Artificial Intelligence Applications in Energy Systems: A Comprehensive Review

Tao Huang, Jingsi Chen, Jun Yin, and Michela Meo

***Abstract*—The global energy sector is transforming under decarbonization, decentralization, and digitalization, introducing complexity and volatility into system operations, and challenging the limits of traditional analytical methods. Artificial intelligence (AI) encompassing machine learning, deep learning, and reinforcement learning has emerged as a critical enabler for building the next generation of energy infrastructure. This review provides a systematic assessment of AI's role across key domains including forecasting, optimization, control, and maintenance. It also examines the emerging potential of large language models from several application perspectives. Our analysis identifies a clear trend from data-driven models toward more sophisticated hybrid and physics-informed architectures. We document this progression through specific advances in forecasting electricity load, renewable generation, and electricity price, in optimizing grid scheduling, energy storage, and demand-side resources, and in enhancing system resilience through intelligent control and maintenance. Beyond current applications, we discuss future research challenges: developing trustworthy AI for high-stakes environments, creating scientific AI that integrates physical principles, managing complex sector-coupled systems, and progressing toward generalized energy foundation models. Overall, this review offers a structured understanding of the synergy between AI and energy systems, outlining current capabilities and the governance required for trustworthy and equitable deployment.**

***Index Terms*—Artificial intelligence, energy systems, machine learning, deep learning, reinforcement learning, large language models.**

ABBREVIATION

AI	artificial intelligence
ML	machine learning
SVM	support vector machines
RF	random forests
ANN	artificial neural networks
DL	deep learning
CNN	convolutional neural networks
LSTM	long short-term memory
NLP	natural language processing
RL	reinforcement learning
DQN	deep Q-networks
DDPG	deep deterministic policy gradient
MARL	multi-agent reinforcement learning
GNNs	graph neural networks
PINNs	physics-informed neural networks
FL	federated learning
LLMs	large language models
GAIA	grid artificial intelligent assistant
ARIMA	autoregressive integrated moving average
BiLSTM	bidirectional LSTM
RMSE	root mean square error
MAPE	mean absolute percentage error
PSO	particle swarm optimization
MHSA	multi-head self-attention
CAM	channel attention modules
LSCAM	long-short term cross attention mechanism
PV	photovoltaic
WPCA	feature-weighted principal component analysis
GRU	gated recurrent unit
3D	three-dimensional
EWPF-Transformer	enhanced wind power forecasting transformer
GAT	graph attention network
GFST-WSF	graph-attentive frequency-enhanced spatial-temporal wind speed forecasting
KDE	kernel density estimation
OPF	optimal power flow
SARIMA	seasonality ARIMA
ARIMAX	exogenous ARIMA
PGDM	pattern-guided diffusion model
DNNs	distributional neural networks
LASSO	least absolute shrinkage and selection operator
VMD	variational mode decomposition
EWT	empirical wavelet transform

Manuscript received 19 September, 2025; revised 7 October, 2025; accepted 24 October, 2025. *(Corresponding author: Tao Huang, e-mail: tao.huang@polito.it)*

T. Huang, J. Chen are with the Department of Energy, Politecnico di Torino, Italy.

J. Yin, M. Meo are with the Department of Electronics and Telecommunications, Politecnico di Torino, Italy.

Color versions of one or more of the figures in this paper are available online at https://doi.org/10.26599/PEF.2025.9650001.

Digital Object Identifier 10.26599/PEF.2025.9650001

MRC	multi resolution convolution
pTFT	penalized temporal fusion transformer
PMU	phasor measurement unit
CCMA	centrally coordinated multi-agent
GA	genetic algorithm
PSO	particle swarm optimization
PSS	power system stabilizers
GC-PPO	graph convolutional proximal policy optimization
BESS	battery energy storage systems
MILP	mixed-integer linear programming
MDP	Markov decision process
SoC	state-of-charge
SSD	second-order stochastic dominance
CvaR	conditional value-at-risk
BRPs	balanced responsible parties
TSO	transmission system operator
FCAS	frequency control ancillary services
FFR	fast frequency response
MIQCP	mixed-integer quadratically constrained program
xAI	explainable AI
IO	inverse optimization
MADRL	multi-agent deep reinforcement learning
RTP	real-time pricing
DRL	deep RL
SCDL	sub-customer directrix load
OLTC	on-load tap changer
AVC	active voltage control
Dec-POMDP	decentralized partially observable Markov decision process
CMDP	constrained Markov decision process
NREL	National Renewable Energy Laboratory
DOE	Department of Energy
EFC	emergency frequency control
RES	renewable energy source
LFC	load-frequency control
PI	proportional–integral
DGA	dissolved gas analysis
DeDn-CNN	denoising convolutional neural network
WAMS	wide-area monitoring systems
RAG	retrieval-augmented generation
SCADA	supervisory control and data acquisition
PNNL	Pacific Northwest National Laboratory

I. Introduction

A. Energy Transition Challenges and Opportunities

With the carbon footprint of energy systems remaining daunting, global energy-related CO_2 emissions hit a record 37.2 Gt in 2023, about 50% higher than that at the turn of the century [1]. This situation underscores the urgency for transforming traditional power grids with fossil fuels into modern energy systems with sustainable sources [2], [3]. The modern energy system is an advanced digital network integrated with unprecedented levels of renewable energy like solar and wind, offering greater flexibility, quicker demand response, and environmentally friendly [4], [5]. However, the fluctuation, intermittent, and stochastic nature of renewable energy also introduces variability and complexity in grid operation [6], [7]. This variability challenges the effectiveness of conventional scheduling and control methods, which are often ill-suited for real-time balancing under conditions of rapid change. Therefore, new strategies are required to maintain supply and demand equilibrium, ensure grid reliability, and stabilize increasingly complex power systems [8], [9].

In this context, AI has emerged as a key enabler. By enhancing forecasting, optimizing operational decisions, and supporting intelligent control, AI methods offer pathways to manage variability, improve efficiency, and unlock new forms of flexibility across the energy value chain. These capabilities are particularly critical in supporting the secure and cost-effective integration of high levels of renewable energy [10], [11].

B. Overview of AI Technologies

AI has evolved along a clear methodological trajectory in energy systems, reflecting increasing capacity to model complexity, uncertainty, and interdependence.

The first wave consisted of traditional ML techniques such as SVM, RF, and shallow ANN. These methods established the feasibility of data-driven forecasting and anomaly detection, outperforming classical statistical baselines but requiring extensive feature engineering and often lacking robustness under rapidly changing conditions [12].

The second wave was marked by the dominance of DL [13], [14]. CNN, particularly LSTM architectures, enabled the capture of long-range temporal dependencies and nonlinear dynamics in load, renewable generation, and price forecasting [15], [16]. To further enhance temporal modeling, attention mechanisms were introduced as add-on modules to recurrent or convolutional structures, allowing models to dynamically weight the relative importance of different time steps and exogenous features. Attention-enhanced CNN-LSTM hybrids have demonstrated improved robustness and accuracy in multivariate forecasting tasks where weather, calendar, and market drivers interact [17], [18], [19], [20]. Building on this trend, transformer-based architectures, originally designed for NLP, have demonstrated superior accuracy in energy time-series forecasting by leveraging self-attention to capture global patterns without recurrence, enabling scalable modeling of long sequences [21].

The third wave centered on RL, which addresses sequential decision-making problems. RL agents have been successfully deployed in simulations of microgrids, storage dispatch, and demand response, learning near-optimal control policies via trial-and-error interaction with dynamic environments. A key milestone was the introduction of DQN, which combine Q-learning with deep neural networks to approximate value functions, enabling scalable control in high-dimensional energy scheduling tasks. Building on this, DDPG extended RL to continuous action spaces, making it suitable for fine-grained control such as inverter reactive power regulation and battery charge/discharge optimization. More recently, MARL has further extended these capabilities to distributed coordination, enabling decentralized optimization of heterogeneous energy resources [22], [23], [24], [25], [26].

Building on this, physics-informed and hybrid AI approaches have emerged as the fourth wave. These models embed physical laws and operational constraints directly into learning architectures, reducing reliance on large historical datasets and ensuring reliability under rare or extreme events. GNNs and PINNs exemplify this trend, capturing topological dependencies and physical consistency in power system modeling [27], [28], [29], [30].

The fifth wave reflects the decentralization of grids, where FL and edge AI enable privacy-preserving, low-latency learning at substations, microgrids, and distribution endpoints. These methods support local autonomy while maintaining coordination through federated optimization and edge offloading [31], [32], [33], [34].

Most recently, LLMs and AI agents have emerged. Beyond text processing, LLMs such as GAIA and eGridGPT show promise in dispatch assistance, tool integration, knowledge retrieval, and operator decision support. Benchmarks like ElecBench are beginning to formalize evaluation for these domain-specific LLMs [35], [36], [37], [38], [39].

Looking ahead, future research will move beyond accuracy to embrace trustworthy AI, where explainability, verifiability, and security are treated as prerequisites for deployment. Equally critical is the rise of scientific AI, which integrates physics-informed learning, causal reasoning, and standardized validation frameworks to ensure that models respect physical laws and generalize reliably. At the system level, sector-coupled and multi-agent optimization will be needed to address the curse of dimensionality in coordinating electricity, heating, transport, and mobility across multiple timescales. These advances must be underpinned by holistic benchmarks that evaluate robustness, fairness, resilience, and computational feasibility, shifting evaluation beyond narrow accuracy metrics. Finally, the prospect of energy foundation models and domain-specialized LLM agents points to a longer-term vision of scalable, adaptive platforms capable of unifying forecasting, optimization, and control. As summarized in Fig. 1, the evolution from traditional ML to LLM-based agents illustrates a consistent progression towards greater adaptability, scalability, and integration, underlining AI's potential to handle the high-dimensional data, uncertainty, and non-linear dynamics inherent in modern energy systems more effectively than traditional heuristics or human operators [40].

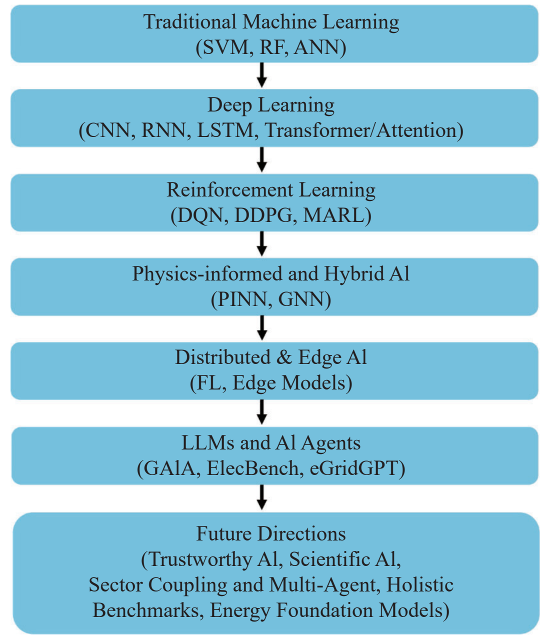


Fig. 1. Evolution of AI methods in energy systems.

C. Objectives and Contributions

The methodological evolution from traditional ML to LLMs is propelled by fundamental shifts in the energy sector's data landscape and operational needs. Although classical ML models excel at structured numerical tasks like time-series forecasting, they are ill-equipped to handle the growing volume of unstructured data, such as technical manuals, maintenance logs, and policy documents. The primary driver for adopting LLMs is their unique ability to perform semantic reasoning on this textual data, creating a holistic understanding of system interdependencies that traditional models cannot [41], [42]. Furthermore, LLMs enable a paradigm shift towards interactive, human-in-the-loop systems, where grid operators can query models in natural language, ask for explanations, and receive co-pilot-like assistance, moving beyond the "black box" nature of earlier AI.

However, the transition from research to reality is impeded by critical non-technical barriers, especially within legacy grid infrastructure. Chief among these are stringent data governance and cybersecurity protocols that restrict the flow of sensitive operational data to external, particularly cloud-based models [40]. Secondly, the deployment of AI in critical infrastructure faces significant regulatory and compliance hurdles, as authorities require absolute proof of reliability and deterministic performance, which can be challenging for probabilistic LLMs [43]. Finally, workforce adoption is hindered by a skill gap and a lack of trust in non-traditional systems, alongside the high cost and complexity of integrating advanced AI with legacy operational technology, representative obstacles to widespread implementation [44].

Faced with both rapid methodological progress and significant implementation barriers, this review aims to bridge the gap by synthesizing recent advances in AI for energy systems. Fig. 2 illustrates a clear thematic classification, covering three primary domains of application: (i) forecasting for electricity load, renewable generation, and electricity price, (ii) optimization for grid scheduling, energy storage, and demand response, (iii) control and maintenance. Next, this review also summarizes six perspectives of LLM applications in energy systems. Within each theme, key subtopics and exemplary techniques are highlighted to illustrate the state-of-the-art. Finally, this review moves beyond applications to discuss cross-cutting challenges and future directions, including trustworthy AI, scientific AI, sector-coupled optimization, holistic benchmarks, and the long-term vision of energy foundation models.

The structure of the review is as follows: Sections II–IV introduce the applications of AI methods in forecasting, optimization, and control and maintenance, respectively. Section V summarizes the applications of LLMs. Section VI explores the future research directions. Section VII concludes the review.

II. AI FOR FORECASTING

A. Electricity Load

Accurate electric load forecasting is essential for efficient and reliable grid operations, particularly with increasing renewable integration. The core challenge in load forecasting is to accurately predict future time-series electricity demand, which can exhibit daily and seasonal patterns as well as randomness due to human activities and weather influences.

Early AI-based load forecasting relied on ANNs, SVMs, and other classical ML models [45], [46]. These methods improved upon traditional ARIMA models by capturing nonlinear relationships and handling higher-dimensional data [47], [48], [49], but often struggled with long-term dependencies and required extensive feature engineering [50], [51]. ANN-based models, especially when hybridized with heuristic optimization, demonstrated robust performance in various settings, but their accuracy was generally surpassed by more advanced deep learning models [52], [53]. Specifically, LSTM architectures have significantly enhanced prediction accuracy by effectively learning complex patterns from historical load data and external influencing factors.

LSTM-based models continue to serve as foundational architectures in load forecasting tasks due to their ability to capture long-term temporal dependencies. For instance, Bohara et al. demonstrated the effectiveness of BiLSTM architectures in a day ahead (24 h) 38 residential homes load forecasting scenarios, achieving approximately 5.6%, 2.85%, and 2.60% reductions in RMSE compared to conventional LSTM, CNN-LSTM, and CNN-BiLSTM models, respectively [54].

To further capture both local and temporal features, CNN-LSTM hybrid architectures combine the local feature extraction capabilities of CNN with the advantages of LSTM in capturing long-term dependencies, making them suitable for multivariate inputs, such as load, weather, and calendar variables [55]. Wang et al. (2024) introduced a CNN combined with multi-layer extended LSTM layers for short-term load forecasting, substantially improving load forecasting accuracy over other DL methods [56]. Zhao et al. (2025) proposed a three-channel CNN-LSTM hybrid, integrating temporal, meteorological, and historical load information through separate LSTM channels, which were then combined using CNN layers. Their approach reduced the MAPE to approximately 0.974%, representing a 23.6% improvement compared to single-channel LSTM models [57].

Additional hybrid models, such as CNN coupled with stacked BiLSTM networks, have been explored in specific scenarios. Akhter et al. (2024) employed this architecture for day-ahead load forecasting in Dhaka, Bangladesh, achieving a MAPE of about 1.64%, outperforming conventional single and hybrid models previously tested on the dataset [58]. Moreover, Jang et al. (2024) performed a comparative study involving multiple single-layer and multi-layer hybrid models, highlighting the overall superior predictive performance of hybrid DL models, especially when multivariate inputs (weather and cal-

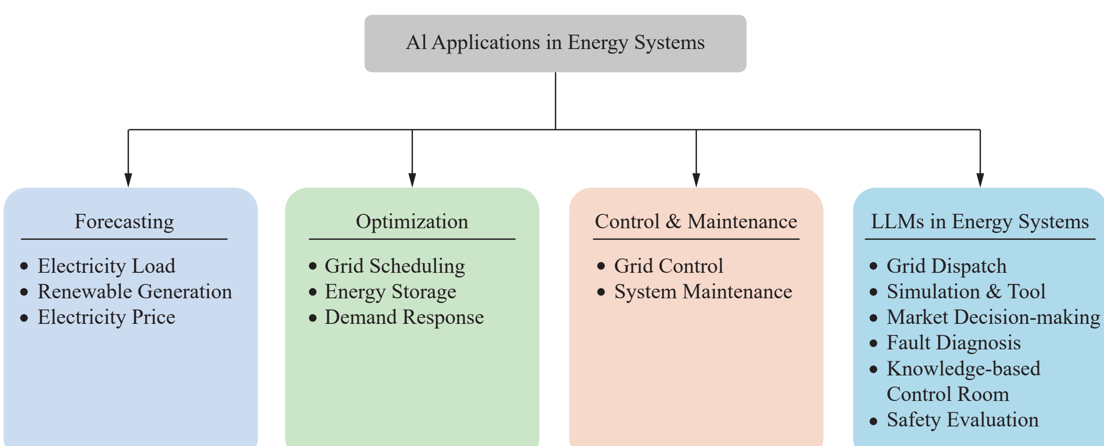


Fig. 2. Overview of AI applications in energy systems.

endar variables) are available [59].

Recently, incorporating attention mechanisms has further enhanced prediction accuracy by dynamically weighting significant features and temporal segments. Quansah and Tenkorang (2023) utilized PSO for hyperparameter tuning combined with CNN-LSTM architectures incorporating MHSA layers, achieving a remarkable MAPE of approximately 1.9376% [60]. Similarly, Guo et al. (2025) designed a complex hybrid CNN-LSTM architecture featuring MHSA and CAM. This sophisticated integration significantly improved forecasting accuracy and robustness compared to traditional CNN-LSTM hybrids [61].

A comparative summary of the evolution of AI methods for electricity load forecasting is provided in Table 1, highlighting their main strengths, weaknesses, and performance improvements.

B. Renewable Generation

In renewable energy forecasting, the inherent intermittency and stochasticity of wind and solar resources have driven a clear progression in AI methodologies from statistical and shallow-learning models to physics-informed deep networks, recurrent and convolutional hybrids, attention-based transformers and graph networks, meta-learning, probabilistic approaches, and finally generative scenario generators [62], [63].

Early work demonstrated that tree-based ensembles and gradient-boosted regressors (e.g., XGBoost) could outperform classical ARIMA/ARMA methods by capturing nonlinear feature interactions between historical power and weather variables, albeit with limited ability to model temporal dependencies beyond a few time steps [64]. Recognizing the value of physical priors, recent studies have embedded governing equations into neural networks. Wang et al. (2025) developed a PINN integrating with a LSCAM as part of the input to predict photovoltaic module temperature and subsequent PV output, outperforming traditional models across different seasons with superior accuracy [65]. Li et al. (2025) applied PINNs to real-time power profiles under gusty conditions, improving prediction accuracy and error bounds during transient events compared with a conventional neural network [66].

Deep recurrent architectures marked the next evolution. Xiao et al. (2023) combined WPCA method with a GRU network, using a PSO algorithm to reduce wind power RMSE by 10%–16% compared to the conventional model [67], while Faruque et al. (2025) fused convolutional layers with dual-layer GRUs and a constraint-aware loss, further tightening very short-term wind forecasts and achieving 3.65% MAPE, outperforming Bi-LSTM and CNN models [68].

To fuse local spatial features with temporal dynamics, hybrid CNN-RNN architectures emerged. Belletreche et al. (2024) introduced a Conv-dual attention LSTM, combining meteorological data via CNN layers with LSTM sequences and dual attention heads, achieving state-of-the-art wind forecast accuracy in desert-region datasets [69]. Similarly, Huang et al. (2023) developed a three-dimensional CNN-GRU model that embeds spatial correlations among neighboring turbines into a 3D tensor, outperforming GRU models by 9.41% in RMSE on the spatial dynamic wind power dataset [70].

Attention mechanisms and transformer architectures then emerged to model long-range dependencies without recurrence. Sarkar et al. (2023) introduced the EWPF-Transformer,

TABLE 1
AI APPLICATIONS FOR ELECTRICITY LOAD FORECASTING

Model Used	Key Advantages	Weaknesses	Performance
ARIMA [48]	Robust under noisy input	Purely linear, no nonlinear learning	Stable, weaker than DL
SARIMA [49]	Good seasonality modeling	Fails for nonlinear/complex data	Baseline model
SVM + ANN hybrid [46]	Hybrid improves short-term load prediction	Early stage, limited scalability	Outperforms single SVM or ANN
ANN, SVM [45]	Captures nonlinearities compared two ML baselines	Needs feature engineering, limited generalization	ANN/SVM better than ARIMA
Iterative memory-driven NN [50]	Learns non-trend fluctuation	Complex, data intensive	Accuracy improved compared to RNN
Highway Network (DL) [51]	Cross-feature collaboration	Requires large datasets	Better long-term online prediction
RNN, ANN [52]	Captures sequential patterns	Gradient vanishing, overfitting risk	Improved over ANN
BiLSTM, CNN-BiLSTM [54]	Bidirectional sequence learning	Overfitting risk	RMSE is 5.6%, 2.85%, and 2.60% lower than LSTM/CNN-LSTM
ES-dRNN (Hybrid) [55]	Combines exponential smoothing + dilated RNN	High complexity	Strong robustness
CNN + extended LSTM [56]	Extracts local + long term dependencies features	Heavy model size	Accuracy improved compared to single CNN or LSTM
3-Channel CNN-LSTM [57]	Fuses time, weather, and historical data	Needs multivariate datasets	MAPE reaches 0.974% compared to 1.522% on LSTM
CNN + stacked BiLSTM [58]	Captures deep temporal dependencies	Data hungry	MAPE reaches 1.64%, which is the best in Dhaka case
CNN-LSTM + MHSA (PSO tuned) [60]	Dynamic feature weighting, attention	Complex hyperparameter tuning	MAPE reaches the lowest compared to other algorithms
CNN-LSTM + MHSA + GAM + CAM [61]	Multi-modal learning, channel attention	Computationally heavy	Better generalization compared to CNN-LSTM

which used multi-head self-attention to learn global patterns and achieved 5%–20% accuracy gains over LSTM and GRU [71]. Liu et al. (2023) hybridized a GAT for turbine-level spatial relations with a temporal transformer encoder in their GFST-WSF model, delivering the best 6–24-h wind speed forecasts by jointly modeling inter-turbine and temporal dependencies [72]. Alongside transformers, GNNs proved effective. He et al. (2025) built a GNN model combined with ramp feature recognition and error correction mechanisms, achieving an approximate 28% improvement compared to traditional models in Belgian wind farm datasets [73].

Meta-learning and few-shot methods have addressed data scarcity at new sites. Chen et al. (2024) presented WG-Reptile, a meta-learning algorithm for ultra-short-term wind power forecasting that adapts pre-trained models to new farms with minimal data, achieving accuracy comparable to fully trained models [74]. Boutahir et al. (2024) extended this idea to solar domains, showing that meta-learning notably improves model generalization in limited-data regimes [75].

In solar forecasting, convolutional models have likewise advanced performance. Ladjal et al. (2025) evaluated four solar radiance predictors, including CNN-LSTM hybrids, and found that the hybrid outperformed pure CNN and pure RNN on multi-hour horizons [76], while Al-Ali (2023) combined CNN, LSTM, and a Transformer to push PV power forecast and achieved highest accuracy compared to LSTM-CNN models on the Fingrid dataset [77]. Beyond point forecasts, probabilistic forecasting techniques have matured in parallel. Bai et al. (2021) coupled a ConvLSTM network with KDE to produce full predictive distributions, yielding reliable confidence intervals and improve forecast accuracy by more than 20% compared to conventional models [78].

Finally, generative AI methods now enable scenario generation for stochastic planning. Zhang et al. (2020) trained a conditional Wasserstein GAN on wind forecast residuals to produce realistic power trajectories for Monte Carlo simulations [79], while Dong et al. (2025) proposed a diffusion-based model to generate controllable renewable energy scenarios capturing manifold space feature distributions [80]. Variational autoencoder-GAN hybrids have also been employed to produce wind and solar real-time series data for probabilistic OPF studies [81].

A comparative summary of the evolution of AI methods for renewable generation forecasting is provided in Table 2.

TABLE 2
AI APPLICATIONS FOR RENEWABLE GENERATION FORECASTING

Model Used	Key Advantages	Weaknesses	Performance
XGBoost (tree ensemble) [64]	Captures nonlinear weather–power interactions, faster training	Limited temporal dependency modeling	Better accuracy than ARIMA/ARMA, efficient probabilistic forecasts
WPCA + PSO + GRU [67]	Combines feature reduction + optimized GRU	Model complexity, depends on feature selection	RMSE reduces 10%–16% compared to conventional GRU
CNN + dual GRU + constraint-aware loss [68]	Integrates spatio-temporal info, physics constraints	Requires tuning, heavier computation	Achieved 3.65% MAPE, better than BiLSTM/CNN
3D CNN-GRU [70]	Captures spatial-temporal features	Needs large 3D tensor dataset	RMSE decreases 9.41% compared to GRU
Conv-Dual Attention LSTM [69]	Fuses meteorological data	High computation cost	State-of-the-art desert-region accuracy
CNN-LSTM hybrid [76]	Captures temporal + spatial solar dynamics	More complex than CNN/RNN alone	Outperformed CNN/RNN in solar DNI forecast
ConvLSTM + KDE [78]	Produces probabilistic intervals	Needs KDE tuning	More than 20% accuracy gain, reliable confidence bounds
Transformer (EWPF) [71]	Captures long-range dependencies, global patterns	Training cost, requires large data	5%–20% accuracy gains over LSTM/GRU
CNN + LSTM + Transformer [77]	Combines convolution, recurrence, and attention	Higher computation	Best accuracy on Fingrid PV dataset
Transformer + Graph Attention Network (GFST-WSF) [72]	Models turbine-level spatial + temporal dependencies	Complexity of graph + transformer	Best 6–24-h forecasts among benchmarks
GCN + ramp/error correction [73]	Ramp detection + error correction, robust	Needs high-quality SCADA data	~ 28% improvement compared to traditional wind models
PINN (thermal modeling for borehole storage) [66]	Real-time physical consistency, better transient accuracy	Needs detailed system parameters	Results outperform NN
PINN + LSCAM [65]	Embeds physical priors (temperature–PV link), improved generalization	Complexity, requires physics knowledge	Outperformed classical PV models across seasons
Meta-learning (WG-Reptile) [74]	Few-shot adaptation for new wind farms	Relies on pre-trained models	Comparable accuracy to fully trained models
Meta-learning + Voting Regressor (solar) [75]	Improves generalization with limited data	Sensitive to pre-training choice	Better robustness in solar radiation
Conditional Wasserstein GAN [79]	Generates realistic stochastic scenarios	Training instability	Provides Monte Carlo-ready trajectories
PGDM [80]	Captures manifold distributions, controllable	Very high compute demand	Realistic renewable scenarios for planning
Improved VAEGAN [81]	Generates renewable scenarios with interpretable features	GAN training instability	Effective for OPF scenario generation

C. Electricity Price

In recent years, classical econometric and statistical models such as ARIMA, SARIMA, and ARIMAX inputs remained the baseline for short-term electricity price forecasting due to their transparency and interpretability [82]. However, as energy markets grew more volatile and external features (e.g., weather, demand, and renewables) became central, machine learning methods like random forests, support vector regression, and gradient boosting began to outperform traditional models, offering greater flexibility for modelling complex, non-linear relationships.

Building on that, Marcjasz et al. (2023) developed DNNs with probabilistic outputs, which replace single-point predictions with full predictive distributions, e.g., normal or Johnson's SU, a critical advance for risk-aware applications in markets such as Germany's day-ahead market, where the model significantly beats LASSO and quantile regression averaging approaches [83].

Building on probabilistic forecasting outputs, the hybrid sequential deep architectures have emerged. Notably, Hajigholam et al. (2024) introduced a hybrid architecture combining Bi-LSTM followed by GRU layers, where bidirectional LSTM first captures past-and-future context and GRU refines it, applied to New York City day-ahead prices. This Bi-LSTM + GRU hybrid outperformed ARIMA, RF, XGB, basic LSTM, CNN-LSTM, and GRU alone [84]. Around the same time, Ghimire et al. (2025) proposed a signal-decomposition and deep-learning hybrid framework, which integrated VMD and EWT to reduce noise and extract features, followed by MRC and Bi-LSTM, to process half hourly price data from South Australia spanning January 2018 to December 2022. The results notably improved compared to all conventional models [85].

The frontier of Transformer-based methods entered the EPF field in 2024–2025, enhancing the model capacity to capture long-range dependencies via self-attention mechanisms. Specifically, Llorente and Portela (2024) introduced a pure Transformer model, relying exclusively on attention layers without recurrent units, for EPF, demonstrating superior performance across multiple open datasets and enhancing reproducibility through open-source code [86]. In parallel, Jiang et al. (2024) proposed a probabilistic forecasting framework based on a pTFT, demonstrating improved interval predictions in volatile deregulated markets [20]. Additionally, Han et al. proposed TransGraph-Opt model that integrates Transformer, GNN, and PCGrad, deployed it on the PJM interconnection market data and PMU measurements of IEEE 39-bus power system model datasets, and the results is outperformed traditional models [87].

Finally, the latest innovation involves leveraging local-temporal convolutional transformers, which fuse convolutional layers for local dependency extraction with Transformer architectures for longer-range attention, enabling highly accurate day-ahead wholesale price forecasts and supporting energy system flexibility and sustainability [88]. These developments mark a layered progression from deterministic point to probabilistic output, through sequential deep hybrids and domain-aware ensembles, reaching pure and enhanced Transformer architectures.

A comparative summary of the evolution of AI methods for electricity price forecasting is provided in Table 3.

III. AI FOR OPTIMIZATION

A. Grid Scheduling

Grid scheduling optimization represents a foundational application area where AI methodologies have demonstrated significant advances in addressing the multi-objective nature of power system operations. Traditional mathematical programming approaches, although effective for well-defined problems, face limitations when dealing with the increasing complexity introduced by renewable energy integration and real-time operational requirements.

RL has emerged as a particularly promising approach for real-time grid operational problems, offering the ability to learn optimal policies through interaction with dynamic environments. DQN and policy gradient algorithms have been suc-

TABLE 3
AI APPLICATIONS FOR ELECTRICITY PRICE FORECASTING

Model Used	Key Advantages	Weaknesses	Performance
Probabilistic DNN [83]	Outputs full predictive distributions (normal, Johnson's SU), risk-aware	Needs large data, complex calibration	Outperformed LASSO, QRA in Germany DA market
Bi-LSTM + GRU Hybrid [84]	Captures and refines bi-directional temporal context	Model complexity	Outperformed ARIMA, RF, XGB, LSTM, CNN-LSTM
VMD + EWT + MRC + Bi-LSTM [85]	Removes noise, extracts robust features	Multi-stage pipeline, heavier compute	Best results on South Australia half-hourly price forecasts
Pure Transformer (attention only) [86]	Captures long-range dependencies, reproducible open datasets/code	Data hungry, high compute	Superior accuracy across multiple EPF datasets
pTFT [20]	Probabilistic intervals, interpretable temporal attention	Needs penalization tuning	Improved forecast intervals in volatile markets
TransGraph-Opt (Transformer + GNN + PCGrad) [87]	Models graph (market topology) + temporal sequences, mitigates gradient conflicts	Complex architecture	Outperformed traditional ML/DL on PJM + IEEE 39-bus
Local-temporal convolutional transformer [88]	CNN layers capture local features, Transformer handles long-term	High complexity, training cost	Best day-ahead wholesale EPF accuracy, supports flexibility

cessfully applied to frequency regulation and congestion management tasks, enabling systems to adapt to changing grid conditions [89]. Xiao et al. (2025) proposed a fast-converged deep RL framework using DDPG to solve transient security-constrained optimal power flow, demonstrating its efficacy on IEEE 39-bus and a 710-bus regional grid [90]. The development of physics-informed RL represents a significant advancement in grid control applications. Mahapatra et al. (2022) integrated physics-based information into RL agent's neural network loss function, achieving 2.7× faster convergence in generator damping control [91].

MARL has shown particular promise for distributed grid control scenarios. The CCMA architecture proposed by De Mol et al. (2025) for power grid topology control demonstrates superior sample efficiency and performance compared to single-agent approaches. This hierarchical framework enables regional agents to propose actions while a coordinating agent makes final decisions, effectively addressing the combinatorial nature of grid control action spaces [92].

The integration of evolutionary algorithms with real-time systems has proven particularly effective for unit commitment problems with high renewable penetration. Alakel et al. (2024) employed GA and PSO methods for the optimal tuning of gain and time constant parameters of PSS in a power network with integrated PV energy conversion systems [93]. The results reveal that the PSO and GA-based PSS are significantly better than the PI-PSS, and further, PSO-PSS is marginally more effective than the GA-PSS in damping the rotor angle oscillations. Zhu et al. (2024) demonstrated the effectiveness of PSO-DQN hybrid approaches in clean energy scheduling, increasing renewable utilization from 62.4% to 87.7% and reducing scheduling costs by 22% [89].

GNNs have gained traction for capturing the topological structure of power networks in optimization problems. Lee et al. (2024) developed a GC-PPO algorithm to determine how much power individual buses dispatch to reduce frequency fluctuations across a power grid. The results demonstrated effectively determine optimal power dispatch across decentralized power grid nodes [94].

A comparative summary of the evolution of AI methods for grid scheduling optimization is provided in Table 4.

B. Energy Storage

BESS has become central to grid flexibility, enabling energy arbitrage, reserve provision, and fast frequency services while absorbing renewable variability. Recent work shows a clear methodological progression, which is from early end-to-end DRL with explicit cycle ageing to optimization-ready degradation costing that can be embedded in MILP and dispatch, and finally to risk-sensitive, forecast-aware, and multi-market decision-making frameworks.

The first wave established that model-free controllers are viable if battery wear is priced realistically. Cao et al. (2020) formulated arbitrage as a MDP and trained a DRL agent that combined CNN-LSTM price features with an explicit lithium-ion degradation model. The learned policy was competitive against an optimization baseline under realistic cycling costs, demonstrating feasibility beyond stylized simulations [95]. In parallel, the theory matured on how to price degradation within system optimization. Bansal et al., (2020) proposed one line embedded Rainflow cycle-counting in economic dispatch, showing that co-optimizing generation and storage with cycle-depth costs materially changes incentives and can be decisive for profitability [96]. He et al. (2021) constructed a complementary thread which is derived from a marginal degradation cost, a time-varying shadow price that couples cell ageing, time value of money, and system conditions, so that BESS can be dispatched on equal economic footing with generators [97]. Crucially, Lee and Kim (2022) further proposed a degradation-cost formulation that replaces direct Rainflow with a one-cycle cost function and an auxiliary SoC variable that tracks discharge-induced wear. This preserves depth-of-discharge effects while remaining MILP-friendly for day-ahead scheduling [98].

As BESS revenues diversified, risk management became central. Using SSD, Khaloie et al. (2023) proposed a risk-aware day-ahead and intraday bidding scheme with a benchmark-selection procedure, showing how SSD can steer portfolios toward improved downside protection relative to risk-neutral baselines [99]. García-Miguel et al. (2024) compared CVaR and other risk measures with Rainflow-based degradation for the Iberian market. The results show that some risk

TABLE 4
AI APPLICATIONS FOR GRID SCHEDULING OPTIMIZATION

Model Used	Key Advantages	Weaknesses	Performance
PSO-DQN hybrid RL [89]	Improves renewable utilization in scheduling, adapts to dynamic grid	Training complexity, scalability challenges	Renewable share increases from 62.4% to 87.7%, scheduling cost reduces 22%
Fast-converged DRL + DDPG-CL-PE-ED [90]	Solves transient security-constrained OPF, scalable to large grids	Sensitive to hyperparameters	Validated on IEEE 39-bus + 710-bus grid
Physics-informed RL + augmented random search [91]	Embeds physics in loss, accelerates convergence	Requires detailed system modeling	2.7× faster convergence in generator damping control
CCMA RL [92]	Efficiently handles combinatorial action space, better sample efficiency	High coordination overhead	Outperformed single-agent RL in grid topology control
GA + PSO for power system stabilizers [93]	Effective parameter tuning of PSS with PV integration	Heuristic, no learning ability	PSO-PSS better than GA-PSS and PI-PSS in damping oscillations
GC-PPO [94]	Captures grid topology in dispatch decisions, scalable	Needs graph embedding + large compute	Effective decentralized frequency regulation

清華大学出版社 Tsinghua University Press | SciOpen

metrics can underperform without careful calibration, underscoring the need to tune risk parameters jointly with wear models [100].

Learning methods also became risk-sensitive and forecast-aware. Madahi et al. (2024) used distributional DRL to learn arbitrage policies under an imbalance-settlement mechanism. The distributional critics allow BRPs to adjust their risk preferences, and the proposed control framework outperformed DQN and SAC baselines on Belgian TSO data [101]. Recognizing that purely reactive agents cannot "plan", Sage et al. (2024) augmented DRL with time-series forecasting, providing the agent with short-horizon price predictions. On highly non-stationary Alberta prices, combining multiple forecasts increased DQN accumulated rewards by about 60% versus no-forecast agents, illustrating the value of forecast-aware control [102]. Beyond single markets, Li et al. (2024) developed a Transformer-based temporal feature extractor within DRL to jointly bid in energy and FCAS across seven markets in Australia, reporting substantial out-performance versus optimization and DRL benchmarks, evidence that temporal-aware architectures help in multi-market settings [103].

With degradation now central to revenues, life-cycle-consistent operations emerged. Ma et al. (2024) coordinated long-term capacity bounds (quarterly) with short-term energy and regulation services, updating bid limits to balance revenues with cell health. This health-aware co-optimization is practical when bids must reflect evolving usable capacity [104]. Zhao et al. (2022) proposed that mobile BESS control with knowledge-assisted DRL stabilizes learning under spatiotemporal uncertainties [105], and He et al. (2025) extended applicability by FFR with modular batteries models module-level efficiency heterogeneity and cycle ageing within a MIQCP, reporting 28%–57% power-loss cost reduction and 4%–15% ageing-cost reduction [106].

A comparative summary of the evolution of AI methods for energy storage optimization is provided in Table 5.

C. Demand Response

Early DR studies used optimization formulations to approximate appliance behavior and device constraints, producing executable baselines against which data-driven methods. Antunes et al. (2022) proposed the modular MILP library of appliance models for home energy management, which is a representative foundation and formalizes shiftable, interruptible, and thermostatic loads and user discomfort into tractable building blocks for DR scheduling [107].

As smart meters diffused, more recent work uses smart-meter traces to learn the latent relations between user routines and load shapes, then links those patterns to DR suitability. For example, Vasilis et al. (2024) proposed a new ML-based framework to achieve optimal load profiling, utilizing ~ 5000 households data in London and 4 different clustering algorithms, with xAI to improve the interpretability of the solution. The results screened out internally inconsistent clusters as poor DR candidates, explicitly tying learned behavior classes to actionable DR program design [108]. To address the "aggregate-meter only" observability gap, Esteban-Perez et al. (2024) used an IO to decompose net demand into baseload, flexible, and self-generation components and estimated their nonlinear price responses, making the behavior–consumption link explicit from smart-meter aggregates [109]. Together, these works shifted DR's foundations from rule-based "executable approximations" to data-grounded, explainable behavior associations.

TABLE 5

AI APPLICATIONS FOR ENERGY STORAGE OPTIMIZATION

Model Used	Key Advantages	Weaknesses	Performance
DRL (CNN-LSTM + degradation model) [95]	End-to-end arbitrage, includes lithium-ion ageing	Needs accurate degradation data	DRL policy competitive with MILP baseline under cycling costs
Rainflow cycle-counting in economic dispatch [96]	Explicit degradation-aware dispatch, simple to embed	Not MILP-friendly, coarse cycle model	Shows that generation-centric dispatch needs to consider storage degradation for profitability
Marginal degradation cost model [97]	Couples ageing, time-value, system conditions	Needs calibrated shadow price	Dispatch aligned with generators on economic basis
MILP-friendly degradation cost formulation [98]	One-cycle SoC-based function, computationally efficient	Approximation of real Rainflow	Practical for day-ahead scheduling
SSD-based risk-aware bidding [99]	Improves downside risk control, flexible portfolio selection	Needs benchmark selection	Reduced risk exposure compared to risk-neutral baseline
CVaR + Rainflow [100]	Combines risk + wear	Risk metrics may underperform if mis-tuned	Iberian market: sensitive trade-off between risk and ageing
Distributional RL (DQN, SAC baseline) [101]	Risk-sensitive arbitrage, adjustable preferences	Higher sample complexity	Outperformed DQN/SAC in Belgian TSO data
Forecast-augmented DRL (DQN + time-series) [102]	Uses forecasts to improve DRL planning	Needs accurate forecasts	Alberta market: rewards increase about 60% compared to no-forecast
Transformer-enhanced DRL for multi-market bidding [103]	Jointly optimizes across markets, temporal feature extraction	High complexity	Outperformed DRL and optimization in Australian FCAS markets
Health-aware co-optimization [104]	Coordinates long-term capacity with short-term services	Needs accurate degradation models	Maintained capacity health while ensuring revenue
Mobile BESS + knowledge-assisted DRL [105]	Stabilizes training under spatiotemporal uncertainty	Relies on expert knowledge input	Accelerates learning process by 30% compared to baseline
MIQCP + performance aware schedule [106]	Models heterogeneity and FFR, integrates degradation	Computationally expensive	Power-loss cost drops 28%–57%, ageing cost drops 4%–15%

Building on heterogeneity, the field moved from generic tariffs to segmentation-driven coordination and differentiated strategies. Meng et al. (2023) proposed an adaptive clustering-based customer segmentation framework and developed customized demand models for each group. A profit maximization problem has been formulated and the results improved cost and peak metrics relative to uniform tariffs [110]. For settings with many premises and local information, multi-agent formulations stabilize coordination under preference heterogeneity. Xie et al. (2023) proposed a new MADRL-based approach with an agent assigned to individual buildings to facilitate demand response programs with diverse loads. The shared attention mechanism has been used to obtain decentralized cooperative policies for electricity costs minimization and efficient load shaping without knowledge of building energy systems [111]. Meanwhile, Cheng et al. (2024) developed privacy-preserving FL to train response/incentive models without centralizing raw data and still reaches economic equilibria at aggregation [112].

Pricing and incentives then became algorithmically adaptive. Wang et al., (2024) proposed bi-level RTP model, which embeds an upper-level retailer objective and lower-level user responses. The study solved the coupled game with RL, explicitly considering distributed energy resources, carbon constraints, and grid fluctuations [113]. Fraija et al. (2024) applied DRL to dynamic pricing under market and supply constraints. The results allowed DR Aggregator to improve its profit per sale of electricity by 35% [114]. To avoid risky online exploration, Kumar et al. (2024) used offline RL with historical smart-meter logs to learn deployable, individual-level RTP policies, enabling "no-trial-and-error" roll-out [115].

Recognizing that enrollment and response are not purely economic, recent studies integrated behavioral economics and social factors to improve take-up, persistence, and fairness. A behavioral-science synthesis argues that effective DR must separately address investment, participation, and the actual real-time response separately, and maps specific interventions (feedback, framing, and defaults) to each [116]. Ryu et al. (2023) modeled loss aversion and uncertainty in DR adoption and exit decisions based on prospect-theoretic game approach, explaining divergent responses under identical incentives [117]. System-level evidence and reviews emphasize that although monetary incentives dominate, information, trust, and convenience substantively affect small-user participation, with fairness for vulnerable groups a design requirement [118]. Through analyzing data from more than 200 thousand households, Wang et al. showed that randomized emergency DR can cause significant peak-time reductions without additional economic burden to vulnerable households, offering quantitative support for "welfare-neutral or improving" program design [119].

At the individual level, the field is converging toward end-to-end personalization by fusing structural knowledge with online learning and recommendations. Shao et al. (2024) extracted a "SCDL" to characterize each customer, then solved a bi-level (aggregator–customer) optimization while defining multi-dimensional portraits that evaluate customers' performance in DR programs [120]. Eirinaki et al., (2022) proposed a framework integrated with real-time recommendation of efficient appliance use to help households improve energy efficiency. The results on UK-DALE dataset showed that the framework can help reduce their energy consumption by 2%–17% [121].

A comparative summary of the evolution of AI methods for demand response optimization is provided in Table 6.

IV. AI for Control and Maintenance

A. Grid Control

A key control challenge in active distribution networks is

TABLE 6
AI APPLICATIONS FOR DEMAND RESPONSE OPTIMIZATION

Model Used	Key Advantages	Weaknesses	Performance
Modular MILP appliance library [107]	Provides formal, executable DR baselines (shiftable, thermostatic loads)	Requires detailed device parameters	Foundation models for DR scheduling
ML clustering + XAI [108]	Learns latent user load profiles, interpretable clusters	Needs large smart-meter datasets	Identified ~ 5000 London households, screened poor DR candidates
Adaptive clustering + segmentation [110]	Customized group demand models, profit maximization	Sensitive to cluster quality	Achieves better profit compared to uniform pricing
Multi-agent DRL + shared attention [111]	Decentralized DR across heterogeneous buildings	Requires many agents, training overhead	Reduces the load demand by 6% compared to standard RL approach
FL for DR [112]	Protects privacy, avoids centralizing raw data	Requires communication infra	Achieved equilibria while preserving privacy
Bi-level real-time pricing + RL [113]	Models retailer–user game, considers DERs and carbon	Complexity of solving bilevel problem	Stable RTP under fluctuating grid
RL-based DRA for dynamic pricing [114]	Learns adaptive incentives under market constraints	Data intensive	Profit increases 35% for DRA
Offline RL for RTP [115]	Learns deployable pricing from historical logs	Limited adaptability to new states	Eliminates risky online trial-and-error
Prospect-theoretic game model [117]	Captures loss aversion and uncertainty in DR participation	Needs behavioral calibration	Small scale ESS average benefit-cost ratio increases from 0.78 to 1.08
SCDL + Bi-level optimization [120]	Personalized DR with structural + behavioral features	Needs per-customer profiling	Improved customer portraits, better DR targeting

maintaining voltage quality while minimizing tap operations and losses. Cao et al. (2021) proposed a DRL two-timescale Volt/VAR control (VVC) that coordinates fast inverter reactive power with slower OLTC actions without relying on exact feeder models. Across multiple feeders, the learned policy reduced voltage violations and switching actions relative to model-based benchmarks, evidencing feasibility of physical-model-free control [122]. Building on this paradigm, Kabir et al. (2023) formulated a DRL two-timescale VVC, reporting minimized voltage violation costs and system operation costs over conventional rule-based VVC across realistic scenarios, thus reinforcing the generality of learning-based coordination for distribution systems [123]. Scalability and heterogeneity motivated MARL. Wang et al. (2021) cast AVC as a Dec-POMDP and released the open-source environment. The results demonstrated that attention-based agents coordinating many distributed energy resources outperform local droop-like heuristics on feeder-level voltage and loss objectives, providing a community benchmark and MARL algorithms for reproducible evaluation of real-world applications [124].

Despite these advances, MARL-based voltage control introduces the risk of emergent and destabilizing behaviors, such as oscillatory actions, unsafe exploration, or adversarial interactions among agents. The risks of emergent behaviors that must be formally modeled and constrained, both technically and institutionally. Several technical strategies have been proposed to mitigate these risks. A primary mitigation strategy lies in sophisticated reward engineering. Instead of rewarding only local voltage regulation, the reward function must include system-level penalty terms that disincentivize rapid control actions, excessive power exchange, or behaviors that reduce overall stability margins. This encourages agents to learn not just effective, but also smooth and safe control policies [125]. Second, safe RL frames the problem as a CMDP, enforcing operational safety limits during both training and deployment. Approaches such as projection layers and barrier-function critics ensure that actions leading to voltage collapse or transient instability are excluded even during exploration [126]. Moreover, to provide formal stability guarantees, researchers are integrating MARL with traditional control-theoretic principles. For example, Lyapunov stability analysis can be used to define a “safe region” of operation. The MARL system is then designed to operate strictly within this region, with a classic, provably stable controller acting as a fallback supervisor that intervenes if the AI attempts to exit the safe envelope [127]. These techniques complement physics-informed MARL, which has been introduced by Zhang et al. (2024), that encodes grid sensitivities and network structure into the policy, accelerating learning and enhancing robustness on fast voltage control tasks in distribution networks [128].

Beyond technical safeguards, effective governance frameworks are required for safe deployment of MARL-based controllers. The NREL (2024) emphasized that a critical governance component is mandatory certification within high-fidelity digital-twins of the power system, where agents are rigorously stress-tested against contingency scenarios, communication failures, and potential cyberattacks to formally verify resilience and fail-safe behaviors [129]. The DOE (2024) highlighted the need for hierarchical oversight in which MARL agents function only as level-1 or advisory controllers. Human operators must retain ultimate authority and the ability to override AI decisions, often supported by automated, physics-based safety nets that preemptively block unsafe control actions [130]. ENTSO-E (2024) further stressed the importance of auditability and explainability: MARL systems should generate immutable logs of observations, decisions, and internal states, while explainable AI techniques provide operators with transparent rationales, thereby fostering trust and enabling effective post-event analysis [131]. Together, these technical safeguards and governance measures provide a multi-layered defense, making MARL a more credible candidate for real-world deployment.

Frequency regulation under large disturbances is another frontier where AI demonstrates value. Xie et al. (2021) developed distributional DRL for EFC, showing on large-scale test systems that distributional critics stabilize frequency faster and more robustly than classical under-frequency load shedding logic, particularly under high RES [132]. Complementarily, Zheng et al. (2023) integrated linear active disturbance rejection control with DRL for LFC in multi-area systems, reporting improved settling time and overshoot over PI baselines across renewable variability cases [133].

A comparative summary of the evolution of AI methods for grid control is provided in Table 7.

TABLE 7
AI APPLICATIONS FOR GRID CONTROL

Model Used	Key Advantages	Weaknesses	Performance
DRL for two-timescale Volt/VAR control (VVC) [122]	Model-free, coordinates inverters with OLTC without feeder model	Training needs large data, stability concerns	Reduced voltage violations and tap actions across feeders
DRL two-timescale VVC (degradation-aware inverters) [123]	Incorporates inverter degradation cost, robust coordination	Relies on realistic inverter models	Reduced violation and system cost vs. rule-based VVC
Multi-agent RL (Dec-POMDP for AVC) [124]	Decentralized agents, open-source benchmark, scalable	Training complexity, requires communication	Outperformed droop heuristics on feeder voltage/loss
Physics-informed MARL (PV inverters) [128]	Encodes sensitivities + topology, faster, more stable	Needs accurate physics models	Robust fast-voltage control, accelerated convergence
Distributional DRL for EFC [132]	Distributional critics stabilize control, robust to RES	Computationally heavy	Faster stabilization vs. UFLS, robust under high RES
DRL + active disturbance rejection control for LFC [133]	Hybrid improves disturbance rejection, shorter settling	Complexity in integration	Improved settling and overshoot vs. PI baseline

B. System Maintenance

On the maintenance side, PdM benefits from fusing established diagnostics with modern machine learning. For power transformers, Reference [134] reviewed DGA alongside vibration and thermal methods, detailing how supervised learning and ensemble models improve fault typing and early warning despite class imbalance and overlapping signatures. Sutikno et al. (2024) refined DGA interpretation with multi-method machine-learning ensembles, reporting higher diagnostic consistency against legacy ratio methods on curated utility datasets [135].

Computer vision has shifted asset inspection from manual patrols to automated detection. Qiu et al. (2022) proposed an improved YOLOv4 pipeline for insulator defect detection using aerial imagery, achieving higher precision recall than baseline detectors on composite datasets and establishing a pragmatic recipe (data augmentation + sharpening + attention) for transmission assets [136].

Chen et al. (2024) further provided an effective identification model to solve the problem of electricity transmission line defects based on the improved YOLOv4 network. The results achieved better efficiency and accuracy while reducing unnecessary false positives [137].

Beyond visible imagery, Tang et al. (2024) introduced a deformable convolution module into the DeDn-CNN with image denoising algorithm to thermal fault diagnosis for infrared images of complex electrical equipment. The results yielded accurate hot-spot localization on substation equipment, enabling actionable thermal diagnostics with reported component-level detection accuracies above 85% on curated sets [138].

Event analytics from PMU data is maturing into robust monitoring. Liu et al. (2023) developed a real-world PMU event classification workflow that addresses noisy logs and missing data via preprocessing, event localization, and feature engineering. The framework has been validated on Western Interconnection data, and the results showed that lightweight models trained under this framework surpass neural baselines when labels are imperfect, enabling scalable deployment in WAMS [139].

Methodologically, Gök et al. (2023) generate physics-plausible synthetic PMU frequency trajectories to enlarge the training set of deep classifiers and show that the "synthetically supported" set yields higher transient-event classification accuracy than training on field data alone, directly addressing label scarcity in real grids [140]. In parallel, MansourLakouraj et al. (2022) exploit multi-rate PMU streams from PV-penetrated active distribution feeders and apply a spectral GNN, which encodes network topology via the graph Laplacian, to jointly classify event type and identify the affected region, demonstrating improved robustness to reporting-rate heterogeneity and distribution-grid operating variability [141].

For localization, GNNs encode network topology and mixed node/edge attributes. Sun et al. (2021) demonstrated GNN-based distribution fault location, outperforming conventional impedance-based estimators in simulations that incorporate measurement uncertainty [142].

A comparative summary of the evolution of AI methods for system maintenance is provided in Table 8.

V. LLMs in Energy Systems

Beyond traditional machine learning architectures, the recent emergence of LLMs represents a paradigm shift in AI's potential role within the energy sector. These models, pretrained on vast textual and code datasets, possess an unprecedented ability to understand, generate, and reason with unstructured human language, opening new frontiers for operational efficiency and human–machine interaction. This section surveys the emerging applications of LLMs across energy systems, which can be broadly categorized into several key areas: grid dispatch, simulation and tool integration, market decision-making, fault diagnosis, knowledge-based control room assistance, and safety evaluation.

TABLE 8
AI applications for system maintenance

Model Used	Key Advantages	Weaknesses	Performance
ML ensemble for DGA [135]	Higher diagnostic consistency, overcomes ratio limits	Needs curated utility datasets	Improved early fault typing vs. legacy DGA
YOLOv4 (improved) for insulator defect detection [136]	High precision/recall, robust pipeline (augmentation + attention)	Relies on large annotated datasets	Outperformed baseline detectors
YOLOv4 (lightweight, CV-assisted) [137]	Efficient defect detection, reduces false positives	May underperform on unseen asset types	Higher efficiency + accuracy than prior YOLO models
Deformable CNN + DeDn-CNN for infrared images [138]	Accurate hotspot localization, > 85% component detection	Sensitive to infrared noise and preprocessing of image denoising	Effective thermal fault diagnosis for infrared images of complex electrical equipment
Event classification workflow (PMU data) [139]	Handles noise + missing data, scalable to WAMS	Preprocessing overhead	Validated on Western Interconnection data, better than NN baselines
Deep learning + synthetic PMU data augmentation [140]	Addresses label scarcity, improves transient classification	Synthetic data realism limits	Higher accuracy vs. field-only training
Spectral GNN + multi-rate PMU [141]	Encodes topology, robust to reporting-rate heterogeneity	High computation demand	Improved robustness in PV-penetrated feeders
GNN for fault location [142]	Encodes node/edge attributes, better under uncertainty	Needs graph construction, less tested in field	Outperformed impedance-based estimators in simulation

A. Applications

The complexity of power dispatch presents a prime target for LLM specialization. Cheng et al. (2025) introduced GAIA, a domain-specialized LLM for power dispatch, reporting that GAIA was fine-tuned on multi-source dispatch corpora and evaluated on ElecBench. The results outperformed baseline LLaMA-2 configurations on multiple dispatch tasks [36]. The ElecBench benchmark explicitly targets power dispatch tasks and identifies GAIA as the first dispatch-domain LLM with multiple parameter scales for controlled evaluation [35]. This work demonstrates that through domain-specific adaptation, LLMs can learn to execute tasks requiring precise understanding of grid physics and operational protocols.

A significant barrier to AI in energy is the inability to interact with specialized software. Jia et al. (2024) proposed a modular, feedback-driven multi-agent framework that equips LLMs with enhanced RAG with stepwise reasoning and environment-acting to run power-system simulations across previously unseen tools such as DALINE and MATPOWER. On 34 tasks in DALINE, the proposed framework achieved 96.07% coding accuracy compared to 0% in GPT-4o and 33.8% in GPT-4o web interface. The results highlighted the potential of LLMs agents in power systems [143]. This breakthrough shows that with structured reasoning and tool-calling capabilities, LLM agents can autonomously operate complex power system software, dramatically lowering the expertise barrier for simulation and analysis.

LLMs are also being trialed as cognitive components within market trading agents. Zhang et al. (2024) modeled BESS bidding in Australia's FCAS market as risk-aware DRL. Then they added an LLM-assisted AI agent for conditional hybrid decision-making and self-reflection to mitigate hallucinations. The results reported higher bidding profitability and improved robustness versus DRL-only baselines under uncertain market conditions [144]. This suggests that the reasoning and contextualizing abilities of LLMs can complement the optimization power of DRL, leading to more robust and profitable strategies in non-stationary environments.

The explainability of LLMs is particularly valuable for diagnostic tasks. Liu et al. (2024) explored prompt-engineered LLMs (e.g., GPT-4) for power grid fault diagnosis from sensor/asset descriptions, reporting accuracy and explainability gains relative to naive prompting and tree/chain-of-thought variants [145]. Similarly, Dave et al. (2024) embedded an LLM agent into a diagnostic system that grounds explanations in physical models and sensor data to reduce hallucination risk [146]. For real-time control-room use, Liu et al. (2024) combined a DNN with a pre-trained LLM to analyze SCADA alarm sequences for decision support [147]. Collectively, these studies highlight a key trend that marrying the semantic reasoning of LLMs with the certainty of physical models or sensor data is a promising path toward trustworthy and interpretable diagnostics.

For real-time operations, LLMs are being tailored as next-generation decision support tools. Choi et al. (2024) described eGridGPT, a trustworthy-AI concept where RAG grounds LLM outputs in utility procedures, SCADA/PMU context, and operator checklists to reduce hallucinations for "control room of the future" use cases [37]. PNNL published LLM use cases that process alarm sequences, perform root-cause narratives, and generate operator recommendations with domain-adapted LLMs [148]. The core innovation here is the construction of a verified informational context, comprising manuals, real-time data, and procedures, that constrains the LLM's outputs and grounds them in operational reality, thereby minimizing hallucination for "control room of the future" use cases.

As LLMs are considered for increasingly critical roles, proactive safety assessment becomes paramount. Ruan et al. (2024) systematically examined vulnerabilities such as prompt injection, unstable decision-making, and data leakage in LLM-based grid operations, urging the establishment of countermeasures and standards [149]. Majumder et al. (2024) further evaluated LLMs on electrical-engineering tasks, including power-flow-type reasoning, both their promise and limitations, and underscoring the need for rigorous benchmarks before widespread deployment [150]. These pioneering studies establish a crucial research frontier. The need for rigorous, standardized red-teaming and benchmarking ensures that the deployment of LLMs is accompanied by a thorough understanding of their failure modes and security risks.

B. Hallucination Risk

Although these pioneering applications highlight the promise of LLMs, they also repeatedly expose a fundamental obstacle, which is the risk of hallucination, preventing the reliable deployment of general-purpose LLMs in high-stakes energy system operations [151]. This phenomenon, wherein a model generates plausible-sounding but factually incorrect, inconsistent, or entirely fabricated information, stems from their core design as statistical generators rather than symbolic reasoners.

In the energy sector, the manifestation of hallucination phenomena is particularly dangerous where decisions have physical consequences. Models may invent nonexistent equipment, misrepresent the thermal limits of transmission lines, or most critically, generate hazardous switching sequences that cause equipment damage or trigger cascading blackouts. These risks are amplified by the fact that LLMs lack a grounded understanding of physics and are trained on internet-scale data that may contain inaccuracies or lack specific, up-to-date grid knowledge.

C. Pathways to Trustworthy LLMs

In safety-critical energy applications, mitigating hallucinations requires not only improving language fluency but also system designs that guarantee determinism and verifiability. Current research is converging on two complementary approaches [152]. The first is to embed LLMs within hybrid neuro-symbolic architectures, where the model's generative flexibility is bounded by formal rules and deterministic solvers. The second is to apply external validation paradigms, treating the LLM as a black-box generator whose outputs must pass independent checks before they are acted upon. Together, these strategies shift the burden of reliability away from the LLM itself and onto surrounding safety mechanisms.

The hybrid neuro-symbolic pathway redefines the LLM's role from decision-maker to language interface [153]. Operators' queries are translated by the LLM into structured inputs for certified tools such as power-flow simulators or optimization engines. These solvers remain the sole source of numerical truth, while the LLM only converts results into human-readable summaries. For more complex tasks, such as drafting switching sequences, the LLM may generate candidate plans, but they are then screened by a symbolic verifier or safety shield that encodes hard grid constraints like thermal and voltage limits, frequency stability, or N-1 criteria. RAG can further constrain outputs by forcing the model to ground responses in a curated and versioned knowledge base containing SCADA/PMU data, operating procedures, and equipment manuals. In this way, the model's creativity is strictly "anchored" to documented facts and real-time system states.

The external validation pathway adds a second line of defense. Recommendations that imply physical actions are replayed in digital-twin simulations, which act as empirical "reality checks" by comparing simulated outcomes with operational safety thresholds [154]. In parallel, cross-model consensus leverages diversity [153], which means that LLMs, rule-based expert systems, and GNNs can all be tasked with the same query, and results are accepted only when independent systems agree. Disagreement is not treated as failure but as a signal of complexity or ambiguity, automatically escalating the issue for human review. Combined with immutable audit trails that record data, model versions, and decisions, these mechanisms ensure that no unsafe action can bypass deterministic checks, positioning LLMs as assistants that enhance human oversight rather than uncontrolled decision-makers.

VI. Future Research Directions

A. Trustworthy AI

In high-stakes grid operations, explainability alone does not define trustworthiness. Although the ability to understand a model's reasoning is essential, operators and regulators increasingly demand broader guarantees that go beyond transparency. The first of these is uncertainty quantification [155]. A model that issues confident predictions without indicating when it is operating outside its training domain is not reliable. Approaches such as BNNs and conformal prediction provide a way to attach calibrated confidence intervals to outputs, alerting operators to potential blind spots [156]. Closely linked to this is the requirement for provable safety and robustness. In power system control, no model should ever recommend an action that violates hard operational limits. Here, architectural innovations such as neuro-symbolic designs, with a neural module proposing actions and a symbolic safety shield enforcing physical laws, illustrate how safety can be embedded by design rather than added as an afterthought [157].

Trustworthiness also requires that decisions be reproducible and auditable [158]. Every control action must be traceable back to the exact data, code, and model version that generated it. Emerging MLOps practices make such immutable audit trails feasible and provide regulators with the accountability they expect [159]. At the same time, fairness and equity must be recognized as integral dimensions of trust [160]. As AI increasingly manages distributed resources and demand response, it must not systematically disadvantage vulnerable groups, for example, by curtailing low-income households more often than others. Fairness-aware training, where equity metrics are embedded directly in the model objective, offers one route to mitigate such risks. Finally, resilience against adversarial and cyber threats is indispensable [161]. Attacks such as data poisoning or malicious sensor manipulation are not hypothetical. They are real vulnerabilities that must be addressed through both algorithmic defenses and certification processes, drawing on long-standing practices in protection relays and SCADA systems.

Taken together, these requirements show that trustworthiness is not a single property but a multidimensional standard. Meeting it will require architectural innovations that integrate uncertainty quantification, safety, auditability, fairness, and resilience into AI systems from the ground up. Only then can AI move from promising prototypes to trusted partners in critical grid operations.

B. Scientific AI

The growing emphasis on explainability and resilience reflects a broader shift away from purely data-driven, correlational models toward a paradigm of "scientific AI". At the core of this transition lies physics-informed AI, which reframes machine learning not as a black-box predictor but as a hybrid modeling tool that combines statistical learning with first-principles knowledge [162]. Rather than replacing physics, this approach leverages domain laws to guide learning, producing models that are more data-efficient, robust to out-of-distribution conditions, and inherently interpretable because their outputs are constrained by physical consistency [163].

Realizing this shift, however, requires co-evolution in both data infrastructure and validation standards. The infrastructure must move beyond passive "data lakes" toward context-rich digital-twins that serve as dynamic, queryable representations of the physical grid [164]. In such twins, data are intrinsically linked to their physical origin, including asset identity, topology position, and engineering specifications. Coupling these with advanced physics-based simulators enables the generation of targeted synthetic datasets for rare but critical scenarios, as well as systematic "what-if" testing for robust model validation.

Equally important are new validation standards. Traditional statistical metrics such as mean absolute error are insufficient when models are deployed in safety-critical environments. Future standards must evaluate physical consistency (e.g., adherence to Kirchhoff's laws and operational limits), include standardized extrapolation benchmarks that test generalization to extreme contingencies and unprecedented weather patterns, and adopt causal validation frameworks that probe whether models have captured genuine cause–effect relationships rather than spurious correlations.

By jointly investing in these evolutions of infrastructure and standards, the energy sector can lay the groundwork for a new generation of scientific AI, which means that models not only achieve predictive accuracy but also embody the robustness, interpretability, and causal understanding required for reliable use in critical grid operations.

C. Sector-Coupled and Multi-Agent Optimization

Modern energy systems are no longer isolated silos. They form a tightly interwoven fabric of electricity, heating, cooling, mobility, and gas networks, where decisions unfold on vastly different timescales, from seconds in frequency regulation to months in storage planning. The technical challenge is that sector coupling multiplies the dimensionality of the optimization space. Millions of states and interdependent actions must be coordinated across heterogeneous domains. Traditional approaches, such as mixed-integer programming, often collapse under this burden, as computational complexity and non-convexities grow exponentially with system scale. This is the essence of the curse of dimensionality in sector-coupled optimization [165].

AI offers a way forward by replacing monolithic formulations with structured decomposition. MARL provides a natural framework. Instead of solving one massive optimization, the problem is distributed across agents that represent distinct sectors (e.g., district heating, EV fleets, and distribution grids) or even different temporal layers within a sector. Each agent learns specialized policies for its local sub-problem, while coordination emerges through carefully designed interaction protocols. Within this backbone, other AI techniques add further scalability, for instance, GNNs compress complex infrastructures into lower-dimensional representations, approximate dynamic programming and DRL make exploration of high-dimensional state spaces tractable, and federated or distributed learning allows collaboration across stakeholders without requiring sensitive data to be centralized. Together, these methods transform dimensionality from an obstacle into an opportunity for modular learning.

Looking ahead, distributed and edge intelligence will be essential to complement cloud-based coordination. Substations, microgrids, and local aggregators must retain autonomy to act in real time, particularly under communication failures or cyberattacks. Edge AI, strengthened by federated learning, enables such local decision-making while preserving system-wide coherence. In our view, the integration of multi-agent, multi-sector, and multi-timescale optimization remains one of the most underexplored but potentially high-impact frontiers for AI in energy systems.

D. Holistic Benchmarks

A persistent barrier to responsible AI deployment in energy systems is how performance is evaluated. At present, most studies still rely on narrow accuracy metrics such as RMSE on historical datasets. Although useful for basic validation, such numbers reveal little about whether a model can be trusted under real operating conditions. In critical infrastructures, the decisive question is not only "How accurate is the average prediction?" but also "Will the system remain reliable, fair, and resilient under stress?" Similar concerns have recently motivated the creation of multi-dimensional evaluation suites such as ElecBench, which assesses factuality, stability, security, and fairness for power-dispatch LLMs rather than accuracy alone [35].

Moving forward, the community needs shared, open testbeds built on high-fidelity digital-twins. Digital-twins are increasingly recognized as dynamic, queryable representations of power systems, capable of linking data to physical assets, topology, and engineering specifications [166]. In the energy domain, they have been identified as essential tools for benchmarking AI under diverse operating conditions, from high-inertia transmission systems to distribution grids with high shares of distributed resources [167]. Advances in high-performance computing now make it feasible to couple such twins with detailed simulators, generating synthetic datasets for rare contingencies while supporting systematic "what-if" testing [168].

Benchmarking should also incorporate a diverse library of challenge cases that push models beyond routine conditions. These cases should cover scenarios such as noisy sensors, missing data, or communication delays to test model robustness. High-impact, low-probability events (e.g., N-k failures, cascading failures, or extreme weather) must be incorporated to assess resilience. Fairness audits should reveal whether AI-driven decisions systematically disadvantage specific customer groups. Finally, evaluation needs to be multi-dimensional. A meaningful benchmark should resemble a "scorecard" that reports on efficiency (accuracy and cost), resilience (load shed, recovery time, and stability margins), robustness (performance under noise or attack), fairness (metrics such as demographic parity), and computational feasibility (inference time and training cost).

Creating such benchmarks will require collective effort across academia, utilities, and regulators. With them, the field can move toward a culture of rigorous, transparent evaluation, which is an essential step for AI to become a reliable partner in the operation of future energy systems.

E. Energy Foundation Models

Beyond the incremental steps above, a more ambitious frontier is the development of "foundation models" for energy. Just as GPT-type models provide a shared backbone for language tasks. The idea is borrowed from natural language processing, where large pre-trained models now serve as the backbone for countless applications. In the energy domain, such a model would aim to encode universal physical principles, such as Kirchhoff's laws, stability constraints, and network flow dynamics into a single, pre-trained architecture that can then be adapted to different grid topologies and regional contexts.

The viable pathway toward this vision rests on several pillars. First, physics-informed architectures will be essential. A natural choice is to use GNNs that reflect the structure of power grids, combined with differentiable power flow solvers to ensure physical consistency in predictions. This approach aligns with broader advances in physics-informed learning, which embed governing equations and domain knowledge directly into neural architectures [169], [170]. Second, multimodal pretraining can allow the model to learn from diverse sources, such as time-series data from SCADA and PMUs, thousands of grid topology examples, and even unstructured text such as technical manuals [171]. Self-supervised tasks, for instance, predicting the outcome of N-1 contingencies, could serve as pretraining objectives that teach the model fundamental grid dynamics. Third, efficient fine-tuning would allow utilities to adapt the foundation model with relatively little local data, enabling tasks such as optimal power flow, fault detection, or asset management without starting from scratch each time.

In our judgment, realizing truly universal, multi-domain foundation models will not be achieved overnight. It demands advances in model design, large-scale multimodal datasets, shared benchmarks, and governance frameworks for trustworthy deployment. Scaling to this level is likely to be a decade-long effort, requiring close collaboration between academia, industry, and regulators. Nonetheless, this "grand vision" offers a unifying pathway for the next generation of AI in energy: Models that are not only powerful predictors but also scientifically grounded, adaptable, and deployable across diverse energy systems.

VII. Conclusion and Policy Implications

This review synthesized the state-of-the-art applications of AI in energy systems, charting a consistent methodological progression from classical ML toward more sophisticated, physically-grounded, and agentic architectures. The evidence confirms that AI has evolved from an auxiliary analytical tool into a foundational enabler of the energy transition. By enhancing forecasting, optimizing operations, and enabling autonomous control, AI directly contributes to the reliability and efficiency required to accommodate high penetrations of renewable energy, empowering a more flexible and decarbonized grid.

Looking forward, the integration of AI into energy systems is no longer a question of if, but of how. Realizing this technology's full potential hinges on addressing three core imperatives for the next decade of development. First is achieving trustworthiness through explainable, verifiable, and cyber-secure systems that earn regulatory and operator acceptance. Second is ensuring deep physical and social integration, where AI decision-making respects both network constraints and equity considerations. Third is delivering distributed scalability through edge intelligence and federated learning to manage millions of decentralized assets. Mastering these imperatives is essential for creating energy systems that are not only sustainable but also equitable and secure.

To navigate this transition, a proactive policy and governance framework is essential. From a market perspective, regulators must mandate open data interfaces and transparent evaluation standards to unlock efficiency gains, while using targeted incentives to de-risk initial AI investments for smaller entities. Operationally, policymakers should champion cross-sector standards to leverage AI for resilience across coupled electricity, transport, and heating systems. Finally, to ensure governance and accountability, clear requirements for explainability, auditability, and minimum cybersecurity standards are indispensable for all critical AI applications. Together, these actions will build the trust and structure needed to establish AI as an operational cornerstone of the sustainable energy future.

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.

Funding

This work is carried out without the support of any funding agency.

AI Usage Statement

During the preparation of this manuscript, the authors used AI-assisted tools for language polishing only. The authors have reviewed and edited the content and take full responsibility for the final version of the manuscript.

References

[1] IEA. (2024, Mar.). CO_2 emissions in 2023. [Online]. Available: https://www.iea.org/reports/co2-emissions-in-2023

[2] F. S. Al-Ismail, S. Alam, M. Shafiullah, I. Hossain, and S. M. Rahman, "Impacts of renewable energy generation on greenhouse gas emissions in Saudi Arabia: a comprehensive review," *Sustainability*, vol. 15, no. 6, pp. 5069, Mar. 2023.

[3] S. Yi, K. R. Abbasi, K. Hussain, A. Albaker, and R. Alvarado, "Environmental concerns in the United States: can renewable energy, fossil fuel energy, and natural resources depletion help?" *Gondwana Res.*, vol. 117, pp. 41–55, May 2023.

[4] T. Hargreaves and L. Middlemiss, "The importance of social relations in shaping energy demand," *Nat. Energy*, vol. 5, no. 3, pp. 195–201,

Feb. 2020.

[5] M. S. Nazir, F. Alturise, S. Alshmrany, H. M. J. Nazir, M. Bilal, A. N. Abdalla, P. Sanjeevikumar, and Z. M. Ali, "Wind generation forecasting methods and proliferation of artificial neural network: a review of five years research trend," *Sustainability*, vol. 12, no. 9, pp. 3778, May 2020.

[6] J. Y. Zhou, J. Shi, and G. Li, "Fine tuning support vector machines for short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 52, no. 4, pp. 1990–1998, Apr. 2011.

[7] C. Koliba, M. DeMenno, N. Brune, and A. Zia, "The salience and complexity of building, regulating, and governing the smart grid: lessons from a statewide public–private partnership," *Energy Policy*, vol. 74, pp. 243–252, Nov. 2014.

[8] Z. Y. Zhang, J. Z. Wang, D. X. Wei, T. R. Luo, and Y. R. Xia, "A novel ensemble system for short-term wind speed forecasting based on Two-stage Attention-Based Recurrent Neural Network," *Renew. Energy*, vol. 204, pp. 11–23, Mar. 2023.

[9] S. Onwusinkwue, F. Osasona, I. A. I. Ahmad, A. C. Anyanwu, S. O. Dawodu, O. C. Obi, and A. Hamdan, "Artificial intelligence (AI) in renewable energy: a review of predictive maintenance and energy optimization," *World J. Adv. Res. Rev.*, vol. 21, no. 1, pp. 2487–2499, Jan. 2024.

[10] L. Chen, Z. H. Chen, Y. B. Zhang, Y. F. Liu, A. I. Osman, M. Farghali, J. M. Hua, A. Al-Fatesh, I. Ihara, D. W. Rooney, and P. S. Yap, "RETRACTED ARTICLE: artificial intelligence-based solutions for climate change: a review," *Environ. Chem. Lett.*, vol. 21, no. 5, pp. 2525–2557, Jun. 2023.

[11] S. Kumari and P. Muthulakshmi, "SARIMA model: an efficient machine learning technique for weather forecasting," *Procedia Comput. Sci.*, vol. 235, pp. 656–670, Jan. 2024.

[12] C. J. Ejiyi, D. S. Cai, D. Thomas, S. Obiora, E. Osei-Mensah, C. Acen, F. O. Eze, F. Sam, Q. X. Zhang, and O. O. Bamisile, "Comprehensive review of artificial intelligence applications in renewable energy systems: current implementations and emerging trends," *J. Big Data*, vol. 12, no. 1, pp. 169, Jul. 2025.

[13] A. L'Heureux, K. Grolinger, and M. A. M. Capretz, "Transformer-based model for electrical load forecasting," *Energies*, vol. 15, no. 14, pp. 4993, Jul. 2022.

[14] N. Khan, F. U. M. Ullah, I. U. Haq, S. U. Khan, M. Y. Lee, and S. W. Baik, "AB-Net: a novel deep learning assisted framework for renewable energy generation forecasting," *Mathematics*, vol. 9, no. 19, pp. 2456, Oct. 2021.

[15] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, "Forecasting day-ahead electricity prices: a review of state-of-the-art algorithms, best practices and an open-access benchmark," *Appl. Energy*, vol. 293, pp. 116983, Jul. 2021.

[16] W. C. Kong, Z. Y. Dong, Y. W. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[17] J. Lin, J. Ma, J. G. Zhu, and Y. Cui, "Short-term load forecasting based on LSTM networks considering attention mechanism," *Int. J. Electr. Power Energy Syst.*, vol. 137, pp. 107818, May 2022.

[18] Y. F. Gou, C. Guo, and R. S. Qin, "Ultra short term power load forecasting based on the fusion of Seq2Seq BiLSTM and multi head attention mechanism," *PLoS One*, vol. 19, no. 3, pp. e0299632, Mar. 2024.

[19] S. Ding, D. Y. He, and G. R. Liu, "Improving short-term load forecasting with multi-scale convolutional neural networks and transformer-based multi-head attention mechanisms," *Electronics*, vol. 13, no. 24, pp. 5023, Dec. 2024.

[20] H. Jiang, S. Pan, Y. Dong, and J. Z. Wang, "Probabilistic electricity price forecasting based on penalized temporal fusion transformer," *J. Forecast.*, vol. 43, no. 5, pp. 1465–1491, Feb. 2024.

[21] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecast.*, vol. 37, no. 4, pp. 1748−1764, Oct. /Dec. 2021.

[22] K. Kumar, S. Kwon, and S. Bae, "Deep reinforcement learning-based control strategy for integration of a hybrid energy storage system in microgrids," *J. Energy Storage*, vol. 108, pp. 114936, Feb. 2025.

[23] L. Pinciroli, P. Baraldi, M. Compare, and E. Zio, "Optimal operation and maintenance of energy storage systems in grid-connected microgrids by deep reinforcement learning," *Appl. Energy*, vol. 352, pp. 121947, Dec. 2023.

[24] A. T. Perera and P. Kamalaruban, "Applications of reinforcement learning in energy systems," *Renew. Sustain. Energy Rev.*, vol. 137, pp. 110618, Mar. 2021.

[25] D. Cao, W. H. Hu, J. B. Zhao, G. Z. Zhang, B. Zhang, Z. Liu, Z. Chen, and F. Blaabjerg, "Reinforcement learning and its applications in modern power and energy systems: a review," *J. Mod. Power Syst. Clean Energy*, vol. 8, no. 6, pp. 1029–1042, Nov. 2020.

[26] T. Yang, L. Y. Zhao, W. Li, and A. Y. Zomaya, "Reinforcement learning in sustainable energy and electric systems: a survey," *Annu. Rev. Control*, vol. 49, pp. 145–163, Apr. 2020.

[27] R. Nellikkath and S. Chatzivasileiadis, "Physics-informed neural networks for AC optimal power flow," *Electr. Power Syst. Res.*, vol. 212, pp. 108412, Nov. 2022.

[28] H. F. Zhang, X. L. Lu, X. Ding, and X. M. Zhang, "Physics-informed line graph neural network for power flow calculation," *Chaos*, vol. 34, no. 11, pp. 113123, Nov. 2024.

[29] Q. H. Ngo, B. L. H. Nguyen, T. V. Vu, J. H. Zhang, and T. Ngo, "Physics-informed graphical neural network for power system state estimation," *Appl. Energy*, vol. 358, pp. 122602, Mar. 2024.

[30] A. Deihim, D. Apostolopoulou, and E. Alonso, "Initial estimate of AC optimal power flow with graph neural networks," *Electr. Power Syst. Res.*, vol. 234, pp. 110782, Sep. 2024.

[31] A. Grataloup, S. Jonas, and A. Meyer, "A review of federated learning in renewable energy applications: potential, challenges, and future directions," *Energy AI*, vol. 17, pp. 100375, Sep. 2024.

[32] L. F. Tang, H. P. Xie, X. Y. Wang, and Z. H. Bie, "Privacy-preserving knowledge sharing for few-shot building energy prediction: a federated learning approach," *Appl. Energy*, vol. 337, pp. 120860, May 2023.

[33] S. Lee and D. H. Choi, "Federated reinforcement learning for energy management of multiple smart homes with distributed energy resources," *IEEE Trans. Ind. Inf.*, vol. 18, no. 1, pp. 488–497, Jan. 2022.

[34] H. Zhou, Z. Y. Zhang, D. W. Li, and Z. Su, "Joint optimization of computing offloading and service caching in edge computing-based smart grid," *IEEE Trans. Cloud Comput.*, vol. 11, no. 2, pp. 1122–1132, Apr. /Jun. 2023.

[35] X. Y. Zhou, H. Zhao, Y. H. Cheng, Y. J. Cao, G. Q. Liang, G. L. Liu, W. X. Liu, Y. Xu, and J. H. Zhao, "ElecBench: a power dispatch evaluation benchmark for large language models," arXiv: 2407.05365, 2024.

[36] Y. H. Cheng, H. Zhao, X. Y. Zhou, J. H. Zhao, Y. J. Cao, C. Yang, and X. L. Cai, "A large language model for advanced power dispatch," *Sci. Rep.*, vol. 15, no. 1, pp. 8925, Mar. 2025.

[37] S. L. Choi, R. Jain, P. Emami, K. Wadsack, F. Ding, H. F. Sun, K. Gruchalla, J. Hong, H. M. Zhang, X. Q. Zhu, and B. Kroposki, "eGridGPT: trustworthy AI in the control room," National Renewable Energy Laboratory, Golden, No. NREL/TP-5D00-87740, 2024.

[38] S. L. Choi, R. Jain, C. Feng, P. Emami, H. M. Zhang, J. Hong, T. Kim, S. Park, F. Ding, M. Baggu, and B. Kroposki, "Generative AI for power grid operations," No. National Renewable Energy Laboratory, Golden, NREL/TP-5D00-91176, 2024.

[39] H. F. Hamann, B. Gjorgiev, T. Brunschwiler, L. S. A. Martins, A. Puech, A. Varbella, J. Weiss, J. Bernabe-Moreno, A. B. Massé, S. L. Choi, I. Foster, B. M. Hodge, R. Jain, K. Kim, V. Mai, F. Mirallès, M. De Montigny, O. Ramos-Leaños, H. Suprême, L. Xie, E. N. S. Youssef, A. Zinflou, A. Belyi, R. J. Bessa, B. P. Bhattarai, J. Schmude, and S. Sobolevsky, "Foundation models for the electric power grid," *Joule*, vol. 8, no. 12, pp. 3245–3258, Dec. 2024.

[40] F. Henao, R. Edgell, A. Sharma, and J. Olney, "AI in power systems: a systematic review of key matters of concern," *Energy Inf.*, vol. 8, no. 1, pp. 76, Jun. 2025.

[41] G. Antonesi, T. Cioara, I. Anghel, V. Michalakopoulos, E. Sarmas, and L. Toderean, "From Transformers to Large Language Models: a systematic review of AI applications in the energy sector towards Agentic Digital Twins," arXiv: 2506.06359, 2025.

[42] S. Madani, A. Tavasoli, Z. K. Astaneh, and P.-O. Pineau, "Large language models integration in smart grids," *Energy Reports*, vol. 14, pp. 1562–1577, 2025.

[43] C. Park, "Addressing challenges for the effective adoption of artificial intelligence in the energy sector," *Sustainability*, vol. 17, no. 13, pp. 5764, Jun. 2025.

[44] R. Alsaigh, R. Mehmood, and I. Katib, "AI explainability and governance in smart energy systems: a review," *Front. Energy Res.*, vol. 11, pp. 1071291, Jan. 2023.

[45] G. Mitchell, S. Bahadoorsingh, N. Ramsamooj, and C. Sharma, "A comparison of artificial neural networks and support vector machines for short-term load forecasting using various load types," in *Proceedings of 2017 IEEE Manchester PowerTech*, 2017, pp. 1–4.

[46] D. X. Niu, Q. Wanq, and J. C. Li, "Short term load forecasting model using support vector machine based on artificial neural network," in *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, 2005, pp. 4260–4265.

[47] C. Tarmanini, N. Sarma, C. Gezegin, and O. Ozgonenel, "Short term load forecasting based on ARIMA and ANN approaches," *Energy Rep.*, vol. 9 Suppl 3, pp. 550−557, May 2023.

[48] E. Chodakowska, J. Nazarko, and Ł. Nazarko, "ARIMA models in electrical load forecasting and their robustness to noise," *Energies*, vol. 14, no. 23, pp. 7952, Nov. 2021.

[49] K. Goswami and A. B. Kandali, "Electricity demand prediction using data driven forecasting scheme: ARIMA and SARIMA for real-time load data of Assam," in *Proceedings of 2020 International Conference on Computational Performance Evaluation*, 2020, pp. 570–574.

[50] B. Yang, X. H. Yuan, and F. Tang, "Iterative memory-driven load forecast network model for accuracy improvement," *Energy Rep.*, vol. 9 Suppl 10, pp. 388−395, Oc. t 2023.

[51] J. M. Fan, M. W. Zhong, Y. P. Guan, S. Q. Yi, C. C. Xu, Y. P. Zhai, and Y. W. Zhou, "An online long-term load forecasting method: hierarchical highway network based on crisscross feature collaboration," *Energy*, vol. 299, pp. 131459, Jul. 2024.

[52] Z. H. Pang, F. X. Niu, and Z. O'Neill, "Solar radiation prediction using recurrent neural network and artificial neural network: a case study with comparisons," *Renew. Energy*, vol. 156, pp. 279–289, Aug. 2020.

[53] M. Abdelsattar, M. A. Ismeil, K. Menoufi, A. AbdelMoety, and A. Emad-Eldeen, "Evaluating Machine Learning and Deep Learning models for predicting Wind Turbine power output from environmental factors," *PLoS One*, vol. 20, no. 1, pp. e0317619, Jan. 2025.

[54] B. Bohara, R. I. Fernandez, V. Gollapudi, and X. P. Li, "Short-term aggregated residential load forecasting using BiLSTM and CNN-BiLSTM," in *Proceedings of 2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies*, 2022, pp. 37–43.

[55] S. Smyl, G. Dudek, and P. Pełka, "ES-dRNN: a hybrid exponential smoothing and dilated recurrent neural network model for short-term load forecasting," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 8, pp. 11346–11358, Aug. 2024.

[56] C. Wang, X. Li, Y. Shi, W S. Jiang, Q. Song, and X. L. Li, "Load forecasting method based on CNN and extended LSTM," *Energy Rep.*, vol. 12, pp. 2452–2461, Dec. 2024.

[57] X. J. Zhao, H. M. Peng, L. Y. Zhang, and H. W. Ma, "Research on a short-term power load forecasting method based on a three-channel LSTM-CNN," *Electronics*, vol. 14, no. 11, pp. 2262, May 2025.

[58] K. F. B. Akhter, S. Mobasshira, S. N. Haque, M. A. K. Hesham, and T. Ahmed, "Short-term electricity demand forecasting of Dhaka city using CNN with stacked BiLSTM," arXiv: 2406.06651, 2024.

[59] J. Jang, B. Kim, and I. Kim, "Comparative analysis of deep learning techniques for load forecasting in power systems using single-layer and hybrid models," *Int. Trans. Electr. Energy Syst.*, vol. 2024, no. 1, pp. 5587728, Jun. 2024.

[60] P. K. Quansah and E. K. A. Tenkorang, "Short-term load forecasting using a particle-swarm optimized multi-head attention-augmented CNN-LSTM network," arXiv: 2309.03694, 2023.

[61] W. Y. Guo, S. J. Liu, L. G. Weng, and X. Y. Liang, "Power grid load forecasting using a CNN-LSTM network based on a multi-modal attention mechanism," *Appl. Sci.*, vol. 15, no. 5, pp. 2435, Feb. 2025.

[62] E. A. Tuncar, Ş. Sağlam, and B. Oral, "A review of short-term wind power generation forecasting methods in recent technological trends," *Energy Rep.*, vol. 12, pp. 197–209, Dec. 2024.

[63] M. Yang, Y. T. Huang, C. Y. Xu, C. Y. Liu, and B. Z. Dai, "Review of several key processes in wind power forecasting: mathematical formulations, scientific problems, and logical relations," *Appl. Energy*, vol. 377, pp. 124631, Jan. 2025.

[64] X. L. Li, L. F. Ma, P. Chen, H. Xu, Q. J. Xing, J. H. Yan, S. Y. Lu, H. H. Fan, L. Yang, and Y. Q. Cheng, "Probabilistic solar irradiance forecasting based on XGBoost," *Energy Rep.*, vol. 8 Suppl 5, pp. 1087−1095, Aug. 2022.

[65] K. Q. Wang, L. J. Wang, Q. Meng, C. Yang, Y. S. Lin, J. Y. Zhu, Z. Y. Zhao, C. Zhou, C. H. Zheng, and X. Gao, "Accurate photovoltaic power prediction via temperature correction with physics-informed neural networks," *Energy*, vol. 328, pp. 136546, Aug. 2025.

[66] P. C. Li, F. Guo, Y. F. Li, X. J. Yang, and X. D. Yang, "Physics-informed neural network for real-time thermal modeling of large-scale borehole thermal energy storage systems," *Energy*, vol. 315, pp. 134344, Jan. 2025.

[67] Y. L. Xiao, C. Z. Zou, H. T. Chi, and R. C. Fang, "Boosted GRU model for short-term forecasting of wind power with feature-weighted principal component analysis," *Energy*, vol. 267, pp. 126503, Mar. 2023.

[68] O. Faruque, A. Hossain, S. M. M. Alam, and M. Khalid, "Constraint-aware wind power forecasting with an optimized hybrid machine learning model," *Energy Convers. Manage. : X*, vol. 27, pp. 101026, Jul. 2025.

[69] M. Belletreche, N. Bailek, M. Abotaleb, K. Bouchouicha, B. Zerouali, M. Guermoui, A. Kuriqi, A. H. Alharbi, D. S. Khafaga, M. El-Shimy, and E. S. M. El-Kenawy, "Hybrid attention-based deep neural networks for short-term wind power forecasting using meteorological data in desert regions," *Sci. Rep.*, vol. 14, no. 1, pp. 21842, Sep. 2024.

[70] X. S. Huang, Y. B. Zhang, J. Z. Liu, X. J. Zhang, and S. C. Liu, "A short-term wind power forecasting model based on 3D convolutional neural network–gated recurrent unit," *Sustainability*, vol. 15, no. 19, pp. 14171, Sep. 2023.

[71] R. Sarkar, S. G. Anavatti, T. Dam, M. Pratama, and B. Al Kindhi, "Enhancing wind power forecast precision via multi-head attention transformer: an investigation on single-step and multi-step forecasting," in *Proceedings of 2023 International Joint Conference on Neural Networks*, 2023, pp. 1–8.

[72] H. Liu, H. M. Ma, and T. Y. Hu, "Enhancing short-term wind speed forecasting using graph attention and frequency-enhanced mechanisms," arXiv: 2305.11526, 2023.

[73] X. He, Y. C. Ma, J. C. Xie, G. Zhang, and T. Xie, "Enhanced wind power forecasting using graph convolutional networks with ramp characterization and error correction," *Energies*, vol. 18, no. 11, pp. 2763, May 2025.

[74] F. H. Chen, J. Yan, Y. Q. Liu, Y. M. Yan, and L. B. Tjernberg, "A novel meta-learning approach for few-shot short-term wind power forecasting," *Appl. Energy*, vol. 362, pp. 122838, May 2024.

[75] M. K. Boutahir, A. Hessane, Y. Farhaoui, M. Azrour, M. S. Benyeogor, and N. Innab, "Meta-learning guided weight optimization for enhanced solar radiation forecasting and sustainable energy management with VotingRegressor," *Sustainability*, vol. 16, no. 13, pp. 5505, Jun. 2024.

[76] B. Ladjal, M. Nadour, M. Bechouat, N. Hadroug, M. Sedraoui, A. Rabehi, M. Guermoui, and T. F. Agajie, "Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance: a study on solar potential in Ghardaia, Algeria," *Sci. Rep.*, vol. 15, no. 1, pp. 15404, May 2025.

[77] E. M. Al-Ali, Y. Hajji, Y. Said, M. Hleili, A. M. Alanzi, A. H. Laatar,

and M. Atri, "Solar energy production forecasting based on a hybrid CNN-LSTM-transformer model," *Mathematics*, vol. 11, no. 3, pp. 676, Jan. 2023.

[78] M. L. Bai, X. Y. Zhao, Z. H. Long, J. F. Liu, and D. R. Yu, "Short-term probabilistic photovoltaic power forecast based on deep convolutional long short-term memory network and kernel density estimation," arXiv: 2107.01343, 2021.

[79] Y. F. Zhang, Q. Ai, F. Xiao, R. Hao, and T. F. Lu, "Typical wind power scenario generation for multiple wind farms using conditional improved Wasserstein generative adversarial network," *Int. J. Electr. Power Energy Syst.*, vol. 114, pp. 105388, Jan. 2020.

[80] X. C. Dong, Y. Y. Sun, Y. Yang, and Z. H. Mao, "Controllable renewable energy scenario generation based on pattern-guided diffusion models," *Appl. Energy*, vol. 398, pp. 126446, Nov. 2025.

[81] Z. L. Li, X. G. Peng, W. B. Cui, Y. L. Xu, J. A. Liu, H. L. Yuan, C. S. Lai, and L. L. Lai, "A novel scenario generation method of renewable energy using improved VAEGAN with controllable interpretable features," *Appl. Energy*, vol. 363, pp. 122905, Jun. 2024.

[82] C. O'Connor, M. Bahloul, S. Prestwich, and A. Visentin, "A review of electricity price forecasting models in the day-ahead, intra-day, and balancing markets," *Energies*, vol. 18, no. 12, pp. 3097, Jun. 2025.

[83] G. Marcjasz, M. Narajewski, R. Weron, and F. Ziel, "Distributional neural networks for electricity price forecasting," *Energy Econom.*, vol. 125, pp. 106843, Sep. 2023.

[84] A. Hajigholam Saryazdi, "A novel hybrid deep learning model for electricity price forecasting," *SSRN Electronic Journal*, Art. no. 5166713, 2025.

[85] S. Ghimire, T. Nguyen-Huy, R. C. Deo, D. Casillas-Pérez, A. A. M. Ahmed, and S. Salcedo-Sanz, "Novel deep hybrid model for electricity price prediction based on dual decomposition," *Appl. Energy*, vol. 395, pp. 126197, Oct. 2025.

[86] O. Llorente and J. Portela, "A transformer approach for electricity price forecasting," arXiv: 2403.16108, 2024.

[87] L. C. Han and J. G. Wang, "A regional electricity price prediction method based on Transformer and Graph Neural Networks," *Alexandria Eng. J.*, vol. 128, pp. 52–64, Sep. 2025.

[88] B. W. Zhang, H. D. Tian, A. Berry, and A. C. Roussac, "A local-temporal convolutional transformer for day-ahead electricity wholesale price forecasting," *Sustainability*, vol. 17, no. 12, pp. 5533, Jun. 2025.

[89] T. Y. Zhu, X. Guan, C. Chen, X. J. Cao, C. M. Wang, and J. R. Liao, "Application of artificial intelligence based on state grid ESG platform in clean energy scheduling optimization," *Sci. Rep.*, vol. 14, no. 1, pp. 31304, Dec. 2024.

[90] T. N. Xiao, Y. Chen, H. Diao, S. W. Huang, and C. Shen, "Fast-converging deep reinforcement learning for optimal dispatch of large-scale power systems under transient security constraints," *J. Mod. Power Syst. Clean Energy*, vol. 13, no. 5, pp. 1495–1506, Sep. 2025.

[91] K. Mahapatra, X. Y. Fan, X. Y. Li, Y. Z. Huang, and Q. H. Huang, "Physics informed reinforcement learning for power grid control using augmented random search," in *Proceedings of the 55th Hawaii International Conference on System Sciences*, 2022.

[92] B. de Mol, D. Barbieri, J. Viebahn, and D. Grossi, "Centrally coordinated multi-agent reinforcement learning for power grid topology control," in *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, 2025, pp. 460–475.

[93] F. Alakel and S. Kumar, "GA and PSO techniques based optimal tuning of power system stabilizers in a PV integrated power network," *Eur. J. Eng. Technol. Res.*, vol. 9, no. 5, pp. 1–7, Sep. 2024.

[94] Y. Lee, H. Choi, L. Pagnier, C. H. Kim, J. Lee, B. Jhun, H. Kim, J. Kurths, and B. Kahng, "Reinforcement learning optimizes power dispatch in decentralized power grid," *Chaos Solitons Fractals*, vol. 186, pp. 115293, Sep. 2024.

[95] J. Cao, D. Harrold, Z. Fan, T. Morstyn, D. Healey, and K. Li, "Deep reinforcement learning-based energy storage arbitrage with accurate lithium-ion battery degradation model," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4513–4521, Sep. 2020.

[96] R. K. Bansal, P. C. You, D. F. Gayme, and E. Mallada, "Storage degradation aware economic dispatch," in *Proceedings of 2021 American Control Conference*, 2021, pp. 589–595.

[97] G. N. He, S. Kar, J. Mohammadi, P. Moutis, and J. F. Whitacre, "Power system dispatch with marginal degradation cost of battery storage," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3552–3562, Jul. 2021.

[98] J. O. Lee and Y. S. Kim, "Novel battery degradation cost formulation for optimal scheduling of battery energy storage systems," *Int. J. Electr. Power Energy Syst.*, vol. 137, pp. 107795, May 2022.

[99] H. Khaloie, J. Faraji, F. Vallée, C. S. Lai, J. F. Toubeau, and L. L. Lai, "Risk-aware battery bidding with a novel benchmark selection under second-order stochastic dominance," *IEEE Trans. Ind. Appl.*, vol. 59, no. 3, pp. 3009−3018, May/Jun. 2023.

[100] P. L. García-Miguel, J. Alonso-Martinez, S. A. Gómez, and J. L. Rodríguez-Amenedo, "Impact of risk measures and degradation cost on the optimal arbitrage schedule for battery energy storage systems," *Int. J. Electr. Power Energy Syst.*, vol. 157, pp. 109883, Jun. 2024.

[101] S. S. K. Madahi, B. Claessens, and C. Develder, "Distributional reinforcement learning-based energy arbitrage strategies in imbalance settlement mechanism," *J. Energy Storage*, vol. 104, pp. 114377, Dec. 2024.

[102] M. Sage, J. Campbell, and Y. F. Zhao, "Enhancing battery storage energy arbitrage with deep reinforcement learning and time-series forecasting," in *Proceedings of the ASME 2024 18th International Conference on Energy Sustainability Collocated with the ASME 2024 Heat Transfer Summer Conference and the ASME 2024 Fluids Engineering Division Summer Meeting*, 2024.

[103] J. H. Li, C. L. Wang, Y. R. Zhang, and H. Wang, "Temporal-aware deep reinforcement learning for energy storage bidding in energy and contingency reserve markets," *IEEE Trans. Energy Markets Pol. Reg.*, vol. 2, no. 3, pp. 392–406, Sep. 2024.

[104] Q. L. Ma, W. Wei, and S. W. Mei, "Health-aware coordinate long-term and short-term operation for BESS in energy and frequency regulation markets," *Appl. Energy*, vol. 356, pp. 122363, Feb. 2024.

[105] H. Zhao, Z. F. Liu, X. Mai, J. H. Zhao, J. Qiu, G. L. Liu, Z. Y. Dong, and A. M. Y. M. Ghias, "Mobile battery energy storage system control with knowledge-assisted deep reinforcement learning," *Energy Convers. Econ.*, vol. 3, no. 6, pp. 381–391, Dec. 2022.

[106] Y. T. He, G. C. Ruan, and H. W. Zhong, "Performance-aware control of modular batteries for fast frequency response," *IEEE Trans. Sustain. Energy*, vol. 16, no. 4, pp. 2547–2559, Oct. 2025.

[107] C. H. Antunes, M. J. Alves, and I. Soares, "A comprehensive and modular set of appliance operation MILP models for demand response optimization," *Appl. Energy*, vol. 320, pp. 119142, Aug. 2022.

[108] V. Michalakopoulos, E. Sarmas, I. Papias, P. Skaloumpakas, V. Marinakis, and H. Doukas, "A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs," *Appl. Energy*, vol. 361, pp. 122943, May 2024.

[109] A. Esteban-Perez, D. Bunn, and Y. Ghiassi-Farrokhfal, "Estimating the unobservable components of electricity demand response with inverse optimization," arXiv: 2410.02774, 2024.

[110] F. L. Meng, Q. Ma, Z. X. Liu, and X. J. Zeng, "Multiple dynamic pricing for demand response with adaptive clustering-based customer segmentation in smart grids," *Appl. Energy*, vol. 333, pp. 120626, Mar. 2023.

[111] J. H. Xie, A. Ajagekar, and F. Q. You, "Multi-Agent attention-based deep reinforcement learning for demand response in grid-responsive buildings," *Appl. Energy*, vol. 342, pp. 121162, Jul. 2023.

[112] H. Y. Cheng, T. G. Lu, R. Hao, J. M. Li, and Q. Ai, "Incentive-based demand response optimization method based on federated learning with a focus on user privacy protection," *Appl. Energy*, vol. 358, pp. 122570, Mar. 2024.

[113] J. Q. Wang, Y. Gao, and R. J. Li, "Reinforcement learning based bilevel real-time pricing strategy for a smart grid with distributed energy resources," *Appl. Soft Comput.*, vol. 155, pp. 111474, Apr. 2024.

[114] A. Fraija, N. Henao, K. Agbossou, S. Kelouwani, M. Fournier, and S.

H. Nagarsheth, "Deep reinforcement learning based dynamic pricing for demand response considering market and supply constraints," *Smart Energy*, vol. 14, pp. 100139, May 2024.

[115] S. R. Kumar, A. Easwaran, B. Delinchant, and R. Rigo-Mariani, "Real-time retail electricity pricing using offline reinforcement learning," in *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*, 2024, pp. 454–458.

[116] D. Sloot, N. Lehmann, A. Ardone, and W. Fichtner, "A behavioral science perspective on consumers' engagement with demand response programs," *Energy Research Letters*, vol. 4, no. 1, 2023.

[117] J. Ryu and J. Kim, "A prospect-theoretic game approach to demand response market participation through energy sharing in energy storage systems under uncertainty," *Energy Rep.*, vol. 9, pp. 1093–1103, Dec. 2023.

[118] Federal Energy Regulatory Commission. (2024, Dec.). 2024 Assessment of Demand Response and Advanced Metering: Staff Report Pursuant to Section 1252(e)(3) of the Energy Policy Act of 2005, Washington, DC, USA, Dec. 2024. [Online]. Available: https://www.ferc.gov/sites/default/files/2024-11/Annual%20Assessment%20of%20Demand%20Response_1119_1400.pdf

[119] Z. H. Wang, B. Lu, B. Wang, Y. M. Qiu, H. Shi, B. Zhang, J. Y. Li, H. Li, and W. H. Zhao, "Incentive based emergency demand response effectively reduces peak load during heatwave without harm to vulnerable groups," *Nat. Commun.*, vol. 14, no. 1, pp. 6202, Oct. 2023.

[120] Y. F. Shao, S. Fan, Y. H. Meng, K. Q. Jia, and G. Y. He, "Personalized demand response based on sub-CDL considering energy consumption characteristics of customers," *Appl. Energy*, vol. 374, pp. 123964, Nov. 2024.

[121] M. Eirinaki, I. Varlamis, J. Dahihande, A. Jaiswal, A. A. Pagar, and A. Thakare, "Real-time recommendations for energy-efficient appliance usage in households," *Front. Big Data*, vol. 5, pp. 972206, Jun. 2022.

[122] D. Cao, J. B. Zhao, W. H. Hu, N. P. Yu, F. Ding, Q. Huang, and Z. Chen, "Deep reinforcement learning enabled physical-model-free two-timescale voltage control method for active distribution systems," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 149–165, Jan. 2022.

[123] F. Kabir, N. P. Yu, Y. Q. Gao, and W. Y. Wang, "Deep reinforcement learning-based two-timescale Volt-VAR control with degradation-aware smart inverters in power distribution systems," *Appl. Energy*, vol. 335, pp. 120629, Apr. 2023.

[124] J. H. Wang, W. K. Xu, Y. J. Gu, W. B. Song, and T. C. Green, "Multi-agent reinforcement learning for active voltage control on power distribution networks," in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2021, pp. 250.

[125] R. Hossain, M. Gautam, M. MansourLakouraj, H. Livani, and M. Benidris, "Multi-agent deep reinforcement learning-based volt-VAR control in active distribution grids," in *Proceedings of 2023 IEEE Power & Energy Society General Meeting*, 2023, pp. 1–5.

[126] P. P. Yu, H. C. Zhang, Y. H. Song, Z. Y. Wang, H. Y. Dong, and L. Ji, "Safe reinforcement learning for power system control: a review," *Renew. Sustain. Energy Rev.*, vol. 223, pp. 116022, Nov. 2025.

[127] Y. J. Wang and Z. Wu, "Control Lyapunov-barrier function-based safe reinforcement learning for nonlinear optimal control," *AIChE J.*, vol. 70, no. 3, pp. e18306, Mar. 2024.

[128] B. Zhang, D. Cao, W. H. Hu, A. M. Y. M. Ghias, and Z. Chen, "Physics-Informed Multi-Agent deep reinforcement learning enabled distributed voltage control for active distribution network using PV inverters," *Int. J. Electr. Power Energy Syst.*, vol. 155, pp. 109641, Jan. 2024.

[129] S. Choi, R. Jain, and H. M. Zhang, "Digital twin+ AI: control room of the future," National Renewable Energy Laboratory, Golden, No. NREL/PR-5D00-89725, 2024.

[130] U.S. Department of Energy. (2024, Apr.). AI for Energy: Opportunities for a Modern Grid and Clean Energy Economy, Washington, DC, USA, 2024. [Online]. Available: https://www.energy.gov/sites/default/files/2024-04/AI%20EO%20Report%20Section%205.2g%28i%29_043024.pdf

[131] ENTSO-E. (2024). ENTSO-E RDI Roadmap 2024 – 2034. Innovation missions to build the power system for a carbon-neutral Europe. [Online]. Available: https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/Publications/RDC%20publications/entso-e_RDI_roadmap_2024–2034_240710.pdf

[132] J. Xie and W. Sun, "Distributional deep reinforcement learning-based emergency frequency control," *IEEE Trans. Power Syst.*, vol. 37, no. 4, pp. 2720–2730, Jul. 2022.

[133] Y. M. Zheng, J. Tao, Q. L. Sun, H. Sun, Z. Q. Chen, and M. W. Sun, "Deep reinforcement learning based active disturbance rejection load frequency control of multi-area interconnected power systems with renewable energy," *J. Franklin Inst.*, vol. 360, no. 17, pp. 13908–13931, Nov. 2023.

[134] H. O. Cao, C. B. Zhou, Y. H. Meng, J. X. Shen, and X. Y. Xie, "Advancement in transformer fault diagnosis technology," *Front. Energy Res.*, vol. 12, pp. 1437614, Jul. 2024.

[135] Suwarno, H. Sutikno, R. A. Prasojo, and A. Abu-Siada, "Machine learning based multi-method interpretation to enhance dissolved gas analysis for power transformer fault diagnosis," *Heliyon*, vol. 10, no. 4, pp. e25975, Feb. 2024.

[136] Z. B. Qiu, X. Zhu, C. B. Liao, D. Z. Shi, and W. Q. Qu, "Detection of transmission line insulator defects based on an improved lightweight YOLOv4 model," *Appl. Sci.*, vol. 12, no. 3, pp. 1207, Jan. 2022.

[137] L. H. Chen, "Defect identification of electricity transmission line insulators based on the improved lightweight network model with computer vision assistance," *Heliyon*, vol. 10, no. 10, pp. e30405, May 2024.

[138] Z. B. Tang and X. Jian, "Thermal fault diagnosis of complex electrical equipment based on infrared image recognition," *Sci. Rep.*, vol. 14, no. 1, pp. 5547, Mar. 2024.

[139] Y. C. Liu, L. Yang, A. Ghasemkhani, H. Livani, V. A. Centeno, P. Y. Chen, and J. S. Zhang, "Robust event classification using imperfect real-world PMU data," *IEEE Internet Things J.*, vol. 10, no. 9, pp. 7429–7438, May 2023.

[140] G. Gök, Ö. Salor, and M. C. Taplamacıoğlu, "Transient event classification using pmu data with deep learning techniques and synthetically supported training-set," *IET Gener. Trans. Dist.*, vol. 17, no. 6, pp. 1287–1297, Mar. 2023.

[141] M. MansourLakouraj, M. Gautam, H. Livani, and M. Benidris, "A multi-rate sampling PMU-based event classification in active distribution grids with spectral graph neural network," *Electr. Power Syst. Res.*, vol. 211, pp. 108145, Oct. 2022.

[142] H. B. Sun, S. Kawano, D. Nikovski, T. Takano, and K. Mori, "Distribution fault location using graph neural network with both node and link attributes," in *Proceedings of 2021 IEEE PES Innovative Smart Grid Technologies Europe*, 2021.

[143] M. S. Jia, Z. Y. Cui, and G. Hug, "Enabling large language models to perform power system simulations with previously unseen tools: a case of Daline," arXiv: 2406.17215, 2024.

[144] B. R. Zhang, C. J. Li, G. Chen, and Z. Y. Dong, "Large language model assisted optimal bidding of BESS in FCAS market: an AI-agent based approach," arXiv: 2406.00974, 2024.

[145] L. Jing and A. Rahman, "Fault diagnosis in power grids with large language model," arXiv: 2407.08836, 2024.

[146] A. J. Dave, T. N. Nguyen, and R. B. Vilim, "Integrating LLMs for explainable fault diagnosis in complex systems," arXiv: 2402.06695, 2024.

[147] C. Liu, G. Dalzell, E. Kranz, X. H. Yu, A. Kelly, and M. Almashor, "Real-time machine learning for power grid SCADA alarm event detection decision support," in *Proceedings of the IECON 2024–50th Annual Conference of the IEEE Industrial Electronics Society*, 2024, pp. 1–6.

[148] Y. S. Chen and A. Anderson, "Connecting Minds: AI Use Cases to Bridge Power Systems and Large Language Models for Practical Applications," Pacific Northwest National Laboratory, Richland, PNNL-38003, 2025.

[149] J. Q. Ruan, G. Q. Liang, H. Zhao, G. L. Liu, X. Z. Sun, J. Qiu, Z. Xu, F. S. Wen, and Z. Y. Dong, "Applying large language models to power

systems: potential security threats," *IEEE Trans. Smart Grid*, vol. 15, no. 3, pp. 3333–3336, May 2024.

[150] S. Majumder, L. Dong, F. Doudi, Y. T. Cai, C. Tian, D. Kalathil, K. Ding, A. A. Thatte, N. Li, and L. Xie, "Exploring the capabilities and limitations of large language models in the electric energy sector," *Joule*, vol. 8, no. 6, pp. 1544–1549, Jun. 2024.

[151] P. Zablocki and Z. Gajewska, "Assessing hallucination risks in large language models through internal state analysis," *Authorea Preprints*, Jul. 17, 2024.

[152] U. Nawaz, M. Anees-ur-Rahaman, and Z. Saeed, "A review of neuro-symbolic AI integrating reasoning and learning for advanced cognitive systems," *Intell. Syst. Appl.*, vol. 26, pp. 200541, Jun. 2025.

[153] M. Gaur and A. Sheth, "Building trustworthy NeuroSymbolic AI Systems: consistency, reliability, explainability, and safety," *AI Mag.*, vol. 45, no. 1, pp. 139–155, Feb. 2024.

[154] R. J. Somers, J. A. Douthwaite, D. J. Wagg, N. Walkinshaw, and R. M. Hierons, "Digital-twin-based testing for cyber–physical systems: a systematic literature review," *Inf. Software Technol.*, vol. 156, pp. 107145, Apr. 2023.

[155] Y. Shi, P. F. Wei, K. Feng, D. C. Feng, and M. Beer, "A survey on machine learning approaches for uncertainty quantification of engineering systems," *Mach. Learn. Computat. Sci. Eng.*, vol. 1, no. 1, pp. 11, Jan. 2025.

[156] J. Arbel, K. Pitas, M. Vladimirova, and V. Fortuin, "A primer on Bayesian neural networks: review and debates," *Statistical Science*, vol. 41, no. 2, pp. 316–353, 2026.

[157] A. Younesi, P. Siano, A. Moradpour, and A. Mehrizi-Sani, "Hybrid neuro-symbolic learning and reasoning for resilient load restoration in smart microgrids," *Renewable Energy*, vol. 256, pp. 124401, Jan. 2026.

[158] M. Mora-Cantallops, S. Sánchez-Alonso, E. García-Barriocanal, and M. A. Sicilia, "Traceability for trustworthy AI: a review of models and tools," *Big Data Cogn. Comput.*, vol. 5, no. 2, pp. 20, May 2021.

[159] B. Eken, S. Pallewatta, N. Tran, A. Tosun, and M. A. Babar, "A multivocal review of MLOps practices, challenges and open issues," *ACM Comput. Surv.*, vol. 58, no. 2, pp. 39, Jan. 2026.

[160] D. Pessach and E. Shmueli, "A review on fairness in machine learning," *ACM Comput. Surv.* (*CSUR*), vol. 55, no. 3, pp. 51, Mar. 2023.

[161] Z. Y. Zhang, M. X. Liu, M. Y. Sun, R. L. Deng, P. Cheng, D. Niyato, M. Y. Chow, and J. M. Chen, "Vulnerability of machine learning approaches applied in IoT-based smart grid: a review," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 18951–18975, Jun. 2024.

[162] Z. K. Hao, S. M. Liu, Y. C. Zhang, C. Y. Ying, Y. Feng, H. Su, and J. Zhu, "Physics-informed machine learning: a survey on problems, methods and applications," arXiv: 2211.08064, 2022.

[163] Y. W. Xu, S. Kohtz, J. Boakye, P. Gardoni, and P. F. Wang, "Physics-informed machine learning for reliability and systems safety applications: state of the art and challenges," *Reliab. Eng. Syst. Saf.*, vol. 230, pp. 108900, Feb. 2023.

[164] R. Kabir, D. Halder, and S. Ray, "Digital twins for IoT-driven energy systems: a survey," *IEEE Access*, vol. 12, pp. 177123–177143, Nov. 2024.

[165] S. Keren, C. Essayeh, S. V. Albrecht, and T. Morstyn, "Multi-agent reinforcement learning for energy networks: computational challenges, progress and open problems," arXiv: 2404.15583, 2024.

[166] M. A. M. Yassin, A. Shrestha, and S. Rabie, "Digital twin in power system research and development: principle, scope, and challenges," *Energy Rev.*, vol. 2, no. 3, pp. 100039, Sep. 2023.

[167] M. M. Thwe, A. Ştefanov, V. S. Rajkumar, and P. Palensky, "Digital twins for power systems: review of current practices, requirements, enabling technologies, data federation, and challenges," *IEEE Access*, vol. 13, pp. 105517–105540, Jun. 2025.

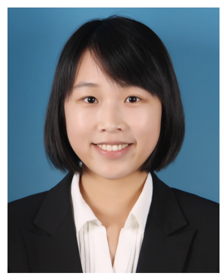
[168] E. Iraola, M. García-Lorenzo, F. Lordan-Gomis, F. Rossi, E. Prieto-Araujo, and R. M. Badia, "HP2C-DT: high-precision high-performance computer-enabled digital twin," *Future Generation Computer Systems*, vol. 180, Art. no. 108333, 2026.

[169] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. F. Wang, and L. Yang, "Physics-informed machine learning," *Nat. Rev. Phys.*, vol. 3, no. 6, pp. 422–440, May 2021.

[170] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric deep learning: grids, groups, graphs, geodesics, and gauges," arXiv: 2104.13478, 2021.

[171] C. Y. Li, Z. Gan, Z. Y. Yang, J. W. Yang, L. J. Li, L. J. Wang, and J. F. Gao, "Multimodal foundation models: from specialists to general-purpose assistants," *Found. Trends® Comput. Graph. Vision*, vol. 16, no. 1–2, pp. 1–214, May 2024.

Tao Huang received his B.S. and M.E. degrees from Shanghai Jiao Tong University, Shanghai, China, in 2002 and 2007, respectively, and his Ph.D. degree from Politecnico di Torino, Italy, in 2011. He is currently an Associate Professor in the Department of Energy at Politecnico di Torino. His research interests focus on electricity market, smart grids, power system operation, and optimization.



Jingsi Chen received her M.Sc. degree in ICT Engineering for Smart Societies from Politecnico di Torino, Italy, in 2024. She is currently pursuing the Ph.D. degree in the Department of Energy, Politecnico di Torino. Her research interests focus on electricity market, optimization and the intersection of artificial intelligence and power system.

Jun Yin received his M.Sc. degree in ICT Engineering for Smart Societies from Politecnico di Torino, Italy, in 2024. He is currently pursuing the Ph.D. degree in the Department of Electronics and Telecommunications, Politecnico di Torino. His research interests lie at the intersection of artificial intelligence and energy systems, including edge artificial intelligence, electricity markets, and agent-based modeling.

Michela Meo received her Ph.D. degree in Electronic Engineering from Politecnico di Torino, Italy, in 1997. She is currently a Full Professor (since June 2016) with the Department of Electronics and Telecommunications at Politecnico di Torino. Her research interests include green networking, energy efficiency in communication networks, performance evaluation, traffic modeling, and wireless networks.