long-term temporal structure. | Captures seasonal and frequency patterns effectively; strong long-horizon performance. | Added architectural complexity; requires careful evaluation; not designed for text inputs. |
| [45] | NSTransformer | Forecasting under nonstationary renewable and demand patterns by modelling distribution shift and changing temporal statistics. | Better handling of non-stationarity and regime changes than vanilla transformers. | Can be harder to train; improvements may be dataset-dependent; not language-centric. |
| [28, 46] | TimesNet | Wind, solar, and load forecasting via multi-periodicity modelling and temporal pattern extraction across horizons. | Effective at capturing multiscale temporal patterns; strong general-purpose baseline. | Requires careful hyperparameter selection; not designed for uncertainty or text integration by default. |
| [47, 48] | PatchTST | Multivariate renewable and demand forecasting using patch-based tokenisation to improve learning efficiency and generalisation. | Strong accuracy and efficiency; widely adopted in recent benchmarks. | Purely numeric forecaster; interpretability remains limited without post-hoc analysis. |
| [49] | TSMixer | High-throughput renewable and load forecasting using temporal and channel mixing as an efficient alternative to attention. | Fast inference with competitive accuracy and lower compute cost. | May underperform on complex long-range or spatiotemporal dependencies. |
| [50] | TimeMixer | Renewable and demand forecasting with improved temporal feature mixing across scales for long-horizon stability. | Competitive performance with an efficient architecture. | Newer model with fewer domain-specific validations; sensitive to protocol and forecast horizon. |
| [51] | TimeXer | Transformer-style forecasting is designed for stronger cross-variable interaction modelling in multivariate time series. | Strong multivariate modelling and competitive benchmark performance. | Requires careful tuning; not designed for text integration; compute can exceed simpler mixer baselines. |

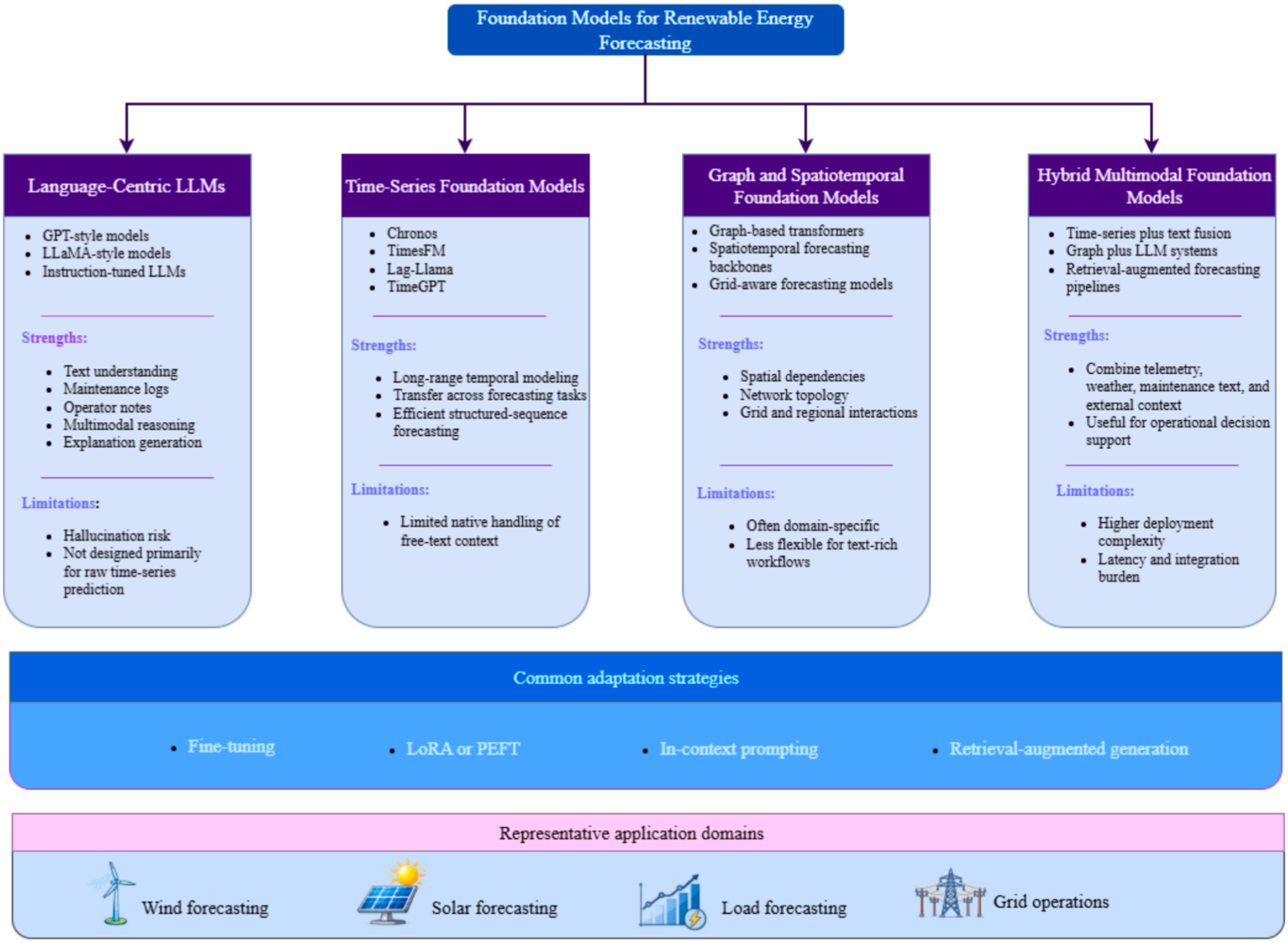


**Fig. 3.** Taxonomy of foundation models for renewable energy forecasting.

selection in renewable energy forecasting should be driven by task characteristics rather than broad assumptions about model superiority. If the objective is narrowly defined, short-term prediction on structured numerical data, conventional ML and specialised forecasting architectures may remain the most practical choice. If the objective involves multimodal fusion, explanation, knowledge retrieval, anomaly interpretation, or interaction with operators through natural language, LLM-enabled pipelines may offer additional value. Accordingly, a rigorous comparison between model classes should use matched datasets, clearly stated forecasting horizons, transparent evaluation metrics, and strong baseline selection.

Although conventional ML and LLM-enabled approaches differ in their primary strengths, they should not be viewed as mutually exclusive alternatives in renewable energy applications. LLM-enabled methods are mainly used for language-centric and knowledge-intensive tasks, such as literature mining, information extraction from technical documents, chatbot-style interfaces for grid interaction, and automated report or code generation for simulation and optimisation workflows [69–73]. By contrast, conventional ML remains more established for structured numerical tasks, including fault detection in photovoltaic and wind systems, energy demand forecasting from time-series data, and optimisation of smart-grid operation through supervised and reinforcement learning pipelines [74]. Nevertheless, both families have been applied to pattern recognition, anomaly detection, renewable energy data analysis, decision support, and modelling of climate-sensitive system behaviour [8,69,75]. Overall, the literature suggests that their greatest practical value often emerges in complementary or hybrid workflows, where ML supports core numerical prediction while LLM-enabled components contribute contextual interpretation, reasoning over unstructured information, and operator-facing explanation.

## 4. Applications of LLM-enabled and hybrid approaches

In this section, we distinguish direct language-model contributions from gains produced by broader hybrid forecasting pipelines. Benefits are attributed specifically to LLMs when they stem from language-centric capabilities such as prompting, retrieval over textual evidence, instruction-following, or text-aware multimodal fusion. By contrast, when performance gains are driven primarily by remote sensing, spatio-temporal forecasting, computational fluid dynamics, optimisation, or uncertainty-aware numerical modelling, we refer to these as outcomes of LLM-enabled hybrid systems or adjacent forecasting model classes rather than of standalone LLMs.

### 4.1. Solar energy forecasting

Solar energy forecasting remains one of the most active areas of LLM- and hybrid-enabled forecasting research, yet the reported evidence suggests that the contribution of language-centric models is not uniform across tasks. As summarised in Table 5, some studies use foundation-model architectures directly for photovoltaic forecasting. In contrast,

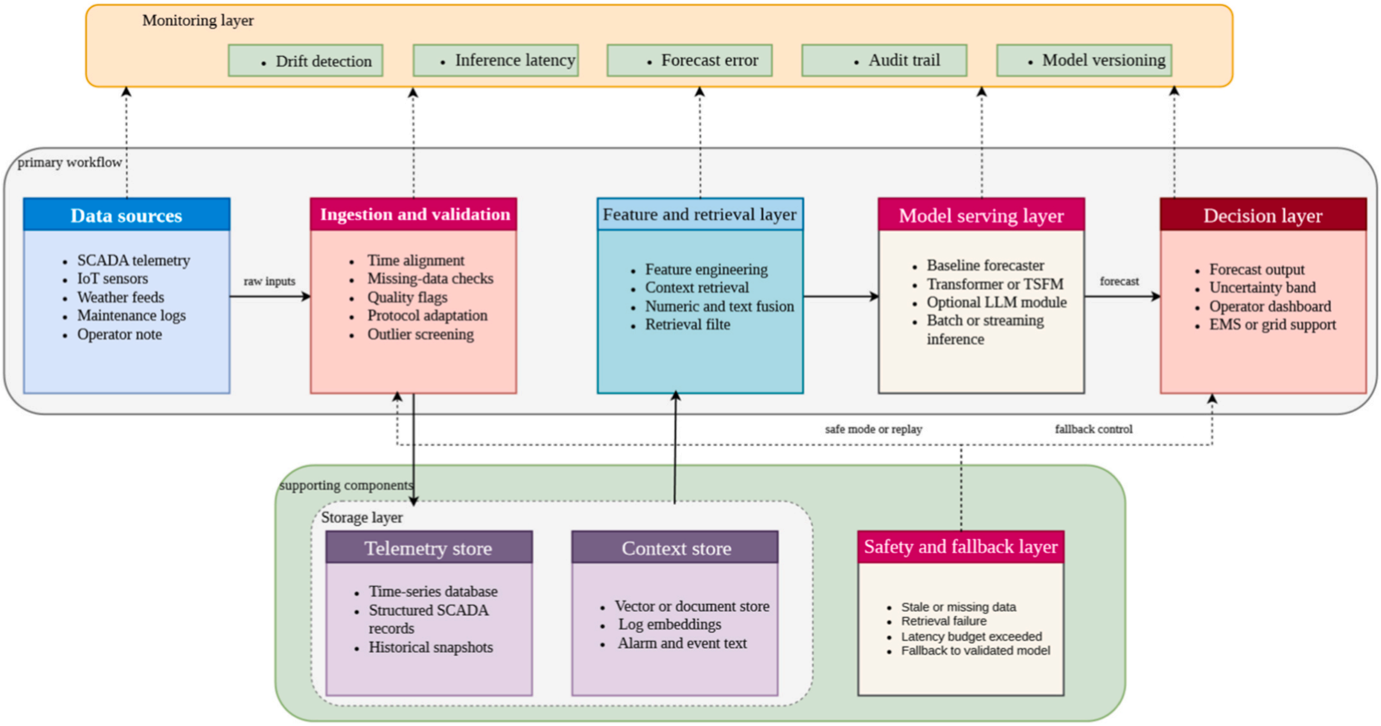


**Fig. 4.** Implementation-oriented deployment blueprint for LLM-enabled renewable energy forecasting in operational SCADA environments.

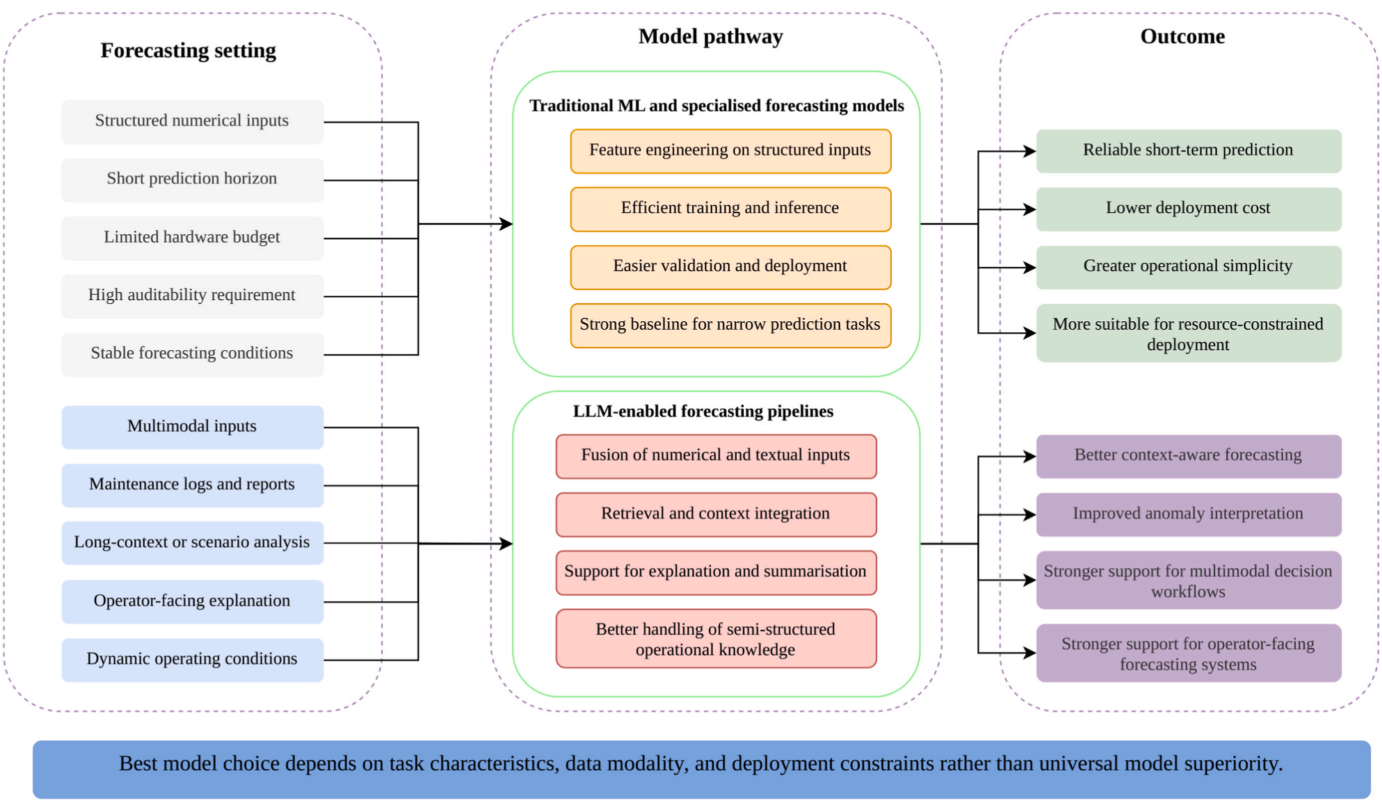


**Fig. 5.** Task-dependent comparison of traditional ML and LLM-enabled approaches in renewable energy forecasting.

others use multimodal large models for adjacent tasks such as photovoltaic panel detection, localisation, or contextual interpretation. This distinction is important because not all reported gains correspond to direct forecasting improvement.

For direct solar forecasting, several recent studies reformulate photovoltaic prediction as a foundation-model or language-model task and report competitive quantitative performance. For example, Chronos-based zero-shot solar forecasting outperformed multiple statistical and neural baselines across 30 min, 1 h, and 2 h horizons, achieving a 58.97% reduction in 30 min RMSE relative to SeasonalNaive

**Table 4**
Comparison of traditional ML and LLM-enabled approaches in renewable energy forecasting.

| Dimension | Traditional ML and conventional deep learning | LLM-enabled approaches | Review interpretation |
|---|---|---|---|
| Forecasting performance | Often strong on structured numerical time-series, especially when task definition, horizon, and input variables are well specified [27,55] | May provide advantages when forecasting is combined with text, metadata, retrieval, explanation, or multimodal context [56] | Reported gains are context-dependent and not universal; fair comparisons require matched datasets, horizons, and strong baselines. |
| Computational cost | Typically lower training and inference cost; easier to deploy in constrained environments [57,58] | Usually higher computational and memory demands, especially for fine-tuning or real-time multimodal use [59,60] | Operational feasibility depends on latency budgets, hardware, and maintenance burden |
| Data requirements | Can perform well with moderately sized structured datasets when feature design is appropriate [61–63] | Often benefits from larger corpora, domain adaptation, and access to unstructured or semi-structured inputs [64,65] | The value of LLMs is more plausible when additional text-rich context is genuinely available. |
| Interpretability and auditability | Often easier to inspect, validate, and operationalise in regulated environments settings [66] | Can support natural-language explanations, but internal decision processes may remain difficult to audit [67] | Explainability should be assessed separately from predictive performance. |
| Deployment suitability | Well-suited to stable, narrowly defined forecasting pipelines [68] | Potentially useful in hybrid workflows involving forecasting, retrieval, summarisation, operator support, and cross-modal reasoning [16] | Method choice should depend on task scope, risk tolerance, and deployment constraints rather than model class alone |

[79] on one dataset. Likewise, RF-PVELLM-DBN combined a photovoltaic large language model, domain knowledge, and dynamic Bayesian network constraints, achieving an MAPE of 1.983%, RMSE of 2.98, and $R^2 = 0.958$ using only one month of training data, while also maintaining strong weather-specific performance under sunny, rainy, and cloudy conditions [79]. Other studies also support the potential of language-model-based architectures under transfer-constrained settings. Pegasus-based prompt learning achieved full-sample RMSE/MAE/MSE/$R^2$ of 0.2960/0.1894/0.0876/0.9736 and remained robust in few-shot and zero-shot scenarios [80], while StairLLM-Net reported the best MSE in five of six zero-shot cross-station groups and the best MAE in four of six groups [81].

At the same time, the solar literature also includes studies whose main contribution is not forecasting itself but rather multimodal perception or contextual support. For instance, the PVAL framework based on GPT-4o addressed solar panel detection and quantification from satellite imagery rather than generation forecasting, achieving a solar F1-score of 82.87%, solar recall of 99.15%, and weighted average F1-score of 94.96% [76]. Such results are important for solar asset mapping and downstream planning, but they should not be interpreted as direct evidence of improved photovoltaic power forecasting. This reinforces the need to distinguish between forecasting-oriented evidence and detection- or assessment-oriented evidence when reviewing LLM applications in the solar domain.

Overall, the merged evidence suggests that LLM-enabled solar

**Table 5**
Synthesis of representative studies comparing limitations of conventional ML with the novelty and reported evidence of LLM-enabled or hybrid approaches across solar, wind, and load demand forecasting.

| Domain | Limitation of conventional ML | Main novelty of the LLM-enabled or hybrid approach | Reported within-study performance and caveats | Ref. |
|---|---|---|---|---|
| Solar | Conventional ML and CNN-based PV detection methods require large annotated datasets, frequent retraining, and often struggle to generalise across diverse rooftop structures, geographic regions, and environmental conditions; they also misclassify visually similar objects such as shadows, parking lots, and flat surfaces | Proposes the PVAL framework based on GPT-4o for solar panel detection, localisation, and quantification from satellite imagery. The main novelty is the combination of multimodal LLM reasoning with task decomposition, standardised JSON output, five-shot prompting, domain-specific fine-tuning, and a confidence-driven auto-labelling mechanism for scalable PV assessment across heterogeneous regions. | This study used a detection-oriented rather than forecasting-oriented evaluation protocol. The fine-tuned PVAL model reported a solar F1-score of 82.87%, solar recall of 99.15%, and a weighted average F1-score of 94.96%. In the Orlando case study, the fine-tuned version reported an F1-score of 97.52%, compared with 92.42% for the prompted version. The study also reported a test cross-entropy loss of 0.12 ± 0.02 across the evaluated methods. For downstream assessment tasks, location prediction accuracy was reported as 87.38% on the fine-tuned dataset and 71.24% on the test dataset. The confidence-based auto-labelling policy was estimated to label approximately 65–70% of images with more than 95% precision. Because this evaluation focuses on PV detection, localisation, and assessment rather than power-output forecasting, these results should not be interpreted as direct evidence of improved renewable-energy forecasting performance. | [76] |
| Solar | Statistical models struggle with nonlinear weather driven variability, while conventional ML and neural models require substantial | Introduces a Chronos-based LLM framework for short-term solar forecasting by discretising continuous time series into tokens, enabling zero-shot | The study evaluated Chronos on four solar datasets against seven statistical baselines, DeepAR, and PatchTST under its selected | [77] |

(*continued on next page*)

**Table 5** (*continued*)

| Domain | Limitation of conventional ML | Main novelty of the LLM-enabled or hybrid approach | Reported within-study performance and caveats | Ref. |
|---|---|---|---|---|
| | historical data and are weaker in data-scarce settings | forecasting without fine-tuning on the target solar datasets | deterministic and probabilistic evaluation settings. Within that study protocol, Chronos reported favourable results across several 30 min, 1 h, and 2 h forecasting horizons. Example reported values include RMSE/MAE = 3.119/1.360 MW for Dataset 1 at the 30 min horizon and QS/CRPS = 0.377/0.734 MW for Dataset 3 at the 30 min horizon. The study also reported a 58.97% RMSE reduction on Dataset 1 at the 30 min horizon relative to the Seasonal Naive baseline. These results should be interpreted as study-specific evidence rather than as directly comparable benchmark values across the wider solar-forecasting literature. | |
| Wind | Existing statistical, ML, and standard deep-learning methods struggle with the nonlinear, nonstationary, and changing spatio-temporal structure of wind power data. | Fine-tunes a pre-trained BERT model for spatiotemporal wind forecasting using patch-based encoding, redesigned temporal and spatial embeddings, LoRA adaptation, and multi-stage training | Within the study's evaluation protocol, BERT4ST reported favourable MAE/RMSE values on the Chuzhou wind-power dataset across the tested horizons: 1.20/2.19 at 10 min, 3.03/5.58 at 1 h, and 6.78/11.91 at 4 h. On the NorthSea dataset, the reported MAE/RMSE values were 0.378/0.550, 0.575/0.827, and 0.980/1.395 across the corresponding horizons. These results were reported to be lower than those of the baselines included in that study, but they should be interpreted within the specific datasets, horizons, preprocessing choices, and baseline configurations used by the authors. | [78] |
| Solar | Conventional black-box ML models for PV forecasting rely heavily on large, high-quality historical datasets, ignore domain expert knowledge, increase computational cost, and offer limited interpretability and adaptability under complex multi-weather conditions. | Proposes RF-PVELLM-DBN, a knowledge-enhanced hybrid forecasting framework for zero-carbon residential buildings that fine-tunes a DeepSeek-R1-based photovoltaic LLM with domain knowledge, uses a PVQA workflow with scenario engineering and question filtering, encodes expert knowledge into DBN whitelist and blacklist constraints, and combines this with Random-Forest feature selection for weather-adaptive forecasting | Reported within-study forecasting evidence was based on one year of 15-min PV data from a zero-carbon residential building in Tibet. Under the study's small-sample setting, using only one month of training data, RF-PVELLM-DBN reported MAPE = 1.983%, RMSE = 2.98, and $R^2$ = 0.958. The authors reported that these values were favourable relative to the LSTM and Transformer baselines included in their experimental protocol. Weather-specific results were also reported: RMSE = 1.29, MAPE = 0.661%, and $R^2$ = 0.9946 under sunny conditions; RMSE = 1.66, MAPE = 0.877%, and $R^2$ = 0.991 under rainy conditions; and RMSE = 1.45, MAPE = 0.741%, and $R^2$ = 0.982 under cloudy conditions. Compared with PVELLM-DBN within the same study, RF-PVELLM-DBN reportedly improved $R^2$ by 2.86% in sunny weather, 7.0% in rainy weather, and 4.9% in cloudy weather. These values should be interpreted within the study's specific dataset, weather grouping, small-sample setup, and baseline configuration rather than as general benchmark evidence across PV | [79] |

(*continued on next page*)

**Table 5** (*continued*)

| Domain | Limitation of conventional ML | Main novelty of the LLM-enabled or hybrid approach | Reported within-study performance and caveats | Ref. |
|---|---|---|---|---|
| | | | forecasting studies. | |
| Solar | Existing PV forecasting models are often scenario-specific and lose robustness under diverse, few-shot, and zero-shot settings. | Uses Pegasus with prompt-learning, cosine-similarity cross-attention, and DQ-LoRA two-stage fine-tuning to cast PV forecasting as a text-generation task with stronger semantic and cross-modal alignment | Within the study's evaluation protocol, full-sample results on the Sanyo dataset were reported as RMSE/MAE/MSE/ $R^2 = 0.2960/0.1894/0.0876/0.9736$. The corresponding fixed Informer results were reported as 0.3203/0.1976/ 0.1026/0.9719. Under the same study settings, the few-shot and zero-shot results were reported as 0.4257/0.2789/ 0.1812/0.9453 and 0.5523/ 0.3993/0.3051/ 0.9108, respectively. These results suggest favourable performance within the authors' experimental setup, but they should be interpreted in the context of the specific Sanyo dataset, sampling conditions, baseline configuration, and the few-shot/zero-shot evaluation design used in that study. | [80] |
| Solar | Existing PV forecasting models have weak zero-shot transfer across stations and rely heavily on hyperparameter tuning. | Introduces StairLLM-Net, which uses a GPT-2 backbone, patch-based embedding, wavelet-enhanced multiscale inputs, and a Stair Decoder to support cross-station zero-shot PV forecasting without finetuning | The study evaluated the model across 12 PV stations organised into six zero-shot train–test settings. Within that study protocol, the model reported the lowest MSE in 5 of the 6 groups and the lowest MAE in 4 of the 6 groups among the baselines considered. The reported MAE values across the six groups were 3.800, 3.727, 3.844, 3.655, 3.710, and 4.928, while the corresponding MSE values were 67.94, 62.82, 68.25, 61.18, 63.70, and 102.03. These results indicate favourable zero-shot transfer performance within the authors' experimental design. Still, they should not be interpreted as directly comparable benchmark values beyond the specific PV stations, grouping strategy, horizons, and baseline settings used in that study. | [81] |
| Wind | Statistical, physical, and standard deep-learning models struggle with nonlinear, non-stationary wind power dynamics and often generalise poorly across sites and horizons | Introduces CMLLM, which converts SCADA data into text for frozen pre-trained LLM forecasting and adds a prior-knowledge prompt prefix with the dataset and statistical information | Within the study's evaluation protocol, CMLLM reported favourable eMAE/ eRMSE values across the three wind-farm datasets considered by the authors. For Dataset 1, the reported eMAE/ eRMSE values were 0.3246/ 0.4776 at 15 min and 1.7121/ 2.3493 at 12 h. For Dataset 2, the corresponding values were 0.2394/0.4629 and 1.3641/ 2.2311, while for Dataset 3 they were 0.3579/ 0.5628 and 1.6653/2.2791. These values were reported to be lower than those of the comparison models included in that study, but they should be interpreted within the specific wind-farm datasets, forecast horizons, metric definitions, preprocessing choices, and baseline configurations used by the authors. | [82] |
| Wind | Existing statistical, physics-based, and deep-learning models struggle with non- | Introduces STELLM, which combines seasonal-trend decomposition, spatio-temporal | Within the study's evaluation protocol, STELLM reported lower MAE/RMSE values than the baselines | [83] |

(continued on next page)

**Table 5** (*continued*)

| Domain | Limitation of conventional ML | Main novelty of the LLM-enabled or hybrid approach | Reported within-study performance and caveats | Ref. |
|---|---|---|---|---|
| | stationary wind series and often fail to capture both long-range temporal structure and spatial turbine dependencies. | prompting, autoregressive GPT-2 fine-tuning, and a localised spatial module for wind forecasting | considered by the authors across four public wind datasets. On DSWE1, the reported MAE/RMSE values were 0.7454/0.9610 at Horizon 1 and 2.1041/2.6978 at Horizon 12. On DSWE2, the corresponding values were 0.8109/1.1035 and 2.0282/2.4734. On NREL, they were 1.2685/1.7971 and 3.9789/5.0547, while on SDWPF they were 0.2678/0.3817 and 0.9360/1.2901. These results indicate favourable performance within the authors' experimental setting, but they should be interpreted in relation to the specific datasets, forecast horizons, metric definitions, preprocessing steps, and baseline configurations used in that study. | |
| Wind | Conventional statistical models have weak nonlinear learning capability and restrictive distribution assumptions, standard ML methods struggle with high-dimensional data and changing meteorological conditions, and existing DL models show limited generalisation to unseen wind farms, inflexible fixed horizon prediction, and weak robustness to mutation events in wind power | Proposes WindPromptCast, a hard-soft hybrid prompt-learning framework for multi-location multistep wind power forecasting. Its main novelty is reformulating wind power forecasting as a language modeling task through hard prompt generation, then adapting a frozen pretrained LLM using a gated soft-prompt adapter and a hard-soft prompt fusion mechanism to enable parameter-efficient zero-shot forecasting across unseen wind farms and varying prediction horizons | Reported within-study forecasting evidence was based on 14 wind farms. Within the authors' evaluation protocol, WindPromptCast reported average MAE/RMSE/MAX values of 11.670/18.827/316.92 MW. Seasonal MAE/RMSE values were reported as 11.5/18.6 MW in spring, 11.1/18.3 MW in summer, 10.8/17.9 MW in autumn, and 13.3/20.4 MW in winter. In the zero-shot setting, the model was trained on 2019 data from wind farms A–D for a 16-step horizon and tested on unseen wind farms H, I, J, and N in 2020. Under this setting, the study reported RMSE reductions of 12.03% relative to DRNet-BiLSTM and 13.64% relative to X-TCN at 16 steps, as well as an MAE reduction of up to 13.1% relative to X-TCN at 16 steps, 7.355 MW versus 8.517 MW. The study also reported favourable performance up to 30-step forecasts. These findings should be interpreted within the specific wind-farm selection, zero-shot split, forecast horizon, seasonal grouping, and baseline configuration used by the authors rather than as general benchmark evidence across wind-forecasting studies. | [84] |
| Wind | Existing wind-power forecasting methods often rely only on structured numerical data, underutilise the semantic understanding capability of LLMs, and lack task-specific cross-modal designs for jointly representing wind power output and numerical weather prediction inputs | Proposes M2WLLM, a multi-modal multi-task ultrashort-term wind power forecasting framework that combines tailored textual prompts with temporal numerical data through a Prompt Embedder, a Data Embedder, a Semantic Augmenter based on cross attention with pre-trained word embeddings, and LoRA-based fine-tuning of a GPT-2 backbone | Reported within-study forecasting evidence was based on three Chinese provincial datasets evaluated at 15 min, 1 h, 2 h, and 4 h horizons. On the Inner Mongolia dataset, M2WLLM reported MAE/RMSE values of 3.36/5.24, 6.59/9.57, 8.76/12.21, and 11.37/15.26 MW across the four horizons. On the Gansu dataset, the corresponding values were 6.68/10.74, 12.10/19.29, 15.57/23.56, and 18.13/26.66 MW. On the Yunnan dataset, they were 2.58/4.64, 4.48/7.40, 5.99/9.27, and 7.87/11.59 MW. The authors reported lower errors than the baselines included in their study, including LSTM, TCN, Transformer, Informer, Autoformer, | [85] |

(*continued on next page*)

**Table 5** (*continued*)

| Domain | Limitation of conventional ML | Main novelty of the LLM-enabled or hybrid approach | Reported within-study performance and caveats | Ref. |
|---|---|---|---|---|
| | | | Adaptive-GCN, CNN-BiLSTM-Att, and GPT4TS, and also reported favourable few-shot performance within their experimental protocol. These results should be interpreted in relation to the specific provincial datasets, forecast horizons, input modalities, baseline settings, and few-shot evaluation design used by the authors. | |
| Load | Statistical and standard deep-learning load forecasters struggle with nonlinear, noisy, and nonstationary demand patterns, and existing Transformer models remain limited for modern short-term load forecasting | Introduces LFLLM, a Llama 2-7B-based multi-voltage load forecasting model using parameter-efficient finetuning with layer freezing to adapt a pre-trained LLM to load time series and zero-shot transfer tasks | Reported within-study results on the AEMO dataset were RMSE/MAE/MAAPE/SMAPE = 253.649/180.062/0.053/0.027. On EWELD, the reported RMSE/MAE values were 26.737/18.312, while on REFIT the reported RMSE/MAE/MAAPE values were 409.939/217.022/0.415. The study also reported average zero-shot MAAPE/SMAPE values of 0.291/0.303 across datasets. These results were favourable relative to the baselines considered by the authors, but they should be interpreted within the study's specific datasets, forecast horizons, baseline configurations, zero-shot setting, and evaluation protocol rather than as directly comparable benchmark values across load-forecasting studies. | [86] |
| Load | Statistical feature-selection methods miss dynamic temporal dependencies, classic load forecasters generalise poorly, and Transformer models trained from scratch remain computationally demanding | Introduces EnergyGPT, a GPT2-based multi-energy load forecasting model with global and local dynamic feature selection plus frequency, temporal, and spatial adapters for efficient LLM fine-tuning | Within the study's evaluation protocol, EnergyGPT reported lower error than the baselines considered by the authors across the evaluated seasons and four load types using the Arizona State University dataset. As an example, the reported winter results were MAPE = 1.426, 2.573, 2.302, and 1.486; RMSE = 403.154, 411.298, 100.174, and 644.388 kW; and $R^2$ = 92.747, 98.844, 96.024, and 97.957 for electricity, cooling, heating, and combined loads, respectively. These values indicate favourable performance within the authors' experimental setting, but they should be interpreted in relation to the specific dataset, seasonal grouping, load categories, baseline configurations, and evaluation protocol used in that study rather than as directly comparable benchmark evidence across load-forecasting studies. | [87] |
| Load | Existing load forecasters cannot effectively use textual socio-demographic information. | Introduces LLMSA, which combines a socio-related adapter and a socially aware reprogramming layer to fuse socio-demographic text with load time series in an LLM-based forecasting framework | Within the study's evaluation protocol, LLMSA reported lower errors than the baselines considered by the authors, including PatchTST, iTransformer, TimeLLM, PatchTST wSoc, and BERT, on the Irish CER dataset. The study reported up to a 4.96% reduction in MSE. Example reported results include MSE/MAE/$R^2$ = 0.4728/0.3779/0.3527 for the over-age-55 household group and 0.4668/0.3728/0.3420 for unemployed or retired households. These findings suggest | [88] |

(*continued on next page*)

**Table 5** (*continued*)

| Domain | Limitation of conventional ML | Main novelty of the LLM-enabled or hybrid approach | Reported within-study performance and caveats | Ref. |
|---|---|---|---|---|
| | | | favourable performance within the authors' socio-demographic load-forecasting setup, but they should be interpreted in relation to the specific Irish CER dataset, household grouping strategy, baseline configurations, and evaluation protocol used in that study. | |
| Load | Conventional machine-learning load forecasters require substantial historical target data, transfer-learning approaches depend on sufficiently similar source-domain load data and can suffer from negative transfer, and existing graph-based spatial models usually lack inductive capability when the number of buildings changes between training and deployment | Proposes BlackInter, a memory-efficient black-box tuning an inductive adapter for pre-trained LLMs in building-level load forecasting. Its main novelty is the combination of an input embedding layer, an inductive spatial embedding layer, a feature extraction layer based on dynamic mode decomposition, and a Mix-Adam optimisation method that updates the adapter using LLM inference results without requiring access to the backbone LLM architecture or parameters | Reported within-study forecasting evidence was based on few-shot and zero-shot building-level load-forecasting settings. Under the authors' evaluation protocol, BlackInter reportedly reduced prediction error to 42.88% of the best-performing alternative method in the few-shot setting and to 69.44% in the zero-shot setting. The authors also reported lower errors than the comparison methods included in their study and observed that the model remained effective when the number of buildings differed between training and testing. These findings suggest favourable inductive transfer performance within the study design, but they should be interpreted in relation to the specific building datasets, the few-shot and zero-shot splits, the comparison methods, and the evaluation protocol used by the authors. | [89] |
| Load | Existing multi-energy forecasting methods tend to treat electrical, heat, and gas loads as homogeneous inputs, which prevents accurate modelling of their distinct physical evolution patterns; they also primarily minimise statistical error while neglecting asymmetric decision costs and disruptive risks in downstream power distribution network operation. | Proposes RSynLLM, a risk-aware routing mixture of-experts large language model for large-scale multi-energy load forecasting. Its main novelty is the combination of tailored electrical, heat, and gas load experts within a routing MoE encoder, customised prompt instructions for Vicuna-7B, marker alignment for continuous prediction, and a risk-aware loss that combines asymmetric error penalisation with CVaR to better align forecasts with operational decision-making objectives. | Reported within-study forecasting evidence was based on a large-scale multi-energy power distribution network benchmark. Within the authors' evaluation protocol, RSynLLM reported MAE/RMSE/MAPE values of 1.64/2.69/15.29 at the 3 h horizon, 1.79/2.83/17.81 at the 6 h horizon, and 1.85/2.85/18.06 at the 12 h horizon. The authors reported lower MAE and RMSE than the baselines considered in their study, including GRU, GCN, T-GCN, AGCRN, MegaCRN, MPGTN, UrbanGPT, and STCInterLLM, across the reported horizons. They also reported the lowest MAPE at 3 h and 12 h, and the near-lowest at 6 h, within their comparison setting. These findings suggest favourable performance within the study's multi-energy forecasting setup, but they should be interpreted in relation to the specific PDN benchmark, forecast horizons, load types, risk-aware objective, baseline configurations, and evaluation protocol used by the authors. | [90] |

**Note:** The numerical values in this table are reported results from the original studies. They are not normalised across datasets, forecast horizons, temporal resolutions, baseline configurations, or evaluation protocols. Accordingly, they should be interpreted as study-specific evidence rather than as a unified benchmark of LLM-enabled forecasting performance.

applications are most convincing when embedded in hybrid pipelines that combine time-series forecasting, weather information, and domain knowledge, especially in small-sample, few-shot, or zero-shot settings. However, the review also shows that some reported advantages come from multimodal integration, transfer learning, prompt design, or knowledge-enhanced hybridisation rather than from language-centric reasoning alone. Table 5, therefore, presents solar studies at the paper level, including their main novelty and quantitative evidence, to make this distinction explicit.

### 4.2. Wind energy forecasting

Wind energy forecasting has also seen a growing number of LLM-enabled and foundation model-based approaches, particularly for problems involving strong temporal variability, non-stationarity, and cross-site generalisation. In contrast to earlier narrative summaries that focused mainly on potential contextual benefits, the revised synthesis in Table 5 shows that many recent wind studies now report direct forecasting evidence with explicit numerical comparisons against strong deep-learning baselines.

Several representative studies have adapted pre-trained language models or language-model-inspired architectures to spatio-temporal wind forecasting, but their reported gains should be interpreted as study-specific results rather than general evidence of LLM superiority. BERT4ST fine-tuned a pre-trained BERT model using patch-based encoding and redesigned temporal and spatial embeddings, reporting the lowest MAE/RMSE among the models evaluated at the tested horizons on the Chuzhou and NorthSea datasets [78]. CMLLM converted SCADA data into textual representations and combined a frozen pre-trained LLM with prior-knowledge prompts, reporting lower eMAE/eRMSE than the selected comparison models on three wind-farm datasets [82]. STELLM combined seasonal-trend decomposition, spatio-temporal prompting, autoregressive GPT-2 fine-tuning, and a localised spatial module, and reported improved performance relative to the baselines considered on four public wind datasets [83]. However, these results do not by themselves establish that LLM-based models are universally superior to conventional ML or specialised time-series forecasters, because the studies differ in datasets, forecast horizons, preprocessing pipelines, baseline selection, and the extent of hyperparameter tuning reported for competing methods.

The strongest evidence in this domain comes from studies that explicitly test generalisation and zero-shot transfer across wind farms. Wind Prompt Cast reformulated multi-location multi-step wind forecasting as a language-modelling problem and used hybrid prompt learning to adapt a frozen pre-trained LLM. Across 14 wind farms, it reported average MAE/RMSE/MAX of 11.670/18.827/316.92 MW and reduced RMSE by 12.03% relative to DRNet-BiLSTM and by 13.64% relative to X-TCN in a zero-shot setting on unseen farms [84]. Similarly, M2WLLM combined textual prompts with numerical weather prediction and temporal data via a cross-modal architecture, achieving consistently strong results across multiple Chinese provincial datasets and outperforming LSTM, TCN, Transformer, Informer, AutoFormer, Adaptive-GCN, CNN-BiLSTM-Att, and GPT4TS [85].

These results indicate that the wind literature has moved beyond purely conceptual claims about contextual reasoning and now includes substantial direct forecasting evidence. In most cases, performance improvements arise from hybrid architectural choices such as prompt engineering, spatio-temporal adaptation, cross-modal embedding, parameter-efficient fine-tuning, or the conversion of structured turbine data into forms that pre-trained models can exploit more effectively. Table 5 synthesises wind studies in terms of both methodological novelty and quantitative performance, rather than only describing general capabilities.

### 4.3. Load demand forecasting

Load demand forecasting is one domain where LLM-enabled methods have been more frequently evaluated as direct forecasting components, particularly when demand time series are combined with contextual, socio-demographic, operational, or heterogeneous information sources. As shown in Table 5, recent studies report within-study improvements for short-term, building-level, and multi-energy load prediction, suggesting that LLM-enabled approaches may be useful beyond auxiliary interpretation roles in selected settings. However, these findings should be interpreted cautiously because the reported gains depend on dataset characteristics, forecast horizons, baseline selection, hyperparameter tuning, preprocessing choices, and whether the language-model component was isolated through ablation experiments. Therefore, the current evidence supports a context-dependent role for LLM-enabled load forecasting rather than a general claim of superiority over conventional ML or specialised time-series baselines.

A first group of studies adapts pre-trained large language models to structured load-time series using parameter-efficient tuning. LFLLM used a Llama 2-7B backbone with layer freezing for multi-voltage short-term load forecasting and reported lower errors than the selected comparison models on several datasets, including AEMO, EWELD, and REFIT, while also reporting favourable average zero-shot MAAPE/SMAPE across datasets [86]. EnergyGPT likewise used a GPT-2-based architecture with global and local dynamic feature selection and multiple lightweight adapters, reporting lower errors than the baselines considered across seasons and four load types in the Arizona State University dataset [87]. However, these results should be interpreted as within-study findings rather than definitive evidence that LLM-based models are generally superior to conventional ML or specialised time-series forecasting baselines. The strength of the evidence depends on the competitiveness of the selected baselines, the transparency of hyperparameter tuning, the consistency of data preprocessing, and whether equivalent training and evaluation budgets were applied across models.

A second group of studies shows that LLM-enabled frameworks may be particularly valuable when non-traditional contextual or heterogeneous inputs are incorporated. LLMSA fused socio-demographic text with load-time series via a socio-related adapter and a socially aware reprogramming layer, outperforming PatchTST, iTransformer, TimeLLM, PatchTST wSoc, and BERT on the Irish CER dataset, achieving up to 4.96% reduction in MSE [88]. BlackInter addressed building-level few-shot and zero-shot forecasting under changing numbers of buildings and reported that its prediction error fell to 42.88% of the best alternative method in the few-shot setting and 69.44% in the zero-shot setting [89]. RSynLLM extended the scope further to risk-aware multi-energy forecasting, achieving the best MAE and RMSE across all reported horizons compared with GRU, GCN, T-GCN, AGCRN, MegaCRN, MPGTN, UrbanGPT, and STCInterLLM, while also incorporating asymmetric decision risk via a CVaR-aware objective [90].

Overall, the reviewed studies suggest that load demand forecasting is a promising empirical setting for LLM-enabled methods, particularly in short-term, building-level, multi-energy, few-shot, and zero-shot forecasting contexts. However, the current evidence should be interpreted as within-study performance reporting rather than definitive proof that LLM-enabled models are generally superior to conventional ML or specialised time-series forecasting baselines. Reported improvements appear to depend on multiple factors, including architectural adaptation, parameter-efficient fine-tuning, contextual or heterogeneous data fusion, baseline selection, hyperparameter-tuning transparency, and the availability of comparable training and evaluation budgets. Therefore, the most defensible interpretation is that LLM-enabled methods may serve as useful components within hybrid load-forecasting pipelines when numerical demand series are complemented by broader contextual information, while conventional ML, statistical models, and specialised time-series architectures remain essential and potentially competitive baselines.

### 4.4. Cross-cutting issues in forecasting models

Beyond individual domains, cross-cutting issues such as forecast horizons, robustness, and interpretability shape the broader adoption of LLM-enabled and related forecasting approaches. While solar, wind, and load demand forecasting highlight domain-specific benefits of LLMs, several cross-cutting challenges and opportunities apply across all renewable energy applications. These issues transcend individual domains and influence the broader effectiveness of forecasting models in practice. Specifically, three dimensions forecast horizon suitability, robustness under varying weather conditions, and interpretability are critical for evaluating the reliability, scalability, and trustworthiness of LLM-based systems. Addressing these dimensions helps assess when LLM-enabled approaches may offer advantages over conventional ML, and when those advantages remain limited or context-dependent. Because cross-cutting challenges are conceptually distinct yet equally critical, we structure this section into sub-subsections. This ensures each issue is treated with sufficient depth while maintaining clarity.

#### 4.4.1. Forecast horizon suitability of selected models

Forecast horizon is a key practical consideration when selecting a model for renewable energy prediction. Short-term forecasts, spanning minutes to hours ahead, are essential for real-time grid operations and dispatch optimisation, whereas medium-term forecasts, spanning days to weeks, and long-term forecasts, spanning months to years, support maintenance scheduling, market planning, and investment decisions [91,92].

Traditional machine learning models such as Random Forests and Gradient Boosting Machines often perform well for short-to medium-term forecasting when the inputs are structured, and temporal dependencies are relatively limited. However, their performance commonly degrades over longer horizons because they are less effective at modelling complex temporal dynamics and long-range dependence [93]. Recurrent sequence models such as Long Short-Term Memory (LSTM) networks are better suited to capturing temporal structure and can therefore remain competitive for medium-term forecasting, although their effectiveness depends strongly on data quality and sequence length [94].

Transformer-based forecasters and time-series foundation models (TSFMs), by contrast, are designed to capture long-range temporal dependence more effectively and are often better suited to multi-horizon forecasting, including medium- and long-term settings, albeit with higher computational cost [29,95]. Language-centric LLMs should be distinguished from these numerical forecasting models. Their role is most relevant when forecasting also depends on textual or contextual information, such as maintenance logs, policy updates, market signals, or operator reports, in which case they are better suited to LLM-enabled hybrid pipelines rather than treated as generic substitutes for time-series forecasters. CNNs, particularly when applied to spatially rich inputs such as satellite imagery or two-dimensional representations of time-series data, are effective for short-term and event-driven forecasting. Still, they usually require hybridisation with temporal models for extended forecasting horizons [96].

A detailed summary of the relative suitability of selected model classes for different forecasting horizons is provided in Supplementary Table S1. To maintain taxonomy clarity, transformer forecasters and time-series foundation models are separated from language-centric LLM-enabled hybrid workflows.

#### 4.4.2. Model generalisation and robustness under varying weather conditions

An essential aspect of forecasting model evaluation is the ability to generalise beyond the conditions represented in the training data. Generalisation refers to whether a model maintains acceptable predictive performance under unseen or poorly represented environmental conditions, including storms, sustained cloud cover, abrupt wind-speed shifts, and temperature anomalies. Traditional ML algorithms, such as Random Forest and Gradient Boosting, can perform well under moderately varying conditions, but their performance often degrades when rare or extreme events are not adequately represented in the training set. Deep learning models such as LSTM and CNN architectures generally adapt better to complex nonlinear weather patterns, though they may still be prone to overfitting when the training data lacks sufficient variability.

Transformer-based forecasters and time-series foundation models (TSFMs) have shown strong potential for improved generalisation because they can capture long-range temporal dependencies and, in some cases, benefit from large-scale pretraining on diverse numerical sequences. Their robustness under unusual weather conditions, however, still depends on the breadth, quality, and representativeness of the data used during pretraining and adaptation. Language-centric LLMs should be distinguished from these numerical forecasting models. In forecasting workflows, LLMs are more appropriately used within hybrid pipelines that incorporate textual or contextual information such as meteorological reports, operational logs, or policy-related signals, where they may support reasoning, retrieval, and decision-oriented interpretation rather than acting as standalone weather forecasters. Hybrid approaches that combine data-driven models with physics-based simulations can further improve robustness by embedding physical constraints that remain valid under extreme conditions. A detailed comparison of the relative generalisation capabilities of the reviewed model classes and their suitability under varying or extreme weather conditions is provided in Supplementary Table S2.

#### 4.4.3. Interpretability of reviewed models

In renewable energy forecasting, interpretability is essential for building stakeholder trust, supporting operational decision-making, and satisfying regulatory expectations. Forecast outputs must therefore be translated into explanations that practitioners can understand and act upon. Although general explainable artificial intelligence (XAI) techniques such as LIME and SHAP can be applied across multiple model families, the most suitable interpretability method depends on the underlying model architecture and the type of reasoning it performs.

A detailed mapping of representative interpretability techniques for the forecasting model classes discussed in this review is provided in Supplementary Table S3. To maintain taxonomy clarity, conventional tree-based models, recurrent neural networks, convolutional models, transformer forecasters or time-series foundation models (TSFMs), LLM-enabled hybrid workflows, and hybrid physics-informed approaches are treated as distinct categories. The table maps each model class to suitable XAI methods and highlights the type of insight each pairing can provide, including feature attribution, temporal relevance, spatial saliency, token- or embedding-level attention, and adherence to physical constraints. These explanations can help practitioners understand model behaviour, identify the drivers of forecast outcomes, and assess the reliability and trustworthiness of predictions in renewable energy applications.

These cross-cutting interpretability issues also illustrate both the potential and the limitations of current forecasting approaches. To better understand the broader functions that may explain reported improvements across solar, wind, and load demand applications, the next section examines the cross-domain roles of LLM-enabled and hybrid forecasting pipelines in renewable energy systems.

## 5. Cross-domain capabilities of LLM-enabled and hybrid forecasting pipelines in renewable energy

Building on the domain-specific evidence discussed above, this section synthesises the cross-domain roles of LLM-enabled and hybrid forecasting pipelines in renewable energy systems. Rather than treating all transformer-based forecasting capabilities as intrinsic to LLMs, we distinguish language-centric contributions from those of specialised numerical, spatiotemporal, graph-based, and uncertainty-aware forecasting models. Across wind, solar, and load-demand applications, LLMs mainly contribute through text interpretation, retrieval over unstructured evidence, contextual reasoning, explanation generation, and operator-facing decision support. At the same time, dedicated forecasting components typically handle the primary numerical forecasting task. Accordingly, this section focuses on recurring cross-domain functions, including multimodal data integration, predictive maintenance

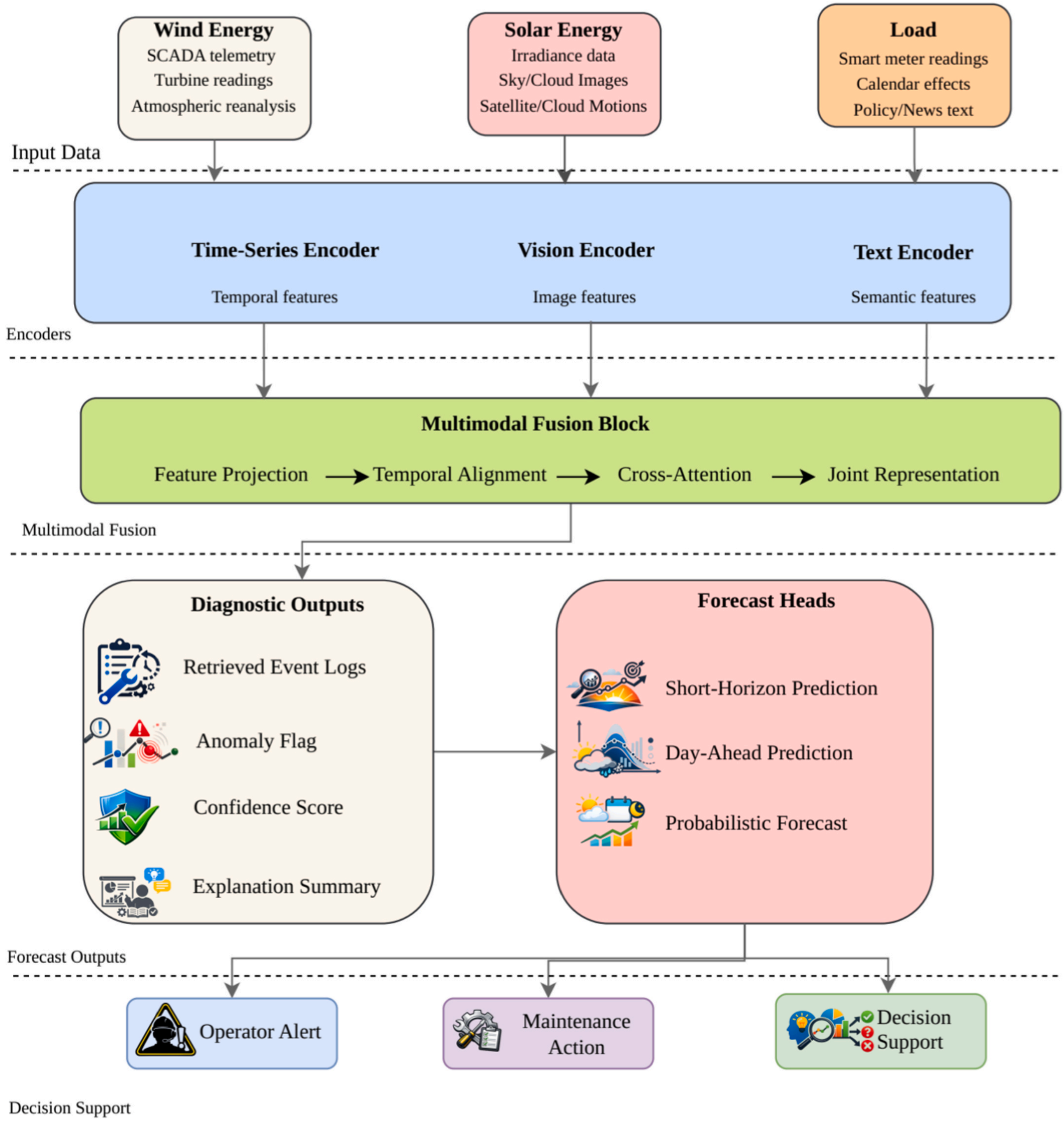


**Fig. 6.** LLM-enabled multimodal data integration pipeline for renewable energy forecasting (Reproduced from Refs. [84,100]).

support, scenario-based decision analysis, and data-quality enhancement.

### 5.1. *Multimodal data integration in LLM-enabled and hybrid pipelines*

A key role of LLM-enabled forecasting pipelines is to integrate structured energy data with unstructured contextual information. Renewable energy forecasting increasingly draws on heterogeneous sources, including SCADA telemetry, turbine and inverter measurements, irradiance and weather records, maintenance logs, market documents, and operator notes [97,98]. In these settings, language-centric LLMs are most useful for retrieving, interpreting, and summarising textual or semi-structured evidence. At the same time, specialised numerical, vision, or spatiotemporal models usually remain responsible for processing physical measurements and generating forecasts.

In wind, solar, and load-demand applications, multimodal pipelines can combine time-series data with weather narratives, maintenance records, remote-sensing products, socio-economic indicators, and market or policy signals [8,99]. The main benefit, therefore, is not that LLMs

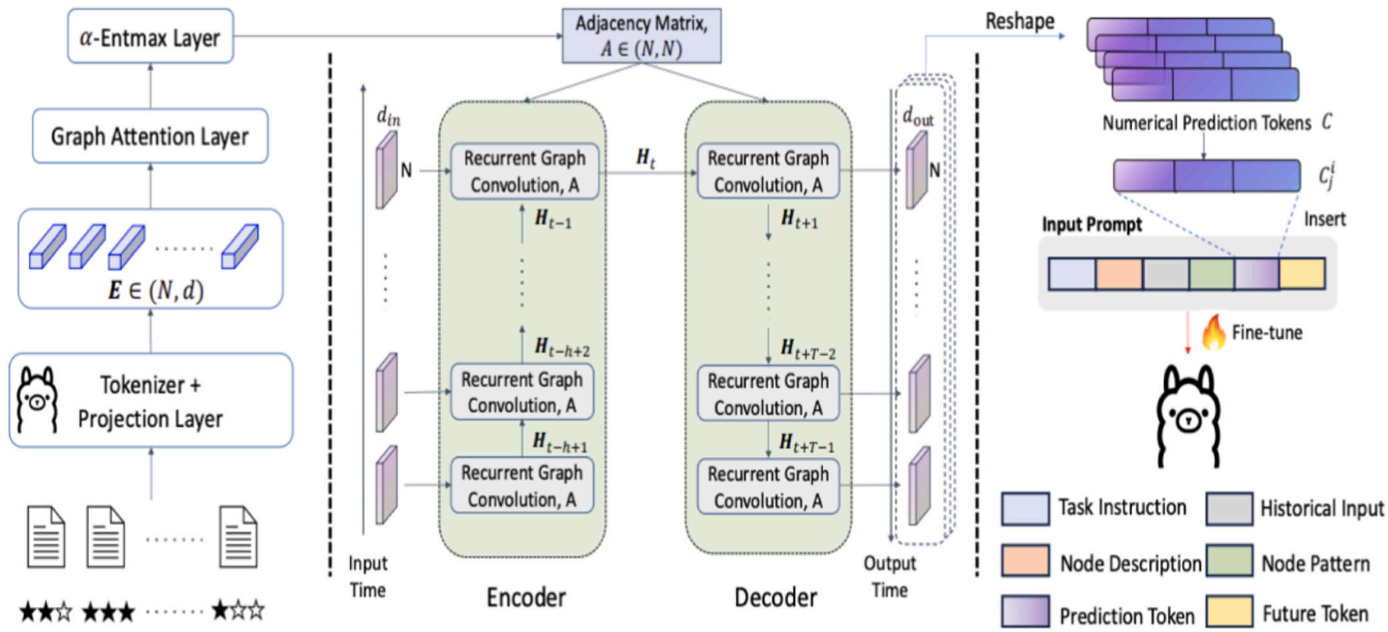


**Fig. 7.** Architecture of a hybrid Spatiotemporal Graph–LLM framework for energy forecasting [108].

replace conventional forecasting models, but that they can support context-aware fusion, anomaly interpretation, explanation generation, and operator-facing decision support. Fig. 6 illustrates this general architecture, where modality-specific encoders process numerical, visual, and textual inputs before a fusion layer supports forecasting, diagnostic outputs, uncertainty information, and explanation summaries.

### *5.2. Temporal and spatial pattern learning in hybrid spatiotemporal and Graph-LLM frameworks*

Accurate renewable energy forecasting depends strongly on the ability to model complex temporal and spatial structure. In practice, this capability is usually delivered by spatiotemporal forecasting architectures, graph-based encoders, and other specialised sequence models rather than by language-centric LLMs alone. Transformer-based timeseries models and graph-aware hybrid systems can capture long-range temporal dependence, non-linear interactions, and distributed spatial relationships in renewable energy data [98,101]. Where LLMs are incorporated, their role is generally to add contextual reasoning, prompt-based conditioning, textual retrieval, or operator-facing explanation on top of these spatiotemporal representations.

This distinction is important because many renewable energy forecasting tasks involve nonstationary, highly variable temporal processes. Compared with traditional models that rely on fixed time windows or struggle with temporal drift, spatiotemporal and transformer-based forecasting models can better represent diurnal cycles, seasonal weather patterns, regime changes, and delayed cross-variable effects [102]. In wind forecasting, for example, hybrid spatiotemporal pipelines may model turbulence, ramp events, and topographic influences across sites and altitudes [103,104]. In solar forecasting, temporal dependencies may include irradiance variation, lagged temperature effects, cloud-cover dynamics, and particulate accumulation [105,106]. In load forecasting, the relevant temporal structure may involve demand shifts associated with holidays, economic activity, mobility changes, or other behaviourally mediated events.

Spatial pattern learning is likewise more appropriately attributed to graph-based and multimodal forecasting architectures than to LLMs alone. With suitable geospatial encoders and spatial embeddings, hybrid frameworks can model wind shear across turbine clusters, cloud shadow propagation over solar farms, and geographically varying demand spikes across urban grids [98]. When combined with LLM modules, the resulting system can leverage these numerical and spatial representations alongside textual context, policy constraints, and domain instructions to support better interpretation and decision-making. Thus, the overall spatiotemporal capability should be viewed as a property of the hybrid forecasting framework, with the LLM playing a complementary role rather than acting as a universal forecaster in isolation.

Recent work on hybrid graph-LLM architectures illustrates this design principle. As shown in Fig. 7, graph-based encoders can capture node-level spatial dependencies, such as relationships among turbines, solar sites, or grid regions. In contrast, recurrent or graph convolutional layers capture temporal evolution across those nodes. These representations may then be converted into sequence-like tokens and combined with textual context, domain knowledge, or task instructions for downstream reasoning and decision support [107]. This type of fusion can improve generalisation and contextual interpretation, but the underlying spatial and temporal learning should still be attributed primarily to the spatiotemporal forecasting stack.

### *5.3. Predictive maintenance and anomaly detection in LLM-enabled monitoring workflows*

Another important cross-domain role of LLM-enabled workflows lies in predictive maintenance and anomaly detection, particularly when structured sensor streams must be interpreted together with unstructured maintenance evidence. In renewable energy systems, unexpected failures can cause costly downtime, repair expenditure, and lost generation. Traditional machine-learning approaches often depend on domain-specific feature engineering and fault-labelled datasets. By contrast, hybrid monitoring systems that include LLM modules can improve scalability by combining numerical anomaly detection outputs with maintenance logs, manuals, event reports, and operator observations [109].

In wind energy systems, monitoring pipelines may combine SCADA measurements such as rotor speed, gearbox temperature, and vibration readings with historical failure reports and technical manuals [110]. In such settings, the LLM can help interpret textual maintenance evidence, relate sensor anomalies to past failure descriptions, and support the generation of maintenance recommendations. Similarly, in solar installations, hybrid monitoring systems may combine inverter fault logs,


**Fig. 8.** Workflow diagram of an LLM-enabled predictive-maintenance methodology.

environmental data, and panel telemetry to detect potential degradation mechanisms, such as microcracks, shading losses, or soiling [111]. Specialised monitoring models generally produce the forecasting or anomaly signal itself, while the LLM strengthens contextual explanation and operational reasoning.

In load demand systems, LLM-enabled monitoring may also help interpret irregular consumption behaviour related to faulty smart meters, cyber incidents, or unusual behavioural shifts [112]. Because language-centric models can process both temporal summaries and textual reports, they are well-suited to supporting root-cause interpretation and operator-facing explanation within broader anomaly-detection workflows. This interpretability is particularly valuable when monitoring systems must be integrated into real-time decision environments.

Fig. 8 presents an implementation-oriented predictive-maintenance pipeline for renewable energy systems. The workflow begins with heterogeneous data ingestion, including SCADA telemetry, alarm and event logs, maintenance records, inspection notes, manuals, and operating context, followed by parsing, standardisation, and quality control. The processed information is then organised into complementary knowledge and storage layers, including a time-series database for operational history and a vector store with a knowledge graph for manuals, historical incidents, maintenance procedures, and fault relationships. On top of this foundation, retrieval and feature-construction modules extract anomaly windows, residual trends, similar past incidents, and relevant procedures, which are passed to a hybrid predictive-maintenance engine that combines baseline fault models with retrieval-augmented large-language-model reasoning. This design enables evidence-grounded fault diagnosis, root-cause explanation, maintenance action recommendations, and confidence-aware decision support. At the same time, human-in-the-loop validation and safety fallback mechanisms improve robustness for practical deployment.

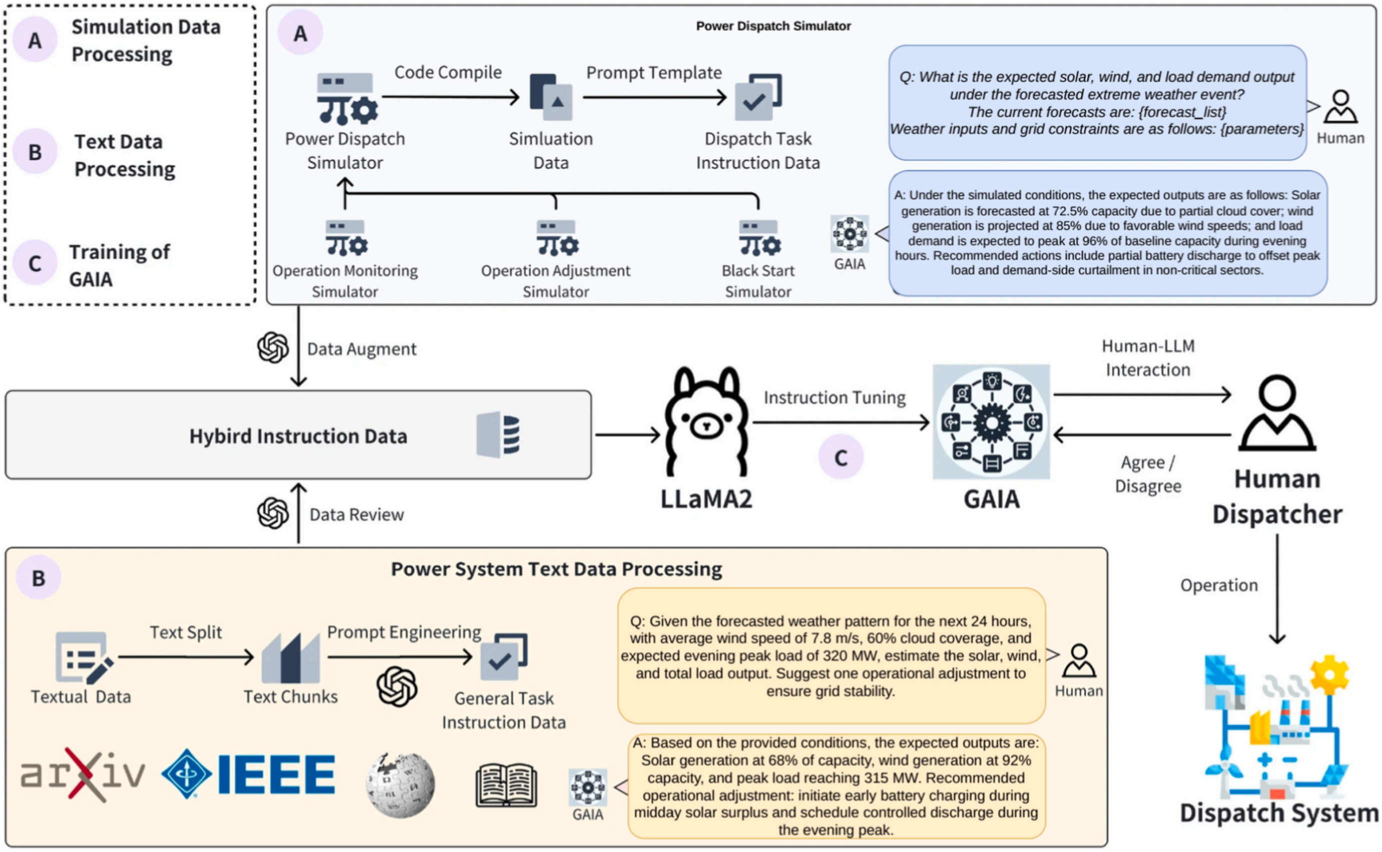


**Fig. 9.** LLM-enabled framework for renewable energy scenario simulation and decision optimisation [117].

**Table 6**
Cross-domain roles of LLM-enabled and hybrid forecasting pipelines in renewable energy systems.

| Capability in the hybrid pipeline | Cross-domain role | Wind Energy | Solar Energy | Load Demand |
|---|---|---|---|---|
| Multimodal data integration | Combines structured sensing, text, market, and environmental context within a unified workflow | SCADA logs, turbine telemetry, maintenance records, and weather narratives | Irradiance records, atmospheric reports, maintenance logs, remote-sensing context | Smart meters, socioeconomic indicators, market conditions, policy documents |
| Temporal and spatial pattern learning | Captures temporal dependence and spatial variability through spatiotemporal and graph-based forecasting components, with LLMs adding contextual reasoning | Turbulence patterns, wind ramps, terrain effects | Seasonal irradiance cycles, lagged temperature effects, and cloud dynamics | Diurnal cycles, peak load shifts, holiday or event-driven demand |
| Predictive maintenance and anomaly detection | Combines sensor-based monitoring with textual evidence for fault interpretation and maintenance planning | Rotor or gearbox anomalies, vibration-based faults, and maintenance reports | Inverter degradation, shading or soiling signals, operational logs | Smart-meter anomalies, irregular usage patterns, and incident reports |
| Scenario modelling and decision support | Supports what-if analysis under uncertainty using forecasting and optimisation outputs, plus operator-facing interpretation | Grid balancing during storms or ramp events | Planning under extreme weather or policy changes | Dynamic pricing strategies and demand-side management under shocks |
| Data quality enhancement | Improves dataset consistency, completeness, and interpretability through harmonisation, parsing, and context-aware preprocessing | Harmonises SCADA formats; flags missing or inconsistent telemetry | Cleans noisy irradiance or sensor logs; supports missing-data handling | Detects inconsistencies in consumption data; improves integration of smart meter and IoT sources |

### *5.4. Scenario modelling and decision optimisation in LLM-enabled energy management*

LLM-enabled workflows can also support decision-making through scenario-based analysis, especially when operators need to compare multiple possible futures under uncertainty about weather, markets, policies, or infrastructure. Renewable energy systems operate under highly variable conditions, including climatic variability, grid constraints, price fluctuations, and regulatory interventions. In such settings, scenario modelling is usually driven by forecasting, optimisation, and uncertainty-aware numerical components, while LLM modules help formulate queries, interpret scenario assumptions, explain outcomes, and communicate recommendations to operators [113,114].

In wind and solar forecasting, hybrid decision-support systems may simulate high-impact events such as storms, prolonged overcast periods, curtailment conditions, or storage constraints. These systems can then evaluate alternative operational responses, including storage dispatch, load curtailment, reserve planning, or market participation strategies. When language-centric models are included, they can help translate structured inputs and scenario outputs into human-readable recommendations, policy-aware reasoning, and operator-facing summaries, rather than serving as the sole engine of probabilistic forecasting.

**Table 7**
Addressing data quality challenges in renewable energy datasets using LLM-enabled preprocessing and hybrid data-engineering workflows.

| Ref. | Challenge | LLM-enabled solution | Expected impact |
|---|---|---|---|
| [123] | Missing values | Context-aware support for imputation using temporal, spatial, and cross-variable relationships learned from multimodal data | Preserves data continuity and improves training stability |
| [124] | Inconsistencies across sources | Automatic harmonisation of variable names, units, and formats through language-aware metadata processing | Reduces preprocessing errors and improves dataset integration |
| [125] | Bias in datasets | Bias identification through statistical and semantic inspection of data distributions, with support for targeted augmentation or rebalancing | Improves the representativeness and reliability of forecasts |
| [126] | Noisy or unstructured inputs | LLM-based parsing, cleaning, and entity extraction from free-text logs and irregular formats | Improves signal-to-noise ratio and supports more reliable downstream forecasting |

For load-demand applications, LLM-enabled scenario analysis can support planning for pricing interventions, weather shocks, changes in consumer behaviour, or regulatory adjustments [115]. For example, a hybrid framework may evaluate how a time-of-use tariff affects evening peaks or how an extreme heatwave alters regional demand profiles. When combined with forecasting, optimisation, and causal analysis modules, LLMs can help interpret the implications of these scenarios and present the results in a form that supports demand-side management and resilient grid operation [116].

More generally, scenario modelling in renewable energy requires the simulation of alternative futures under varying weather, operational, market, and policy conditions. Hybrid pipelines that combine historical generation data, meteorological inputs, operational logs, and optimisation constraints can generate multi-variable forecasts under different assumptions. Within these systems, LLMs are most useful as contextual reasoning and communication layers that help operators compare alternatives, understand trade-offs, and align forecasting outputs with real-world operational decisions.

Fig. 9 shows a representative LLM-enabled scenario simulation and decision optimisation workflow. Meteorological forecasts, historical generation data, and grid constraints are combined to generate alternative solar, wind, and load demand scenarios. The forecasting and optimisation layers provide the quantitative basis for scenario evaluation, while the LLM supports human-in-the-loop interpretation and recommendation generation. This design enhances decision-making efficiency by enabling operators to explore multiple what-if scenarios without altering physical grid configurations.

Table 6 provides a cross-domain summary of how LLM-enabled and hybrid forecasting pipelines support renewable energy systems in wind, solar, and load demand applications. The table highlights the main roles these pipelines can play, including multimodal data integration, temporal and spatial pattern learning, predictive maintenance and anomaly detection, scenario modelling and decision support, and data quality enhancement. It also makes clear that, in most cases, LLMs contribute mainly through contextual reasoning, text interpretation, and operator support. At the same time, specialised forecasting components handle the main numerical and spatiotemporal forecasting tasks.

### 5.5. *Mitigating data quality issues in renewable energy datasets*

A further role of LLM-enabled forecasting pipelines is to support data quality improvement in real-world renewable energy datasets. These datasets often contain missing values, inconsistent labels, heterogeneous metadata, and sampling biases. In such cases, LLM modules can contribute through natural-language understanding, metadata harmonisation, contextual parsing, and knowledge-guided preprocessing, while the actual numerical correction or imputation may still rely on dedicated statistical or forecasting components [69,118].

First, LLM-enabled workflows can assist context-aware imputation by recognising patterns and relationships across multimodal sources such as SCADA logs, weather records, and market data. Second, their language-processing capability can help harmonise metadata, variable names, and format conventions across diverse datasets, reducing integration inconsistencies [119–121]. Third, hybrid data-quality pipelines may use LLM-guided anomaly detection and bias analysis, supported by historical records and domain ontologies, to identify systematic errors or under-represented operating regimes that warrant targeted correction or augmentation [122]. In this sense, the overall framework does not merely consume existing data but can also support preprocessing steps that improve dataset reliability before downstream forecasting.

To illustrate these roles, Table 7 summarises common data-quality challenges in renewable energy datasets, the corresponding LLM-enabled solutions, and their expected impact on forecasting reliability. The table is intended to show how language-centric components can strengthen preprocessing and interpretation within broader data-engineering pipelines.

## 6. Challenges, future directions, and implementation guidance

### 6.1. *Technical challenges*

This section is organised into separate sub-subsections because the principal technical barriers, including data quality, interpretability, computational complexity, and hallucination risk with uncertainty evaluation, are conceptually distinct and require dedicated treatment. Presenting them individually helps clarify their specific implications for the reliable deployment of LLM-enabled renewable energy forecasting systems while maintaining the clarity and readability of the review.

#### 6.1.1. *Data quality*

One of the primary challenges in applying LLMs is ensuring high-quality, unbiased, and representative datasets for their training and validation. Forecasting for Renewable Energy involves diverse data sources such as meteorological reports, IoT sensor readings, satellite images, and grid operational data. Inconsistencies, missing data and biases in historical records can challenge LLM predictions [127,128]. In addition, forecasting models may experience uncertainty due to data gaps caused by sensor malfunctions, weather anomalies, or under-reported consumption trends.

Moreover, merging heterogeneous data from various sources requires comprehensive data preprocessing to ensure datasets are organised and cleaned [129]. These take the form of missing values, anomalies, and differences in data formats, all of which will become more frequent and complex, and for which there is a goal to improve the efficiency of ML pipelines [130]. Cross-border energy cooperation, especially, needs standards to address discrepancies in data collection across regions.

Additionally, acquiring real-time information and synchronising data across diverse energy actors (power plants, grid operators, policy-makers, etc.) introduces logistical hurdles that require secure, scalable data sharing [131]. Integrating smart contracts and payment channels would enhance the efficiency and transparency of energy transactions [132]. In addition, continuous monitoring of data and automated data quality assessment techniques, such as anomaly detection models and

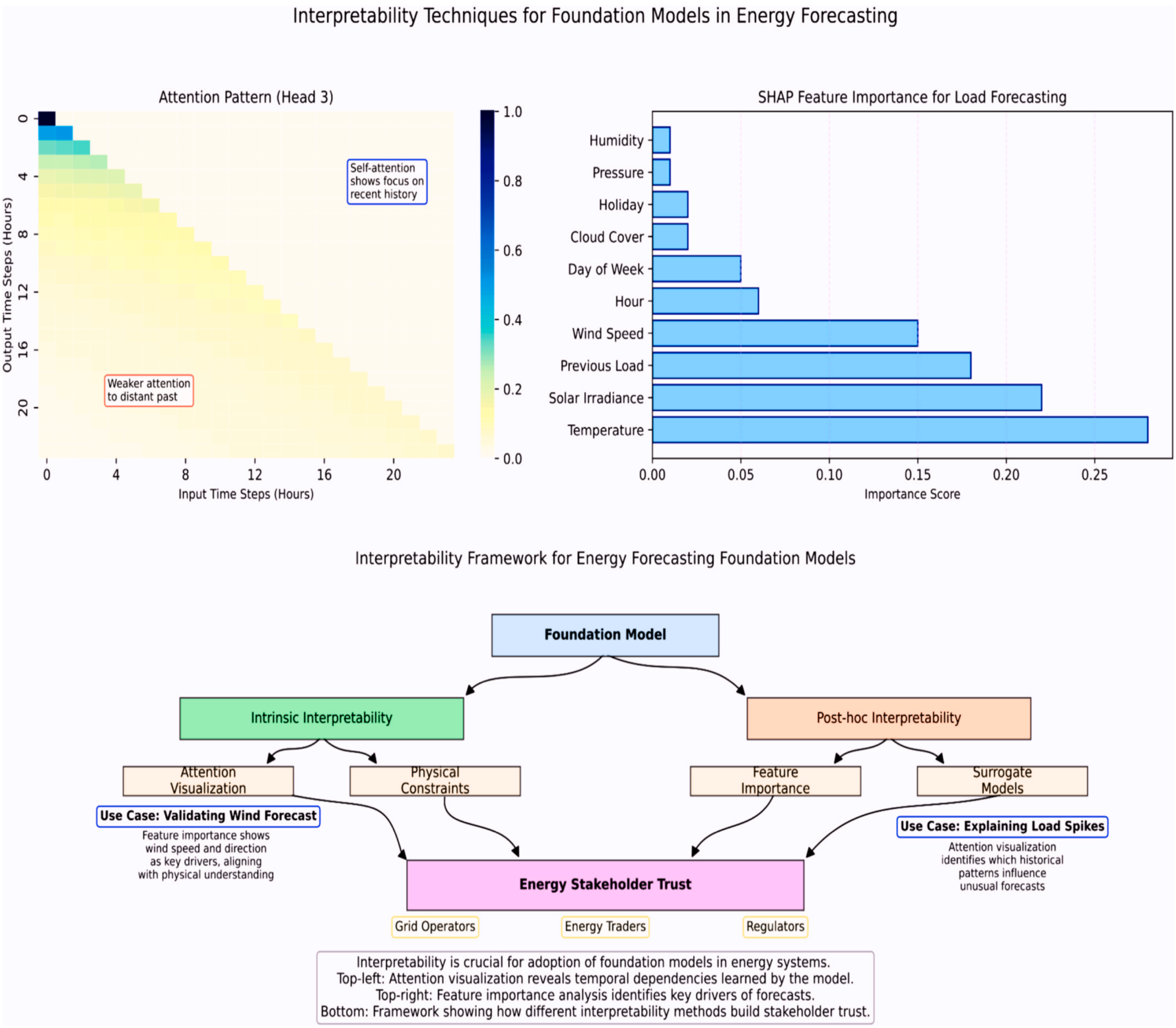


**Fig. 10.** Interpretability techniques for LLM-based forecasting in renewable energy [146].

**Table 8**
Operational interpretability mechanisms in the proposed LLM-based forecasting framework.

| Ref. | Mechanism | Output Format | Benefit for Decision-Makers |
|---|---|---|---|
| [147] | Natural Language Justifications | Concise summaries of key variables, weather events, and system states influencing forecasts. | Builds trust by linking predictions to domain-relevant factors |
| [148] | Scenario-based "What-If" Analysis | Forecast adjustments with causal explanations for hypothetical conditions | Supports contingency planning and operational readiness |
| [149] | Interactive Visualisation Dashboards | Attention maps, SHAP feature attributions, and temporal saliency plots | Enables intuitive exploration of model reasoning |
| [150] | Audit and Compliance Logging | Archived forecasts with associated explanations and metadata | Facilitates regulatory review and ensures accountability |
| [151] | Domain-specific Constraints | Integration of physical and operational limits into predictions | Ensures outputs are realistic and policy-aligned |

uncertainty quantification methods, may alleviate the problems surrounding data reliability [133].

Limited access to open–access, high–resolution data for the training of AI models is also a barrier, restricting researchers and developers from creating robust forecasting models [134]. By promoting and supporting data-sharing initiatives and collaborations between governments, research institutions, and energy companies, we could pave the way for more accurate and generalisable LLM-driven forecasting solutions [135]. Resolving these challenges is crucial for improving the robustness and accuracy of LLM-based renewable energy forecasting.

### *6.1.2. Model interpretability and explainability*

Interpretability remains a pressing issue, as LLMs often operate as black boxes, complicating trust and regulatory acceptance, making it difficult to understand how they arrive at their predictions. Also, interpretability is essential for trust and decision-making, for energy grid operators and policymakers [136]. As LLMs employ deep learning architectures that operate on massive amounts of data through multiple layers of abstraction, our understanding of them is far more complex and nuanced by comparison, perhaps explaining the difficulties involved in unpacking causal relationships [137].

In response to this, researchers are developing XAI methods specifically tailored for energy forecasting, such as attention mechanisms, feature attribution techniques, and surrogate models that capture LLM behaviour in a more interpretable way. Visualisation techniques from the attention domain (heatmaps, attention scores, etc.) help stakeholders understand which features used in the prediction process contribute the most to LLM-based results [138]. To further enhance interpretability, feature attribution methods such as SHAP and LIME are being employed to decompose intricate LLM decision-making into more easily understandable building blocks [139].

Moreover, integrating domain-specific constraints and knowledge graphs can enhance the transparency of LLM-based forecasting systems [140]. By embedding physics-based modelling constraints into LLMs, researchers aim to align AI-driven predictions with established scientific principles governing energy generation and consumption [141]. Additionally, knowledge graphs enable LLMs to establish causal relationships between energy consumption patterns, weather anomalies, and market fluctuations, thereby improving decision-making accuracy [142].

Enabling energy stakeholders with interpretable outputs (e.g., confidence intervals, uncertainty quantification, natural-language justifications) will be necessary for real-world adoption and regulatory compliance in renewable energy governance [143]. Moreover, hybrid AI strategies that combine LLMs with rule-based approaches and reinforcement learning are an active area of research aimed at achieving better explainability without sacrificing predictive performance [144]. The adoption of LLMs for renewable energy solutions can be improved by advancing explainability frameworks and interpretability mechanisms to make them more trustworthy tools in the hands of energy providers, policymakers, and regulatory structures [145].

Fig. 10 provides a multi-layered view of interpretability techniques applied to LLM-based renewable energy forecasting. The top-left panel presents an attention pattern heatmap, revealing that the model assigns stronger attention weights to recent time steps, thereby capturing short-term temporal dependencies essential for accurate forecasting. The top-right panel shows SHAP-derived feature importance scores for a load forecasting scenario, quantifying the influence of variables such as temperature, solar irradiance, wind speed, and cloud cover on the predicted outputs. The bottom section synthesises these approaches into a comprehensive interpretability framework, categorising them as intrinsic (e.g., attention visualisation, physical constraints) or post-hoc (e.g., feature attribution, surrogate models). This framework explicitly links interpretability outputs to the trust and confidence of key energy stakeholders, including grid operators, energy traders, and policymakers, thereby addressing the need for transparent, explainable, and actionable LLM-driven forecasts in operational decision-making.

Table 8 presents the operational interpretability mechanisms incorporated into the proposed LLM-based forecasting framework, along with their corresponding output formats and practical benefits for energy sector decision-makers. By linking each mechanism to supporting literature, the table demonstrates how advanced XAI techniques, such as natural language justification, scenario-based "what-if" analysis, and SHAP-based visualisations, are adapted to address the specific transparency needs of grid operators, policymakers, and regulatory bodies. The inclusion of audit logging and domain-specific constraints ensures that model explanations are not only comprehensible but also actionable within the context of renewable energy governance and compliance.

### 6.1.3. Computational complexity, inference latency, and energy consumption

Large language models introduce substantial computational, memory, and energy requirements that can constrain their practical use in renewable energy forecasting. Compared with conventional machine learning models and lighter transformer-based forecasters, general-purpose LLMs typically require high-performance GPUs, larger

**Table 9**
Reported strategies and evidence for balancing computational burden, latency, and interpretability in LLM- and AI-enabled energy forecasting workflows.

| Ref. | Challenge | Concrete evidence | Mitigation strategy | Operational effect |
|---|---|---|---|---|
| [155] | High training energy consumption | Example use of a smaller domain-specific model, such as Llama 3.1 8B | Use smaller domain-specific models instead of very large general-purpose models LLMs | Reduces computational load, memory demand, and carbon footprint while retaining task relevance |
| [156] | Extensive parameter updates during fine-tuning | Parameter-efficient fine-tuning methods such as LoRA or prefix-tuning can reduce GPU hours and memory usage by up to 90% | Fine-tune only a small subset of parameters | Lowers adaptation cost and makes customisation more feasible for utilities with limited compute budgets |
| [157] | Heavy inference cost in real-time forecasting | Distilled student models are reported to reduce inference burden and improve serving efficiency relative to larger teacher models. | Deploy compressed student models for inference-time use | Preserves much of the predictive utility with lower latency and energy consumption |
| [158] | Carbon-intensive compute infrastructure | Cloud or HPC resources can be selected based on renewable-powered Or lower carbon supply | Use a renewable-powered cloud or HPC infrastructure | Aligns model development and deployment with sustainability objectives and reduces lifecycle emissions. |
| [159] | Underutilised hardware during training | Shared scheduling and off-peak execution can improve accelerator utilisation. | Schedule training during low-demand periods and share HPC resources | Improves resource utilisation and lowers the effective energy cost per experiment |
| [160] | Explainability overhead in near-real-time forecasting | Combined prediction and explanation time, including SHAP and LIME, is approximately 250 ms per instance in a REST API plus IoT simulation setting | Use efficient TreeSHAP computation, API-based deployment, and caching or surrogate deployment through distillation | Demonstrates that interpretable forecasting can remain viable for near-real-time smart energy applications |
| [161] | Need for transparent forecasting without sacrificing practical usability | LightGBM is identified as the best trade-off among evaluated ML models, while SHAP and LIME provide global and local explanations for PV, wind, and electric demand forecasting | Combine accurate tabular forecasting models with posthoc XAI methods rather than relying on opaque black-box outputs. | Improves trust, situational awareness, and decision support in renewable-rich operational settings |

**Table 10**
Concrete evidence on model scale, computational cost, latency, and interpretability-related signals from representative studies.

| Ref. | Task/domain | Concrete model-size or compute evidence | Latency/deployment evidence | Interpretability-related evidence | Use in this review |
|---|---|---|---|---|---|
| [160] | Multi-scale household energy forecasting | Holt-Winters + XGBoost + XAI framework; reports very high predictive accuracy with $R^2$ = 0.9999 and RMSE = 0.00344 | Combined prediction and explanation time is approximately 250 ms per instance using a REST API integrated with an IoT simulation; also reports Streamlit dashboard and Flask deployment. | SHAP, LIME, permutation feature importance, and partial dependence plots; LIME stability analysis is also reported | Strongest evidence in this review for combining concrete explainability with deployment-oriented latency evidence |
| [161] | PV, wind, and electric-demand forecasting | Evaluates multiple ML models, including XGBoost, Random Forest, NGBoost, CatBoost, SVR, and KNN, with LightGBM identified as the best trade-off between accuracy, stability, and interpretability | No explicit hardware or numerical latency reported, but the framework is positioned for scalable deployment in renewable-rich operational settings. | SHAP quantifies global and local feature contributions; LIME provides case-specific explanations; the paper includes SHAP-based global and local interpretation and instance-level LIME explanations | Strong evidence for practical post-hoc XAI integration in renewable energy forecasting, but weaker on compute-cost quantification |
| [162] | Load, solar, and wind multi-horizon forecasting | Reports workstation-level compute setup: Intel Core i9 CPU, 32 GB RAM, NVIDIA RTX 3090 with 24 GB VRAM; batch size = 64, learning rate = 0.001, dropout = 0.2–0.3 | Simulated near-real-time setting with sub-second inference latency of approximately 0.12 s per sample; interpretability adds only about 0.01 s | SHAP, LIME, and Transformer attention heatmaps are explicitly used; SHAP identifies influential features, LIME provides local explanations, and attention maps highlight salient time steps and variables | Useful supporting evidence for combined XAI, hardware, and latency reporting in hybrid deep-learning energy forecasting, although deployment is evaluated in a controlled simulated environment rather than a full operational field setting |
| [163] | Load forecasting/ privacy-preserving smart grids | Uses LoRA for parameter-efficient fine-tuning; updates only low-rank matrices while freezing the base LLM; reports r = 64, α = 64, and dropout = 0.1 | States that LoRA introduces minimal additional inference latency because Wx and (AB)x can be computed in parallel; no absolute latency value reported | No SHAP or LIME reported | Strong evidence for parameter-efficient adaptation and qualitative latency mitigation in LLM-based load forecasting |
| [164] | PV production and consumption-load forecasting | Reports training time per epoch across models: CNN 28 s, LSTM 41 s, GRU 40 s, TCN 41 s, Transformer 33 s, CNN-TCN 44 s, CNN-GRU 43 s, CNN-LSTM 46 s, and CT-Transformer 49 s | Provides direct training-time evidence, but not inference latency | Attention mechanism is described as learning sequence dependence and enabling parallel GPU-based processing, but no post-hoc SHAP or attention-visualisation output is reported | Strong evidence for concrete computational-cost comparison and accuracy-cost trade-off analysis |
| [165] | Park-level integrated energy systems/green power forecasting | Explicitly uses LLAMA-7B for photovoltaic and wind power prediction in the smart management system | Applied in an intelligent park-management setting, but no numerical latency value is reported | No formal XAI method reported | Useful evidence for model scale in an applied energy-management setting |
| [166] | Wind speed forecasting | Reports optimiser Adam, learning rate 0.0002, number of multi-head attention = 2, dropout = 0.3, epochs = 100, batch size = 32, and MRI blocks = 3 | No hardware or explicit latency reported | Provides architectural interpretability signals through temporal attention and interactive attention across multi-resolution subsequence, but not post-hoc XAI visualisation | Useful as evidence of transformer design complexity and attention-based architectural reasoning |
| [167] | Multiple wind-farm forecasting | Reports results averaged over 100 runs; for 8 h-ahead forecasting, Transformer + achieves an MAPE of 8.4%; the paper states that the parallelised input structure improves model training efficiency. | No explicit latency or hardware reported | Multi-head attention is used to capture cross-farm correlations, but no SHAP, LIME, or explicit attention-map visualisation is shown. | Useful as supporting evidence that transformer parallelism and attention can improve forecasting quality and efficiency |

**Table 11**
Integration challenges for deploying LLMs in renewable energy forecasting.

| Ref. | Integration Area | Challenge Description |
|---|---|---|
| [168] | SCADA Systems | Integration with supervisory control and data acquisition (SCADA) pipelines is impeded by proprietary protocols, high-frequency streams, and limited interoperability with modern AI data buses. |
| [169] | IoT and Smart Devices | Explosive growth of IoT sensors and smart meters produces large, continuous streams that LLMs must ingest and infer on securely and in (near) real time. |
| [69] | Grid Infrastructure | Legacy grid assets may lack compute, bandwidth, and standardised interfaces to host or interact with LLM-based forecasting, creating deployment bottlenecks. |
| [170] | Policy and Regulation | Compliance with evolving market rules, cybersecurity baselines, and data governance (privacy/retention/ sovereignty) can lag behind technical capability, constraining rollout. |

memory budgets, and more costly cloud or high-performance computing infrastructure, which may limit adoption by utilities, grid operators, and smaller organisations [152,153]. The sustainability dimension is also important because training and serving large models in energy-intensive data centres can increase the carbon footprint of AI-enabled forecasting workflows [154]. However, the reviewed literature indicates that these burdens can be reduced through model-size selection, parameter-efficient adaptation, model compression, and hardware-aware deployment. As summarised in Table 12, reported mitigation strategies include the use of smaller domain-specific models such as Llama 3.1 8B, parameter-efficient fine-tuning methods that can reduce GPU hours and memory usage by up to 90%, distilled student models for lower-cost inference, and renewable-powered or resource-shared computing infrastructure [155–159].

The interpretability side of this trade-off also requires concrete treatment rather than purely qualitative discussion. In this review, Fig. 10 already illustrates representative explainability mechanisms, while Table 11 summarises operator-facing interpretability tools, including temporal saliency analysis, natural-language justifications, scenario-based what-if analysis, and audit logging. Recent forecasting

**Table 12**
Recommended benchmarking and reporting template for renewable energy forecasting studies.

| Item | Recommended reporting fields |
|---|---|
| Dataset description | Source, site or region, asset type, date range, sampling interval, number of assets, missing-data rate, and whether data are from SCADA, IoT, weather feeds, maintenance logs, market signals, or mixed modalities [171]. |
| Forecast task | Target variable, forecast horizon, update frequency, point or probabilistic forecasting, and whether the task includes anomaly detection or decision support [172]. |
| Data split | Train, validation, and test periods; rolling-origin or blocked split strategy; leakage controls; and whether cross-site or cross-season generalisation is evaluated [173]. |
| Baselines | Persistence model, statistical baseline, conventional ML model, recent transformer forecaster, and ablation of the proposed LLM-enabled components [16]. |
| Evaluation metrics | MAE, RMSE, MAPE, $R^2$, calibration error for probabilistic outputs, and event-specific metrics where ramp or fault events are relevant [174]. |
| Statistical testing | Mean and standard deviation across runs, confidence intervals, and statistical significance testing where repeated experiments are feasible [174]. |
| Compute and hardware | GPU or CPU type, memory, parameter count, training time, inference batch size, and whether quantisation, pruning, or distillation is used [175]. |
| Operational metrics | End-to-end latency, throughput, model warm-start time, retrieval time, data freshness, and system availability [173]. |
| Robustness checks | Performance under missing data, delayed telemetry, sensor drift, distribution shift, and unseen weather regimes [173]. |
| Deployment context | Edge, on-premises, cloud, or hybrid deployment; model serving framework; fallback policy; and human oversight requirements [16]. |

studies provide stronger empirical support for this discussion. For example, XAI-enabled household energy forecasting frameworks have reported SHAP, LIME, permutation feature importance, and partial dependence plots together with deployment-oriented evidence such as REST API integration and response times of approximately 250 ms per instance [160]. Similarly, renewable forecasting studies based on LightGBM have combined SHAP-based global and local explanations with instance-level LIME analysis for photovoltaic, wind, and electric-demand prediction [161]. In addition, an interpretable hybrid deep-learning framework for load, solar, and wind forecasting has reported explicit SHAP, LIME, and attention-heatmap analysis together with workstation-level hardware details and sub-second inference latency, with only a small interpretability overhead [162]. These studies demonstrate that explainability in forecasting can be operationalised through both post-hoc feature-attribution methods and attention-based visual mechanisms rather than remaining an abstract design goal.

To complement the mitigation-oriented summary in Table 9, Table 10 consolidates representative evidence on model scale, computational burden, latency, deployment context, and interpretability support across selected studies. The resulting pattern is clear. LLM- and transformer-based forecasting systems can provide richer multimodal reasoning and more flexible operator support, but they must be carefully matched to the timing and resource constraints of the deployment environment. Not every task in a SCADA- or IoT-connected forecasting pipeline requires continual invocation of a large model. In practice, lightweight baselines, compressed student models, or efficient gradient-boosting predictors may be more appropriate for high-frequency prediction loops, while larger language-enabled components may be reserved for multimodal retrieval, event summarisation, anomaly explanation, and decision support.

This design principle aligns with the implementation-oriented deployment blueprint in Section 6.4, which treats latency monitoring, fallback logic, and selective LLM invocation as first-class operational requirements. Model pruning, quantisation, distillation, and parameter-efficient fine-tuning should therefore be viewed not only as sustainability measures, but also as practical enablers for real-time or near-real-time deployment [152]. Additional efficiency gains may come from specialised hardware accelerators such as TPUs and energy-efficient AI chips [176], renewable-powered cloud or HPC infrastructure, and, in the longer term, federated or decentralised AI pipelines that shift part of the computational load toward localised energy nodes [177]. Overall, the reviewed literature suggests that the most viable pathway is not unrestricted model scaling, but careful matching of model size, inference pathway, and interpretability support to the latency, cost, and governance constraints of the target forecasting workflow.

### 6.1.4. Hallucination risk and uncertainty evaluation

Reliable deployment of language-centric models in renewable energy forecasting requires explicit treatment of two closely related issues, namely hallucination and predictive uncertainty. In this context, hallucination refers to outputs that appear fluent and plausible but are not adequately supported by the available numerical evidence, retrieved documents, or physical operating constraints. Within renewable forecasting workflows, hallucinations may appear as fabricated causal explanations for forecast changes, incorrect summarisation of maintenance logs, invented links between weather events and asset behaviour, or overconfident recommendations that are inconsistent with SCADA telemetry or validated baseline models [16]. These failure modes are especially problematic in operator-facing settings because natural-language explanations can create a false sense of confidence even when the underlying forecast or recommendation is unreliable.

Hallucination risk can be reduced, although not eliminated, through grounding and verification strategies. In LLM-enabled renewable energy forecasting, retrieval-augmented generation (RAG) should be restricted to trusted operational sources such as SCADA records, maintenance logs, validated weather feeds, and documented grid events, with retrieved evidence filtered by timestamp, asset identifier, and relevance score before being used in model responses [178,179]. Structured output constraints, tool-based validation, and schema checking can further reduce unsupported responses by requiring the model to report specific variables, units, timestamps, event codes, and uncertainty bands rather than producing unrestricted free-text explanations. For high-impact operator-facing recommendations, generated explanations should be cross-checked against numerical forecasts, physical operating limits, rule-based safety constraints, and recent telemetry before they are displayed or acted upon. Where evidence is incomplete or inconsistent, the system should abstain, flag low confidence, or fall back to the last validated forecast or a lightweight baseline model, with human review required for abnormal, safety-critical, or dispatch-relevant cases [180]. Empirical grounding should also be reported through measurable indicators such as retrieval precision, evidence-support rate, unsupported-claim rate, physical-constraint violation rate, abstention or fallback frequency, expert agreement on generated explanations, and performance under missing or delayed telemetry. These measures make hallucination mitigation auditable and allow future studies to evaluate whether LLM-generated explanations and recommendations are genuinely supported by operational evidence rather than merely plausible.

Uncertainty quantification is equally important because forecasting quality in renewable energy systems cannot be adequately assessed using point-error metrics alone. Wind, solar, and load forecasting are affected by meteorological variability, sensor noise, missing data, regime shifts, and site-specific non-stationarity, all of which introduce both data uncertainty and model uncertainty. Accordingly, studies should report probabilistic forecasts whenever possible, for example, through prediction intervals, quantile forecasts, distributional outputs, or ensemble-based uncertainty estimates [181,182]. Conformal prediction is particularly relevant in this setting because it can provide calibrated prediction intervals or sets with finite-sample coverage guarantees under relatively weak assumptions, making it attractive for operational forecasting tasks where distribution shift and limited labelled data are common [74]. Calibration analysis is also necessary to determine whether predicted confidence levels are consistent with observed frequencies rather than merely appearing numerically plausible.

From an evaluation perspective, future benchmarking should include both accuracy and reliability criteria. In addition to MAE, RMSE, or MAPE, studies should report interval coverage, mean interval width, pinball loss for quantile forecasts, calibration error or reliability diagrams, and proper scoring rules such as the continuous ranked probability score when distributional forecasts are available [26,183]. The evaluation should also examine whether uncertainty remains well-calibrated under missing telemetry, delayed data ingestion, seasonal shifts, unseen sites, and rare ramp events. For LLM-enabled forecasting systems, these reliability checks should be complemented by hallucination-oriented evaluation measures, including retrieval precision, evidence-support rate, unsupported-claim rate, physical-constraint violation rate, abstention or fallback frequency, and expert agreement on generated explanations. This is important because an uncertainty-aware and explanation-generating model is useful only when both its confidence estimates and its natural-language outputs remain grounded, calibrated, and informative under the same operational stressors that degrade forecasting performance.

More broadly, hallucination mitigation and uncertainty evaluation should be treated as linked requirements for trustworthy deployment rather than as separate add-ons. A forecast explanation that is not grounded in verified evidence is unsafe, and a numerically accurate forecast without calibrated uncertainty may still be misleading for dispatch, reserve scheduling, or operator intervention. Therefore, reliable renewable forecasting pipelines should combine grounded retrieval, constrained reasoning, explicit uncertainty reporting, fallback logic, and auditability.

### 6.2. Integration challenges

Beyond technical barriers, integrating LLMs into existing energy infrastructure and regulatory frameworks presents additional hurdles. While technical challenges focus on LLM performance, a separate set of obstacles arises when embedding these models into real-world energy systems. However, renewable energy forecasting doesn't operate in isolation; it has to be aligned with the operational SCADA pipeline, IoT sensor networks, existing power grids, and evolving regulatory frameworks. LLMs' success depends not only on their technological sophistication but also on how well they can be integrated into the energy sector's existing digital and physical ecosystem. To summarise the key integration barriers and their implications for LLM deployment in renewable energy forecasting, Table 11 consolidates the main challenge areas together.

### 6.3. Future directions

We structure this section into sub-subsections to highlight each research agenda item, such as hybrid AI approaches, benchmarks and datasets, SCADA integration, and responsible AI, as independent directions. This ensures that the breadth of future work is conveyed clearly without compressing diverse opportunities into a single narrative.

#### 6.3.1. Hybrid AI approaches

Enhanced precision and resilience through the integration of LLMs and physics-based models [184]. This technique enables LLMs to leverage basic energy principles and domain-specific constraints, leading to more robust, physically meaningful predictions. Hybrid AI models that combine machine learning-based pattern recognition with physics-driven modelling can improve renewable energy predictions, particularly in cases of limited training data or rapidly changing environmental variables [185]. Moreover, hybrid AI models enable more comprehensible scenario-based simulations by integrating NWP models, real-time sensor measurements, and CFD simulations [186]. This combination enhances assessments of wind and solar energy resources, optimising energy generation strategies under various meteorological and grid conditions [187–189].

Furthermore, hybrid AI frameworks can utilise reinforcement learning to adjust forecasting models based on their past performance, thereby improving subsequent predictions [190]. These paradigms enable self-learning for models by analysing and correcting systematic errors over time through parameter adjustment. Hybrid models are particularly useful for embedding expert knowledge and decision frameworks based on well-established rules (thus ensuring from the very start that AI predictions adhere to established physical and engineering principles) [191,192]. We believe that such an integration limits the potential for incorrect predictions, while improving model interpretability and thereby the broader trustworthiness of AI-guided energy forecasts for decision-makers, such as politicians, energy grid managers, and investors [193].

In addition, hybrid models enable multi-objective optimisation of energy efficiency, cost-effectiveness, and sustainability objectives [194–196]. These are all important applications, such as optimising energy storage strategies, minimising power losses in transmission networks, and improving the economic viability of integrated renewable energy projects. This synergy between LLMs and physics-based models represents a new frontier in AI-driven energy forecasting, offering the potential for high reliability, adaptability, and explainability to support informed decision-making in renewable energy management.

#### 6.3.2. Explainability and interpretability

Techniques to augment transparency and trust in predictions made by LLMs for stakeholders of the energy sector using attention visualisations, SHAP (Shapley Additive Explanations), or LIME (Local Interpretable Model-Agnostic Explanations) [197,198]. Such methods can therefore be helpful for interpreting LLM-generated predictions and for identifying key influencing variables in energy forecasting [199]. Additionally, introducing loss functions that incorporate domain knowledge of constraints and physics into LLM architectures will ensure that outputs are consistent with the behaviours of real-world energy systems, in that forecasts reflect the physical fundamentals of electromechanical energy production and loads [73,200].

**Table 13**
Example reporting fields for SCADA-connected and context-aware renewable energy forecasting studies.

| Field | Type | Description |
|---|---|---|
| timestamp | datetime | UTC timestamp for each record after time synchronisation. |
| asset id | string | Unique identifier for turbine, inverter, feeder, plant, or smart meter. |
| site id | string | Plant or region identifier for multi-site forecasting. |
| target variable | float | Measured power output, load demand, or other prediction target. |
| weather features | vector | Forecast or measured exogenous variables such as wind speed, irradiance, temperature, humidity, and pressure. |
| operational _features | vector | SCADA telemetry such as rotor speed, blade pitch, inverter status, voltage, current, frequency, or storage state of charge. |
| quality flag | categorical | Indicator for missingness, sensor fault, manual correction, or out-of-range values. |
| event code | string | Alarm, curtailment, outage, or maintenance event identifier. |
| event text | text | Free-text maintenance note, operator comment, incident summary, or weather bulletin. |
| retrieval id | string | Key linking the time-stamped record to retrieved documents or vector-store entries. |
| forecast horizon | integer | Prediction horizon in minutes or hours. |
| prediction | float | Model output for the specified horizon. |
| uncertainty band | vector | Prediction interval or quantile outputs when uncertainty is reported. |
| model version | string | Version tag for auditability, rollback, and reproducibility. |

Additionally, integrating explainability tools such as counterfactual reasoning, feature attribution, and causal analysis can further enhance the trustworthiness of LLM-driven forecasts [201,202]. Future research must strive to design user interfaces that enable grid operators and decision-makers to explore, analyse, and verify LLM-generated insights using interactive dashboards, natural language queries, and mechanisms for in-context feedback. A future challenge in bridging theoretical AI models and practical decision-making for energy management will be ensuring that decision-makers can see and act on the predictions generated by advanced AI models.

#### 6.3.3. Development of energy-specific LLMs

Fine-tuning LLMs specifically for energy data by integrating domain knowledge, physics-based constraints, and renewable energy system dynamics [203,204]. This is achieved by incorporating essential elements, such as weather-dependent variability, grid infrastructure parameters, and operational limits, into model architectures to enhance predictive accuracy [205]. Physics-informed neural networks (PINNs) can creatively design inputs for predictions that align with the fundamental laws governing the respective energy systems [206,207]. Also, Self-supervised learning techniques, which have gained traction in recent years, would allow LLMs to be updated in tandem to keep pace with changes in the energy marketplace, regulation, and renewable energy technologies without requiring large amounts of labelled data [13,208]. This potentially improves the model's generalisation and robustness across different energy modes.

In particular, future work should investigate the application of specialised LLMs to sector-specific energy areas, such as demand response management, where LLMs can be used to develop optimised energy consumption patterns and strategies based on real-time grid conditions and underlying economics. From a grid operation perspective, LLMs can be key to identifying faults, cybersecurity threats, and inefficiencies on the grid before they materialise into catastrophic failure by building an understanding of anomalies [209,210]. For example, we could use LLMs to power predictive maintenance of energy assets by analysing sensor data, historical performance logs, and environmental variables to schedule equipment servicing in advance, minimising downtime.

By integrating these advanced techniques, energy-specific LLMs can significantly improve grid resilience, reduce operational costs, and promote a more sustainable and intelligent energy ecosystem.

### 6.4. Evidence-informed reporting guidance for SCADA-connected LLM-enabled forecasting studies

To avoid presenting a speculative deployment architecture, this subsection reframes the implementation discussion as evidence-informed reporting guidance derived from recurring themes in the reviewed literature. Across the reviewed studies, practical deployment concerns repeatedly include data interfaces, SCADA and IoT integration, inference latency, serving configuration, fallback logic, monitoring, auditability, and human oversight [211–213]. However, only a limited number of studies provide full operational validation in live SCADA environments. Therefore, the benchmarking template, schema fields, and serving considerations presented below should be interpreted as recommended reporting items and synthesis-based deployment considerations rather than as a validated universal SCADA deployment blueprint.

#### 6.4.1. Benchmarking and reporting template

For fair comparison between conventional ML, transformer forecasters, time-series foundation models, and LLM-enabled pipelines, studies should report the items listed in Table 12. This structure is intended to reduce ambiguity caused by heterogeneous datasets, horizons, and evaluation protocols.

#### 6.4.2. Example reporting fields for SCADA and contextual data

Because the reviewed literature reports heterogeneous SCADA, IoT, weather, and contextual data inputs, future studies would benefit from clearer and more consistent reporting of the data fields used in forecasting pipelines. Table 13, therefore, provides example reporting fields that can help describe SCADA-connected, context-aware forecasting studies more transparently. The table is not intended to prescribe a universal schema, but to summarise the types of structured and unstructured information that are commonly relevant when linking renewable-energy forecasting models with operational data streams.

#### 6.4.3. Serving architecture, latency, and deployment considerations

Building on the evidence-informed reporting perspective above, this subsection summarises serving, storage, fallback, and monitoring considerations that recur across the reviewed literature and should be reported more transparently in future SCADA-connected forecasting studies. Rather than prescribing a validated deployment architecture, these considerations can be organised into five reporting dimensions.

1. **Ingestion layer:** SCADA and IoT streams are collected continuously, validated for missing values and range violations, and aligned with lower-frequency meteorological and market data.
2. **Feature and retrieval layer:** Structured telemetry is written to a time-series or relational store, while maintenance notes, incident reports, and policy documents are indexed in a vector or document store to support retrieval-augmented inference.
3. **Model serving layer:** Conventional machine learning or transformer-based forecasters generate the primary numerical forecast, while the LLM module is invoked selectively for multimodal fusion, anomaly explanation, summarisation, or event-aware adjustment rather than for every control-cycle decision.
4. **Decision and fallback layer:** If telemetry is delayed, retrieval fails, or the LLM response exceeds the latency budget, the system falls back to the last validated forecast or to a lightweight baseline model.
5. **Monitoring layer:** The deployment logs model version, input freshness, inference time, forecast error, drift indicators, and operator overrides to support auditing and periodic recalibration.

Because latency tolerance is application dependent, no single threshold is appropriate for all renewable energy systems. Nevertheless, an illustrative operational envelope for short-horizon forecasting can be reported as follows: data ingestion and validation within 1–2 s, feature refresh and retrieval within 1 s, primary model inference within 1–3 s, and total advisory response time within 5–10 s [214, 215]. These values should be treated as reporting targets rather than universal requirements.

Key deployment risks include timestamp misalignment across sources, packet loss in telemetry streams, schema drift, retrieval of irrelevant textual context, excessive inference latency during peak load, model degradation under seasonal shift, and limited auditability of generated natural-language explanations. Accordingly, production deployment should include schema validation, queue back-pressure handling, confidence-based abstention, human review for high-impact recommendations, and scheduled re-evaluation against strong non-LLM baselines.

Overall, this evidence-informed guidance clarifies that future work should report not only predictive accuracy, but also data-interface assumptions, latency constraints, model-serving choices, fallback behaviour, monitoring procedures, and the extent of operational validation. This framing avoids treating the proposed fields as a validated architecture and instead positions them as reporting recommendations derived from the limitations and recurring deployment concerns identified in the reviewed literature.

## 7. Conclusion

Building on the literature synthesised in this review, LLMs and related foundation model approaches are emerging as promising tools for renewable energy forecasting and optimisation across wind, solar, and load demand applications. Their main potential lies in the ability to integrate heterogeneous information sources, including historical energy data, meteorological observations, operational records, and textual context, thereby supporting forecasting workflows that extend beyond conventional single-source prediction pipelines. Across the studies reviewed, some papers report improvements in forecasting quality, interpretability, or decision support, particularly in multimodal and hybrid settings. However, these benefits are not universal and remain strongly dependent on dataset characteristics, model design, evaluation protocol, and operational context. Accordingly, this review does not claim that LLM-based methods exhibit superior standardised performance to traditional machine learning approaches, because the included studies differ substantially in datasets, forecasting horizons, baselines, and evaluation procedures, which limit direct quantitative comparisons.

This review should be understood as a targeted extension of the existing LLM-energy review literature rather than as a first-of-its-kind survey. While prior reviews have examined LLM applications across energy systems, power systems, smart grids, and building energy contexts, the present review narrows the focus to renewable energy forecasting and optimisation, particularly wind, solar, and load-demand applications. Its main contribution is therefore the consolidation of this narrower forecasting-oriented evidence base, with emphasis on distinguishing language-centric LLM pipelines from time-series foundation models and general transformer-based forecasters, and on highlighting methodological and deployment-related issues such as benchmarking practice, evaluation design, SCADA/IoT integration, uncertainty, and reliability.

The review further shows that, although progress is encouraging, the application of LLMs in renewable energy forecasting remains at an early stage of maturity. Major challenges persist in relation to data quality, interpretability, computational cost, reproducibility, regulatory compliance, and robust real-time deployment. In addition, hallucination risk and uncertainty quantification remain critical but under-addressed issues. In safety- and reliability-sensitive forecasting settings, fabricated explanations, unsupported recommendations, or poorly calibrated confidence estimates can undermine trust even when average predictive accuracy appears favourable. For this reason, future studies should evaluate not only point prediction performance but also hallucination resilience, calibration quality, probabilistic reliability, and interval-based uncertainty behaviour under realistic operating conditions.

Looking ahead, future research should prioritise open and reproducible benchmarks, stronger comparative evaluation against conventional and transformer-based baselines, and clearer reporting of data regimes, forecast horizons, and evaluation metrics. Greater emphasis is also needed on grounded and trustworthy deployment practices, including retrieval validation, structured output constraints, fallback logic, human oversight, and uncertainty-aware forecasting frameworks such as probabilistic prediction, calibration analysis, and conformal prediction. Additional research is warranted on physics-informed and hybrid artificial intelligence frameworks, transferability across regions and energy markets, and low-latency integration into streaming operational environments. Embedding these models into practical forecasting pipelines, including Supervisory Control and Data Acquisition workflows and real-time decision support systems, will be essential for translating research progress into operational value. With these advances, LLM-enabled systems may become useful supporting components of renewable energy forecasting and management, provided that their benefits are assessed rigorously, their uncertainty is reported transparently, and their limitations are managed explicitly.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rser.2026.117171.

## Data availability

No data was used for the research described in the article.

## References

[1] Ahmed SD, Al-Ismail FS, Shafiullah M, Al-Sulaiman FA, El-Amin IM. Grid integration challenges of wind energy: a review. IEEE Access 2020;8:10857–78.
[2] Ahmed M, Mirsaeidi S, Koondhar MA, Karami N, Tag-Eldin EM, Ghamry NA, et al. Mitigating uncertainty problems of renewable energy resources through efficient integration of hybrid solar PV/wind systems into power networks. IEEEThesis MSc Comput Sci Special: Data Sci Access 2024;12:30311–28.
[3] Liu L, He G, Wu M, Liu G, Zhang H, Chen Y, et al. Climate change impacts on planned supply–demand match in global wind and solar energy systems. Nat Energy 2023;8:870–80.
[4] Cavus M. Advancing power systems with renewable energy and intelligent technologies: a comprehensive review on grid transformation and integration. Electronics 2025;14:1159.
[5] Ejuh Che E, Roland Abeng K, Iweh CD, Tsekouras GJ, Fopah-Lele A. The impact of integrating variable renewable energy sources into grid-connected power systems: challenges, mitigation strategies, and prospects. Energies 2025;18:689.
[6] Feng T, Shen H, Yang X, Nianga J-M, Wang Z. Integration of large language models with IoT in smart agriculture to improve efficiency, yield, and quality. Ind Sci Eng 2024;1:15–35.
[7] Senthil Kumar S, Srinivasan C, Sridhar P. Enhancing grid stability and efficiency in buildings through forecasting and intelligent energy management of distributed energy resources. Electr Eng 2024:1–18.
[8] Mahmud D, Hajmohamed H, Almentheri S, Alqaydi S, Aldhaheri L, Khalil RA, Saeed N. Integrating llms with its: recent advances, potentials, challenges, and future directions. IEEE Trans Intell Transport Syst 2025;26(5):5674–709.
[9] Boumaiza A. A blockchain-based scalability solution with microgrids peer-to-peer trade. Energies 2024;17:915.
[10] Alazemi T, Darwish M, Radi M. Renewable energy sources integration via machine learning modelling: a systematic literature review. Heliyon 2024;10:e26088.
[11] Krasnopolsky V. Applying machine learning in numerical weather and climate modeling systems. Climate 2024;12:78.
[12] SaberiKamarposhti M, Kamyab H, Krishnan S, Yusuf M, Rezania S, Chelliapan S, Khorami M. A comprehensive review of AI-enhanced smart grid integration for hydrogen energy: advances, challenges, and future prospects. Int J Hydrogen Energy 2024;67:1009–25.
[13] Zhao, H., Wu, L., Shan, Y., Jin, Z., Sui, Y., Liu, Z., et al. A comprehensive survey of large language models in management: Applications, challenges, and opportunities. SSRN preprint. https://doi.org/10.2139/ssrn.4932769.
[14] Zaboli A, Kasimalla SR, Park K, Hong Y, Hong J. A comprehensive review of behind-the-meter distributed energy resources load forecasting: models, challenges, and emerging technologies. Energies 2024;17:2534.
[15] Giotopoulos KC, Michalopoulos D, Vonitsanos G, Papadopoulos D, Giannoukou I, Sioutas S. Dynamic workload management system in the public sector: a comparative analysis. Future Internet 2025;17:119.
[16] Mirshekali H, Reza Shadi M, Ghanadi Ladani F, Reza Shaker H. A review of large language models for energy systems: applications, challenges, and future prospects. IEEE Access 2025;13:163162–88.
[17] Amjad F, Korõtko T, Rosin A. Review of LLMs applications in electrical power and energy systems. IEEE Access 2025;13:150951–69.
[18] Ghafari M, Sami A, Rezapour H, Saravani Ghayour S, Lin F, Basaran K, et al. A comprehensive review on the application of large language models (LLMs) in power systems. IEEE Access 2025;13:209450–86.
[19] Zhou X, Xu Y, Zhao J, Zhang R. Large language model applications in power systems: a comprehensive review and outlook. J Mod Power Syst Clean Energy 2026:1–18.
[20] Liu M, Zhang L, Chen J, Chen W-A, Yang Z, Lo LJ, et al. Large language models for building energy applications: opportunities and challenges. Springer; 2025. p. 225–34.
[21] Bressane A, Garcia AJdS, Castro MVd, Xerfan SD, Ruas G, Negri RG. Fuzzy machine learning applications in environmental engineering: does the ability to deal with uncertainty really matter? Sustainability 2024;16:4525.
[22] Bilgili A, Gül K. Forecasting power generation of wind turbine with real-time data using machine learning algorithms. Clean Technol Recycl 2024;4:108–24.

[23] Liang W, Wang G, Jiao K, Ning X, Zhang J, Guo X, et al. Conversion mechanism and gasification kinetics of biomass char during hydrothermal carbonization. Renew Energy 2021;173:318–28.
[24] Ma Z, Mei G. A hybrid attention-based deep learning approach for wind power prediction. Appl Energy 2022;323:119608.
[25] Sri Preethaa K, Muthuramalingam A, Natarajan Y, Wadhwa G, Ali AAY. A comprehensive review on machine learning techniques for forecasting wind flow pattern. Sustainability 2023;15:12914.
[26] Toubeau J-F, Dapoz P-D, Bottieau J, Wautier A, De Greve Z, Vallée F. Recalibration of recurrent neural networks for short-term wind power forecasting. Elec Power Syst Res 2021;190:106639.
[27] Masini RP, Medeiros MC, Mendes EF. Machine learning advances for time series forecasting. J Econ Surv 2023;37:76–111.
[28] Yu J, Li X, Yang L, Li L, Huang Z, Shen K, et al. Deep learning models for PV power forecasting. Energies 2024;17:3973.
[29] Antonesi G, Cioara T, Anghel I, Michalakopoulos V, Sarmas E, Toderean L. From transformers to large language models: a systematic review of AI applications in the energy sector towards agentic digital twins. arXivpreprint; 2025. arXiv: 250606359.
[30] Zhou T, Ma Z, Wen Q, Wang X, Sun L, Jin R. Fedformer: frequency enhanced decomposed transformer for long-term series forecasting. PMLR; 2022. p. 27268–86.
[31] Liang Z, Chai R, Sun Y, Jiang Y, Sun D. Cold wave recognition and wind power forecasting technology considering sample scarcity and meteorological periodicity characteristics. Appl Sci 2025;15:4312.
[32] Cai H, Zhang Y, Zhang J, Chen J, Liu J, Xu J. A novel multi-scale and adaptive multi-period deep learning with compression-fusion attention for cold storage load prediction. Electronics 2025;15:160.
[33] Kim J, Kim H, Kim H, Lee D, Yoon S. A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges. Artif Intell Rev 2025;58:216.
[34] Du X, Lang Z, Liu M, Wu J. Regression analysis and prediction of monthly wind and solar power generation in China. Energy Rep 2024;12:1385–402.
[35] Hasanpour MA, Engholm R, Fafoutis X. Pump cavitation detection with machine learning: a comparative study of SVM and deep learning. IEEE; 2024. p. 219–25.
[36] Peng P. Predicting residential building cooling load with a machine learning random forest approach. Int J Interact Des Manuf 2024:1–14.
[37] Asghar R, Fulginei FR, Quercio M, Mahrouch A. Artificial neural networks for photovoltaic power forecasting: a review of five promising models. IEEE Access 2024;12:90461–85.
[38] Khan SZ, Muzammil N, Ghafoor S, Khan H, Zaidi SMH, Aljohani AJ, et al. Quantum long short-term memory (QLSTM) vs. classical LSTM in time series forecasting: a comparative study in solar power forecasting. Front Phys 2024;12: 1439180.
[39] Singh U, Rizwan M, Alaraj M, Alsaidan I. A machine learning-based gradient boosting regression approach for wind power production forecasting: a step towards smart grid environments. Energies 2021;14:5196.
[40] Ruan Z, Sun W, Yuan Y, Tan H. Accurately forecasting solar radiation distribution at both spatial and temporal dimensions simultaneously with fully-convolutional deep neural network model. Renew Sustain Energy Rev 2023;184:113528.
[41] Al Mamun A, Sohel M, Mohammad N, Sunny MSH, Dipta DR, Hossain E. A comprehensive review of the load forecasting techniques using single and hybrid predictive models. IEEE Access 2020;8:134911–39.
[42] Zou Y, Chen H, Wang J, Gong C, Wang Z, Jin S. Data center computing load forecasting based on feature modeling and the informer model. 2025. p. 739–44.
[43] Liu Y, Wang Z. Robust and efficient multihorizon photovoltaic power forecasting with a dilated multi-scale transformer. Sol Energy 2025;302:114077.
[44] Yang X, Li H, Huang X, Feng X. FEDAF: frequency enhanced decomposed attention free transformer for long time series forecasting. Neural Comput Appl 2024;36:16271–88.
[45] Wang J, Gao S, Zhao X, Huang X, Lu J, Wu C. A cross-domain information fusion method for non-stationary distributed photovoltaic forecasting. Appl Energy 2025;402:126867.
[46] Su Y, Zhang W, Zhou N, Deng G, Wang Z. A short-term time-series prediction approach for photovoltaic power generation in residential energy systems. Springer; 2025. p. 1–20.
[47] Suresh V. Benchmarking transformer variants for hour-ahead PV forecasting: PatchTST with adaptive conformal inference. Energies 2025;18:5000.
[48] Karpontinis D, Alexandridis G. Patch-based transformers for long-term energy consumption forecasting. Int J Artif Intell Tools 2026;35:2540013.
[49] Lee Y, Jeong J. TSMixer-and transfer learning-based highly reliable prediction with short-term time series data in small-scale solar power generation systems. Energies 2025;18:765.
[50] Gong M, Wu Z, Zhang F, Sun M, Huang J, Tian Y, et al. Adaptive multi-scale short term wind power forecasting framework based on TimeMixer and mix-of-experts. Elec Power Syst Res 2026;251:112196.
[51] Xu S, Gao M. A hybrid model for wind power forecasting based on multi-scale temporal channel mixing, PatchTST, and Chebyshev KAN. IEEE Access 2026;14: 20111–25.
[52] Adams JS, Altantzis C, Beucler T, Bingler J, Brown S, Brunschwiler T, et al. How to use the power of AI to reduce the impact of climate change on Switzerland. 2024.
[53] Oprea S-V, Bâra A. A recommendation system for prosumers based on large language models. Sensors 2024;24:3530.
[54] Ferdaus MM, Dam T, Sarkar MR, Uddin M, Anavatti SG. Foundation models for clean energy forecasting: a comprehensive review. Renew Sustain Energy Rev 2026;226:116452.
[55] Fatima SSW, Rahimi A. A review of time-series forecasting algorithms for industrial manufacturing systems. Machines 2024;12:380.
[56] de Zarzà I, de Curtò J, Roig G, Calafate CT. LLM multimodal traffic accident forecasting. Sensors 2023;23:9225.
[57] Shuvo MMH, Islam SK, Cheng J, Morshed BI. Efficient acceleration of deep learning inference on resource-constrained edge devices: a review. Proc IEEE 2023;111:42–91.
[58] Rodrigues TK, Suto K, Nishiyama H, Liu J, Kato N. Machine learning meets computation and communication control in evolving edge and cloud: challenges and future perspective. IEEE Commun Surv Tutor 2020;22:38–67.
[59] Sapkota R, Qureshi R, Usman Hadi M, Zohaib Hassan S, Sadak F, Shoman M, et al. Multi-modal LLMs in agriculture: a comprehensive review. IEEE Trans Autom Sci Eng 2025;22:22510–40.
[60] Luo H, Liu Y, Zhang R, Wang J, Sun G, Niyato D, et al. Toward edge general intelligence with multiple-large language model (Multi-LLM): architecture, trust, and orchestration. IEEE Trans Cogn Commun Netw 2025;11:3563–85.
[61] Eldeeb H, Elshawi R. Empowering machine learning with scalable feature engineering and interpretable AutoML. IEEE Trans Artif Intell 2025;6:432–47.
[62] Kulkarni SA, Gurupur VP, King C. Impact analysis of stacked machine learning algorithms based feature selections for deep learning algorithm applied to regression analysis. 2022. p. 269–75.
[63] Naser M. Why does machine learning work really well in many engineering problems? Machine learning for computational science and engineering, 1; 2025. p. 30.
[64] Adhikari S. Large language models in modern data engineering: a systematic review of architectures, use cases, and limitations, 2. International Journal of Business & Computational Science; 2025.
[65] Smajić A, Karlović R, Bobanović Dasko M, Lorencin I. Large language models for structured and semi-structured data, recommender systems and knowledge base engineering: a survey of recent techniques and architectures. Electronics 2025;14: 3153.
[66] Alangari N, El Bachir Menai M, Mathkour H, Almosallam I. Exploring evaluation methods for interpretable machine learning: a survey. Information 2023;14:469.
[67] Ghaffar Nia N, Amiri A, Luo Y, Kline A. Ethical perspectives on deployment of large language model agents in biomedicine: a survey. AI and Ethics 2026;6:32.
[68] Shivashankar K, Al Hajj G, Martini A. Maintainability and scalability in machine learning: challenges and solutions. ACM Comput Surv 2025;57:1–36.
[69] Madani S, Tavasoli A, Astaneh ZK, Pineau P-O. Large language models integration in smart grids. arXiv preprint; 2025. arXiv:250409059.
[70] Onile AE, Belikov J, Petlenkov E, Levron Y. Leveraging digital twins and demand side recommender chatbot for optimizing smart grid energy efficiency. IEEE 2023:1–5.
[71] Alaharju H. Ensuring performance and reliability in LLM-based applications: a case study. 2024.
[72] Zhang L, Chen Z. Opportunities of applying large language models in building energy sector. Renew Sustain Energy Rev 2025;214:115558.
[73] Zhang L, Chen Z, Ford V. Advancing building energy modeling with large language models: exploration and case studies. Energy Build 2024;323:114788.
[74] Rosca C-M, Stancu A. A comprehensive review of machine learning models for optimizing wind power processes. Appl Sci 2025;15:3758.
[75] Gu Y, You H, Cao J, Yu M, Fan H, Qian S. Large language models for constructing and optimizing machine learning workflows: a survey. arXiv preprint; 2024. arXiv:241110478.
[76] Guo M, Weng Y. Solar photovoltaic assessment with large language model. Appl Energy 2025;402:126835.
[77] Liao W, Yang Y, Yang Z, Bak-Jensen B, Fang J, Porté-Agel F. Short-term solar power forecasting with large. Language Model 2025:1–6.
[78] Lai Z, Wu T, Fei X, Ling Q. BERT4ST:: fine-tuning pre-trained large language model for wind power forecasting. Energy Convers Manag 2024;307:118331.
[79] Chen B, Yu J, Zhou M, Wang M. Research on PV power generation forecasting in zero-carbon residential buildings under multi-weather conditions based on a PV system large language model. Eng Res Express 2026;8:025328.
[80] He H, Hong R, Luo D, Huang S, Dai X. Photovoltaic power forecasting based on large language models: a prompt-learning and two-stage fine-tuning approach. Elec Power Syst Res 2026;252:112461.
[81] Sun J, Long H, Zhuo Y, Zheng S. StairLLM-Net: a large language model based approach for photovoltaic power forecasting. 2025. p. 388–93.
[82] Zhu G, Jia W, Xing Z, Xiang L, Hu A, Hao R. CMLLM: a novel cross-modal large language model for wind power forecasting. Energy Convers Manag 2025;330: 119673.
[83] Wu T, Ling Q. STELLM: spatio-temporal enhanced pre-trained large language model for wind speed forecasting. Appl Energy 2024;375:124034.
[84] Duan Z, Bian C, Yang S, Li C. Prompting large language model for multi-location multi-step zero-shot wind power forecasting. Expert Syst Appl 2025;280:127436.
[85] Fan H, Li M, Zhang Z, Cheng L, Ye Y, Liu W, et al. M2WLLM: multi-modal multi-task ultra-short-term wind power prediction algorithm based on large language model. Inf Fusion 2025:103541.
[86] Liu G, Bai Y, Wen K, Wang X, Liu Y, Liang G, et al. LFLLM: a large language model for load forecasting. TechRxiv 2025.
[87] Liang M, Hu Y, Weng H, Xi J, Yin B. EnergyGPT: fine-tuning large language model for multi-energy load forecasting. Renew Energy 2025;251:123313.
[88] Chen W, Zhang X, Zhu L, Shi J, Chang Z, Han Z, et al. Socially aware load forecasting utilizing large language models. IEEE Trans Ind Inf 2026;22:231–42.

[89] Zhou Y, Wang M. Empower pre-trained large language models for building-level load forecasting. IEEE Trans Power Syst 2025;40:4220–32.
[90] Li S, Wu L, Peng T, Huang J, Jiang H, Xue Y, et al. RSynLLM: a risk-aware routing mixture-of-experts large language model for multi-energy load forecasting in large-scale distribution networks. Appl Energy 2026;406:127227.
[91] Misiurek K, Olkuski T, Zyśk J. Review of methods and models for forecasting electricity consumption. Energies 2025;18:4032.
[92] Elmousalami H, Peng Hui FK, Alnaser AA. Enhancing smart and zero-carbon cities through a hybrid CNN-LSTM algorithm for sustainable AI-Driven solar power forecasting (SAI-SPF). Buildings 2025;15:2785.
[93] O'Connor C, Bahloul M, Prestwich S, Visentin A. A review of electricity price forecasting models in the day-ahead, intra-day, and balancing markets. Energies 2025;18:3097.
[94] Camacho M, Maldonado-Correa J, Torres-Cabrera J, Martín-Martínez S, Gómez-Lázaro E. Short-medium-term solar irradiance forecasting with a CEEMDAN-CNN-ATT-LSTM hybrid model using meteorological data. Appl Sci 2025;15:1275.
[95] Hassan M, Kabir ME, Jusoh M, An HK, Negnevitsky M, Li C. Large language models in transportation: a comprehensive bibliometric analysis of emerging trends, challenges and future research. IEEE Access 2025;13:132547–98.
[96] Liu M, Yu F, Diao Z, Huang Z, Wu H, Chen Y. Hyperspectral image classification based on dual-feature transformation fusion transformer. Earth Sci Inf 2025;18:486.
[97] Nguyen T, Shah R, Bansal H, Arcomano T, Maulik R, Kotamarthi R, et al. Scaling transformer neural networks for skillful and reliable medium-range weather forecasting. Adv Neural Inf Process Syst 2025;37:68740–71.
[98] Sapkota R, Qureshi R, Hassan SZ, Shutske J, Shoman M, Sajjad M, et al. Multi-modal LLMs in agriculture: a comprehensive review. Authorea Preprints; 2024.
[99] Kim J, Kim H, Kim H, Lee D, Yoon S. A comprehensive survey of time series forecasting: architectural diversity and open challenges. arXiv preprint; 2024. arXiv:241105793.
[100] Lin H, Yu M. A novel distributed PV power forecasting approach based on Time-LLM. arXiv preprint; 2025. arXiv:250306216.
[101] Qiu Z, Li C, Wang Z, Xie R, Zhang B, Mo H, et al. EF-LLM: energy forecasting LLM with AI-assisted automation, enhanced sparse prediction, hallucination detection. arXiv preprint arXiv:2411.00852v2 2024. https://doi.org/10.48550/arXiv.2411.00852.
[102] Yang Y, Lou H, Wu J, Zhang S, Gao S. A survey on wind power forecasting with machine learning approaches. Neural Comput Appl 2024;36:12753–73.
[103] Bai M, Zhou Z, Wang R, Yang Y, Qin Z, Chen Y, et al. HouYi: an open-source large language model specially designed for renewable energy and carbon neutrality field. arXiv preprint; 2023. arXiv:230801414.
[104] Alnegheimish S, Nguyen L, Berti-Equille L, Veeramachaneni K. Large language models can be zero-shot anomaly detectors for time series? arXiv preprint; 2024. arXiv:240514755.
[105] Chen S, Long G, Jiang J, Liu D, Zhang C. Foundation models for weather and climate data understanding: a comprehensive survey. arXiv preprint; 2023. arXiv:231203014.
[106] Wen X, Shen Q, Zheng W, Zhang H. AI-driven solar energy generation and smart grid integration: a holistic approach to enhancing renewable energy efficiency. Academia Nexus J 2024;3.
[107] Shi H, Fang L, Chen X, Gu C, Ma K, Zhang X, et al. Review of the opportunities and challenges to accelerate mass-scale application of smart grids with large-language models. IET Smart Grid 2024;7:737–59.
[108] Fan H, Chai Y, Liu C, Liu W, Zhang Z, Run W, et al. EV-STLLM: electric vehicle charging forecasting based on spatio-temporal large language models with multi-frequency and multi-scale information fusion. arXiv preprint; 2025. arXiv:250709527.
[109] Han M, Canli I, Shah J, Zhang X, Dino IG, Kalkan S. Perspectives of machine learning and natural language processing on characterizing positive energy districts. Buildings 2024;14:371.
[110] Aghaei R, Kiaei AA, Boush M, Vahidi J, Zavvar M, Barzegar Z, et al. Harnessing the potential of large language models in modern marketing management: applications, future directions, and strategic recommendations. arXiv preprint; 2025. arXiv:250110685.
[111] Gao M, Zhou S, Gu W, Wu Z, Hu Z, Zhu H, et al. LFPLM: a general and flexible load forecasting framework based on pre-trained language model. arXiv e-prints; 2024. arXiv–2406.
[112] Hu Y, Kim H, Ye K, Lu N. Applying fine-tuned LLMs for reducing data needs in load profile analysis. Appl Energy 2025;377:124666.
[113] Wu Y, Xue W. Data-driven weather forecasting and climate modeling from the perspective of development. Atmosphere 2024;15:689.
[114] Mammadov K, Novruzova I, Hasanova N, Mammadov S. Survey of machine learning applications in wind power implementation: a comprehensive view. IEEE 2024:1–8.
[115] Su J, Jiang C, Jin X, Qiao Y, Xiao T, Ma H, et al. Large language models for forecasting and anomaly detection: a systematic literature review. arXiv preprint; 2024. arXiv:240210350.
[116] Aghaei R, Kiaei AA, Boush M, Vahidi J, Barzegar Z, Rofoosheh M. The potential of large language models in supply chain management: advancing decision-making, efficiency, and innovation. arXiv preprint; 2025. arXiv:250115411.
[117] Cheng Y, Zhao H, Zhou X, Zhao J, Cao Y, Yang C, et al. A large language model for advanced power dispatch. Sci Rep 2025;15:8925.
[118] Zheng S, Pan K, Liu J, Chen Y. Empirical study on fine-tuning pre-trained large language models for fault diagnosis of complex systems. Reliab Eng Syst Saf 2024;252:110382.
[119] Marfoglia A, Nardini F, Arcobelli VA, Moscato S, Mellone S, Carbonaro A. Towards real-world clinical data standardization: a modular FHIR-driven transformation pipeline to enhance semantic interoperability in healthcare. Comput Biol Med 2025;187:109745.
[120] Esperança M, Freitas D, Paixão PV, Marcos TA, Martins RA, Ferreira JC. Proactive complaint management in public sector informatics using AI: a semantic pattern recognition framework. Appl Sci 2025:15. 2076-3417.
[121] Zhao Y, Chen G, Liu J. Polymer data challenges in the AI era: bridging gaps for next-generation energy materials. arXiv preprint; 2025. arXiv:250513494.
[122] Atlam HF. LLMs in cyber security: bridging practice and education. Big Data and Cognitive Computing 2025;9:184.
[123] Yu R, Chen S, Xie Y, Yao H, Willard J, Jia X. Foundation models for environmental science: a survey of emerging frontiers. arXiv preprint; 2025. arXiv:250404280.
[124] Santos A, Pena EH, Lopez R, Freire J. Interactive data harmonization with LLM agents: opportunities and challenges. 2025.
[125] Li M, Chen H, Wang Y, Zhu T, Zhang W, Zhu K, et al. Understanding and mitigating the bias inheritance in llm-based data augmentation on downstream tasks. arXiv preprint; 2025. arXiv:250204419.
[126] Wiest IC, Wolf F, Leßmann M-E, van Treeck M, Ferber D, Zhu J, et al. A software pipeline for medical information extraction with large language models, open source and suitable for oncology. npj Precis Oncol 2025;9:313.
[127] Schoenegger P, Park PS, Karger E, Trott S, Tetlock PE. Ai-augmented predictions: llm assistants improve human forecasting accuracy. ACM Trans Interact Intell Syst 2025;15:1–25.
[128] Aldahiri A, Alrashed B, Hussain W. Trends in using IoT with machine learning in health prediction system. Forecasting 2021;3:181–206.
[129] Dhawas P, Dhore A, Bhagat D, Pawar RD, Kukade A, Kalbande K. Big data preprocessing, techniques, integration, transformation, normalisation, cleaning, discretization, and binning. In: Big data analytics techniques for market intelligence. IGI Global Scientific Publishing; 2024. p. 159–82.
[130] Nandan Prasad A. Data quality and preprocessing. Introduction to data governance for machine learning systems: fundamental principles, critical practices, and future trends. Springer; 2024. p. 109–223.
[131] Pereira A, Hovstø A, Rodrigues E, Silvia F, Suciu G, AL-Naday M, et al. Edge driven digital twins in distributed energy systems: role and opportunities for hybrid data driven solutions. 2024.
[132] He Q, Wang J, Shi R, He Y, Wu M. Enhancing renewable energy certificate transactions through reinforcement learning and smart contracts integration. Sci Rep 2024;14:10838.
[133] Elouataoui W. AI-Driven frameworks for enhancing data quality in big data ecosystems: error_detection, correction, and metadata integration. 2024. arXiv preprint arXiv:240503870.
[134] Gangwal A, Ansari A, Ahmad I, Azad AK, Sulaiman WMAW. Current strategies to address data scarcity in artificial intelligence-based drug discovery: a comprehensive review. Comput Biol Med 2024;179:108734.
[135] Guo Z, Tang F, Luo L, Zhao M, Kato N. A survey on applications of large language model-driven digital twins for intelligent network optimization. IEEE Commun Surv Tutor 2025;28:3388–411.
[136] Barbierato E, Gatti A. The challenges of machine learning: a critical review. Electronics 2024;13:416.
[137] Raiaan MAK, Mukta MSH, Fatema K, Fahad NM, Sakib S, Mim MMJ, et al. A review on large language models: architectures, applications, taxonomies, open issues and challenges. IEEE Access 2024;12:26839–74.
[138] Mumuni F, Mumuni A. Explainable artificial intelligence (XAI): from inherent explainability to large language models. arXiv preprint; 2025. arXiv:250109967.
[139] Zhao H, Yang F, Shen B, Lakkaraju H, Du M. Towards uncovering how large language model works: an explainability perspective. arXiv preprint; 2024. arXiv:240210688.
[140] Some L, Yang W, Bain M, Kang B. A comprehensive survey on integrating large language models with knowledge-based methods. arXiv preprint; 2025. arXiv:250113947.
[141] Kokott S, Marek A, Merz F, Karpov P, Carbogno C, Rossi M, et al. Towards efficient and accurate input for data-driven materials science from large-scale all-electron density functional theory (DFT) simulations. Model Simulat Mater Sci Eng 2024;32:28–31.
[142] Ye W, Zhang Y, Yang W, Tang L, Cao D, Cai J, et al. Beyond forecasting: compositional time series reasoning for end-to-end task execution. 2024. arXiv preprint arXiv:241004047.
[143] Ukoba K, Olatunji KO, Adeoye E, Jen T-C, Madyira DM. Optimizing renewable energy systems through artificial intelligence: review and future prospects. Energy Environ 2024:0958305X241256293.
[144] Sakshi TM, Tyagi P, Jain V. Emerging trends in hybrid information systems modeling in artificial intelligence. Hybrid Inform Syst: Non Lin Optim Strat Artif Intell 2024:115.
[145] Lukianykhin O, Shendryk V, Shendryk S, Malekian R. Promising AI applications in power systems: explainable AI (XAI), transformers, LLMs. Springer; 2024. p. 66–76.
[146] Ferdaus MM, Dam T, Sarkar MR, Uddin M, Anavatti SG. Foundation models for clean energy forecasting: a comprehensive review. 2025. arXiv preprint arXiv:250723147.
[147] Sirri N, Guyeux C. Factor analysis of weather conditions impact on firefighter interventions. Springer; 2024. p. 267–82.
[148] Balzan F, Santos PP, Gabbrielli M, Albarracin M, Lopes M. A computational model of inclusive pedagogy: from understanding to application. 2025. arXiv preprint arXiv:250502853.

[149] Matthew B. Interpretable attention mechanisms in human activity recognition: visualizing and understanding MSANet's decision process. 2025.
[150] Mathew TA. Enhancing data platform observability with AI-driven metadata analytics. World J Adv Eng Technol Sci 2025;15:39–47.
[151] Jia L, Yen N, Pei Y. A physics-coupled deep learning framework for hydrodynamic diffusion modeling in watershed systems: integrating spatiotemporal networks and environmental constraints. IEEE Access 2025;13:34985–5003.
[152] Bai G, Chai Z, Ling C, Wang S, Lu J, Zhang N, et al. Beyond efficiency: a systematic survey of resource-efficient large language models. 2024. arXiv preprint arXiv:240100625.
[153] Chen D, Youssef A, Pendse R, Schleife A, Clark BK, Hamann H, et al. Transforming the hybrid cloud for emerging AI workloads. 2024. arXiv preprint arXiv:241113239.
[154] Wu C-J, Acun B, Raghavendra R, Hazelwood K. Beyond efficiency: scaling ai sustainably. IEEE Micro; 2024.
[155] Nguyen S, Zhou B, Ding Y, Liu S. Towards sustainable large language model serving. ACM SIGENERGY Energy Inform Rev 2024;4:134–40.
[156] Gogineni K, Suvizi A, Venkataramani G. LLMs on a budget: system-level approaches to power-efficient and scalable fine-tuning. IEEE Open J Comput Soc 2025;6:987–1000.
[157] Xu W, Chen J, Zheng P, Yi X, Tian T, Zhu W, et al. Deploying foundation model powered agent services: a survey. 2024. arXiv preprint arXiv:241213437.
[158] Beena B, Prashanth C, Manideep TSS, Saragadam S, Karthik G. A green cloud-based framework for energy-efficient task scheduling using carbon intensity data for heterogeneous cloud servers. IEEE Access 2025;13:73916–38.
[159] Kumar M, Kaur G, Rana PS. Performance, portability, productivity, and security in HPC cloud: a systematic literature review. J Supercomput 2025;81:1197.
[160] Devanathan B, Jnana Varshitha K, Pavan Kumar L, Lakshmanan SA, Krishna Prakash N. Explainable AI framework using XGBoost with SHAP and LIME for multi-scale household energy forecasting. IEEE Access 2025;13:149750–64.
[161] Varshitha KJ, Kumar LP, Manellore P, Devanathan B, S AL, Prakash NK. Explainable AI-Enhanced energy forecasting using LightGBM with SHAP and LIME interpretability. 2025. p. 369–73.
[162] Khan MA, Zulcaffle TMA, Ahmad M, Alkhatib AAT. Interpretable hybrid deep learning for real-time multi-horizon forecasting of load, solar, and wind generation using multi-source energy and weather data. J Comput Biomed Informatics 2025;10.
[163] Zaman Z, Gauravaram P, Jha S, Hu W. Bloom-LLM: privacy-preserving large language model for load forecasting. 2025. p. 128–38.
[164] Zarghami M, Niknam T, Aghaei J, Nezhad AH. Concurrent PV production and consumption load forecasting using CT-Transformer deep learning to estimate energy system flexibility. IET Renew Power Gener 2024;18:2139–61.
[165] Li Y, Hu W, Zhang F, Li Y. Multi-objective collaborative operation optimization of park-level integrated energy system clusters considering green power forecasting and trading. Energy 2025;319:135055.
[166] Yu C, Yan G, Yu C, Liu X, Mi X. MRIformer: a multi-resolution interactive transformer for wind speed multi-step prediction. Inf Sci 2024;661:120150.
[167] Qu K, Si G, Shan Z, Kong X, Yang X. Short-term forecasting for multiple wind farms based on transformer model. Energy Rep 2022;8:483–90.
[168] Rao S, Neethirajan S. Computational architectures for precision dairy nutrition digital twins: a technical review and implementation framework. Sensors 2025;25:4899.
[169] Khatiwada K, Hopper J, Cheatham J, Joshi A, Baidya S. Large language models in the IoT Ecosystem–A survey on security challenges and applications. 2025. arXiv preprint arXiv:250517586.
[170] Beridze T, Lomidze G. A policy-centered framework for cybersecurity management: ensuring information assurance through governance and oversight. Int J Adv Comput Method Emerg Technol 2024;14:1–13.
[171] Oyekan M, Enyejo JO. Harnessing data analytics to maximize renewable energy asset performance. International Journal of Scientific Research and Modern Technology 2023;2:64–80.
[172] Russell-Gilbert A, Sommers A, Thompson A, Cummins L, Mittal S, Rahimi S, et al. AAD-LLM: adaptive anomaly detection using large language models. 2024. p. 4194–203.
[173] Howlader AS. Machine learning applications in renewable energy: predictive analytics for solar cell performance optimization and energy yield forecasting. Am J Sch Res Innovat 2025;4:392–427.
[174] Piotrowski P, Rutyna I, Baczyński D, Kopyt M. Evaluation metrics for wind power forecasts: a comprehensive review and statistical analysis of errors. Energies 2022;15:9657.
[175] Ngo D, Park H-C, Kang B. Edge intelligence: a review of deep neural network inference in resource-limited environments. Electronics 2025;14:2495.
[176] Kumar PV, Kulkarni A, Mendhe D, Keshar DK, Babu SBT, Rajesh N. AI-Optimized hardware design for internet of things (IoT) devices. IEEE 2024:21–6.
[177] Ding N, Ouyang X, Gao L, Huang J, Xing G. An overview on economic analysis of Internet of Everything. IEEE Commun Surv Tutor 2025;27(6):3742–71.
[178] Li Y, Guo X, Gao J, Chen G, Zhao X, Zhang J, et al. LLMs trust humans more, that'sa problem! unveiling and mitigating the authority bias in retrieval-augmented generation. 2025. p. 28844–58.
[179] Dimitrova M. Retrieval-augmented generation (RAG): advances and challenges. Probl Eng Cybern Robot 2025;83:32–57.
[180] Kull K, Khan MA, Asad B, Naseer MU, Kallaste A, Vaimann T. Adaptive deep belief networks and LightGBM-Based hybrid fault diagnostics for SCADA-managed PV systems: a real-world case study. Electronics 2025;14:3649.
[181] Ahmed R, Sreeram V, Mishra Y, Arif M. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. Renew Sustain Energy Rev 2020;124:109792.
[182] Eren Y, Küçükdemiral İ. A comprehensive review on deep learning approaches for short-term load forecasting. Renew Sustain Energy Rev 2024;189:114031.
[183] Yan Z, Zhenyuan D, Xu Y, Zhou Z. Probabilistic PV power forecasting by a multi-modal method using GPT-agent to interpret weather conditions. IEEE 2024:1–6.
[184] Cherian A, Corcodel R, Jain S, Romeres D. Llmphy: complex physical reasoning using large language models and world models. 2024. arXiv preprint arXiv:241108027.
[185] Guido R, Ferrisi S, Lofaro D, Conforti D. An overview on the advancements of support vector machine models in healthcare applications: a review. Information 2024;15:235.
[186] Andrews S, Cathcart J, Chen A, Cornec H, Kumar S, Majholm J, et al. Understanding contrail management: opportunities, challenges, and insights. RMI Contrail Impact Task Force; 2024.
[187] Safaei Pirooz AA, Sanjari MJ, Kim Y-J, Moore S, Turner R, Weaver WW, et al. Adaptation of high spatio-temporal resolution weather/load forecast in real-world distributed energy-system operation. Energies 2023;16:3477.
[188] Perera SM, Putrus G, Conlon M, Narayana M, Sunderland K. Wind energy harvesting and conversion systems: a technical review. Energies 2022;15:9299.
[189] Barhmi K, Heynen C, Golroodbari S, van Sark W. A review of solar forecasting techniques and the role of artificial intelligence. MDPI; 2024. p. 99–135.
[190] Zhuang D, Gan VJ, Tekler ZD, Chong A, Tian S, Shi X. Data-driven predictive control for smart HVAC system in IoT-integrated buildings with time-series forecasting and reinforcement learning. Appl Energy 2023;338:120936.
[191] Xu E, Wang H, Zhang Y, Li S, Tang Y, Zhou Z, et al. A survey of physics-informed AI for complex urban systems. 2025. arXiv preprint arXiv:250613777.
[192] Emmanouilidis C, Waschull S, Zotelli J. Integrating knowledge and data-driven artificial intelligence for decisional enterprise interoperability. Springer; 2024. p. 372–98.
[193] Olivia D, Khan Z, Shetty S. A machine learning and explainable artificial intelligence approach for insurance fraud classification. Intel Artif 2025;28:140–69.
[194] Güven AF, Yörükeren N, Mengi OÖ. Multi-objective optimization and sustainable design: a performance comparison of metaheuristic algorithms used for on-grid and off-grid hybrid energy systems. Neural Comput Appl 2024;36:7559–94.
[195] Luo L, Wei H, Lin Z, Wu J, Wang W, Sun Y. Multi-objective optimal energy-efficient retrofit determination using hybrid urban building energy model. Considering uncertainties between models. Springer; 2025. p. 183–206.
[196] Liang Z, Liang Y, Luo X, Yu Z, Chen J, Chen Y. Multi-objective optimization of proton exchange membrane fuel cell based methanol-solar-to-X hybrid energy systems. Appl Energy 2024;373:123828.
[197] de Carvalho Souza ME, Weigang L. Unveiling the black box: the significance of XAI in making LLMs transparent. INTELIGENCIA ARTIFICIAL 2025;2:1.
[198] Mohammadi H, Bagheri A, Giachanou A, Oberski DL. Explainability in practice: a survey of explainable NLP across various domains. arXiv preprint; 2025. arXiv:250200837.
[199] Ali SS, Ali M, Bhatti DMS, Choi BJ. Explainable clustered federated learning for solar energy forecasting. Energies 2025;18:2380.
[200] Barman KG, Caron S, Sullivan E, de Regt HW, de Austri RR, Boon M, et al. Large physics models: towards a collaborative approach with large language models and foundation models. 2025. arXiv preprint arXiv:250105382.
[201] Cohrs K-H, Diaz E, Sitokonstantinou V, Varando G, Camps-Valls G. Large language models for causal hypothesis generation in science. Mach Learn: Sci Technol 2025;6:013001.
[202] Li D, Jiang B, Huang L, Beigi A, Zhao C, Tan Z, et al. From generation to judgment: opportunities and challenges of llm-as-a-judge. 2024. arXiv preprint arXiv:241116594.
[203] Majumder S, Dong L, Doudi F, Cai Y, Tian C, Kalathil D, et al. Exploring the capabilities and limitations of large language models in the electric energy sector. Joule 2024;8:1544–9.
[204] Hamann HF, Brunschwiler T, Gjorgiev B, Martins LS, Puech A, Varbella A, et al. A perspective on foundation models for the electric power grid. 2024. arXiv preprint arXiv:240709434.
[205] Das O, Zafar MH, Sanfilippo F, Rudra S, Kolhe ML. Advancements in digital twin technology and machine learning for energy systems: a comprehensive review of applications in smart grids, renewable energy, and electric vehicle optimisation. Energy Convers Manag X 2024:100715.
[206] Farea A, Yli-Harja O, Emmert-Streib F. Understanding physics-informed neural networks: techniques, applications, trends, and challenges. AI, 5; 2024. p. 1534–57.
[207] Baisthakur S, Fitzgerald B. Physics-informed neural network surrogate model for bypassing blade element momentum theory in wind turbine aerodynamic load estimation. Renew Energy 2024;224:120122.

[208] Liao W, Wang S, Yang D, Yang Z, Fang J, Rehtanz C, et al. TimeGPT in load forecasting: a large time series model perspective. Appl Energy 2025;379:124973.
[209] Yu J, Shvetsov AV, Alsamhi SH. Leveraging machine learning for cybersecurity resilience in Industry 4.0: challenges and future directions. IEEE Access 2024;12: 159579–96.
[210] Ajimon ST, Kumar S. Applications of LLMs in quantum-aware cybersecurity leveraging LLMs for real-time anomaly detection and threat intelligence. Leveraging large language models for quantum-aware cybersecurity. IGI Global Scientific Publishing; 2025. p. 201–46.
[211] Palma G, Cecchi G, Rizzo A. Large language models for predictive maintenance in the leather tanning industry: multimodal anomaly detection in compressors. Electronics 2025;14:2061.
[212] Girija SS, Kapoor S, Arora L, Pradhan D, Raj A, Shetgaonkar A. Optimizing LLMs for resource-constrained environments: a survey of model compression techniques. 2025. arXiv preprint arXiv:250502309.
[213] Sadeepa S, Kavinda K, Hashika E, Sandeepa C, Gamage T, Liyanage M. DisLLM: distributed LLMs for privacy assurance in resource-constrained environments. IEEE 2024:1–9.