Towards to a AI-based energy transition, what kind of a platform is needed?

Hou Shengren
h.shengren@tudelft.nl
Intelligent Electrical Power Grids
Building 36, Room LB03.170

February 24, 2023

1 Motivation

The energy system is currently undergoing a significant transition, moving from a centralized, fuel-based, and automatic control-dominated system to a decentralized, highly renewable energy penetration, and autonomous entity. This transition necessitates advanced artificial intelligence (AI) technologies to reformulate the operation framework and simulation models, and fully explore the multi-dimensional, historical, and operational data within energy systems. These data contain hidden paradigms and valuable information that were previously ignored due to various reasons, such as a shortage of computing power and insufficient ability of representative models.

AI-based energy transition relies heavily on a large amount of high-quality data. This data includes information about existing power stations, capacity, yearly electricity consumption, and ancillary service requirements, as well as (hourly) time series of load, wind, and solar power generation, heat demand, and more. However, data collection can be a tedious process. The bits and pieces of data are sometimes difficult to find, often poorly documented, and almost always tedious to process. Files may be provided in different formats, downloading requires repetitive manual clicking, data structures between different sources are incompatible, daylight savings time and leap years are treated differently, URLs change frequently, and older data is updated without informing users (and sometimes deleted altogether).

Doing double work is inefficient. Currently, dozens if not hundreds of modeling teams in Europe spend significant resources gathering and processing data, all doing essentially the same thing. Highly skilled people waste a lot of time gathering data, time that would be better spent doing actual research. For instance, a standard workflow of the data-driven model includes data collection, pre-processing (exploration, sampling, dimensionality reduction, cleaning, and integration, etc.), and analytical processing (building models and performance evaluation), as shown in Fig. 1. It is often said that the actual research part may only consist of 10% of the core work (learning process), while 90% is related to setting up the workflow, data collection, and management. However, the core components of energy transition research involve building data-driven models to find patterns in data that may provide insights into the phenomena described by data and solve challenges faced by the system. In this regard, wasting time on other steps is impractical. We have gone through this process ourselves, which can be a frustrating experience.

Moreover, data and models should be integrated together with a standard framework for solving specific tasks in data-driven energy research, such as forecasting and management. That is to say, the entity should be stored in this way: several task-related data-sets are used to update the data-driven model for a specific research task. Thus, the data-based model should first feed the original data into the model, then preprocess the data, then update the parameters based on the data. This is called the "directly use feature," which is vital for individuals from different backgrounds to become familiar with, learn, and explore the magic of data-based models in energy systems.

We believe an open-source platform that standardizes the workflow, links data and models tightly and solves representative tasks in the energy system will lead and help users focus on the model's core, which can significantly promote the progress of AI-based energy transition.

1.1 Concepts for Data and Models

In this section, we present concepts related to the collection and use of open-sourced data and models in the context of power systems and energy transition.

To facilitate the integration of AI technologies in power systems and energy transition, we collected and investigated existing open-sourced data and model platforms. We also explored and listed how these platforms stored data, models, and other features. The information was stored in a meta-data structure in JSON format.

1.1.1 Concept for Power System Data

Power system data is essential for AI-based energy transition. The databases may provide information on national power plant fleets, renewable generation assets, transmission networks, time series for electricity loads, dispatch, spot prices, cross-border trades, weather information, and other relevant factors. Additionally, they may offer other energy statistics, such as fossil fuel imports and exports, gas, oil, and coal prices, emissions certificate prices, and information on energy efficiency costs and benefits.

With the integration of AI technologies and energy transition, high-quality, structured, and high-resolution data are urgently needed. We explored the current data platforms and found that they have already collected a vast amount of power system data, including electricity consumption from hourly to higher resolutions, such as minutes, renewable generation, market data, including time-series spot prices, forward contracts, technical data, including installed capacity, the profile of power plants, specific network topology, and related parameters.

1.1.2 Concept for Power System Models

Power system models describe the behavior of the power system and its components, such as generators, transmission lines, transformers, and loads. They are used to analyze the performance of the power system and assess its reliability, security, and economic efficiency.

Power system models typically have a temporal resolution of one hour or less. Some models concentrate on the engineering characteristics of the system, including a good representation of high-voltage transmission networks and AC power flow. Others depict electricity spot markets and are known as dispatch models. While other models embed autonomous agents to capture, for instance, bidding decisions using techniques from bounded rationality.

Recently, data-driven-based power system models that employ AI technologies have become increasingly popular. These models use data to identify patterns and relationships and derive insights that may not be apparent through traditional modeling approaches. AI technologies have been implemented to solve various tasks in power systems, including optimal power flow, economic dispatch, energy management, false data injection, electricity forecasting, and more.

2 Artificial intelligence applications in Power Systems

The energy transition depends heavily on the information and communication infrastructure and efficient management of the massive amount of data generated from various sources such as smart meters, phase measurement units, and a range of sensors. The integration of different renewable and intermittent DG systems and the increasing communication between power system entities present a host of technical challenges. These include dealing with the growing volumes of data generated throughout the grid network, the consequent need for data storage and processing, and managing highly non-linear power systems. Traditional approaches based solely on mathematical equations derived from physics may not be suitable for handling these challenges in modern power systems due to their limitations.

Therefore, there is an urgent need for novel approaches that can handle big data, extract values and information from data, and formulate an implicit paradigm of the system. Data-driven models, as representative AI technologies, have been applied to different areas of power systems, which can be categorized as 1. Forecasting, 2. Asset monitoring and management, 3. Power consumption behavior analysis, and 4. Control and Operation, as summarized in Table. 1.

Table 1: Summary of AI applications on power systems

Application domain		
Forecasting	Renewable generation [1]	Solar intensity prediction
		PV forecasting
		Wind power forecasting
		Wind power ramp forecasting
		Wind speed forecasting
	Load and price forecasting [2, 3]	Short-term forecasting
		Medium-term forecasting
		Long-term forecasting
Asset monitor and management	Transformer health management [4]	
	Cable partial discharge monitor [5]	
	System fault detection and diagnosis [6]	
	System anomaly detection [7]	
	System resilience [8]	
	Cyber-attack detection [9]	DoS attacks
		FDI attacks
		Cyber deception attacks
		Spoofing attacks
Power consumption behavior analysis		
Control and Operation	Power market trading [10]	
	Congestion management [11]	Network reconfiguration
		Active and reactive power control
	Demand side management [12]	Tariff price design
		load shedding management
		Electricity pricing
		home energy management
	Optimal energy system scheduling [13]	energy dispatch
		distributed energy storage system control
		Integrated energy system management
		microgrid energy management

3 Case Study

In this section, we will explore a significant task in power system research: congestion management. We will use this example to demonstrate the frameworks and details of data and models used in both model-based and data-driven research methods.

3.1 Congestion Management

Due to increasing load demand and high penetration of renewable energies, distribution networks frequently operate under over-loading and bilateral power flow conditions, leading to serious congestion problems. Congestion management is, therefore, critical to ensuring safe, cost-effective, and sustainable power system operation.

Congestion problems in distribution networks can be classified into voltage problems and overloading problems. Voltage problems include low voltage, caused by overloading on the end of a feeder, and over-voltage, caused by high penetration of PV systems on the coupling bus. Overloading occurs when loading is close to or exceeds the thermal limits of power components. Modern distribution networks often have long feeders, where voltage problems are more common than thermal limits. Thus, we will use voltage regulation as an example of congestion management research.

Voltage drop occurs when delivering active and reactive power through a feeder, while the penetration of local renewable energies causes over-voltage problems. The control methods for voltage regulation in distribution networks consist of network reconfiguration, reactive power control, and active power control. Network reconfiguration refers to changing the distribution network structure by adjusting the status of the switches to deliver power to customers in an efficient and regulated voltage condition. Reactive and active power control methods refer to changing power flow to meet voltage constraints through demand response, power generation curtailment, or energy storage systems.

Intelligent data-driven and model-based approaches to voltage problems are vital for the energy transition in a cost-effective and timely manner. In this section, ongoing research aims to identify which methods are most reliable, efficient, and why. We identify two categories of approaches, for each of which we indicate the information required for thorough analysis, such that the outcomes of the research provide trustworthy and clear directions for impact and performance in practice. We analyze existing work and extrapolate what would be needed to continue the respective line of research. The commonalities and differences between the two approaches are summarized, which are illustrated based on model, data, and algorithm categories. We identify the boundaries between the model, data, and algorithm below.

- The model is a surrogate of the realistic entity, which should be created using accurate parameters and mathematical formulation. In voltage regulation problems, the model is an electricity network with a specific topology and parameters. The parameters used for creating a model can also be seen as a kind of data.

- Data refers to the information that can be measured and collected explicitly, used to record the model condition during electricity generation, transmission, and load consumption dynamics. In voltage regulation problems, input data includes reactive and active load consumption of each node, while output data is the voltage of each node, which is calculated by solvers executing power flow estimation.

- The algorithm is used to determine the values of variables with a specific goal, given the input data and model. In the voltage regulation problem, the decision variables can be power curtailment of PV generation, charge/discharge of storage systems, and the goal is to limit the voltage to predefined boundaries like $0.95pu - 1.05pu$.

Note that regardless of whether using model-based or data-driven approaches, the goal is to develop a new algorithm (method) that can meet voltage regulation demands in a realistic electricity network, i.e. real-time response, optimal solution, etc.

3.1.1 Case study for Model-Based Voltage Regulation

In the research paper [14], the authors propose a model-based approach for voltage regulation in an active distribution network using smart photovoltaic (PV) inverters to control power flow. The first step of the approach involves building a model of the active distribution network and its energy units, which includes parameters for the distribution transformer, PV inverters, and the line. The model is based on mathematical formulations and parameters.

The authors then input one day's load consumption data and PV generation data for each node into the model to determine the voltage violation situation without control of active power. They find that the voltage drop problem of one node can be optimally solved by reducing the active power output of the neighboring node with low voltage. On the other hand, for voltage rise problems, adjusting the node itself or the neighbor nodes yields similar performance.

Based on this analysis, the authors propose a decentralized and real-time method for voltage regulation that relies on the knowledge and analysis of the model rather than the input data. The input data is used only once and is not essential to the method's development. The method is called model-based because finding the optimal solution depends on the model. The estimation of model dynamics and the determination of the optimal values of decision variables are performed during the online-schedule stage, which is closely linked to the model.

3.1.2 Case Study for Supervise Learning Based Voltage Regulation

Supervised learning-based voltage regulation problems require a large amount of labeled data [15]. These methods aim to build a mapping between input (load consumption and PV generation of each node) and output (the optimal value of active power control) based on the weights and parameters of the surrogate function, which can be a linear function or a deep neural network. Figure 2 illustrates the workflow for data-driven voltage regulation, which involves collecting data sets with the same resolution, including load consumption, renewable generation, and price data, as well as profile data of the corresponding power network and assets. To improve performance and scalability, a broad range of data representing different scenarios is necessary to update the weights and bias of the deep neural networks. The optimal value of decision variables, which can be gathered through iterative solving of the model-based optimization approaches, serves as the label for the data.

During the online scheduling stage, the trained deep neural network can directly obtain the optimal value of decision variables without building and solving a model, as in the previous case. The deep neural network has built a black box function based on the combination of weights and biases that represents the model dynamics and the correspondence between voltage and decision variables, as well as the optimal goal.

Therefore, supervised learning-based voltage regulation algorithms rely heavily on data rather than a model. They use labeled data to train the parameters of deep neural networks, indirectly building a surrogate model. In contrast, model-based approaches rely on accurate parameters and mathematical formulations.

3.1.3 Discussion

In order to solve voltage regulation problems, researchers often use either model-based or data-driven approaches. These two methods can be divided into three components: the model, data, and al-

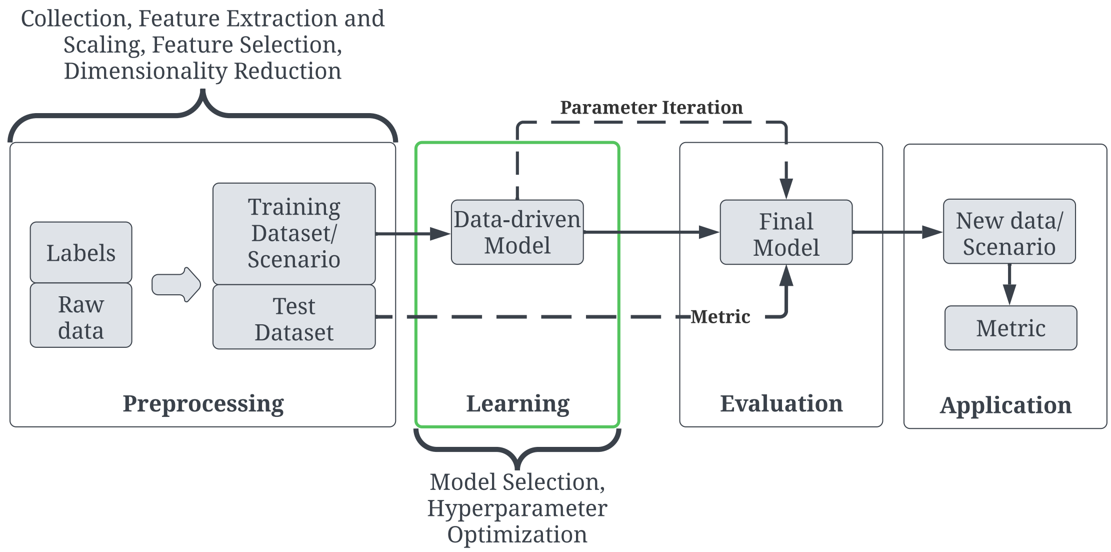

Figure 1: Working Pipeline for data-driven model

gorithms. Model-based algorithms are heavily reliant on the accuracy of the nominal model, which is derived from physical rules. This means that the performance of the algorithm is limited by the quality of the nominal model used to test it. In practice, full knowledge, information, and parameters of a distribution network are rarely collected, which means that algorithms developed for the nominal model cannot be directly used in a realistic distribution network. To address this issue, the use of digital twins is becoming increasingly urgent as they accurately represent the physical distribution network when developing the model-based approach.

On the other hand, researchers are increasingly turning to data-driven methods to accelerate the AI-based energy transition. This trend can be attributed to the accumulation of large amounts of data in energy systems, which can now be leveraged with the help of flourishing machine learning technologies, especially deep learning. Data-driven approaches are theoretically capable of overcoming the inaccuracies of model-based approaches by drawing information from specific energy networks. However, the performance of data-driven approaches is significantly impacted by the quality of data, such as completeness and accuracy. In addition, the representative ability of the chosen model is the performance bottleneck for the specific energy task. Many previous results are limited to a particular model or data set, extrapolating claims to broader settings where this may not always be appropriate. Furthermore, comparisons between competing ideas are often done in limited settings or not performed at all. To draw reliable and universal conclusions, easily available and representative models and data are essential.

Figure 1 shows a working pipeline for data-driven models. Overall, both model-based and data-driven approaches have their advantages and limitations, and their suitability depends on the specific energy task at hand.

4 Comparison and summary for existing platforms

4.1 Overview for Data platforms

Table 2 presents a list of representative open power system data platforms such as Open Power System Data and ENTSO-E Transparency. These platforms include only raw or reorganized data without any models. However, the data provided by these platforms cannot be directly used for different tasks.

Users need to identify the required resolution of data for specific tasks and then clean and process the data again for the model used in that task.

4.2 Overview of Model platforms

Table 2: Power system data platforms

Name	Scope	Description
Open power system data	Mixed data platform	Standarized open power system data platform
ENTSO-E Transparency	Mixed data platform	The central data platform of the European transmission system operators
AG Energiebilanzen.	Electricity consumption	Yearly energy balances for Germany, including yearly electricity consumption by sector.
Statistisches Bundesamt	Electricity consumption	Monthly electricity, heat and gas consumption for Germany
50 Hertz	Electricity consumption	Hourly wind generation (part of open power system data)
IRENA RE Capacity Statistics	Capacity	Installed renewable capacity since 2006 for all countries world-wide. PDF.
Electric power statistics	Capacity	2015 combined heat and power (CHP) data
EurObserv'ER Database.	Capacity	Different renewable-related statistics in all EU members
IEA PVPS	Capacity	Global solar PV capacities dating back to 1992
United Nations Statistics Division Energy Statistics Database.	Capacity	Annual capacity data
IEA monthly electricity statistics	Generation	Monthly electricity generation since 2007 by fuel category for OECD
ENTSO-E production data	Generation	Monthly generation by fuel
ENTSO-E Exchange data	Cross-border trade	Monthly electricity trade between EU countries
ENTSO-E Statistical Factsheet	Cross-border trade	Annual data on installed capacity as well as generation and physical flows by technology and country since 2009. PDF files.
energinet.dk	Price data	Fabulous website of the Danish TSO, providing day-ahead spot prices for DE/AT, NO, SE, DK (all bidding zones), dating back to 2002.
Power Exchanges	Price data	
DE imbalance price (amprion)	Balancing price data	Unbalanced data and price for 15mins
RTE	Mixed data platform	Generation, consumption, market and network data
Ninja renewable energy	Wind and solar time series data	Run simulations of the hourly power output from wind and solar power plants located anywhere in the world.

Table 3: Power System Models

Project	Host	License	Access	Coding	Documentation	Scope/type
AMIRIS	German Aerospace Center	Apache 2.0	GitLab	Java	wiki	Agent-based electricity market modeling
Breakthrough Energy Model	Breakthrough Energy Foundation	MIT	GitHub	Python, Julia	website, GitHub	Power sector modeling
DIETER	DIW Berlin	MIT	Download	GAMS	Publicated paper	Dispatch and investment
Dispa-SET	EC Joint Research Centre	EUPL 1.1	GitHub	GAMS, Python	website	European transmission and dispatch
PowerGridModel	Alliander	Apache 2.0	GitHub	Python, C++	Website	
EMMA	Neon Neue Energieökonomik	CC BY-SA 3.0	Download	GAMS	website	electricity market
GENESYS	RWTH Aachen University	LGPLv2.1	GitHub	C++	website	European electricity system
NEMO	University of New South Wales	GPLv3	GitHub	Python	website, list	Australian NEM market
OnSSET	KTH Royal Institute of Technology	MIT	GitHub	Python	website, GitHub	cost-effective electrification
pandapower	University of Kassel,	BSD-new	GitHub	Python	website	automated power system analysis
PowerMatcher	Flexiblepower Alliance Network	Apache 2.0	GitHub	Java	website	smart grid
Power TAC	Erasmus Centre for Future Energy Business Rotterdam School of Management Erasmus University	Apache 2.0	GitHub	Java	website, forum	automated retail electricity trading simulation
renpass	University of Flensburg	GPLv3	Invitation	R, MySQL	manual	renewables pathways
SciGRID	DLR Institute of Networked Energy Systems	Apache 2.0	GitHub	Python	website, newsletter	European transmission grid
SIREN	Sustainable Energy Now	AGPLv3	GitHub	Python	website	renewable generation
SWITCH	University of Hawai'i	Apache 2.0	GitHub	Python	website	optimal planning
URBS	Technical University of Munich	GPLv3	GitHub	Python	website	distributed energy systems
RLOESS	TU Delft	MIT License	Github	Python	Github	distributed energy systems
Europe Gaint	binational project	Proprietary software	Website	-	Website	Data and model platform
FlexMeasures	Seita	Apache 2.0	Github	Python	Website	Energy management problem

Table 3 shows a list of power system model platforms such as AMIRIS, DIETER, and PowerGridModel. These platforms provide models for power system simulation and analysis. The models can be used to simulate various scenarios such as dispatch and investment analysis, electricity market modeling, and European transmission and dispatch. The coding languages used in these models include Java, Python, Julia, and GAMS, and the documentation of these models is available on their respective websites, wikis, or publications.

It is important to note that these platforms may require users to have some expertise in coding and power system analysis. Therefore, it is essential to choose a platform that meets the user's needs and level of expertise.

4.3 Overview of integration platforms

Integration platforms like *OpenML* and *Europe Giant* aim to provide a consolidated platform for accessing both data and models/algorithms for various tasks. However, despite their advantages, these platforms still have some drawbacks.

- Meta-data is rough, and not clear. The metadata is often incomplete or unclear, which can make it difficult to understand the structure of the data and how it should be used. This can increase the learning cost and make it challenging to use the data effectively.

- Integration between data and models and algorithms. The integration between data, models, and algorithms is often encapsulated, which can limit flexibility and make it difficult to customize the platform to meet specific needs. This can be particularly problematic for researchers who need to modify or adapt algorithms to suit their research questions.

- Data management problem. How to manage and access open-source and private data together, while respecting the data sovereignty of owners, can be a challenging problem. It's crucial to have strong data governance policies in place to protect sensitive data and ensure that all users have the appropriate permissions to access and use the data.

- Data fragmentation problem. In these platforms, data fragmentation can make it difficult to integrate datasets from different sources. Each dataset may have a different format or structure, which can make it challenging to combine and analyze them effectively. To address this issue, it's important to develop standards for data sharing and integration, so that data can be easily combined and used across different platforms.

4.4 Feature comparison of representative platforms

ENTSO-E transparency is an energy data platform that contains electricity generation, load, transmission, and balancing data. It is built according to the regulation (EU) No 5 43/2013 of 14 June 2013 on submission and publication of data in electricity markets. With this platform, users can collect energy-related data in the EU and visualize it with the dashboard provided. The data can be filtered with simple metadata labels such as generation, load or time-series, and countries. RTE and energinetdk also provide similar functions and features as data platforms, promoting the transparency of electricity markets. However, data provided by these platforms are dispersed over various repositories and formatted inconsistently, which makes it very hard to use directly in energy research. Additionally, poorly documented data and lack of metadata cause challenges in combining data from different sources, with data quality problems such as inconsistencies, errors, and missing widely existing.

To overcome these challenges, Open Power System Data (OPSD) collects data provided on these platforms and individual sources, processes, documents, and republishes these data. The republished data have a standardized format, providing CSV files and a text file containing structured metadata in the JSON format. OPSD provides scripts and detailed documents containing all information needed to generate these data, increasing the transparency and reliability of data. OPSD categorizes data into consumption, renewable generation, and market subgroups, with extra descriptions of resolution, region, and length, among others. However, for data-driven-based congestion management of a specific network, technical data of this network, such as topology and parameters of energy assets, and time-series data in the network are required. This is important to realize the 'directly use feature' or frictionless we mentioned, which is lacking in the current OPSD platform.

Open Machine Learning (OML) platform is built to realize the frictionless of machine learning research, which can tensely integrate models and datasets with different tasks. It abstracts machine learning research into four concepts: datasets, tasks, flows and runs. The data-driven research is organized by:

- Finding which tasks (e.g., classification) need to be solved for every dataset.
- Finding all evaluation runs that people did and how well their models performed for every task.
- Finding model details, evaluations, and the exact algorithm pipelines used for every run.
- Finding all the evaluation runs for every flow (pipeline) to see how well it performed on different tasks.

Behind such a structure, the primary concept is that data is the cornerstone while tasks are goals. Datasets have well-documented metadata. With these metadata, users can filter by many properties such as size, type, tasks, and more. OML also provides automated analysis for datasets, computing a range of data quality characteristics. To fit the character of machine learning, OML uses Run to record the details of algorithms and related hyper-parameters. This platform mainly solved the 'directly use feature' problem of data-driven research by providing functions.

However, OML was restricted by the primary design concepts and caused several serious issues listed below:

- Disorderly increased amounts of datasets and tasks.
- High-level encapsulation of data processing and algorithms.
- Overly distributed data uploading without management.

- Neglect of data sovereignty.

Disorderly increased numbers of datasets and tasks increased the cost of organizing, checking and investigating the platform. High-level encapsulation of workflow for data-driven research increased the learning cost for beginners. Also, it harms the transparency of datasets. Overly distributed data uploading without management caused a serious data fragment problem. For instance, users uploaded different datasets with different labels, ids. However, these datasets actually contain the same data. Each dataset can be counted as a data fragment in the data lake of the platform. In OML, these small datasets are separated in the whole lake because of inadequate management. Actually, similar datasets can continuously be integrated together if sufficient metadata is provided. Moreover, the neglect of data sovereignty damaged the enthusiasm of users from private companies to share their valuable data on the platform especially when the thought: "data is an asset" is popular. A positive example platform, Europe Giant, handles open and restricted data together using data market and federation training technology.

In summary, existing data platforms for energy research promoted the concept of data transparency and facilitated energy research in the model-based energy research stage. However, the separation of datasets and models in these platforms extremely damaged the 'directly use feature', especially in the AI-based energy research stage. In this stage, data-driven-based research is more popular, which requires a higher demand for the integration of data and models. OML provided a good template for an open data-driven research platform, defining a tense link between data and model, connecting these two components with tasks. However, it is focused on machine learning instead of energy research. Moreover, it contains issues like high-level encapsulation, poor data management, etc. motioned before. Inspired by the flourishing of AI-based energy transition research, We absorb the good features of existing platforms from different areas to develop an open data and model platform to promote the progress of AI-based energy transition.

5 Platform design and features

This section illustrates the functions and features of the designed platform. Fig 2 shows the working flow of congestion management tasks on the designed platform. We first choose compatible data sets from different big data lakes based on the specific task's demand. Then, these data are integrated with a model chosen from the model lake to solve the task.

5.1 Data quality check

The platform checks the data uploaded from different users, ensuring that it meets the requirements of meta-data, original data link, and clear description.

5.2 Data retrieval

The following features are available for data retrieval:

- High-quality meta-data: The meta-data should include more information based on different energy system data features. It should include time resolution, range, the task used for the specific data, owner of the data, meta-data table, and license information, etc.

- List and search: Users can filter and search data based on the meta-data, keywords, etc.

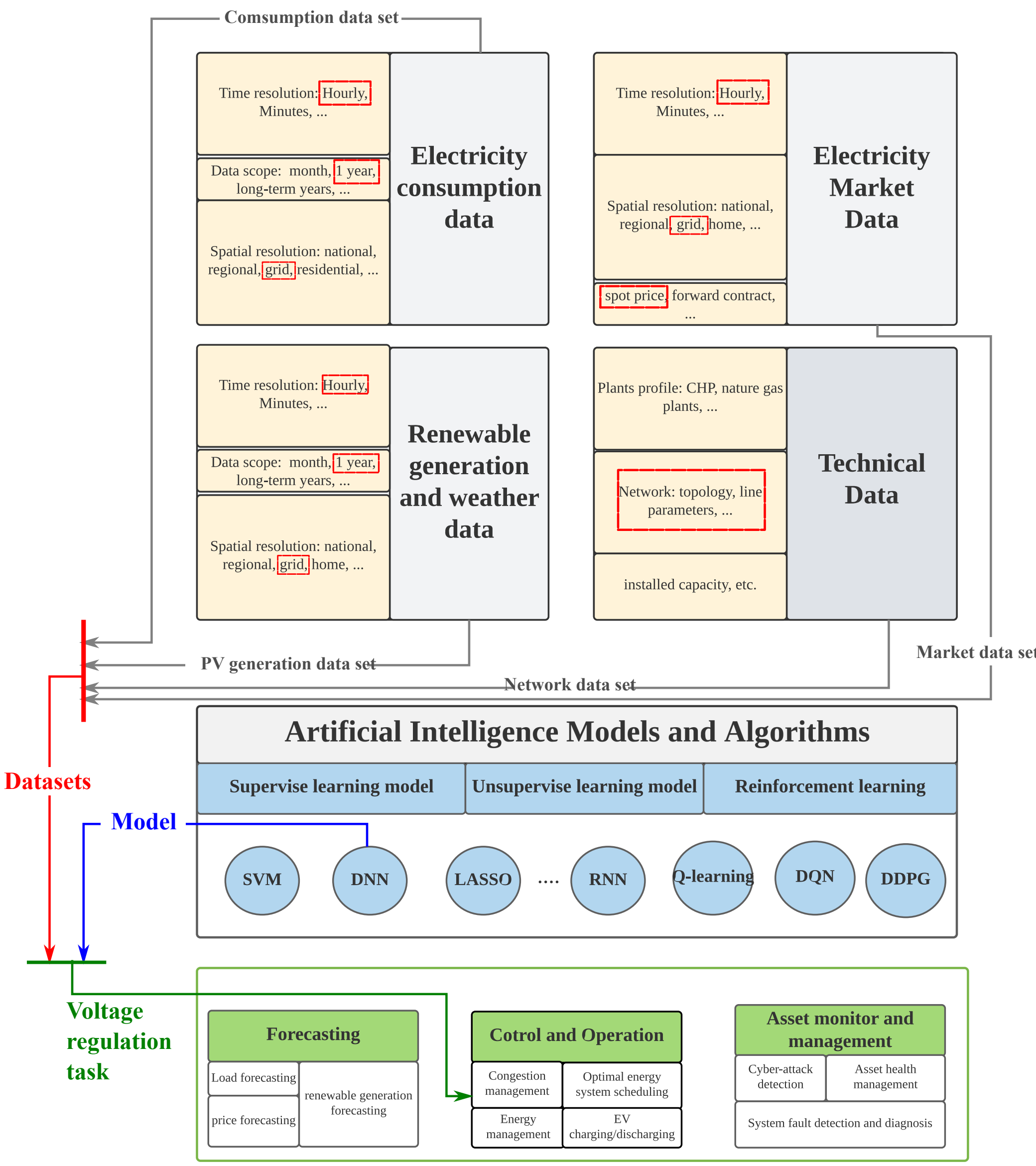

Figure 2: Working Pipeline for data-driven model

5.3 Data Visualization

The platform contains a data dashboard to illustrate datasets and evaluation metrics with figures. The dashboard contains two types of contexts. First, raw open-sourced data can be illustrated by figures in the dashboard. Second, after training the model with data, the performance and other metrics can be plotted by the dashboard.

5.4 Data Management

The platform manages both open and restricted data. It includes the following features:

- Federation data training for restricted data: Private data is controlled by the owners. Trusted algorithms and models are brought to the data, so privacy-preserving computation can take place in the data owners' edge environment. As a result, unnecessary replication, transport, and storage of data are avoided without compromising sovereignty.
- Version control of open data: Open-sourced datasets can be updated with different versions by the owner. After the data quality is checked and validated, it can be published again with the same label but using a different version signal.
- Data integration: Open-sourced datasets can be integrated by the platform based on meta-data compatibility to eliminate the data fragmentation problem.

5.5 Community function

A group or forum is contained in the platform to provide the following functions:

- Users can upload, update data, and share data in an open or not open way.
- Users can share experiences of using data or models on this platform.
- Users can make use of these data from open sources or from data owners.

5.6 Documents and external functions

The platform includes the following features:

- Clean implementation for model and data with a specific task listed in Table 1.
- Guidance and documents for the platform.
- Beginner-friendly interface for training model with data: The platform contains a graphical operation interface for beginners to pick up a task, then choose a model and data related to the task to solve it.
- External resources links: External data and model resources and websites are listed and updated in our platform.

Table 4: Features Comparison for Representative Platforms

Platforms	Data quality check	Meta data	Filter	Data visualization	Task and model	Fedeartion Training	Version control	Data Integration	Discussion group	Distributed data uploading	Documents
OpenPowerSystemData	✓	✓	O	O	O	O	✓	O	O	✓	✓
ENTSO-E Transparency	O	O	✓	✓	O	O	O	✓	✓	O	O
energinet dk	✓	O	✓	O	O	O	O	✓	O	O	O
RTE	✓	O	✓	✓	O	O	O	✓	O	O	O
Ninja renewable energy	O	O	✓	O	O	O	O	✓	O	O	✓
Europe Gaint	O	✓	✓	O	✓	✓	✓	O	O	O	O
FlexMeasures	✓	O	O	✓	✓	O	O	O	O	✓	✓
RLOESS	O	O	O	✓	✓	O	O	O	O	O	✓
OpenML	O	✓	✓	O	✓	O	✓	✓	✓	✓	✓

Our proposed platform for AI-based energy transition research will be an open data and model platform that integrates data, models, and tasks. The platform will be designed to promote data sharing, transparency, and collaboration among researchers, practitioners, and industry players.

To achieve this goal, the platform will be built on four key pillars:

Data integration and management: The platform will have a centralized data lake that integrates various datasets from different sources, such as open data portals, industry partners, and research institutions. The data lake will be managed by a dedicated team to ensure data quality, metadata management, and data sovereignty.

Model repository: The platform will also have a model repository that contains various AI models and algorithms for energy transition research. The repository will be open to the community, and researchers can contribute their models to the platform. The repository will have a version control system to ensure that the latest and most relevant models are available to users.

Task management and collaboration: The platform will have a task management system that enables researchers to create, share, and collaborate on research tasks. Tasks can range from data preparation and cleaning to model training and validation. The task management system will also include a project management tool that enables users to track the progress of their research projects.

AI-powered analysis and visualization: The platform will use AI algorithms to perform advanced analysis and visualization of the data and models. Users will be able to explore the data and models in a user-friendly interface and gain insights into energy transition research. The platform will also support interactive visualization tools to enable users to visualize and communicate their research results.

The platform will be accessible through a web-based interface that is easy to use and navigate. It will be open to the community, and users will be able to register and create their profiles. The platform will also have a forum section where users can share their ideas, ask questions, and discuss their research findings.

In conclusion, our proposed platform for AI-based energy transition research will be a comprehensive and collaborative platform that integrates data, models, and tasks. It will be designed to promote data sharing, transparency, and collaboration among researchers, practitioners, and industry players. The platform will use AI algorithms to perform advanced analysis and visualization of the data and models, enabling users to gain insights into the energy transition research.

References

[1] C. Croonenbroeck and G. Stadtmann, "Renewable generation forecast studies–review and good practice guidance," *Renewable and Sustainable Energy Reviews*, vol. 108, pp. 312–322, 2019.

[2] I. K. Nti, M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya, "Electricity load forecasting: A systematic review," *Journal of Electrical Systems and Information Technology*, vol. 7, no. 1, pp. 1–19, 2020.

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

[4] R. R. Sharma, "Design of distribution transformer health management system using iot sensors," *Journal of Soft Computing Paradigm*, vol. 3, no. 3, pp. 192–204, 2021.

[5] S. S. Refaat and M. A. Shams, "A review of partial discharge detection, diagnosis techniques in high voltage power cables," in *2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018)*. IEEE, 2018, pp. 1–5.

[6] Y. Zhao, T. Li, X. Zhang, and C. Zhang, "Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future," *Renewable and Sustainable Energy Reviews*, vol. 109, pp. 85–101, 2019.

[7] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, and A. Amira, "Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives," *Applied Energy*, vol. 287, p. 116601, 2021.

[8] J. Jasiūnas, P. D. Lund, and J. Mikkola, "Energy system resilience–a review," *Renewable and Sustainable Energy Reviews*, vol. 150, p. 111476, 2021.

[9] G. M. Messinis and N. D. Hatziargyriou, "Review of non-technical loss detection methods," *Electric Power Systems Research*, vol. 158, pp. 250–266, 2018.

[10] Y. Ye, D. Qiu, M. Sun, D. Papadaskalopoulos, and G. Strbac, "Deep reinforcement learning for strategic bidding in electricity markets," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1343–1355, 2019.

[11] S. Gumpu, B. Pamulaparthy, and A. Sharma, "Review of congestion management methods from conventional to smart grid scenario," *International Journal of Emerging Electric Power Systems*, vol. 20, no. 3, 2019.

[12] D. Mariano-Hernández, L. Hernández-Callejo, A. Zorita-Lamadrid, O. Duque-Pérez, and F. S. García, "A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis," *Journal of Building Engineering*, vol. 33, p. 101692, 2021.

[13] H. Shengren, E. M. Salazar, P. P. Vergara, and P. Palensky, "Performance comparison of deep rl algorithms for energy systems optimal scheduling," *arXiv preprint arXiv:2208.00728*, 2022.

[14] A. Fu, M. Cvetković, and P. Palensky, "Distributed cooperation for voltage regulation in future distribution networks," *IEEE Transactions on Smart Grid*, 2022.

[15] P. Vongvilasack, S. Premrudeepreechacharn, and K. Ngamsanroaj, "The application of machine learning for the voltage and reactive power control in power distribution network," in *2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2022, pp. 1–6.