

Review

A Survey on the Use of Synthetic Data for Enhancing Key Aspects of Trustworthy AI in the Energy Domain: Challenges and Opportunities

Michael Meiser *  and Ingo Zinnikus

German Research Center for Artificial Intelligence (DFKI), Saarland Informatics Campus (SIC),
66123 Saarbruecken, Germany; ingo.zinnikus@dfki.de

* Correspondence: michael.meiser@dfki.de

Abstract: To achieve the energy transition, energy and energy efficiency are becoming more and more important in society. New methods, such as Artificial Intelligence (AI) and Machine Learning (ML) models, are needed to coordinate supply and demand and address the challenges of the energy transition. AI and ML are already being applied to a growing number of energy infrastructure applications, ranging from energy generation to energy forecasting and human activity recognition services. Given the rapid development of AI and ML, the importance of Trustworthy AI is growing as it takes on increasingly responsible tasks. Particularly in the energy domain, Trustworthy AI plays a decisive role in designing and implementing efficient and reliable solutions. Trustworthy AI can be considered from two perspectives, the Model-Centric AI (MCAI) and the Data-Centric AI (DCAI) approach. We focus on the DCAI approach, which relies on large amounts of data of sufficient quality. These data are becoming more and more synthetically generated. To address this trend, we introduce the concept of Synthetic Data-Centric AI (SDCAI). In this survey, we examine Trustworthy AI within a Synthetic Data-Centric AI context, focusing specifically on the role of simulation and synthetic data in enhancing the level of Trustworthy AI in the energy domain.

Keywords: Trustworthy AI; synthetic data; Data-Centric AI; technical robustness; transparency; explainability; reproducibility; privacy; fairness; sustainability



Citation: Meiser, M.; Zinnikus, I. A Survey on the Use of Synthetic Data for Enhancing Key Aspects of Trustworthy AI in the Energy Domain: Challenges and Opportunities. *Energies* **2024**, *17*, 1992. <https://doi.org/10.3390/en17091992>

Academic Editors: Ershun Du, Qin Wang and Mingjian Cui

Received: 29 February 2024

Revised: 19 April 2024

Accepted: 20 April 2024

Published: 23 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Awareness of energy and energy efficiency is growing rapidly in the face of global climate change, making the energy transition a major societal issue. To achieve the energy transition, efficient, reliable and sustainable energy technologies are needed [1–3].

A basic measure of energy efficiency is first of all the ratio between the input and output of a particular energy generation or conversion system. On the basis of this ratio, further efficiency measures and performance indices can be defined that allow two systems or different versions of a particular system to be compared. The aim is usually to increase the efficiency of a given system. Sustainability plays an important role, as the energy source should be based on renewable and environmentally friendly resources. Finally, there is a strong, reciprocal relationship between reliability and efficiency, particularly in the case of renewable energies. This is illustrated by the fact that performance measures for photovoltaic systems include temporal and geographic factors (e.g., actual insolation) that influence the practical efficiency of a system [4].

As demand and supply are becoming increasingly interlinked with the use of renewable energies, new methods are required to adjust supply and demand. Artificial Intelligence (AI) and Machine Learning (ML) models are considered to be a key factor in accomplishing the energy transition [5–7] and are already widely used in various areas of the energy sector. These areas include cybersecurity analysis and simulation for power system protection [8,9], simulation-based studies on grid stability, reliability, and resilience [10–13],

simulation and analysis of smart grid technologies and architectures [14,15], as well as advanced simulation techniques for power system modeling and analysis [16,17]. Offering new and promising opportunities, AI technology continues to expand not only in the energy sector [18–20] but also in many other domains such as healthcare [21,22] and finance [23–25].

With the ongoing development and popularization of AI and ML models, these models are more and more used to perform highly responsible tasks. Since the functionality of ML models is not always fully explainable and reliable—known as the black-box problem [26]—the topic of *Trustworthy AI* [27–29] is becoming increasingly important. When AI approaches are used to improve the efficiency of applications in the energy sector, the question of the reliability and trustworthiness of the AI approaches themselves becomes a crucial issue.

In general, two primary approaches exist for establishing a certain level of trust in AI (see Figure 1). On the one hand, we can focus on the ML algorithms themselves and work towards achieving “fairness” for the models, also referred to as a model-centric approach or *Model-Centric AI* (MCAI) [30]. On the other hand, we can analyze the data on which these ML algorithms are trained, commonly known as the data-centric approach or *Data-Centric AI* (DCAI) [31]. The DCAI approach is often disregarded in current AI research [32]. However, ML models require large amounts of data to become robust and functional [33,34] and high-quality datasets are essential. Even the best ML models are unable to perform well if trained on insufficient data [35]. As the interest in AI and ML models continues to grow, the availability of large amounts of data is becoming increasingly important.

Because data are essential, the DCAI approach is a very promising method for ensuring Trustworthy AI. Data are typically collected in real-time in the physical world. However, collecting a sufficient amount of sensor data this way to be able to meaningfully train ML algorithms is a time-consuming and cumbersome task. In addition, collected data can contain gaps [36] or incorrect samples due to sensor measurement errors, the data themselves or even the ground-truth data are not always annotated or there can be inaccuracies in labeling [37]; data are biased in some way [38], or they cannot be collected at all due to privacy regulations. Since data collection is cumbersome, alternative methods of data collection are already being explored, such as the participatory collection of data from people [39–41].

Simulations and synthetic data provide the potential to address and solve real-world data collection problems. For instance, synthetic data allow for reduced data collection costs or to compensate for biases in datasets [42]. Furthermore, synthetic data can be generated fully labeled and annotated with ground truth, without any data gaps or sampling rate variations [17]. As a result, simulations and the generation of synthetic data have grown in popularity and are now utilized in various AI applications across various domains [43], such as energy [44–46], healthcare [47–49], finance [50,51] or manufacturing [52,53].

However, developing simulations and generating synthetic data poses challenges. Synthetic data are highly domain-dependent, as each domain has its own set of characteristics that must be addressed in order for synthetic data to be sufficiently representative and meaningful. The generation of unbiased synthetic data is a developing discipline and requires further research to take full advantage of this technology [54].

Consequently, we examine the potential and impact of synthetic data in the energy domain from the perspective of Trustworthy AI and investigate the following research question:

Given that Trustworthy AI encompasses the aspects of efficiency, reliability and sustainability, what benefits do synthetic energy data contribute to the development of these aspects of Trustworthy AI in the energy domain?

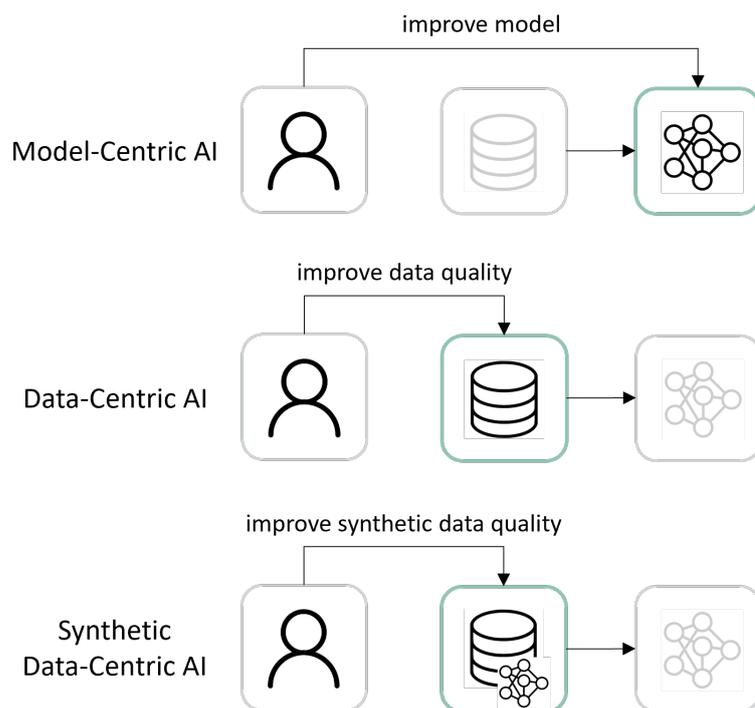


Figure 1. High-level comparison between the Model-Centric AI, Data-Centric AI and Synthetic Data-Centric AI approach based on the graphic presented in [31].

Because synthetic data offer such high potential to improve Trustworthy AI, we introduce the term *Synthetic Data-Centric AI* (SDCAI). As illustrated in Figure 1, the SDCAI approach expands upon the Data-Centric AI approach. Using an SDCAI approach, a developer or researcher attempts to enhance the performance of a ML model by generating and refining the synthetic data used to train the model.

This survey is structured as follows: in Section 2, we first examine the different definitions and aspects of Trustworthy AI and determine what we understand by this term. We then provide a more detailed analysis of the individual aspects of Trustworthy AI, including *technical robustness and generalization* in Section 4.1, *transparency and explainability* in Section 4.2, *reproducibility* in Section 4.3, *fairness* in Section 4.4, *privacy* in Section 4.5 and *sustainability* in Section 4.6. For each of these aspects, we will elaborate on how synthetic data are able to contribute to increasing the level of trust in the energy domain. In Section 5, we finally examine the key features of Trustworthy AI that contribute to improving the quality of synthetic data and identify the areas in which synthetic data have the greatest potential to enhance trust.

2. Trustworthy AI

While a final definition for Trustworthy AI has not yet been established, the proposals converge in several aspects that are crucial for fostering user acceptance of reliable AI systems. The European Commission’s High-Level Expert Group on AI (HLEG-AI) with the “Assessment List for Trustworthy Artificial Intelligence (ALTAI)” [55] as well as other experts and institutions defined several factors [56–62]. As mentioned, these definitions for Trustworthy AI overlap in various factors and can generally be divided into technical and non-technical (ethical and other) factors (see Figure 2).

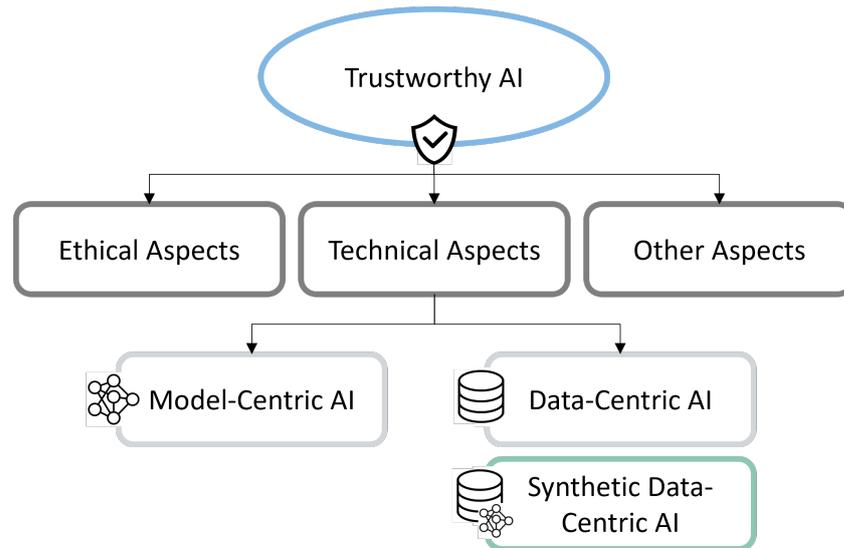


Figure 2. Schematic view on Trustworthy AI and the involvement of the Synthetic Data-Centric AI approach.

This survey approaches the topic of ensuring Trustworthy AI using synthetic data from a technical perspective, as we are convinced that this is where synthetic data have the most potential to improve the level of trust. In our understanding, enhancing efficiency, reliability, and sustainability requires technical considerations as well. Therefore, these aspects are the most suitable for improvement by using synthetic data.

Figure 3 illustrates the technical facets of Trustworthy AI from an SDCAI perspective. In total, we extracted eight key aspects from the previously mentioned definitions: *technical robustness, generalization, transparency, explainability, reproducibility, fairness, privacy and sustainability*.

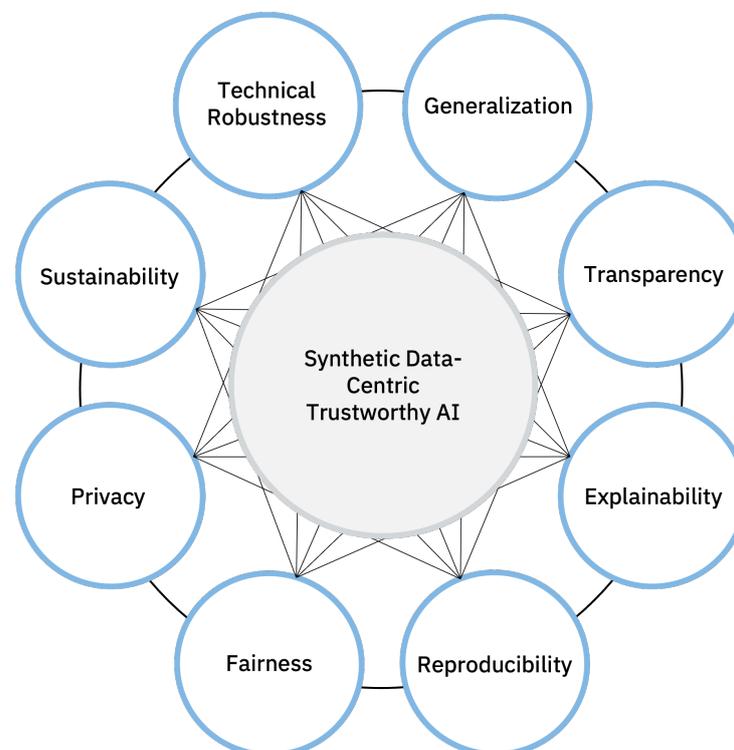


Figure 3. Eight key aspects of the Synthetic Data-Centric AI approach for enhancing Trustworthy AI.

Technical robustness refers to the ability of ML models to provide accurate results even when faced with data that differ from the training data. To accomplish this, ML models require high-quality training data, which involves several factors, such as the availability of ground truth data, the absence of data gaps, and appropriate data labeling (see Section 4.1).

Generalization is closely linked to the aspect of technical robustness and refers to the ability of ML models to perform effectively on unseen data given a limited amount of training data (see Section 4.1).

The principles of *transparency* and *explainability* are closely interrelated and emphasize the need for comprehensive visibility and the ability to understand the behavior of an AI system. Transparency, as well as explainability, are also highly dependent on data quality since these factors can only be guaranteed if the training data sufficiently represent the underlying real-world data (see Section 4.2).

If the data quality is inadequate, issues may arise with the *reproducibility* of experiments and studies (see Section 4.3). Reproducibility requires that ML models developed in scientific work, but also commercial ML models, can be replicated by other researchers at any point in time. Ideally, the experiments described should yield comparable or identical results.

The aspect of *fairness* is also an essential part of data quality, as imbalanced and biased data may hinder the creation of robust ML models that require balanced datasets (see Section 4.4).

Further, it is important to ensure *privacy* and security as components of data quality to prevent personal data from being traced back from synthetic data (see Section 4.5).

Sustainability in the context of Trustworthy AI focuses on specific aspects of the overall concept of sustainability, in the sense of renewable and environmentally friendly resources. In this survey, *sustainability* refers to the fact that AI systems and ML models should be trained and refined on data that are collected/generated under the most environmentally friendly conditions possible, e.g., by using renewable energy resources.

Naturally, the quality of the training data is a critical factor for the optimal functioning and robustness of ML models (see Section 1). As part of the DCAI approach, it is crucial to comprehend the requisite data quality for effective and efficient ML training. Providing high-quality datasets for training ML models is at least as crucial as improving the algorithms themselves. If the data contain biases, it is practically impossible for the trained algorithms to be unbiased.

Therefore, if ML models are to be improved through understanding and optimizing data using the DCAI approach, it is crucial to ensure the data meet certain quality standards. This involves labeling the data, eliminating data gaps, preventing and minimizing bias, and ensuring an adequate quantity is available. Improving dataset quality is a crucial aspect addressed by Trustworthy AI.

Almost all of the technical cornerstones that we have defined to ensure Trustworthy AI are closely related to the quality of the data.

The problems associated with the DCAI approach for ML using real-world data prompted an investigation into whether synthetic data have the potential to reasonably extend this approach in the context of Trustworthy AI. Synthetic data have the ability to provide answers to various issues in several areas related to Trustworthy AI, such as data augmentation for robustness, private data release, data de-biasing and fairness [54].

We are convinced that synthetic data have the potential to be a key enabler in the development of Trustworthy AI. Therefore, this survey focuses on understanding how synthetic energy data can contribute to ensuring Trustworthy AI. To this end, we introduce the new term *Synthetic Data-Centric AI* (SDCAI) approach, which is an extension of the DCAI approach (see Figure 1). The SDCAI approach addresses the question of how to train ML algorithms on synthetic datasets in a meaningful and reliable way.

In this contribution, we focus specifically on the energy domain. As argued before, there exist serious technological challenges that can be addressed by using ML and AI systems. Among these challenges are the storage and distribution of energy through

grids, which play a crucial role in attaining a reliable energy supply in the future. The energy domain is a suitable application example for Trustworthy AI because it shares many characteristics with other domains, which allows the findings of this survey to be widely applicable beyond this particular field into other domains.

To the best of our knowledge, no study has specifically addressed the advances, challenges, and opportunities of synthetic data for the development of Trustworthy AI in the energy domain. Ref. [63] provide a review of advances, challenges, and opportunities in generating data for some aspects of Trustworthy AI [63], but the authors do not address key aspects that we consider in this survey, such as explainability, reproducibility, and sustainability. Furthermore, we are not aware of any research exploring a Data-Centric AI approach for Trustworthy AI, specifically in the energy sector, nor referencing a Synthetic Data-Centric AI approach.

3. Synthetic Data

To explore synthetic data and the SDCAI approach, it is essential to have a clear understanding of the concept of synthetic data. The idea of synthetic data reaches back at least as far as to the Monte Carlo Simulation [64] and can be defined as follows [54]:

Definition 1. “Synthetic data are data that have been generated using a purpose-built mathematical model or algorithm, with the aim of solving a (set of) data science task(s).” [54].

In the energy domain, the amount of image data processed is not as large as in other domains, such as the healthcare domain [65,66]. The majority of the data that are processed in the context of the energy domain are time-series data, specifically consumption data for different types of energy, including electricity, wind, water, and solar. Additionally, there are time-series data collected from sensors that monitor variables such as temperature, humidity, or motion.

There are many approaches to synthetically generate time series data in the energy domain, e.g., based on using ML models and different neural architectures [44,67–71]. This includes AI applications that make decisions, such as an energy forecast for distribution grids that controls energy supply and demand. This enables the determination of the amount of energy allocated to households and buildings.

The methods that can be used to generate synthetic data, especially in the energy domain, are discussed further in Section 3.2. However, it is necessary to collect and prepare real and synthetic data before generating data.

3.1. Data Preparation

Data preparation is an important component for all types and uses of synthetic data, but especially for ensuring Trustworthy AI principles, and thus for the Synthetic Data-Centric AI approach as we define it.

Data Preparation can be divided into two general sub-areas: *data collection* (Section 3.1.1) and *data preprocessing* (Section 3.1.2) [72].

3.1.1. Real-World Data Collection

ML models frequently lack a sufficient amount of labeled data for training [33]. Consequently, collecting real-world data is an essential task in the development of such models. As previously mentioned in Section 1, collecting data in the real world causes many problems in principle, but especially in the energy domain due to its cumbersome and time-consuming nature.

The majority of the energy data collected are about electricity in private households, but other sensor data are also collected in a household, such as gas, temperature, or humidity.

Freely available real-world datasets are often published without appropriate documentation, making them difficult to use [73]. These datasets often suffer from the fact that the

data are not fully labeled and there is no guarantee that the labels are correct. Specifically, the ground-truth data can lack annotations or labels [37] or even be non-existent. However, ground-truth data are indispensable for numerous supervised learning problems, such as Non-Intrusive Load Monitoring (NILM) algorithms [74]. In particular, freely available real-world energy data suffer from the fact that never all the ground truth data that make up the smart meter data are available. Various datasets exemplify this problem, such as REFIT [75], GeLaP [76], ENERTALK [77], GREEND [78], IEDL [79], UK-DALE [80] or [81].

The IDEAL dataset contains electrical and gas data for private households, including individual room temperature and humidity readings and temperature readings from the boiler [82]. The available sensor data are augmented by anonymized survey data and metadata including occupant demographics, self-reported energy awareness and attitudes, and building, room and appliance characteristics. Energy data were collected for both consumption and PV generation [83]. Indoor climate variables such as temperature, airflow, relative humidity, CO₂ level and illuminance were also collected.

The majority of the available datasets listed contain metadata, but the metadata are incomplete. In some cases, for example, device types are available, but there is no information about a manufacturer or year of construction. All of the datasets listed are systemically biased in some way (see Section 4.4) since they were all collected locally in a single country, or at least on a single continent. Each country has people with their own country-specific habits, population groups and consumer behavior, which ultimately results in their own energy consumption. Furthermore, publicly available datasets pose the issue of representing only a small subset of a population. Due to the large amount of effort involved in collecting the data, such a dataset typically contains only a few households over a short period of time, which leads to statistical bias. To substantially reduce both systemic and statistical biases, it is necessary to obtain data from a larger subsample of a population that is more representative. For example, the dataset should include information from a wider range of countries, population groups, and ethnicities. However, this is challenging to achieve in practice and would require a considerable amount of time.

Synthetic data could help in this case. Simulation tools such as [84–86] for electricity or heating [87–91], allow the simulation of all sorts of human behavior and habits and thus also the generation of synthetic energy consumption data. If generated using a simulation, synthetic data have the advantage over collected real-world data of being fully labeled and ensuring “ground truth” for all appliances used without the existence of data holes [17]. However, this human behavior is very complex to simulate, making human behavior one of the most critical parameters in energy models. Nevertheless, there are a number of works that propose and develop concepts for the simulation of human behavior within the energy domain, such as [92–94].

Irrespective of the methods used to generate synthetic data, a key challenge in using synthetic data is evaluating the quality of the data and how accurately the synthetic data represent the real data (see Section 4.1). The quality of the synthetic energy data must be guaranteed because it is pointless to generate synthetic energy data that do not adequately reflect the domain (see Section 3.2).

3.1.2. Data Preprocessing

The amount of effort needed for data preprocessing is growing and is already a very large part of the ML model development process, consuming over 80 percent of the time and resources before the actual model can finally be developed [95].

The data preprocessing steps include all steps that are necessary before data can be fed into an AI system and thus used for training or testing ML models. These steps include a number of aspects such as data cleaning, anomaly detection, data anonymization, and data privacy [72].

When discussing the individual aspects of Trustworthy AI throughout this survey (see Section 4), we will occasionally encounter data preprocessing steps, which will be addressed in more detail in that specific aspect.

3.2. Data Generation

There are basically three ways to generate synthetic data: based on real data, without real data, or as a hybrid combination of these two types [96]. These approaches can be applied to the energy domain as well, resulting in synthetic data derived from any or all of the three methods.

For example, ML models can be trained on pure real-world temperature or electricity time series data, which in turn, generate synthetic data [97–100].

With different simulations, it is also possible to generate synthetic data using both real and synthetic data. For instance, in an electricity simulation of a household, the human behavior of the residents is not generated based on real data, but randomly [84–86]. This means that residents randomly turn on and off appliances in an apartment and whenever an appliance is turned on, the simulation uses the real power consumption of the appliance to calculate the total power consumption of the apartment.

It is also possible to generate synthetic energy data without the direct use of real-world data. For example, a simulation can be used to customize the human behavior of residents when turning on appliances, as well as to synthetically generate the energy consumption of individual appliances [101–103].

The *utility*, i.e., the extent to which a synthetic dataset is an exact substitute for real data, depends on the fidelity of the underlying generation model [96]. There is no universal method for measuring the utility of synthetic data [104], instead, there are two different concepts, referred to as general and specific utility measures [105]. A general utility measure concept for synthetic data that is most frequently described is the propensity score [105,106]. This involves developing a classification model that distinguishes between real and synthetic data. If the model cannot distinguish between the two datasets, the synthetic data have a high degree of utility. Since synthetic data are ultimately intended to be used to train and test ML models, it should be ensured that this type of model can be trained and tested on these data. [17] describes a methodology that ensures the quality of the synthetic data by using ML models. The authors use exemplary NILM models trained on both synthetic and real data and then compare their results. They demonstrate that ML models trained on synthetic data can even outperform models trained on comparable real data. Specific measures for the utility of synthetic data are confidence interval overlap [107] and standardized bias [108], which work with statistical methods.

However, there are also risks in using synthetic data to train AI systems such as data quality (including data pollution or data contamination), bias propagation, security risks and misuse [42]. This survey is well aware of the risks of synthetic data, and therefore, addresses these risks for the conditions in the energy domain and develops and presents solutions for them.

In the following sections, the opportunities for using synthetic data to develop Trustworthy AI in the energy domain are discussed in more detail for each of the aspects of Trustworthy AI considered in this survey.

4. Aspects of Trustworthy Synthetic Data-Centric AI

In the following, we consider the individual aspects that are required for an AI application to be considered trustworthy. For each aspect, we will discuss in more detail how synthetic data can be used to improve these aspects of Trustworthy AI (see Section 2) and thereby develop trust in AI (SDCAI approach).

The content of the different aspects under consideration often overlaps, as they are interdependent and partially cover related topics. Figure 4 illustrates a graphical representation of the literature fields discussed in this review, visualized by VOSviewer, a tool for constructing and visualizing bibliometric networks (VOSviewer: <https://www.vosviewer.com/>, accessed on 15 April 2024).

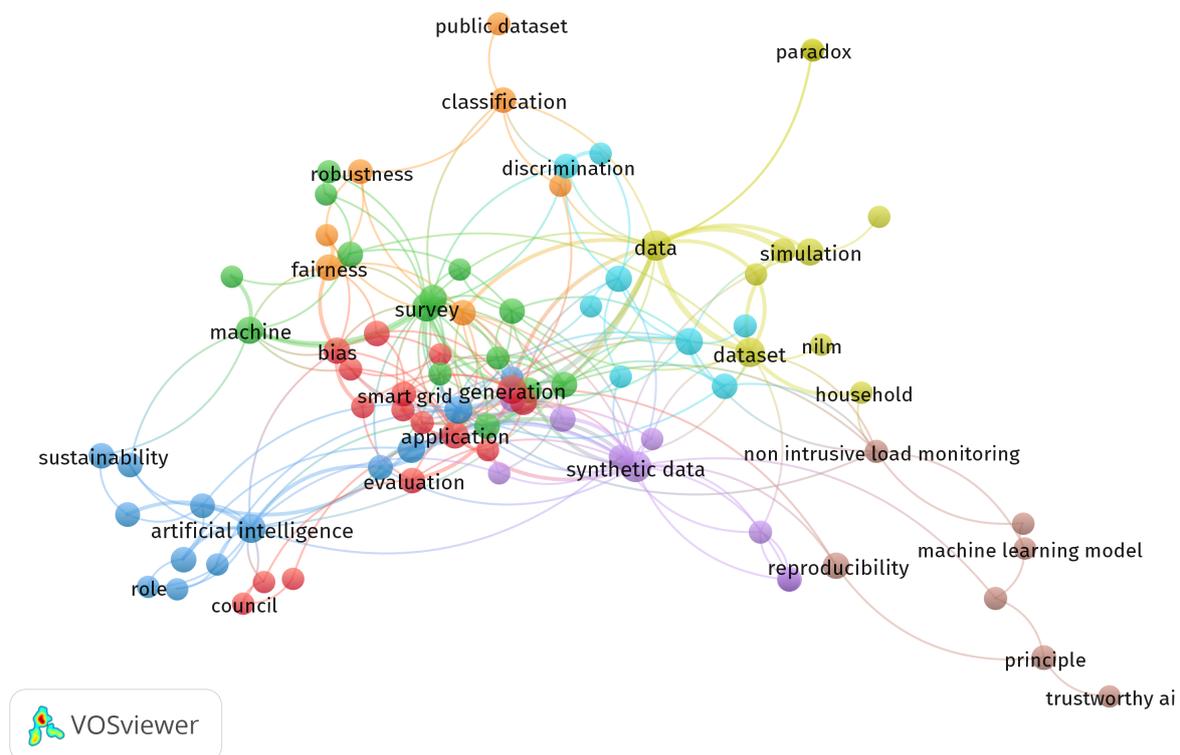


Figure 4. Graphical representation of the literature fields discussed in this review, visualized by VOSviewer.

4.1. Technical Robustness and Generalization

To ensure trustworthiness and to prevent and minimize unintended malfunctions, AI systems should have an appropriate level of *technical robustness*. The term ‘technical robustness’ of AI systems is broad and covers many aspects. Among other things, it refers to the ability of ML models to perform on unseen data or robustness against samples that are not very similar to the data on which a model was trained [109]. The concept of technical robustness is an important cornerstone for ensuring Trustworthy AI. The improvement of balanced and robust training techniques and datasets can enhance not only fairness (see Section 4.4) but also explainability (see Section 4.2) [110].

The aspect of *generalization* is closely related to the aspect of technical robustness and represents the ability of an AI system to make accurate predictions about unknown data based on limited training data [72]. It is preferable for ML models to maximize generalizability while minimizing the amount of training data required, as data collection is both time-consuming and resource-intensive (see Section 4.6).

Although both concepts of technical robustness and generalization are closely related, there is still no general consensus on whether greater robustness is beneficial or disadvantageous for the ability of ML models to generalize. In the literature, there are arguments both in favor and against such an argumentation [111–113].

Data design is able to ensure the technical robustness, reliability and generalizability of an AI model trained on those data [114]. It is well understood that variant data with different distributions and different scenarios are important for robustness since training an AI model without such data can seriously affect its performance [72].

Synthetic data offer the powerful advantage of allowing the generation of hypothetical scenarios, such as critical scenarios. AI systems and data-driven models should perform effectively in situations that do not occur sufficiently often in real-world data, such as critical situations, to be considered trustworthy [54]. Including situations that may occur in the real world but are extremely rare and should not be intentionally induced is extremely important, because ML models should be trained and tested with every possible situation

that could occur in the real world in order to achieve maximum robustness and reliability. For instance, an AI assistance system designed to assist elderly individuals in private households using various sensor data cannot learn how to behave appropriately in a critical situation if such a scenario is not included in the training data. Moreover, it is not possible to test and verify the performance of the system in such a situation without adequate testing data. Synthetic data have already been applied across multiple domains to increase the robustness of AI systems, including visual machine learning [115] and churn prediction [116]. Synthetic data are also used in other domains, such as covering nuclear power plant accidents [117], ensuring safe drone landings [118], and generating critical autonomous driving situations to improve AI-based systems. A number of techniques have already been used to generate safety-critical driving scenarios, such as those based on accident sketches [119], based on a search algorithm that iteratively optimizes behavior action sequences of the surrounding traffic participants [120], based on influential behavior patterns [121] or based on reinforcement learning [122]. The approaches mentioned here for representing critical situations are domain-specific. The exact definition of what constitutes a critical situation is typically dependent on a particular domain. It remains an open question whether the creation of these critical situations with the methods mentioned can be transferred to other domains. Furthermore, it has not yet been clarified how many and what kind of critical situations must be present in a dataset for ML models to be considered trustworthy.

Especially in the energy domain, it is essential for ML models to be both technically robust and generalizable in order to achieve a high level of transferability. This characteristic is particularly important for ML models, like assistance systems or NILM models, that are trained and developed based on a restricted sample of households but required to be functional and robust in other households [123,124]. In this context, transferability refers to the ability of ML models to be both robust and highly generalizable. This means that the model is able to produce accurate results on households that were not included in the training dataset of the model [125].

However, to be effective in training ML algorithms, synthetic data must be of consistently high quality without bias. Previous research has demonstrated that biased data can negatively impact the generalization properties of ML algorithms [126–128]. For instance, if an energy consumption prediction model is only trained using energy data collected from private households in Europe, the model will have difficulties accurately predicting the energy consumption of households in South America or Asia. The energy consumption patterns in these regions are different, as people there have other relevant characteristics, such as different consumer behaviors, different climatic conditions, and different appliances in their households, which result in other forms of energy consumption.

Also, in the energy domain, synthetic data can enhance the robustness and generalizability of ML models by producing a multitude of diverse training datasets. Simulations are capable of generating substantial amounts of high-quality energy data from a variety of demographic groups such as denomination, nationality, gender, age, occupation, income, and different life circumstances (see Section 3.2). This allows ML models to be better prepared for real-world scenarios and substantially enhances the accuracy of predictions and outcomes of AI systems.

In an ideal world, the data for all the described scenarios would be freely available as benchmark datasets. Once available, these datasets can be used to train and test ML models, increasing their robustness and generalizability. The objective of benchmark datasets, which are already used in the ML domain, is to represent the real world as reliably as possible [127] and can also be generated synthetically. The idea of synthetic benchmark datasets is closely related to the concept of transparency and explainability and is, therefore, described in more detail in the following Section.

4.2. Transparency and Explainability

Another key factor in achieving Trustworthy AI is ensuring the transparency, explainability and accountability of ML models and the data used to train and test them. In academic literature, the term *transparency* is typically understood in the sense that all components of an AI system should be visible and explainable, including the training data [55].

Definition 2. *Transparency can also be defined “as any information provided about an AI system beyond its model outputs.” [129].*

The concept of transparency is one of the key topics addressed by *Explainable AI* (XAI) [130]. The purpose of XAI is to enable individuals to understand precisely and in detail why and on what basis an AI system has or has not made a decision [131]. Ensuring the transparency of the data utilized for training and testing ML models is an important aspect of XAI.

ML models and the data on which they are trained often suffer from a lack of transparency and explainability. For instance, achieving transparency and explainability in black-box models, which are generative ML algorithms where the input and output are known but the functionality is unknown [132], poses significant challenges [133]. Understandably, the unknown nature of the functionality does not encourage trust in the algorithms.

The use of synthetic data can be beneficial since it allows black-box models to be trained on specifically designed data. Training black-box models on high-quality, unbiased data can increase the level of trust in the results of the models [134].

As mentioned previously, a key challenge in generating synthetic data is to evaluate the quality of the data and the accuracy with which it represents the real data. In essence, it is necessary to measure the utility of the data [96] (see Section 3.2). There are a number of XAI methods that allow to measure the extent to which the synthetic data represent real-world scenarios [135]. Auditing methods need to be developed to determine the reliability and representativeness of synthetic data in the energy domain. To achieve this, techniques such as dimensionality reduction [136] or correlation analysis [137] can be used.

To ensure maximum transparency and explainability, it is necessary to carefully consider the methods used for generating synthetic data. The use of generative black-box models to create synthetic datasets can lead to a lack of trust in the generated data, as it is not possible to fully understand and reconstruct how exactly the synthetic data were generated. There are methods and metrics that can be used to evaluate synthetic data generated by black-box models [138].

Another attempt to ensure transparent synthetic energy data is to utilize a human-in-the-loop approach from a data perspective [139]. Involving humans at various stages of data generation can be useful in many different processes, such as data extraction, data integration, data cleaning, data annotation, and iterative labeling [95]. Thus, this approach can also contribute to the development of synthetic data. When using simulations to generate synthetic data, a human-in-the-loop approach is often necessary anyway. This is because the data must be generated by an expert, or at least consulted by one, to achieve high data quality since simulations often require knowledge that a non-expert user does not necessarily possess. However, involving humans in the development process can also introduce risks, such as multiple errors during annotation or data extraction, which can lead to serious consequences for synthetic datasets. These types of risks caused by humans are known as *human bias* [61] (see Section 4.4).

Another essential component to ensure transparency in data science, in general, and synthetic data, in particular, is the provision of metadata [140]. Synthetic data should always be made publicly available in order to achieve maximum transparency. This includes the synthetic data themselves and all information related to the generation of the data, which is also essential for the reproducibility of the generation process (see Section 4.3). The metadata for a synthetic dataset should describe, among other things, what data were used

to generate the dataset, where the original data used came from, how and when the data were collected, what methods and techniques (or ML models) were used to generate the synthetic data and other relevant specifics. It is important to be able to reconstruct where synthetic data originated from and what methods and concepts were used to generate it [141]. Therefore, it is essential to describe exactly how the data are constructed and what data were utilized to generate it. Furthermore, it is critical to explain and publish the data that served as the source for the synthetic data.

Ensuring the quality of this metadata is crucial since this meta information is not useful if it can be misinterpreted or if it is incorrect. Inaccurate or missing meta information can be harmful, but even correct metadata have the potential to cause harm. For example, metadata for synthetic data generation that include information about the ethnicity of the original data can lead to discriminatory behavior [142]. Therefore, the content of metadata should be carefully considered. Since sensitive meta information requires adequate protection, there are techniques available to protect its confidentiality [143].

Besides their beneficial qualities for technical robustness and generalizability (see Section 4.1), synthetic benchmark datasets are a promising approach to ensure both transparency and XAI aspects. Such benchmark datasets allow independent developers of ML algorithms to effectively evaluate and compare the performance of their models on a transparent database [144]. In addition to testing, benchmark datasets allow for the training of ML models. It is particularly important that benchmark datasets maintain a high quality. Any data gaps (see Section 4.1), biases (see Section 4.4), or privacy violations (see Section 4.5) in such datasets would compromise the quality of any ML models trained on them. However, creating high-quality benchmark datasets is typically complex and time-consuming. Thus, it is crucial to make such datasets freely available for public use in order to maximize accessibility to a broad audience [144].

Synthetic benchmark datasets are a relatively new idea with a growing presence across diverse domains, such as geoscience [145], face recognition [146], visual domain adaptation [147], and nighttime dehazing [148]. Nevertheless, benchmark datasets with available ground truth data, referred to as *attribution benchmark datasets*, are still rare [145]. This is also problematic for the development of Trustworthy AI systems in the energy domain since ground truth data are necessary for the evaluation of different XAI methods as well as for the development of diverse ML models such as NILM [74].

As a result, it would be highly desirable to have synthetic benchmark datasets available for the energy data domain as well. For instance, potential datasets could include energy consumption data alongside reliable ground truth data from households or industry, or energy consumption data stemming from solar, wind, or hydropower plants. To avoid the risk of data bias, these synthetic benchmark datasets should be as diverse as possible. This means that different ethnic groups should be represented, as well as different demographic groups (denomination, nationality, gender, age, occupation, income, etc.) and different standards of living. Further research would be needed to provide high-quality synthetic benchmark datasets for different areas of the energy domain, as to our knowledge no work has been conducted in this domain yet. We are only aware of datasets that consist of real data (see Section 3.1).

4.3. Reproducibility

A further key aspect of Trustworthy AI is the *reproducibility* of AI systems, which is closely related to the concepts of transparency and Explainable AI (see Section 4.2).

The purpose of reproducibility is to ensure that scientific work and publications can be replicated by other scientists at any given time. Furthermore, it is desirable that the experiments described are capable of producing the same or comparable outcomes.

Reproducibility can be defined as follows:

Definition 3. “Reproducibility is the ability of independent investigators to draw the same conclusions from an experiment by following the documentation shared by the original investigators” [149].

Although reproducibility is an essential issue in the scientific community, it is unfortunately becoming increasingly difficult to replicate experiments in science in general [150]. In ML research, in particular, it is becoming increasingly difficult to replicate experiments presented and conducted in scientific publications without discrepancies [151,152]. When presenting the results of a trained ML model, there can be a number of reasons for a lack of reproducibility, such as not having access to the training data, not having the code to run the experiments publicly available, or not having conducted a sufficient number of experiments to be able to draw a robust conclusion [151].

The concept of reproducibility can be divided into two categories: the reproducibility of methods and the reproducibility of results [153]. From the perspective of the DCAI approach, the category of methods can include, for example, the reproducibility of data collection and data preprocessing. Reproducibility of results is more likely to be considered on the model side, including, for example, the reproducibility of model settings such as parameters and weights.

Synthetic data are capable of ensuring the reproducibility of data and have already been used for this purpose, for example, by replacing missing or sensitive data with simulated data and then analyzing these data alongside the original data [154]. A variety of techniques are available to reconstruct missing data in time series data. For example, artificial intelligence and multi-source reanalysis data were used to fill gaps in climate data [155], ML methods were used to fill gaps in streamflow data [156] and a multidimensional context autoencoder was used to fill gaps in building energy data [157]. Synthetic data have also been used in other domains to increase reproducibility, such as in biobehavioral sciences [158], in the health data [159], and in synthetic biology [160].

According to [161], an important reproducibility standard is that datasets are transparent and should be published (see Section 4.2). When synthetic datasets are publicly available, researchers are able to ensure that their results are reproducible [158]. This highlights the importance of developing and providing synthetic benchmark datasets, as outlined in Section 4.2. In some cases, however, publication is not possible due to privacy constraints [151]. The use of synthetic data could help in the anonymization of datasets, which will be discussed in more detail later (see Section 4.5).

As already mentioned, also in the energy domain, synthetic data are mainly developed and generated to train and test ML models (see Section 3). Nonetheless, reproducing the outcomes of these ML models can be challenging as they may produce distinct results despite using identical parameters and data for training. This phenomenon is observed for non-intrusive load monitoring models, for example, [17]). Therefore, when measuring the quality of synthetic data using ML models, these inconsistent results of ML models should be explicitly taken into account. One way to address the issue of unstable results of ML models is by running them multiple times and calculating the average of all the results. This approach provides a reasonable level of certainty that the mean value of the results is stable [17]. The number of times experiments must be repeated to achieve a certain level of confidence can be calculated by using Cochran's sample size [162].

4.4. Fairness

The aspect of *fairness* is also a fundamental part of Trustworthy AI and is closely linked to the concepts of transparency and explainable AI.

However, the term 'fairness' of a dataset is not clearly defined. Numerous definitions of fairness exist, including those presented in [163–165]. Nonetheless, it is nearly impossible to simultaneously satisfy all constraints mentioned in the literature [166]. For this survey, we understand fair data to be data that are not biased in any way.

Due to their functionality and characteristics, in general, ML models inherit biases from their training data [167–171]. This means that without fair data, it is very difficult to develop fair ML models. Therefore, research was conducted on the creation of fair data [172–174].

Bias in data can be understood as an unfairness that results from data collection, sampling, and measurement, whereas discrimination can be understood as an unfairness arising from human prejudice and stereotyping based on sensitive attributes, which can occur intentionally or unintentionally [114]. This Section focuses on discussing data fairness. However, other works consider discrimination theory in much greater detail [175–178].

There are numerous types of biases in data that, when used for training ML algorithms, may result in biased algorithmic outcomes. These biases include *measurement bias*, *omitted variable bias*, *representation bias*, *aggregation bias*, *sampling bias*, *longitudinal data fallacy*, and *linking bias* [114]. According to [61], AI bias can be divided into three main categories: *human bias*, *systemic bias* and *statistical/computational bias*. Human bias is a phenomenon that occurs when individuals exhibit systematic errors in their thinking, often stemming from a limited set of heuristic principles. Systemic bias arises from the procedures of certain organizations that have the effect of favoring certain social groups and disfavoring others. Statistical and computational biases are biases that occur when a data sample is not a reasonable representation of the population as a whole.

Many instances exist wherein biased systems have been evaluated for their ability to discriminate against specific populations and subgroups, such as facial recognition [179] and recommender systems [180]. There are numerous instances of data biases, including datasets like ImageNet [181] and Open Images [182], which are used in the majority of studies in this field and consequently exhibit representation bias [183]. Additionally, there exist facial datasets like IJB-A [184] and Adience [185], which lack balance in terms of race, resulting in systemic bias [171].

When creating synthetic data based on real data, there is of course the risk that biases of the real data are unintentionally transferred to the synthetic data [47,169]. For example, if a real-world energy dataset consists only of European energy data, the synthetic households will also reflect the characteristics of European households. This is referred to as the *out-of-distribution* (OOD) generalization problem [186], a well-known challenge when working with synthetic data [187]. The OOD problem describes a situation where the data distribution of the test dataset is not identical to the data distribution of the training dataset when developing ML models. Synthetic data allow augmenting data, thus, reducing the OOD problem [187].

The problem of unfair datasets is known in the literature, and there are already existing approaches for generating high-quality, fair synthetic data from ‘unfair’ source data [169,188,189]. Methods like the one described in [169] achieve fairness in synthetic data by removing edges between features. This approach can be applied to time series data, but adapting it to image data is challenging.

When generating synthetic data, it is crucial to prevent data bias caused by the absence of underprivileged groups in simulator development. It is also essential to avoid performance degradation when an AI model trained on synthetic data is applied to real-world data. Synthetic datasets were already used in different domains to reduce bias in datasets, including face recognition [190], robotics [191] and healthcare [43].

The design of the data, in general, and thus synthetic data, is crucial for minimizing bias and increasing trustworthiness. There are several ways to avoid data bias and ensure the generation of a fair dataset. These methods involve providing comprehensive documentation of metadata for dataset creation, including the techniques used to produce the dataset, its motivations, and its characteristics [141,192]. This metadata should also include the information about where the data originate and how they were collected [193]. There is also the idea of using labels to better categorize datasets [193] or using methods to detect statistical biases such as Simpson’s paradox [194,195]. To eliminate bias in a dataset, also various preprocessing techniques can be used, including suppression, massaging the dataset, reweighing, and sampling [196].

In general, it is challenging to identify biases in synthetic data during post-processing. Therefore, efforts should be made during pre-processing to ensure that synthetic data are not generated with bias. This can be achieved, for example, by ensuring that the underlying

data are unbiased and that the methods for data generation are also unbiased. Nevertheless, biases in synthetic data can also be addressed, for instance, through human checkers who oversee the data generation process [197], or with the assistance of data augmentation techniques [54,198]. There are also approaches that measure the fairness of synthetic data, such as using a two covariate-level disparity metric [199].

4.5. Privacy

Strictly speaking, *privacy* is not a purely technical aspect of Trustworthy AI. However, technical methods and concepts, particularly on the data side, can substantially contribute to privacy. Hence, we also included this aspect in this survey.

Developing robust and effective AI requires large amounts of data [33], which is not always straightforward to obtain due to data protection regulations. The majority of datasets contain people's personal information, which is justifiably strictly protected in many countries. Furthermore, the collection of real-world data is generally under strict protection. As exemplified by the collection of facial datasets, privacy concerns can be incredibly complex [200]. Privacy regulations pose a challenge due to differing implementations across countries, necessitating specialized legal expertise for any privacy assessment. In academic research, it is recognized that there exists a known issue regarding the possible disclosure of attributes in datasets. To address this concern, concepts and approaches have already been proposed, such as described in [201,202].

Synthetic data have the potential to provide a solution to the privacy problem described. Nonetheless, it is a widespread misconception to assume that synthetic data always satisfy privacy regulations [54]. If synthetic data are derived from real-world data, they may disclose information about the original data that underlie it, potentially due to comparable distributions, outliers, and low-probability incidents. As a result, producing synthetic data that ensures privacy requires considerable effort.

In general, two objectives can be distinguished for synthetic data and privacy: generating synthetic data to enhance privacy and ensuring privacy in synthetic data.

Privacy in synthetic data can be achieved by using certain techniques, such as data anonymization [203] or data concealment [201]. By synthesizing data, data anonymization can be achieved by removing or anonymizing personal information from the original real-world data to protect privacy [204]. When generating synthetic data, they can often be useful to hide certain information in the data to protect sensitive or confidential data.

According to [205] there are privacy-enhancing technologies exist that allow for legally compliant processing and analysis of personal data, such as *federated learning* (FL) [206] and *differential privacy* (DP) [207].

The concept of FL was originally designed and developed to enable the training of ML algorithms while adhering to privacy regulations [208]. FL attempts to protect the security and privacy of local raw training data by maintaining it at its source or storage location, without ever transferring it to a central server [209].

In the energy domain, this means that one way to comply with privacy regulations would be to create synthetic energy data at the edge, such as in a private household or in an industrial building itself. In order to be able to create synthetic data on the edge, the framework used to create it must run on the edge as well. Moreover, the framework should have the capability to store and save the synthetic data within the edge environment. However, not only would the synthetic data need to remain in this environment, but also the ML models that are trained or refined on this synthetic data. ML models such as Non-Intrusive Load Monitoring disaggregation [74], energy forecasting [210] or human activity recognition models [211] would then be trained and fine-tuned by using FL concepts [212–214]. As a result, the original real-world data, the synthetic data themselves, as well as the ML models used can remain within the edge environment.

However, there are potential risks associated with utilizing FL systems, as the trained ML models may become vulnerable to attacks if exported from the edge environment [215]. Although this does not allow access to the actual model data, it is still possible to obtain the

parameters and weights of the trained ML models [216]. To improve security, FL can be combined with other privacy-enhancing technologies such as differential privacy [217].

Differential Privacy (DP) is a mathematical concept of ensuring privacy by adding noise to data in order to protect personally identifiable information [202]. Increasing the amount of noise added makes it increasingly difficult to recognize the original data, resulting in a greater protection of privacy. The concept of DP has been well-established and applied to the concept of FL [218–220]. DP has also been utilized for various use cases involving synthetic data [221–224].

A major challenge in ensuring privacy is the reliable evaluation of whether the synthesized data are sufficiently anonymous after implementing concepts such as DP, i.e., whether personal data can be derived. Controversial opinions exist in the literature regarding this matter.

According to [225], there are no robust and objective methods to determine whether a synthetic dataset appears sufficiently different from its real-world counterpart to be classified as an anonymous dataset.

Despite this opinion, there are also studies that propose criteria to determine the quality of synthetic data in terms of privacy. According to [226], there are different criteria to measure the quality of the synthetic data in terms of privacy, including the *exact match score*, the *neighbors' privacy score*, and the *membership inference score*.

The *exact match score* indicates whether the synthetic data contain any copies of the real-world data [227]. A score of zero implies that there are no duplicates of the authentic data in the synthetic data. However, this score is problematic when synthetic data are generated based on real-world data. For example, if real energy data from freely available datasets are used to create synthetic data, the *exact match score* will be very high due to the (intended) copies of the real-world data within the synthetic data. However, if the real-world data are anonymized, the synthetic data will also be anonymous, even if the *exact match score* is high.

Related to the previous score is the *neighbors' privacy score*, which measures whether there are similarities between the synthetic data and the real data. Although these are not direct copies, they are potential indicators of information disclosure. For generating synthetic energy data based on real-world data, the *neighbors' privacy score* potentially encounters similar issues as those previously addressed in relation to the *exact match score*.

A *membership inference attack* aims to uncover the data used for generating synthetic data, even without the attackers having access to the original data [228–230]. The *membership inference score* represents the probability that such an attack will be successful. A high score indicates that it is unlikely that a particular dataset was used to generate synthetic data. Conversely, a low score indicates that it is likely that a particular dataset was used to generate synthetic data. If a dataset is identified by such an attack, private information could be exposed. Despite this, freely available energy datasets have been usually protected by omitting any direct personal information [231–234].

Before synthetic data can be widely utilized in the energy domain, it is necessary to carefully consider all relevant privacy concerns that have been previously discussed. Moreover, it is important to understand the relative positioning of the produced synthetic data with respect to the original data in relation to the information it potentially reveals.

When generating synthetic energy data using real-world data, the energy consumption data of machines or devices can be included. Depending on the extent to which synthetic data permits inferences to be drawn from real data, there are privacy regulations that must be followed. If the energy consumption patterns of machines cannot be linked to humans, it is likely that data protection regulations are met. On the other hand, if it is possible to identify and reconstruct when the machines were activated, it is possible to draw conclusions about human behavior from the real data. For instance, it is feasible to ascertain how an individual behaves at their home based on their usage of electronic devices in their personal space, or even to determine if an individual is at home at all.

However, human behavior is protected by a much higher level of data privacy, which is very strict and well-protected by data protection regulations in most countries. For instance, as stated in Article 4 (1) of the EU's General Data Protection Regulation (GDPR) [235], *personal data* is understood as "any information relating to an identified or identifiable natural person".

When generating synthetic energy data, it is crucial to ensure that no data protection regulations are violated. This involves preventing the derivation of human behavior that could be traced back to any specific individual based on the synthetic data. If synthetic energy data are generated independently of the behavior of a real individual, and cannot be traced back to a specific person, then they generally do not violate any privacy rights. For instance, to prevent the disclosure of human behavior in synthetic data, it can be anonymized or randomized in a way that maximizes the difficulty of tracing it back to the original real data. Furthermore, the behavior of simulated individuals designed to generate data for critical situations that do not occur in the real world can be entirely custom-built and not based on real behavior. Therefore, the behavior of a real individual cannot be disclosed.

The use of synthetic data in combination with the strong privacy protection of the underlying original data allows a balance between transparency (see Section 4.2), data protection and research objectives [236].

4.6. Sustainability

The concept of *sustainability* can be considered as a technology part of AI, including the methods to train AI and the actual processing of data by AI [237].

The field of AI sustainability can be divided into two categories: on the one hand, AI methods and concepts that aim to reduce energy consumption and emissions, and on the other hand, the development of environmentally friendly AI itself [237]. This survey focuses on the second category of sustainability, hereafter referred to as the concept of *sustainability of AI*.

Definition 4. "*Sustainability of AI is focused on sustainable data sources, power supplies, and infrastructures as a way of measuring and reducing the carbon footprint from training and/or tuning an algorithm.*" [237].

AI systems generally require a lot of computing resources over a long period of time to achieve robustness and provide valuable outcomes (see Section 4.1). Large models, such as natural language processing models, are dependent on large-scale data centers, which consume vast amounts of energy and resources and thus emit considerable amounts of CO₂ [238]. For instance, training the complex architecture of the ChatGPT model requires a considerable amount of computing resources like GPUs over a period of months [239].

The purpose of the concept of sustainability of AI is to highlight these previously mentioned issues and to ensure sustainable and environmentally friendly AI development. With the transition towards sustainable energy and the growing scarcity of resources, it has become crucial to focus on reducing energy and computational resource usage. Therefore, there is a necessity to develop model architectures and training techniques that are more energy-efficient [239]. The carbon footprint of developing and training ML models should be in a healthy proportion to its benefits.

In order to develop more environmentally friendly ML models, understanding the amount of energy, resources, and CO₂ emissions consumed in the model development process is crucial. This also covers the emissions of the server during the model training, including the energy consumption of the hardware and the energy grid of the server [240]. There are methods and frameworks available for tracking emissions [240,241]. To achieve sustainability of AI, researchers and developers should publish the energy consumption and carbon footprint data of their ML models [242]. This would enable other researchers and developers to compare the energy usage of their models, which would encourage

healthy competition and also be an important contribution to the transparency aspect of Trustworthy AI (see Section 4.2).

There are already some concepts that address how AI systems, and computer systems, in general, can become more environmentally friendly and thus ultimately consume fewer resources. These include green AI [243], cloud computing [244], and power-aware computing [245].

Synthetic data can contribute substantially to the advancement of sustainable development and sustainability itself by reducing the need for data collection in the real world, which can cause numerous problems. Especially in the energy domain, data collection is not only time-consuming but also resource-intensive (see Section 3.1). This is due to the fact that acquiring data in this domain primarily involves the utilization of hardware such as sensors, which consume energy themselves [246] as well as requiring considerable amounts of cost and resources for their production [247,248].

If synthetic data are sufficiently well designed and generated, many of the time- and resource-consuming steps involved in preparing and preprocessing real-world data can be eliminated. These include various steps such as filling gaps in data, annotating data with human assistance, or debiasing data (see Section 4.4).

Following the arguments presented throughout this survey, synthetic data, if generated appropriately, offer the potential to develop more robust and effective ML models. Consequently, this could potentially lead to process improvements, ultimately resulting in energy and resource efficiencies in the long term. This includes ML models focused on improving energy efficiency and sustainability. For instance, there are models designed to enhance sustainability in the food [249] and smart cities domains [250], as well as energy prediction models for buildings [251].

However, the development of simulations and thus the generation of synthetic data is initially associated with development efforts and costs energy and resources [252–254]. The greater the amount of data generated via an established simulation framework, the greater the benefit compared to real data. However, this is only valid if the synthetic data prove to be sufficiently useful.

Therefore, it is crucial to ensure that generating synthetic data consumes less energy than collecting data in the real world. This applies to the entire process, including both the development of the simulation framework used to generate the data and the generation of the synthetic data themselves. If it can be ensured that generating synthetic data consumes less energy than collecting real-world data, then synthetic data will be a powerful cornerstone to improving the sustainability and environmental friendliness of AI systems.

5. Discussion: Open Issues and Further Directions

This survey focused on analyzing how synthetic data can contribute to accelerating the development of efficient, reliable, and sustainable aspects of *Trustworthy AI* in the energy domain. To address the trend of using synthetic data for training ML models, we introduced the term *Synthetic Data-Centric AI* (SDCAI) as an extension of the DCAI approach. Further, we analyzed different aspects of Trustworthy AI, selected technical factors, and considered them from the perspective of the SDCAI approach. We examined the potential(s), opportunities and risks of synthetic data in the energy domain for each of the technical aspects of Trustworthy AI mentioned in more detail. Although this work focuses on the energy domain, we are convinced that many of the results are transferable to other domains due to their characteristics.

Altogether, we identified a total of eight technical factors in the areas of efficiency, reliability and sustainability that should be satisfied in order for data and the resulting ML models trained on those data to be classified as trustworthy: *technical robustness and generalization* (see Section 4.1), *transparency and explainability* (see Section 4.2), *reproducibility* (see Section 4.3), *fairness* (see Section 4.4), *privacy* (see Section 4.5) and *sustainability* (see Section 4.6). Table 1 summarizes the literature and the main potentials of synthetic data identified in this review to improve the considered technical aspects of Trustworthy AI.

Table 1. Overview of the literature and the main potentials of synthetic data identified in this review to improve the considered technical aspects of Trustworthy AI.

Technical Aspect	Literature	Potentials by Using Synthetic Data
Technical Robustness and Generalization	[54,72,109–128]	Data of critical and unusual situations Divers training and testing datasets Synthetic Benchmark datasets
Transparency and Explainability	[55,61,74,95,96,129–148]	Training Black Box Models Provision of Metadata Synthetic Benchmark datasets
Reproducibility	[17,149–162]	Replacing missing data Synthetic Benchmark datasets
Fairness	[43,47,54,61,114,141,163–199]	Data augmentation Fair Data design
Privacy	[33,54,74,200–236]	Federated Learning Differential Privacy Data anonymization and randomization
Sustainability	[237–254]	Reducing Real-World Data Collection Reducing real-world Data Preparing and Preprocessing

Considering the results of our analysis, a general and essential feature of synthetic data is their configurable nature and the control afforded by the design-to-generation process, which distinguishes them from real-world data that are only collectible. The circumstances under which real-world data are recorded are often not transparent and reproducible in retrospect. In contrast, due to the controllability of the design-to-generation process, the production of synthetic data is reproducible in structure, repeatable, and widely accessible. This is crucial as synthetic data that are generated properly are not only technically accurate but also transparent and reliable, ensuring key attributes such as correct annotation and labeling of data.

Generating synthetic data allows ML models to be trained and tested on theoretically any amount of data with high variability, resulting in increased technical robustness. In particular, carefully designed synthetic data enable the generation of critical situations that do not occur in real-world data but are essential for achieving a high generalization capability. If generated properly, synthetic data can be used to improve the performance of ML models in the energy domain, thus increasing the technical robustness as well as the generalizability [17].

Additionally, synthetic data enable the development of transparent and more explainable black-box models. The reason for this is that properly designed synthetic datasets have a high level of transparency because it is possible to understand exactly what methods were used to generate them. In general, more transparent and explainable data can increase trust in the ML models trained on these data [134], thus making them more reliable. Synthetic benchmark datasets offer improved transparency by providing completely labeled and annotated data without issues such as data gaps or differing frequencies. Specifically, synthetic benchmark datasets allow for improved ML model reproducibility as transparent and publishable data are a necessary prerequisite for reproducibility [161]. As a consequence, it would be highly desirable to have synthetic benchmark datasets available for the energy data domain as well.

Fairness can be improved through the use of synthetic data and thoughtful data design, which allows avoiding the biases that are often present in real-world datasets. One approach to conducting this is to design data that are representative of a wide range of ethnic and demographic groups. Moreover, synthetic data can also be used to protect privacy by removing or anonymizing personal information from real datasets [204].

Synthetic data allow for minimizing data collection in the physical world, as fewer sensors are required to collect data. Moreover, synthetic data can foster the development of more resilient ML models that are designed to optimize processes to advance sustainability.

Due to the objective of this survey, the aspects of Trustworthy AI considered have a strong technical perspective and are not exhaustive. There are additional non-technical aspects of Trustworthy AI that were not addressed in this survey due to their diminished relevance to synthetic data. Moreover, synthetic data may only be partially effective in addressing non-technical aspects of Trustworthy AI, if effective at all.

The development of methods and concepts for generating trustworthy synthetic data is an ongoing process, and further research is needed to fully exploit the potential of this emerging technology. This survey has identified several open questions that need to be addressed in the future in order to realize the full potential of synthetic data for accelerating the development of Trustworthy AI. These questions involve determining a reliable method of measuring the utility of synthetic data, as well as how to provide a reliable measure that synthetic data are not biased in any way. An additional question that requires further investigation is the optimal balance between real and synthetic data when augmenting data in order to achieve the best possible results. Furthermore, it is necessary to determine which real data should be available in order to generate useful synthetic data. This applies both to the energy domain and across domains. Another question that requires clarification is the extent to which the training and testing of ML models benefit from the inclusion of critical situations in a dataset and whether it is possible to quantify this benefit. Furthermore, it is essential to define the minimum number of critical situations that should be present in a dataset and to determine their composition. With additional research, it may be possible to establish generalizable and cross-domain principles for generating critical situations.

Irrespective of the domain, the decision to use synthetic data for training ML models should be made with caution, as it takes a considerable amount of development time to ensure that the data are useful and can adequately represent a domain. It is essential that synthetic data are of a certain quality and properly designed to achieve Trustworthy AI.

However, once this initial development time has been invested, synthetic data, if generated and used correctly, has great potential to substantially increase the level of trust in AI in any of the considered technical factors. Trustworthy AI principles and methods are capable of improving the quality of synthetic data, so they should be a key consideration when generating synthetic data. As this survey showed when of sufficient quality, synthetic data allow an increase in the level of trust in the *technical robustness, generalizability, transparency, explainability, reproducibility, fairness, privacy, and sustainability* of AI applications in the energy domain.

Author Contributions: Conceptualization, M.M. and I.Z.; investigation, M.M. and I.Z.; writing—original draft preparation, M.M. and I.Z.; writing—review and editing, M.M. and I.Z.; visualization, M.M. and I.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) as part of the ForeSightNEXT project and by the German Federal Ministry of Education and Research (BMBF) as part of the ENGAGE project.

Data Availability Statement: No new data were created in this review. Data sharing is not applicable to this review.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chu, S.; Majumdar, A. Opportunities and challenges for a sustainable energy future. *Nature* **2012**, *488*, 294–303. [[CrossRef](#)] [[PubMed](#)]
2. Steg, L.; Perlaviciute, G.; Van der Werff, E. Understanding the human dimensions of a sustainable energy transition. *Front. Psychol.* **2015**, *6*, 805. [[CrossRef](#)]
3. Dominković, D.F.; Bačeković, I.; Pedersen, A.S.; Krajačić, G. The future of transportation in sustainable energy systems: Opportunities and barriers in a clean energy transition. *Renew. Sustain. Energy Rev.* **2018**, *82*, 1823–1838. [[CrossRef](#)]

4. Khalid, A.M.; Mitra, I.; Warmuth, W.; Schacht, V. Performance ratio—Crucial parameter for grid connected PV plants. *Renew. Sustain. Energy Rev.* **2016**, *65*, 1139–1158. [[CrossRef](#)]
5. Višković, A.; Franki, V.; Jevtić, D. Artificial intelligence as a facilitator of the energy transition. In Proceedings of the 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 23–27 May 2022; pp. 494–499.
6. Griffiths, S. Energy diplomacy in a time of energy transition. *Energy Strategy Rev.* **2019**, *26*, 100386. [[CrossRef](#)]
7. Jimenez, V.M.M.; Gonzalez, E.P. The Role of Artificial Intelligence in Latin Americas Energy Transition. *IEEE Lat. Am. Trans.* **2022**, *20*, 2404–2412. [[CrossRef](#)]
8. Sulaiman, A.; Nagu, B.; Kaur, G.; Karuppaiah, P.; Alshahrani, H.; Reshan, M.S.A.; AlYami, S.; Shaikh, A. Artificial Intelligence-Based Secured Power Grid Protocol for Smart City. *Sensors* **2023**, *23*, 8016. [[CrossRef](#)]
9. Chehri, A.; Fofana, I.; Yang, X. Security risk modeling in smart grid critical infrastructures in the era of big data and artificial intelligence. *Sustainability* **2021**, *13*, 3196. [[CrossRef](#)]
10. Xie, J.; Alvarez-Fernandez, I.; Sun, W. A review of machine learning applications in power system resilience. In Proceedings of the 2020 IEEE Power & Energy Society General Meeting (PESGM), Montreal, QC, Canada, 2–6 August 2020; pp. 1–5.
11. Shi, Z.; Yao, W.; Li, Z.; Zeng, L.; Zhao, Y.; Zhang, R.; Tang, Y.; Wen, J. Artificial intelligence techniques for stability analysis and control in smart grids: Methodologies, applications, challenges and future directions. *Appl. Energy* **2020**, *278*, 115733. [[CrossRef](#)]
12. Omitaomu, O.A.; Niu, H. Artificial intelligence techniques in smart grid: A survey. *Smart Cities* **2021**, *4*, 548–568. [[CrossRef](#)]
13. Song, Y.; Wan, C.; Hu, X.; Qin, H.; Lao, K. Resilient power grid for smart city. *iEnergy* **2022**, *1*, 325–340. [[CrossRef](#)]
14. Massaoudi, M.; Abu-Rub, H.; Refaat, S.S.; Chihi, I.; Oueslati, F.S. Deep learning in smart grid technology: A review of recent advancements and future prospects. *IEEE Access* **2021**, *9*, 54558–54578. [[CrossRef](#)]
15. Bose, B.K. Artificial intelligence techniques in smart grid and renewable energy systems—Some example applications. *Proc. IEEE* **2017**, *105*, 2262–2273. [[CrossRef](#)]
16. Tang, Y.; Huang, Y.; Wang, H.; Wang, C.; Guo, Q.; Yao, W. Framework for artificial intelligence analysis in large-scale power grids based on digital simulation. *CSEE J. Power Energy Syst.* **2018**, *4*, 459–468. [[CrossRef](#)]
17. Meiser, M.; Duppe, B.; Zinnikus, I. Generation of meaningful synthetic sensor data—Evaluated with a reliable transferability methodology. *Energy AI* **2024**, *15*, 100308. [[CrossRef](#)]
18. Jin, D.; Ocone, R.; Jiao, K.; Xuan, J. Energy and AI. *Energy AI* **2020**, *1*, 100002. [[CrossRef](#)]
19. Tomazzoli, C.; Scannapieco, S.; Cristani, M. Internet of things and artificial intelligence enable energy efficiency. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 4933–4954. [[CrossRef](#)]
20. Aguilar, J.; Garces-Jimenez, A.; R-moreno, M.; García, R. A systematic literature review on the use of artificial intelligence in energy self-management in smart buildings. *Renew. Sustain. Energy Rev.* **2021**, *151*, 111530. [[CrossRef](#)]
21. Yu, K.H.; Beam, A.L.; Kohane, I.S. Artificial intelligence in healthcare. *Nat. Biomed. Eng.* **2018**, *2*, 719–731. [[CrossRef](#)]
22. Panch, T.; Mattie, H.; Celi, L.A. The “inconvenient truth” about AI in healthcare. *NPJ Digit. Med.* **2019**, *2*, 77. [[CrossRef](#)]
23. Cao, L. AI in finance: Challenges, techniques, and opportunities. *ACM Comput. Surv. (CSUR)* **2022**, *55*, 1–38.
24. Buchanan, B.G. *Artificial Intelligence in Finance*; The Alan Turing Institute: London, UK, 2019.
25. Hilpisch, Y. *Artificial Intelligence in Finance*; O’Reilly Media: Sebastopol, CA, USA, 2020.
26. Castelvechi, D. Can we open the black box of AI? *Nat. News* **2016**, *538*, 20. [[CrossRef](#)]
27. Kaur, D.; Uslu, S.; Rittichier, K.J.; Durresi, A. Trustworthy artificial intelligence: A review. *ACM Comput. Surv. (CSUR)* **2022**, *55*, 1–38. [[CrossRef](#)]
28. Thiebes, S.; Lins, S.; Sunyaev, A. Trustworthy artificial intelligence. *Electron. Mark.* **2021**, *31*, 447–464. [[CrossRef](#)]
29. Floridi, L. Establishing the rules for building trustworthy AI. In *Ethics, Governance, and Policies in Artificial Intelligence*; Springer: Cham, Switzerland, 2021; pp. 41–45.
30. Hamid, O.H. From model-centric to data-centric AI: A paradigm shift or rather a complementary approach? In Proceedings of the 2022 8th International Conference on Information Technology Trends (ITT), Dubai, United Arab Emirates, 25–26 May 2022; pp. 196–199.
31. Zha, D.; Bhat, Z.P.; Lai, K.H.; Yang, F.; Hu, X. Data-centric AI: Perspectives and challenges. In Proceedings of the 2023 SIAM International Conference on Data Mining (SDM), Minneapolis, MN, USA, 27–29 April 2023; pp. 945–948.
32. Sambasivan, N.; Kapania, S.; Highfill, H.; Akrong, D.; Paritosh, P.; Aroyo, L.M. “Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; pp. 1–15.
33. Roh, Y.; Heo, G.; Whang, S.E. A survey on data collection for machine learning: A big data-ai integration perspective. *IEEE Trans. Knowl. Data Eng.* **2019**, *33*, 1328–1347. [[CrossRef](#)]
34. Taori, R.; Dave, A.; Shankar, V.; Carlini, N.; Recht, B.; Schmidt, L. Measuring robustness to natural distribution shifts in image classification. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 18583–18599.
35. Whang, S.E.; Roh, Y.; Song, H.; Lee, J.G. Data collection and quality challenges in deep learning: A data-centric ai perspective. *VLDB J.* **2023**, *32*, 791–813. [[CrossRef](#)]
36. Najeh, H.; Singh, M.P.; Ploix, S.; Chabir, K.; Abdelkrim, M.N. Automatic thresholding for sensor data gap detection using statistical approach. In *Sustainability in Energy and Buildings: Proceedings of SEB 2019*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 455–467.

37. Klemenjak, C.; Reinhardt, A.; Pereira, L.; Makonin, S.; Bergés, M.; Elmenreich, W. Electricity consumption data sets: Pitfalls and opportunities. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA, 13–14 November 2019; pp. 159–162.
38. Ma, B.; Zheng, X. Biased data revisions: Unintended consequences of China’s energy-saving mandates. *China Econ. Rev.* **2018**, *48*, 102–113. [[CrossRef](#)]
39. de Vos, A.; Preiser, R.; Masterson, V.A. Participatory data collection. In *The Routledge Handbook of Research Methods for Social-Ecological Systems*; Taylor & Francis: Abingdon, UK, 2021; p. 119.
40. Xu, Y.; Maitland, C. Participatory data collection and management in low-resource contexts: A field trial with urban refugees. In Proceedings of the Tenth International Conference on Information and Communication Technologies and Development, Ahmedabad, India, 4–7 January 2019; pp. 1–12.
41. Shilton, K. Participatory personal data: An emerging research challenge for the information sciences. *J. Am. Soc. Inf. Sci. Technol.* **2012**, *63*, 1905–1915. [[CrossRef](#)]
42. Marwala, T.; Fournier-Tombs, E.; Stinckwich, S. The Use of Synthetic Data to Train AI Models: Opportunities and Risks for Sustainable Development. *arXiv* **2023**, arXiv:2309.00652.
43. Nikolenko, S.I. *Synthetic Data for Deep Learning*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 174.
44. Zhang, C.; Kuppannagari, S.R.; Kannan, R.; Prasanna, V.K. Generative adversarial network for synthetic time series data generation in smart grids. In Proceedings of the 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Aalborg, Denmark, 29–31 October 2018; pp. 1–6.
45. Klemenjak, C.; Kovatsch, C.; Herold, M.; Elmenreich, W. A synthetic energy dataset for non-intrusive load monitoring in households. *Sci. Data* **2020**, *7*, 108. [[CrossRef](#)]
46. Reddy, T.; Claridge, D. Using synthetic data to evaluate multiple regression and principal component analyses for statistical modeling of daily building energy consumption. *Energy Build.* **1994**, *21*, 35–44. [[CrossRef](#)]
47. Giuffrè, M.; Shung, D.L. Harnessing the power of synthetic data in healthcare: Innovation, application, and privacy. *NPJ Digit. Med.* **2023**, *6*, 186. [[CrossRef](#)]
48. Benaim, A.R.; Almog, R.; Gorelik, Y.; Hochberg, I.; Nassar, L.; Mashiach, T.; Khamaisi, M.; Lurie, Y.; Azzam, Z.S.; Khoury, J.; et al. Analyzing medical research results based on synthetic data and their relation to real data results: Systematic comparison from five observational studies. *JMIR Med. Inform.* **2020**, *8*, e16492. [[CrossRef](#)]
49. Ive, J.; Viani, N.; Kam, J.; Yin, L.; Verma, S.; Puntis, S.; Cardinal, R.N.; Roberts, A.; Stewart, R.; Velupillai, S. Generation and evaluation of artificial mental health records for natural language processing. *NPJ Digit. Med.* **2020**, *3*, 69. [[CrossRef](#)]
50. Assefa, S.A.; Dervovic, D.; Mahfouz, M.; Tillman, R.E.; Reddy, P.; Veloso, M. Generating synthetic data in finance: Opportunities, challenges and pitfalls. In Proceedings of the First ACM International Conference on AI in Finance, New York, NY, USA, 15–16 October 2020; pp. 1–8.
51. Da Silva, B.; Shi, S.S. Style transfer with time series: Generating synthetic financial data. *arXiv* **2019**, arXiv:1906.03232.
52. Papacharalampopoulos, A.; Tzimanis, K.; Sabatakakis, K.; Stavropoulos, P. Deep quality assessment of a solar reflector based on synthetic data: Detecting surficial defects from manufacturing and use phase. *Sensors* **2020**, *20*, 5481. [[CrossRef](#)]
53. Manettas, C.; Nikolakis, N.; Alexopoulos, K. Synthetic datasets for Deep Learning in computer-vision assisted tasks in manufacturing. *Procedia CIRP* **2021**, *103*, 237–242. [[CrossRef](#)]
54. Jordon, J.; Szpruch, L.; Houssiau, F.; Bottarelli, M.; Cherubin, G.; Maple, C.; Cohen, S.N.; Weller, A. Synthetic Data—what, why and how? *arXiv* **2022**, arXiv:2205.03257.
55. Ala-Pietilä, P.; Bonnet, Y.; Bergmann, U.; Bielikova, M.; Bonefeld-Dahl, C.; Bauer, W.; Bouarfa, L.; Chatila, R.; Coeckelbergh, M.; Dignum, V.; et al. *The Assessment List for Trustworthy Artificial Intelligence (ALTAI)*; European Commission: Luxembourg, 2020.
56. TAILOR EU Project. The TAILOR Handbook of Trustworthy AI. 2022. Available online: <http://tailor.isti.cnr.it/handbookTAI/TAILOR.html#id1> (accessed on 15 April 2024).
57. Yeung, K. Recommendation of the Council on Artificial Intelligence (OECD). *Int. Leg. Mater.* **2020**, *59*, 27–34. [[CrossRef](#)]
58. The White House, Guidance for Regulation of Artificial Intelligence Applications. In Memorandum for the Heads of Executive Departments and Agencies. 2020. Available online: <https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf> (accessed on 15 April 2024).
59. National Institute of Standards and Technology, U.S. Department of Commerce. AI Risks and Trustworthiness. Available online: https://airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF/Foundational_Information/3-sec-characteristics (accessed on 15 April 2024).
60. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework. 2023. Available online: <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf> (accessed on 15 April 2024).
61. Schwartz, R.; Vassilev, A.; Greene, K.; Perine, L.; Burt, A.; Hall, P. *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*; NIST Special Publication; US Department of Commerce, National Institute of Standards and Technology: Washington, DC, USA, 2022; Volume 1270.
62. Bundesamt für Sicherheit in der Informationstechnik. *AI Cloud Service Compliance Criteria Catalogue (AIC4)*; Federal Office for Information Security: Bonn, Germany, 2021. Available online: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/CloudComputing/AIC4/AI-Cloud-Service-Compliance-Criteria-Catalogue_AIC4.html (accessed on 15 April 2024).

63. Liang, W.; Tadesse, G.A.; Ho, D.; Fei-Fei, L.; Zaharia, M.; Zhang, C.; Zou, J. Advances, challenges and opportunities in creating data for trustworthy AI. *Nat. Mach. Intell.* **2022**, *4*, 669–677. [CrossRef]
64. Harrison, R.L. Introduction to monte carlo simulation. In *AIP Conference Proceedings*; American Institute of Physics: College Park, MD, USA, 2010; Volume 1204, pp. 17–21.
65. Rahane, W.; Dalvi, H.; Magar, Y.; Kalane, A.; Jondhale, S. Lung cancer detection using image processing and machine learning healthcare. In Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, India, 1–3 March 2018; pp. 1–5.
66. Qayyum, A.; Qadir, J.; Bilal, M.; Al-Fuqaha, A. Secure and robust machine learning for healthcare: A survey. *IEEE Rev. Biomed. Eng.* **2020**, *14*, 156–180. [CrossRef]
67. Shi, J.; Guo, J.; Zheng, S. Evaluation of hybrid forecasting approaches for wind speed and power generation time series. *Renew. Sustain. Energy Rev.* **2012**, *16*, 3471–3480. [CrossRef]
68. Sharadga, H.; Hajimirza, S.; Balog, R.S. Time series forecasting of solar power generation for large-scale photovoltaic plants. *Renew. Energy* **2020**, *150*, 797–807. [CrossRef]
69. Hossain, M.S.; Mahmood, H. Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *IEEE Access* **2020**, *8*, 172524–172533. [CrossRef]
70. Yoon, J.; Jarrett, D.; Van der Schaar, M. Time-series generative adversarial networks. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
71. Ribeiro, M.H.D.M.; da Silva, R.G.; Moreno, S.R.; Mariani, V.C.; dos Santos Coelho, L. Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. *Int. J. Electr. Power Energy Syst.* **2022**, *136*, 107712. [CrossRef]
72. Li, B.; Qi, P.; Liu, B.; Di, S.; Liu, J.; Pei, J.; Yi, J.; Zhou, B. Trustworthy AI: From principles to practices. *ACM Comput. Surv.* **2023**, *55*, 1–46. [CrossRef]
73. Minh, D.; Wang, H.X.; Li, Y.F.; Nguyen, T.N. Explainable artificial intelligence: A comprehensive review. *Artif. Intell. Rev.* **2022**, *55*, 3503–3568. [CrossRef]
74. Kaselimi, M.; Protopapadakis, E.; Voulodimos, A.; Doulamis, N.; Doulamis, A. Towards trustworthy energy disaggregation: A review of challenges, methods, and perspectives for non-intrusive load monitoring. *Sensors* **2022**, *22*, 5872. [CrossRef]
75. Firth, S.; Kane, T.; Dimitriou, V.; Hassan, T.; Fouchal, F.; Coleman, M.; Webb, L. REFIT Smart Home Dataset. 2017. Available online: https://repository.lboro.ac.uk/articles/dataset/REFIT_Smart_Home_dataset/2070091/1 (accessed on 15 April 2024).
76. Wilhelm, S.; Jakob, D.; Kasbauer, J.; Ahrens, D. GeLaP: German labeled dataset for power consumption. In Proceedings of the Sixth International Congress on Information and Communication Technology: ICICT 2021, London, UK, 25–26 February 2021; Springer: Singapore, 2022; Volume 1, pp. 21–33.
77. Shin, C.; Lee, E.; Han, J.; Yim, J.; Rhee, W.; Lee, H. The ENERTALK dataset, 15 Hz electricity consumption data from 22 houses in Korea. *Sci. Data* **2019**, *6*, 193. [CrossRef] [PubMed]
78. Monacchi, A.; Egarter, D.; Elmenreich, W.; D’Alessandro, S.; Tonello, A.M. GREEND: An energy consumption dataset of households in Italy and Austria. In Proceedings of the 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), Venice, Italy, 3–6 November 2014; pp. 511–516.
79. Chavan, D.R.; More, D.S.; Khot, A.M. IEDL: Indian Energy Dataset with Low frequency for NILM. *Energy Rep.* **2022**, *8*, 701–709. [CrossRef]
80. Kelly, J.; Knottenbelt, W. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Sci. Data* **2015**, *2*, 150007. [CrossRef]
81. Schlemminger, M.; Ohrdes, T.; Schneider, E.; Knoop, M. Dataset on electrical single-family house and heat pump load profiles in Germany. *Sci. Data* **2022**, *9*, 56. [CrossRef]
82. Pullinger, M.; Kilgour, J.; Goddard, N.; Berliner, N.; Webb, L.; Dzikovska, M.; Lovell, H.; Mann, J.; Sutton, C.; Webb, J.; et al. The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes. *Sci. Data* **2021**, *8*, 146. [CrossRef] [PubMed]
83. Sartori, I.; Walnum, H.T.; Skeie, K.S.; Georges, L.; Knudsen, M.D.; Bacher, P.; Candanedo, J.; Sigounis, A.M.; Prakash, A.K.; Pritoni, M.; et al. Sub-hourly measurement datasets from 6 real buildings: Energy use and indoor climate. *Data Brief* **2023**, *48*, 109149. [CrossRef] [PubMed]
84. Delfosse, A.; Hebrail, G.; Zerroug, A. Deep learning applied to nilm: Is data augmentation worth for energy disaggregation? In *ECAI 2020*; IOS Press: Amsterdam, The Netherlands, 2020; pp. 2972–2977.
85. Chen, D.; Irwin, D.; Shenoy, P. Smartsim: A device-accurate smart home simulator for energy analytics. In Proceedings of the 2016 IEEE International Conference on Smart Grid Communications (SmartGridComm), Sydney, NSW, Australia, 6–9 November 2016; pp. 686–692.
86. Meiser, M.; Duppe, B.; Zinnikus, I. SynTiSeD—Synthetic Time Series Data Generator. In Proceedings of the 2023 11th Workshop on Modelling and Simulation of Cyber-Physical Energy Systems (MSCPES), San Antonio, TX, USA, 9 May 2023; pp. 1–6.
87. Long, L.; Ye, H. The roles of thermal insulation and heat storage in the energy performance of the wall materials: A simulation study. *Sci. Rep.* **2016**, *6*, 24181. [CrossRef] [PubMed]
88. Wei, S.; Jones, R.; De Wilde, P. Driving factors for occupant-controlled space heating in residential buildings. *Energy Build.* **2014**, *70*, 36–44. [CrossRef]
89. Ji, R.; Zhang, Z.; He, Y.; Liu, J.; Qu, S. Simulating the effects of anchors on the thermal performance of building insulation systems. *Energy Build.* **2017**, *140*, 501–507. [CrossRef]

90. Pérez-Andreu, V.; Aparicio-Fernández, C.; Vivancos, J.L.; Cárcel-Carrasco, J. Experimental data and simulations of performance and thermal comfort in a typical mediterranean house. *Energies* **2021**, *14*, 3311. [[CrossRef](#)]
91. Badieli, A.; Allinson, D.; Lomas, K. Automated dynamic thermal simulation of houses and housing stocks using readily available reduced data. *Energy Build.* **2019**, *203*, 109431. [[CrossRef](#)]
92. Gaetani, I.; Hoes, P.J.; Hensen, J.L. Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy. *Energy Build.* **2016**, *121*, 188–204. [[CrossRef](#)]
93. Chen, S.; Wu, J.; Pan, Y.; Ge, J.; Huang, Z. Simulation and case study on residential stochastic energy use behaviors based on human dynamics. *Energy Build.* **2020**, *223*, 110182. [[CrossRef](#)]
94. Peng, C.; Yan, D.; Wu, R.; Wang, C.; Zhou, X.; Jiang, Y. Quantitative description and simulation of human behavior in residential buildings. *Build. Simul.* **2012**, *5*, 85–94. [[CrossRef](#)]
95. Chai, C.; Li, G. Human-in-the-loop Techniques in Machine Learning. *IEEE Data Eng. Bull.* **2020**, *43*, 37–52.
96. El Emam, K.; Mosquera, L.; Hoptroff, R. *Practical Synthetic Data Generation: Balancing Privacy and the Broad Availability of Data*; O'Reilly Media: Sebastopol, CA, USA, 2020.
97. Binderbauer, P.J.; Kienberger, T.; Staubmann, T. Synthetic load profile generation for production chains in energy intensive industrial subsectors via a bottom-up approach. *J. Clean. Prod.* **2022**, *331*, 130024. [[CrossRef](#)]
98. Sandhaas, A.; Kim, H.; Hartmann, N. Methodology for Generating Synthetic Load Profiles for Different Industry Types. *Energies* **2022**, *15*, 3683. [[CrossRef](#)]
99. Hong, T.; Macumber, D.; Li, H.; Fleming, K.; Wang, Z. Generation and representation of synthetic smart meter data. *Build. Simul.* **2020**, *13*, 1205–1220. [[CrossRef](#)]
100. Behm, C.; Nolting, L.; Praktiknjo, A. How to model European electricity load profiles using artificial neural networks. *Appl. Energy* **2020**, *277*, 115564. [[CrossRef](#)]
101. Reinhardt, A.; Klemenjak, C. How does load disaggregation performance depend on data characteristics? Insights from a benchmarking study. In Proceedings of the eleventh ACM International Conference on Future Energy Systems, Virtual Event, 22–26 June 2020; pp. 167–177.
102. Harell, A.; Jones, R.; Makonin, S.; Bajić, I.V. TraceGAN: Synthesizing appliance power signatures using generative adversarial networks. *IEEE Trans. Smart Grid* **2021**, *12*, 4553–4563. [[CrossRef](#)]
103. Buneeva, N.; Reinhardt, A. AMBAL: Realistic load signature generation for load disaggregation performance evaluation. In Proceedings of the 2017 IEEE International Conference on Smart Grid Communications (smartgridcomm), Dresden, Germany, 23–26 October 2017; pp. 443–448.
104. Dankar, F.K.; Ibrahim, M. Fake it till you make it: Guidelines for effective synthetic data generation. *Appl. Sci.* **2021**, *11*, 2158. [[CrossRef](#)]
105. Snoke, J.; Raab, G.M.; Nowok, B.; Dibben, C.; Slavkovic, A. General and specific utility measures for synthetic data. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2018**, *181*, 663–688. [[CrossRef](#)]
106. Woo, M.J.; Reiter, J.P.; Oganian, A.; Karr, A.F. Global measures of data utility for microdata masked for disclosure limitation. *J. Priv. Confid.* **2009**, *1*. [[CrossRef](#)]
107. Schenker, N.; Gentleman, J.F. On judging the significance of differences by examining the overlap between confidence intervals. *Am. Stat.* **2001**, *55*, 182–186. [[CrossRef](#)]
108. Loong, B.; Zaslavsky, A.M.; He, Y.; Harrington, D.P. Disclosure control using partially synthetic data for large-scale health surveys, with applications to CanCORS. *Stat. Med.* **2013**, *32*, 4139–4161. [[CrossRef](#)] [[PubMed](#)]
109. Majumdar, S. Fairness, explainability, privacy, and robustness for trustworthy algorithmic decision-making. In *Big Data Analytics in Chemoinformatics and Bioinformatics*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 61–95.
110. Balagopalan, A.; Zhang, H.; Hamidieh, K.; Hartvigsen, T.; Rudzicz, F.; Ghassemi, M. The road to explainability is paved with bias: Measuring the fairness of explanations. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, 21–24 June 2022; pp. 1194–1206.
111. Xu, H.; Mannor, S. Robustness and generalization. *Mach. Learn.* **2012**, *86*, 391–423. [[CrossRef](#)]
112. Tsipras, D.; Santurkar, S.; Engstrom, L.; Turner, A.; Madry, A. Robustness may be at odds with accuracy. *arXiv* **2018**, arXiv:1805.12152.
113. Raghunathan, A.; Xie, S.M.; Yang, F.; Duchi, J.C.; Liang, P. Adversarial training can hurt generalization. *arXiv* **2019**, arXiv:1906.06032.
114. Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. A survey on bias and fairness in machine learning. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–35. [[CrossRef](#)]
115. Tsirikoglou, A. Synthetic Data for Visual Machine Learning: A Data-Centric Approach. Ph.D. Thesis, Linköping University, Linköping, Sweden, 2022.
116. Wang, A.X.; Chukova, S.S.; Nguyen, B.P. Data-Centric AI to Improve Churn Prediction with Synthetic Data. In Proceedings of the 2023 3rd International Conference on Computer, Control and Robotics (ICCCR), Shanghai, China, 24–26 March 2023; pp. 409–413.
117. Qi, B.; Xiao, X.; Liang, J.; Po, L.C.C.; Zhang, L.; Tong, J. An open time-series simulated dataset covering various accidents for nuclear power plants. *Sci. Data* **2022**, *9*, 766. [[CrossRef](#)]

118. Marcu, A.; Costea, D.; Licaret, V.; Pirvu, M.; Slusanschi, E.; Leordeanu, M. SafeUAV: Learning to estimate depth and safe landing areas for UAVs from synthetic data. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Milano, Italy, 8–14 September 2018.
119. Gambi, A.; Nguyen, V.; Ahmed, J.; Fraser, G. Generating critical driving scenarios from accident sketches. In Proceedings of the 2022 IEEE International Conference On Artificial Intelligence Testing (AITest), Newark, CA, USA, 15–18 August 2022; pp. 95–102.
120. Kaufmann, D.; Klampfl, L.; Klück, F.; Zimmermann, M.; Tao, J. Critical and challenging scenario generation based on automatic action behavior sequence optimization: 2021 IEEE autonomous driving ai test challenge group 108. In Proceedings of the 2021 IEEE International Conference On Artificial Intelligence Testing (AITest), Oxford, UK, 23–26 August 2021; pp. 118–127.
121. Tian, H.; Wu, G.; Yan, J.; Jiang, Y.; Wei, J.; Chen, W.; Li, S.; Ye, D. Generating critical test scenarios for autonomous driving systems via influential behavior patterns. In Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering, Rochester, MI, USA, 10–14 October 2022; pp. 1–12.
122. Ding, W.; Chen, B.; Xu, M.; Zhao, D. Learning to collide: An adaptive safety-critical scenarios generating method. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 2243–2250.
123. Murray, D.; Stankovic, L.; Stankovic, V.; Lulic, S.; Sladojevic, S. Transferability of neural network approaches for low-rate energy disaggregation. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 8330–8334.
124. Martinez-Soto, A.; Jentsch, M.F. A transferable energy model for determining the future energy demand and its uncertainty in a country’s residential sector. *Build. Res. Inf.* **2020**, *48*, 587–612. [[CrossRef](#)]
125. Klemenjak, C.; Faustine, A.; Makonin, S.; Elmenreich, W. On metrics to assess the transferability of machine learning models in non-intrusive load monitoring. *arXiv* **2019**, arXiv:1912.06200.
126. Tommasi, T.; Patricia, N.; Caputo, B.; Tuytelaars, T. A deeper look at dataset bias. In *Domain Adaptation in Computer Vision Applications*; Springer, Cham, Switzerland, 2017; pp. 37–55.
127. Torralba, A.; Efros, A.A. Unbiased look at dataset bias. In Proceedings of the CVPR 2011, Washington, DC, USA, 20–25 June 2011; pp. 1521–1528.
128. Khosla, A.; Zhou, T.; Malisiewicz, T.; Efros, A.A.; Torralba, A. Undoing the damage of dataset bias. In Proceedings of the Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; Proceedings, Part I 12; Springer: Berlin/Heidelberg, Germany, 2012; pp. 158–171.
129. Zerilli, J.; Bhatt, U.; Weller, A. How transparency modulates trust in artificial intelligence. *Patterns* **2022**, *3*. [[CrossRef](#)] [[PubMed](#)]
130. Xu, F.; Uszkoreit, H.; Du, Y.; Fan, W.; Zhao, D.; Zhu, J. Explainable AI: A brief survey on history, research areas, approaches and challenges. In Proceedings of the Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, 9–14 October 2019; Proceedings, Part II 8; Springer: Berlin/Heidelberg, Germany, 2019; pp. 563–574.
131. Pearl, J. The limitations of opaque learning machines. *Possible Minds* **2019**, *25*, 13–19.
132. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should i trust you?” Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 13–17 August 2016; pp. 1135–1144.
133. Holm, E.A. In defense of the black box. *Science* **2019**, *364*, 26–27. [[CrossRef](#)] [[PubMed](#)]
134. Hassija, V.; Chamola, V.; Mahapatra, A.; Singal, A.; Goel, D.; Huang, K.; Scardapane, S.; Spinelli, I.; Mahmud, M.; Hussain, A. Interpreting black-box models: A review on explainable artificial intelligence. *Cogn. Comput.* **2024**, *16*, 45–74. [[CrossRef](#)]
135. Holzinger, A.; Saranti, A.; Molnar, C.; Biecek, P.; Samek, W. Explainable AI methods—a brief overview. In *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*; Springer: Cham, Switzerland, 2022; pp. 13–38.
136. Reddy, G.T.; Reddy, M.P.K.; Lakshmana, K.; Kaluri, R.; Rajput, D.S.; Srivastava, G.; Baker, T. Analysis of dimensionality reduction techniques on big data. *IEEE Access* **2020**, *8*, 54776–54788. [[CrossRef](#)]
137. Gogtay, N.J.; Thatte, U.M. Principles of correlation analysis. *J. Assoc. Physicians India* **2017**, *65*, 78–81. [[PubMed](#)]
138. Alaa, A.; Breugel, B.; Saveliev, E.; Schaar, M. How Faithful Is Your Synthetic Data? Sample-Level Metrics for Evaluating and Auditing Generative Models. In *International Conference on Machine Learning*; PMLR: Baltimore, MD, USA, 2022; pp. 290–306.
139. Wu, X.; Xiao, L.; Sun, Y.; Zhang, J.; Ma, T.; He, L. A survey of human-in-the-loop for machine learning. *Future Gener. Comput. Syst.* **2022**, *135*, 364–381. [[CrossRef](#)]
140. Stoyanovich, J.; Howe, B. Nutritional labels for data and models. *Q. Bull. Comput. Soc. IEEE Tech. Comm. Data Eng.* **2019**, *42*, 13–23.
141. Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J.W.; Wallach, H.; Iii, H.D.; Crawford, K. Datasheets for datasets. *Commun. ACM* **2021**, *64*, 86–92. [[CrossRef](#)]
142. Weller, A. Transparency: Motivations and challenges. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 23–40.
143. Kilbertus, N.; Gascón, A.; Kusner, M.; Veale, M.; Gummadi, K.; Weller, A. Blind justice: Fairness with encrypted sensitive attributes. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 2630–2639.
144. Sarkar, A.; Yang, Y.; Vihinen, M. Variation benchmark datasets: Update, criteria, quality and applications. *Database* **2020**, *2020*, baz117. [[CrossRef](#)] [[PubMed](#)]

145. Mamalakis, A.; Ebert-Uphoff, I.; Barnes, E.A. Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. *Environ. Data Sci.* **2022**, *1*, e8. [[CrossRef](#)]
146. Colbois, L.; de Freitas Pereira, T.; Marcel, S. On the use of automatically generated synthetic image datasets for benchmarking face recognition. In Proceedings of the 2021 IEEE International Joint Conference on Biometrics (IJCB), Shenzhen, China, 4–7 August 2021; pp. 1–8.
147. Peng, X.; Usman, B.; Kaushik, N.; Wang, D.; Hoffman, J.; Saenko, K. Visda: A synthetic-to-real benchmark for visual domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 2021–2026.
148. Zhang, J.; Cao, Y.; Zha, Z.J.; Tao, D. Nighttime dehazing with a synthetic benchmark. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 2355–2363.
149. Gundersen, O.E. The fundamental principles of reproducibility. *Philos. Trans. R. Soc. A* **2021**, *379*, 20200210. [[CrossRef](#)] [[PubMed](#)]
150. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **2016**, *533*, 452–454. [[CrossRef](#)] [[PubMed](#)]
151. Pineau, J.; Vincent-Lamarre, P.; Sinha, K.; Larivière, V.; Beygelzimer, A.; d’Alché Buc, F.; Fox, E.; Larochelle, H. Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). *J. Mach. Learn. Res.* **2021**, *22*, 7459–7478.
152. Henderson, P.; Islam, R.; Bachman, P.; Pineau, J.; Precup, D.; Meger, D. Deep reinforcement learning that matters. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
153. Goodman, S.N.; Fanelli, D.; Ioannidis, J.P. What does research reproducibility mean? *Sci. Transl. Med.* **2016**, *8*, 341ps12. [[CrossRef](#)] [[PubMed](#)]
154. Grund, S.; Lüdtke, O.; Robitzsch, A. Using synthetic data to improve the reproducibility of statistical results in psychological research. *Psychol. Methods* **2022**. [[CrossRef](#)] [[PubMed](#)]
155. El Hachimi, C.; Belaqziz, S.; Khabba, S.; Ousanouan, Y.; Sebbar, B.e.; Kharrou, M.H.; Chehbouni, A. ClimateFiller: A Python framework for climate time series gap-filling and diagnosis based on artificial intelligence and multi-source reanalysis data. *Softw. Impacts* **2023**, *18*, 100575. [[CrossRef](#)]
156. Arriagada, P.; Karelavic, B.; Link, O. Automatic gap-filling of daily streamflow time series in data-scarce regions using a machine learning algorithm. *J. Hydrol.* **2021**, *598*, 126454. [[CrossRef](#)]
157. Fu, C.; Quintana, M.; Nagy, Z.; Miller, C. Filling time-series gaps using image techniques: Multidimensional context autoencoder approach for building energy data imputation. *Appl. Therm. Eng.* **2024**, *236*, 121545. [[CrossRef](#)]
158. Quintana, D.S. A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. *Elife* **2020**, *9*, e53275. [[CrossRef](#)] [[PubMed](#)]
159. Chen, R.J.; Lu, M.Y.; Chen, T.Y.; Williamson, D.F.; Mahmood, F. Synthetic data in machine learning for medicine and healthcare. *Nat. Biomed. Eng.* **2021**, *5*, 493–497. [[CrossRef](#)] [[PubMed](#)]
160. Jessop-Fabre, M.M.; Sonnenschein, N. Improving reproducibility in synthetic biology. *Front. Bioeng. Biotechnol.* **2019**, *7*, 18. [[CrossRef](#)]
161. Heil, B.J.; Hoffman, M.M.; Markowitz, F.; Lee, S.I.; Greene, C.S.; Hicks, S.C. Reproducibility standards for machine learning in the life sciences. *Nat. Methods* **2021**, *18*, 1132–1135. [[CrossRef](#)] [[PubMed](#)]
162. Cochran, W.G. *Sampling Techniques*; John Wiley & Sons: Hoboken, NJ, USA, 1977.
163. Kusner, M.J.; Loftus, J.; Russell, C.; Silva, R. Counterfactual fairness. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
164. Hardt, M.; Price, E.; Srebro, N. Equality of opportunity in supervised learning. *Adv. Neural Inf. Process. Syst.* **2016**, *29*.
165. Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; Zemel, R. Fairness through awareness. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, Cambridge, MA, USA, 8–10 January 2012; pp. 214–226.
166. Kleinberg, J.; Mullainathan, S.; Raghavan, M. Inherent trade-offs in the fair determination of risk scores. *arXiv* **2016**, arXiv:1609.05807.
167. Dastin, J. Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of Data and Analytics*; Auerbach Publications: Boca Raton, FL, USA, 2022; pp. 296–299.
168. Segal, B.; Rubin, D.M.; Rubin, G.; Pantanowitz, A. Evaluating the clinical realism of synthetic chest x-rays generated using progressively growing gans. *SN Comput. Sci.* **2021**, *2*, 321. [[CrossRef](#)]
169. van Breugel, B.; Kyono, T.; Berrevoets, J.; van der Schaar, M. Decaf: Generating fair synthetic data using causally-aware generative networks. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22221–22233.
170. Lu, K.; Mardziel, P.; Wu, F.; Amancharla, P.; Datta, A. Gender bias in neural natural language processing. In *Logic, Language, and Security: Essays Dedicated to Andre Scedrov on the Occasion of His 65th Birthday*; Springer: Cham, Switzerland, 2020; pp. 189–202.
171. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018; pp. 77–91.
172. Calmon, F.; Wei, D.; Vinzamuri, B.; Natesan Ramamurthy, K.; Varshney, K.R. Optimized pre-processing for discrimination prevention. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
173. Feldman, M.; Friedler, S.A.; Moeller, J.; Scheidegger, C.; Venkatasubramanian, S. Certifying and removing disparate impact. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; pp. 259–268.

174. Zhang, L.; Wu, Y.; Wu, X. A causal framework for discovering and removing direct and indirect discrimination. *arXiv* **2016**, arXiv:1611.07509.
175. Bohren, J.A.; Imas, A.; Rosenberg, M. The dynamics of discrimination: Theory and evidence. *Am. Econ. Rev.* **2019**, *109*, 3395–3436. [[CrossRef](#)]
176. Willborn, S.L. The disparate impact model of discrimination: Theory and limits. *Am. UL Rev.* **1984**, *34*, 799.
177. Romei, A.; Ruggieri, S. A multidisciplinary survey on discrimination analysis. *Knowl. Eng. Rev.* **2014**, *29*, 582–638. [[CrossRef](#)]
178. Marshall, R. The economics of racial discrimination: A survey. *J. Econ. Lit.* **1974**, *12*, 849–871.
179. Raji, I.D.; Buolamwini, J. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, HI, USA, 27–28 January 2019; pp. 429–435.
180. Schnabel, T.; Swaminathan, A.; Singh, A.; Chandak, N.; Joachims, T. Recommendations as treatments: Debiasing learning and evaluation. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 20–22 June 2016; pp. 1670–1679.
181. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
182. Krasin, I.; Duerig, T.; Alldrin, N.; Ferrari, V.; Abu-El-Haija, S.; Kuznetsova, A.; Rom, H.; Uijlings, J.; Popov, S.; Veit, A.; et al. Openimages: A public dataset for large-scale multi-label and multi-class image classification. *Dataset* **2017**, *2*, 18. Available online: <https://github.com/openimages> (accessed on 15 April 2024).
183. Shankar, S.; Halpern, Y.; Breck, E.; Atwood, J.; Wilson, J.; Sculley, D. No classification without representation: Assessing geodiversity issues in open data sets for the developing world. *arXiv* **2017**, arXiv:1711.08536.
184. Klare, B.F.; Klein, B.; Taborsky, E.; Blanton, A.; Cheney, J.; Allen, K.; Grother, P.; Mah, A.; Jain, A.K. Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1931–1939.
185. Eidinger, E.; Enbar, R.; Hassner, T. Age and gender estimation of unfiltered faces. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 2170–2179. [[CrossRef](#)]
186. Liu, J.; Shen, Z.; He, Y.; Zhang, X.; Xu, R.; Yu, H.; Cui, P. Towards out-of-distribution generalization: A survey. *arXiv* **2021**, arXiv:2108.13624.
187. Moller, F.; Botache, D.; Huseljic, D.; Heidecker, F.; Bieshaar, M.; Sick, B. Out-of-distribution detection and generation using soft brownian offset sampling and autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 46–55.
188. Xu, D.; Yuan, S.; Zhang, L.; Wu, X. Fairgan: Fairness-aware generative adversarial networks. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 570–575.
189. Xu, D.; Wu, Y.; Yuan, S.; Zhang, L.; Wu, X. Achieving causal fairness through generative adversarial networks. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019.
190. Kortylewski, A.; Egger, B.; Schneider, A.; Gerig, T.; Morel-Forster, A.; Vetter, T. Analyzing and reducing the damage of dataset bias to face recognition with synthetic data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 15–20 June 2019.
191. Srivastava, S.; Li, C.; Lingelbach, M.; Martín-Martín, R.; Xia, F.; Vainio, K.E.; Lian, Z.; Gokmen, C.; Buch, S.; Liu, K.; et al. Behavior: Benchmark for everyday household activities in virtual, interactive, and ecological environments. In Proceedings of the Conference on Robot Learning, Auckland, New Zealand, 14–18 December 2022; pp. 477–490.
192. Bender, E.M.; Friedman, B. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Trans. Assoc. Comput. Linguist.* **2018**, *6*, 587–604. [[CrossRef](#)]
193. Holland, S.; Hosny, A.; Newman, S.; Joseph, J.; Chmielinski, K. The dataset nutrition label. *Data Prot. Priv.* **2020**, *12*, 1.
194. Kievit, R.A.; Frankenhuis, W.E.; Waldorp, L.J.; Borsboom, D. Simpson’s paradox in psychological science: A practical guide. *Front. Psychol.* **2013**, *4*, 513. [[CrossRef](#)] [[PubMed](#)]
195. Alipourfard, N.; Fennell, P.G.; Lerman, K. Can you trust the trend? discovering simpson’s paradoxes in social data. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA, USA, 5–9 February 2018; pp. 19–27.
196. Kamiran, F.; Calders, T. Data preprocessing techniques for classification without discrimination. *Knowl. Inf. Syst.* **2012**, *33*, 1–33. [[CrossRef](#)]
197. Mannino, M.; Abouzied, A. Is this real? Generating synthetic data that looks real. In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, New Orleans, LA, USA, 20–23 October 2019; pp. 549–561.
198. Georgopoulos, M.; Oldfield, J.; Nicolaou, M.A.; Panagakis, Y.; Pantic, M. Mitigating demographic bias in facial datasets with style-based multi-attribute transfer. *Int. J. Comput. Vis.* **2021**, *129*, 2288–2307. [[CrossRef](#)]
199. Bhanot, K.; Qi, M.; Erickson, J.S.; Guyon, I.; Bennett, K.P. The problem of fairness in synthetic healthcare data. *Entropy* **2021**, *23*, 1165. [[CrossRef](#)] [[PubMed](#)]
200. Van Noorden, R. The ethical questions that haunt facial-recognition research. *Nature* **2020**, *587*, 354–359. [[CrossRef](#)] [[PubMed](#)]
201. Hittmeir, M.; Mayer, R.; Ekelhart, A. A baseline for attribute disclosure risk in synthetic data. In Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy, New Orleans, LA, USA, 16–18 March 2020; pp. 133–143.

202. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In Proceedings of the Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, 4–7 March 2006; Proceedings 3; Springer: Berlin/Heidelberg, Germany, 2006; pp. 265–284.
203. Majeed, A.; Lee, S. Anonymization techniques for privacy preserving data publishing: A comprehensive survey. *IEEE Access* **2020**, *9*, 8512–8545. [[CrossRef](#)]
204. Stadler, T.; Oprisanu, B.; Troncoso, C. Synthetic data–anonymisation groundhog day. In Proceedings of the 31st USENIX Security Symposium (USENIX Security 22), Boston, MA, USA, 10–12 August 2022; pp. 1451–1468.
205. Brauneck, A.; Schmalhorst, L.; Kazemi Majdabadi, M.M.; Bakhtiari, M.; Völker, U.; Baumbach, J.; Baumbach, L.; Buchholtz, G. Federated machine learning, privacy-enhancing technologies, and data protection laws in medical research: Scoping review. *J. Med. Internet Res.* **2023**, *25*, e41588. [[CrossRef](#)]
206. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial intelligence and Statistics, Fort Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
207. Dwork, C. Differential privacy. In *International Colloquium on Automata, Languages, and Programming*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1–12.
208. Liu, Y.; Zhang, L.; Ge, N.; Li, G. A systematic literature review on federated learning: From a model quality perspective. *arXiv* **2020**, arXiv:2012.01973.
209. Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A.N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. Advances and open problems in federated learning. *Found. Trends[®] Mach. Learn.* **2021**, *14*, 1–210. [[CrossRef](#)]
210. Hong, T.; Pinson, P.; Wang, Y.; Weron, R.; Yang, D.; Zareipour, H. Energy forecasting: A review and outlook. *IEEE Open Access J. Power Energy* **2020**, *7*, 376–388. [[CrossRef](#)]
211. Gu, F.; Chung, M.H.; Chignell, M.; Valaee, S.; Zhou, B.; Liu, X. A survey on deep learning for human activity recognition. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–34. [[CrossRef](#)]
212. Zhang, Y.; Tang, G.; Huang, Q.; Wang, Y.; Wu, K.; Yu, K.; Shao, X. Fednilm: Applying federated learning to nilm applications at the edge. *IEEE Trans. Green Commun. Netw.* **2022**, *7*, 857–868. [[CrossRef](#)]
213. Savi, M.; Olivadese, F. Short-term energy consumption forecasting at the edge: A federated learning approach. *IEEE Access* **2021**, *9*, 95949–95969. [[CrossRef](#)]
214. Xiao, Z.; Xu, X.; Xing, H.; Song, F.; Wang, X.; Zhao, B. A federated learning system with enhanced feature extraction for human activity recognition. *Knowl.-Based Syst.* **2021**, *229*, 107338. [[CrossRef](#)]
215. Lyu, L.; Yu, H.; Yang, Q. Threats to federated learning: A survey. *arXiv* **2020**, arXiv:2003.02133.
216. Mugunthan, V.; Polychroniadou, A.; Byrd, D.; Balch, T.H. Smpai: Secure multi-party computation for federated learning. In Proceedings of the NeurIPS 2019 Workshop on Robust AI in Financial Services, Vancouver, BC, Canada, 13 December 2019; MIT Press: Cambridge, MA, USA, 2019; pp. 1–9.
217. Brundage, M.; Avin, S.; Wang, J.; Belfield, H.; Krueger, G.; Hadfield, G.; Khlaaf, H.; Yang, J.; Toner, H.; Fong, R.; et al. Toward trustworthy AI development: Mechanisms for supporting verifiable claims. *arXiv* **2020**, arXiv:2004.07213.
218. Xin, B.; Geng, Y.; Hu, T.; Chen, S.; Yang, W.; Wang, S.; Huang, L. Federated synthetic data generation with differential privacy. *Neurocomputing* **2022**, *468*, 1–10. [[CrossRef](#)]
219. Rodríguez-Barroso, N.; Stipicich, G.; Jiménez-López, D.; Ruiz-Millán, J.A.; Martínez-Cámara, E.; González-Seco, G.; Luzón, M.V.; Véganzones, M.A.; Herrera, F. Federated Learning and Differential Privacy: Software tools analysis, the Sherpa. ai FL framework and methodological guidelines for preserving data privacy. *Inf. Fusion* **2020**, *64*, 270–292. [[CrossRef](#)]
220. Xin, B.; Yang, W.; Geng, Y.; Chen, S.; Wang, S.; Huang, L. Private fl-gan: Differential privacy synthetic data generation based on federated learning. In Proceedings of the ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 2927–2931.
221. McClure, D.; Reiter, J.P. Differential Privacy and Statistical Disclosure Risk Measures: An Investigation with Binary Synthetic Data. *Trans. Data Priv.* **2012**, *5*, 535–552.
222. Varma, G.; Chauhan, R.; Singh, D. Sarve: Synthetic data and local differential privacy for private frequency estimation. *Cybersecurity* **2022**, *5*, 26. [[CrossRef](#)] [[PubMed](#)]
223. Rosenblatt, L.; Liu, X.; Pouyanfar, S.; de Leon, E.; Desai, A.; Allen, J. Differentially private synthetic data: Applied evaluations and enhancements. *arXiv* **2020**, arXiv:2011.05537.
224. Jordon, J.; Yoon, J.; Van Der Schaar, M. PATE-GAN: Generating synthetic data with differential privacy guarantees. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
225. Arora, A.; Arora, A. Synthetic patient data in health care: A widening legal loophole. *Lancet* **2022**, *399*, 1601–1602. [[CrossRef](#)]
226. Haddad, F. How to Evaluate the Quality of the Synthetic Data. In AWS Machine Learning Blog. 2022. Available online: <https://aws.amazon.com/blogs/machine-learning/how-to-evaluate-the-quality-of-the-synthetic-data-measuring-from-the-perspective-of-fidelity-utility-and-privacy/> (accessed on 15 April 2024).
227. Puri, R.; Spring, R.; Patwary, M.; Shoeybi, M.; Catanzaro, B. Training question answering models from synthetic data. *arXiv* **2020**, arXiv:2002.09599.
228. van Breugel, B.; Sun, H.; Qian, Z.; van der Schaar, M. Membership inference attacks against synthetic data through overfitting detection. *arXiv* **2023**, arXiv:2302.12580.

229. Carlini, N.; Chien, S.; Nasr, M.; Song, S.; Terzis, A.; Tramer, F. Membership inference attacks from first principles. In Proceedings of the 2022 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 22–26 May 2022; pp. 1897–1914.
230. Shokri, R.; Stronati, M.; Song, C.; Shmatikov, V. Membership inference attacks against machine learning models. In Proceedings of the 2017 IEEE symposium on security and privacy (SP), San Jose, CA, USA, 22–24 May 2017; pp. 3–18.
231. Arjunan, P.; Poolla, K.; Miller, C. EnergyStar++: Towards more accurate and explanatory building energy benchmarking. *Appl. Energy* **2020**, *276*, 115413. [[CrossRef](#)]
232. Chen, Y.; Hong, T.; Luo, X.; Hooper, B. Development of city buildings dataset for urban building energy modeling. *Energy Build.* **2019**, *183*, 252–265. [[CrossRef](#)]
233. Ribeiro, M.; Pereira, L.; Quintal, F.; Nunes, N. SustDataED: A public dataset for electric energy disaggregation research. In *ICT for Sustainability 2016*; Atlantis Press: Amsterdam, The Netherlands, 2016; pp. 244–245.
234. Filip, A. Blued: A fully labeled public dataset for event-based nonintrusive load monitoring research. In Proceedings of the 2nd Workshop on Data Mining Applications in Sustainability (SustKDD), San Diego, CA, USA, 21 August 2011; Volume 2012.
235. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons. In Official Journal of the European Union. European Union. 2016. Available online: <http://data.europa.eu/eli/reg/2016/679/oj> (accessed on 15 April 2024).
236. Young, M.; Rodriguez, L.; Keller, E.; Sun, F.; Sa, B.; Whittington, J.; Howe, B. Beyond open vs. closed: Balancing individual privacy and public accountability in data sharing. In Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 29–31 January 2019; pp. 191–200.
237. Van Wynsberghe, A. Sustainable AI: AI for sustainability and the sustainability of AI. *AI Ethics* **2021**, *1*, 213–218. [[CrossRef](#)]
238. Strubell, E.; Ganesh, A.; McCallum, A. Energy and policy considerations for deep learning in NLP. *arXiv* **2019**, arXiv:1906.02243.
239. Ray, P.P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber-Phys. Syst.* **2023**, *2*, 121–154. [[CrossRef](#)]
240. Lacoste, A.; Luccioni, A.; Schmidt, V.; Dandres, T. Quantifying the carbon emissions of machine learning. *arXiv* **2019**, arXiv:1910.09700.
241. Henderson, P.; Hu, J.; Romoff, J.; Brunskill, E.; Jurafsky, D.; Pineau, J. Towards the systematic reporting of the energy and carbon footprints of machine learning. *J. Mach. Learn. Res.* **2020**, *21*, 10039–10081.
242. Patterson, D.; Gonzalez, J.; Hölzle, U.; Le, Q.; Liang, C.; Munguia, L.M.; Rothchild, D.; So, D.R.; Texier, M.; Dean, J. The carbon footprint of machine learning training will plateau, then shrink. *Computer* **2022**, *55*, 18–28. [[CrossRef](#)]
243. Yigitcanlar, T.; Mehmood, R.; Corchado, J.M. Green artificial intelligence: Towards an efficient, sustainable and equitable technology for smart cities and futures. *Sustainability* **2021**, *13*, 8952. [[CrossRef](#)]
244. Kumar, S.; Buyya, R. Green cloud computing and environmental sustainability. In *Harnessing Green IT: Principles and Practices*; Wiley: Hoboken, NJ, USA, 2012; pp. 315–339.
245. Graybill, R.; Melhem, R. *Power Aware Computing*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
246. Sachan, V.K.; Imam, S.A.; Beg, M.T. Energy-efficient communication methods in wireless sensor networks: A critical review. *Int. J. Comput. Appl.* **2012**, *39*, 35–48.
247. Ali, A.S.; Zanzinger, Z.; Debose, D.; Stephens, B. Open Source Building Science Sensors (OSBSS): A low-cost Arduino-based platform for long-term indoor environmental data collection. *Build. Environ.* **2016**, *100*, 114–126. [[CrossRef](#)]
248. Lovett, T.; Gabe-Thomas, E.; Natarajan, S.; Brown, M.; Padget, J. Designing sensor sets for capturing energy events in buildings. In Proceedings of the 5th International Conference on Future Energy Systems, Cambridge, UK, 11–13 June 2014; pp. 229–230.
249. Abdella, G.M.; Kucukvar, M.; Onat, N.C.; Al-Yafay, H.M.; Bulak, M.E. Sustainability assessment and modeling based on supervised machine learning techniques: The case for food consumption. *J. Clean. Prod.* **2020**, *251*, 119661. [[CrossRef](#)]
250. De Las Heras, A.; Luque-Sendra, A.; Zamora-Polo, F. Machine learning technologies for sustainability in smart cities in the post-covid era. *Sustainability* **2020**, *12*, 9320. [[CrossRef](#)]
251. Pham, A.D.; Ngo, N.T.; Truong, T.T.H.; Huynh, N.T.; Truong, N.S. Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. *J. Clean. Prod.* **2020**, *260*, 121082. [[CrossRef](#)]
252. So, H.Y.; Chen, P.P.; Wong, G.K.C.; Chan, T.T.N. Simulation in medical education. *J. R. Coll. Physicians Edinb.* **2019**, *49*, 52–57. [[CrossRef](#)] [[PubMed](#)]
253. de Paula Ferreira, W.; Armellini, F.; De Santa-Eulalia, L.A. Simulation in industry 4.0: A state-of-the-art review. *Comput. Ind. Eng.* **2020**, *149*, 106868. [[CrossRef](#)]
254. Kato, T.; Kamoshida, R. Multi-agent simulation environment for logistics warehouse design based on self-contained agents. *Appl. Sci.* **2020**, *10*, 7552. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.