Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

PEAKSAFE-AI: Explainable Multi-Agent Peak-Demand Forecasting with Uncertainty-Aware Assurance and Human-in-the-Loop Decision Support

Stephen Kamran

kamran.msbf@gmail.com

Government College University, Lahore https://orcid.org/0009-0000-3063-5611

Farah Saman

University of Exeter https://orcid.org/0000-0003-0933-4368

Research Article

Keywords: Electricity demand forecasting, peak-load warning, XGBoost, SHAP, prediction intervals, multi-agent systems, human-in-the-loop, information latency

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11574304/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

PEAKSAFE-AI: Explainable Multi-Agent Peak-Demand Forecasting with Uncertainty-Aware Assurance and Human-in-the-Loop Decision Support

Stephen Kamran
Government College University Lahore
Lahore, Pakistan
kamran.msbf@gmail.com

Farah Saman
University of Exeter
Exeter, United Kingdom
fs531@exeter.ac.uk

Abstract—Accurate electricity-demand forecasts do not necessarily translate into trustworthy peak-risk decisions, particularly when demand measurements are delayed and uncertainty overlaps an operational warning threshold. This study develops PEAKSAFE-AI, an explainable three-agent decision-support prototype integrating XGBoost forecasting, TreeSHAP evidence attribution, and validation-calibrated interval-based assurance. The retrospective evaluation uses 29,998 half-hourly records from the UK National Energy System Operator for 2025–2026, with chronological training, validation, winter and summer evaluation periods. Two forecast-origin information regimes distinguish current-demand availability from a one-period delay, each compared with its information-matched persistence baseline. Winter mean absolute error is 298.39 MW under immediate information and 531.34 MW under delayed information, versus corresponding persistence errors of 803.04 and 1,540.06 MW. At a training-defined extreme-demand threshold of 37,565.30 MW, assurance escalation recognizes 68/70 and 65/70 winter episode onsets, compared with 47/70 and 43/70 from point-forecast alerts. The gains require additional warning periods and simulated human-review workload. PEAKSAFE-AI demonstrates a transparent forecast-to-escalation workflow, while remaining an offline prototype rather than a live grid-control system.

Index Terms—*electricity demand forecasting, peak-load warning, XGBoost, SHAP, prediction intervals, multi-agent systems, human-in-the-loop, information latency.*

I. INTRODUCTION

Electricity-demand forecasts guide operational preparedness, reserve planning, and risk-sensitive decision support. Yet a small point-error metric does not reveal whether a model detects the start of an extreme-demand episode, how much uncertainty remains near a warning threshold, or which information was actually available at forecast origin. Recent research emphasizes peak-load forecasting, interpretable electricity-demand prediction, and trustworthy human–AI energy management [1]–[6].

PEAKSAFE-AI addresses a focused question: can a transparent, coordinated forecasting and assurance workflow improve the recognition of extreme-demand episode onsets while exposing the effect of delayed measurements and human-review workload? Unlike autonomous dispatch or control, the system supplies alerts and review recommendations for retrospective testing only.

The contributions are fourfold: (i) information-matched immediate/delayed forecasting and persistence baselines; (ii) chronological validation-based uncertainty intervals and an explicit three-way risk rule; (iii) winter episode-onset analysis with warning-workload trade-offs; and (iv) a reproducible three-agent prototype with TreeSHAP summaries and controlled demand-shift sensitivity.

II. RELATED WORK AND RESEARCH GAP

Contemporary short-term and peak-demand studies employ boosted trees, hybrid learning, and SHAP-style explanations [1]–[4], [7], [8]. Related grid-operations and smart-meter XAI studies demonstrate applications beyond predictive accuracy [9], [10], while research on uncertainty-sensitive explanations and counterfactual cyberattacks highlights additional robustness considerations [11], [12]. Probabilistic load forecasting research further emphasizes adaptive and local conformal interval methods [13], [14]. Although these streams motivate accurate or interval-aware forecasts, reporting their metrics separately does not establish the operational consequence of a threshold-adjacent prediction.

Recent research has also extended electricity-demand forecasting through explainable data-driven architectures [15] and quantile-based neural prediction methods [16]. Adaptive conformal calibration further offers a complementary approach to managing forecast uncertainty across electricity-network levels [17]. However, accurate point forecasts, predictive uncertainty, and model interpretability do not independently establish whether a decision-support system can recognize the onset of extreme-demand episodes while managing unnecessary warnings and human-review requirements.

Human–AI energy management and agentic-systems taxonomies distinguish analysis from authorized intervention [5], [18], [19]. Trustworthy-AI and systemic-assurance studies similarly motivate inspectable response policies [20], [21]. More broadly, risk governance, operational XAI monitoring, Industry 5.0, and human–AI collaboration motivate transparent accountability [22]–[25]. The specific gap addressed here is the joint comparison of forecast-origin information, interval-based warning decisions, episode-onset detection, and simulated operator-review burden using a common chronological electricity-demand evaluation.

III. PEAKSAFE-AI METHODOLOGY

- *A. Data and temporal partitioning*

We use the UK National Energy System Operator (NESO) historic demand files, containing 17,520 half-hour records from 2025 and 12,478 from 2026 (through 17 September), totalling 29,998 raw settlement records [26]. After lag/rolling-feature warm-up and next-period target alignment, the design

partitions the observations by forecast-target date as in Table I. Training-only quantiles define high (Q80 = 29,362.20 MW), very high (Q90 = 33,926.60 MW), and extreme (Q95 = 37,565.30 MW) demand. These fixed cutoffs are not recalibrated on test data.

TABLE I
CHRONOLOGICAL EVALUATION PARTITIONS

Period	Target dates	N
Training	08 Jan–31 Oct 2025	14,255
Validation	01 Nov–31 Dec 2025	2,928
Winter holdout	01 Jan–31 Mar 2026	4,318
Summer holdout	01 Apr–17 Sep 2026	8,160

One additional training row is lost when constructing delayed rolling features; the final comparison fits on 14,254 observations, whereas thresholds retain the original training-only 14,255-row source. This explains the 1.10-MW Q90 quantile difference between the two training subsets. NESO historic records are published with reporting latency, so the immediate-information regime is a retrospective observability assumption, not established live availability [26].

- *B. Forecast origin and model training*

The forecast target is next-period national demand, y(t+1). The immediate regime includes ND(t) and historical/calendar features; the delayed regime excludes ND(t), using information available through ND(t−1) and correspondingly shifted rolling statistics. The immediate and delayed models have 22 and 21 predictors, respectively. Each XGBoost regressor uses 350 estimators, depth 5 and learning rate 0.05; other implementation settings are fixed in the notebook. Both are evaluated against regime-matched persistence predictions. Model forecasts and baseline forecasts are compared via MAE, RMSE and signed bias. This design probes one-period measurement latency, not future covariate availability guaranteed in live operations.

$$\hat{y}_{t+1}^{(r)} = f_r(X_t^{(r)}), \quad r \in \{\text{immediate, delayed}\} \quad (1)$$

- *C. Uncertainty-aware assurance*

On the held-out 2025 validation period (n = 2,928), each regime calibrates a symmetric absolute-residual radius r to target approximately 90% marginal interval coverage. For forecast ŷ(t+1), the interval is [L,U] = [ŷ−r, ŷ+r], with r = 660.91 MW (immediate) or 1,228.46 MW (delayed). At threshold τ = Q95, the rule is SAFE if U < τ; AUTO ALERT if L ≥ τ; otherwise HUMAN REVIEW. "Escalation" means automatic alert or human review. These intervals use a fixed validation residual quantile; adaptive conformal work on renewable-energy forecasting provides methodological context [27], but coverage under seasonal distribution shift is not guaranteed.

$$I_t^{(r)} = [\hat{y}_{t+1}^{(r)} - q_r, \hat{y}_{t+1}^{(r)} + q_r] \quad (2)$$

$$D_t = \begin{cases} \text{SAFE}, & U_t < \tau \\ \text{AUTO ALERT}, & L_t \geq \tau \\ \text{HUMAN REVIEW}, & L_t < \tau \leq U_t \end{cases} \quad (3)$$

- *D. Coordinated agents and explainability*

The Forecast Agent produces point and validation-calibrated interval forecasts; the XAI Agent generates TreeSHAP contributions and identifies dominant predictive features; and the Assurance Agent translates forecast uncertainty into three traceable decision labels: SAFE, AUTO ALERT, and HUMAN REVIEW. This modular architecture integrates multi-agent coordination principles [18], [19] with explainability and systemic AI assurance [20], [21]. By separating predictive estimation, evidence attribution, and risk-sensitive decision governance, PEAKSAFE-AI extends conventional accuracy-focused forecasting toward transparent, uncertainty-aware operational decision support.

Importantly, XAI evidence supports interpretation and auditability without modifying the predefined threshold gate, preserving the distinction between explanatory information and decision authority [10], [23]. Prediction intervals overlapping the training-derived extreme-demand threshold trigger simulated human review, consistent with human–AI collaboration and accountable oversight principles [5], [25]. Each agent produces structured outputs that support reproducible case-level traceability, uncertainty assessment, and retrospective evaluation. The framework remains an offline decision-support prototype: no actual operator decisions, autonomous grid interventions, or real-world control actions are performed (Fig. 1).

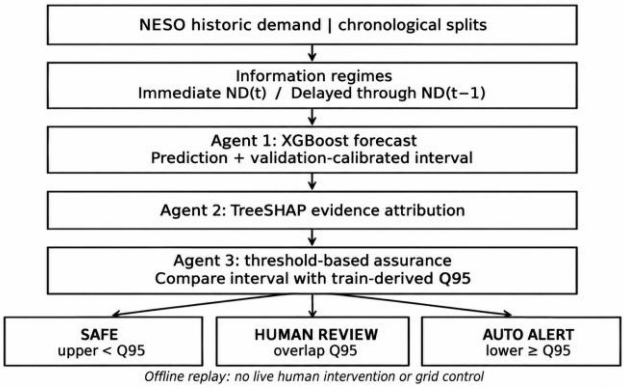


Fig. 1. PEAKSAFE-AI offline three-agent forecasting and assurance architecture.

- *E. Peak episodes and sensitivity checks*

An extreme-demand episode is a contiguous run of winter observations with actual ND ≥ τ. The winter holdout contains 632 such periods across 70 episodes (mean duration: 9.03 half-hours). We compare episode-onset recognition under point-forecast threshold alerts and uncertainty-aware interval escalation, assessing their ability to identify emerging extreme-demand events. Additionally, demand-history predictors are synthetically shifted by ±250/500 MW to examine forecast sensitivity and decision-label stability; these controlled perturbations do not represent validated physical disturbances or adversarial attack scenarios.

$$\text{Onset recognition} = \text{detected episode onsets} / 70 \quad (4)$$

IV. EXPERIMENTAL RESULTS

- *A. Information-matched forecast performance*

TABLE II
FORECAST PERFORMANCE (MAE / RMSE, MW)

Holdout	Model	MAE	RMSE
Val.	Imm. persistence	764.43	991.94
Val.	Imm. XGBoost	314.05	399.78
Val.	Delay persistence	1,468.25	1,891.49
Val.	Delay XGBoost	585.09	740.90
Winter	Imm. persistence	803.04	1,033.99
Winter	Imm. XGBoost	298.39	393.07
Winter	Delay persistence	1,540.06	1,975.34
Winter	Delay XGBoost	531.34	685.79
Summer	Imm. persistence	591.55	764.52
Summer	Imm. XGBoost	287.98	368.92
Summer	Delay persistence	1,118.64	1,427.05
Summer	Delay XGBoost	460.23	587.79

Relative to the appropriate persistence comparator, winter MAE falls by 62.84% (immediate) and 65.50% (delayed). Delayed measurements increase XGBoost winter MAE from 298.39 to 531.34 MW; direct comparisons of the models without matching baseline information would obscure this constraint. The summer distribution differs markedly: no Q95 event occurs among 8,160 summer observations, so summer extreme-peak recall is undefined rather than 100%.

- *B. Assurance performance and workload*

TABLE III
WINTER Q95 ASSURANCE (4,318 PERIODS)

Metric	Immediate	Delayed
Overall coverage	91.43%	92.36%
Peak-only coverage	85.13%	88.77%
Safe decisions	3,551	3,437
Auto alerts	501	407
Human reviews	266	474
Unsafe safe releases	2	5
False auto alerts	0	1
Peak escalation recall	99.68%	99.21%

The immediate regime routes 266/4,318 winter cases (6.16%) to review, versus 474 (10.98%) with delayed information. Of these reviews, 129 and 221 concern actual Q95 periods, respectively. Because the method deliberately escalates uncertain borderline cases, these are simulated review demands rather than measured human workload or intervention efficacy. Summer yields no Q95 auto alerts or reviews under either regime.

- *C. Operational interpretation of interval routing*

At the Q95 cutoff, the gate explicitly separates three different informational states. When the upper interval bound remains below 37,565.30 MW, SAFE communicates that the entire calibrated range is below the extreme threshold; a rare extreme actual value can nevertheless occur outside that range. When the lower bound exceeds the cutoff, AUTO ALERT provides a high-confidence threshold warning. A crossing interval is routed to HUMAN REVIEW because a point prediction alone would suppress ambiguity. This behavior should be interpreted as an audit-friendly prototype policy rather than a cost-optimal operating rule. The immediate winter false-auto-alert count is zero; this does not imply perfect alert specificity for other risk thresholds, seasons, or grids.

The asymmetric error cost is central, particularly when incomplete or delayed demand information affects forecasting reliability [28]. Related probabilistic energy-forecasting research also emphasizes uncertainty quantification as an important consideration in risk-sensitive prediction [13], [14], [27]. In electricity-system operations, missed extreme-demand events may undermine timely preparedness, whereas excessive precautionary warnings can increase operator workload and reduce the practical usefulness of decision-support systems. These competing consequences motivate evaluating not only predictive accuracy but also how uncertainty influences the assignment of operational warning categories.

In the immediate winter evaluation, only two actual Q95 observations were labelled SAFE, while 137 of the 266 review cases were not extreme. The delayed regime routed five actual extremes to SAFE, and 253 of 474 reviews were non-extreme. These findings illustrate how restricted information availability coincides with wider prediction intervals and greater simulated review demand. Importantly, HUMAN REVIEW indicates unresolved threshold uncertainty rather than a confirmed extreme-demand event, while SAFE does not guarantee that an extreme observation cannot occur. Consistent with trustworthy-AI assurance and human–AI collaboration principles [5], [20], [21], operational deployment would require explicit warning lead-time requirements, acceptable missed-peak rates, operator-review capacity, and an assessment of the asymmetric costs of false warnings and missed events. PEAKSAFE-AI therefore provides a transparent basis for examining safety–workload trade-offs rather than an empirically optimized grid-intervention policy.

- *D. Extreme-demand episode recognition*

TABLE IV
WINTER EXTREME-DEMAND EPISODE POLICIES

Regime / policy	Episodes	Onsets	False warn.
Imm. point	65/70	47/70	31
Imm. assure	69/70	68/70	137
Delay point	61/70	43/70	43
Delay assure	69/70	65/70	254

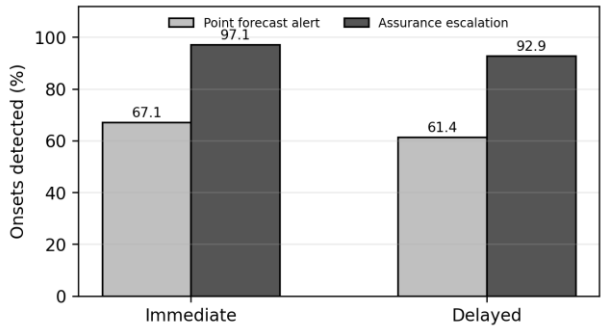


Fig. 2. Winter Q95 episode-onset recognition (70 episodes per regime).

Onset recognition rises from 67.14% to 97.14% under immediate information (+30.00 percentage points) and from

61.43% to 92.86% under delayed information (+31.43 points). These improvements are accompanied by 106 and 211 extra false-warning periods, respectively, relative to point alerts. Since the escalation rule contains the point-alert decision set, nonnegative onset gains are partly structural; reported paired tests should therefore be viewed as exploratory policy comparisons rather than evidence of independent causal superiority.

- *E. Explanations and synthetic perturbations*

TreeSHAP attribution identifies current ND as a dominant driver in the immediate information regime: 73.56% of the summed mean absolute attribution across all winter observations and 81.53% for winter extreme periods. The delayed regime necessarily lacks ND(t). A +500-MW correlated shift in historical-demand predictors raises mean forecasts by 480.72 MW (immediate) and 476.33 MW (delayed), changing 4.24% and 4.75% of assurance labels. These observations describe sensitivity to specified input edits, not causal demand effects or verified grid resilience.

- *F. Three-agent case-level traceability*

Representative winter cases demonstrate the joined workflow. Under immediate information, a true extreme of 38,278 MW receives a forecast of 38,041.97 MW, dominant SHAP evidence from ND(t) (79.43% of local absolute contribution), and a HUMAN REVIEW outcome because its interval [37,381.06, 38,702.89] MW overlaps Q95. Under delayed information, the same target receives a forecast of 36,846.57 MW, with nd_lag_1 explaining 69.02% of absolute local contribution; its wider interval also crosses Q95 and triggers review. A normal observation of 25,107 MW is classified SAFE under both regimes. These trace examples establish reproducible routing, but local SHAP contributions are descriptive and cannot establish physical causation.

The complete prototype audit contains 24,956 records across two regimes and both 2026 seasonal partitions, including 740 simulated review entries. Internal decision-count checks passed in the notebook. These audit trails illustrate inspectability and reproducibility of a software pipeline; they should not be described as field-validated compliance or evidence of an actual human-in-the-loop intervention.

V. DISCUSSION AND LIMITATIONS

The results distinguish predictive accuracy from decision suitability. Relative to matched baselines, boosted forecasting is consistently more accurate; nevertheless, a one-period observation delay widens validation-calibrated uncertainty substantially and increases review demand. Interval escalation recognizes more Q95 episode onsets than point-only alerts, but creates more warnings on ordinary periods. The useful output is consequently a transparent risk–workload trade-off, not a claim that the assurance policy is universally optimal.

Four limitations are important. First, NESO historical values do not prove that current settlement measurements were observable at forecast origin; only the delayed regime approximates a stricter information constraint. Second, residual-quantile calibration provides empirical validation-based intervals but no universal coverage guarantee under shift, particularly in extreme winter demand. Third, thresholds and post-hoc experimental variants require future prospective preregistration or rolling-origin replication. Fourth, agents are orchestrated software modules, the review queue is simulated, and no human decisions, grid control, operational benefits, or protection against genuine attacks were evaluated.

The PEAKSAFE-AI framework also holds practical relevance for electricity systems in the Global South, particularly Pakistan, where ongoing investments in grid stability and transmission infrastructure underscore the need for reliable, transparent, and risk-aware power-system decision support [29], [30]. As electricity networks modernize, frameworks such as PEAKSAFE-AI can inform the integration of explainable AI-driven demand forecasting, uncertainty-calibrated peak-risk assessment, information-latency awareness, and human-in-the-loop governance into operational planning and grid-monitoring processes. By distinguishing routine demand conditions from high-risk and uncertain peak events, the proposed three-agent architecture offers a transparent basis for prioritizing warnings and supporting operator oversight without replacing human decision authority. Although empirically evaluated using UK electricity-demand data, its potential application to emerging-market power systems requires further validation using local demand patterns, measurement availability, operational constraints, and operator requirements.

Future work should use as-published availability timestamps, rolling-origin recalibration, weather and calendar forecasts with verified issue times, cost-sensitive alarm calibration, and operator-in-the-loop trials. The framework could ultimately support energy-system decision assurance, including emerging-market grids, but cross-system performance must be validated independently.

VI. CONCLUSION

PEAKSAFE-AI integrates next-period electricity-demand forecasting, TreeSHAP evidence, and uncertainty-aware three-way assurance within a three-agent offline research prototype. On the winter holdout, XGBoost MAE was 298.39 MW under immediate information and 531.34 MW under delayed information, outperforming information-matched persistence baselines. Assurance escalation recognized 68/70 and 65/70 extreme-demand episode onsets, versus 47/70 and 43/70 for point alerts, while increasing warning burden and simulated review activity. These findings motivate jointly reporting forecast-origin constraints, uncertainty, episode onset, and review workload before operational deployment.

REFERENCES

[1] S. Dai, F. Meng, H. Dai, Q. Wang, X. Chen, W. Bai, et al., "Machine learning in peak demand forecasting: Foundations, trends, and insights," Renew. Sustain. Energy Rev., vol. 227, Art. no. 116500, 2026, doi: 10.1016/j.rser.2025.116500.

[2] S. Vasudevan and K. Jothinathan, "Real-time efficient short-term peak load and day-ahead electricity load forecasting system using machine learning approach," Energy, vol. 342, Art. no. 139508, 2026, doi: 10.1016/j.energy.2025.139508.

[3] J. Wang, B. Yu, X. Chen, G. Dai, G. Dai, W. Liu, et al., "An interpretable short-term electrical load forecasting model based on SHapley Additive exPlanations—A case study in Haidian, Beijing," Electr. Power Syst. Res., vol. 247, Art. no. 111769, 2025, doi: 10.1016/j.epsr.2025.111769.

[4] J. Park, J. Lee, and E. Kim, "Month-conditioned boosting framework with SHAP-in-the-loop for short-term electricity load forecasting,"

Comput. Mater. Contin., vol. 88, no. 1, Art. no. 50, 2026, doi: 10.32604/cmc.2026.079734.
[5] S. An, J. Qiu, J. Lin, Z. Yuan, W. Tian, and Z. Yao, "A review of human–AI synergy in smart energy management: Concepts, functions, applications, and future frontiers," Energy Internet, vol. 3, no. 1, pp. 5–22, 2026, doi: 10.1049/ein2.70021.
[6] D. Ramos, P. Faria, and Z. Vale, "Linking short-term electricity demand forecasting and explainable AI: A review for building energy applications," Appl. Energy, vol. 412, Art. no. 127658, 2026, doi: 10.1016/j.apenergy.2026.127658.
[7] M. Shafiuzzaman, M. S. Islam, T. M. R. Bashar, M. Munem, M. Nahiduzzaman, M. Ahsan, and J. Haider, "Enhanced very short-term load forecasting with multi-lag feature engineering and Prophet-XGBoost-CatBoost architecture," Energy, vol. 335, Art. no. 137981, 2025, doi: 10.1016/j.energy.2025.137981.
[8] P. Szeląg, T. Popławski, and M. Adamusiński, "Day-ahead XGBoost forecasting of aggregated residential load: Accuracy and SHAP ranking agreement across experimental configurations," Energies, vol. 19, no. 18, Art. no. 4426, 2026, doi: 10.3390/en19184426.
[9] S. Duan, X. Liu, Y. Lei, J. Yu, Y. Yu, and Z. Yang, "Interpretable machine learning-based intelligent high-frequency static security adjustment method for power system operations," IET Smart Grid, vol. 9, no. 1, Art. no. e70074, 2026, doi: 10.1049/stg2.70074.
[10] B. Zhang, C. Ma, X. Li, K. Yang, H. Shi, H. Hao, and X. Gui, "Interpretable smart meter anomaly detection based on Bayesian-optimized XGBoost and SHAP," Information, vol. 17, no. 8, Art. no. 810, 2026, doi: 10.3390/info17080810.
[11] L. Christodoulou and C. Sun, "The impact of machine learning uncertainty on the robustness of counterfactual explanations," Expert Syst. Appl., vol. 309, Art. no. 131198, 2026, doi: 10.1016/j.eswa.2026.131198.
[12] G. Paphitis, S. Guarino, A. K. Marnerides, and M. Panteli, "Explainability-guided counterfactual cyberattacks through surrogate modeling in power systems," Sustain. Energy Grids Netw., vol. 47, Art. no. 102500, 2026, doi: 10.1016/j.segan.2026.102500.
[13] V. Suresh, A. Swain, B. S. Revathi, and J. M. Guerrero, "Mamba based adaptive conformal inference for probabilistic short-term load forecasting," Knowl.-Based Syst., vol. 328, Art. no. 114222, 2025, doi: 10.1016/j.knosys.2025.114222.
[14] L. Shi, Z. Shen, and T. Lyu, "Local online conformal prediction for power load forecasting intervals," Electr. Power Syst. Res., vol. 261, Art. no. 113412, 2026, doi: 10.1016/j.epsr.2026.113412.
[15] J. N. Mupenzi, D. Witarsyah, A. Kusnadi, and A. S. Sunge, "An explainable data-driven framework for short-term residential energy forecasting," Energy Build., vol. 366, Art. no. 117713, 2026, doi: 10.1016/j.enbuild.2026.117713.
[16] S. A. H. Shah, U. Ahmed, M. Bilal, A. R. Khan, S. Razzaq, I. Aziz, and A. Mahmood, "Improved electric load forecasting using quantile long short-term memory network with dual attention mechanism," Energy Rep., vol. 13, pp. 2343–2353, 2025, doi: 10.1016/j.egyr.2025.01.058.
[17] H. Hong, S. Xu, Y. Yang, J. Hu, Y. Chen, and J. Wu, "Hierarchical reconciliation and per-bus adaptive conformal calibration for distribution-level residential load forecasting: A critical assessment of topology priors," Energy Informatics, 2026, doi: 10.1186/s42162-026-00690-1.
[18] R. Sapkota, K. I. Roumeliotis, and M. Karkee, "AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges," Inf. Fusion, vol. 126, Art. no. 103599, 2026, doi: 10.1016/j.inffus.2025.103599.
[19] M. A. Farahani, M. I. Khan, and T. Wuest, "Hybrid agentic AI and multi-agent systems in smart manufacturing," J. Manuf. Syst., vol. 86, pp. 612–623, 2026, doi: 10.1016/j.jmsy.2026.04.002.
[20] A. Hämäläinen, A. Karhinen, and R. Viitala, "Building trustworthy AI systems for condition monitoring," IEEE Access, vol. 14, pp. 67023–67037, 2026, doi: 10.1109/ACCESS.2026.3688909.
[21] Y. Huang, "Towards trustworthy AI-enabled systems: From model-level reliability to systemic assurance," in *Proc. 34th ACM Int. Conf. Foundations of Software Engineering Companion (FSE Companion)*, 2026, pp. 17–20, doi: 10.1145/3803437.3804867.
[22] S. Raza, R. Sapkota, M. Karkee, and C. Emmanouilidis, "TRiSM for Agentic AI: A review of trust, risk, and security management in LLM-based agentic multi-agent systems," AI Open, vol. 7, pp. 71–95, 2026, doi: 10.1016/j.aiopen.2026.02.006.
[23] S. Tekkesinoglu, M. Wagner, and P. Runeson, "The role of explainability throughout the MLOps lifecycle: Review and research agenda," Front. Comput. Sci., vol. 8, Art. no. 1737008, 2026, doi: 10.3389/fcomp.2026.1737008.
[24] E. Vyhmeister and G. G. Castane, "When industry meets trustworthy AI: A systematic review of AI for Industry 5.0," AI Ethics, vol. 6, Art. no. 200, 2026, doi: 10.1007/s43681-026-01026-1.
[25] M. Kumar, R. Rani, G. Epiphaniou, and C. Maple, "Adaptive trust-aware SOC human–AI teaming for resilient operations," Comput. Secur., vol. 169, Art. no. 104971, 2026, doi: 10.1016/j.cose.2026.104971.
[26] National Energy System Operator, "Historic Demand Data," NESO Data Portal. [Online]. Available: https://www.neso.energy/data-portal/historic-demand-data. Accessed: Oct. 8, 2026.
[27] V. Suresh, B. S. Revathi, and J. M. Guerrero, "A non-parametric adaptive conformal inference based probabilistic hour-ahead solar PV power forecasting method," Sci. Rep., vol. 16, Art. no. 11730, 2026, doi: 10.1038/s41598-026-40911-x.
[28] R. Hu, K. Zhou, and X. Lu, "Integrated loads forecasting with absence of crucial factors," Energy, vol. 322, Art. no. 135632, 2025, doi: 10.1016/j.energy.2025.135632.
[29] World Bank, *Appraisal Environmental and Social Review Summary (ESRS): BEST-PAK MPA Program Phase 1—Grid Stability Enhancement Project (P504536)*, Rep. ESRSA04732, Washington, DC, USA, Jan. 2026. [Online]. Available: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099012726092014169
[30] Asian Development Bank, Second Power Transmission Strengthening Project: Report and Recommendation of the President, Project No. 59002-001, Manila, Philippines, Oct. 2025. [Online]. Available: https://www.adb.org/projects/documents/pak-59002-001-rrp.