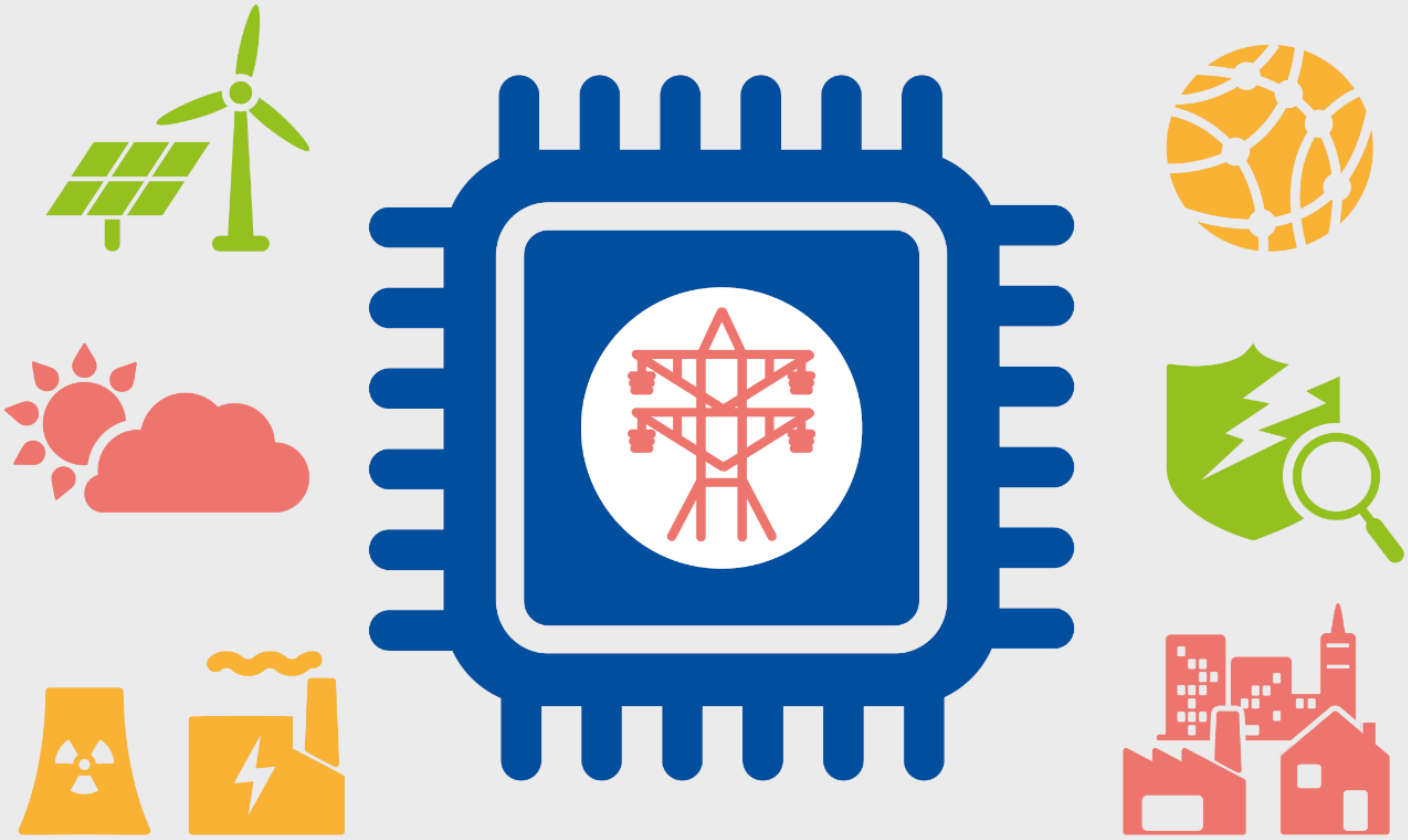
EUROPEAN UNION AGENCY
FOR CYBERSECURITY



CYBERSECURITY AND PRIVACY IN AI – FORECASTING DEMAND ON ELECTRICITY GRIDS

JUNE 2023

ABOUT ENISA

The European Union Agency for Cybersecurity, ENISA, is the Union's agency dedicated to achieving a high common level of cybersecurity across Europe. Established in 2004 and strengthened by the EU Cybersecurity Act, the European Union Agency for Cybersecurity contributes to EU cyber policy, enhances the trustworthiness of ICT products, services and processes with cybersecurity certification schemes, cooperates with Member States and EU bodies, and helps Europe prepare for the cyber challenges of tomorrow. Through knowledge sharing, capacity building and awareness raising, the Agency works together with its key stakeholders to strengthen trust in the connected economy, to boost resilience of the Union's infrastructure, and, ultimately, to keep Europe's society and citizens digitally secure. More information about ENISA and its work can be found here: www.enisa.europa.eu.

CONTACT

For contacting the authors, please use info@enisa.europa.eu
For media enquiries about this paper, please use press@enisa.europa.eu.

EDITORS

Monika Adamczyk, Apostolos Malatras, Ioannis Agrafiotis, ENISA

LEGAL NOTICE

COPYRIGHT NOTICE

ISBN: 978-92-9204-620-0 – DOI: 10.2824/92851 – Catalogue Nr: TP-09-23-010-EN-N

TABLE OF CONTENTS

1. INTRODUCTION 6

1.1 STUDY OBJECTIVES 6

1.2 METHODOLOGY 7

1.2.1 Description of the scenario 7
1.2.2 Identification of cybersecurity and privacy threats and vulnerabilities 7
1.2.3 Identification of cybersecurity and privacy controls 7

1.3 TARGET AUDIENCE 7

1.4 USING THIS DOCUMENT 8

2. SCENARIO DESCRIPTION 9

2.1 PURPOSE AND CONTEXT 10

2.2 HIGH-LEVEL DESCRIPTION 10

2.3 ACTORS AND ROLES 11

2.4 PROCESSED DATA 12

2.5 MACHINE LEARNING ALGORITHMS 14

2.6 ASSETS 14

2.7 OVERALL PROCESS 15

2.8 PRIVACY AND CYBERSECURITY REQUIREMENTS 20

3. SECURITY AND PRIVACY THREATS AND VULNERABILITIES 24

3.1 THREAT CONTEXTUALISATION 24

3.1.1 Compromise of ML application components 25
3.1.2 Poisoning 25
3.1.3 Human error 26
3.1.4 Data disclosure 26
3.1.5 Unlawful Processing 26
3.1.6 Unfair processing 26
3.1.7 Lack of transparency 27
3.1.8 Diversion of purpose 27
3.1.9 No respect of data minimisation 27
3.1.10 No respect of storage limitation 27
3.1.11 Synthesis of possible impacts and associated threats 27

3.2 VULNERABILITIES ASSOCIATED TO THREATS AND AFFECTED ASSETS 28

4. CYBERSECURITY AND PRIVACY CONTROLS 33

4.1 IMPLEMENT A SECURITY BY DESIGN PROCESS 34

4.2 DOCUMENT THE ELECTRICAL FORECAST SYSTEM 34

4.3 CHECK THE VULNERABILITIES OF THE ML COMPONENTS AND IMPLEMENT PROCESSES TO MAINTAIN THEIR SECURITY LEVELS OVER TIME 35

4.4 CHOOSE AND DEFINE A MORE RESILIENT MODEL DESIGN 35

4.5 INTEGRATE POISONING CONTROL IN THE TRAINING DATASET 36

4.6 ENLARGE THE TRAINING DATASET 36

4.7 SECURE THE TRANSIT OF THE COLLECTED DATA 37

4.8 CONTROL ALL DATA USED BY THE ML MODEL 37

4.9 ENSURE RELIABLE SOURCES ARE USED 38

4.10 IMPLEMENT ACCESS RIGHT MANAGEMENT PROCESS 38

4.11 ENSURE ALL SYSTEMS AND DEVICES COMPLY WITH AUTHENTICATION, AND ACCESS CONTROL POLICIES 39

4.12 REDUCE THE AVAILABLE INFORMATION ABOUT THE MODEL 40

4.13 IDENTIFY A DATA CONTROLLER FOR THE ENERGY CONSUMPTION ANTICIPATION DATA PROCESSING 40

4.14 PROPERLY COLLECT AND MAINTAIN USER CONSENT WHEN NEEDED FOR DETAILED ENERGY CONSUMPTION USAGE 41

4.15 ANONYMIZE DATA COMING FROM THE CONCENTRATOR 41

4.16 GENERATE LOGS AND PERFORM INTERNAL AUDIT 42

4.17 PERFORM A PRIVACY IMPACT ASSESSMENT 42

4.18 DEFINE AND IMPLEMENT A DATA RETENTION POLICY 43

4.19 STUDY ON DATA FIELDS NECESSITY AND JUSTIFICATION IN THE PRIVACY POLICY 43

4.20 FORMALISE A LIA (LEGITIMATE INTEREST ASSESSMENT) 44

4.21 MINIMISE DATA AT EACH STEP OF THE PROCESSING; COLLECT ONLY WHAT IS NEEDED WHEN NEEDED 44

4.22 IMPLEMENT A PRIVACY BY DESIGN PROCESS 45

4.23 RAISE AWARENESS OF SECURITY AND PRIVACY ISSUES AMONG ALL STAKEHOLDERS 45

4.24 SUMMARY 46

5. CONCLUSION 49

ANNEX I: SECURITY AND PRIVACY SCALES AND REQUIREMENTS 50

A.1 CYBERSECURITY AND PRIVACY SEVERITY SCALES 50

A.2 CYBERSECURITY SCALE OF IMPACT 51

A.3 PRIVACY SCALE OF IMPACT 51

A.4 PRIVACY REQUIREMENTS CRITERIA 52

EXECUTIVE SUMMARY

Given the great influence of artificial intelligence (AI) in people's daily lives due to the key role it plays in digital transformation through its automated decision-making capabilities, ENISA aims to raise awareness of cybersecurity and privacy threats related to various scenarios using artificial intelligence. To this end, ENISA, with the support of the Ad-Hoc Working Group on Artificial Intelligence Cybersecurity, has published two reports in the last two years: **Cybersecurity Challenges of Artificial Intelligence**[1] and **Securing Machine Learning Algorithms**[2].

ENISA continues its momentum with a new report on cybersecurity and privacy in forecasting demand on electricity grids. An in-depth study of the scenario has been conducted by identifying first the assets, the actors and their roles, relevant processes, the AI algorithms used, as well as the requirements in terms of cybersecurity and privacy needed for it. Building upon previous ENISA work such as the “Securing Machine Learning Algorithms” report cited above, in addition to legislation such as GDPR and literature searches, this report has identified cybersecurity and privacy threats and vulnerabilities that can be exploited in the examined scenario. While focus is on ML-related threats and vulnerabilities, broader AI considerations were also taken into account. Lastly, corresponding cybersecurity and privacy controls that consider the context of the scenario and the impact of the associated threats/vulnerabilities were defined. The specificities within the implementation of these controls are described, including possible trade-offs between cybersecurity, privacy, and performance. Each control is classified as a cybersecurity control, a privacy control, or a mixture of both, depending on the threats it mitigates and their associated impacts (cybersecurity impacts, privacy impacts, or both).

This report allows better assessment of the reality that artificial intelligence brings its own set of threats, which consequently insists on the search for new security measures to counter them. Finally, it should be noted that this guide strongly emphasises privacy issues in the same way as cybersecurity issues, privacy being one of the most important challenges facing society today. Security and privacy are intimately related, but both equally important, and a balance must be made specific to each use. As a result, as seen in this report, efforts to optimise security and privacy can often come at the expense of system performance.

[1] See https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges
[2] See https://www.enisa.europa.eu/publications/securing-machine-learning-algorithms

1. INTRODUCTION

Abuse of Artificial Intelligence (AI), which has developed significantly in recent years, has been identified by ENISA as one of the top emerging threats[3]. By providing new opportunities to solve decision-making problems intelligently and automatically, AI is being applied to more and more business cases in a growing number of sectors. The associated benefits are significant. However, the development of AI is also accompanied by new threats which project teams will have to face.

In many projects, it is apparent that these new threats are related to aspects of cybersecurity in addition to privacy, particularly when AI is used for innovative projects whereby processing personal data takes place. In this spirit, on 21 April 2021, the Commission published the AI Act proposal[4], which sets out requirements (including cybersecurity and privacy) for AI systems deemed to be high-risk (e.g., used in biometric identification or critical infrastructure management and operation) in order to mitigate threats to health, safety, and fundamental rights.

To go further into these high-risk AI systems, ENISA proposes this new report: "Cybersecurity and Privacy in AI - Forecasting Demand on Electricity Grids", which builds directly on the work already initiated by ENISA since 2020 on the identification of risks associated with AI. **This new report analyses cybersecurity and privacy requirements and measures in use of AI in forecasting demand on electricity grids. The report describes the scenario fundamental principles (assets, actors processes etc.), identifies the security and privacy risks it poses, and finally cybersecurity and privacy controls, which counteract the identified risks.**

1.1 STUDY OBJECTIVES

Findings from the ENISA's report on securing machine learning algorithms[5] indicate that there is no uniform strategy in applying a specific set of security controls to protect machine learning algorithms and in some cases, deployed security controls may result in trade-offs in security and performance. ENISA therefore recommends that organisations which use AI systems should perform detailed analysis of their own AI systems, and conduct targeted risk assessments to find the appropriate balance between cybersecurity, privacy and performance.

The objectives of this publication are as follows:

- Provide a detailed description of a **Forecasting Demand on Electricity Grids** scenario.
- Identify AI cybersecurity and privacy measures taking into account requirements, threats and vulnerabilities defined for this scenario and practical guidance on how to implement them.
- Provide recommendations on how to balance the trade-offs between cybersecurity, privacy, and performance in this scenario.

The following section outlines the methodology followed in producing this report.

[3] See https://www.enisa.europa.eu/news/cybersecurity-threats-fast-forward-2030
[4] See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206
[5] See https://www.enisa.europa.eu/publications/securing-machine-learning-algorithms

1.2 METHODOLOGY

Production of this publication, was undertaken in three stages:

- Identify and describe in detail the forecasting demand on electricity grids scenario.
- Identify cybersecurity and privacy threats associated with the scenario.
- Identify relevant cybersecurity and privacy controls.

1.2.1 Description of the scenario

The scenario description is aligned with ENISA's previous work on AI[6],[7] and provides the following information:

- The purpose and the context
- A high-level description of the scenario, highlighting the data encountered
- The involved actors and their associated roles
- A detailed description of the data
- Machine learning algorithms
- Other assets (besides data) associated with the described scenario
- The overall process of the scenario
- Security and privacy requirements

1.2.2 Identification of cybersecurity and privacy threats and vulnerabilities

This section of the report focusses on the threats and vulnerabilities related to use of AI in forecasting demand on electricity grids scenario. Based on previously mentioned ENISA work, legislation such as GDPR, and desk research, this report identified the cybersecurity and privacy threats and vulnerabilities that can be exploited. Given the prevalence of machine learning (ML) and ENISA's past work on the topic, there is more emphasis placed on threats and vulnerabilities related to ML. Nonetheless, wider considerations of AI have been taking into account when identifying security threats and vulnerabilities as well as dedicated to privacy, for which GDPR data protection principles[8] were used as a starting point of our analysis.

1.2.3 Identification of cybersecurity and privacy controls

Following the analysis of forecasting electricity on electricity grids scenario, the identification of threats (and their impact), and associated vulnerabilities, this section of the report presents the corresponding cybersecurity and privacy controls that:

- Take into account the context of the scenario
- Take into account the security and privacy impact of the threat/vulnerabilities (as described in this report)

The specificities of implementation of such controls are described including the possible trade-offs between cybersecurity, privacy and performance. Each control can be either a cybersecurity control, a privacy control, or a mixture of both, depending on the threats it mitigates and their associated impacts (cybersecurity impacts, privacy impacts, or both). Many of the controls are of technical nature, but when appropriate (e.g., per GDPR requirements), organizational controls have also been identified.

1.3 TARGET AUDIENCE

The target audience of this report can be divided into the following categories:

[6] See https://www.enisa.europa.eu/publications/securing-machine-learning-algorithms
[7] See https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges
[8] See Article 5 of GDPR

- **All actors (private or public):** to help them in their risk analysis, in the identification of cybersecurity and privacy threats and in the identification of the appropriate security and privacy controls to mitigate the threats related to this scenario
- **AI technical community, AI cybersecurity and privacy experts and AI experts** (designers, developers, ML experts, data scientists, etc.) with an interest in developing secure solutions and in integrating security and privacy by design in their solutions.
- **Cybersecurity and privacy community**: to help in identifying cybersecurity and privacy threats related to forecasting demand on electricity grids scenario and in identifying the appropriate security and privacy controls to mitigate the threats.

1.4 USING THIS DOCUMENT

Although we based the scenario close to reality, some assumptions have been to made for the purposes of its analysis. These assumptions should be reassessed by the reader when doing their own analysis. Please note this list is non-exhaustive:

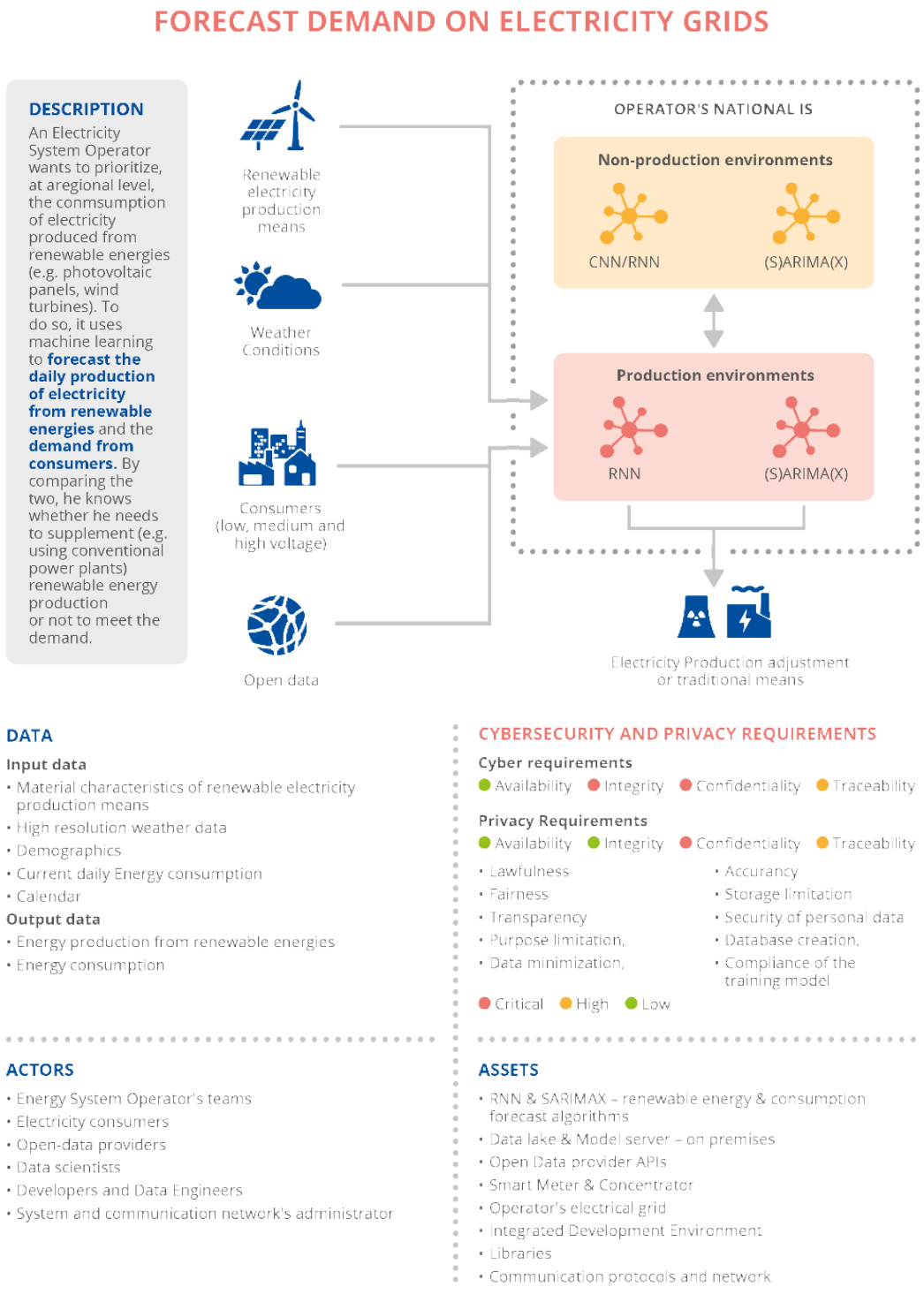
- The algorithms chosen for the scenario are based on desk research, and there may be other algorithms that are better suited for it.
- Regarding the privacy aspects, we put ourselves in the shoes of the data controller, but only based on the hypothesis that we have made. It is left to the attention of the audience that the privacy requirements as well as the control measures must always be adapted to a context and a situation.
- After identifying potential threats, this report identifies security and privacy controls that could be applied to the scenario. However, as is the case for any application of ML, one must also consider traditional security standards (e.g., ISO 27001/2, NIST) because ML applications are also subject to more global threats.

It should also be kept in mind that the elements of this report are valid as of the date of publication, and could evolve over time.

2. SCENARIO DESCRIPTION

The following figure presents an overview of the topics that will be addressed in this chapter.

Figure 1: Scenario overview



2.1 PURPOSE AND CONTEXT

Consumption and production balance is the core challenge of any smart electricity grid. This is particularly the case since the rise of intermittent renewable energies that help to rebalance the energy mix and reduce reliance on energy production from fossil fuels. However, the inherent intermittence of renewable energies requires distributors to take into account precise predictions to ensure that production is able to meet consumption, even during peak hours. Moreover, to this day there is no way to store energy on a large scale that could allow the temporary separation of production to meet current consumption.

For this reason, a national electricity system electricity supplier[9] has defined the following purposes:

- reduce the production of electricity from non-renewable energies (conventional power plants, etc.),
- prioritise the consumption of electricity produced from renewable energies (photovoltaic panels, wind turbines, etc.),
- Ensure all technical conditions for the supply of electricity, including the need to maintain supply with fluctuating demand.

To reach these objectives, the national electricity system electricity supplier seeks to make daily predictions on the amount of non-renewable energy production required to supplement the amount of energy produced by renewable sources so that the total energy produced matches the estimated consumer demand. This is done on a daily basis in three steps:

- Predict the production of renewable sources using meteorological data (history and forecast), data from the materials (mechanical / electronical characteristics of the solar panels and the windmills) and historical data concerning the previous power production from power farms.
- Forecast consumption of the consumers with demographic data, calendar data (events, holidays), and energy consumption.
- Compare predicted renewable production against consumption forecast. This provides an important key performance indicator (KPI) for the national electricity system electricity supplier teams to make a decision as to whether additional power plants may be required to supplement the renewable energy, or if the energy produced by renewable sources may be stored.

2.2 HIGH-LEVEL DESCRIPTION

The technical core of the scenario is therefore prediction, with the use of artificial intelligence algorithms. To be more precise, machine learning is used along with the **supervised learning** paradigm (e.g., multivariate regression or neural networks).

To achieve these objectives, a dataset is used, amongst which, material characteristics of solar panels and wind turbines, weather data, energy consumption of consumers, calendar data and demographics (regional information about the average age of the population, employment, number of inhabitants etc.). Most of these data are obtained through public application programming interfaces (APIs), with the remaining data already in possession of the electricity provider. Once **collected in a data lake** via data engineers and network administrators, the data must be explored and prepared. This is aimed at better understanding the data, evaluating the quality and quantity at disposal and adopt a modelling strategy. The data scientist and engineering teams are responsible for these steps. The electricity supplier's teams, especially

[9] In this analysis, we have assigned to one actor the roles of several actors. In many countries, there are different actors such as: the distributed system operator, market operator, energy producers

business experts, may be required to give insights about the data collected and help decide which strategy to choose.

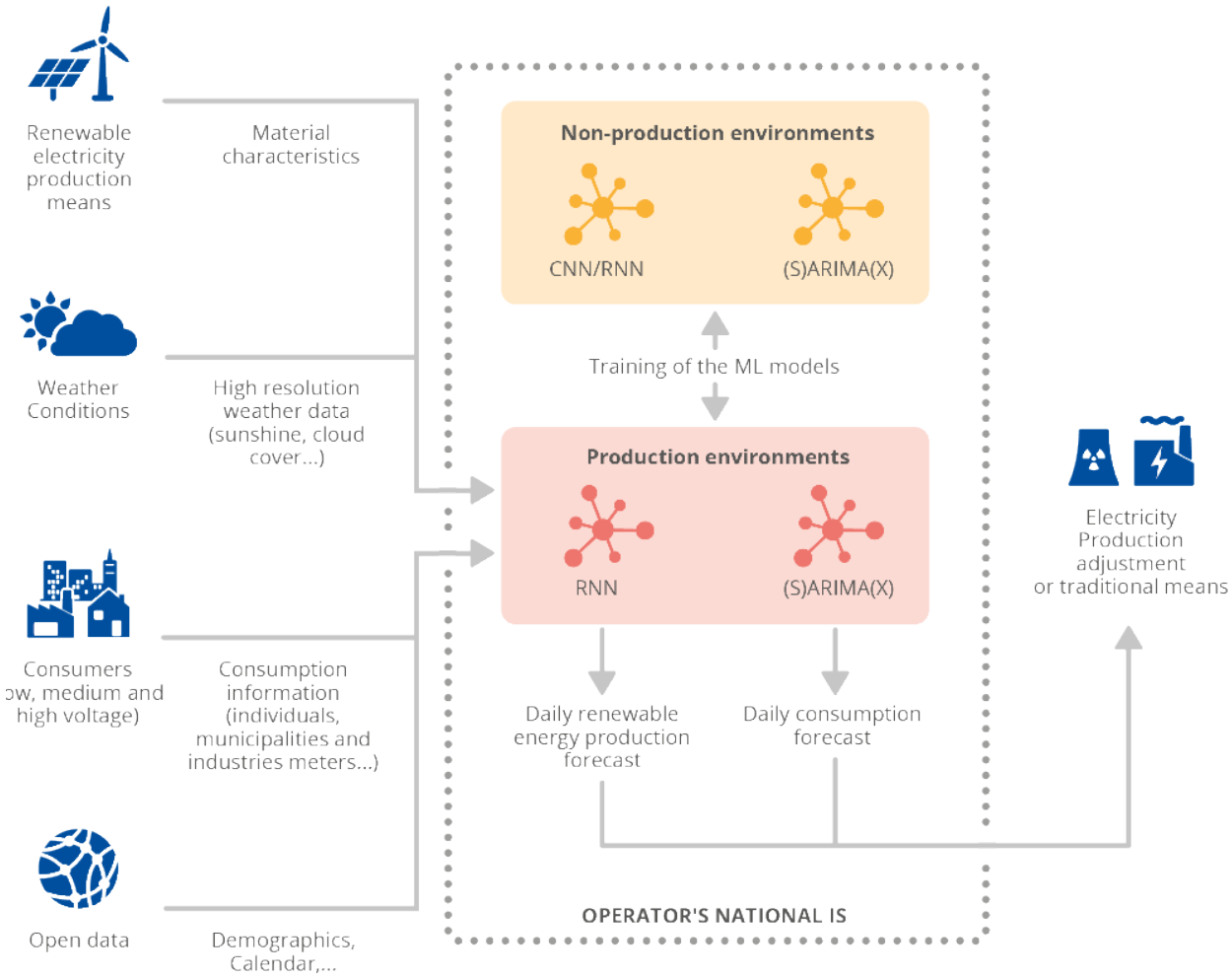
Selected data are then used to create two machine learning models. A data scientist team is responsible for this step. One model is built to determine the next day's electricity production from renewable electricity production facilities, while the other is built to determine the next day's electricity consumption. The two models are described in detail in 1.9. **These two models produce the following outputs**:

- a daily prediction of the renewable energy production
- a daily energy consumption of facilities (inhabitants, industrials etc.)

At the end of the prediction processes, predictions are fed back into a **dashboard tool** used by the electricity provider teams to adjust their power procurement decisions.

The following sections aim to detail such scenario by describing actors and their roles, data processed, the assets that allow the project to exist, the explanation of the scenario lifecycle, and the privacy and cybersecurity requirements applied to it.

Figure 2: High-level description



2.3 ACTORS AND ROLES

The following actors are involved in the forecasting demand on electricity grids scenario.

Figure 3: Actors, roles, and their description

Actor	Role	Description
Electricity supplier's teams	*End Users and Data Owner (**Data Controller**)*	Electricity System Electricity supplier's teams oversee controlling the electricity supply so that it is optimal at all times, knowing that energy not consumed is energy wasted. In this scenario, the team interprets the model's results, provides requirements to data scientists, and has control over the solar and wind infrastructure that provides data inputs into the model.[10]
Electricity consumers	*Data Provider*	Electricity consumers are indirect participants. Their electricity consumption, whether individual or not, is reported through the network of electricity meters, energy suppliers, etc.
Open-data providers	*Data Provider*	Open data providers are an external entity providing open-data, including demographics, and weather forecasts of solar irradiance, cloud cover and wind speed, amongst other metrics.
Data scientists	*Data scientist*	Data scientists are the spearhead of the algorithms. They oversee cleaning and preparing of the data, building the models and ensuring the relevance of the results. They work with the Electricity Distribution Electricity supplier's teams to build a relevant model.
Developers and Data Engineers	*Developers and Data Engineers*	Developers and Data Engineers are responsible for routing data, transformation, production environments and other technical operations as required.
System and communication network administrators	*Network administrators*	System and communication network's administrator are responsible for the installation and the maintenance of digital systems supporting the machine learning application, including network (flows etc.) and hosting of on-premises servers that make the application work.

2.4 PROCESSED DATA

As previously mentioned, there are two models, and each uses separate data. It should also be noted that the output data of one model (predicted data) becomes the input data of the same model the next day.

Figure 4: Data needed to build the model and input data that the model will use for the electricity production from renewable energies

[10] Usually, these renewable sources are dispatched by the system operator.

Data	Data type	Source / data provider	Data Procurement
Model 's input data			
Material characteristics of renewable electricity production means (e.g., photovoltaic panels, wind turbines)	Structured data	Owners of the renewable electricity production means, i.e., **the electricity system electricity supplier**	These data can be obtained from the results of the regular maintenance of the electricity supplier's equipment.
High resolution weather data (solar irradiance, cloud cover and wind speed over time)	Time series	Official meteorological and climatological services, governments as **Open-data provider**	Several APIs can provide these data for every renewable electricity production means. They can provide historical data and prediction over the coming hours or days[11].
Model's output data			
Energy production from renewable energies	Time series	Owners of the renewable electricity production means, **i.e., the electricity system electricity supplier**	These data are known for every renewable electricity production means. The collection of this information for every plant will give a global overview of the electricity data.

Figure 5: Data needed to build the model and input data that the model will use the consumers' energy demand

Data	Data type	Source / data provider	Data Procurement
Model's input data			
Demographics (statistical data relating to the size of the population attending the area)	Structured data	Public APIs from national statistical institutes as **Open-data provider**	Several APIs can provide these data for every city and occasionally for each neighbourhood. For instance, the INSEE[12] provide the number of inhabitants and their repartition for several geographical scales in France. The difficulty is to determine which data are impactful to avoid the model becoming over-parameterised.
Calendar (calendar of events, holydays... to help predict times of the year when consumption is high or lower than average)	Structured data	**Open-data provider**	These data are public and can be found in open data platforms. The difficulty is to determine which data are impactful to avoid the model becoming over-parameterised. This information can be found and must be expressed in a way that the system can understand[13].
Current daily energy consumption	Time series	Consumers' (including industrials, municipalities, and individuals) smart meters belonging to **the electricity system electricity suppliers**. These smart meters provide surname, first name, address, meter number, telephone number, and electricity consumption.[14]	The electricity supplier can collect the energy data from smart meters installed at the consumers' site. Smart meter data is sent to a concentrator, located at a secondary substation (medium voltage - low voltage) where all data from all the individuals are gathered[15].
Model's output data			

[11] See https://api.meteo-concept.com/documentation_openapi
[12] See https://api.gouv.fr/les-api/api_donnees_locales
[13] **Elamin, Niematallah et Fukushige, Mototsugu.** Modeling and forecasting hourly electricity demand by SARIMAX with interactions. 2018. https://www.sciencedirect.com/science/article/abs/pii/S0360544218319297
[14] Smart meters have an ID, which is related to a name, addressee etc at the concentrator level.
[15] See https://api.meteo-concept.com/documentation_openapi

Energy consumption	Time series	Consumers' (including industrials, municipalities, and individuals) smart meters belonging to **the electricity system electricity suppliers**. These smart meters provide surname, first name, address, meter number, telephone number, electricity consumption[16]	As highlighted, the data that will be used to train the model are collected in the following way: the electricity supplier can collect the energy data from smart meters installed at the consumers' site. Smart meter data is sent to a concentrator, located at a secondary substation (medium voltage - low voltage) where all data from all the individuals are gathered[17].

2.5 MACHINE LEARNING ALGORITHMS

The following algorithms are used in our scenario. The first algorithm (RNN) is used to forecast the daily production of electricity from renewable energies. The second algorithm (SARIMAX) is used to forecast the daily consumer demand.

Figure 6: Machine learning algorithms used

Learning paradigm	Subtype	Algorithm	Type of data ingested	Description
Supervised Learning	Regression	RNN	Time series	A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. This allows it to exhibit temporal dynamic behaviour[18] [19] [20].
Supervised Learning	Regression	SARIMAX	Time series Structured Data	Given a time series, the ARIMA model is a tool to understand and predict the future values of this series. The model is composed of two parts: an autoregressive part (AR) and a moving average part (MA). A SARIMAX algorithm has additional seasonality modelling (S) and exogenous variables (X) in the model[18] [19] [20].

2.6 ASSETS

In addition to the previously described data, the scenario is supported by the following additional assets.

[16] As we will see, the output training data comes from personal data sources. However, once the model is in production, it only predicts the overall energy consumption.
[17] See https://api.meteo-concept.com/documentation_openapi
[18] **Neda Tavakoli; Sima Siami-Namini; Akbar Siami Namin.** A Comparison of ARIMA and LSTM in Forecasting Time Series https://scholars.ttu.edu/en/publications/a-comparison-of-arima-and-lstm-in-forecasting-time-series-5
[19] **B, Prabadevi, et al.** Deep Learning for Intelligent Demand Response and Smart Grids: A Comprehensive Survey https://arxiv.org/abs/2101.08013
[20] **Abualig, Laith, et al.** Wind, Solar, and Photovoltaic Renewable Energy Systems with and without Energy Storage Optimization: A Survey of Advanced Machine Learning and Deep Learning Techniques https://www.mdpi.com/1996-1073/15/2/578

Figure 7: Asset's description

Type of asset	Asset	Description
Models	RNN, daily renewable energy prediction forecast	The model takes as inputs: history of energy production from renewable energies, meteorological data, and material characteristics of renewable electricity production means and the model learns to predict energy production using these data as standalone inputs.
	SARIMAX, daily consumption forecast	The model takes as inputs: real-time energy consumption of inhabitants, demographics and calendar data and the model learns to predict energy production using these data as standalone inputs.
Environment tools	Data lake – *on-premises*	The data ingestion platforms that enable data engineers to store data and exploit them.
	Open-data provider APIs	The different APIs used by the model to obtain data (weather, demographic, calendar).
	Smart meters	The smart meter is a meter that measures accurately and in real-time the various energy consumptions (water, electricity, gas). The meter directly transmits the various information to a backup system allowing the analysis of the collected data.
	Concentrator	A termination point where all data collected by the smart meters in the same area are recorded. The data are then aggregated at this point to form a global consumption.
	Model server – *on-premises*	The model server is used to store the model and process the training and retraining phases.
	Dashboard tools – *on-premises*	Tool monitoring electricity production and consumption. The two models are incorporated in this tool.
	Electricity supplier's electrical grid	Data collected by smart meters is communicated through the electrical grid (Power Line Communications – PLC) before reaching the concentrator.
	Integrated Development Environment	Software application that provide comprehensives facilities to computer programmers.
	Programming Libraries (with algorithms for transformation…)	A collection of precompiled codes that can be used in a project to realise well-defined operations.
	Communication protocol and communication Networks	A system of rules allowing a group of entities to communicate with each other to share or ask information.

2.7 OVERALL PROCESS

As a reminder, the national electricity system electricity supplier seeks to predict the amount of daily "traditional" energy to produce in addition to that produced by renewable energies to satisfy consumer estimated demand.

To carry out this scenario, electricity supplier's teams, data scientists, data engineers and the system administrator work together to understand the final objectives and define the needs.

In the following paragraphs, the scenario is described by incorporating it into the different stages of the machine learning lifecycle:

Data collection

The first step is to determine how to obtain the needed data and how to incorporate it in the model. Data engineers/developers and network administrators route the data used in the models. The former to format it, to know where to retrieve it, to validate its quality and the latter to build the necessary flows as these data are routed through various channels (public APIs, individual consumer meters provided by the electric company). Of course, even once in production, the data must be routed to the models to make their predictions. Data are routed to an existing on-premises (belonging to the electricity supplier) data lake, which is a set of servers and resources specifically designed to store and orchestrate the data, and to define a set of data tables.

Most of these data **(material characteristics, and production history**) are owned by the **electricity supplier**. The electricity supplier can collect this data since it knows the amount of equipment at its disposal, the state of these facilities, and its production history. To carry out its project, the electricity supplier collects the data it needs from the results of the regular maintenance of its equipment and from its production history. As mentioned earlier, **weather data** has an important part to play in this scenario. It greatly influences the amount of electricity produced by renewable energies. The prediction of these data is the expertise of the **meteorological services**. The latter make these data accessible through APIs that they provide and maintain. **This data is considered as open-data and is therefore non-proprietary and free to use.** The electricity supplier collects this data using the APIs made available.

An important source of data collected is the **electricity consumption of the inhabitants**. The consumption data is considered personal data because it includes personal information such as name, surname, address, smart meter number and phone number of account holder in addition to the consumption. **The electricity supplier collects their electricity consumption from smart meters installed locally at the consumers' place**[21]. By default, only the daily consumption data (global consumption of the household over a day) is collected to allow the consumers to consult the history of his consumption free of charge. However, smart meters can collect more detailed consumption information. This detailed consumption information is the hourly or half-hourly consumption data. The collection of this detailed consumption data is not automatic. The electricity supplier does not collect detailed consumption data for all households by default, however these may be collected with the user's consent.

This data then passes through the Electricity supplier's electrical grid - PLC to a concentrator located at a secondary sub-station (medium voltage - low voltage). Once this information arrives at the concentrator, a first processing is done by data engineers. To protect this **personal data**, in accordance with the General Data Protection Regulation (GDPR), the data engineer must filter the data so that it cannot be used to directly identify individual consumers. This includes:

- the name of the person
- the address of the person
- the meter number of the person
- the phone number of the person

It is also worth noting that smart meters have geolocation, and such information could be used to track the location of consumers living in low density areas. To circumvent this risk, the electricity supplier collects data in areas with a large number of households[22].

The default or detailed consumption values (if the consumer has agreed to share this information for the purpose of the processing) are kept and then aggregated (i.e., summed with all other consumption data) in a large consumption database maintained by

[21] See https://www.cnil.fr/fr/linky-gazpar-quelles-donnees-sont-collectees-et-transmises-par-les-compteurs-communicants
[22] See https://ec.europa.eu/eurostat/documents/3859598/5935825/KS-GQ-13-003-EN.PDF/baa96509-3f4b-4c7a-94dd-feb1a31c7291

the electricity supplier**. This aggregated data does not allow for the retrieval of consumer data. It is therefore anonymised data.**

In addition, open data are also used in this prediction stage. This includes statistical data related to the size of the population present in the area (**demographics**), and the **calendar** of events and holidays, specific to the residents to help to predict times of the year when consumption is high or below average. Demographics can be provided by several APIs and calendars are public, allowing them to be manually located.

Data cleaning and data pre-processing

From there, data scientists use various tools ranging from on-premises servers (to perform calculations) to programming interfaces on their workstations which call upon data science libraries. These tools allow the data scientists to work with the data and build models. The **collected data** is **cleaned** (i.e., ensuring that the data collected is of high quality and exploitable by the model) by data scientists for **pre-processing**[23].

The pre-processing phase consists first of filling the not available (NA) values (interpolation, average, median, etc.) before encoding the categorical variables into numerical values. Finally, it is ensured that time scales are appropriate for the problem at hand (if it is desired for predictions to be made for every hour of the day, data must be scaled to the hour). Moreover, for both models' needs, structured data are converted into time series where possible, keeping in mind the static nature of some variables, such as material characteristics, which are constant over time.

Model design and implementation

Once the data pre-processing done, the types of models to be built must be selected. The models selected are the following:

- **A Recurrent Neuronal Network (RNN)**: to forecast the daily production of electricity from renewable energies.
    - This model was selected because of the known efficiency of these recurrent networks to make predictions for sequential data such as text, sound, or time series. Several papers[18] [19] [20] [24] [25] mention different experiments in which RNN, and in particular Long Short-Term Memory (LSTM), are efficient in forecasting electricity production.
- **A Seasonal AutoRegressive Integrated Moving Average with eXogenous variables (SARIMAX)**: to forecast consumers' demand.
    - Auto-regressive models are very efficient for modelling a temporal phenomenon. Publications[26] [27] expose the benefits of using SARIMAX models to forecast electricity consumption. This approach was chosen because consumer consumption prediction is less complex than meteorological forecasting or production forecasting, and seasonality and impacts of contrasting events such as weekends vs weekdays can be anticipated. This model does not necessarily need a neural network, which can be harder to build and heavier to put in production.

[23] **Brownlee, Jason.** Data Preparation for Machine Learning, Data Cleaning, Feature Selection, and Data Transforms in Python. 2020 https://books.google.fr/books?id=uAPuDwAAQBAJ&printsec=frontcover&hl=fr&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

[24] **Lee, Donghun and Kim, Kwanho.** Recurrent Neural Network-Based Hourly Prediction of Photovoltaic Power Output Using Meteorological Information. 2019. https://www.mdpi.com/1996-1073/12/2/215

[25] **Pavicevic, Milutin and Popovic, Tomo**. Forecasting Day-Ahead Electricity Metrics with Artificial Neural Networks. 2022. https://www.mdpi.com/1424-8220/22/3/1051

[26] **Elamina, Niematallah et Fukushige, Mototsugu**. Modeling and forecasting hourly electricity demand by SARIMAX with interactions. 2018. https://ideas.repec.org/a/eee/energy/v165y2018ipbp257-268.html

[27] **Sim, Sze En, et al.** Forecasting Electricity Consumption Using SARIMA Method in IBM SPSS Software. 2019. https://www.hrpub.org/journals/article_info.php?aid=8614

For the prediction of consumer demand, the **model's parameters** are determined by statistical tests and analysis of the time series for the SARIMAX algorithm.

The characteristics of the RNN input layer (e.g., batch size, time step, number of sequences/features) are defined. The output of the RNN consists of a real value corresponding to the consumer's consumption for the next day.

Model training, model testing and optimisation
During the **training method** phase, model training is realised on-premises. The data is prepared to the same scale at which predictions are to be made. For our scenario, predictions will be made at the hourly level. During the training phase, data scientists collaborate with data engineers and the electricity supplier to improve their understanding of the data. The data is then divided in different categories for the needs of machine learning, 70% is used for training, 15% for validation and 15% for testing[28].

Once the data is prepared, the data scientists choose and elaborate the models which will learn from the inputted data. In our case, two models are elaborated: the first to predict renewable energy production of the next day, the other to predict electricity consumption of the next day.

The prediction of energy production from renewable energies is highly dependent on the weather and the environment in general. To be prepared for any eventuality, the electricity supplier must also train the model for **extreme weather conditions**.

Model Evaluation
To **evaluate the model**, metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are considered.

Model Deployment
Finally, the **model deployment** is done on-premises and the model can be used to predict the future values using new data.

Once the model is ready, the data scientists, data engineers and network teams work together to industrialise the model. It is not only the model that is put to production, but all the pipeline work, from pre-processing the data to model evaluation.

Deploying the environments necessary for its proper functioning: a pre-production environment for the model testing phases in real conditions and a production environment for later. Technically, this consists in creating flows:

- between the data lake and the servers where the model is hosted with the data displayed earlier,
- between the model servers and the existing dashboarding tool with only the prediction results.

When the pre-production environment is stable, the model is put into production and it can be used through the dashboarding tool. From there, operation teams can now use the dashboard tool to adjust their power procurement decisions. The dashboard allows for a comparison of the results of the two predictions from the separate models and provides key performance indicators (KPIs) that help operation teams decide for the next day whether certain power plants should be turned on or whether energy generated from renewable sources can be stored.

Monitoring and inference

[28] **Baheti, Pragati.** Train, Validation, and Test Sets: How to Split Your Machine Learning Data. *v7labs.* [En ligne] 22 May 2022. https://www.v7labs.com/blog/train-validation-test-set.

When **monitoring**, performance is measured every time that new data is acquired, comparing predictions to real data. Data engineers / scientists oversee monitoring of the models in addition to the distribution of the acquired data to infer any potential drift. As the scenario here has an impact on critical infrastructure, there cannot afford to be any underperforming results, even for a short period of time. Considering this, as predictions are made daily, it is reasonable to action weekly re-training of the models. Scheduling of model training processes on a weekly basis is done using a job scheduler. All processing pipeline and model training is automated. Through these automated tests and model evaluation, the model is continuously integrated into the production environment.

Figure 8: Synthesis of the ML lifecycle and involved actors

Steps	Description	Actors	Assets
Data Collection	Data is collected from various sources such as individual meters, public weather and demographics APIs or historical consumption and production data. They are placed in a data pool on-premises before being manipulated. All data collected from the individual smart meters are anonymous after the concentrator and cannot identify the consumer.	Data Engineer/Developers Electricity consumers Electricity System Electricity supplier's teams System and communication network's administrator Open Data provider	Data lake, Concentrator Electricity supplier's electrical Grid Smart meters Open data provider APIs Programming Libraries Integrated development environment
Data Cleaning	The data is cleaned. Useful data is recovered. Irrelevant or incorrect data is removed.	Data Engineer/Developers Data Scientist Electricity System Electricity supplier's teams	Programming Libraries Data lake, Model servers Integrated development environment
Data pre-processing	NA values from cleaned data are filled and all the categorical variables are encoded into numerical values. The characteristics of the input layer of the algorithms are defined. Data that could identify the consumer is removed at the concentrator level.	Data Engineer/Developers Data Scientist Electricity System Electricity supplier's teams	Programming Libraries Data lake, Model servers Integrated development environment
Model design and implementation	The model's designs and parameters are determined by our needs, statistical tests, and analysis of the time series.	Data Scientist	Programming Libraries Data lake, Model servers Integrated development environment
Model training	Data is prepared to enable hourly predictions. Initial training is done using 70% of the historical data and processed in regional data centres next to supervision centres. 15% of the data is used for evaluation of the model.	Data Scientist	Programming Libraries Data lake, Model servers Integrated development environment
Model testing	15% of the most recent data is used for testing.	Data Scientist	Programming Libraries Data lake, Model servers Integrated development environment
Optimization	The hyperparameters of the model are optimised through different techniques.	Data Scientist	Programming Libraries Data lake, Model servers Integrated development environment

Model evaluation	From a Machine Learning perspective, the model is evaluated on dedicated data with metrics such as Mean Absolute Error or Mean Squared Error. If the model's performance does not meet the required outcomes, previous steps must be repeated.	Data Scientist	Programming Libraries Data lake, Model servers Integrated development environment
Model deployment	The model is put into production on dedicated servers and incorporated in a dashboarding tool. These servers are hosted on-premises.	Data Engineer/Developers Electricity System Electricity supplier's teams System and communication network's administrator, Cloud provider	Data lake Model server, Dashboarding tool Integrated development environment
Monitoring and inference	The model, once in production, provides predictions to electricity suppliers. These predictions can be displayed via dashboarding tools or directly incorporated into existing tools for reuse.	Data scientist Electricity System Electricity supplier's teams System and communication network's administrator	Data lake, Model server Dashboarding tool Concentrator Electricity supplier's electrical grid Smart meters Open data provider APIs

2.8 PRIVACY AND CYBERSECURITY REQUIREMENTS

Cybersecurity requirements

The context information given in the previous section enables evaluation of the application's cybersecurity and privacy requirements. The following table summarises the cybersecurity requirements.

Figure 9: Cybersecurity requirements

	Level	Explanation
Availability	**Low**	Predictions made by the model are made **every day**. An unavailability during **half a week would be tolerable** as old processes could be reused until the application is back in service. **Longer unavailability** would be too time consuming for the Electricity supplier.
Integrity	**Critical**	Data, whether used for training or made by the model, must be **accurate with a high level of quality** at all steps of the lifecycle. The alteration of the manipulated data can cause a very **large imbalance** between the different means of energy production (non-renewable and renewable). Moreover, a poorly trained model could cause an **underproduction or an overproduction** of energies with very non-negligible consequences.
Confidentiality	**Critical**	Not all data is confidential, as some exists in the public domain. However, consumption information of individuals used in the context of our scenario include **personal data (upstream of the concentrator)**. This data is confidential and could highlight the behaviour of individuals whose information is retrieved.
Traceability	**High**	Actions linked to the process must be logged (traced and dated) to enable full traceability of changes made to the algorithm. As personal data is used in the process, actions linked must be imputable. i.e. all actions must be traced (even consultation actions) and attributable without possibility of repudiation.

Privacy requirements

The energy consumption of individuals, which contains personal data (name, address, phone number, etc.) is collected through smart meters. This data is processed in a concentrator so that only aggregated daily consumption values are routed to the data lake. This means there is no longer any personal data in the data lake because the sum of the data consumption of a concentrator does not permit individual consumption to be recognised. In this step it is crucial that this transformation from personal data to anonymised data is performed recognising the goals identified of the WP29 guidelines on anonymisation techniques (i.e., avoiding singling out, likability and inference)[29].

It is important to note that the billing functions are not considered in our case, this topic being out of scope.

Because **our scenario handles personal data in the data collection phase**, it is subject to the GDPR. **The following data protection requirements and recommendations should be satisfied**.

Figure 10: Data protection principles

Requirements	Explanation
Lawfulness, fairness, and transparency[30]	Personal data collected (both in training and in production) in the purpose of the scenario must be processed lawfully. That is, the data processing must be based on one of the legal bases provided by Article 6 of GDPR. In addition, individuals' personal data must be processed only in the way they reasonably expect, and any unexpected but justified processing must be clearly explained. Data processing must also comply with the transparency requirements. **Lawfulness**: for the specificities of this scenario, and as explained in its description, if the electricity supplier wants to extract the detailed consumption data of the users (consumption per hour or half hour), the electricity supplier must obtain the consent of the consumers[31]. If the user does not consent to the retrieval of this hourly or half-hourly consumption data, the electricity supplier may retrieve his daily consumption data for legitimate interest. The electricity supplier has a legitimate interest in at least retrieving consumers' daily consumption data. Without this data, the prediction of the users' energy demand is impossible. **Fairness:** even if the electricity supplier demonstrates that the data processing has a legal basis, it must only process personal data in line with consumers reasonable expectations, i.e., for the prediction of energy demand, and not use it in a way that has an unjustified negative effect on them. **Transparency:** the electricity supplier must also be transparent with consumers and clearly inform how and why their personal data are used throughout the machine learning lifecycle (for instance as part of the consumer agreement).
Purpose limitation	Personal data collected (both in training and in production) needs to be processed for specific, explicit, and legitimate purposes and is not further processed in a manner that is incompatible with those purposes. This means that the energy consumption collected and processed should only be used to predict the energy demand.
Data minimisation	Personal data collected (both in training and in production) needs to be adequate, relevant, and limited to what is necessary in relation to the purpose of predicting energy consumption demand by using Machine Learning. The electricity supplier should justify the data to be used and should indicate the procedures in place to ensure that only essential data such as electricity consumption is processed for this purpose.

[29] See https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
[30] The choice of a legal basis is final, and a treatment cannot involve two legal bases. However, in our case, as explained in the body of the report, two treatments are made: a treatment based on the fact that the data is extracted daily and a treatment based on the fact that the data is extracted every hour or half hour. Depending on which of the two treatments is accepted by the consumers, a corresponding legal basis will be chosen.
[31] See https://www.cnil.fr/fr/linky-gazpar-quelles-donnees-sont-collectees-et-transmises-par-les-compteurs-communicants

Accuracy	Personal data collected (both in training and in production) must be accurate and kept up to date. The electricity supplier, from the collection phase to the production phase, should put in place all appropriate measures to promote data accuracy such as keeping the data up to date and ensuring that they remain correct. For example, putting in place a clear and straightforward process for consumers to update their addresses so that consumption data are always associated with the right locations.
Storage limitation	Personal data collected (both in training and in production) must be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. Specifically, in our case, the personal data collected is anonymised before exiting the concentrator.
Security of personal data (Integrity and Confidentiality)	Personal data collected (both in training and in production) must be processed in a manner that ensures appropriate security, in particular integrity and confidentiality of the data. In our case, the risks from data corruption are minimal for consumers (as a reminder, billing data is not in scope of our scenario), and we consider that the data extracted by the smart meter has no other purpose than the prediction of the energy demand). On the other hand, the electricity supplier must put in place adequate security measures to ensure the confidentiality of the data, as the leakage of the latter can indicate the user's habits.

As a complement to the GDPR data protection principles listed above, other key privacy topics related to AI systems need to be addressed (some of which are specifically mentioned by national data protection supervisory authorities)[32]:

Figure 11: Data protection supervisory authorities' recommendations for AI systems

Recommendations	Explanation
Database creation	During the collection phase, the data passing through the electricity supplier network to the concentrator are personal data. A process aiming at deleting direct identifier and aggregate the consumption between all the consumer is made at the level of the concentrator. However, before this processing, data are stored in the database of the concentrator. The electricity supplier must ensure that only duly authorised persons have access to this data and prevent data loss. The electricity supplier must also ensure that only the right data (energy consumption) are injected into the concentrator.
Compliance of the training model (i.e., before production)	The electricity supplier must justify that the data it collects, the tools it uses and the training model it has chosen to achieve its objectives is relevant and free from bias. The electricity supplier must ensure that before the system is put into production, appropriate steps are taken to prevent discrimination and that a human electricity supplier can verify the quality of the algorithm's outputs.

Finally, the project requires a data protection impact assessment[33] as it completes at least one criterion in the following table[34]:

Figure 12: Is a Data Protection Impact Assessment (DPIA) necessary?

Criteria	Does it match the criteria?	Justification
Evaluation or scoring		
Automated decision making with legal or similar significant effect		
Systematic monitoring	X	The processing is not intended to do this. Nevertheless, the data collected may allow to determine whether a person is present at home during certain time periods.

[32] See https://www.cnil.fr/fr/intelligence-artificielle/la-cnil-publie-ressources-grand-public-professionnels
[33] See https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/#:~:text=DPIAs%20are%20a%20legal%20requirement,trust%20and%20engagement%20with%20individuals
[34] The condition for a DPIA to occur is that at least two conditions are satisfied.

Sensitive data or data of highly personal nature		
Data processed on a large scale	X	Personal data related to energy consumption are used on a large scale by the electricity supplier.
Matching or combining datasets		
Data concerning vulnerable data subjects		
Innovative use or applying new technological or organizational solutions	X	The use of artificial intelligence in this scenario: predicting energy production capacity and demand is innovative.
Preventing data subjects from exercising a right or using a service or contract		

To assess the risks in terms of privacy, personal data processing must consider the following requirements:

Figure 13: Privacy requirements regarding data

	Level	Explanation
Availability	**Low**	The data, for the user, does not necessarily have to be available. An unavailability of several days is acceptable.
Integrity	**Low**	The output data of the concentrator is not considered as personal data. Indeed, at the concentrator output, it is the sum of the consumption data (from pseudonymised individual consumption data) that is transmitted to the data scientists. The alteration of this data has no influence on the users, such as a more expensive billing.
Confidentiality	**Critical**	The personal data processed, in particular the detailed (hourly), is confidential. The leakage of such data (possible upstream of the concentrator) would be detrimental to the consumer since it would leak his address for instance. Moreover, in case of data leakage from smart meters, it is possible to know the life habits of consumers and when they are present (high consumption) and when they are absent (low consumption) from home.
Traceability	**High**	At a minimum, the actions related to the process must be recorded (traced and dated) to be able to follow the actors involved during the stages of use of personal data (upstream of the concentrator). Particularly in the case of court cases, it is necessary to be able to attribute actions to an individual and to be able to impute them criminally.

3. SECURITY AND PRIVACY THREATS AND VULNERABILITIES

3.1 THREAT CONTEXTUALISATION

This section highlights the threats applicable in the forecasting demand on electricity grids scenario. Examples in the context of this scenario are also provided. Also outlined are the impacts that each threat can have in terms of security and privacy as well as their severity in the table.

There are two major feared events to consider for our scenario which are the loss of integrity of the data, and the loss of confidentiality of the detailed electricity consumption of inhabitant data.

The loss of integrity of the data could lead to an **electrical production disruption**. Indeed, the impact would be maximum if too little energy was produced, and there was insufficient energy supplied to meet demand. It should be noted, however, that the impact is lessened because the model only assists and does not make decisions directly. Decision-makers have other indicators with them to make decisions (minimum value of energy to be produced etc.).

For the company, the loss of confidentiality of the detailed consumption of inhabitant data could lead to **reputation degradation,** which can result in loss of trust of the consumer.

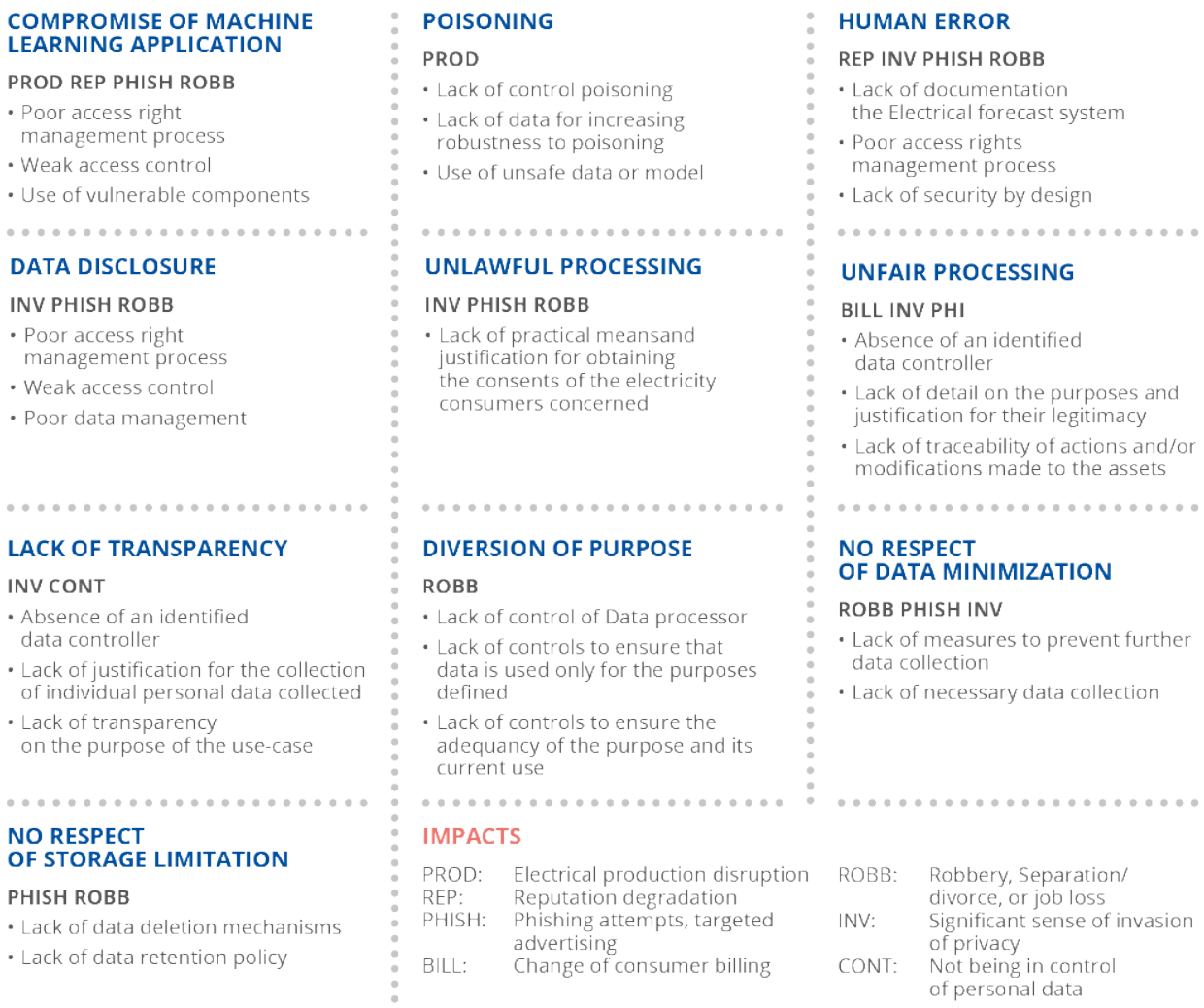
For the consumers, the loss of confidentiality of the detailed consumption could have 2 impacts. First, stolen personal data (home address, for example) could be used to perform **phishing attempts or targeted advertising**. Also, if detailed consumption of an inhabitant (either hourly or half hourly) were to leak, people would be able to identify whether the inhabitant is at home or not, and it could lead, depending on the case, to unwanted consequences such as **robbery** (a burglar might determine the hours at which a person is not a home using his energy consumption data)**, separation/divorce** (much like phone bills or credit card details might contain data that the person would consider private, even from his or her spouse)**, or job loss** (in the case where an employer uses energy consumption data).

Moreover, non-compliance with certain privacy requirements by the electricity supplier can lead to impacts on the user such as the **significant feeling of invasion of privacy, feeling out of control of their personal data**, or **change in energy consumption billing** if the data is processed for other purposes, besides the directly related with this case.

Considering these feared events and associated impacts, the following threats are associated with our scenario.

Before going into the details of the threats and vulnerabilities to which this scenario is subject, the following figure summarises all the threats that will be explained subsequently:

Figure 14: Summary of threats and vulnerabilities



3.1.1 Compromise of ML application components

An external attacker can directly compromise smart meters, the concentrator, or the API to retrieve weather or demographic data if they are exposed. If one of these components is compromised, the manipulated data can be modified, leading to an **electrical production disruption**.

Data scientists, developers and the various stakeholders involved could have their workstations compromised by an external attacker (directly by using insecure code libraries, for example) which could compromise the ML components. The attacker would then be able either to modify the manipulated data, leading to an **electrical production disruption**, or steal confidential data (detailed consumption of inhabitant), leading to **reputation degradation** for the company, **significant feeling of invasion of privacy, phishing attempts, or targeted advertising** or even **robbery, separation/divorce, or job loss** for the consumers.

3.1.2 Poisoning

During the **data collection** phase, an external attacker or a malicious user could compromise one or more smart meters (via a worm for instance) to create outliers which would then

constitute **historical consumption data**. Once the model used to predict the next day consumption has been trained with such data, it could overestimate electricity consumption, leading to an **electrical production disruption.**

In this idea of data modification and after **having collected the data**, any attacker with sufficient rights and access to the data lake could modify the training data to make the model operate with undesired characteristics, and could even render the model dysfunctional. This could lead to an **electrical production disruption**.

3.1.3 Human error

A data engineer could route the consumption data from the concentrator to the data lake without aggregating and anonymising it. Such a threat could lead to a potential data disclosure which could entail **reputation degradation** for the company, a **significant feeling of invasion of privacy, phishing attempts, or targeted advertising** or even **robbery, separation/divorce, or job loss** for the consumers.

3.1.4 Data disclosure

Throughout the scenario lifecycle, an attacker can try to compromise the concentrator to collect stored personal data relating to energy consumption, even where there is no data retention after sending the aggregated data to the data lake. This can be done by physical tampering with the concentrator, or by exploiting vulnerabilities through access to the electricity supplier's electrical network via its smart meters. Resultingly, an attacker with multiple smart meters connected to the electricity supplier’s electrical grid could directly attack other smart meters on the same network to collect consumption data.

For the company, these data leaks can lead to **reputation degradation.** For the electricity consumers, they can lead to **significant feeling of invasion of privacy, phishing attempts, targeted advertising,** or even **robbery, separation, or divorce and/or job loss.**

3.1.5 Unlawful Processing

There are two cases where two different legal bases should be applied. First, when the electricity supplier wishes to extract the energy consumption of the inhabitants on an hourly or half-hourly basis, it should ask for the users' consent, and proceed to process only when this consent has been given freely, clearly, specifically and knowingly by the user. Second, in the case of daily extraction, the electricity supplier extracts the consumption data based on legitimate interest.

In the case where the electricity supplier was to collect data on an hourly or half-hourly basis without having obtained the user’s prior consent beforehand (or otherwise offering opportunity to withdraw consent), the impact on consumers would be a **significant sense of invasion of privacy**. Their data extracted without their consent would lead to a total lack of trust in the electricity supplier. In addition, the legitimate interest in the day-to-day processing of consumer data may not be adequately justified by the electricity supplier or worse, no justification may be provided. In this scenario, users may question the purpose of the use of their personal data and feel **a significant feeling of invasion of privacy**.

3.1.6 Unfair processing

Even if the electricity supplier demonstrates that the data processing has a lawful basis (depending on the level of data extraction: daily or hourly), it might not process the personal data in the way consumers expect it to be i.e. for energy demand forecasting only. They could use it in a way that has an unjustified negative effect on them, such as **unknowingly changing their billing** based on consumption forecasts made by the algorithm.

3.1.7 Lack of transparency

The electricity supplier may also not be transparent with users about how their personal data is used and the goal of the use of their data. The terms governing these aspects may not be included or may be deliberately omitted by the electricity supplier in, for example, the contract with its customers. The impact this could have on consumers could result in a **feeling of being not in control of personal data**.

3.1.8 Diversion of purpose

The user data extracted in our case is intended to predict consumer energy demand. However, the electricity supplier may not respect this principle and use the consumer data for ambiguous and non-explicit purposes unrelated to the prediction of energy demand, such as using their data to offer other services or to resell them. For instance, users might be profiled based on their energy consumption and could receive targeted advertisements. Using this data to learn about users' routines and habits could be considered a serious invasion of users' privacy. In the wrong hands, this data could even be used against the physical security of consumers.

An even more extreme diversion from the purpose of this information collection could facilitate the surveillance of users; the person in possession of the model and the data could examine the habits of the users to work out whether or not the consumer is at home. The impact this could have on consumers could lead to **robbery, or separation/divorce, or job loss.**

3.1.9 No respect of data minimisation

This could materialise in the form of collecting more data than necessary at the smart meter level or by using more data than necessary at the hub level. For instance, smart meters could collect much more accurate data such as real-time consumption that can be used to track the consumer. The impact this could have on consumers could lead to a **significant feeling of invasion of privacy**, **separation/divorce, job loss in case of data leakage** or to **potential phishing attempts, targeted advertising,** or even **robbery.**

3.1.10 No respect of storage limitation

Data extracted from smart meters of individual households are routed to the concentrator level and amalgamated. This means that no household specific data should be stored after amalgamation. Without clear rules for storing and removing personal data at the concentrator level, the electricity supplier could decide to store user data in the concentrator database for several months when it is not operationally justified. If this data were to be stolen or leaked, the impact to the user could be **separation/divorce, job loss,** potential **phishing attempts, targeted advertising,** or even **robbery.**

3.1.11 Synthesis of possible impacts and associated threats

The following table sums up the severity of each impact and the associated threats.

Figure 1: Synthesis of possible impacts and associated threats[35]

Impact	Severity	Type	Associated Threats
Electrical production disruption	**High**	Cybersecurity	Compromise of ML application components Poisoning
Reputation degradation	**High**	Cybersecurity	Compromise of ML application components Data disclosure
Phishing attempts, targeted advertising	**Moderate**	Privacy	Compromise of ML application components Unlawful processing Unfair processing

[35] See the severity scales in the Annex

Robbery, Separation/divorce, or job loss	**High**	Privacy	Compromise of ML application components Unlawful processing
Significant feeling of invasion of privacy	**Moderate**	Privacy	Unlawful processing Unfair processing lack of transparency Diversion of purpose No respect of data minimization No respect of storage limitation
Not being in control of personal data	**Moderate**	Privacy	Lack of transparency
Change of consumer billing	**High**	Privacy	Unfair processing

3.2 VULNERABILITIES ASSOCIATED TO THREATS AND AFFECTED ASSETS

This figure cross evaluates each threat to a set of associated vulnerabilities. The actors involved and the assets possibly affected by the vulnerabilities are also highlighted:

Figure 2: Mapping vulnerabilities to threats and assets/actors on which they rely

Vulnerabilities	Threats	Actors	Assets involved
Absence of an identified data controller	Unlawful processing Unfair processing Lack of transparency Diversion of purpose No respect of data minimization No respect of storage limitation	Electricity supplier's teams	Data
Absence of mechanisms to ensure that processing of consumer electricity affected by consent cannot be carried out without consent	Unlawful processing	Electricity supplier's teams	Smart meters Concentrator Electricity supplier's electrical grid
Disclosure of sensitive data for ML algorithm training	Data disclosure	Data scientists	Model
Existing biases in the ML model or in the data	Diversion of purpose	Data scientists	Model Data
Lack of anonymisation	No respect of data minimization	Electricity supplier's teams Data engineers	Data Concentrator
Lack of auditability of processing	Unlawful processing Unfair processing Lack of transparency Diversion of purpose No respect of data minimization	Electricity supplier's teams Data scientist Developers and data engineers System and communication network administrators	N/A

	No respect of storage limitation		
Lack of control for poisoning	Poisoning	Data Scientists	Model
Lack of control of Data processor[36]	Diversion of purpose	Electricity supplier's teams	Data
Lack of controls to ensure that data is used only for the purposes defined	Diversion of purpose	Electricity supplier's teams	Smart meters Concentrator
Lack of controls to ensure the adequacy of the purpose and its current use	Diversion of purpose	Electricity supplier's teams	Smart meters Concentrator
Lack of data deletion mechanisms	No respect of storage limitation	Electricity supplier's teams	Data
Lack of data for increasing robustness to poisoning	Poisoning	Electricity supplier's teams Data engineers and developers	Data
Lack of data retention policy	No respect of storage limitation	Electricity supplier's teams	Data
Lack of detail on the purposes and justification for their legitimacy	Unlawful processing Unfair processing Lack of transparency	Electricity supplier's teams	N/A
Lack of documentation	Human error Unlawful processing Unfair processing Lack of transparency Diversion of purpose No respect of data minimization No respect of storage limitation	Electricity supplier's teams Data scientists Developers and Data Engineers System and communication network administrators	N/A
Lack of justification for the collection of individual personal data collected	Unlawful processing Lack of transparency No respect of data minimization	Electricity supplier's teams	Data
Lack of legal basis related to users' consent when their detailed consumption data (per hour or half hour) are processed or that legitimate interest related to the daily processing of the data is not properly justified or that no justification is provided at all	Unlawful processing	Electricity supplier's teams	N/A

[36] Strictly speaking, this vulnerability is outside the scope of our analysis since the company oversees the production and distribution of energy -so there is no external stakeholder which manipulate personal data. However, this is not necessarily the case. We therefore draw the reader's attention to this point. Nevertheless, we will not deal with it in the following section

Lack of measures to prevent further data collection	No respect of data minimization	Electricity supplier's teams	Data
Lack of necessary data selection	No respect of data minimization	Electricity supplier's teams Data scientist Developers and data engineers	Data
Lack of practical means and justification for obtaining the consents of the electricity consumers concerned (those who have a half-hourly view of their electricity consumption)	Unlawful processing	Electricity supplier's teams	N/A
Lack of security by design	Compromise of ML application components Poisoning Human error Data disclosure	Electricity supplier's teams	All assets
Lack of privacy by design	Unlawful processing Unfair processing Lack of transparency Diversion of purpose No respect of data minimization Diversion of purpose	Electricity supplier's teams	All assets
Lack of security process to maintain a good security level of the components of the Electrical forecast system	Compromise of ML application components Data disclosure	Electricity supplier's teams	Data lake Open-data provider APIs Smart meters Concentrator Model server Dashboard tools Electricity supplier's electrical grid Integrated Development Environment Programming Libraries Communication protocol and communication Networks
Lack of traceability of actions and/or modifications made to the assets	Unlawful processing Unfair processing Lack of transparency No respect of storage limitation Compromise of ML application components	Electricity supplier's teams	Smart meters Electricity supplier's electrical grid Concentrator
Lack of transparency on the purpose of the processing, the exact consumption data that are	Lack of transparency	Electricity supplier's teams	N/A

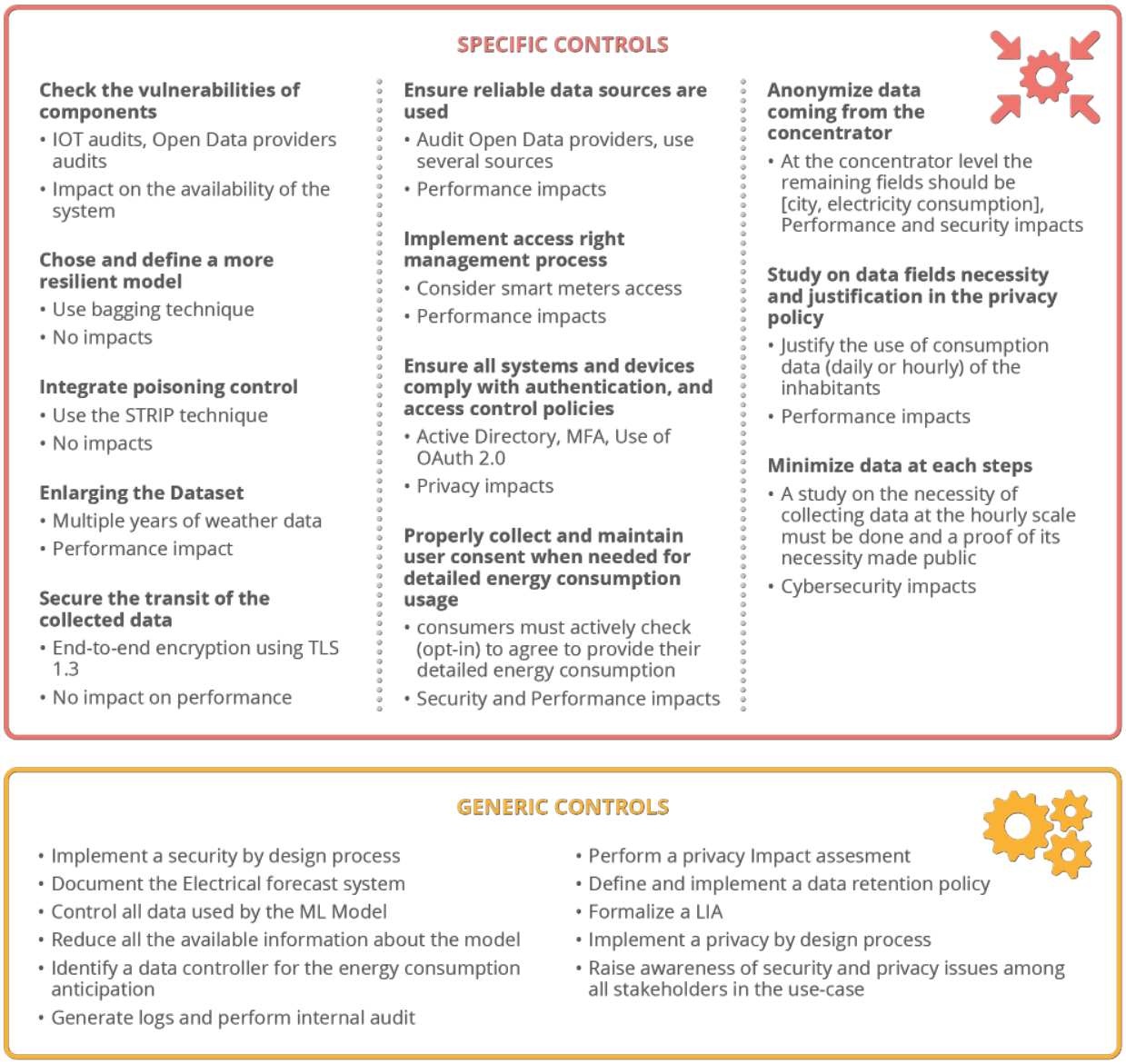
extracted, and how they are processed.			
Lack of verification that the data is adequate, relevant and not excessive for the purpose of estimating electricity consumption	No respect of data minimization	Electricity supplier's teams	Data
Model easy to poison	Poisoning	Electricity supplier's teams Data scientists Developers and data engineers	Model
No detection of poisoned samples in the training dataset	Poisoning	Electricity supplier's teams Data scientist Developers and data engineers	Data
Poor access rights management process	Compromise of ML application components Poisoning Data disclosure Human error	Electricity supplier's teams	Data lake Open-data provider APIs Smart meters Concentrator Model server Dashboard tools Electricity supplier's electrical grid Integrated Development Environment
Poor data management	Poisoning Data disclosure Diversion of purpose No respect of data minimization No respect of storage limitation	Data scientists Developers and data engineers	Data
Excessive information available on the model	Compromise of ML application components	Electricity supplier's teams Data scientist	Model
Unprotected sensitive data on test environments	Data disclosure	Electricity supplier's teams Data scientists Developers and Data Engineers System and communication network administrators	Data
Use of uncontrolled data	Poisoning No respect of data minimization No respect of storage limitation	Electricity consumers Data engineers Open-data providers	Data
Use of unsafe data or models (e.g., with transfer learning)	Poisoning	Electricity consumers, Open-data providers	Data

Use of vulnerable components (Among the whole supply chain)	Compromise of ML application components	Electricity supplier's teams Data scientist Developers and data engineers Open-data providers	Data lake Open-data provider APIs Smart meters Concentrator Model server Dashboard tools Electricity supplier's electrical grid Integrated Development Environment Programming Libraries Communication protocol and communication Networks
Weak access protection mechanisms for ML model components and for personal data (encryption, access control mechanism…)	Compromise of ML application components Poisoning Data disclosure Human error	Electricity supplier's teams	Data lake, Open-data provider APIs Smart meters Concentrator Model server Dashboard tools Electricity supplier's electrical grid Integrated Development Environment

4. CYBERSECURITY AND PRIVACY CONTROLS

Before expanding on the details of the cybersecurity and privacy controls applied to the scenario, the following figure summarizes all the controls that will be described:

Figure 17: Summary of cybersecurity and privacy controls



4.1 IMPLEMENT A SECURITY BY DESIGN PROCESS

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity	Lack of Security by Design	• Compromise of ML application components • Poisoning • Human error • Data disclosure

By default, Security by Design is a methodology to strengthen the cybersecurity of the organisation by automating its data security controls and developing a robust IT infrastructure. This approach focuses on implementing the security protocols from the basic building blocks of the entire IT infrastructure design. The goal of this control is to ensure that risks are identified as early as possible, mitigated before systems go live, and appropriate security controls are implemented.

The lack of Security by Design increases the likelihood of all threats related to our case. Therefore, the electricity supplier must ensure, from the development phase, to put in place adequate controls to limit the cybersecurity risk. This starts with a global risk analysis in which all risks associated with all assets are identified. In our case, the electricity supplier will try to reduce the attack surface (for example by updating the smart meters' operating system continuously), to apply the principle of least privilege (by implementing access rights management), to take care of the confidentiality and integrity of the collected data (by encrypting the users' consumption data for example).

Conceiving a project following a Security by Design methodology requires less effort than adding security on top of an existing project. However, there could be an impact on the functionality of the use if the outcome of the Security by Design methodology would lead to some functionality not being implemented due to the security risk it generates (e.g., installation of unsecure smart meters).

4.2 DOCUMENT THE ELECTRICAL FORECAST SYSTEM

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity & Privacy	Lack of Electrical forecast system documentation	• Human error • Unlawful processing • Unfair processing • Lack of transparency • Diversion of purpose • No respect of data minimisation • No respect of storage limitation

Project and system documentation must be produced to preserve knowledge on decisions made during the project phase, application architecture, configuration, maintenance, ability to maintain effectiveness over time, and assumptions made about the model's use.

The system is complex with multiple assets (Open-Data provider, smart meters, concentrator, etc.) which enhance the fact that an exhaustive documentation is mandatory for all assets. This documentation should also include the changes that will be applied, including to the documentation throughout the system life cycle.

Therefore, it is necessary that the Open-Data Provider, Data Scientists, Developers, Data Engineers, and System and communication network administrators work together to create and sustain the documentation.

This control does not impact system performance, cybersecurity, or privacy.

4.3 CHECK THE VULNERABILITIES OF THE ML COMPONENTS AND IMPLEMENT PROCESSES TO MAINTAIN THEIR SECURITY LEVELS OVER TIME

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity & Privacy	• Use of vulnerable components • Lack of security process to maintain a good security level of the components of the Electrical forecast system	• Compromise of ML application components • Data disclosure

For smart meters and the concentrator, it is necessary for the company to ensure that regular security audits are carried out to ensure there are no vulnerabilities. Regular vulnerability scans and automatic patch management processes should also be implemented to maintain a good security level. Preparation of a remediation plan that can be implemented quickly must be committed to, with the capacity to review this plan over time.

One differentiating factor for smart meters is that they can be considered as IoT devices, for which audits need specific expertise. Although these devices are connected to the network as regular devices, they also have an interface with the physical world. In our case, smart meters interface with the consumer's electrical system, which could be a source of vulnerability.

For Open-Data provider APIs, it is necessary to define a process to ensure that the level of Open-Data provider security is sufficient over time to avoid data recovery. This could mean an annual audit of the APIs and the completion of an annual security questionnaire.

This control would have an impact on the availability of the system, and thus its performance, as it may be audited or updated. This could lead to periodic unavailability of the system for several hours at a time. However, as the availability need of the system is low, the impact of this control is only moderate.

Moreover, updates could cause the system to malfunction, leading to errors in the prediction systems that could have a major impact.

4.4 CHOOSE AND DEFINE A MORE RESILIENT MODEL DESIGN

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity	Model easy to poison	Poisoning

Poisoning of the model could cause electrical production disruption, which would have a high impact for the company. For this reason, the company needs to ensure that the model is fully resilient against this type of attack.

For the electrical forecast system, a methodology to harden the system against poisoning threats could be use of the "Bagging" (bootstrap aggregating) technique[37]. This is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms, which combines several versions of the model (e.g., models trained with

[37] https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.695.4667&rep=rep1&type=pdf

different datasets) to limit the impacts related to the poisoning of one of the models. In this case, the company could generate multiple models from different data sets (weather data and consumption of each different years for example) and use a Bagging algorithm with these models to predict the consumption of electricity. If an attacker were to succeed in poisoning one of the models, the effect on the outputs of the system would be greatly reduced, or even mitigated completely, due to the prediction also relying on the other healthy models.

This could have an impact on performance of the model as the most representative data is from the most recent year, and using data from previous years would make the prediction less accurate.

4.5 INTEGRATE POISONING CONTROL IN THE TRAINING DATASET

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity	• No detection of poisoned samples in the training dataset • Lack of control for poisoning	Poisoning

In addition to the previous control, and to reduce to a minimum the probability of the models being poisoned, the training dataset should be checked for poisoning.

The RNN algorithm used in our scenario can be checked for poisoning using the STRIP technique. This principle relies on disturbing the inputs and observing the randomness of the predictions. For example, the company could change 1 week of weather data, replacing summer data with winter data, and observe the variance in predictions versus unperturbed. Without high variance, the model could be considered as poisoned. These tests should be performed before the production phase, during the training phase, ensuring that the models used in production are healthy.

This control doesn't have impact on system performance, cybersecurity, or privacy.

4.6 ENLARGE THE TRAINING DATASET

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity	Lack of data for increasing robustness to poisoning	Poisoning

Large timeframes (multiple years) of weather data can be collected and used to train the algorithm to prevent it being susceptible to poisoning attacks. Modifying small amount of weather data would then have a low impact of the overall prediction.

Using large timeframes (multiple years) of inhabitant consumption data could also reduce the probability that poising attacks on this data could impact the prediction. This could still have an impact on the model accuracy, considering that weather systems can shift over long timeframes too, but considering the high impact of disruption to electrical production, this control should still be implemented. Long timeframes of consumption data would not impact consumer privacy as any stored data would be the anonymised after the concentrator.

4.7 SECURE THE TRANSIT OF THE COLLECTED DATA

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity	Poor data management	Poisoning

Considering the possible impacts of Poisoning, the transit of the collected data should be fully protected against a loss of integrity. Within the collection phase of weather data or inhabitant consumption data, the electrical company should ensure that protocols used for the transfer of data are encrypted and otherwise state of the art (without any known vulnerabilities), such as TLS 1.3. As such, an attacker would not be able to modify the data during transit, and would not be able to poison the model through that route.

Generally, Open-Data suppliers comply with this level of requirement. This control would not impact on privacy or performance of the system.

4.8 CONTROL ALL DATA USED BY THE ML MODEL

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity	Use of uncontrolled data	Poisoning

Considering the high impacts of Poisoning attacks, several means to control the data must be applied, some specific to data science and others more to business- common-practice. These techniques can be applied at the concentrator level and then in the data lake. At the concentrator level, the incoming data is not yet aggregated, so an initial check can be carried out to avoid inconsistent data.

There are two common cases of inconsistent data – inconsistency of data type (e.g., alphanumeric vs number) and inconsistency of data value. To prevent this, checks should be performed for the two different cases. An example of an inconsistent value would be to determine whether a consumption figure is abnormally high in relation to previous data, contract data, data from other consumers. Once abnormal data is detected, it can be readjusted before following the data aggregation procedure to the data lake.

At the data lake level, more controls can be carried out:

- Non-technical checks such as checking the overall consistency of the data (for example checking variability of the weather data)
- Technical checks specific to data science and the data sanitisation process to detect possible anomalies (otherwise known as outlier detection).

This control could add latency to the system in case of many reported anomalies on the data. But our scenario does not have a strong need for availability, and the control is important to reduce the probability of model poisoning.

The checks that are applied at the concentrator level could have an impact on privacy, because they would manipulate personal data from consumers. To remediate, the systems (software, scripts, excel sheets, etc.) used to perform these checks should be regularly audited, documented, secured by design, and follow access control rules (authentication to use the system, proper access management, etc.).

4.9 ENSURE RELIABLE SOURCES ARE USED

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity	Use of unsafe data or models (e.g., with transfer learning)	Poisoning

Considering the possible disruption of electrical production due to poisoning attacks, it is mandatory for the company to ensure the reliability of the sources used to collect the data. For weather and demographic data, which is collected from an Open-Data Provider, this should be of a reliable standard, ideally audited once a year to ensure that it doesn't have any vulnerability that could lead to a loss of integrity of the data. Moreover, and considering the criticality of the poisoning threat, the data can be crossed between several sources to reduce the impact when one of the sources is compromised.

However, crossing data from multiple sources could introduce errors within the data set which could alter the model and its performance. Our scenario does not have a strong need for availability.

4.10 IMPLEMENT ACCESS RIGHT MANAGEMENT PROCESS

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity & Privacy	Poor access rights management	• Compromise of ML application components • Poisoning • Data disclosure • Human error

Threats related to poor access rights management, such as data disclosure, can have a high impact on the company (electrical production disruption) or on the consumers (robbery, separation/divorce or job loss). This security measure defines and assigns roles to users, respecting the principle of least privilege, and limits the access scenarios for processes and assets only to those for which the users have a justified reason.

For all the classical systems of our scenario, such as Concentrator, Data lake, Model server, Dashboard tools, and Integrated Development Environment, it is essential to create roles for each user group with associated access rules. These work alongside a process for providing access only to the actors that need it, and to prevent access to unauthorised persons. Each role must respect the principle of least privilege, in particular the roles that can connect to the concentrator as they allow access to all the consumption data and associated personal data. By default, only administrators and data engineers would have this access. It is also very important

to have a process for modifying or revoking user access in timely manner, as long processes for access management generally result in neglect by an organisation.

Considering the high volume of smart meter devices, the administration account would likely need to be generic, with the account password stored within an administration Bastion. Then, only mandated electricity supplier's teams would have the ability to have access this Bastion and connect to the smart meters.

In case of an incident or failure, the complexity of the management of these accesses, particularly for the smart meters, could prevent teams from accessing the system to correct potential problems, which could impact the availability of the system, thus its performance.

4.11 ENSURE ALL SYSTEMS AND DEVICES COMPLY WITH AUTHENTICATION, AND ACCESS CONTROL POLICIES

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity & Privacy	• Weak access protection mechanisms for machine learning model components • Weak access protection mechanisms of the personal data (encryption, access control mechanism…)	• Compromise of ML application components • Poisoning • Data disclosure • Human error

In this control there must be a distinction made between user access and device access. For user access, the Electrical Company must ensure that proper user authentication and access controls regulate access to Smart meters, the Concentrator, the data lake, the Model server, the Dashboard tools, and the Integrated Development Environment. Such access authentication and access control should be managed centrally, with a dedicated solution (e.g., Active Directory) linked to all these assets, except in the case of smart meters, where authentication should remain local due to the high volume of devices. Moreover, for sensitive assets manipulating personal data, such as the concentrator, the electrical company should implement multi-factor authentication for the data engineers and administrators that are granted access. The multi-factor authentication solution must be linked to the central authentication system.

For device access, the electrical company should ensure that all device-to-device requests are properly authenticated. To deal with threats, it is important to use good market practices. For example, use of the OAuth 2.0 protocol for all exposed APIs with control via an API gateway. In addition, the certificates for the authorisation server keys should be managed through the company's own Public Key Infrastructure in order to avoid any identity theft scenario. It is also necessary to respect an authentication policy (based on the current recommendations of bodies such as ENISA[38] or NIST[39]) for the authentication of devices on the authorisation server. To protect the secrets in the smart meters, it is necessary to secure them by ensuring that their OS version is up-to-date, and that non-essential ports are closed, amongst other factors. To protect access to the company electric network via smart meters, it is necessary to carry out Network access control with good practices[40]. However, in case of an incident or a failure, the complexity of the authentication, especially for the smart meters, could slow down the remediation of the problem (for example if administrators had trouble authenticating on the system due to the reinforced authentication) and have an impact on the of availability of the system, thus its performance.

Additionally, personally authenticating the employees on the system can have an impact on privacy, because personal data (identifiers, mail addresses, etc) would be recorded in these

[38] https://www.enisa.europa.eu/news/enisa-news/tips-for-secure-user-authentication
[39] See https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63-3.pdf
[40] See https://cipherwire.net/wp-content/uploads/2013/06/802.1X_and_NAC___Best_Practices_for_Effective_Network_Access_Control.pdf

authentication systems. These authentication systems must respect the principles of Article 5 of the GDPR (data minimisation etc).

4.12 REDUCE THE AVAILABLE INFORMATION ABOUT THE MODEL

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity & Privacy	Too much information available on the model	Compromise of ML application components

This control lowers the likelihood of Compromise of ML application components by limiting the knowledge of a malicious user about his target and consequently making it more difficult to launch a cyberattack. It is important to understand that this security control alone is not enough, as security by obscurity does not provide sufficient protection.

Therefore, all documentation concerning the model itself should be protected. They should be encrypted and protected by a Data Right Management mechanism, allowing only the Data Scientists to access them from their computers. Specifically, the documentation can be stored in a dedicated file server directory to which access rights are regularly reviewed and linked to the access rights management mechanisms described in 4.10 and 4.11. This server must be encrypted by default and located in a network zone dedicated to the company's sensitive resources. In addition, user workstations accessing this file server must be sufficiently protected. In addition to the rights management mechanisms, it is necessary to provide, at least, encryption of the workstations, a confidentiality filters fitted to the screens of workstations in nomadic situations, and securing (OS version up-to-date, closure of non-essential ports etc) of the workstation to prevent the opening of certain ports and the insertion of USB keys. In addition, to avoid the propagation of documents relating to the model, it is recommended that a data loss protection solution be deployed.

This control could have an impact on privacy aspect because it may be considered as a contradiction to the principle of transparency required by the GDPR. Therefore, it is important to define documentation that can be communicated to external users (i.e., customers), explaining that their data is not directly used by the model. Moreover, this control could impact the performance of the system, because in case of an incident or a failure, the complexity of the access management could prevent data scientists to have access to the documentation to correct the problem and have an impact of availability of the system, thus its performance.

4.13 IDENTIFY A DATA CONTROLLER FOR THE ENERGY CONSUMPTION ANTICIPATION DATA PROCESSING

Type	Associated Vulnerabilities	Threats it mitigates
Privacy	Absence of an identified data controller	• Unlawful processing • Unfair processing • Lack of transparency

Identifying a data controller is essential to prevent various high impacts for the company and its employees due to unlawful processing, unfair processing, or lack of transparency. Self-identifying as the data controller means performing accountability actions (documentation, assessments, etc.) and making sure that the energy consumption anticipation data processing is compliant with personal data protection GDPR obligations and does not infringe on privacy rights of data subjects. Self-identifying as the data controller means also taking responsibility in case of non-compliance or adverse privacy effects on data subjects.

Such a measure does not have a negative impact in our case. On the contrary, it improves the privacy of users without impacting the performance of the system or its security.

4.14 PROPERLY COLLECT AND MAINTAIN USER CONSENT WHEN NEEDED FOR DETAILED ENERGY CONSUMPTION USAGE

Type	Associated Vulnerabilities	Threats it mitigates
Privacy	• Absence of mechanisms to ensure that the processing of consumer electricity affected by consent cannot be carried out without consent • Lack of legal basis related to user's consent when their detailed consumption data (per hour or half hour) are processed or that the legitimate interest related to the daily processing of the data is not properly justified or that no justification is provided at all • Lack of practical means and justification for obtaining the consent of the electricity consumers concerned	Unlawful processing

When the electricity supplier wants to extract the energy consumption of the inhabitants on an hourly or half-hourly basis, it should ask first for the users' consent and go on with the processing of the granular data only when this consent has been given. This means that there must be an opt-in function in the user interface (e.g., a checkbox) provided by the energy electricity supplier that the consumers must actively check (opt-in) to agree to provide their detailed energy consumption. Then, the consent of each user would be stored within a dedicated database belonging to the company and used accordingly for the collection (or not) of detailed consumption data. It is as important to ensure that when a subject revokes their consent, their detailed consumption data is no longer processed.

However, when this control is properly implemented, it is easier for data subjects to withhold consent for use of their personal data. Therefore, the data collected could be less rich with proper consent management and could be limited to daily energy consumption for a larger number of data subjects. In addition, there is a (relatively small) additional cost to implementing proper consent management. This measure impacts on security, because the company must protect not only the energy consumption data that the user gives with his consent but also the storage of the consent over time in a dedicated and secure space. The collection and storage must have a level of security adapted to the principles of non-repudiation (encryption of the data, traceability of access to the equipment, signatures via daily private keys in the database of the dedicated space, etc.).

4.15 ANONYMIZE DATA COMING FROM THE CONCENTRATOR

Type	Associated Vulnerabilities	Threats it mitigates
Privacy	Lack of anonymization	No respect of data minimization

The data that could identify consumers are anonymised at the concentrator level. Anonymisation requires an analysis to correctly identify a subset of fields that irreversibly prevents re-identification of data subjects and should be performed as early in the processing as possible. After this identification, a dedicated solution can be used to anonymise the data. It should also be remembered that the collected data are deleted in the concentrator after the aggregation of all data from different households.

In this case the data that need to be anonymised at the concentrator level is the Energy consumption of inhabitants containing these fields [surname, first name, address, meter number, telephone number, electricity consumption]. After anonymisation, the remaining fields should be [city, electricity consumption], with the city field being derived from the address of the concentrator and the electricity consumption being the sum of consumption by all customers associated with the concentrator. This is the data that is sent to the data lake.

In this case, in the event of a security breach, it may be more difficult to perform a cybersecurity forensic analysis on anonymised data than on personally identifiable data. This has the impact of significantly slowing down the work of those employed to determine the cause of a security issue. Ultimately, finding security measures for these kinds of vulnerabilities would also create other security challenges.

4.16 GENERATE LOGS AND PERFORM INTERNAL AUDIT

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity & Privacy	• Lack of auditability of processing • Lack of traceability of actions and/or modifications made to the assets	• Unlawful processing • Unfair processing • Lack of transparency • Diversion of purpose • No respect of data minimization • Compromise of Machine Learning application • No respect of storage limitation

Generating logs and performing internal audit allows for improvements in supervision of all assets of the model, creating a better understanding of the decisions made in real time by the algorithm. This also helps in understanding individual incidents, whilst auditing internal processes and questioning the processing carried out in the frame of the project reduces the likelihood of many threats in this scenario.

Therefore, the process must be auditable in the sense that certain questions must always be answered clearly: Who accessed the data (especially at the concentrator level)? Why did the algorithm make a particular decision? How exactly was the data processed? To this end, the technical and organisational parts of the process must generate traces (computer logs, activity reports, etc.) so that they can be audited. These logs should be stored in a dedicated tool, such as a log management solution. The logs provided by the equipment must be signed (using their own private key) and their signature stored by the server that centralises the logs to ensure a principle of non-repudiation. In addition, at each (usually daily) backup of the log server, these backups must also be signed by the log sink to ensure once again the principle of non-repudiation. Moreover, these logs should be analysed by a SOC and rules should be defined to detect such anomalies as many calls from a smart meter to the concentrator or an abnormal response. In addition, all processes must be regularly audited by the company's internal control teams to ensure compliance with the requirements. As far as privacy is concerned, these audits should at least ensure the lawfulness of the processing, its transparency, its application, the data minimisation, the respect of the storage limitation and the fairness aspect. These checks should be based in part on the logs collected.

This control could have an impact on privacy because the generated logs can contain personal data. This log management system should be included in the privacy impact assessment scope of the system. Moreover, the logging system should be carefully implemented to avoid unnecessary overhead due to the extra operations it executes.

4.17 PERFORM A PRIVACY IMPACT ASSESSMENT

Type	Associated Vulnerabilities	Threats it mitigates

Privacy	• Lack of controls to ensure that data is used only for the purposes defined • Lack of controls to ensure the adequacy of the purpose and its current use • Lack of detail on the purposes and justification for their legitimacy	• Diversion of purpose • Unlawful processing • Unfair processing • Lack of transparency

This privacy measure allows for an in-depth analysis of the impact of the processing on the privacy of users, compliance with the GDPR and to identify whether the associated risks are well addressed by the proposed privacy and security measures.

The privacy impact assessments and general accountability actions described above (in the internal audit process control) help ensure that the purpose of the processing is well defined (the prediction of residents' consumption) and that the actual use of the data remains within the scope of this case. To do so, the electricity supplier should use the requirements described below, and asses the vulnerabilities of the implemented system, to illustrate privacy threats and counter measures.

Such analysis may possibly impact the performance of the scenario if it results in one of the functionalities not allowing to minimise the data protection risks to the rights and freedoms of natural persons.

4.18 DEFINE AND IMPLEMENT A DATA RETENTION POLICY

Type	Associated Vulnerabilities	Threats it mitigates
Privacy	• Lack of data deletion mechanisms • Lack of data retention policy	No respect of storage limitation

This privacy control requires a storage duration value to be defined for each personal data involved in the processing, including individual detailed consumption data, individual daily consumption data etc. When this storage duration value is implemented, the privacy control acts to minimise the risk of keeping the data longer than it is strictly necessary. One way to implement it is by having a script that automatically deletes the data based on predefined criteria.

However, once all that data is deleted, it cannot be used later for forensic analysis, as may be required when a company faces a security breach.

4.19 STUDY ON DATA FIELDS NECESSITY AND JUSTIFICATION IN THE PRIVACY POLICY

Type	Associated Vulnerabilities	Threats it mitigates
Privacy	• Lack of justification for the collection of individual personal data collected • Lack of transparency on the purpose of the processing, the exact consumption data that are extracted, and how they are processed.	• Unlawful processing • Lack of transparency • No respect of data minimization

A lack of justification on how data are collected, correlated with a lack of transparency on the purpose of the processing or on the accuracy of the data collected could result in unlawful processing which would have moderate impact for the consumer (significant sense of invasion

of privacy). Therefore, the electricity supplier must study the necessity of different data fields and justify this in the privacy policy.

The personal data used is the consumption data (daily or hourly) of the inhabitants. The electricity supplier must therefore implement a privacy control in which they could explain why it is necessary to use this data specifically in the context of this scenario. The supplier must also formalise the explanation and make it freely available to consumers.

Nevertheless, this practice may have some impact on the relevance of the data collected. Taking time to justify the need for data may lead to the abandonment of the collection of some data for which justification is difficult, which in turn, may reduce the quality of the data at hand.

4.20 FORMALISE A LIA (LEGITIMATE INTEREST ASSESSMENT)

Type	Associated Vulnerabilities	Threats it mitigates
Privacy	Lack of legal basis related to legitimate interest	Unlawful processing

The Legitimate Interest Impact Assessment (LIA)[41] is used to determine if an organisation can process data using the legitimate interest lawful basis. In our case, the legal basis for the extraction of daily consumption data is legitimate interest. If the electricity supplier does not justify this legal basis, it faces unlawful processing. To justify the legal basis, the electricity supplier must put in place a LIA. In this LIA, the electricity supplier must formalise a reflection on how the data processing is necessary, and how it balances with the rights and freedoms of consumers.

The impact this could have on the electricity supplier would be a loss of time and energy spent formalising the LIA.

4.21 MINIMISE DATA AT EACH STEP OF THE PROCESSING; COLLECT ONLY WHAT IS NEEDED WHEN NEEDED

Type	Associated Vulnerabilities	Threats it mitigates
Privacy	• Lack of necessary data selection • Lack of verification that the data is adequate, relevant, and not excessive for the purpose of estimating electricity consumption	No respect of data minimisation

Minimisation requires an analysis to correctly identify the subset of fields which are required, and should be performed as early as possible during processing. Non-minimisation of data is the lack of selection of necessary, adequate, and relevant data.

For example, in our case, the electricity supplier only requires the consumption data of individuals. The electricity supplier extracts two types of consumption data: the daily consumption and the consumption at the half-hour scale. A study on the necessity of collecting data at the hourly scale must be done and a proof of its necessity made public. Then, the

[41] See https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/#:~:text=There's%20no%20defined%20process%2C%20but,(consider%20the%20individual's%20interests).

extracted individual data should be anonymised at the hub level and aggregated with other data from other households before being sent to the model.

However, data minimisation also involves data anonymisation. In this case, in the event of a security breach, it may be more difficult to perform a cybersecurity forensic analysis on anonymised data than on personally identifiable data. This has the impact of significantly slowing down the work of those employed to determine the cause of a security issue.

4.22 IMPLEMENT A PRIVACY BY DESIGN PROCESS

Type	Associated Vulnerabilities	Threats it mitigates
Privacy	Lack of Privacy by Design	• Unlawful processing • Unfair processing • Lack of transparency • Diversion of purpose • No respect of data minimization

The electricity supplier must ensure that a compliance study and privacy risk assessment is formalised in a Privacy Impact Analysis document, and that the identified action plan is implemented before the scenario is put into operation.

Conceiving a project following a Privacy be Design methodology requires less effort than adding privacy on top of an existing project. However, there could be an impact on the functionality of the scenario if the outcome of the Privacy by Design methodology leads to functionality not being implemented due to the privacy risk it generates (bear in mind that in our case, personal data is anonymised and then aggregated and deleted at the concentrator). Another impact of note is related to security, as it may not be possible to collect data useful for forensic analysis following a cyber-attack due to privacy limitations.

4.23 RAISE AWARENESS OF SECURITY AND PRIVACY ISSUES AMONG ALL STAKEHOLDERS

Type	Associated Vulnerabilities	Threats it mitigates
Cybersecurity & privacy	Lack of consideration of attacks to which the energy forecasting system could be exposed	• Poisoning • Data disclosure • Unlawful processing • Unfair processing • Lack of transparency • Diversion of purpose • No respect of data minimization

There are many threats with high impact in this scenario, and the company must ensure that such threats are comprehended by the teams. For instance, the company must ensure that all its teams (including Data Scientists) are trained in the specificities of machine learning and the new associated cyber and privacy risks.

This can take the form of regular training sessions or local cyber and privacy risks reporting. This raises awareness of issues to teams, and in some cases can present solutions tailored to the specificities of forecasting demand on electricity grids scenario.

This control does not directly impact performance of the system.

4.24 SUMMARY

The following table summarises every control described in the previous section. Each control is associated with vulnerabilities, mitigated threats, and addressed privacy and security requirements.

Figure 18: Summary of controls and mitigated threats

Control name and type	Associated Vulnerabilities	Threat mitigated	Privacy and security requirements addressed
Implement a Security by Design process	Lack of Security by Design	Compromise of ML application components Poisoning Human error Data disclosure	Integrity of the data Availability of the data Confidentiality of the data Traceability of the data
Document the Electrical forecast system	Lack of Electrical forecast system documentation	Human error Unlawful processing Unfair processing Lack of transparency Diversion of purpose No respect of data minimization No respect of storage limitation	Integrity of the data Availability of the data Confidentiality of the data Traceability of the data Lawfulness of the process Fairness of the process Transparency of the process Purpose limitation of the process Data minimization of the process Accuracy of the data Storage limitation of the data
Check the vulnerabilities of the components used and Implement processes to maintain security levels of ML components over time	Use of vulnerable components Lack of security process to maintain a good security level of the components of the Electrical forecast system	Compromise of ML application components Data disclosure	Integrity of the data Availability of the data Confidentiality of the data Traceability of the data
Choose and define a more resilient model design	Model easy to poison	Poisoning	Integrity of the data
Integrate poisoning control in the training dataset	No detection of poisoned samples in the training dataset Lack of control for poisoning	Poisoning	Integrity of the data
Enlarge the training dataset	Lack of data for increasing robustness to poisoning	Poisoning	Integrity of the data
Secure the transit of the collected data	Poor data management	Poisoning	Integrity of the data
Control all data used by the ML Model	Use of uncontrolled data	Poisoning	Integrity of the data
Ensure reliable sources are used	Use of unsafe data or models (e.g., with transfer learning)	Poisoning	Integrity of the data

Implement access right management process	Poor access rights management	Compromise of ML application components Poisoning Data disclosure Human error	Integrity of the data Availability of the data Confidentiality of the data Traceability of the data
Ensure all systems and devices comply with authentication, and access control policies	Weak access protection mechanisms for machine learning model components Weak access protection mechanisms of the personal data (encryption, access control mechanism…)	Compromise of ML application components Poisoning Data disclosure Human error	Integrity of the data Availability of the data Confidentiality of the data Traceability of the data
Reduce the available information about the model	Too much information available on the model	Compromise of ML application components	Confidentiality of data
Identify a data controller for the energy consumption anticipation data processing	Absence of an identified data controller	Unlawful processing Unfair processing Lack of transparency	Lawfulness of the process Fairness of the process Transparency of the process
Properly collect and maintain user consent when needed for detailed energy consumption usage	Absence of mechanisms to ensure that the processing of consumer electricity affected by consent cannot be carried out without consent Lack of legal basis related to user's consent when their detailed consumptions data (per hour or half hour) are processed or that the legitimate interest related to the daily processing of the data is not properly justified or that no justification is provided at all Lack of practical means and justification for obtaining the consents of the electricity consumers concerned	Unlawful processing	Lawfulness of the process
Anonymise data coming from the concentrator	Lack of anonymization	No respect of data minimization	Confidentiality of the data Data minimization of the process
Generate Log generation and perform Internal audit process	Lack of auditability of processing Lack of traceability of actions and/or modifications made to the assets	Unlawful processing Unfair processing Lack of transparency Diversion of purpose No respect of data minimisation Compromise of Machine Learning application No respect of storage limitation	Lawfulness of the process Fairness of the process Transparency of the process Purpose limitation of the process Data minimisation of the process Accuracy of the data Storage limitation of the data
Perform a privacy Impact Assessment	Lack of controls to ensure that data is used only for the purposes defined Lack of controls to ensure the adequacy of the purpose and its current use	Diversion of purpose Unlawful processing Unfair processing Lack of transparency	Purpose limitation of the process Lawfulness of the process Data minimisation of the process Transparency of the process

	Lack of detail on the purposes and justification for their legitimacy		
Define and implement a data retention policy	Lack of data deletion mechanisms Lack of data retention policy	No respect of storage limitation	Storage limitation of the data
Study on data fields necessity and justification in the privacy policy	Lack of justification for the collection of individual personal data collected Lack of transparency on the purpose of the processing, the exact consumption data that are extracted, and how they are processed.	Unlawful processing Lack of transparency No respect of data minimisation	Lawfulness of the process Data minimization of the process Transparency of the process
Formalize a LIA (Legitimate Interest Assessment)	Lack of legal basis related to legitimate interest	Unlawful processing	Lawfulness of the process
Minimise data at each step of the processing; collect only what is needed when needed	Lack of necessary data selection Lack of verification that the data is adequate, relevant, and not excessive for the purpose of estimating electricity consumption	No respect of data minimisation	Data minimisation of the process
Implement a Privacy by Design process	Lack of Privacy by Design	Unlawful processing Unfair processing Lack of transparency Diversion of purpose No respect of data minimisation	Lawfulness of the process Fairness of the process Transparency of the process Purpose limitation of the process Data minimisation of the process
Raise awareness of security and privacy issues among all stakeholders	Lack of consideration of attacks to which home assistant could be exposed	Poisoning Data disclosure Unlawful processing Unfair processing Lack of transparency Diversion of purpose No respect of data minimisation	Integrity of the data Availability of the data Confidentiality of the data Traceability of the data Lawfulness of the process Fairness of the process Transparency of the process Purpose limitation of the process Data minimisation of the process Accuracy of the data

5. CONCLUSION

In this report, analysis of **forecasting demand on electricity grids** supported by Artificial Intelligence (AI) is presented. AI is often used as an umbrella term that encompasses the technology behind many smart solutions and devices. It is a constantly evolving field where new innovations appear regularly in different areas of activity.

Even though AI can be very beneficial for the industries areas it applies to, it can also have quite a significant impact for security and privacy, especially when these systems have sensitive functionalities, as for the scenario that is described in this report. Indeed, AI comes with a wide range of privacy and security vulnerabilities causing threats with heavy impacts for organisations.

Regardless how AI is being used in support of business functionality, it should not be a surprise that the many identified threats are similar. However, we must be aware of the fact that depending on the context of the scenario, the same threats apply differently and have different levels of impact. Regarding the impact of each threat, it must be also noted that even in the same scenario every instance is unique, and a proper study must be carried out by each company using AI to maintain a proper security and privacy level.

This guide helps in the identification and evaluation of threats in a specific business scenario which uses AI, but it is important to remember that while the analyzed scenario is based on real life examples, it includes assumptions which may not match the business context in which other organizations would like to implement in. Therefore, the entire cybersecurity and privacy context (requirements, threats, vulnerabilities, and controls) must be adapted to the context and reality of the individual business. In addition, the controls proposed in this document are not sufficient on their own, and must be complemented by the standard security measures that already exist.

While security and privacy are not necessarily the same, they are intimately related, and equally important. In their management, a balance must be found between the two in the sense that one must always make sure that the regulations and recommendations concerning the two aspects are always respected. Unfortunately, these two parameters are regularly at the expense of performance. It is therefore an equation with three variables, two of which respond to the need for regulation and risk, which need to be correctly balanced to achieve the desired effect.

ANNEX I: SECURITY AND PRIVACY SCALES AND REQUIREMENTS

A.1 CYBERSECURITY AND PRIVACY SEVERITY SCALES

Availability	
Level	Definition
Low	Service provided must be restored within **few days or less**.
Moderate	Service provided must be restored within **a day or less**.
High	Service provided must be restored within **half a day or less**.
Critical	Service provided must be restored within **few hours or less**.

Integrity	
Level	Definition
Low	A loss of integrity in the process **does not need to be identified or corrected**.
Moderate	Any degradation in the process **must be identified but not necessarily corrected**.
High	Any degradation in the process **must be identified and corrected**.
Critical	**No degradation** in the process is tolerated at any time.

Confidentiality	
Level	Definition
Low	Process-related data can be **accessed by everyone**.
Moderate	Access to process-related data is **restricted to internal staff and trusted partners**.
High	Access to process-related data is **restricted to employees having an organisation or functional link with the process**.
Critical	Access to process-related data is **restricted to a very limited number of individuals**.

Traceability	
Level	Definition
Low	The **absence of traces** of actions on the service provided **is acceptable**.
Moderate	**Actions** related to the service provided must be **identified**. They must be traced and detected.
High	The **actions** related to the process and their **actors** must be **identified and dated**. They must be **imputable**.
Critical	Service provided **actions** must be **legally enforceable** and **time stamped**. They must have a **probative value**.

A.2 CYBERSECURITY SCALE OF IMPACT

Severity[42]	
Level	Definition
1 - Low	No operational impact on business performance or on the safety of people and property. The company/entity will overcome the situation without too much difficulty.
2 - Moderate	Degradation of business performance without impact on safety of people and property. The company/entity will overcome the situation despite some difficulties (operation in degraded mode).
3 - High	Severe deterioration in the performance of the business, with possible significant impacts on the safety of people and property. The company/entity will overcome the situation with serious difficulties (operation in very degraded mode).
4 - Critical	Inability of the company/entity to carry out all or part of its business, with possible serious impacts on the safety of people and property. The company/entity is unlikely to overcome the situation (its survival is threatened).

A.3 PRIVACY SCALE OF IMPACT

Severity[43]	
Level	Definition
1 - Low	The persons concerned affected will not be affected or may experience some inconvenience, which they overcome without difficulty.
2 - Moderate	The persons concerned may experience significant inconvenience significant inconvenience, which they be able to overcome despite some difficulties.
3 - High	The persons concerned may experience significant consequences, which they should be able be able to overcome, but with real and significant difficulties.
4 - Critical	The people concerned may experience significant consequences, if not irreparable irremediable consequences, that they may not overcome.

[42] Based on « Agence Nationale de la Securité des Systèmes d'Information » (ANSSI). See: https://www.ssi.gouv.fr/uploads/2019/11/anssi-guide-ebios_risk_manager-en-v1.0.pdf
[43] Metrics based on the National Commission on Informatics and Liberty (CNIL) -an independent French Administrative regulatory body. See: https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-3-en-knowledgebases.pdf

A.4 PRIVACY REQUIREMENTS CRITERIA

Regarding the privacy, the applied requirements are based on the GDPR data protection principles. The figure below summarises such requirements.

Requirements	Article
Lawfulness, fairness, and transparency	Art. 5.a
Purpose limitation	Art. 5.b
Data minimisation	Art. 5.c
Accuracy	Art. 5.d
Storage limitation	Art. 5.e
Security of personal data (integrity and confidentiality)	Art. 5.f

Moreover, as a complement to the GDPR requirements listed above, some key privacy topics related to AI systems need to be addressed (some of them are specifically mentioned by national data protection supervisory authorities[44]). These additional privacy requirements are listed below:

Recommendations	Details
Database creation	Projects must ensure the compliance of the data collected and injected into the database for training and in production. Moreover, they must ensure that only legitimate persons have access to the data in the database, prevent data loss and hide personal data.
Compliance of the training model (i.e., before production)	The data collected and the training model must be processed in accordance with the state of the art of Machine Learning development to guarantee control of all processes carried out. This consists, for instance, of good justifications for the chosen learning method, the reliability of the third-party tools used, paying attention to open-source data, conducting a meticulous training protocol based on the state of the art, and finally verifying the quality of the system once in the experimentation phase.[45]

[44] Such as the CNIL. It proposes an analysis grid to enable organizations to assess the maturity of their artificial intelligence systems regarding the GDPR available on https://www.cnil.fr/fr/intelligence-artificielle/guide.
[45] see the various CNIL method sheets on this subject, such as this one: https://www.cnil.fr/fr/intelligence-artificielle/guide/developper-et-entrainer-un-algorithme

TP-09-23-010-EN-N

ABOUT ENISA

The European Union Agency for Cybersecurity, ENISA, is the Union's agency dedicated to achieving a high common level of cybersecurity across Europe. Established in 2004 and strengthened by the EU Cybersecurity Act, the European Union Agency for Cybersecurity contributes to EU cyber policy, enhances the trustworthiness of ICT products, services and processes with cybersecurity certification schemes, cooperates with Member States and EU bodies, and helps Europe prepare for the cyber challenges of tomorrow. Through knowledge sharing, capacity building and awareness raising, the Agency works together with its key stakeholders to strengthen trust in the connected economy, to boost resilience of the Union's infrastructure, and, ultimately, to keep Europe's society and citizens digitally secure. More information about ENISA and its work can be found here: www.enisa.europa.eu.

ENISA
European Union Agency for Cybersecurity

Athens Office
Agamemnonos 14, Chalandri 15231, Attiki, Greece

Heraklion Office
95 Nikolaou Plastira
700 13 Vassilika Vouton, Heraklion, Greece

enisa.europa.eu

ISBN 978-92-9204-620-0
doi 10.2824/92851