

Machine Learning for Sustainable Energy Systems

Priya L. Donti^{1,2} and J. Zico Kolter^{1,3}

¹Computer Science Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA; email: pdonti@cs.cmu.edu

²Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

³Bosch Center for Artificial Intelligence, Pittsburgh, Pennsylvania 15222, USA

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Annu. Rev. Environ. Resour. 2021. 46:719–47

First published as a Review in Advance on August 4, 2021

The *Annual Review of Environment and Resources* is online at environ.annualreviews.org

<https://doi.org/10.1146/annurev-environ-020220-061831>

Copyright © 2021 by Annual Reviews. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See credit lines of images or other third-party material in this article for license information



Keywords

sustainable energy, energy transition, climate change, machine learning, artificial intelligence

Abstract

In recent years, machine learning has proven to be a powerful tool for deriving insights from data. In this review, we describe ways in which machine learning has been leveraged to facilitate the development and operation of sustainable energy systems. We first provide a taxonomy of machine learning paradigms and techniques, along with a discussion of their strengths and limitations. We then provide an overview of existing research using machine learning for sustainable energy production, delivery, and storage. Finally, we identify gaps in this literature, propose future research directions, and discuss important considerations for deployment.

Contents

1. INTRODUCTION	720
2. WHAT IS MACHINE LEARNING?	721
2.1. Notable Paradigms	722
2.2. Common Settings	724
2.3. Common Classes of Techniques	724
2.4. Strengths, Limitations, and Alternatives	725
3. APPLICATIONS OF MACHINE LEARNING FOR SUSTAINABLE ENERGY SYSTEMS	726
3.1. Predicting Electricity Supply and Demand	726
3.2. Optimizing Energy Systems	729
3.3. Maximizing Renewable Power Generation	732
3.4. Reducing Fossil Fuel Impacts	733
3.5. Predictive Maintenance and Fault Detection	733
3.6. Planning Sustainable Energy Infrastructure	734
3.7. Managing Energy Systems Data	734
3.8. Developing Next-Generation Sustainable Energy Technologies	735
3.9. Informing Policy	736
4. DISCUSSION	736
4.1. Research Directions	736
4.2. Considerations for Deployment	738

1. INTRODUCTION

Energy systems are the backbone of modern society and are a major focus of many strategies to promote environmental, economic, and social sustainability. For instance, moving to renewable and low-carbon¹ energy sources will be critical to achieve both climate-change and air-quality targets (2), and energy access is a key pillar of economic development (3). Given the imperative to move quickly on these fronts, it is no surprise that researchers and practitioners have sought to leverage a wide variety of tools from many different areas, including machine learning (ML).

ML refers to a set of techniques that can automatically extract patterns in data—usually, large quantities of data. Due to recent improvements in methods, computing infrastructure, and data availability, these techniques have become pervasive in applications such as targeted advertising, the creation of smartphone voice assistants, and the analysis of medical images to aid diagnosis decisions.

In this review, we consider the ways in which ML has been applied to the development and operation of sustainable energy systems. In particular, following the designation by Bruckner et al. (2),

Machine learning (ML): a subfield of artificial intelligence concerned with computational techniques “that improve[] [their] performance at some task through experience” (4)

¹While we often use the terms renewable and low-carbon interchangeably because of the multiple axes of sustainability we address, they are in fact distinct terms with different implied objectives. In particular, the term renewable energy is often used in service of general sustainability and circular economy goals, whereas low-carbon energy is largely used in service of climate change goals. Although many forms of energy (such as solar and wind power) fall under both categories, these categories are indeed distinct; for instance, biomass-based energy may be renewable but not necessarily low-carbon (1), whereas nuclear energy is low-carbon but not renewable.

we focus on applications in the energy supply sector, covering electric power systems, heating and cooling networks, and fuel supply systems such as natural gas networks. (Due to the great breadth of this topic, we do not cover energy demand sectors such as transportation, buildings, and industry, though these are of course also important components of sustainable energy systems.)

Within this scope, we consider applications that serve the UN Sustainable Development Goals (SDGs) (5) in one or more of the following ways: (a) facilitating the expansion and use of renewable and low-carbon energy sources; (b) reducing the negative environmental and climate impacts of existing fossil fuel and energy infrastructure; and (c) increasing energy availability, reliability, and resilience to aid economic development and climate change adaptation. In particular, these actions directly address the goals of mitigating and adapting to climate change (SDG 13); ensuring human health and well-being (SDG 3) via improvements in air, water, and soil quality; and ensuring universal access to clean, reliable, and affordable energy (SDG 7). Many of these actions also have implications for environmental justice—for example, due to the inequitable distribution of health impacts from pollution attributed to fossil fuels (6)—though strong frameworks will be needed to ensure that the energy systems to which we transition are themselves just (7), particularly in the era of big data (8).

Prior related research includes that by Perera et al. (9), who survey applications of ML for the production and integration of renewable energy, with a focus on forecasting, siting, and plant configuration. Ramchurn et al. (10) describe applications of artificial intelligence (AI) and ML to smart grids, covering aspects including flexible demand, virtual power plants, and self-healing networks. Rolnick et al. (11) survey applications of ML to climate change mitigation and adaptation, with a section focused on electricity systems. In the area of sustainability more broadly, Vinuesa et al. (12) use expert elicitation to understand the role of AI in the SDGs, Gomes et al. (13) describe research in the area of computational sustainability, and Kaack (14) surveys applications of AI in climate change mitigation—all touching on energy systems. We consolidate, update, and augment the perspectives presented in these publications, with a focus on aspects at the intersection of ML and sustainable energy systems.

As far as we are aware, the review closest in scope to ours is that by Mosavi et al. (15). These authors survey research at the intersection of ML and energy systems, classifying it according to the type of ML technique employed (e.g., neural networks versus support vector machines). While the literature reviewed is extensive, we find that this technique-based grouping does not necessarily elucidate broader patterns in the research. For instance, neural networks can be used within supervised, unsupervised, or reinforcement learning settings (see Section 2.1), which represent very different ML paradigms. Instead, we adopt a more conceptual view, summarizing research with respect to its energy systems application area and describing the ML paradigms employed.

The rest of this review is organized as follows: We first present an overview of ML, providing a taxonomy of methods as well as a discussion of their strengths and limitations. We then survey past research employing ML for sustainable energy systems, organized by area of application within the sustainable energy systems literature. We conclude by posing considerations and future directions for both research and deployment.

2. WHAT IS MACHINE LEARNING?

Suppose that you are tasked with the problem of predicting electricity consumption for a given region throughout the day tomorrow. How would you go about writing a computer program that could generate those predictions? One approach might entail manually specifying a set of logical rules that describe how people consume electricity through the day, and then simulating the behavior of a large number of people. For instance, there might be rules describing when

Sustainable energy systems: energy systems consistent with the principles of environmental, economic, and social sustainability, encompassing a wide variety of concepts including environmental conservation, climate action, poverty alleviation, health promotion, and environmental justice

Artificial intelligence (AI): a set of computational techniques that perform complex tasks often associated with human intelligence

Supervised learning: refers to techniques that learn a function from inputs to outputs given a dataset of input/output pairs

someone wakes up, their activities during the day, and their main drivers of energy use. However, accurately enumerating such rules would be a challenging task, even for an energy systems expert.

In reality, a much more common approach is to collect data detailing electricity consumption during past days, along with features that correlate with this consumption (such as temperature or day of the week). One could then write a program that attempts to find correlations between the past consumption data and their corresponding features, and then uses these correlations to predict future consumption given the relevant features (or estimates of them) at future times. This approach illustrates the concept of ML, a form of data-driven programming that automatically learns programs based on examples.

While there are many different types of ML techniques, at their core, most ML algorithms are based on only three components:

1. A model or hypothesis class that specifies the set of functions the ML algorithm can represent. Informally, this can be thought of as the skeleton of the program that the algorithm produces. These models often have free parameters that can be adjusted to specialize to the task at hand.
2. An objective or loss function that specifies the desirable behavior of the model.
3. An optimization or training procedure that specifies how to choose or adjust the parameters of the model in order to improve performance on the objective.

For instance, in our forecasting example, the model might be a low-degree polynomial of temperature and day of the week, where the free parameters are the coefficients of the polynomial. The objective might be to minimize the absolute error of our predictions of future electricity consumption.² The training procedure might involve making small incremental adjustments to the parameters to iteratively improve the objective (e.g., via gradient descent, a common procedure in many ML algorithms).

Before diving further into the details, we first clarify the relationship of ML to other relevant fields. ML is a subfield of AI, which describes a set of techniques concerned with making computers perform complex tasks traditionally associated with human intelligence (such as perception, speech, movement, and logical reasoning). ML also has a deep relationship to statistics, with a significant overlap in both history and techniques. The difference between these two fields is largely one of perspective (16), as ML is generally more concerned with performance on the task at hand (i.e., optimizing the objective), whereas statistics is generally more concerned with discovering some notion of truth in the underlying data (i.e., understanding the quality of the learned model parameters). ML also has close ties to optimization (given its reliance on optimization procedures) and control theory (see Section 2.1.1).

2.1. Notable Paradigms

While we provide a relatively general definition of ML above, we now describe several notable ML paradigms pertaining to different settings in which ML can be used.

2.1.1. Supervised, unsupervised, and reinforcement learning. The above-described load forecasting setting, in which we provided input/output pairs to an ML algorithm, is an example of a specific paradigm called supervised learning. In this setting, the goal is for the ML algorithm to

²Note that performance on the training data used to construct the model is related to, but different from, performance on data the model has not yet seen. In particular, it is important to avoid overfitting, which refers to the phenomenon where a data-driven model makes good predictions on training data but does not generalize well to unseen data.

learn a function mapping from inputs (features) to their desired outputs (labels) given some supervision on what these input/output pairs should look like. (This process is referred to as regression when the outputs are continuous, and classification when the outputs are discrete.) Supervised learning has found much success in areas such as image classification, automated speech recognition, and machine translation. Unfortunately, this paradigm is not applicable in all settings; in many cases, it is prohibitively expensive to get enough labeled data to use supervised learning, or the system of interest involves a decision-making process that cannot be sufficiently described by single input/output pairs.

The framework of unsupervised learning, in contrast, requires only that we provide inputs to the ML algorithm, without any corresponding outputs. As there is no output to produce, the algorithm merely attempts to find some form of structure over the inputs. For instance, clustering techniques aim to group data into similar categories (clusters). Another major paradigm is dimensionality reduction, which aims to find a low-dimensional subspace that captures most of the variation in the data (similarly to techniques such as principal component analysis). While unsupervised methods are useful for analyzing and/or partitioning data, a notable caveat is that key attributes (such as the number of clusters or dimensions) are typically picked by the algorithm designer; as a result, the learned outputs may be an artifact of the algorithm itself rather than representing true attributes of the underlying data.

A third major paradigm is reinforcement learning (RL), a setting where an agent must learn how to act in a sequential environment to maximize some reward (17). Unlike the paradigms of supervised and unsupervised learning, RL algorithms do not operate over a fixed dataset but rather within a setting where the algorithm can take an action that affects future states of some system. This setting is similar to that considered in adaptive control, and indeed these fields have a great degree of shared history, though they often differ in the types of structural assumptions they make about the underlying system (18). RL is also closely related to the area of agent-based modeling (ABM), though agent-based models often involve manually specifying behavioral rules, whereas RL aims to learn such rules automatically. While RL has had some notable successes, such as beating humans in complex games like Go (19), there have been comparatively few deployments of RL on real-world physical systems. This stems from the fact that RL agents must often act suboptimally (potentially for a long time) during the learning process; most successful applications thus require, at the very least, a (realistic) simulation environment on which to train the agent.

2.1.2. Online learning. The paradigms described above (with the possible exception of RL) typically occur in the offline (or batch) setting, where the ML algorithm is provided with a complete dataset up front over which to learn. In contrast, in the online (or streaming) setting, data points arrive one at a time, and the algorithm must make a prediction before receiving the next data point. Since online learning models update their parameters as data are processed, they often require different evaluation metrics than offline algorithms do.

2.1.3. Transfer learning. The above discussion has implicitly assumed that ML models are trained in settings similar to those in which they ultimately operate, and indeed, ML models often have trouble generalizing to settings they have not yet encountered. The paradigm of transfer learning (20), as well as the related areas of multi-task and meta-learning, therefore focuses on the ability of ML models to adapt to new tasks. In particular, modern ML techniques are data hungry, meaning that in practice these models require a prohibitively large quantity of data for training in order to reach a suitable level of performance. Transfer learning techniques aim to address low-data settings by transferring pretrained ML models (built potentially with a great deal of data) to new tasks where fewer data are available, for instance, by fine-tuning the models on a small quantity of data from the new task.

Unsupervised learning: refers to techniques that aim to find some structure over a dataset of inputs

Reinforcement learning (RL): refers to techniques that aim to optimize some agent via interaction with a sequential environment

2.2. Common Settings

There are several notable settings in which ML algorithms have found success and which are of particular relevance to the sustainable energy systems literature. While these represent a relatively small slice of ongoing applied ML, they also represent some of the more visible application areas, and are at least partly responsible for some of its recent prominence.

One such area is computer vision, which is concerned with deriving insights from images, videos, or other visual data. This area has been largely reshaped in recent years by advances in ML, and in turn, computer vision challenges such as image classification have become some of the most standard benchmarks in ML. In particular, methods based on deep learning and convolutional neural networks have outperformed humans on the seminal ImageNet classification challenge (21), achieved state-of-the-art performance in object detection and semantic segmentation, and been successfully applied in settings such as remote sensing (22) and autonomous driving.

Another notable setting is natural language processing (NLP), which is concerned with the analysis of (written and spoken) human language. Alongside tools from linguistics, ML models have been used extensively for NLP tasks. For instance, topic modeling is a commonly used unsupervised learning technique that aims to discover thematic clusters of words or phrases (i.e., topics) within text documents. Recent years have also seen advances in machine translation and automated question/answer systems, as well as realistic text generation by language models (23), driven by advances in deep learning.

Finally, ML has been widely applied to time-series analysis problems, which are concerned with uncovering patterns in temporal data. These include tasks such as automated speech recognition (24) and temporal forecasting (e.g., load forecasting). While many ML techniques are applicable, there also exist particular approaches (such as recurrent neural networks) that are specialized for time-series problems.

2.3. Common Classes of Techniques

We now briefly describe some of the major algorithmic approaches used within ML.

2.3.1. Deep learning. Deep learning (25, 26), currently one of the more prominent approaches to ML, broadly refers to a class of models based on the composition of linear and nonlinear functions (layers). These layers are optimized using gradient-based methods based on what are called backpropagation techniques. This paradigm of composable layers trained using backpropagation has proven extremely powerful, capable of expressing very complex functions while also generalizing well in practice when presented with new data. Several specialized forms of deep learning models, or architectures, have also emerged; these include generic feedforward networks (for unstructured data), convolutional networks (for image data), recurrent networks (for time-series data), and graph networks (for graph-structured data). Deep learning methods have been widely applied within many different ML paradigms, including supervised learning, unsupervised learning, and RL.

2.3.2. Decision trees. Decision trees are a class of models that recursively partition input data along individual features of importance. That is, decision tree methods analyze patterns in input data to produce trees with decision nodes; for a given data point, these decision nodes can be followed on the basis of the values of the features in order to obtain a prediction. For instance, in our load forecasting example, one decision node might encode whether a day is a weekday and a subsequent node might encode whether the temperature is above 80°F, where two yeses might imply a prediction that peak power consumption will be high. These methods can use multiple different loss functions and are typically optimized using a greedy strategy (as exact and gradient-based

methods are intractable for this model class). Although they were among the very first ML approaches, decision trees have seen renewed interest in recent years owing to (a) their success when integrated with ensemble methods (see Section 2.3.4) and (b) the extent to which their predictions are often viewed as more interpretable than those of alternative algorithms (though the ultimate scope and value of this interpretability are very much a point of contention in the literature).

2.3.3. Support vector machines. Support vector machines (SVMs) are another type of classical ML model based on a linear model class (i.e., the prediction is a linear function of the input) or a nonlinear extension known as a kernel hypothesis class. These methods typically use a type of loss function called a (regularized) hinge loss, which gives them the geometric property of being a so-called max-margin classifier: If we view inputs to the model as points in n -dimensional space, then an SVM finds a (linear, or kernel-based) separator that lies directly between the two classes. The resultant optimization problem can be efficiently solved globally, with different methods being used in practice to solve the problem depending on the particular problem instance.

2.3.4. Ensemble learning. Although many of the methods described above can perform well alone, in settings where one wants to obtain the absolute highest level of performance for a given problem (and where computational constraints are of no concern), then a common approach is to use a group, or ensemble, of ML algorithms. The results of these algorithms can then be combined using some form of (often weighted) averaging. Common paradigms include bagging, defined as training similar ML models and then averaging their results in some deterministic way; stacking, which refers to training heterogeneous ML models and then combining their results using some form of meta-model; and boosting, which learns and then combines a series of similar ML models, but in a way such that the training of subsequent models depends on the results of previous models.

2.3.5. Bayesian models and Gaussian processes. Broadly speaking, Bayesian learning in the ML setting refers to methods that attempt to model uncertainty in the data and parameters of the model. These methods often employ so-called priors over the parameters and data, which encode beliefs about the nature of the data and parameters prior to observing any data. Bayesian methods are prevalent in many areas of ML—in fact, it is possible to derive Bayesian versions of most of the methods presented above—but a canonical approach in this class involves using Gaussian processes (GPs) (27). GPs are nonlinear regression models that capture the similarity between two points using a covariance or kernel function (and can be interpreted as putting a Bayesian prior over the functions that are fit to observations). A primary advantage of GPs is that they not only model predictions but also give a measure of uncertainty based on the quantity of similar data seen by the model so far. This property has made GPs particularly practical in the context of Bayesian optimization, a global search procedure that uses GPs to quantify possible upper and lower bounds for some function that is to be optimized. Because Bayesian optimization is a data-driven approach, no information about the “true” underlying function being optimized is required in practice, making these techniques amenable to many practical scenarios where one wants to optimize a function that is not known to the practitioner.

2.4. Strengths, Limitations, and Alternatives

As evidenced by the discussion above, ML is a powerful paradigm for data-driven programming and can facilitate the analysis of large and heterogeneous data streams in cases where they would be impossible to analyze manually. While conceptually simple, this paradigm can manifest in many forms. For instance, ML can be used to scale human intuition by identifying patterns in comparatively small quantities of labeled data, and then applying these learned patterns at a much

Variable renewable energy sources:

renewable energy sources whose output changes over time based on external factors, such as solar irradiation or wind speed

larger scale. It can also be used to glean actionable insights from unstructured data streams, such as satellite imagery or text documents, and to optimize complex systems based on observations of the systems' behavior, among many other applications.

At the same time, ML has several major limitations. For instance, ML algorithms are extremely dependent on the quality of the data they receive ("garbage in, garbage out"). More broadly, ML is fundamentally an amplifier of the systems in which it is deployed, meaning that while it is capable of amplifying the benefits of these systems, it is also equally capable of exacerbating biases (28), inequities (29), and market failures (30) through its data, design, and applications. ML methods also generally assume that the data on which they are trained and tested are similar in distribution to one another, and they have difficulty dealing with scenarios where this is not the case (known as distribution shift). In addition, ML tends to have difficulty enforcing any physics or hard constraints associated with the domains in which it operates, and many methods also suffer from a lack of interpretability. Finally, like most statistical methods, ML tends to find correlations in data as opposed to discovering causal relationships. Many of these topics represent active areas of ML research (see Section 4).

We also note that while ML is broadly powerful, complex or cutting-edge ML techniques may not always be best suited or needed for every problem. For instance, linear regression may be a better alternative to more complex supervised ML techniques in cases where only small quantities of data are available, or where the structure of the relationships between the inputs is well known. Techniques from classical control theory may be more appropriate than RL when the dynamics of the underlying environment are simple or well structured; ABM techniques may be more appropriate when the rules governing agent behavior are well known and do not need to be learned.

In general, we encourage researchers and practitioners not to view ML as a black box or a silver bullet but rather to employ it in a principled manner that is guided by an understanding of its strengths, limitations, and underlying assumptions, as well as of relevant technical and contextual considerations surrounding the problem at hand. For instance, what kinds of data are actually available, and what are their volume and quality? How and by whom will this application be deployed, and what are the associated logistical or regulatory needs? (For instance, is it important that solutions be physically realistic for use in engineering workflows, or that they be interpretable by decision makers who may use or audit them?) Are complex techniques required, or will simple techniques suffice? Such questions can help both select the most appropriate techniques and ensure that the tools that are ultimately developed are well tailored to the needs of the problem at hand.

3. APPLICATIONS OF MACHINE LEARNING FOR SUSTAINABLE ENERGY SYSTEMS

We now present an overview of the literature using ML for sustainable energy systems, organized by energy systems application area. **Figure 1** maps these application areas to the main ML paradigms discussed in the previous section.

3.1. Predicting Electricity Supply and Demand

Given the stochastic nature of both energy supply and demand, a large body of prior research has attempted to provide better forecasts of these quantities, both to enable the management of electric grids with large shares of variable renewable energy sources and to guide energy system planning. In particular, as described by Hong & Fan (31), real-time estimates and short-term forecasts (on the scale of minutes to weeks) enable better power system optimization and demand response, and medium- to long-term forecasts (on the scale of months to years) can inform planning and energy

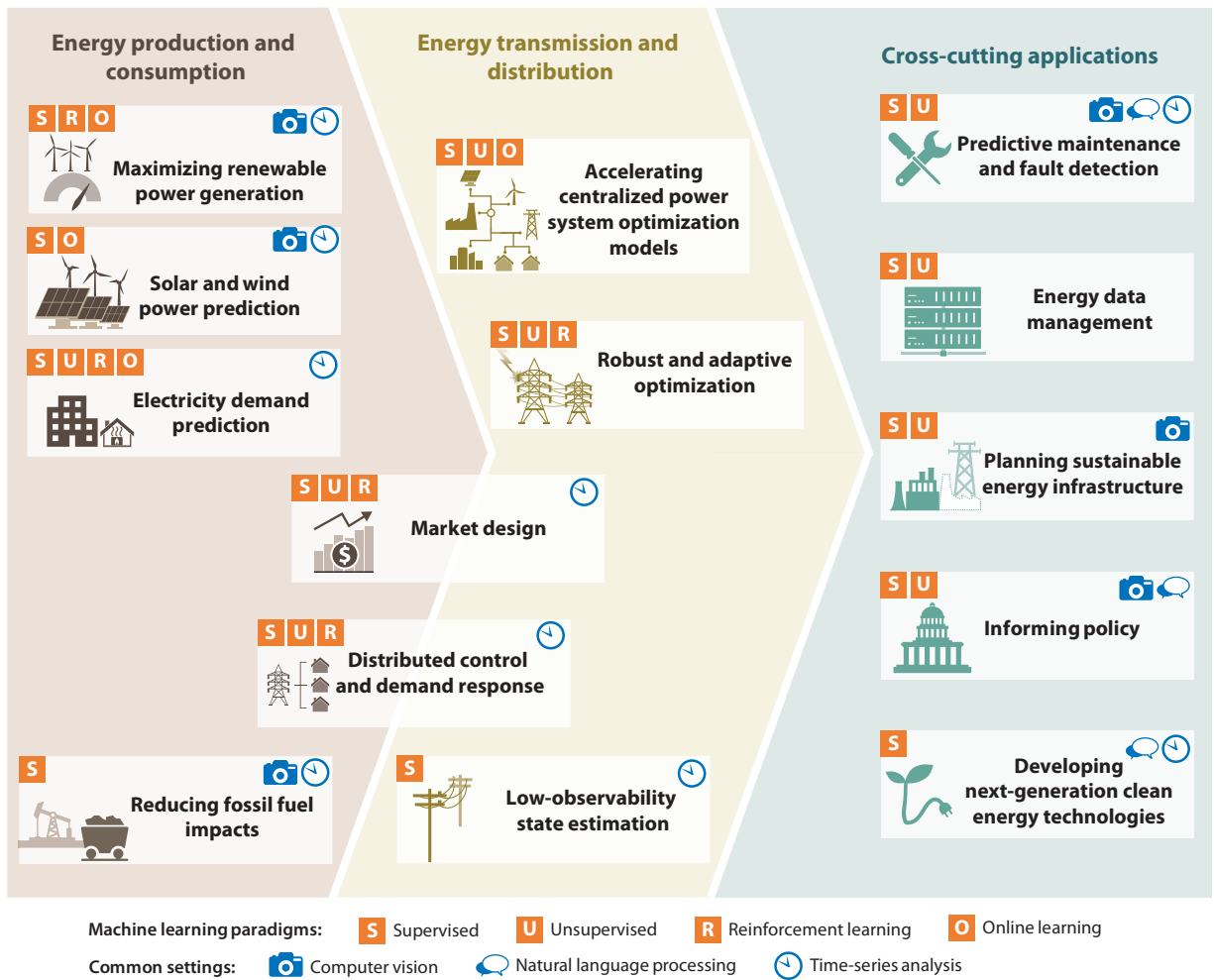


Figure 1

An overview of sustainable energy systems applications where machine learning has been applied, alongside common machine learning paradigms used within each setting.

policy. Likewise, estimates at different spatial scales—for example, at the distribution transformer level versus at the level of individual generators or buildings—can inform optimization and planning decisions by different sets of entities. Due to its strengths in time-series analysis, ML has been used extensively to construct such estimates on both the supply and demand sides.

3.1.1. Demand estimation. A plethora of papers have been written on electrical load estimation, in turn prompting many reviews. For instance, Kuster et al. (32) present a taxonomy of the prior load forecasting literature with respect to spatial scale, temporal resolution, and type of method used, finding that ML models are most prevalent in short-term forecasting applications, whereas regression is more prevalent for longer-term forecasts. Hong & Fan (31) critically review prior research in both point load forecasting and probabilistic load forecasting, spanning methods from both statistics and ML; they argue that probabilistic forecasts in particular will be necessary to manage modern, renewable grids (for additional reviews, see 33 and 34).

One trend in this literature involves obtaining granular real-time load measurements in cases where they may not be immediately available—for instance, due to a shortage of grid communication infrastructure—but where they may be useful for demand response. For instance, Ledva et al. (35) disaggregate real-time power signals at a distribution feeder into residential air conditioning and other loads, using an online learning algorithm that employs external weather data as well as load models learned from historical data.

Another trend involves the construction of short-term load forecasts using supervised learning alongside additional structural information. For instance, Kell et al. (36) use k -means clustering on smart meter data to group customers with similar attributes, and then construct separate load forecasts for each cluster via supervised learning. Bogomolov et al. (37) analyze telecommunications data via Fourier analysis to characterize human behavioral dynamics and apply the resultant frequency-domain features within a decision tree regression to forecast average and peak daily energy consumption for the next week. Wang et al. (38) forecast load in multienergy systems, using a long short-term memory (LSTM) model, alongside information about correlations among electricity, heating, and cooling loads, to make coupled forecasts of all three. While this research incorporates structural information about the features informing load forecasts, other research has sought to incorporate structural information about the context within which these forecasts are deployed. For instance, Donti et al. (39) consider the context of power system dispatch, embedding a differentiable economic dispatch model within a neural network to produce load forecasts that are tuned not for accuracy but for the quality of dispatch decisions made on their basis.

A third trend involves constructing load estimates in situations where labeled load data are limited, in particular by using ML to transfer insights from contexts with data to contexts without. For instance, Mocanu et al. (40) forecast load for buildings without historical load data, first using deep RL to construct building load profiles for buildings with historical data and then using transfer learning to extrapolate these insights. Yuan et al. (41) address the context where hourly smart meter data are available for some residential consumers but only monthly billing statements are available for others. These authors use a deep learning approach to learn how to downsample from monthly to hourly load for metered customers, and then apply these insights to unmetered customers via a Bayesian deep learning approach.

3.1.2. Solar and wind power estimation. Due to the variable nature of solar and wind power generation, ML has been used extensively to construct short-term forecasts of these quantities, as described in other reviews (42–45). Mirroring the above discussion of trends in demand estimation, some trends in solar and wind power estimation include disaggregating power generation to obtain real-time estimates, as well as employing innovative features or structural information for short-term forecasts. For instance, Kara et al. (46) disaggregate solar panel loads from real-time distribution feeder-level power measurements, using a learning method incorporating contextual information about the solar load signals. Sun et al. (47) train a supervised convolutional neural network to forecast near-real-time solar panel power production, using video recordings of the sky. Wang et al. (48), Hao & Tian (49), and Du et al. (50) forecast wind power production at granularities of 10–15 min, using signal decomposition techniques in order to deal with the fluctuations in and nonstationarity of wind time-series data.

Given the close relationship between weather and solar/wind power output, a related body of research has aimed to improve weather and climate forecasting models for use as input to power production forecasts. Voyant et al. (51) review different ML methods used in solar irradiance forecasting. McGovern et al. (52) survey applications of ML to the prediction of storms and other high-impact weather events. Hwang et al. (53) predict weather variables such as temperature and precipitation at a 3–6-week subseasonal scale, using an ensemble of clustering and linear

regression models. Rolnick et al. (11) discuss ways in which ML can help accelerate climate models to enable them to provide more granular forecasts; we conjecture that this, in turn, may help assess long-term renewable energy potential (or energy demand) in order to inform generation planning processes.

Building on these two bodies of research, recent studies have looked at ways to better integrate weather predictions into power production forecasting models. For example, Haupt et al. (54) construct several physics-based numerical weather prediction (NWP) models to estimate solar irradiance from 0 to 72 h ahead (using ML at various points to detect clouds in images of the sky, correct NWP outputs, and intelligently ensemble different models). These multitimescale NWP outputs are then provided as inputs to supervised ML models, in order to enable them to produce probabilistic, multitimescale power forecasts. Employing similar techniques, Kosovic et al. (55) use NWP outputs as features within supervised wind power forecasting models. Mathe et al. (56) construct a solar forecasting model that explicitly incorporates spatiotemporal information about NWP data, employing a long-term recurrent convolutional network for this purpose. While these studies take an important first step, we opine that it may be fruitful for future research to more deeply integrate weather and power prediction—for example, by incorporating (reduced-form) NWP physics directly into ML-based power forecasting models via hybrid physical modeling techniques (57).

While supervised power forecasting techniques assume reasonably complete knowledge of historical power production, this assumption does not always hold for distributed energy resources. In particular, the sizes and locations of distributed solar energy systems are not always available to power system operators. As a result, several projects have attempted to automatically map attributes of solar power systems by applying ML to satellite and aerial imagery (for an overview, see 58). These estimates in turn could be used in the loop of engineering-based models for power output prediction.

3.2. Optimizing Energy Systems

The optimization of electric power grids poses several fundamental challenges. In particular, the amount of power injected into the grid must equal the amount of power consumed at every moment, and the resultant power flows must satisfy physical constraints dictated by the grid topology. Unfortunately, optimization problems that enforce these constraints, such as unit commitment and optimal power flow (OPF), are large and slow to solve. This challenge is exacerbated in power grids with large amounts of variable renewables whose output varies on the basis of weather and environmental conditions, as power system optimization must be performed more frequently and under greater uncertainty to accommodate these renewables. It will also become increasingly important to ensure that power and energy grids are robust to unexpected events, both because of this variability and in consideration of climate change adaptation-related needs (59). As a result, a large body of research has attempted to accelerate, distribute, robustify, or otherwise improve power and energy system optimization. While the vast majority of the approaches we describe here focus exclusively on electric power systems, there has also been recent interest in addressing the co-optimization of multienergy systems (60), with proposed directions for ML (61).

3.2.1. Accelerating centralized power system optimization models. One set of approaches to accelerating power system optimization procedures has involved using ML to approximate these procedures. In particular, in the context of OPF, these approaches have involved generating a dataset by running multiple instances of OPF, training a supervised ML model on the corresponding input/output pairs, and then using this model to generate (approximate) OPF solutions. As described in a review by Hasan et al. (62), early naïve approaches in this vein could

Multienergy systems: energy systems integrating multiple energy grids (e.g., power, gas, and heating grids)

not guarantee the feasibility or optimality of their solutions, limiting their viability in practice. As a result, recent approaches have attempted to incorporate pertinent structure from OPF into deep learning-based approximators, in order to increase their chances of success. For instance, Fioretto et al. (63) propose approximation methods combining deep learning and Lagrangian duality, incorporating information about OPF dual variables into the neural network loss function to incentivize feasible solutions. However, such methods merely encourage—rather than enforce—feasibility, potentially limiting their utility. As such, recent research has aimed to more directly enforce OPF constraints within approximators. For instance, Zamzam & Baker (64) use a neural network to predict a partial set of OPF outputs, and then solve for the remaining outputs explicitly using power flow equations representing the physics of the grid. Donti et al. (65) build on this approach, directly incorporating the power flow equations into the neural network itself.

A parallel set of approaches has attempted to use ML within power system optimization models in order to speed them up, rather than attempting to replace these models entirely (for a more in-depth discussion, see 62). For instance, Misra et al. (66) propose a streaming algorithm to learn which OPF constraints are active at any given time, thereby reducing the complexity of solving the OPF problem. Baker (67) proposes to use a supervised learning approach to learn warm-start points for OPF optimizers in order to help these solvers converge more quickly. Similarly, Dong et al. (68) use a physics-integrated supervised learning approach to learn warm-start points, specifically predicting primal and dual variables that were identified as most relevant to their solver's performance. Xavier et al. (69) propose a framework to speed up security-constrained unit commitment models, using a combination of clustering and supervised learning approaches to identify redundant constraints, learn warm-start points, and identify subspaces to which the solver can likely be restricted.

As of now, it is not yet clear (at least to us) which of these classes of approaches is more likely to succeed—or whether they will be surpassed by approaches from, for instance, traditional optimization (70) or circuit simulation (71). In addition, many ML approaches in this area assume a static grid topology rather than accounting for the fact that the grid's topology changes frequently as a result of, for instance, outages. This challenge will need to be surmounted for these approaches to achieve real-world applicability.

3.2.2. Distributed control and demand response. While the above-described research focuses on improving centralized optimization methods, another body of work has looked to improve decentralized power grid optimization techniques in the context of both large-scale electrical grids and microgrids. These techniques aim to implement intelligent control strategies for distributed devices such as solar inverters, batteries, and load devices in order to help them balance supply and demand or provide ancillary services such as frequency and voltage regulation. In this vein, Antonopoulos et al. (72) review academic and commercial uses of AI and ML for demand response in applications such as controlling and scheduling devices, or selecting optimal sets of customers to respond to particular grid events. Lopez-Garcia et al. (73) survey applications of neural networks to microgrid control, including the control of energy storage and distributed generation devices for energy balancing, frequency regulation, and voltage regulation.

One main approach to distributed control has involved designing controllers that mimic or otherwise account for centralized grid optimization procedures. For instance, Dobbe et al. (74) employ a supervised learning approach that trains distributed grid controllers to mimic the actions they would have taken under OPF, but using only locally available grid measurements (given limitations in power grid communication infrastructure). Karagiannopoulos et al. (75) employ a similar approach in the setting of chance-constrained OPF. In a slightly different vein, Hassan et al. (76) propose a hierarchical demand response framework for multienergy systems,

which couples the decentralized control actions taken by groups of devices (based on Markov decision processes) with the centralized control actions taken by a central utility (based on chance-constrained OPF). While these methods address the important problem of coordinating between different distributed controllers, a potential challenge may arise in incentivizing controllers to actually follow their desired protocols—as opposed to, for instance, deviating for game theoretic reasons—depending on how they are operated and regulated.

Another set of approaches has employed ML techniques to optimize distributed devices based on market signals or power prices, as opposed to knowledge of a centralized optimization scheme. For instance, several reviews (77–79) describe applications of (deep) RL to the control of distributed devices, within both microgrids and large-scale power grids. In their review, Vázquez-Canteli & Nagy (80) focus on applications of RL to demand response, describing methods to control different classes of devices such as electric vehicles, heating and cooling systems, and smart appliances. These authors identify the need for multiagent RL methods that coordinate between different demand response agents, a theme that harkens back to earlier demand response techniques based on agent-based models (81). Finally, to highlight an example in the multienergy systems setting, Sheikhi et al. (82) propose a demand response method to reduce peak load in both electricity and natural gas networks, using RL to model customer preferences, residential appliance characteristics, and energy prices. While we highlight RL approaches here, we note that the decision regarding when to use RL versus traditional control theoretic approaches is not always clear-cut (for a more in-depth discussion, see 18). Acknowledging this trade-off, two reviews in particular (77, 79) identify the need for RL methods to be more closely integrated with traditional control methods and domain-specific knowledge.

While this set of approaches relies on electricity price signals (83), most electricity pricing schemes do not currently reflect criteria that are directly relevant to mitigating climate change and/or air pollution—such as the emissions intensity of the grid at a given time. In the interim, forecasts of power grid emissions intensity (in terms of both greenhouse gases and criteria air pollutants) may be useful in guiding sustainability-oriented control schemes. While emissions intensity estimates can be derived using power system optimization models, as discussed above, these models may be overly expensive to run, prompting the use of ML-based methods. For instance, Bruce & Ruff (84) forecast the average hourly regional carbon dioxide (CO₂) intensity of the UK power grid up to 4 days ahead, using a supervised deep learning approach. Leerbeck et al. (85) forecast hourly marginal CO₂ emissions intensities (which reflect the emissions intensities associated with marginal changes in demand) up to 1 day ahead, using ML to inform the forecasting model via trend extraction and feature selection. We note, however, that many of these methods suffer from a lack of ground truth on emissions factors, instead using proxies to this ground truth to train supervised ML models. As a result, it is often difficult to evaluate the quality of these emissions forecasting models.

3.2.3. Market design. Balancing power grids and multienergy systems with large shares of (distributed) renewables will require innovative market designs to guide distributed controllers. While traditional economic and game theoretic approaches will be important in designing such markets, ML approaches may also be able to help. For instance, a review by Zhang et al. (77) briefly discusses applications of RL to market design, including the setting of dynamic prices and the finding of equilibria in energy trading markets. In one example of research in this area, Du & Li (86) propose to use RL to set prices in a multi-microgrid setting in order to decrease the peak-to-average power ratio and maximize profits. Antonopoulos et al. (72) review applications of ML to designing markets and power prices for demand response, including applications for learning customer preferences and for incentivizing participation in demand response programs.

3.2.4. Robust and adaptive optimization. Several ML techniques have been proposed to help power grids operate in a robust and adaptive manner. Echoing themes discussed above, these include (a) methods to accelerate or decentralize traditional robust power system optimization techniques and (b) methods to replace these traditional optimization techniques altogether. Along the first theme, as discussed above, Xavier et al. (69) propose ML methods to speed up the problem of security-constrained unit commitment, and Karagiannopoulos et al. (75) propose a method to distribute the problem of chance-constrained OPF. Relatedly, King et al. (87) approximate stability constraints within transient stability-constrained OPF using a neural network, and Halilbašić et al. (88) approximate constraints with security-constrained OPF using decision trees. Along the second theme, Glavic (78) reviews several RL applications for preventive, emergency, and restorative control of power grids. As an additional example, Ni & Paul (89) formulate secure power system optimization as an attacker-defender game, and attempt to find a Nash equilibrium of this game using RL. Participants in the recent “Learning to Run a Power Network” challenge series proposed RL-based grid topology switching approaches for robust power system control; recent competition winners have used a variety of RL techniques, such as actor-critic or dueling deep Q-network approaches potentially augmented with domain knowledge (90, 91). A problem with most RL techniques, however, is that they are not accompanied by provable guarantees, prompting a need (as discussed above) to integrate RL with areas such as robust control that enforce such guarantees. For instance, recent research has explored incorporating Lyapunov stability guarantees from robust control into RL algorithms for applications such as microgrid management (92).

3.2.5. Low-observability state estimation. As distributed renewable energy resources have become more prevalent, it has become increasingly necessary to monitor and control the voltages of the microgrids and distribution systems with which these resources interact. However, distribution systems in particular have historically not been equipped with many sensors; that is, they are often under low-observability conditions where traditional approaches cannot estimate system voltages (93). As a result, several ML approaches have been proposed to address this problem. For instance, Pertl et al. (94), Jiang & Zhang (95), and Mokhtar et al. (96) use supervised neural network approaches to estimate or forecast distribution system state under low-observability scenarios. Ostrometzky et al. (97) present a physics-informed supervised deep learning method for state estimation, embedding information about the power flow equations governing distribution system behavior into the loss function of a neural network-based state estimator. A caveat with ML-based approaches in this context is that they often require a large and consistently formatted set of historical data, which may or may not be available depending on the system at hand; in data-sparse scenarios, non-ML techniques such as matrix completion (93) may be more appropriate.

3.3. Maximizing Renewable Power Generation

The problem of maximum power point tracking (MPPT) aims to optimize the configurations of (renewable) power generators in order to increase their productivity. Reisi et al. (98) and Abdullah et al. (99) review MPPT algorithms for solar and wind energy systems, respectively, including a discussion of some ML techniques. Reviews by Glavic (78) and Lopez-Garcia et al. (73) contain discussions of several studies using RL and neural networks, respectively, for MPPT. To highlight some examples, Abdelrahman et al. (100) propose an algorithm to adjust the DC voltage applied to solar photovoltaic panels by modeling the power-voltage relationship as a Gaussian process, and then finding the power-maximizing voltage via Bayesian optimization. Abel et al. (101) orient movable solar panels via RL in order to maximize their utilization of direct, reflective, and diffuse solar radiation. Rao et al. (102) reconfigure the topologies of solar panels to account for shading, using a supervised neural network approach to map irradiance values to potential topologies.

Wei et al. (103) use RL to control the shaft rotating speed of a wind turbine to maximize power output without requiring wind speed measurements.

3.4. Reducing Fossil Fuel Impacts

ML has been applied in a variety of contexts to monitor the environmental and climate impacts of fossil fuel extraction activities, which can then inform actions to reduce these impacts. For example, Keramitsoglou et al. (104) train a fuzzy logic-based classifier to automatically detect oil spills from satellite imagery. There has been a particularly great deal of interest in using ML to detect methane leaks from natural gas extraction and transportation, given the extreme potency of methane as a greenhouse gas. For instance, Wan et al. (105) detect methane leaks from natural gas pipelines by applying an SVM model to sensor data. Wang et al. (106) train a convolutional neural network to detect methane leaks from infrared videos at a test facility simulating a natural gas production site. Zukhrufany (107) examines the use of supervised learning techniques to forecast corrosion in natural gas and petroleum pipelines, in order to prevent leaks and explosions before they occur. A challenge with these techniques is that they are all supervised, requiring robust datasets of historical leaks and faults even though these datasets may be unavailable or incomplete. As a result, for example, Wang et al. (106) collect their own (artificial) dataset by inducing controlled methane releases at a test facility. An alternative approach might be to employ unsupervised or physics-integrated methods that use existing knowledge on pipeline physics or the fluid dynamics of methane plumes.

ML may also be used to help reduce the emissions impacts of fossil fuel production. For instance, ML has been used to reduce emissions from freight transportation (11), which may in turn have emissions benefits for the transportation of solid fuels. ML has also been used to improve the energy efficiency of oil and gas production, as described by Narciso & Martins (108). However, we caution practitioners that for any applications improving efficiency in the oil and gas sector, extreme care should be taken to ensure that this research augments—rather than impedes—sustainable energy transition pathways (109).

In scenarios where fossil fuel power plants are being run with carbon capture and sequestration, ML can also help monitor sequestration sites to prevent CO₂ leakage. For instance, Chen et al. (110) use ML to construct computationally inexpensive proxies for physical models describing CO₂ injection and fluid flow processes, for the purposes of monitoring leakage from wells. Similarly, Mo et al. (111) use a convolutional neural network to characterize subsurface pressure and CO₂ saturation fields within a CO₂ storage simulation model. We note, however, that since such methods are simulation based, there is uncertainty as to how they will translate to real-world contexts.

3.5. Predictive Maintenance and Fault Detection

In many cases, quickly detecting and repairing faults in energy systems can help improve the longevity of equipment, reduce waste, and increase system robustness. ML has been used both to detect such faults in real time and to forecast faults before they occur. For instance, as described in Section 3.4, ML has been used to detect and forecast faults in natural gas pipelines. In the context of solar power, Rao et al. (102) use a combination of supervised ML and graph signal processing techniques to detect faults in solar arrays from device measurements. Iyengar et al. (112) use a graphical models approach to examine correlations between the power outputs of nearby residential solar panels, in order to flag potential anomalies. In the context of wind power, Orozco et al. (113) identify wind turbine failures in historical data by building supervised models to predict gearbox component temperatures, and then analyzing the residuals of these models to identify

anomalous temperatures. In the context of nuclear power plants, Calivá et al. (114) propose a method based on supervised learning, clustering, and denoising methods to detect anomalies in simulated nuclear reactor data. Chen & Jahanshahi (115) propose to automatically detect cracks in nuclear power plant infrastructure by applying a convolutional neural network–based technique to video data. In the context of power grids more broadly, Rudin et al. (116) propose multiple applications of ML to proactively suggest power grid maintenance, and Nguyen et al. (117) propose a framework for automatic inspection of power lines that uses ML to analyze images collected by unmanned aerial vehicles.

3.6. Planning Sustainable Energy Infrastructure

There are many cases in which additional information on existing energy infrastructure, future renewable generation capacity, or consumer demand may be useful in improving planning processes. In some cases where this information is not readily available, ML has been used to extract relevant estimates from raw sources of data (such as satellite imagery). For instance, as described in Section 3.1.2, ML can be used to improve climate forecasts as an input to renewable energy siting, as well as to map distributed energy resources (such as solar panels) using satellite or aerial imagery. Similarly, ML has been used to identify buildings (118–120) or estimate building energy consumption (121) in satellite or aerial imagery for the purposes of planning district heating and energy systems (and for planning building efficiency improvements). ML has also been used to map electricity transmission and distribution infrastructure for the purposes of energy access and infrastructure planning (122). That said, other research has found that, especially at lower spatial resolutions, directly detecting distribution lines requires too much manual tagging (123); this work instead combines ground truth on higher-voltage lines with graph algorithms that estimate the locations of low-voltage grids. Other research has used ML to cluster customer data in order to inform rural electrification planning processes (124).

Unfortunately, many of the optimization-based models used for energy planning are slow, often preventing planners from assessing the full range of scenarios they would ideally examine when making planning decisions. As such, ML and related techniques have been used to speed up planning processes. For instance, Mellit et al. (125) use a neural network to predict optimal sizing parameters for solar photovoltaic plants using information on location and average irradiation. Wu et al. (126) propose an efficient model to assess the placement of hydropower dams in the Amazon basin subject to multiple (potentially competing) energy and ecological objectives, using tree-based dynamic programming. Moutis et al. (127) propose a decision tree–based method to size energy storage systems for microgrids. While we have not encountered literature using ML within the loop of existing planning models, in principle, techniques similar to those discussed in Section 3.2.1 may also be directly applicable; in particular, many planning problems fall within the realm of combinatorial optimization, to which ML has been extensively applied (128).

3.7. Managing Energy Systems Data

When conducting data-driven analysis in energy systems applications, it is often necessary to clean and preprocess the data used as input to the analysis. In some cases, ML can facilitate the data cleaning process. For instance, when merging different data records with potentially inconsistent data labeling schemes (particularly at scale), it may be necessary to automatically infer which records should be merged (129). As a result, projects such as the Public Utility Data Liberation Project (130) have used clustering and other ML techniques to automatically match records across different energy systems datasets. Similarly, Rolnick et al. (11) describe applications of ML to data matching and data fusion in smart cities, which can be useful for integrating data originating from

heterogeneous sources. In many energy systems analytics pipelines, ML may also be employed to preprocess data for use within subsequent ML algorithms, for instance, by doing feature engineering or data augmentation (61). However, as automated data processing techniques will never be perfect, it may remain important for human experts to be involved at least to some extent—especially to mitigate against critical errors or biases that could propagate along subsequent data analytics pipelines.

In smart grid applications or other contexts employing large-scale sensing, it may also become necessary to ensure that these data are collected and transmitted in a sustainable manner (9). These kinds of considerations have started to become prevalent particularly in smart city contexts. For instance, Valerio et al. (131) explore a distributed ML approach in the context of smart cities to extract insights from distributed data in situ while transmitting only limited quantities of data. Muhammad et al. (132) provide a framework for producing condensed versions of collected data to reduce the impacts of data transmission in smart cities, using ML and other methods.

3.8. Developing Next-Generation Sustainable Energy Technologies

While many of the generation and storage technologies necessary for sustainable energy systems are readily available, additional innovation can play a role in reducing the costs of existing technologies or in facilitating the development of new technologies to address outstanding challenges. ML has been used in various ways to inform and accelerate science and engineering workflows for the research and development of new technologies.

Materials science can play a key role in the design of next-generation technologies such as photovoltaics, batteries, and electrofuels (133). As described in other reviews (134, 135), ML has been used to guide materials synthesis experiments and to characterize the properties and performance of proposed materials. For instance, Fujimura et al. (136) use supervised ML to predict the ionic conductivities of proposed lithium ion-conducting solids for batteries in order to guide experiments. Similarly, Raccuglia et al. (137) propose a method to guide the synthesis of materials such as organohalide perovskites used for solar energy conversion, training a supervised ML model on the results of previous experiments to predict the results of proposed future experiments. Zhang et al. (138) review additional applications of ML to the design of perovskite materials. Bai et al. (139) and Gomes et al. (140) propose a method combining physics- and AI-based reasoning techniques to scalably characterize crystal structures in proposed solar light absorbers. Zitnick et al. (141) propose the use of ML to create efficient, scalable simulations of potential electrocatalysts for power-to-gas applications, and present the Open Catalyst Dataset to spur research in this area.

Relatedly, ML has been used to accelerate process optimization and deployment workflows for sustainable energy technologies. For instance, Ren et al. (142) use a physics-integrated Bayesian ML method to predict the performance of gallium arsenide solar cells under different growth temperatures, in order to recommend growth temperatures that maximize cell performance. Attia et al. (143) aim to select the parameters of fast-charging protocols that maximize electric vehicle battery lifetimes, using a regression model to predict a battery's remaining useful life from a small number of experimental data and using Bayesian optimization to intelligently guide experimental design.

ML has also been used to facilitate the development of nuclear fusion technologies. Humphreys et al. (144) describe previous research and future directions for the use of AI and ML in this area, including optimizing the planning of experiments, analyzing experimental results, generating data-driven models of fusion systems, detecting plasma disruptions, and contributing to reactor control. For instance, Baltz et al. (145) develop a human-in-the-loop statistical method to guide the setting of experimental parameters for a magnetic confinement fusion reactor. Kates-Harbeck et al. (146) develop a deep learning-based model to forecast plasma disruptions within a tokamak reactor, and show that their model generalizes across multiple reactors.

3.9. Informing Policy

The transition to sustainable energy systems will fundamentally need to be supported by strong policies, regulations, and market frameworks. Many of these decisions often require making normative trade-offs between different (potentially competing) objectives, and must be made under uncertainty about both present circumstances and future outcomes. In certain cases, ML can help provide useful input to these decision-making processes.

For instance, in cases where relevant data are unavailable, ML can help provide estimates of these data, which in turn can be used to inform decision-making and advocacy processes. As described in Sections 3.1.2 and 3.6, ML has been used to map energy infrastructure such as solar panels or power lines from satellite and aerial imagery. Notably, previous research on solar panel mapping used the collected estimates to analyze environmental and socioeconomic factors driving solar deployment (147). Similarly, there has been initial research aiming to monitor real-time greenhouse gas emissions from different entities around the world, including those in the energy sector, in order to guide climate and energy policies. For instance, as described in Section 3.4, ML has been used to detect methane emissions from aerial imagery. There have also been efforts to monitor CO₂ and methane emissions from satellite imagery (148–150), with some efforts beginning to use ML to enable more accurate and targeted monitoring (151). We emphasize that while such ML techniques can provide useful proxies for on-the-ground data, they should not be viewed as a replacement for on-the-ground data where it is possible to obtain them.

ML can also help analyze large bodies of papers, legal documents, or other texts in order to guide science policy or other areas of policy making. For instance, Venugopalan & Rai (152) use topic modeling techniques from NLP to analyze the kinds of technological innovations presented within a dataset of solar photovoltaic patents. Callaghan et al. (153) use topic modeling to analyze climate change research (including that relevant to energy systems) in order to understand potential gaps.

Finally, there are various ways in which ML can help improve tools from economics, the social sciences, and policy analysis to facilitate energy policy. For instance, as discussed in Section 3.2.3, ML can provide input to the process of electricity market design. Additionally, Rolnick et al. (11) describe applications of ML to policy in the context of climate and energy, with applications including incorporating data-driven insights into agent-based models, accelerating policy simulation models, and evaluating the performance of previous policies.

4. DISCUSSION

As evidenced by the discussion above, ML has found wide applicability across a diverse range of sustainable energy systems applications. These include applications in time-series forecasting, optimization, control, data collection, and accelerated science across areas such as energy systems management, planning, innovation, and policy. While we believe that ML has great potential to contribute to advances in these areas, we emphasize that it will be important to develop and deploy ML techniques in a principled manner in order for them to be properly impactful. In particular, we caution that the excitement around ML's capabilities has also come with a great deal of hype, leading ML to at times be applied haphazardly. With these considerations in mind, we propose several directions for research in ML and sustainable energy systems, as well as important considerations for the impactful deployment of research in this area.

4.1. Research Directions

Guided by our assessment of the existing research, we propose several major directions for future work.

4.1.1. Hybrid physical models and robust techniques. Many energy systems problems require that algorithms satisfy hard and fast physical constraints. For instance, failing to satisfy the power flow equations in power system optimization and control contexts may increase operational costs or even lead to blackouts. Unfortunately, most modern ML techniques are not able to satisfy arbitrary constraints, prompting a need for research in hybrid physical models that integrate aspects of traditional physics and engineering models with aspects of ML. Per the discussion in Section 3.2.1 (see also 57), research in this area has taken a few different forms. One direction involves using ML in the loop of existing physics and engineering models, primarily in order to simplify portions of these models. Another direction involves directly embedding constraints from physics and engineering models into ML models, for instance, by creating differentiable layers in neural networks that enforce physical constraints (e.g., 154–156).

Relatedly, given the increasing need for robustness in energy systems, it may become increasingly important to ensure that ML models themselves are robust. One relevant direction in the ML literature has involved the construction of neural networks that are provably robust to certain classes of perturbations (157, 158). Another potentially fruitful but as-yet-underexplored direction may be in the design of RL algorithms accompanied by provable robustness guarantees, particularly in cases where these algorithms are used for energy systems control. For instance, Donti et al. (92) explore methods to enforce Lyapunov-based guarantees from robust control theory within RL algorithms. Buşoniu et al. (18) further discuss potential ways to bridge the gap between RL and control methods.

4.1.2. Interpretable and uncertainty-aware techniques. ML models in the energy sector seldom operate in a vacuum. In particular, they often interact with human users, traditional optimization models, or other automated processes, and may need to be auditable by energy systems regulators in certain contexts. As a result, it is often paramount to ensure that models are interpretable (159, 160), uncertainty-aware (161), or otherwise well integrated within the broader workflows in which they are used. As the term interpretability can mean different things in different contexts, it will likely be necessary to define domain-specific criteria as to when interpretability is needed, what purpose it serves, and what forms of interpretability are acceptable to serve that purpose (159, 160). In particular, as argued by Rudin (162), there are many cases where developing fundamentally interpretable models, rather than fitting retroactive explanations to black-box models, may be most appropriate. In terms of uncertainty-aware ML, Ghahramani (161) discusses research directions in areas such as probabilistic programming and Bayesian optimization. We note that while uncertainty-aware methods may be immediately useful in human-facing contexts, integrating these methods into automated workflows may require additional steps; for instance, many generation scheduling programs used in power systems control rooms do not currently accept probabilistic forecasts. More broadly, we encourage researchers and practitioners to explore additional ways to better integrate ML models into the decision-making contexts in which they are used. For instance, in prior research, Donti et al. (39) explore a method to more closely align the goals of load forecasting models with the goals of the power system optimization procedures that rely on their output (see also 163).

4.1.3. Handling uneven data availability and distribution shift. There are many circumstances in which ML algorithms must operate in circumstances different from those under which they are trained. For instance, as society increasingly experiences the effects of climate change, this will cause shifts in energy supply and demand patterns (59, 164), which means that forecasting models trained on historical data may not generalize well to future scenarios. As another example, while sustainable energy systems innovations will need to be implemented everywhere, the data

available on which to conduct analyses do not uniformly represent global demographics (165), and models may not properly generalize to contexts that are not well represented in the data (see Section 4.2.1 for further discussion). While concerted societal action will be needed to address these challenges, from a methodological perspective, further research in areas such as transfer learning and domain adaptation (20) or physics-integrated modeling (Section 4.1.1) in the context of energy systems may help mitigate some of these distribution shift challenges.

4.2. Considerations for Deployment

Finally, we identify several bottlenecks to the impactful deployment of research in ML and sustainable energy systems, and propose potential directions to address these bottlenecks.

4.2.1. Data availability and access. One bottleneck to the development and deployment of ML methods is lack of access to real-world data. On one hand, real-world data are often proprietary, distributed, incomplete, error ridden, not clearly licensed, or otherwise in a form that is not amenable to ML workflows. On the other hand, synthetic data are often not representative of real use cases. As a result, the development of centralized data repositories, open data initiatives, and accompanying standards (keeping data privacy and security considerations in mind) could help accelerate the progress of ML applications in energy systems contexts.

In addressing the data availability challenge, researchers and practitioners must keep considerations of equity at the forefront. For instance, as discussed by Kaack (14), many organizations do not have the capacity to collect, release, and maintain data, which may exacerbate biases and disparities in available datasets and, therefore, in the insights derived from them. We argue further that this can contribute to widening the (digital) divide between places with and without relevant data (e.g., developed versus developing contexts), as research developments are often (overly) tailored to those contexts where data are available. Addressing these challenges will require strong policies to build data literacy and organizational capacity among both small and large stakeholders around the globe so as to ensure equitable access both to the data economy and to any algorithmic advances developed on its basis.

4.2.2. Translating from research to deployment. While the application areas we discuss are at varying levels of maturity, in practice there are a number of challenges associated with integrating research techniques into real-world deployment workflows (166). For instance, power system optimization procedures are often governed by legacy systems and strict regulatory requirements, often making it difficult to develop or introduce new innovations to enhance these procedures. One means of mitigating these challenges may include the development of realistic benchmarks, test beds, or demonstration projects—for instance, in the context of research challenges. Recent examples of such challenges include the GO Competition run by ARPA-E in the USA (see <https://gocompetition.energy.gov/>) and the Learning to Run a Power Network Challenge run in partnership with RTE France (90). More concerted interdisciplinary platforms for collaboration between researchers and industry practitioners may also help bridge communication and capacity gaps as well as facilitate the scoping and development of projects in a manner amenable to deployment. Regulatory reforms may also be required to accommodate the testing and deployment of new techniques within the energy industry.

4.2.3. Aligning policy and incentives. While many of the applications we describe in this review have the potential to contribute to sustainability goals, they may not automatically do so. For instance, as argued by Victor (30), while ML has the power to make energy markets more efficient, this efficiency can in turn amplify market failures; as a result, strong policy signals (such

as carbon pricing) are needed to ensure that energy markets indeed align with climate and sustainability goals. Relatedly, while ML has the potential to enable and accelerate many sustainability applications, it can also be used in ways that directly oppose sustainability goals (167), and any data infrastructure built to enable these applications may have implications for emissions as well. Fundamentally, ML is a tool that serves to amplify the systems and platforms within which it is deployed—which means that aligning energy systems policy with sustainability goals will be paramount to the success of strategies using ML for sustainable energy.

SUMMARY POINTS

1. Machine learning (ML) is a powerful tool for scaling human insights, optimizing complex systems, and discovering patterns in large datasets.
2. In the context of energy systems, ML has been used in a wide variety of applications. These include supply and demand prediction; system optimization, planning, and maintenance; data management; accelerated science and engineering; and policy analysis.
3. We encourage researchers and practitioners to properly consider the trade-offs between different techniques (ML or others) when scoping energy systems applications, rather than viewing ML as a silver bullet.

FUTURE ISSUES

1. We propose several methodological directions for ML that we believe will be critical for properly integrating ML techniques into energy systems workflows. These include research in hybrid physical modeling and robust ML to address the physical requirements of energy systems; interpretable and uncertainty-aware methods to better integrate ML into deployment workflows; and transfer learning and domain adaptation to address issues of uneven data availability and distribution shift.
2. The energy industry will need to address several bottlenecks to facilitate the deployment of ML models, including resolving issues of data availability and the digital divide, and providing test beds or collaboration platforms to bridge the gap between research and deployment.
3. Strong policy measures will be required to ensure that energy systems incentives are indeed well aligned with sustainability goals, as ML is fundamentally an amplifier of the systems within which it operates.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The writing of this review was supported by a US Department of Energy Computational Science Graduate Fellowship (DE-FG02-97ER25308), the Center for Climate and Energy Decision Making through a cooperative agreement between the National Science Foundation and Carnegie

Mellon University (SES-00949710), the Computational Sustainability Network, and the Bosch Center for Artificial Intelligence. We thank Kyle Bradbury (Duke University), Lauren Kuntz (Gaiascope, Inc.), and Aidan O’Sullivan (University College London) for their thoughtful feedback on the manuscript.

LITERATURE CITED

1. Creutzig F. 2016. Economic and ecological views on climate change mitigation with bioenergy and negative emissions. *GCB Bioenergy* 8(1):4–10
2. Bruckner T, Bashmakov I, Mulugetta Y, Chum H, de la Vega Navarro A, et al. 2014. Energy systems. In *Climate Change 2014: Mitigation of Climate Change. IPCC Working Group III Contribution to the Fifth Assessment Report*, ed. O Edenhofer, R Pichs-Madruga, Y Sokona, E Farahani, S Kadner, et al., pp. 511–97. Cambridge, UK: Cambridge Univ. Press
3. Carley S, Lawrence S, Brown A, Nourafshan A, Benami E. 2011. Energy-based economic development. *Renew. Sustain. Energy Rev.* 15(1):282–95
4. Mitchell TM. 1997. *Machine Learning*. Burr Ridge, IL: McGraw Hill
5. U. N. Gen. Assem. 2015. *Transforming our world: the 2030 Agenda for Sustainable Development*. Publ. A/RES/70/1, U. N., New York
6. Thind MP, Tessum CW, Azevedo IL, Marshall JD. 2019. Fine particulate air pollution from electricity generation in the US: health impacts by race, income, and geography. *Environ. Sci. Technol.* 53(23):14010–19
7. Jenkins K, McCauley D, Heffron R, Stephan H, Rehner R. 2016. Energy justice: a conceptual review. *Energy Res. Soc. Sci.* 11:174–82
8. Mah A. 2017. Environmental justice in the age of big data: challenging toxic blind spots of voice, speed, and expertise. *Environ. Sociol.* 3(2):122–33
9. Perera KS, Aung Z, Woon WL. 2014. Machine learning techniques for supporting renewable energy generation and integration: a survey. In *Data Analytics for Renewable Energy Integration*, ed. WL Woon, Z Aung, S Madnick, pp. 81–96. Berlin: Springer
10. Ramchurn SD, Vytelingum P, Rogers A, Jennings NR. 2012. Putting the ‘smarts’ into the smart grid: a grand challenge for artificial intelligence. *Commun. ACM* 55(4):86–97
11. Rolnick D, Donti PL, Kaack LH, Kochanski K, Lacoste A, et al. 2019. Tackling climate change with machine learning. arXiv:1906.05433 [cs]
12. Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, et al. 2020. The role of artificial intelligence in achieving the sustainable development goals. *Nat. Commun.* 11:233
13. Gomes C, Dietterich T, Barrett C, Conrad J, Dilkina B, et al. 2019. Computational sustainability: computing for a better world and a sustainable future. *Commun. ACM* 62(9):56–65
14. Kaack LH. 2019. *Challenges and prospects for data-driven climate change mitigation*. PhD Thesis, Carnegie Mellon Univ, Pittsburgh, PA
15. Mosavi A, Salimi M, Faizollahzadeh Ardabili S, Rabczuk T, Shamshirband S, Varkonyi-Koczy AR. 2019. State of the art of machine learning models in energy systems, a systematic review. *Energies* 12(7):1301
16. Breiman L. 2001. Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Stat. Sci.* 16(3):199–231
17. Sutton RS, Barto AG. 2018. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press
18. Buşoniu L, de Bruin T, Tolić D, Kober J, Palunko I. 2018. Reinforcement learning for control: performance, stability, and deep approximators. *Annu. Rev. Control* 46:8–28
19. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, et al. 2016. Mastering the game of go with deep neural networks and tree search. *Nature* 529(7587):484–89
20. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, et al. 2020. A comprehensive survey on transfer learning. *Proc. IEEE* 109:43–76
21. Krizhevsky A, Sutskever I, Hinton GE. 2017. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60(6):84–90
22. Ma L, Liu Y, Zhang X, Ye Y, Yin G, Johnson BA. 2019. Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* 152:166–77

23. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, et al. 2020. Language models are few-shot learners. In *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS '20)*, pp. 1877–901. Red Hook, NY: Curran
24. Nassif AB, Shahin I, Attili I, Azzeh M, Shaalan K. 2019. Speech recognition using deep neural networks: a systematic review. *IEEE Access* 7:19143–65
25. Goodfellow I, Bengio Y, Courville A. 2016. *Deep Learning*. Cambridge, MA: MIT Press
26. Raghu M, Schmidt E. 2020. A survey of deep learning for scientific discovery. arXiv:2003.11755 [cs]
27. Rasmussen CE. 2003. Gaussian processes in machine learning. In *Advanced Lectures on Machine Learning*, ed. O Bousquet, U von Luxburg, G Rätsch, pp. 63–71. Berlin: Springer
28. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. 2019. A survey on bias and fairness in machine learning. arXiv:1908.09635 [cs]
29. Gebru T, Denton E. 2020. *Tutorial on fairness accountability transparency and ethics in computer vision at CVPR 2020*. Tutor. Overv., online. <https://sites.google.com/view/fatecv-tutorial>
30. Victor DG. 2019. *Blueprint for the future of AI 2018–2019: how artificial intelligence will affect the future of energy and climate*. Rep., Brookings Inst., Washington, DC. <https://www.brookings.edu/research/how-artificial-intelligence-will-affect-the-future-of-energy-and-climate/>
31. Hong T, Fan S. 2016. Probabilistic electric load forecasting: a tutorial review. *Int. J. Forecast.* 32(3):914–38
32. Kuster C, Rezgui Y, Mourshed M. 2017. Electrical load forecasting models: a critical systematic review. *Sustain. Cities Soc.* 35:257–70
33. Raza MQ, Khosravi A. 2015. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renew. Sustain. Energy Rev.* 50:1352–72
34. Khan AR, Mahmood A, Safdar A, Khan ZA, Khan NA. 2016. Load forecasting, dynamic pricing and DSM in smart grid: a review. *Sustain. Energy Rev.* 54:1311–22
35. Ledva GS, Balzano L, Mathieu JL. 2018. Real-time energy disaggregation of a distribution feeder's demand using online learning. *IEEE Trans. Power Syst.* 33(5):4730–40
36. Kell A, McGough AS, Forshaw M. 2018. Segmenting residential smart meter data for short-term load forecasting. In *Proceedings of the 9th International Conference on Future Energy Systems (e-ENERGY '18)*, pp. 91–96. New York: ACM
37. Bogomolov A, Lepri B, Larcher R, Antonelli F, Pianesi F, Pentland A. 2016. Energy consumption prediction using people dynamics derived from cellular network data. *EPJ Data Sci.* 5:13
38. Wang S, Wang S, Chen H, Gu Q. 2020. Multi-energy load forecasting for regional integrated energy systems considering temporal dynamic and coupling characteristics. *Energy* 195:116964
39. Donti P, Amos B, Kolter JZ. 2017. Task-based end-to-end model learning in stochastic optimization. In *Proceedings of the 30th Conference on Neural Information Processing Systems (NeurIPS '17)*, pp. 5484–94. New York: ACM
40. Mocanu E, Nguyen PH, Kling WL, Gibescu M. 2016. Unsupervised energy prediction in a smart grid context using reinforcement cross-building transfer learning. *Energy Build.* 116:646–55
41. Yuan Y, Dehghanpour K, Bu F, Wang Z. 2019. A multi-timescale data-driven approach to enhance distribution system observability. *IEEE Trans. Power Syst.* 34(4):3168–77
42. Ahmed R, Sreeram V, Mishra Y, Arif M. 2020. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. *Renew. Sustain. Energy Rev.* 124:109792
43. Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, et al. 2018. Forecasting of photovoltaic power generation and model optimization: a review. *Renew. Sustain. Energy Rev.* 81:912–28
44. Maldonado-Correa J, Solano J, Rojas-Moncayo M. 2019. Wind power forecasting: a systematic literature review. *Wind Eng.* 45(2):413–26
45. Giebel G, Kariniotakis G. 2017. Wind power forecasting—a review of the state of the art. In *Renewable Energy Forecasting: From Models to Applications*, ed. G Kariniotakis, pp. 59–109. Amsterdam: Elsevier
46. Kara EC, Roberts CM, Tabone M, Alvarez L, Callaway DS, Stewart EM. 2018. Disaggregating solar generation from feeder-level measurements. *Sustain. Energy Grids Netw.* 13:112–21
47. Sun Y, Szűcs G, Brandt AR. 2018. Solar PV output prediction from video streams using convolutional neural networks. *Energy Environ. Sci.* 11(7):1811–18

48. Wang HZ, Li GQ, Wang GB, Peng JC, Jiang H, Liu YT. 2017. Deep learning based ensemble approach for probabilistic wind power forecasting. *Appl. Energy* 188:56–70
49. Hao Y, Tian C. 2019. A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting. *Appl. Energy* 238:368–83
50. Du P, Wang J, Yang W, Niu T. 2019. A novel hybrid model for short-term wind power forecasting. *Appl. Soft Comput.* 80:93–106
51. Voyant C, Notton G, Kalogirou S, Nivet ML, Paoli C, et al. 2017. Machine learning methods for solar radiation forecasting: a review. *Renew. Energy* 105:569–82
52. McGovern A, Elmore KL, Gagne DJ, Haupt SE, Karstens CD, et al. 2017. Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Am. Meteorol. Soc.* 98(10):2073–90
53. Hwang J, Orenstein P, Cohen J, Pfeiffer K, Mackey L. 2019. Improving subseasonal forecasting in the western US with machine learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2325–35. New York: ACM
54. Haupt SE, Kosovic B, Jensen T, Lee J, Jimenez P, et al. 2016. *The Sun4Cast® solar power forecasting system: the results of the public-private-academic partnership to advance solar power forecasting*. Tech. Rep., Natl. Cent. Atmos. Res./Univ. Corp. Atmos. Res., Boulder CO
55. Kosovic B, Haupt SE, Adriaansen D, Alessandrini S, Wiener G, et al. 2020. A comprehensive wind power forecasting system integrating artificial intelligence and numerical weather prediction. *Energies* 13(6):1372
56. Mathe J, Miolane N, Sebastien N, Lequeux J. 2019. PVNet: a LRCN architecture for spatio-temporal photovoltaic power forecasting from numerical weather prediction. arXiv:1902.01453 [cs.LG]
57. Willard J, Jia X, Xu S, Steinbach M, Kumar V. 2020. Integrating physics-based modeling with machine learning: a survey. arXiv:2003.04919 [physics]
58. de Hoog J, Maetschke S, Ilfrich P, Kolluri RR. 2020. Using satellite and aerial imagery for identification of solar PV: state of the art and research opportunities. In *Proceedings of the 11th ACM International Conference on Future Energy Systems (e-ENERGY 2020)*, pp. 308–13. New York: ACM
59. Ralston Fonseca F, Craig M, Jaramillo P, Bergés M, Severnini E, et al. 2021. Effects of climate change on capacity expansion decisions of an electricity generation fleet in the southeast US. *Environ. Sci. Technol.* 55(4):2522–31
60. Chertkov M, Andersson G. 2020. Multienergy systems. *Proc. IEEE* 108(9):1387–91
61. Duchesne L, Karangelos E, Wehenkel L. 2020. Recent developments in machine learning for energy systems reliability management. *Proc. IEEE* 108(9):1656–76
62. Hasan F, Kargarian A, Mohammadi A. 2020. A survey on applications of machine learning for optimal power flow. In *2020 IEEE Texas Power and Energy Conference (TPEC)*, pp. 37–42. Piscataway, NJ: IEEE
63. Fioretto F, Mak TW, Van Hentenryck P. 2020. Predicting AC optimal power flows: combining deep learning and Lagrangian dual methods. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, Vol. 1, pp. 630–37. Palo Alto, CA: AAAI
64. Zamzam A, Baker K. 2020. Learning optimal solutions for extremely fast AC optimal power flow. In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1–6. Piscataway, NJ: IEEE
65. Donti PL, Rolnick D, Kolter JZ. 2021. *DC3: a learning method for optimization with hard constraints*. Poster presented at 9th International Conference on Learning Representations (ICLR 21), online, May 3–7. <https://iclr.cc/virtual/2021/poster/2868>
66. Misra S, Roald L, Ng Y. 2018. Learning for constrained optimization: identifying optimal active constraint sets. arXiv:1802.09639 [math]
67. Baker K. 2019. Learning warm-start points for AC optimal power flow. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*. Piscataway, NJ: IEEE
68. Dong W, Xie Z, Kestor G, Li D. 2020. Smart-PGSim: using neural network to accelerate AC-OPF power grid simulation. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC '20)*, art. 63. New York: ACM
69. Xavier AS, Qiu F, Ahmed S. 2020. Learning to solve large-scale security-constrained unit commitment problems. *INFORMS J. Comput.* 33(2):739–56

70. Capitanescu F. 2016. Critical review of recent advances and further developments needed in AC optimal power flow. *Electr. Power Syst. Res.* 136:57–68
71. Jereminov M, Pandey A, Pileggi L. 2018. Equivalent circuit formulation for solving ac optimal power flow. *IEEE Trans. Power Syst.* 34(3):2354–65
72. Antonopoulos I, Robu V, Couraud B, Kirli D, Norbu S, et al. 2020. Artificial intelligence and machine learning approaches to energy demand-side response: a systematic review. *Renew. Sustain. Energy Rev.* 130:109899
73. Lopez-Garcia TB, Coronado-Mendoza A, Domínguez-Navarro JA. 2020. Artificial neural networks in microgrids: a review. *Eng. Appl. Artif. Intell.* 95:103894
74. Dobbe R, Sondermeijer O, Fridovich-Keil D, Arnold D, Callaway D, Tomlin C. 2019. Toward distributed energy services: decentralizing optimal power flow with machine learning. *IEEE Trans. Smart Grid* 11(2):1296–306
75. Karagiannopoulos S, Aristidou P, Hug G. 2019. Data-driven local control design for active distribution grids using off-line optimal power flow and machine learning techniques. *IEEE Trans. Smart Grid* 10(6):6461–71
76. Hassan A, Acharya S, Chertkov M, Deka D, Dvorkin Y. 2020. A hierarchical approach to multienergy demand response: from electricity to multienergy applications. *Proc. IEEE* 108(9):1457–74
77. Zhang Z, Zhang D, Qiu RC. 2019. Deep reinforcement learning for power system applications: an overview. *CSEE J. Power Energy Syst.* 6(1):213–25
78. Glavic M. 2019. (Deep) reinforcement learning for electric power system control and related problems: a short review and perspectives. *Annu. Rev. Control* 48:22–35
79. Glavic M, Fonteneau R, Ernst D. 2017. Reinforcement learning for electric power system decision and control: past considerations and perspectives. *IFAC-PapersOnLine* 50(1):6918–27
80. Vázquez-Canteli JR, Nagy Z. 2019. Reinforcement learning for demand response: a review of algorithms and modeling techniques. *Appl. Energy* 235:1072–89
81. Ramchurn S, Vytelingum P, Rogers A, Jennings N. 2011. Agent-based control for decentralised demand side management in the smart grid. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '11)*, pp. 5–12. New York: ACM
82. Sheikhi A, Rayati M, Ranjbar AM. 2016. Demand side management for a residential customer in multi-energy systems. *Sustain. Cities Soc.* 22:63–77
83. Lago J, De Ridder F, De Schutter B. 2018. Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms. *Appl. Energy* 221:386–405
84. Bruce A, Ruff L. 2018. *Deep learning solar PV and carbon intensity forecasts*. Slides, Natl. Grid ESO, Wokingham, UK. <http://powerswarm.co.uk/wp-content/uploads/2018/10/2018.10.18-Bruce-National-Grid-ESO-Deep-Learning-Solar-PV-and-Carbon-Intensity.pdf>
85. Leerbeck K, Bacher P, Junker RG, Goranović G, Corradi O, et al. 2020. Short-term forecasting of CO₂ emission intensity in power grids by machine learning. *Appl. Energy* 277:115527
86. Du Y, Li F. 2019. Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning. *IEEE Trans. Smart Grid* 11(2):1066–76
87. King RTFA, Tu X, Dessaint LA, Kamwa I. 2016. Multi-contingency transient stability-constrained optimal power flow using multilayer feedforward neural networks. In *Proceedings of the 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. Piscataway, NJ: IEEE
88. Halilbašić L, Thams F, Venzke A, Chatzivasilakis S, Pinson P. 2018. Data-driven security-constrained AC-OPF for operations and markets. In *2018 Power Systems Computation Conference (PSCC)*. Piscataway, NJ: IEEE
89. Ni Z, Paul S. 2019. A multistage game in smart grid security: a reinforcement learning solution. *IEEE Trans. Neural Netw. Learn. Syst.* 30(9):2684–95
90. Marot A, Donnot B, Romero C, Donon B, Lerousseau M, et al. 2020. Learning to run a power network challenge for training topology controllers. *Electr. Power Syst. Res.* 189:106635
91. Yoon D, Hong S, Lee BJ, Kim KE. 2021. *Winning the L2RPN Challenge: power grid management via semi-Markov afterstate actor-critic*. Poster presented at 9th International Conference on Learning Representations (ICLR '21), online, May 3–7. <https://iclr.cc/virtual/2021/poster/3003>

92. Donti PL, Roderick M, Fazlyab M, Kolter JZ. 2020. *Enforcing robust control guarantees within neural network policies*. Poster presented at 9th International Conference on Learning Representations (ICLR '21), online, May 3–7. <https://iclr.cc/virtual/2021/poster/2899>
93. Donti PL, Liu Y, Schmitt AJ, Bernstein A, Yang R, Zhang Y. 2019. Matrix completion for low-observability voltage estimation. *IEEE Trans. Smart Grid* 11(3):2520–30
94. Pertl M, Heussen K, Gehrke O, Rezkalla M. 2016. Voltage estimation in active distribution grids using neural networks. In *Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM)*. Piscataway, NJ: IEEE
95. Jiang H, Zhang Y. 2016. Short-term distribution system state forecast based on optimal synchrophasor sensor placement and extreme learning machine. In *Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM)*. Piscataway, NJ: IEEE
96. Mokhtar M, Robu V, Flynn D, Higgins C, Whyte J, et al. 2019. Predicting the voltage distribution for low voltage networks using deep learning. In *Proceedings of the 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. Piscataway, NJ: IEEE
97. Ostrometzky J, Berestizshevsky K, Bernstein A, Zussman G. 2019. Physics-informed deep neural network method for limited observability state estimation. arXiv:1910.06401 [eess]
98. Reisi AR, Moradi MH, Jamasb S. 2013. Classification and comparison of maximum power point tracking techniques for photovoltaic system: a review. *Renew. Sustain. Energy Rev.* 19:433–43
99. Abdullah MA, Yatim A, Tan CW, Saidur R. 2012. A review of maximum power point tracking algorithms for wind energy systems. *Renew. Sustain. Energy Rev.* 16(5):3220–27
100. Abdelrahman H, Berkenkamp F, Poland J, Krause A. 2016. Bayesian optimization for maximum power point tracking in photovoltaic power plants. In *Proceedings of the 2016 European Control Conference (ECC)*, pp. 2078–83. Piscataway, NJ: IEEE
101. Abel D, Williams EC, Brawner S, Reif E, Littman ML. 2018. *Bandit-based solar panel control*. In *Proceedings of the 30th AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI-18)*, pp. 7713–18. Palo Alto, CA: AAAI
102. Rao S, Katoch S, Narayanaswamy V, Muniraju G, Tepedelenlioglu C, et al. 2020. *Machine Learning for Solar Array Monitoring, Optimization, and Control*. San Francisco, CA: Morgan & Claypool
103. Wei C, Zhang Z, Qiao W, Qu L. 2015. Reinforcement-learning-based intelligent maximum power point tracking control for wind energy conversion systems. *IEEE Trans. Ind. Electron.* 62(10):6360–70
104. Keramitsoglou I, Cartalis C, Kiranoudis CT. 2006. Automatic identification of oil spills on satellite images. *Environ. Model. Softw.* 21(5):640–52
105. Wan J, Yu Y, Wu Y, Feng R, Yu N. 2012. Hierarchical leak detection and localization method in natural gas pipeline monitoring sensor networks. *Sensors* 12(1):189–214
106. Wang J, Tchapti LP, Ravikumar AP, McGuire M, Bell CS, et al. 2020. Machine vision for natural gas methane emissions detection using an infrared camera. *Appl. Energy* 257:113998
107. Zukhrufany S. 2018. *The utilization of supervised machine learning in predicting corrosion to support preventing pipelines leakage in oil and gas industry*. Master's Thesis, Univ. Stavanger, Stavanger, Nor.
108. Narciso DA, Martins F. 2020. Application of machine learning tools for energy efficiency in industry: a review. *Energy Rep.* 6:1181–99
109. Victor DG, Yanosek K. 2017. The next energy revolution: the promise and peril of high-tech innovation. *Brookings PlanetPolicy Blog*, June 17. <https://www.brookings.edu/blog/planetpolicy/2017/06/13/the-next-energy-revolution-the-promise-and-peril-of-high-tech-innovation/>
110. Chen B, Harp DR, Lin Y, Keating EH, Pawar RJ. 2018. Geologic CO₂ sequestration monitoring design: a machine learning and uncertainty quantification based approach. *Appl. Energy* 225:332–45
111. Mo S, Zhu Y, Zabarar N, Shi X, Wu J. 2019. Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. *Water Resour. Res.* 55(1):703–28
112. Iyengar S, Lee S, Sheldon D, Shenoy P. 2018. SolarCliques: detecting anomalies in residential solar arrays. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '18)*, art. 38. New York: ACM
113. Orozco R, Sheng S, Phillips C. 2018. *Diagnostic models for wind turbine gearbox components using SCADA time series data*. Conf. Pap. NREL/CP-5000-71166, Natl. Renew. Energy Lab., Boulder, CO

114. Calivá F, De Ribeiro FS, Mylonakis A, Demazière C, Vinai P, et al. 2018. A deep learning approach to anomaly detection in nuclear reactors. In *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*. Piscataway, NJ: IEEE
115. Chen FC, Jahanshahi MR. 2017. NB-CNN: deep learning-based crack detection using convolutional neural network and naive Bayes data fusion. *IEEE Trans. Ind. Electron.* 65(5):4392–400
116. Rudin C, Waltz D, Anderson RN, Boulanger A, Salieb-Aouissi A, et al. 2011. Machine learning for the New York City power grid. *IEEE Trans. Pattern Anal. Mach. Intell.* 34(2):328–45
117. Nguyen VN, Jenssen R, Roverso D. 2018. Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning. *Int. J. Electr. Power Energy Syst.* 99:107–20
118. Maggiori E, Tarabalka Y, Charpiat G, Alliez P. 2017. Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark. In *Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3226–29. Piscataway, NJ: IEEE
119. Demir I, Koperski K, Lindenbaum D, Pang G, Huang J, et al. 2018. DeepGlobe 2018: a challenge to parse the Earth through satellite images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 172–81. Piscataway, NJ: IEEE
120. Yang HL, Yuan J, Lunga D, Laverdiere M, Rose A, Bhaduri B. 2018. Building extraction at scale using convolutional neural network: mapping of the United States. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11(8):2600–14
121. Streltsov A, Malof JM, Huang B, Bradbury K. 2020. Estimating residential building energy consumption using overhead imagery. *Appl. Energy* 280:116018
122. Huang B, Yang J, Streltsov A, Bradbury K, Collins LM, Malof J. 2021. GridTracer: automatic mapping of power grids using deep learning and overhead imagery. arXiv:2101.06390 [cs.CV]
123. Arderne C, Zorn C, Nicolas C, Koks EE. 2020. Predictive mapping of the global power system using open data. *Sci. Data* 7:19
124. Ellman DDA. 2015. *The reference electrification model: a computer model for planning rural electricity access*. PhD Thesis, MIT, Cambridge, MA
125. Mellit A, Benghamem M, Arab AH, Guessoum A. 2003. Modelling of sizing the photovoltaic system parameters using artificial neural network. In *Proceedings of the 2003 IEEE Conference on Control Applications (CCA 2003)*, Vol. 1, pp. 353–57. Piscataway, NJ: IEEE
126. Wu X, Gomes-Selman J, Shi Q, Xue Y, García-Villacorta R, et al. 2018. Efficiently approximating the Pareto frontier: hydropower dam placement in the Amazon basin. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI-18)*, pp. 849–59. Palo Alto, CA: AAAI
127. Moutis P, Skarvelis-Kazakos S, Brucoli M. 2016. Decision tree aided planning and energy balancing of planned community microgrids. *Appl. Energy* 161:197–205
128. Bengio Y, Lodi A, Prouvost A. 2020. Machine learning for combinatorial optimization: a methodological tour d'horizon. *Eur. J. Oper. Res.* 290(2):405–21
129. Guo S, Jiang X, Thornton T, Saunders D. 2019. Approximate string matching of power system substation names. In *Proceedings of the 2019 IEEE Power and Energy Conference at Illinois (PECI)*. Piscataway, NJ: IEEE
130. Catalyst Coop. 2020. *Public Utility Data Liberation Project*. Dataset Updates, Catalyst Coop., Boulder, CO. <https://zenodo.org/communities/catalyst-cooperative/>
131. Valerio L, Passarella A, Conti M. 2016. Hypothesis transfer learning for efficient data computing in smart cities environments. In *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*. Piscataway, NJ: IEEE
132. Muhammad K, Lloret J, Baik SW. 2019. Intelligent and energy-efficient data prioritization in green smart cities: current challenges and future directions. *IEEE Commun. Mag.* 57(2):60–65
133. Chu S, Cui Y, Liu N. 2017. The path towards sustainable energy. *Nat. Mater.* 16:16–22
134. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. 2018. Machine learning for molecular and materials science. *Nature* 559(7715):547–55
135. Liu Y, Zhao T, Ju W, Shi S. 2017. Materials discovery and design using machine learning. *J. Mater.* 3(3):159–77

136. Fujimura K, Seko A, Koyama Y, Kuwabara A, Kishida I, et al. 2013. Accelerated materials design of lithium superionic conductors based on first-principles calculations and machine learning algorithms. *Adv. Energy Mater.* 3(8):980–85
137. Raccuglia P, Elbert KC, Adler PD, Falk C, Wenny MB, et al. 2016. Machine-learning-assisted materials discovery using failed experiments. *Nature* 533(7601):73–76
138. Zhang L, He M, Shao S. 2020. Machine learning for halide perovskite materials. *Nano Energy* 78:105380
139. Bai J, Xue Y, Björck J, Le Bras R, Rappazzo B, et al. 2018. Phase Mapper: accelerating materials discovery with AI. *AI Mag.* 39(1):15–26
140. Gomes CP, Bai J, Xue Y, Björck J, Rappazzo B, et al. 2019. CRYSTAL: a multi-agent AI system for automated mapping of materials’ crystal structures. *MRS Commun.* 9(2):600–8
141. Zitnick CL, Chanussot L, Das A, Goyal S, Heras-Domingo J, et al. 2020. An introduction to electrocatalyst design using machine learning for renewable energy storage. arXiv:2010.09435 [cond-mat.mtrl-sci]
142. Ren Z, Oviedo F, Thway M, Tian SI, Wang Y, et al. 2020. Embedding physics domain knowledge into a Bayesian network enables layer-by-layer process innovation for photovoltaics. *npj Comput. Mater.* 6:9
143. Attia PM, Grover A, Jin N, Severson KA, Markov TM, et al. 2020. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* 578(7795):397–402
144. Humphreys D, Kupresanin A, Boyer M, Canik J, Chang C, et al. 2020. Advancing fusion with machine learning research needs workshop report. *J. Fusion Energy* 39(4):123–55
145. Baltz E, Trask E, Binderbauer M, Dikovskiy M, Gota H, et al. 2017. Achievement of sustained net plasma heating in a fusion experiment with the optometrist algorithm. *Sci. Rep.* 7:6425
146. Kates-Harbeck J, Svyatkovskiy A, Tang W. 2019. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature* 568(7753):526–31
147. Yu J, Wang Z, Majumdar A, Rajagopal R. 2018. DeepSolar: a machine learning framework to efficiently construct a solar deployment database in the United States. *Joule* 2(12):2605–17
148. Nassar R, Hill TG, McLinden CA, Wunch D, Jones DB, Crisp D. 2017. Quantifying CO₂ emissions from individual power plants from space. *Geophys. Res. Lett.* 44(19):10–45
149. Buchwitz M, Reuter M, Schneising O, Bovensmann H, Burrows JP, et al. 2018. Copernicus Climate Change Service (C3S) global satellite observations of atmospheric carbon dioxide and methane. *Adv. Astron. Sci. Technol.* 1:57–60
150. Kuhlmann G, Broquet G, Marshall J, Clément V, Löscher A, et al. 2019. Detectability of CO₂ emission plumes of cities and power plants with the Copernicus anthropogenic CO₂ Monitoring (CO2M) mission. *Atmos. Meas. Tech.* 12(12):6695–719
151. Couture HD, O’Connor J, Mitchell G, Söldner-Rembold I, D’Souza D, et al. 2020. Towards tracking the emissions of every power plant on the planet. In *NeurIPS 2020 Workshop Tackling Climate Change with Machine Learning*, pap. 11. San Diego, CA: Neur. Inf. Process. Syst.
152. Venugopalan S, Rai V. 2015. Topic based classification and pattern identification in patents. *Technol. Forecast. Soc. Change* 94:236–50
153. Callaghan MW, Minx JC, Forster PM. 2020. A topography of climate change research. *Nat. Clim. Change* 10(2):118–23
154. de Avila Belbute-Peres F, Smith K, Allen K, Tenenbaum J, Kolter JZ. 2018. End-to-end differentiable physics for learning and control. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS ’18)*, pp. 7178–89. New York: ACM
155. Chen RTQ, Rubanova Y, Bettencourt J, Duvenaud DK. 2018. Neural ordinary differential equations. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS ’18)*, pp. 6571–83. New York: ACM
156. Greydanus S, Dzamba M, Yosinski J. 2019. Hamiltonian neural networks. In *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS ’19)*, pp. 15379–89. New York: ACM
157. Wong E, Kolter JZ. 2018. Provable defenses against adversarial examples via the convex outer adversarial polytope. *PMLR* 80:5286–95
158. Raghuathan A, Steinhardt J, Liang P. 2018. Certified defenses against adversarial examples. In *Proceedings of the 6th International Conference on Learning Representations (ICLR ’18)*. arXiv:1801.09344 [cs]
159. Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. 2021. Interpretable machine learning: fundamental principles and 10 grand challenges. arXiv:2103.11251 [cs.LG]

160. Doshi-Velez F, Kim B. 2017. Towards a rigorous science of interpretable machine learning. arXiv:1702.08608 [stat.ML]
161. Ghahramani Z. 2015. Probabilistic machine learning and artificial intelligence. *Nature* 521(7553):452–59
162. Rudin C. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1(5):206–15
163. Wilder B, Dilkina B, Tambe M. 2019. Melding the data–decisions pipeline: decision-focused learning for combinatorial optimization. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI-19)*, pp. 1658–65. Palo Alto, CA: AAAI
164. Cronin J, Anandarajah G, Dessens O. 2018. Climate change impacts on the energy system: a review of trends and gaps. *Clim. Change* 151(2):79–93
165. Hilbert M. 2016. Big Data for development: a review of promises and challenges. *Dev. Policy Rev.* 34(1):135–74
166. Moreno JB, Timms M, Wildberger K. 2019. *How artificial intelligence is accelerating the energy transition: an overview of AI activities at E.ON*. Rep., E.ON SE, Munich, Ger.
167. Donaghy T, Henderson C, Jardim E. 2019. *Oil in the cloud: how tech companies are helping Big Oil profit from climate destruction*. Rep., Greenpeace, Washington, DC