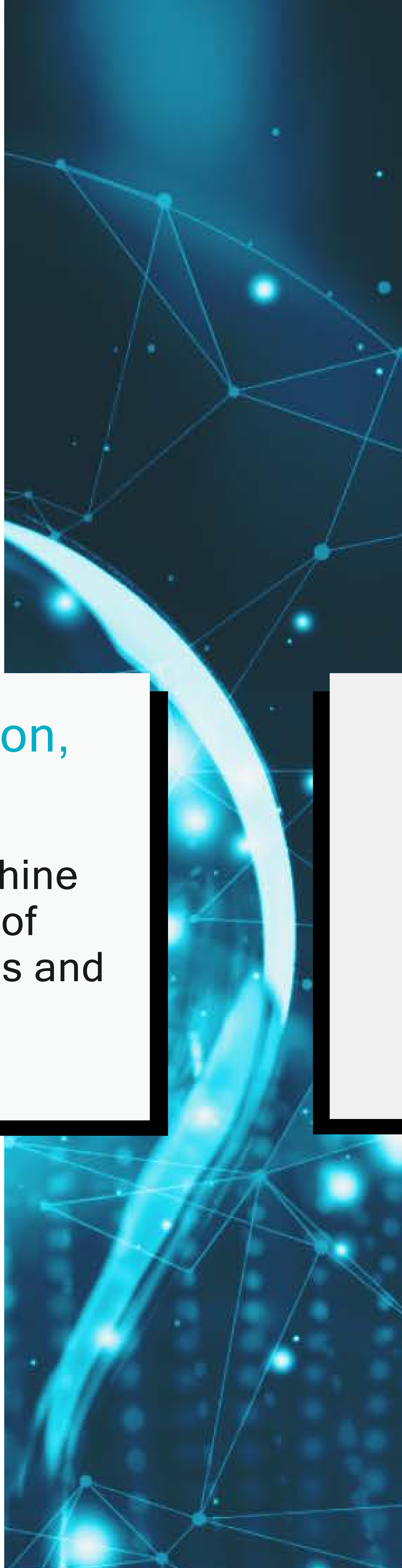
Grid Research, Innovation, and Transformation

Artificial Intelligence and Machine Learning: A Strategic Review of Technology, and Opportunities and Challenges for Adoption



August 2025

Cite: ERCOT. "Artificial Intelligence and Machine Learning: A Strategic Review of Technology, and Opportunities and Challenges for Adoption." ERCOT, August 2025.

Contents

Executive Summary

The rapid transformation of the energy landscape, driven by increasing renewable energy integration and growing demand, has necessitated advanced operational strategies for grid management. The Electric Reliability Council of Texas (ERCOT), managing 90% of the Texas electric load, faces growing challenges as it manages an evolving grid.

The role of Machine Learning (ML) in ERCOT's operational and planning processes can bring about transformative changes in several critical areas, including predictive accuracy, renewable energy management, risk mitigation, resource optimization, decision-making, and regulatory compliance. By integrating ML into its existing framework, ERCOT can move from reactive to more proactive grid management, delivering a more robust and forward-looking system.

One of the core responsibilities of ERCOT is to forecast electricity demand and supply to maintain grid reliability. As the Texas grid becomes more complex, traditional forecasting methods often fall short of accurately predicting the rapidly shifting demand patterns, especially with the unpredictable nature of renewable energy sources like wind and solar. ML offers a significant advantage in this regard by analyzing vast quantities of historical data, weather conditions, consumer behavior, and grid metrics in real-time, thus enabling even more accurate and reliable forecasts.

In real-time grid operations, speed and accuracy are critical. Grid operators must respond quickly to changes in demand, generation, and other factors that affect grid stability. ML-powered decision support systems can assist ERCOT's operators by providing real-time insights and recommendations, allowing them to make more informed and faster decisions during critical moments.

This white paper lists various applications with different ML techniques in Independent System Operator (ISO) or /Regional Transmission Operator (RTO) operations and planning processes, such as those in ERCOT. Artificial Intelligence (AI)/ML techniques have proven their ability to enhance real-time decision-making, improve demand response, and mitigate risks associated with grid reliability and equipment maintenance [1]. This paper outlines some current use cases, the potential for AI/ML integration in ERCOT's operational and planning processes, and the opportunities and challenges in adopting these such innovations.

1. Introduction

ERCOT operates as an ISO, managing the flow of electric power to more than 27 million Texas customers. It handles approximately 1,250 power generation units, with over 142.6 GW of installed capacity, and maintains the delicate balance of energy supply and demand in real-time.

The grid is increasingly dynamic due to the high penetration of renewable energy sources, like wind and solar, which add variability to supply. Accurate forecasting, reliable equipment maintenance, and balancing supply with demand are critical. Traditional approaches rely heavily on historical data and linear models, which fail to capture the complex and nonlinear patterns of modern electricity grids, hence the need for AI/ML techniques to enhance operational efficiency and resilience.

There is widespread recognition across the industry that many organizations are already exploring AI/ML applications, with adoption levels ranging from minimal to full integration [2]. At ERCOT, we see significant potential for AI/ML to address critical operational and planning challenges, such as optimizing energy resource dispatch, improving demand forecasting, and enabling faster decision-making during periods of grid stress. By integrating AI-driven automation into grid management, ERCOT can enhance overall system reliability, minimize manual errors, and improve real-time control. Furthermore, leveraging advancements in computation and AI/ML to adapt to the rapidly evolving ERCOT grid is a key strategic priority outlined in ERCOT’s 2024-2028 strategic plan [3].

2. Definitions

The terms AI, ML, Deep Learning (DL), Generative AI, and Large Language Models (LLM) are all related but represent different levels or subfields within a hierarchy of technology. Here's how they are connected [4, 5, 6, 7]:

- AI is the broadest term, encompassing all techniques that allow machines to mimic or simulate human intelligence. It includes any system or algorithm that can perform tasks that usually require human intelligence, such as reasoning, learning, problem-solving, understanding language, and perception. It can be divided into three types:
  - *Narrow AI*: Designed for specific tasks (e.g., voice assistants, recommendation systems).
  - *General AI*: A theoretical system capable of performing any intellectual task that a human can do.
  - *Super AI:* Another theoretical concept, Super AI would surpass human cognitive abilities, including thinking, reasoning, learning, and even feeling emotions and having desires
- ML is a subset of AI focused on creating systems that can automatically learn and improve from experience without being explicitly programmed. In ML, algorithms are trained on data to make predictions or decisions. There are three main types of ML:
  - *Supervised Learning:* The model is trained on labeled data.
  - *Unsupervised Learning:* The model learns patterns in data without labels.
  - *Reinforcement Learning:* The model learns through trial and error by receiving rewards or penalties.
- DL is a specialized subset of ML that uses neural networks with many layers (hence the term "deep"). These deep neural networks are particularly effective for processing large amounts of data, such as images, audio, and text, and they have significantly advanced AI capabilities in fields like computer vision, speech recognition, and natural language processing. DL models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are often used for tasks that require a high level of abstraction.

- Generative AI is a subfield of both AI and DL that focuses on creating new content, such as images, music, text, or even synthetic data. Generative AI models learn patterns from existing data and use this knowledge to generate new, realistic data or outputs. A key approach in generative AI is Generative Adversarial Networks (GANs), where two neural networks (a generator and a discriminator) work against each other to create increasingly realistic outputs. Transformers (e.g., GPT, BERT) are another class of models commonly used in generative AI, particularly for tasks like text generation and natural language understanding.
- LLMs are a specific type of generative AI that focuses specifically on language understanding and generation. They are trained on vast amounts of text data and use deep learning (often transformers) to generate coherent, human-like text. LLMs are typically built using deep learning architectures, like transformers, which allow them to understand and generate complex natural language. LLMs are generative AI models focused specifically on text. They "generate" language-based outputs, whether it's answering questions, writing articles, or translating languages. They are the most common type of Gen AI model. These models are trained on vast amounts of text to be able to generate natural language after being given prompts. Some examples are ChatGPT, Gemini, Claude Ab, etc.

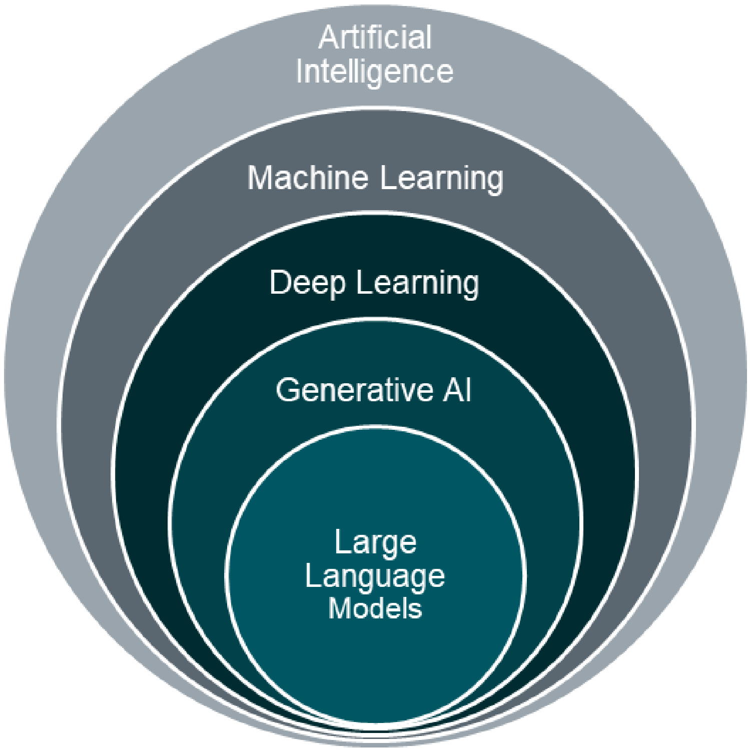


Figure 1. AI, ML, DL, Generative AI, and LLM

In summary, AI is the umbrella term that includes ML, DL, Generative AI, and LLM ("Figure 1. AI, ML, DL, Generative AI, and LLM " above). ML is a method within AI as it represents a way of implementing AI by allowing machines to learn from data. DL is a more specialized form of ML and a more advanced, neural network-based technique within ML that can solve complex tasks like image recognition and natural language processing. Generative AI is a specific application of DL focused on content creation as it applies DL (often through GANs or transformers) to generate new data or content, such as creating images, videos, or text. LLMs are a specific form of generative AI focused on generating human-like text.

3. Commonly Used Machine Learning Techniques

ML is a subset of AI focused on enabling systems to learn from data and improve their performance autonomously. The following are some ML techniques used in the power industry [4, 5, 8]:

1. **Supervised Learning** – In supervised learning, the algorithm learns from a labeled dataset. This means that for every input, there is a corresponding output (or label), and the goal is for the model to learn the mapping from inputs to output. The following are some major supervised learning algorithms:
    - *Linear Regression:* Linear regression is one of the simplest methods to predict continuous values. It works by finding the best-fitting straight line through data points. This line represents the relationship between variables. In the power industry, for example, it helps forecast energy demand based on factors like temperature and time of day.
    - *Logistic Regression:* While called "regression," this is used for classification tasks, where the output is categorical (e.g., yes/no, fail/succeed). It is often used to predict binary outcomes, such as whether a piece of equipment is likely to fail soon. It assigns probabilities to each outcome and then classifies the result into one category or another.
    - *Decision Trees:* A decision tree is a flowchart-like model where each internal node represents a decision based on a feature (such as sensor reading), and each leaf represents the outcome. The tree asks a series of "if-then" questions to decide, making it easy to understand and interpret. For example, in grid monitoring, a decision tree could predict whether a transformer needs maintenance by asking questions based on historical data. Random Forests: A random forest is an ensemble of decision trees. Instead of relying on a single tree, it builds many trees using different parts of the data and combines their results. This makes it more accurate and robust than a single decision tree. In the power industry, it can predict when equipment is likely to fail by

combining predictions from different models, reduing the chance of errors.

2. **Unsupervised Learning** – In unsupervised learning, the algorithm works with unlabeled data, meaning the model doesn't have predefined labels or outcomes. The goal is to find hidden patterns, groupings, or structures in the data. The following are some major unsupervised learning algorithms:
    - *k-Means Clustering*: In k-means clustering, the algorithm groups data points into clusters based on their similarity. It tries to minimize the distance between points in the same cluster and maximize the distance between clusters. This is helpful for segmenting customers based on their energy usage, allowing power companies to create targeted pricing models.
    - *Hierarchical Clustering:* Hierarchical clustering groups data into a hierarchy, creating a tree-like structure where clusters are formed step by step. The algorithm either starts with individual data points and merges them (agglomerative) or starts with all data points in one cluster and splits them (divisive). In the power industry, this technique is useful for grouping similar types of faults in the grid, making fault detection more efficient.
    - *Principal Component Analysis (PCA)*: PCA is a technique used to reduce the number of variables (or dimensions) in a dataset while preserving as much information as possible. It does this by finding new variables (called principal components) that capture most of the variance in the data. This is useful when dealing with large data sets from smart meters or sensors, allowing utilities to focus on the most important features and spot anomalies.
    - *Autoencoders:* Autoencoders are a type of neural network used for compressing data and then reconstructing it. They can capture the essential structure of the data, which makes them excellent for anomaly detection. In the power industry, they can learn the typical behavior of the grid and detect deviations from the norm, such as voltage fluctuations or unusual energy usage, indicating potential issues.
    - *Gaussian Mixture Models (GMM)*: GMM is a probabilistic model that assumes the data is generated from a mixture of several Gaussian distributions. Each distribution represents a different cluster. Unlike k-means, which assigns each data point to one

cluster, GMM assigns probabilities of belonging to multiple clusters. In the power industry, GMM can be used for load profiling by identifying different patterns in energy consumption and modeling uncertainty.
- *DBSCAN (Density-Based Spatial Clustering of Applications with Noise):* DBSCAN is a clustering method that groups data points based on their density. It can identify clusters of varying shapes and sizes and can also detect outliers (noise). In energy usage pattern analysis, DBSCAN can be used to find areas in the grid where energy consumption is significantly different from normal behavior, which might indicate energy theft or inefficiencies.

3. **Reinforcement Learning** – In reinforcement learning, an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties. The agent's goal is to maximize cumulative rewards over time. The following are some of the major reinforcements learning algorithms:
    - *Q-Learning*: Q-learning is a reinforcement learning algorithm where an agent learns the best action to take in each state by receiving rewards (positive or negative feedback) for its actions. Over time, it learns to maximize its total reward. In the power industry, it can be used to optimize energy distribution across the grid by learning how to balance energy supply with varying demand levels, ensuring grid stability.
    - *Deep Q-Networks (DQN):* DQN combines Q-learning with deep neural networks to handle complex environments where the number of possible states and actions is very large. This is useful for managing the integration of renewable energy sources, such as solar or wind, into the grid. The network can learn from past grid data and make real-time decisions to balance traditional and renewable energy sources, ensuring a smooth energy supply.
    - *Policy Gradient Methods:* Instead of learning the value of actions, policy gradient methods directly learn the optimal policy (or strategy) that the agent should follow to maximize its total reward. These methods are useful when the action space is large or continuous. In demand response optimization, for example, the algorithm can learn how to adjust energy pricing dynamically to incentivize users to reduce their energy consumption during peak hours.

- *Actor-Critic Methods:* Actor-critic methods combine the benefits of both value-based and policy-based approaches. The actor decides which action to take, and the critic evaluates how good that action was. Over time, both the actor and the critic improve. In dynamic energy pricing, this method can be used to balance supply and demand in real-time by learning which pricing actions lead to optimal energy usage patterns.

4. Emerging Machine Learning Techniques

As AI/ML capabilities continue to advance, new methods are emerging with the potential to advance decision making, transparency, and explainability. The following subsections highlight two emerging AI/ML developments across a range of applications.

4.1 Large Language Models

The power industry faces challenges such as increased data from sensor integration, renewable energy adoption, and evolving technologies like inverter-Based Resources (IBRs) and Electric Vehicles (EVs). These changes, coupled with rising customer expectations and workforce constraints, demand innovative solutions. LLMs could play a key role by interpreting human prompts, managing complex data, and providing near real-time guidance during crises like extreme weather events. However, understanding their capabilities and limitations is critical for their effective use in power-engineering tasks.

Unlike traditional ML models requiring domain-specific data and labor-intensive setup, foundation models provide versatile, interactive solutions [6]. LLMs have demonstrated strengths in language understanding, prompt engineering, tool embedding, and multi-modal data handling, enabling tasks such as power flow analysis, forecasting, and risk assessments. Enhanced by fine-tuning and prompt engineering, they can process complex queries, integrate with external tools, and interpret non-text data like images. However, LLMs face challenges such as limited domain-specific training data, the absence of safety guardrails, lack of adaptation to physical principles, and cybersecurity risks. Their probabilistic nature and context window limitations restrict their reliability in critical applications. Despite these issues, LLMs offer promise as supportive tools in power systems when combined with governance frameworks and expert oversight [9].

4.2 Explainable Artificial Intelligence

Recent advancements in DL have led to more effective and accurate ML algorithms for power system applications. While these algorithms are claimed to be able to perform better than traditional methods, their complex nature makes it hard to understand how they make decisions and process real-world data. This lack of clarity can make power system experts, who rely on their knowledge and experience, hesitant to trust ML-driven recommendations, especially in situations where reliability is crucial. To address this issue, Explainable Artificial Intelligence (XAI) has been introduced to make ML models easier to understand without reducing their performance or accuracy.

XAI refers to a set of techniques and frameworks designed to make the decisions and inner workings of AI models understandable to humans. Traditional AI models, especially deep learning-based systems, often operate as "black boxes," producing results without offering clear explanations for their predictions [10]. XAI bridges this gap by providing insights into why and how an AI system reaches its conclusions. "Figure 2. High-level Illustration of Explainable AI " on the next page presents a high-level illustration of an explainable AI model. The XAI framework enhances this process by providing an explainable model that works alongside the AI/ML model, ensuring that the results are not only accurate but also understandable and trustworthy.

The key objectives of XAI include:

1. **Transparency:** Making AI models comprehensible by exposing their decision-making processes.
2. **Interpretability:** Ensuring that users can understand how input features contribute to an AI prediction.
3. **Trust and Accountability:** Allowing stakeholders to verify and validate AI-driven outcomes.
4. **Bias Detection:** Identifying and mitigating biases in AI models, ensuring fairness in decision-making.

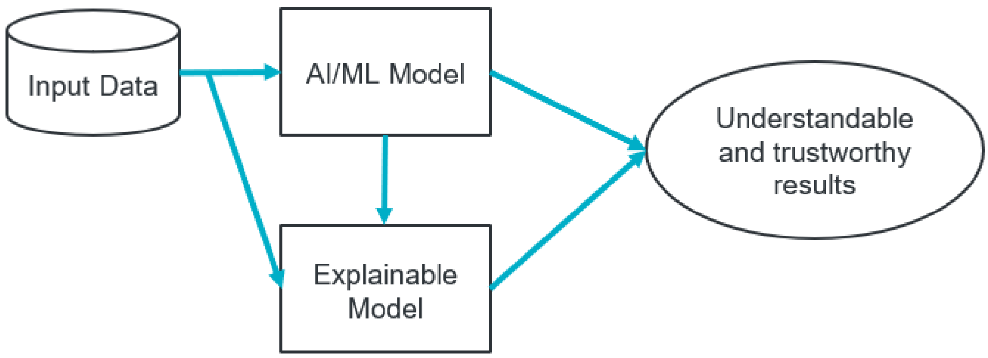


Figure 2. High-level Illustration of Explainable AI

XAI methods can be categorized into the following two main types, each offering different approaches for interpreting complex AI models [10]:

1. **Model-Agnostic Techniques:** These methods can be applied to any AI model without altering its structure. They focus on explaining predictions by analyzing inputs and outputs. Common approaches include:
    - *SHAP (Shapley Additive Explanations):* Inspired by cooperative game theory, SHAP assigns each feature an important value based on its contribution to a prediction. This technique offers both global and local interpretability.
    - *LIME (Local Interpretable Model-Agnostic Explanations*): LIME generates simplified, interpretable models (such as linear regressions) around specific predictions, making it easier to understand why an AI model made a particular decision.
    - *Partial Dependence Plots (PDP) and Accumulated Local Effects (ALE):* These visualization techniques show how individual features impact predictions across the entire dataset, providing insight into model behavior.
    - *Counterfactual Explanations:* These highlight how changes in input variables would alter the output, helping users understand the conditions under which a different decision might occur.
2. **Model-Specific Techniques:** These methods are tailored to specific AI architectures, such as decision trees, neural networks, or support

vector machines. Examples include:

- *Feature Importance Scores:* Used in tree-based models like XGBoost and Random Forests, these scores rank features based on their contribution to the overall prediction.
- *Layer-wise Relevance Propagation (LRP)*: In deep learning models, LRP traces the contribution of each neuron in the network back to the input features, providing detailed explanations for predictions.
- *Saliency Maps:* Commonly used in computer vision models, saliency maps highlight the most influential parts of an input image or dataset, revealing what the model "focused" on when deciding.

5. Possible Use Cases in the Power Industry for Machine Learning

ML offers numerous potential applications in the power industry, particularly within ISO/RTO and transmission operations. These applications can significantly enhance various aspects of grid management and operations. "Table 1: Possible ML Use Cases in the Power Industry " on page 16 illustrates typical use cases for various ML techniques. In practice, multiple techniques have been applied to address the same issues as highlighted in the past ML literature review [4, 5, 8].

Type	Algorithm	Possible Use Case[1]
	Linear Regression	Load forecasting, renewable energy forecasting, energy price prediction, demand response optimization, grid stability and frequency control, spot pricing forecasting, predictive maintenance for power equipment
	Logistic Regression	Fault detection and classification, predictive maintenance for power equipment, outage prediction and classification, renewable energy curtailment classification, voltage stability assessment, contingency classification, distributed energy resource (DER) integration, EV load classification
	Decision Trees	Fault diagnosis and classification, predictive maintenance for power equipment, demand response management, renewable energy generation prediction, power outage prediction, grid congestion management, EV charging load prediction, DER management, contingency analysis
	Random Forests	Load forecasting, renewable energy forecasting, energy price forecasting, equipment failure prediction and maintenance, anomaly detection in grid operations, grid stability and contingency analysis, battery energy storage system (BESS) optimization
Supervised Learning	*Support Vector Machines (SVM)*	Load classification and forecasting, fault detection, power quality disturbance classification, renewable energy integration and forecasting, price forecasting, predictive maintenance of power equipment, weather-driven grid optimization
	Neural Networks	Load forecasting, renewable energy output forecasting, smart grid management and optimization, energy price forecasting, fault detection and diagnosis, predictive maintenance of power equipment, cybersecurity, optimal dispatch of BESS, blackout prediction and prevention,

Type	Algorithm	Possible Use Case[1]
		forecasting EV charging demand, anomaly detection in smart grids, power quality analysis, energy consumption behavior prediction
	k-Nearest Neighbors (k-NN)	Load forecasting, energy theft detection, fault detection in power equipment, power quality monitoring, renewable energy integration, customer segmentation for demand-side management, predictive maintenance of power grid assets, anomaly detection in energy markets, voltage stability prediction, smart grid demand prediction
	Extreme Gradient Boosting (XGBoost)	Load forecasting, renewable energy output forecasting, smart grid management and optimization, energy price forecasting, anomaly detection in grid operations, energy consumption behavior prediction
	k-Means Clustering	Energy consumption behavior analysis, anomaly detection in smart grid operations, fault classification in power equipment, load forecasting for DERs, power quality monitoring, peak load management, load profile classification for grid planning, renewable energy curtailment analysis, seasonal consumption trend identification
	Hierarchical Clustering	Grid reliability analysis, power quality event classification, load profile analysis for grid optimization, renewable energy generation clustering
Unsupervised Learning	*Principal Component Analysis (PCA)*	Load forecasting, anomaly detection in power grids, power quality analysis, renewable energy generation forecasting, predictive maintenance of power equipment, voltage stability monitoring, energy consumption pattern analysis, climate impact analysis on energy consumption, BESS optimization
	Autoencoders	Anomaly detection, energy consumption pattern recognition, power quality monitoring and fault detection, predictive maintenance of grid assets, load forecasting enhancement, renewable energy integration and forecasting, demand response optimization, smart meter data compression, stability monitoring, cybersecurity, EV charging station optimization
	Density-Based Spatial Clustering of Applications with Noise (DBSCAN)	Anomaly detection in meter data, power grid stability and islanding detection, renewable energy production forecasting, EV charging behavior clustering, load profiling and forecasting, outlier detection in power market data
	Q-Learning	BESS optimization, demand response management, grid frequency and voltage control, transmission line congestion management, unit commitment, EV charging optimization, renewable energy curtailment minimization, power market bidding strategy optimization

Type	Algorithm	Possible Use Case[1]
Reinforcement Learning	*Deep Q-Networks (DQN)*	Renewable energy integration, Optimal Power Flow (OPF) with renewable energy integration, EV charging coordination, demand response optimization, unit commitment, power system restoration after blackouts, grid frequency control, preventive maintenance scheduling, renewable energy curtailment minimization
	Policy Gradient Methods	Optimal Power Flow (OPF), real-time control of DERs, BESS optimization, demand response management, grid frequency regulation, EV charging coordination, power system restoration after blackouts, transmission congestion management
	Actor-Critic Methods	Optimal Power Flow (OPF), BESS management, demand response management, fault detection and system resilience, EV charging infrastructure optimization, predictive maintenance of power equipment, real-time grid frequency and voltage control

Table 1: Possible ML Use Cases in the Power Industry

6. Current Machine Learning Uses Within ERCOT

ERCOT has already deployed several ML techniques in the areas of forecasting and market anomaly detection [11, 12, 13]. Below are some details about these implementations.

6.1 SOC and Large Flexible Load (LFL) Forecasting Tool

ERCOT developed two ML-based tools to forecast the system-wide value of battery SOC and LFL in real time. Both tools are integrated into the ERCOT Trend Analysis Tool (TAT) and serve as open-access, web-based platforms within ERCOT to identify and analyze issues that appear in the market based on real-time operational data. The SOC forecasting tool also provides SOC forecasts on each individual battery as well. Both tools will provide forecasts for the short-term (2-hour look-ahead in 5-min resolution), day-ahead (24-hour look-ahead in 1-hour resolution), and week-ahead (168-hour look-ahead in 1-hour resolution) on an hourly update basis. ML models used include Decision Tree (DT), Random Forest (RF) regression, AdaBoost (AB) regression, Bagging (BR) regression, Gradient Boosting (GB) regression, Extreme Gradient Boosting (XGB) regression, and Support Vector Regression (SVR). The Ensemble model calculates average results provided by all the above-mentioned models. Meanwhile, a Deep Learning (DL) model named the Long Short-Term Memory (LSTM) model is also used to provide additional forecast for some extreme weather conditions. See Appendix B for some actual forecasting results.

6.2 Automated Anomaly Detection Tool for Market Price Spike and Separation

The ERCOT market validation team is responsible for anomaly detection across all trading results in the real-time market. Analysis has revealed that most market pricing anomalies were due to erroneous SCED input data from various sources. Currently, ERCOT uses a rule-based model to detect these anomalies, a process that is time-consuming and prone to errors. Some issues could remain undetected for over six months. To address this, ERCOT is developing an automated anomaly detection tool that analyzes SCED input data and real-time prices. This tool leverages both statistical methods (Z-score and Grubb's test) and ML algorithms, including logistic regression, Extreme Gradient Boosting (XGB) regression, Long Short-Term Memory (LSTM), and some transformer-based methods. Early results indicate that the tool can identify and diagnose the root cause of pricing anomalies more efficiently and quickly. More details about this tool can be found in reference item.

7. Potential Uses of Machine Learning at ERCOT

"Table 2:Potential Applications at ERCOT " on page 22 illustrates areas where the increased use of ML could provide benefits to ERCOT [11]. For each item, a brief description is included. The table includes a qualitative assessment of the benefit and level of effort required for implementation. The details and assessments should be considered preliminary, and a more thorough examination will be required before implementation efforts begin.

Application	Sub-topic	Details	Benefits	Possible ML Types
Forecasting for planning and real-time analysis	Price-responsive demand forecasting	ML could be used to analyze customer load telemetry and/or meter data to determine price responsiveness and make predictions on how load will respond to prices in the day-ahead to real-time timeframe.	Risk Mitigation Reliability	Supervised Learning or Reinforcement Learning
	Private Use Network (PUN) forecasting	ML could be used to analyze PUN telemetry and/or meter data to determine behavior and make predictions on how PUNs will behave in the day-ahead to real-time timeframe.	Risk Mitigation Reliability	Supervised Learning or Unsupervised Learning
	BESS behavior prediction	ML could be used to analyze BESS charging and discharging patterns with respect to other variables and make predictions on charge/discharge behavior in the day-ahead to real-time timeframe.	Risk Mitigation Reliability	Supervised Learning or Unsupervised Learning
	DER	ML could be used to	Risk	Supervised

Application	Sub-topic	Details	Benefits	Possible ML Types
		identify DER activity by detecting drops in substation loads when LMP increases.	Mitigation Reliability	Learning or Unsupervised Learning
	Net load forecast risk	ML could be used to analyze weather patterns that present increased risk for net load forecast errors and provide a risk assessment for operators.	Risk Mitigation Reliability	Supervised Learning or Reinforcement Learning
	Forced outage forecasting	ML can learn from past forced outage samples and derive the relationship between forced outage rate and other exogenous factors.	Risk Mitigation Reliability	Supervised Learning or Unsupervised Learning or Reinforcement Learning
	Inertia	ML can provide another layer of intelligence to correct the results generated from the current model-based inertia prediction application.	Risk Mitigation Reliability	Supervised Learning or Reinforcement Learning
	4CP detection and forecasting	ML can be used to detect 4CP activity by monitoring loads at substations for reductions not conforming to the overall system load curve or non-4CP historic behavior.	Risk Mitigation Reliability	Supervised Learning or Reinforcement Learning
	Stressed conditions	The quantity of ancillary services need is calculated on an annual basis. While this provides more advantage to load service entities to hedge against the financial risk, it could underestimate	Risk Mitigation Reliability	Supervised Learning or Reinforcement Learning

Application	Sub-topic	Details	Benefits	Possible ML Types
		the system reliability need at times. ML could forecast the ancillary services need and the impact if sufficient Ancillary Services are unavailable in real time.		
	Dynamic load model parameterization	ML could be used to “learn” the dynamic behavior of loads during disturbances to create a set of dynamic load model parameters for an area.	Risk Mitigation Reliability	Supervised Learning or Reinforcement Learning
Anomaly detection	Telemetry that does not match model specs	ML can be used to characterize telemetry from a source and identify telemetry that does not match those characteristics.	Risk Mitigation Reliability	Supervised Learning
	Oscillation detection	ML could be used to detect the origin of oscillations by “learning” over time how oscillatory signals travel over the multiple network paths and using wave features like amplitude, directionality, phase, etc.	Risk Mitigation Reliability	Supervised Learning
Generation Resource Interconnection or Change Request (GINR) projects forecasting	Predicting which GINR projects will be built.	Accurate prediction of new unit commercialization can provide a clearer picture for long-term planning and seasonal reserve margin planning and, operationally, help anticipate the need for additional ancillary services.	Other (planning reserve margin)	Supervised Learning

Application	Sub-topic	Details	Benefits	Possible ML Types
Real-time optimization	Security-constrained Optimal Power Flow (SCOPF)	Currently, ERCOT SCOPF is a computationally intensive and time-consuming process, which requires us to limit the scope of studies in real time. ML may help to accelerate the process and allow for a broader range of studies.	Risk Mitigation Reliability	Supervised Learning or Unsupervised Learning or Reinforcement Learning

Table 2:Potential Applications at ERCOT

In addition to the potential ML applications listed in Table 2, ERCOT believes that emerging ML techniques, such as LLMs can be instrumental in enhancing the operations and decision-making processes in the following areas:

- **Knowledge Base and Asset Management**
  LLMs can serve as intelligent gateways to enterprise knowledge bases, streamlining access to complex information. Employees can query LLMs for information about assets, operational guidelines, maintenance schedules, or compliance protocols, reducing time spent searching through documentation. Paired with IoT data, LLMs can analyze equipment logs and suggest predictive maintenance schedules or risk assessments.
- **Real-Time Operational Decision Support**
  LLMs assist operators by collecting vast amounts of PMU and SCADA data, identifying potential issues, and suggesting optimal operational adjustments. They can also interpret complex system states and recommend actions during emergencies, such as load shedding or demand response activation. Operators receive real-time, context-aware insights that can improve decision-making speed and accuracy under high-stress conditions.
- **Energy Demand Forecasting and Market Analysis**
  LLMs can analyze historical demand patterns, weather data, and economic indicators to improve load forecasting accuracy. They also

process news, economic reports, and weather predictions to provide real-time insights on how external events could affect demand market prices. Improved forecasting allows ERCOT to balance supply and demand more efficiently and optimize market pricing, helping reduce costs and improve grid stability.

- **Outage Management and Grid Resilience**

  LLMs can assist in outage prediction and management by analyzing grid data, maintenance logs, and historical weather-related outages. During outages, LLMs can process live information to predict restoration times and prioritize repair efforts. Faster, data-driven responses to outages enhance grid resilience and help restore service more quickly.

AI and machine learning have a lot of potential to improve ERCOT's operations and planning, but their adoption has been slow partly due to challenges with interpretability and explainability. ERCOT believes that using new techniques like Explainable AI (XAI) could change how AI-driven decisions are made, explained, and trusted across the energy sector. By making the grid more reliable, helping integrate renewable energy, improving market transparency, and empowering consumers, XAI would support ERCOT's mission of delivering reliable and efficient energy while keeping stakeholders confident. As the grid evolves to include more renewables and decentralized resources, XAI will be key to ensuring that AI-driven innovations are understandable, accountable, and trustworthy. Techniques like iSHAP, LIME, Partial Dependence Plots, Accumulated Local Effects, Counterfactual Explanations, Feature Importance Scores, Layer-wise Relevance Propagation, and Saliency Maps will be crucial in this transformation.

8. Challenges and Barriers to Machine Learning Implementation

Implementation of the increased use of ML applications at ERCOT could face challenges and barriers as described in "Table 3:Challenges and Barriers to Machine Learning Implementation " on the next page below [11].

Challenge	Description
Organization direction, intentionality, OKRs	ERCOT 2025's OKRs emphasize addressing data and data governance, as well as establishing a framework for AI initiatives. We are laying the groundwork to enable AI and data analytics capabilities, empowering data-driven decision-making. However, we face challenges such as ensuring data quality, integrating disparate data sources, and managing data privacy and security. In the upcoming years, we will continue to focus on OKRs related to AI and machine learning to further enhance our capabilities and drive innovation.
Scalability and cost	In some cases, ML initiatives may fit within the existing scope of work and budget of a department. However, other applications could require a significant amount of effort. Some AI/ML solutions can be very costly and resource-intensive, especially in terms of computational power and infrastructure.
Data quality and availability	To train a successful ML model requires a large amount of representative data, relevant to the scenarios trained. ERCOT has data repositories to archive historical operational data, including meter data, telemetry data (SCADA and PMU), and grid and market operation data (market participant submitted data, vendor provided data, ERCOT internal data). Four issues need to be addressed with regards to data access: 1. Segregation of data located on different enterprise systems, 2. Access to data by users in a different department from the department that owns the database, 3. Data governance, and 4. Data availability to meet granularity, latency, and data quality requirements The concept of a data lake may need to be explored to address this challenge. ERCOT applications are generating and accumulating vast amounts of data, but much of it is incomplete, inconsistent, dispersed, and difficult to access. For ML models to work effectively, clean, well-labeled, and high-quality data is crucial. Inconsistent or missing data can lead to poor model performance, inaccurate predictions, or misleading insights. Addressing this issue should be our top priority.

Challenge	Description
Creation and tuning of models	The success of ML needs a good understanding of data itself and how the process is trained. Rather than treating ML as a black-box approach, a significant amount of effort is required to create and tune the ML models, on a continuous basis. This is especially the case for power system applications as the grid is evolving, sometimes at a very fast pace.
Staff training and level of experience	ERCOT has historically emphasized power systems experience in hiring and power systems concepts with on-the-job technical training, which do not necessarily involve knowledge or learning of ML methods. Thus, while ERCOT has access to statistical programming software (both proprietary and open source), not many of its staff are prepared to develop and use these software tools for ML applications.
Limited ability to use cloud services	ERCOT must comply with North American Electric Reliability Corporation (NERC) Critical Infrastructure Plan (CIP) standards. To properly accommodate BES cyber assets and protected cyber assets in cloud computing, existing definitions in NERC CIP standards may need to be revised. Cloud computing is often utilized in other industries for heavy computation, typically for ML processes. While massive amounts of computation can be achieved by purchasing on-site super computers, cost and accessibility are concerns.
ML infrastructure	ML implementation requires consideration of the location of data, network communication bandwidth and speed, processing power requirements, and security of data. Careful planning and investment for all these needs to occur to ensure successful implementation of ML. This is one of ERCOT 2025's OKRs and presents a significant challenge.
Security	Some data used by a new ML application may not have been considered sensitive previously. However, to the extent that the data is being used to make critical operational decisions, the security and redundancy associated with it may need to be revised. This can add cost and complexity to the application.

Table 3:Challenges and Barriers to Machine Learning Implementation

9. Conclusions

This review paper gives an overview of various ML techniques and their applications across the power industry, including use cases relevant to ISO/RTO planning and operations. It highlights selected use cases where ERCOT is currently applying AI/ML techniques. Additionally, the paper outlines some potential use cases for AI/ML integration into ERCOT's operational and planning processes, along with key challenges and opportunities in incorporating these techniques into ERCOT's workflows and decision-making processes.

10. References

[1] Pacific Northwest National Labs, "Artificial Intelligence/Machine Learning Technology in Power System Applications," Technical Report, March 2024.

[2] NERC, "Artificial Intelligence and Machine Learning in Real-Time System Operations White Paper," NERC, White Paper, September 2024.

[3] ERCOT, "STRATEGIC Plan 2024-2028," Electric Reliability Council of Texas, 2023.

[4] A. Alimi, K. Ouahada and A. M. Abu-Mahfouz, "A Review of Machine Learning Approaches to Power System Security and Stability," IEEE Access, June 2020.

[5] A. J. M. Kell, S. McGough and M. Forshaw, "Machine learning applications for electricity market agent-based models: A systematic literature review," June 2022. [Online].

[6] C. Huang, S. Li, R. Liu, H. Wang and Y. Chen, "Large Foundation Models for Power Systems," IEEE Power and Energy Society General Meeting, 2024.

[7] IBM, "Artificial Intelligence Types," IBM, October 2023. [Online].

[8] G. Ruan, H. Zhong, G. Zhang, Y. He, X. Wang and T. Pu, "Review of learning-assisted power system optimization," CSEE Journal of Power and Energy Systems, March 2021.

[9] S. Majumder, L. Dong, F. Doudi, Y. Cai, C. Tian, D. Kalathi, K. Ding, A. A. Thatte, N. Li and L. Xie, "Exploring the Capabilities and Limitations of Large Language Models in the Electric Energy Sector," June 2024. [Online].

[10] R. Machlev, L. Heistrene, K. Levy, J. Belikov, S. Mannor and Y. Levron, "Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities," Energy and AI, August 2022.

[11] J. Billo, P. Du, A. Ahmadi and J. Conto, "ERCOT Machine Learning White Paper," ERCOT, October 2021.

[12] J. Ma, "Forecasting State of Charge (SOC) of Energy Storage Resources (ESR) Based on Machine Learning," ERCOT, 2023.

[13] Q. Zhang, "Regional Price Spike and Separation study towards automated Anomaly Detection - From Data Analysis and Machine learning to End to End Deployment," ERCOT, October 2024.

11. Appendix A

The online SOC forecast and LFL forecast tools fetch historical and real-time data from a database. Data pre-processing examines data quality, including handling inconsistent, missing, and noisy data points, checking for null values and outliers, scaling the data to a given range, and splitting the time-series data into training and testing subset while retaining the temporal order. The cleaned data then is fed into the ML engine to perform training and forecasting. Then, the forecast results are saved in CSV files and a database for visualization and/or post-off-line analysis.

Real-time data and historical data are fed into ML models, each of which will produce a separate forecast result of aggregated SOC and aggregate LFL consumption on system-wide load-zone level and weather-zone level. The ensemble model calculates the average of all the forecast results provided by the ML models as the final forecast results. The ML models are trained at the beginning of every day using the past 240 hours of historical data (equivalent to 10 days length of data) and provide forecast results for the next 24 hours in 1-hour resolution. The forecast is updated every hour and runs 24/7 in an online environment.

The following "Figure 3. BESS SOC Forecast Performance on November 18, 2024 " on the next page and "Figure 4. Large Flexible Load Forecast Performance on November 18, 2024" on the next page show hour-ahead, system-wide SOC and LFL forecasts published on the ERCOT Trend Analysis Tool (TAT) website. "Table 4:Summary of Weekly Performance of System-Wide Forecast " on page 30 shows the performance of both forecast tools in a typical week.

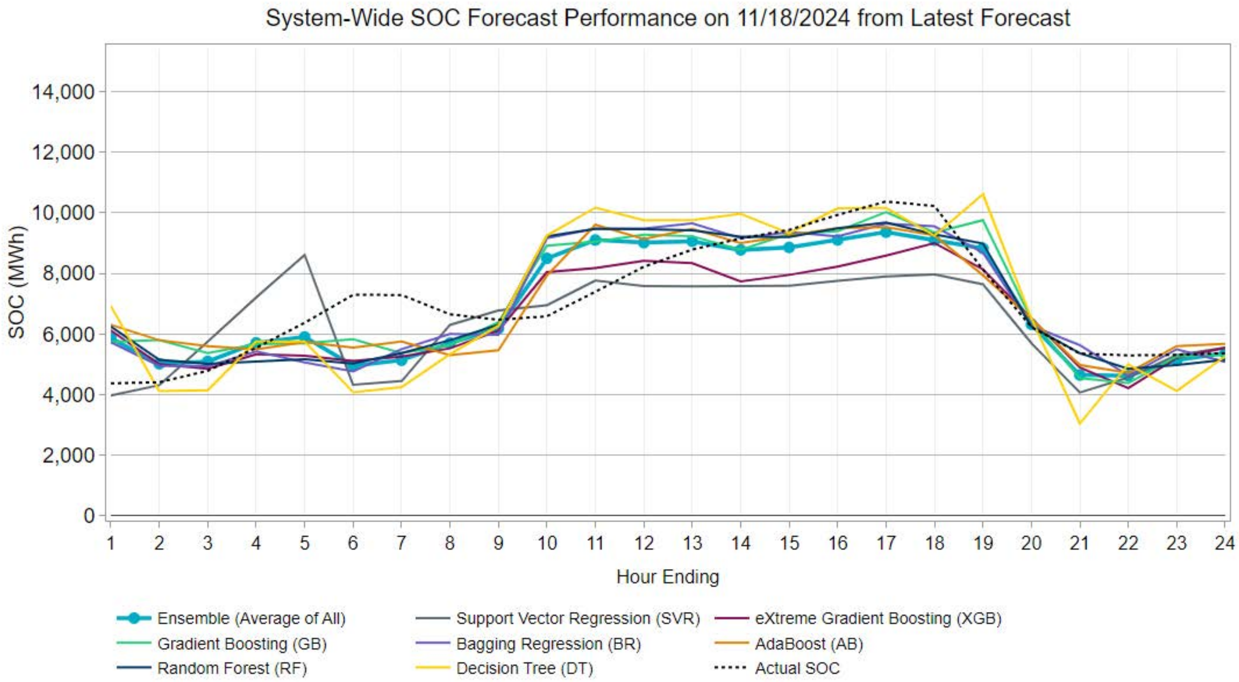


Figure 3. BESS SOC Forecast Performance on November 18, 2024

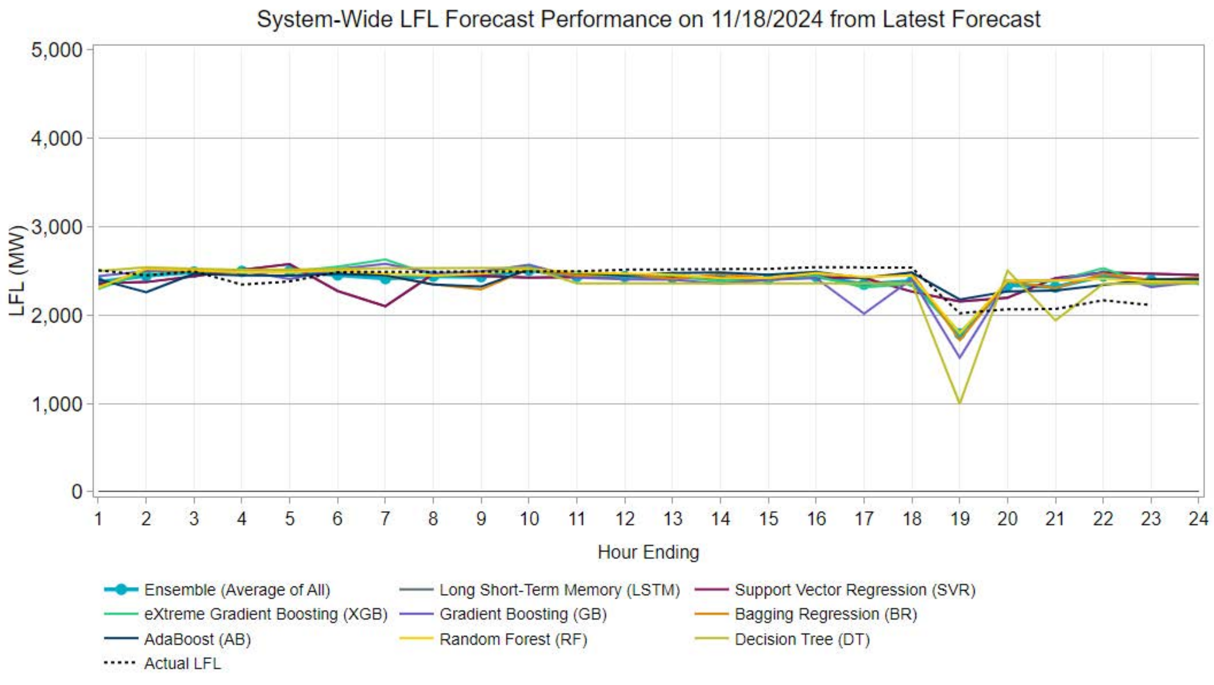


Figure 4. Large Flexible Load Forecast Performance on November 18, 2024

Forecast	Weekly Average		Morning Hour Average (HE07, HE08, & HE09)		Evening Hour Average (HE17, HE18, & HE19)	
	MAPE	MAE	MAPE	MAE	MAPE	MAE
ESR SOC	6.23%	619.34MWh	6.35%	631MWh	4.32%	428.74MWh
LFL	6.62%	185.39MW	4.83%	135.31MW	16.01%	448.51MW

Table 4:Summary of Weekly Performance of System-Wide Forecast

ercot

Taylor
2705 West Lake Drive
Taylor, TX 76574
T 512.248.3000
F 512.248.3095

Austin
8000 Metropolis Dr Building E
Austin, TX 78744
T 512.225.7000
F 512.225.7020

www.ercot.com